1975 Metropolitan Appendix B: Source and Accuracy

Table of Contents

Atlanta 3

Chicago 12

Cincinnati 21

Columbus 30

Hartford 39

Kansas City 48

Madison 57

Miami 66

Milwaukee 75

New Orleans 84

Newport News-Hampton 93

Patterson-Clifton-Passaic 102

Philadelphia 111

Portland 121

Riverside-San Bernadino-Ontario 130

Rochester..... 139

San Antonio 148

San Diego 157

San Francisco- Oakland 166

Springfield-Chicopee-Holyoke 175

Atlanta 1975

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-41
Annual Housing Survey App-41
Selection of the sample. App-41
Building loss sample selection . . . App-42
1970 Census of Population and Housing. App-43
ESTIMATION. App-43
1975 housing inventory App-43
1970-1975 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing. App-44
RELIABILITY OF THE ESTIMATES. App-44
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-45
Rounding errors. App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences. App-47
Illustration of the computation of the standard error of a difference. App-48
Medians App-48
Illustration of the computation of the 95-percent confidence interval of a median App-48

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 21 SMSA's are based on data collected from the 1975 Annual Housing Survey (AHS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development. In each of the 21 SMSA's, the data were collected for the 12-month period from April 1975 through March 1976, with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and a third group of 20 SMSA's were collected from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated for the AHS every three years on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the second group (Year II SMSA's) are: Atlanta, Ga., Chicago, Ill., Philadelphia, Pa.-N.J., and San Francisco-Oakland, Calif.

The remaining SMSA's in the second group are: Cincinnati, Ohio-Ky.-Ind., Colorado Springs, Colo., Columbus, Ohio, Hartford, Conn., Kansas City, Mo.-Kans., Miami, Fla., Milwaukee, Wis., New Orleans, La., Newport News-Hampton, Va., Paterson-Clifton-Passaic, N.J., Portland, Oreg.-Wash., Rochester, N.Y., San Antonio, Tex., San Bernardino-Riverside-Ontario, Calif., San Diego, Calif., Springfield-Chicopee-Holyoke, Mass.-Conn., and Madison, Wis.

In this SMSA, 13,518 units were eligible for interview. Of these sample units, 570 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview 1,484 units were visited but found not to be eligible for interview, because these units were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit issuing was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for those SMSA's which are *not* 100-percent permit issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). The following table indicates which SMSA's are 100-percent permit issuing and which contain a sample from the nonpermit universe.

100 Percent permit-issuing SMSA's	SMSA's with a sample from the nonpermit universe
Hartford, Conn.	Atlanta, Ga.
Miami, Fla.	Chicago, Ill.
Newport News-Hampton, Va.	Cincinnati, Ohio-Ky.-Ind.
Paterson-Clifton-Passaic, N.J.	Colorado Springs, Colo.
San Bernardino-Riverside-Ontario, Calif.	Columbus, Ohio
San Diego, Calif.	Kansas City, Mo.-Kans.
San Francisco-Oakland, Calif.	Madison, Wis.
	Milwaukee, Wis.
	New Orleans, La.
	Philadelphia, Pa.-N.J.
	Portland, Oreg.-Wash.
	Rochester, N.Y.
	San Antonio, Tex.
	Springfield-Chicopee-Holyoke, Mass.-Conn.

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of these sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and the balance of the SMSA, since the

sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—family size	Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000—$5,999 . .		
$6,000—$9,999 . .		
$10,000—$14,999		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. The housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued, and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \frac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Some tables in this report provide estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at enumeration time (i.e., the 1975 housing inventory), and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1975 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1975 housing inventory.—The AHS estimates of characteristics of the 1975 housing inventory employed a three-stage ratio estimation procedure. However, the third-stage ratio estimation procedure was employed in only three SMSA's. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 570 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units and one noninterview cell for the nonpermit universe (if applicable). Sample housing units from the nonpermit universe identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce the variation in sample size for strata used in the sample selection of the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes built after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or by other sources). This procedure was designed to adjust the AHS sample estimates of these types of units to independently derived current estimates. These independent estimates were considered to be the best estimates available. The adjustment was necessary to correct known deficiencies in the AHS sample in the representation of these units (see section on nonsampling error). The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1975 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimates of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units	
New mobile homes	
"Other additions"	

The numerators of the ratios were derived by applying the following factors:

1. For the "conventional new construction units" cell, either a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in the SMSA, or the missed conventional new construction was estimated from the Survey of Construction (SOC). (The table that follows indicates the procedure used to estimate the missing conventional new construction for each SMSA.)
2. For the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied.
3. For the "other additions" cell, the "other additions" rate from SCARF[1] was applied.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

SMSA's employing the SOC estimate of missed conventional new construction	SMSA's employing the national trend for missed conventional new construction
Atlanta, Ga.	Colorado Springs, Colo.
Chicago, Ill.	Hartford, Conn.
Cincinnati, Ohio-Ky.-Ind.	Madison, Wis.
Columbus, Ohio	New Orleans, La.
Kansas City, Mo.-Kans.	Newport News—Hampton, Va.
Miami, Fla.	Paterson-Clifton-Passaic, N.J.
Milwaukee, Wis.	Rochester, N.Y.
Philadelphia, Pa.-N.J.	San Antonio, Tex.
Portland, Oreg.-Wash.	San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.	Springfield-Chicopee-Holyoke, Mass.-Conn.
San Francisco-Oakland, Calif.	

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

For some Year II SMSA's (i.e., Colorado Springs, Colo., Miami, Fla., and San Diego, Calif.) a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1975 housing inventory to an independent estimate of the SMSA's October 1975 housing inventory. This ratio estimate factor equaled the following:

$$\frac{\text{Independent estimate of the October 1975 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using the 1970 census estimate of total housing units for the SMSA in conjunction with the estimate of the change in the housing inventory since the census. The estimate of change was based on either administrative records from utility companies, or, when utility data were not available, on estimates of new construction permits and post-census demolition data. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, are brought into agreement with known estimates of the SMSA housing population.

The third-stage ratio estimation procedure was not employed in all Year II SMSA's since the reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. The independent estimates were only employed in SMSA's where the estimated relative bias was thought to be low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results. A measure of the relative bias of the independent estimate was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1975 independent estimates. In some SMSA's, the measure of the relative bias of the independent estimate was less than 3 percent but the independent estimate was still not employed, since there was evidence that suggested that the procedure for generating the independent estimates was producing substantial overestimates for these SMSA's.

1970-1975 lost units.—The AHS estimate of characteristics of the 1970-1975 lost units employed a one-stage ratio estimation procedure, which is similar to the first-stage ratio estimation procedure mentioned above. The 1970-1975 lost units' file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for State, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and non-

sampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

Nonsampling errors.—Nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1975 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors associated with 1970 census estimates—"coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

AHS-SMSA.—For the 1975 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are not yet available, however, it is believed that they will be similar to the results of the Year I AHS-SMSA reinterview study which were presented in the following Census Bureau memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample: 1974."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year I SMSA's and not for any specific SMSA):

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. The estimated indexes of inconsistency for most of the reinterview items were in the 20 to 50 range. The 20-50 interval is considered moderate indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview program because they had not been previously tested.

3. "Our bias indicator, the net difference rate, shows no signs of serious bias. For only one characteristic was there as much as a 2½ percent probability that the net difference rate is greater than 10 percent."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors.—With errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies in the representation of both conventional new construction and new mobile homes (and trailers) in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later, were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and

therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 10,100 conventional new construction units in this SMSA had permits issued prior to January 1970, and therefore, were missed by the 1975 AHS-SMSA survey.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. Although it is not known exactly, an estimated 5,200 new mobile homes in permit-issuing areas were missed by the 1975 AHS-SMSA survey in this SMSA. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS data still exists.

Rounding errors.—The rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling error, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in tables I and II are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1975 Housing Inventory and for Estimated Number of 1970-1975 Lost Units for the Atlanta, Ga., SMSA, for the Central City and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central city	Not in central city
200.	100	80	120
500.	160	120	190
700.	190	140	220
1,000 . . .	220	170	260
2,500 . . .	350	270	420
5,000 . . .	500	380	590
10,000 . .	710	530	830
25,000 . .	1,110	800	1,300
50,000 . .	1,560	1,060	1,810
100,000. .	2,170	1,260	2,480
250,000. .	3,280	—	3,530
400,000. .	3,950	—	3,910
500,000. .	4,260	—	—

Tables I through IV present the standard errors applicable to estimates of characteristics of the 1975 housing in-

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the Atlanta, Ga., SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
200	5.0	10.9	15.0	21.6	25.0
500	3.1	6.9	9.5	13.7	15.8
700	2.7	5.8	8.0	11.6	13.4
1,000	2.2	4.9	6.7	9.7	11.2
2,500	1.4	3.1	4.2	6.1	7.1
5,000	1.0	2.2	3.0	4.3	5.0
10,000	.7	1.5	2.1	3.1	3.5
25,000	.4	1.0	1.3	1.0	2.2
50,000	.3	.7	.9	1.4	1.6
100,000	.2	.5	.7	1.0	1.1
250,000	.14	.3	.4	.6	.7
400,000	.11	.2	.3	.5	.6
500,000	.10	.2	.3	.4	.5

ventory as well as estimates of characteristics of the 1970-1975 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in these tables.

For ratios, (100) (x/y), where x is not a subclass of y, tables II through IV underestimate the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be obtained by letting the standard error of the ratio be approximately equal to:

$$(100)\,(x/y)\,\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

Let: x = the numerator of the ratio
y = the denominator of the ratio
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

Illustration of the use of the standard error tables.—Table A-1 of part A of this report shows that in 1975 there were 295,300 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 3,480. Consequently, the 68-percent confidence interval, as shown by these data, is from 291,820 to 298,780 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 289,730 to 300,870 housing units with 90 percent confidence; and that the average estimate lies within the interval from 288,340 to 302,260 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 295,300 owner-occupied housing units 68,900, or 23.3 percent, had two bedrooms. Interpolation in table II of this appendix shows that the standard error of the percent is approximately .5 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 22.8 to 23.8 percent; the 90-percent confidence interval is from 22.5 to 24.1 percent; and the 95-percent confidence interval is from 22.3 to 24.3 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; while, if there is a high negative correlation, the formula will underestimate the true standard error.

TABLE III. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units in the Central City

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
200	3.8	8.3	11.4	16.5	19.1
500	2.4	5.3	7.2	10.5	12.1
700	2.0	4.4	6.1	8.8	10.2
1,000	1.7	3.7	5.1	7.4	8.5
2,500	1.1	2.4	3.2	4.7	5.4
5,000	.8	1.7	2.3	3.3	3.8
10,000	.5	1.2	1.6	2.3	2.7
25,000	.3	.7	1.0	1.5	1.7
50,000	.2	.5	.7	1.0	1.2
100,000	.2	.4	.5	.7	.9

TABLE IV. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the Balance of the SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
200	5.9	12.8	17.7	25.5	29.4
500	3.7	8.1	11.2	16.1	18.6
700	3.1	6.9	9.4	13.6	15.7
1,000	2.6	5.7	7.9	11.4	13.2
2,500	1.7	3.6	5.0	7.2	8.3
5,000	1.2	2.6	3.5	5.1	5.9
10,000	.8	1.8	2.5	3.6	4.2
25,000	.5	1.1	1.6	2.3	2.6
50,000	.4	.8	1.1	1.6	1.9
100,000	.3	.6	.8	1.1	1.3
250,000	.2	.4	.5	.7	.8
400,000	.13	.3	.4	.6	.7

Illustration of the computation of the standard error of a difference.—Table A-1 of part A of this report shows that in 1975 there were 155,600 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 86,700. Table I shows the standard error of 68,900 is approximately 1,790 and the standard error of 155,600 is approximately 2,580. Therefore, the standard error of the estimated difference of 86,700 is about

$$3{,}140 = \sqrt{(1{,}790)^2 + (2{,}580)^2}$$

Consequently, the 68-percent confidence interval for the 86,700 difference is from 83,560 to 89,840 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 81,680 to 91,720 housing units, and the 95-percent confidence interval is from 80,420 to 92,980. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval does not include zero or negative values.

Medians.—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From table II determine the standard error of a 50-percent characteristic on the base of the median,
2. add to and subtract from 50 percent the standard error determined in step 1; and
3. using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.1. The base of the distribution from which this median was determined is 295,300 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 295,300 is approximately .7 percentage points.
2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 48.6 and 51.4.
3. From table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 109,200 owner-occupied housing units or 37.0 percent, had one or two persons (for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 64,000 owner-occupied housing units, or 21.7 percent, had three persons (i.e., 2.5 to 3.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{48.6-37.0}{21.7}\right) = 3.0$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{51.4-37.0}{21.7}\right) = 3.2$$

Thus, the 95-percent confidence interval ranges from 3.0 to 3.2 persons.

Chicago 1975

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-41
Annual Housing Survey App-41
Selection of the sample. App-41
Building loss sample selection . . . App-42
1970 Census of Population and Housing. App-43
ESTIMATION. App-43
1975 housing inventory App-43
1970-1975 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing. App-44
RELIABILITY OF THE ESTIMATES. App-44
Nonsampling errors. App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-45
Rounding errors. App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences. App-47
Illustration of the computation of the standard error of a difference. App-47
Medians App-48
Illustration of the computation of the 95-percent confidence interval of a median App-48

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 21 SMSA's are based on data collected from the 1975 Annual Housing Survey (AHS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development. In each of the 21 SMSA's, the data were collected for the 12-month period from April 1975 through March 1976, with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and a third group of 20 SMSA's were collected from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated for the AHS every three years on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the second group (Year II SMSA's) are: Atlanta, Ga., Chicago, Ill., Philadelphia, Pa.-N.J., and San Francisco-Oakland, Calif.

The remaining SMSA's in the second group are: Cincinnati, Ohio-Ky.-Ind., Colorado Springs, Colo., Columbus, Ohio, Hartford, Conn., Kansas City, Mo.-Kans., Miami, Fla., Milwaukee, Wis., New Orleans, La., Newport News-Hampton, Va., Paterson-Clifton-Passaic, N.J., Portland, Oreg.-Wash., Rochester, N.Y., San Antonio, Tex., San Bernardino-Riverside-Ontario, Calif., San Diego, Calif., Springfield-Chicopee-Holyoke, Mass.-Conn., and Madison, Wis.

In this SMSA, 14,442 units were eligible for interview. Of these sample units, 741 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 1,051 units were visited but found not to be eligible for interview, because these units were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit issuing was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for those SMSA's which are *not* 100-percent permit issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). The following table indicates which SMSA's are 100-percent permit issuing and which contain a sample from the nonpermit universe.

100 Percent permit-issuing SMSA's	SMSA's with a sample from the nonpermit universe
Hartford, Conn.	Atlanta, Ga.
Miami, Fla.	Chicago, Ill.
Newport News-Hampton, Va.	Cincinnati, Ohio-Ky.-Ind.
Paterson-Clifton-Passaic, N.J.	Colorado Springs, Colo.
San Bernardino-Riverside-Ontario, Calif.	Columbus, Ohio
San Diego, Calif.	Kansas City, Mo.-Kans.
San Francisco-Oakland, Calif.	Madison, Wis.
	Milwaukee, Wis.
	New Orleans, La.
	Philadelphia, Pa.-N.J.
	Portland, Oreg.-Wash.
	Rochester, N.Y.
	San Antonio, Tex.
	Springfield-Chicopee-Holyoke, Mass.-Conn.

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of these sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and the balance of the SMSA, since the

sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—family size	Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000—$5,999 . .		
$6,000—$9,999 . .		
$10,000—$14,999		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. The housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued, and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \frac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Some tables in this report provide estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at enumeration time (i.e., the 1975 housing inventory), and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1975 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1975 housing inventory.—The AHS estimates of characteristics of the 1975 housing inventory employed a three-stage ratio estimation procedure. However, the third-stage ratio estimation procedure was employed in only three SMSA's. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 741 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units and one noninterview cell for the nonpermit universe (if applicable). Sample housing units from the nonpermit universe identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce the variation in sample size for strata used in the sample selection of the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes built after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or by other sources). This procedure was designed to adjust the AHS sample estimates of these types of units to independently derived current estimates. These independent estimates were considered to be the best estimates available. The adjustment was necessary to correct known deficiencies in the AHS sample in the representation of these units (see section on nonsampling error). The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1975 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimates of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units	
New mobile homes	
"Other additions"	

The numerators of the ratios were derived by applying the following factors:

1. For the "conventional new construction units" cell, either a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in the SMSA, or the missed conventional new construction was estimated from the Survey of Construction (SOC). (The table that follows indicates the procedure used to estimate the missing conventional new construction for each SMSA.)
2. For the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied.
3. For the "other additions" cell, the "other additions" rate from SCARF[1] was applied.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

SMSA's employing the SOC estimate of missed conventional new construction	SMSA's employing the national trend for missed conventional new construction
Atlanta, Ga.	Colorado Springs, Colo.
Chicago, Ill.	Hartford, Conn.
Cincinnati, Ohio-Ky.-Ind.	Madison, Wis.
Columbus, Ohio	New Orleans, La.
Kansas City, Mo.-Kans.	Newport News—Hampton, Va.
Miami, Fla.	Paterson-Clifton-Passaic, N.J.
Milwaukee, Wis.	Rochester, N.Y.
Philadelphia, Pa.-N.J.	San Antonio, Tex.
Portland, Oreg.-Wash.	San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.	Springfield-Chicopee-Holyoke, Mass.-Conn.
San Francisco-Oakland, Calif.	

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

For some Year II SMSA's (i.e., Colorado Springs, Colo., Miami, Fla., and San Diego, Calif.) a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1975 housing inventory to an independent estimate of the SMSA's October 1975 housing inventory. This ratio estimate factor equaled the following:

$$\frac{\text{Independent estimate of the October 1975 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using the 1970 census estimate of total housing units for the SMSA in conjunction with the estimate of the change in the housing inventory since the census. The estimate of change was based on either administrative records from utility companies, or, when utility data were not available, on estimates of new construction permits and post-census demolition data. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, are brought into agreement with known estimates of the SMSA housing population.

The third-stage ratio estimation procedure was not employed in all Year II SMSA's since the reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. The independent estimates were only employed in SMSA's where the estimated relative bias was thought to be low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results. A measure of the relative bias of the independent estimate was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1975 independent estimates. In some SMSA's, the measure of the relative bias of the independent estimate was less than 3 percent but the independent estimate was still not employed, since there was evidence that suggested that the procedure for generating the independent estimates was producing substantial overestimates for these SMSA's.

1970-1975 lost units.—The AHS estimate of characteristics of the 1970-1975 lost units employed a one-stage ratio estimation procedure which is similar to the first-stage ratio estimation procedure mentioned above. The 1970-1975 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for State, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and non-

sampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

Nonsampling errors.—Nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1975 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors associated with 1970 census estimates—"coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

AHS-SMSA.—For the 1975 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are not yet available, however, it is believed that they will be similar to the results of the Year I AHS-SMSA reinterview study which were presented in the following Census Bureau memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample: 1974."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year I SMSA's and not for any specific SMSA):

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. The estimated indexes of inconsistency for most of the reinterview items were in the 20 to 50 range. The 20-50 interval is considered moderate indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview program because they had not been previously tested.

3. "Our bias indicator, the net difference rate, shows no signs of serious bias. For only one characteristic was there as much as a 2½ percent probability that the net difference rate is greater than 10 percent."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors.—With errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies in the representation of both conventional new construction and new mobile homes (and trailers) in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later, were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and

therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 34,400 conventional new construction units in this SMSA had permits issued prior to January 1970, and therefore, were missed by the 1975 AHS-SMSA survey.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. Although it is not known exactly, an estimated 4,800 new mobile homes in permit-issuing areas were missed by the 1975 AHS-SMSA survey in this SMSA. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS data still exists.

Rounding errors.—The rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling error, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in tables I through IV are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1975 Housing Inventory for the Chicago, Ill., SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central city	Not in central city
200	200	200	200
500	310	310	320
700	370	370	370
1,000 . . .	440	440	450
2,500 . . .	700	700	710
5,000 . . .	990	980	1,000
10,000 . .	1,400	1,390	1,410
25,000 . .	2,210	2,180	2,200
50,000 . .	3,130	3,060	3,130
100,000 .	4,400	4,240	4,400
250,000 .	6,880	6,300	6,810
400,000 .	8,590	7,430	8,420
500,000 .	9,530	7,870	9,260
1,000,000	12,890	7,360	12,000
1,500,000	15,030	—	—
2,000,000	16,440	—	—

Tables I through IV present the standard errors applicable to estimates of char-

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory for the Chicago, Ill., SMSA, for the Central City, and for the Balance of the SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
200	10.1	22.0	30.3	43.8	50.6
500	6.4	13.9	19.2	27.7	32.0
700	5.4	11.8	16.2	23.4	27.0
1,000	4.5	9.9	13.6	19.6	22.6
2,500	2.8	6.2	8.6	12.4	14.3
5,000	2.0	4.4	6.1	8.8	10.1
10,000	1.4	3.1	4.3	6.2	7.2
25,000	.9	2.0	2.7	3.9	4.5
50,000	.6	1.4	1.9	2.8	3.2
100,000	.4	1.0	1.4	2.0	2.3
250,000	.3	.6	.9	1.2	1.4
400,000	.2	.5	.7	1.0	1.1
500,000	.2	.4	.6	.9	1.0
1,000,000	.14	.3	.4	.6	.7
1,500,000	.12	.3	.4	.5	.6
2,000,000	.10	.2	.3	.4	.5

acteristics of the 1975 housing inventory as well as estimates of characteristics of the 1970-1975 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in these tables.

For ratios, (100) (x/y), where x is not a subclass of y, tables II and IV underestimate the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be obtained by letting the standard error of the ratio be approximately equal to:

$$(100)\,(x/y)\,\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

Let: x = the numerator of the ratio

y = the denominator of the ratio

σ_x = the standard error of the numerator

σ_y = the standard error of the denominator

Illustration of the use of the standard error tables.—Table A-1 of part A of this report shows that in 1975 there were 1,269,600 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 14,040. Consequently, the 68-percent confidence interval, as shown by these data, is from 1,255,560 to 1,283,640 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 1,247,140 to 1,292,060 housing units with 90 percent confidence; and that the average estimate lies within the interval from 1,241,520 to 1,297,680 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 1,269,600 owner-occupied housing units 344,600, or 27.1 percent, had two bedrooms. Interpolation in table II of this appendix shows that the standard error of the percent is approximately 0.6 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 26.5 to 27.7 percent; the 90-percent confidence interval is from 26.1 to 28.1 percent; and the 95-percent confidence interval is from 25.9 to 28.3 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; while, if there is a high negative correlation, the formula will underestimate the true standard error.

TABLE III. Standard Errors for Estimated Number of 1970-1975 Lost Units for the Chicago, Ill., SMSA, for the Central City, and for the Balance of the SMSA

Size of estimate	Standard error		
	SMSA	In central city	Not in central city
100	120	120	120
200	180	170	170
500	280	270	280
700	330	330	330
1,000	390	390	400
2,500	630	620	650
5,000	890	880	960
10,000	1,270	1,260	1,480
25,000	2,070	2,060	—
50,000	3,060	3,080	—
100,000	4,690	—	—

Illustration of the computation of the standard error of a difference.—Table A-1 of part A of this report shows that in 1975 there were 653,800 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 309,200. Table I shows the standard error

TABLE IV. Standard Errors for Estimated Percentages of 1970-1975 Lost Units for the Chicago, Ill., SMSA, for the Central City, and for the Balance of the SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
100	12.4	27.1	37.3	53.9	62.2
200	8.8	19.2	26.4	38.1	44.0
500	5.5	12.1	16.7	24.1	27.8
700	4.7	10.3	14.1	20.4	23.5
1,000	3.9	8.6	11.8	17.0	19.7
2,500	2.5	5.4	7.5	10.8	12.4
5,000	1.8	3.8	5.3	7.6	8.8
10,000	1.2	2.7	3.7	5.4	6.2
25,000	.8	1.7	2.4	3.4	3.9
50,000	.6	1.2	1.7	2.4	2.8
100,000	.4	.9	1.2	1.7	2.0

of 344,600 is approximately 7,960 and the standard error of 653,800 is approximately 10,560. Therefore, the standard error of the estimated difference of 309,200 is about

$$13{,}220 = \sqrt{(7{,}960)^2 + (10{,}560)^2}$$

Consequently, the 68-percent confidence interval for the 309,200 difference is from 295,980 to 322,420 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 288,050 to 330,350 housing units, and the 95-percent confidence interval is from 282,760 to 335,640. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms, since the 95-percent confidence interval does not include zero or negative values.

Medians.—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From table II determine the standard error of a 50-percent characteristic on the base of the median;
2. add to and subtract from 50 percent, the standard error determined in step 1; and
3. using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.1. The base of the distribution from which this median was determined is 1,269,600 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 1,269,600 is approximately 0.6 percentage points.
2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 48.8 and 51.2.
3. From table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 481,000 owner-occupied housing units or 37.9 percent, had one or two persons (for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 237,000 owner-occupied housing units, or 18.7 percent, had three persons (i.e., 2.5 to 3.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{48.8-37.9}{18.7}\right) = 3.1$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{51.2-37.9}{18.7}\right) = 3.2$$

Thus, the 95-percent confidence interval ranges from 3.1 to 3.2 persons.

Cincinnati 1975

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-41
Annual Housing Survey App-41
Selection of the sample. App-41
Building loss sample selection . . . App-42
1970 Census of Population and Housing. App-43
ESTIMATION. App-43
1975 housing inventory App-43
1970-1975 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing. App-44
RELIABILITY OF THE ESTIMATES. App-44
Nonsampling errors. App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-45
Rounding errors. App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences. App-47
Illustration of the computation of the standard error of a difference. App-47
Medians App-47
Illustration of the computation of the 95-percent confidence interval of a median App-48

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 21 SMSA's are based on data collected from the 1975 Annual Housing Survey (AHS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development. In each of the 21 SMSA's, the data were collected for the 12-month period from April 1975 through March 1976, with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and a third group of 20 SMSA's were collected from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated for the AHS every three years on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the second group (Year II SMSA's) are: Atlanta, Ga., Chicago, Ill., Philadelphia, Pa.-N.J., and San Francisco-Oakland, Calif.

The remaining SMSA's in the second group are: Cincinnati, Ohio-Ky.-Ind., Colorado Springs, Colo., Columbus, Ohio, Hartford, Conn., Kansas City, Mo.-Kans., Miami, Fla., Milwaukee, Wis., New Orleans, La., Newport News-Hampton, Va., Paterson-Clifton-Passaic, N.J., Portland, Oreg.-Wash., Rochester, N.Y., San Antonio, Tex., San Bernardino-Riverside-Ontario, Calif., San Diego, Calif., Springfield-Chicopee-Holyoke, Mass.-Conn., and Madison, Wis.

In this SMSA, 4,721 units were eligible for interview. Of these sample units, 173 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 408 units were visited but found not to be eligible for interview, because these units were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit issuing was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for those SMSA's which are *not* 100-percent permit issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). The following table indicates which SMSA's are 100-percent permit issuing and which contain a sample from the nonpermit universe.

100 Percent permit-issuing SMSA's	SMSA's with a sample from the nonpermit universe
Hartford, Conn.	Atlanta, Ga.
Miami, Fla.	Chicago, Ill.
Newport News-Hampton, Va.	Cincinnati, Ohio-Ky.-Ind.
Paterson-Clifton-Passaic, N.J.	Colorado Springs, Colo.
San Bernardino-Riverside-Ontario, Calif.	Columbus, Ohio
San Diego, Calif.	Kansas City, Mo.-Kans.
San Francisco-Oakland, Calif.	Madison, Wis.
	Milwaukee, Wis.
	New Orleans, La.
	Philadelphia, Pa.-N.J.
	Portland, Oreg.-Wash.
	Rochester, N.Y.
	San Antonio, Tex.
	Springfield-Chicopee-Holyoke, Mass.-Conn.

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of these sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and the balance of the SMSA, since the

sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—family size	Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000—$5,999 . .		
$6,000—$9,999 . .		
$10,000—$14,999		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. The housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued, and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \frac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Some tables in this report provide estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at enumeration time (i.e., the 1975 housing inventory), and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1975 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1975 housing inventory.—The AHS estimates of characteristics of the 1975 housing inventory employed a three-stage ratio estimation procedure. However, the third-stage ratio estimation procedure was employed in only three SMSA's. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 173 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units and one noninterview cell for the nonpermit universe (if applicable). Sample housing units from the nonpermit universe identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce the variation in sample size for strata used in the sample selection of the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes built after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or by other sources). This procedure was designed to adjust the AHS sample estimates of these types of units to independently derived current estimates. These independent estimates were considered to be the best estimates available. The adjustment was necessary to correct known deficiencies in the AHS sample in the representation of these units (see section on nonsampling error). The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1975 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimates of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units.	
New mobile homes	
"Other additions".	

The numerators of the ratios were derived by applying the following factors:

1. For the "conventional new construction units" cell, either a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in the SMSA; or the missed conventional new construction was estimated from the Survey of Construction (SOC). (The table that follows indicates the procedure used to estimate the missing conventional new construction for each SMSA.)
2. For the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied.
3. For the "other additions" cell, the "other additions" rate from SCARF[1] was applied.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

SMSA's employing the SOC estimate of missed conventional new construction	SMSA's employing the national trend for missed conventional new construction
Atlanta, Ga.	Colorado Springs, Colo.
Chicago, Ill.	Hartford, Conn.
Cincinnati, Ohio-Ky.-Ind.	Madison, Wis.
Columbus, Ohio	New Orleans, La.
Kansas City, Mo.-Kans.	Newport News—Hampton, Va.
Miami, Fla.	Paterson-Clifton-Passaic, N.J.
Milwaukee, Wis.	Rochester, N.Y.
Philadelphia, Pa.-N.J.	San Antonio, Tex.
Portland, Oreg.-Wash.	San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.	Springfield-Chicopee-Holyoke, Mass.-Conn.
San Francisco-Oakland, Calif.	

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

For some Year II SMSA's (i.e., Colorado Springs, Colo., Miami, Fla., and San Diego, Calif.) a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1975 housing inventory to an independent estimate of the SMSA's October 1975 housing inventory. This ratio estimate factor equaled the following:

$$\frac{\text{Independent estimate of the October 1975 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using the 1970 census estimate of total housing units for the SMSA in conjunction with the estimate of the change in the housing inventory since the census. The estimate of change was based on either administrative records from utility companies, or, when utility data were not available, on estimates of new construction permits and post-census demolition data. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, are brought into agreement with known estimates of the SMSA housing population.

The third-stage ratio estimation procedure was not employed in all Year II SMSA's since the reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. The independent estimates were only employed in SMSA's where the estimated relative bias was thought to be low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results. A measure of the relative bias of the independent estimate was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1975 independent estimates. In some SMSA's, the measure of the relative bias of the independent estimate was less than 3 percent but the independent estimate was still not employed, since there was evidence that suggested that the procedure for generating the independent estimates was producing substantial overestimates for these SMSA's.

1970-1975 lost units.—The AHS estimate of characteristics of the 1970-1975 lost units employed a one-stage ratio estimation procedure which is similar to the first-stage ratio estimation procedure mentioned above. The 1970-1975 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units + Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for State, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and non-

sampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

Nonsampling errors.—Nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1975 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors associated with 1970 census estimates—"coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

AHS-SMSA.—For the 1975 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are not yet available, however, it is believed that they will be similar to the results of the Year I AHS-SMSA reinterview study which were presented in the following Census Bureau memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample: 1974."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year I SMSA's and not for any specific SMSA):

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. The estimated indexes of inconsistency for most of the reinterview items were in the 20 to 50 range. The 20-50 interval is considered moderate indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview program because they had not been previously tested.

3. "Our bias indicator, the net difference rate, shows no signs of serious bias. For only one characteristic was there as much as a 2½ percent probability that the net difference rate is greater than 10 percent."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors.—With errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies in the representation of both conventional new construction and new mobile homes (and trailers) in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later, were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and

therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 2,300 conventional new construction units in this SMSA had permits issued prior to January 1970, and therefore, were missed by the 1975 AHS-SMSA survey.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. Although it is not known exactly, an estimated 1,700 new mobile homes in permit-issuing areas were missed by the 1975 AHS-SMSA survey in this SMSA. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS data still exists.

Rounding errors.—The rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling error, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in tables I and II are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1975 Housing Inventory and for Estimated Number of 1970-1975 Lost Units for the Cincinnati, Ohio-Ky.-Ind., SMSA

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	110	10,000 . .	1,060
200	150	25,000 . .	1,650
500	240	50,000 . .	2,290
700	280	100,000 .	3,070
1,000 . .	340	250,000 .	3,990
2,500 . .	530	400,000 .	3,640
5,000 . .	750		

Tables I and II present the standard errors applicable to estimates of characteristics of the 1975 housing inventory as well as estimates of characteristics of the 1970-1975 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in these tables.

For ratios, (100) (x/y), where x is not a subclass of y, table II underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approxi-

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the Cincinnati, Ohio-Ky.-Ind., SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
100	10.7	23.3	32.1	46.4	53.5
200	7.5	16.5	22.7	32.8	37.9
500	4.8	10.4	14.4	20.7	23.9
700	4.0	8.8	12.1	17.5	20.2
1,000	3.4	7.4	10.2	14.7	16.9
2,500	2.1	4.7	6.4	9.3	10.7
5,000	1.5	3.3	4.5	6.6	7.6
10,000	1.1	2.3	3.2	4.6	5.4
25,000	.7	1.5	2.0	2.9	3.4
50,000	.5	1.0	1.4	2.1	2.4
100,000	.3	.7	1.0	1.5	1.7
250,000	.2	.5	.6	.9	1.1
400,000	.2	.4	.5	.7	.8

mation of the standard error may be obtained by letting the standard error of the ratio be approximately equal to:

$$(100)\ (x/y)\ \sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

Let: x = the numerator of the ratio

y = the denominator of the ratio

σ_x = the standard error of the numerator

σ_y = the standard error of the denominator

Illustration of the use of the standard error tables.—Table 1 of part A of this report shows that in 1975 there were 276,400 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 3,930. Consequently, the 68-percent confidence interval, as shown by these data, is from 272,470 to 280,330 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 270,110 to 282,690 housing units with 90 percent confidence; and that the average estimate lies within the interval from 268,540 to 284,260 housing units with 95 percent confidence.

Table 1 of part A also shows that of the 276,400 owner-occupied housing units 70,800, or 25.6 percent, had two bedrooms. Interpolation in table II of this appendix shows that the standard error of the percent is approximately .9 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 24.7 to 26.5 percent; the 90-percent confidence interval is from 24.2 to 27.0 percent; and the 95-percent confidence interval is from 23.8 to 27.4 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; while, if there is a high negative correlation, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table 1 of part A of this report shows that in 1975 there were 132,100 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 61,300. Table I shows the standard error of 70,800 is approximately 2,610 and the standard error of 132,100 is approximately 3,270. Therefore, the standard error of the estimated difference of 61,300 is about

$$4{,}180 = \sqrt{(2{,}610)^2 + (3{,}270)^2}$$

Consequently, the 68-percent confidence interval for the 61,300 difference is from 57,120 to 65,480 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 54,610 to 67,990 housing units, and the 95-percent confidence interval is from 52,940 to 69,660. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms, since the 95-percent confidence interval does not include zero or negative values.

Medians.—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated

degree of confidence that the average median from all possible samples lies with the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From table II determine the standard error of a 50-percent characteristic on the base of the median;

2. add to and subtract from 50 percent, the standard error determined in step 1; and

3. using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median.—Table 1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.0. The base of the distribution from which this median was determined is 276,400 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 276,400 is approximately 1.0 percentage points.

2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 48.0 and 52.0.

3. From table 1 of part A, it can be seen by cumulating the frequencies for the first two categories that 111,100 owner-occupied housing units, or 40.2 percent, had one or two persons (for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 48,900 owner-occupied housing units, or 17.7 percent, had three persons (i.e., 2.5 to 3.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{48.0-40.2}{17.7}\right)=2.9$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{52.0-40.2}{17.7}\right)=3.2$$

Thus, the 95-percent confidence interval ranges from 2.9 to 3.2 persons.

Columbus 1975

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-41
Annual Housing Survey App-41
Selection of the sample. App-41
Building loss sample selection . . . App-42
1970 Census of Population and Housing. App-43
ESTIMATION. App-43
1975 housing inventory App-43
1970-1975 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing. App-44
RELIABILITY OF THE ESTIMATES. App-44
Nonsampling errors. App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-45
Rounding errors. App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences. App-47
Illustration of the computation of the standard error of a difference. App-47
Medians App-47
Illustration of the computation of the 95-percent confidence interval of a median App-48

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 21 SMSA's are based on data collected from the 1975 Annual Housing Survey (AHS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development. In each of the 21 SMSA's, the data were collected for the 12-month period from April 1975 through March 1976, with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and a third group of 20 SMSA's were collected from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated for the AHS every three years on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the second group (Year II SMSA's) are: Atlanta, Ga., Chicago, Ill., Philadelphia, Pa.-N.J., and San Francisco-Oakland, Calif.

The remaining SMSA's in the second group are: Cincinnati, Ohio-Ky.-Ind., Colorado Springs, Colo., Columbus, Ohio, Hartford, Conn., Kansas City, Mo.-Kans., Miami, Fla., Milwaukee, Wis., New Orleans, La., Newport News-Hampton, Va., Paterson-Clifton-Passaic, N.J., Portland, Oreg.-Wash., Rochester, N.Y., San Antonio, Tex., San Bernardino-Riverside-Ontario, Calif., San Diego, Calif., Springfield-Chicopee-Holyoke, Mass.-Conn., and Madison, Wis.

In this SMSA, 4,749 units were eligible for interview. Of these sample units, 203 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 464 units were visited but found not to be eligible for interview, because these units were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit issuing was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for those SMSA's which are *not* 100-percent permit issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). The following table indicates which SMSA's are 100-percent permit issuing and which contain a sample from the nonpermit universe.

100 Percent permit-issuing SMSA's	SMSA's with a sample from the nonpermit universe
Hartford, Conn.	Atlanta, Ga.
Miami, Fla.	Chicago, Ill.
Newport News-Hampton, Va.	Cincinnati, Ohio-Ky.-Ind.
Paterson-Clifton-Passaic, N.J.	Colorado Springs, Colo.
San Bernardino-Riverside-Ontario, Calif.	Columbus, Ohio
San Diego, Calif.	Kansas City, Mo.-Kans.
San Francisco-Oakland, Calif.	Madison, Wis.
	Milwaukee, Wis.
	New Orleans, La.
	Philadelphia, Pa.-N.J.
	Portland, Oreg.-Wash.
	Rochester, N.Y.
	San Antonio, Tex.
	Springfield-Chicopee-Holyoke, Mass.-Conn.

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of these sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and the balance of the SMSA, since the

sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—family size	Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000—$5,999 . .		
$6,000—$9,999 . .		
$10,000—$14,999		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. The housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued, and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \frac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Some tables in this report provide estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at enumeration time (i.e., the 1975 housing inventory), and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1975 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1975 housing inventory.—The AHS estimates of characteristics of the 1975 housing inventory employed a three-stage ratio estimation procedure. However, the third-stage ratio estimation procedure was employed in only three SMSA's. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 203 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units and one noninterview cell for the nonpermit universe (if applicable). Sample housing units from the nonpermit universe identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce the variation in sample size for strata used in the sample selection of the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes built after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or by other sources). This procedure was designed to adjust the AHS sample estimates of these types of units to independently derived current estimates. These independent estimates were considered to be the best estimates available. The adjustment was necessary to correct known deficiencies in the AHS sample in the representation of these units (see section on nonsampling error). The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1975 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimates of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units.	
New mobile homes	
"Other additions".	

The numerators of the ratios were derived by applying the following factors:

1. For the "conventional new construction units" cell, either a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in the SMSA, or the missed conventional new construction was estimated from the Survey of Construction (SOC). (The table that follows indicates the procedure used to estimate the missing conventional new construction for each SMSA.)
2. For the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied.
3. For the "other additions" cell, the "other additions" rate from SCARF[1] was applied.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

SMSA's employing the SOC estimate of missed conventional new construction	SMSA's employing the national trend for missed conventional new construction
Atlanta, Ga.	Colorado Springs, Colo.
Chicago, Ill.	Hartford, Conn.
Cincinnati, Ohio-Ky.-Ind.	Madison, Wis.
Columbus, Ohio	New Orleans, La.
Kansas City, Mo.-Kans.	Newport News—Hampton, Va.
Miami, Fla.	Paterson-Clifton-Passaic, N.J.
Milwaukee, Wis.	Rochester, N.Y.
Philadelphia, Pa.-N.J.	San Antonio, Tex.
Portland, Oreg.-Wash.	San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.	Springfield-Chicopee-Holyoke, Mass.-Conn.
San Francisco-Oakland, Calif.	

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

For some Year II SMSA's (i.e., Colorado Springs, Colo., Miami, Fla., and San Diego, Calif.) a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1975 housing inventory to an independent estimate of the SMSA's October 1975 housing inventory. This ratio estimate factor equaled the following:

$$\frac{\text{Independent estimate of the October 1975 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using the 1970 census estimate of total housing units for the SMSA in conjunction with the estimate of the change in the housing inventory since the census. The estimate of change was based on either administrative records from utility companies, or, when utility data were not available, on estimates of new construction permits and post-census demolition data. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, are brought into agreement with known estimates of the SMSA housing population.

The third-stage ratio estimation procedure was not employed in all Year II SMSA's since the reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. The independent estimates were only employed in SMSA's where the estimated relative bias was thought to be low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results. A measure of the relative bias of the independent estimate was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1975 independent estimates. In some SMSA's, the measure of the relative bias of the independent estimate was less than 3 percent but the independent estimate was still not employed, since there was evidence that suggested that the procedure for generating the independent estimates was producing substantial overestimates for these SMSA's.

1970-1975 lost units.—The AHS estimate of characteristics of the 1970-1975 lost units employed a one-stage ratio estimation procedure which is similar to the first-stage ratio estimation procedure mentioned above. The 1970-1975 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and non-

sampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Nonsampling errors.—Nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1975 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors associated with 1970 census estimates—"coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census,** and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

AHS-SMSA.—For the 1975 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are not yet available, however, it is believed that they will be similar to the results of the Year I AHS-SMSA reinterview study which were presented in the following Census Bureau memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample: 1974."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year I SMSA's and not for any specific SMSA):

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."
2. The estimated indexes of inconsistency for most of the reinterview items were in the 20 to 50 range. The 20-50 interval is considered moderate indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview program because they had not been previously tested.
3. "Our bias indicator, the net difference rate, shows no signs of serious bias. For only one characteristic was there as much as a 2½ percent probability that the net difference rate is greater than 10 percent."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors.—With errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies in the representation of both conventional new construction and new mobile homes (and trailers) in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later, were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and

therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 3,700 conventional new construction units in this SMSA had permits issued prior to January 1970, and therefore, were missed by the 1975 AHS-SMSA survey.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. Although it is not known exactly, an estimated 1,800 new mobile homes in permit-issuing areas were missed by the 1975 AHS-SMSA survey in this SMSA. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS data still exists.

Rounding errors.—The rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling error, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in tables I and II are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1975 Housing Inventory and for Estimated Number of 1970-1975 Lost Units for the Columbus, Ohio, SMSA

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100. . . .	90	5,000. . .	620
200. . . .	130	10,000. .	870
500. . . .	200	25,000. .	1,360
700. . . .	230	50,000. .	1,850
1,000 . .	280	100,000 .	2,440
2,500 . .	440	250,000 .	2,790

Tables I and II present the standard errors applicable to estimates of characteristics of the 1975 housing inventory as well as estimates of characteristics of the 1970-1975 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in these tables.

For ratios, (100) (x/y), where x is not a subclass of y, table II underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be obtained by letting the standard error of the ratio be approximately equal to:

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the Columbus, Ohio, SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
100	8.8	19.3	26.5	38.3	44.2
200	6.2	13.6	18.8	27.1	31.3
500	3.9	8.6	11.9	17.1	19.8
700	3.3	7.3	10.0	14.5	16.7
1,000	2.8	6.1	8.4	12.1	14.0
2,500	1.8	3.9	5.3	7.7	8.8
5,000	1.2	2.7	3.8	5.4	6.3
10,000	.9	1.9	2.7	3.8	4.4
25,000	.6	1.2	1.7	2.4	2.8
50,000	.4	.9	1.2	1.7	2.0
100,000	.3	.6	.8	1.2	1.4
250,000	.2	.4	.5	.8	.9

$$(100)\,(x/y)\,\sqrt{\left(\frac{\sigma_x}{x}\right)^2+\left(\frac{\sigma_y}{y}\right)^2}$$

Let: x = the numerator of the ratio
y = the denominator of the ratio
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

Illustration of the use of the standard error tables.—Table 1 of part A of this report shows that in 1975 there were 184,300 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 2,640. Consequently, the 68-percent confidence interval, as shown by these data, is from 181,660 to 186,940 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 180,080 to 188,520 housing units with 90 percent confidence; and that the average estimate lies within the interval from 179,020 to 189,580 housing units with 95 percent confidence.

Table 1 of part A also shows that of the 184,300 owner-occupied housing units 37,300, or 20.2 percent, had two bedrooms. Interpolation in table II of this appendix shows that the standard error of the percent is approximately .9 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 19.3 to 21.1 percent; the 90-percent confidence interval is from 18.8 to 21.6 percent; and the 95-percent confidence interval is from 18.4 to 22.0 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; while, if there is a high negative correlation, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table 1 of part A of this report shows that in 1975 there were 105,100 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedroom and owner-occupied units with three bedrooms is 67,800. Table I shows the standard error of 37,300 is approximately 1,600 and the standard error of 105,100 is approximately 2,450. Therefore, the standard error of the estimated difference of 67,800 is about

$$2{,}930 = \sqrt{(1{,}600)^2 + (2{,}450)^2}$$

Consequently, the 68-percent confidence interval for the 67,800 difference is from 64,870 to 70,730 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 63,110 to 72,490 housing units, and the 95-percent confidence interval is from 61,940 to 73,660. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval does not include zero or negative values.

Medians.—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From table II determine the standard error of a 50-percent characteristic on the base of the median;
2. add to and subtract from 50 percent, the standard error determined in step 1; and
3. using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median.—Table 1 of part A of this report shows the median number of persons for owner-occupied housing units is 2.9. The base of the distribution from which this median was determined is 184,300 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 184,300 is approximately 1.1 percentage points.
2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 47.8 and 52.2.
3. From table 1 of part A, it can be seen by cumulating the frequencies for the first two categories that 78,600 owner-occupied housing units, or 42.6 percent, had one or two persons (for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 33,700 owner-occupied, housing units, or 18.3 percent, had three persons (i.e., 2.5 to 3.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{47.8-42.6}{18.3}\right) = 2.8$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{52.2-42.6}{18.3}\right) = 3.0$$

Thus, the 95-percent confidence interval ranges from 2.8 to 3.0 persons.

Hartford 1975

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-41
Annual Housing Survey App-41
Selection of the sample. App-41
Building loss sample selection . . . App-42
1970 Census of Population and Housing. App-43
ESTIMATION. App-43
1975 housing inventory App-43
1970-1975 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing. App-44
RELIABILITY OF THE ESTIMATES. App-44
Nonsampling errors. App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-45
Rounding errors. App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences. App-47
Illustration of the computation of the standard error of a difference. App-47
Medians App-47
Illustration of the computation of the 95-percent confidence interval of a median App-48

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 21 SMSA's are based on data collected from the 1975 Annual Housing Survey (AHS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development. In each of the 21 SMSA's, the data were collected for the 12-month period from April 1975 through March 1976, with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and a third group of 20 SMSA's were collected from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated for the AHS every three years on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the second group (Year II SMSA's) are: Atlanta, Ga., Chicago, Ill., Philadelphia, Pa.-N.J., and San Francisco-Oakland, Calif.

The remaining SMSA's in the second group are: Cincinnati, Ohio-Ky.-Ind., Colorado Springs, Colo., Columbus, Ohio, Hartford, Conn., Kansas City, Mo.-Kans., Miami, Fla., Milwaukee, Wis., New Orleans, La., Newport News-Hampton, Va., Paterson-Clifton-Passaic, N.J., Portland, Oreg.-Wash., Rochester, N.Y., San Antonio, Tex., San Bernardino-Riverside-Ontario, Calif., San Diego, Calif., Springfield-Chicopee-Holyoke, Mass.-Conn., and Madison, Wis.

In this SMSA, 4,926 units were eligible for interview. Of these sample units, 213 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 322 units were visited but found not to be eligible for interview, because these units were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit issuing was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for those SMSA's which are *not* 100-percent permit issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). The following table indicates which SMSA's are 100-percent permit issuing and which contain a sample from the nonpermit universe.

100 Percent permit-issuing SMSA's	SMSA's with a sample from the nonpermit universe
Hartford, Conn.	Atlanta, Ga.
Miami, Fla.	Chicago, Ill.
Newport News-Hampton, Va.	Cincinnati, Ohio-Ky.-Ind.
Paterson-Clifton-Passaic, N.J.	Colorado Springs, Colo.
San Bernardino-Riverside-Ontario, Calif.	Columbus, Ohio
San Diego, Calif.	Kansas City, Mo.-Kans.
San Francisco-Oakland, Calif.	Madison, Wis.
	Milwaukee, Wis.
	New Orleans, La.
	Philadelphia, Pa.-N.J.
	Portland, Oreg.-Wash.
	Rochester, N.Y.
	San Antonio, Tex.
	Springfield-Chicopee-Holyoke, Mass.-Conn.

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of these sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and the balance of the SMSA, since the

sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—family size	Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000—$5,999 . .		
$6,000—$9,999 . .		
$10,000—$14,999		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. The housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued, and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \frac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Some tables in this report provide estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume 1, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at enumeration time (i.e., the 1975 housing inventory), and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1975 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1975 housing inventory.—The AHS estimates of characteristics of the 1975 housing inventory employed a three-stage ratio estimation procedure. However, the third-stage ratio estimation procedure was employed in only three SMSA's. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 213 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units and one noninterview cell for the nonpermit universe (if applicable). Sample housing units from the nonpermit universe identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce the variation in sample size for strata used in the sample selection of the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes built after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or by other sources). This procedure was designed to adjust the AHS sample estimates of these types of units to independently derived current estimates. These independent estimates were considered to be the best estimates available. The adjustment was necessary to correct known deficiencies in the AHS sample in the representation of these units (see section on nonsampling error). The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1975 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimates of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units. . . . .
New mobile homes . . . . .
"Other additions". . . . .

The numerators of the ratios were derived by applying the following factors:

1. For the "conventional new construction units" cell, either a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in the SMSA, or the missed conventional new construction was estimated from the Survey of Construction (SOC). (The table that follows indicates the procedure used to estimate the missing conventional new construction for each SMSA.)
2. For the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied.
3. For the "other additions" cell, the "other additions" rate from SCARF[1] was applied.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

SMSA's employing the SOC estimate of missed conventional new construction	SMSA's employing the national trend for missed conventional new construction
Atlanta, Ga.	Colorado Springs, Colo.
Chicago, Ill.	Hartford, Conn.
Cincinnati, Ohio-Ky.-Ind.	Madison, Wis.
Columbus, Ohio	New Orleans, La.
Kansas City, Mo.-Kans.	Newport News—Hampton, Va.
Miami, Fla.	Paterson-Clifton-Passaic, N.J.
Milwaukee, Wis.	Rochester, N.Y.
Philadelphia, Pa.-N.J.	San Antonio, Tex.
Portland, Oreg.-Wash.	San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.	Springfield-Chicopee-Holyoke, Mass.-Conn.
San Francisco-Oakland, Calif.	

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

For some Year II SMSA's (i.e., Colorado Springs, Colo., Miami, Fla., and San Diego, Calif.) a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1975 housing inventory to an independent estimate of the SMSA's October 1975 housing inventory. This ratio estimate factor equaled the following:

$$\frac{\text{Independent estimate of the October 1975 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using the 1970 census estimate of total housing units for the SMSA in conjunction with the estimate of the change in the housing inventory since the census. The estimate of change was based on either administrative records from utility companies, or, when utility data were not available, on estimates of new construction permits and post-census demolition data. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, are brought into agreement with known estimates of the SMSA housing population.

The third-stage ratio estimation procedure was not employed in all Year II SMSA's since the reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. The independent estimates were only employed in SMSA's where the estimated relative bias was thought to be low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results. A measure of the relative bias of the independent estimate was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1975 independent estimates. In some SMSA's, the measure of the relative bias of the independent estimate was less than 3 percent but the independent estimate was still not employed, since there was evidence that suggested that the procedure for generating the independent estimates was producing substantial overestimates for these SMSA's.

1970-1975 lost units.—The AHS estimate of characteristics of the 1970-1975 lost units employed a one-stage ratio estimation procedure which is similar to the first-stage ratio estimation procedure mentioned above. The 1970-1975 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units + Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for State, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and non-

sampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

Nonsampling errors.—Nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1975 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors associated with 1970 census estimates—"coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

AHS-SMSA.—For the 1975 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are not yet available, however, it is believed that they will be similar to the results of the Year I AHS-SMSA reinterview study which were presented in the following Census Bureau memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample: 1974."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year I SMSA's and not for any specific SMSA):

1. "The results indicate that the interviewers are doing a good job: From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. The estimated indexes of inconsistency for most of the reinterview items were in the 20 to 50 range. The 20-50 interval is considered moderate indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview program because they had not been previously tested.

3. "Our bias indicator, the net difference rate, shows no signs of serious bias. For only one characteristic was there as much as a 2½ percent probability that the net difference rate is greater than 10 percent."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors.—With errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies in the representation of both conventional new construction and new mobile homes (and trailers) in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later, were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and

therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 4,700 conventional new construction units in this SMSA had permits issued prior to January 1970, and therefore, were missed by the 1975 AHS-SMSA survey.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. Although it is not known exactly, an estimated 300 new mobile homes in permit-issuing areas were missed by the 1975 AHS-SMSA survey in this SMSA. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS data still exists.

Rounding errors.—The rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling error, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in tables I and II are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

Tables I and II present the standard errors applicable to estimates of characteristics of the 1975 housing inventory as well as estimates of characteristics of the 1970-1975 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in these tables.

For ratios, (100) (x/y), where x is not a subclass of y, table II underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1975 Housing Inventory and for Estimated Number of 1970-1975 Lost Units for the Hartford, Conn., SMSA

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	70	5,000 . .	490
200	100	10,000 .	680
500	160	25,000 .	1,050
700	180	50,000 .	1,400
1,000 . .	220	100,000	1,740
2,500 . .	350	200,000	1,530

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the Hartford, Conn., SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
100	6.9	15.2	20.9	30.1	34.8
200	4.9	10.7	14.8	21.3	24.6
500	3.1	6.8	9.3	13.5	15.6
700	2.6	5.7	7.9	11.4	13.2
1,000	2.2	4.8	6.6	9.5	11.0
2,500	1.4	3.0	4.2	6.0	7.0
5,000	1.0	2.1	3.0	4.3	4.9
10,000	.7	1.5	2.1	3.0	3.5
25,000	.4	1.0	1.3	1.9	2.2
50,000	.3	.7	.9	1.3	1.6
100,000	.2	.5	.7	1.0	1.1
200,000	.2	.3	.5	.7	.8

obtained by letting the standard error of the ratio be approximately equal to:

$$(100)\,(x/y)\,\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

Let: x = the numerator of the ratio
y = the denominator of the ratio
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

Illustration of the use of the standard error tables.—Table 1 of part A of this report shows that in 1975 there were 131,200 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 1,670. Consequently, the 68-percent confidence interval, as shown by these data, is from 129,530 to 132,870 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 128,530 to 133,870 housing units with 90 percent confidence; and that the average estimate lies within the interval from 127,860 to 134,540 housing units with 95 percent confidence.

Table 1 of part A also shows that of the 131,200 owner-occupied housing units, 29,400, or 22.4 percent, had two bedrooms. Interpolation in table II of this appendix shows that the standard error of the percent is approximately .9 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 21.5 to 23.3 percent; the 90-percent confidence interval is from 21.0 to 23.8 percent; and the 95-percent confidence interval is from 20.6 to 24.2 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics; the formula will overestimate the true standard error; while, if there is a high negative correlation, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table 1 of part A of this report shows that in 1975 there were 67,200 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 37,800. Table I shows the standard error of 29,400 is approximately 1,110 and the standard error of 67,200 is approximately 1,520. Therefore, the standard error of the estimated difference of 37,800 is about

$$1{,}880 = \sqrt{(1{,}110)^2 + (1{,}520)^2}$$

Consequently, the 68-percent confidence interval for the 37,800 difference is from 35,920 to 39,680 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 34,790 to 40,810 housing units, and the 95-percent confidence interval is from 34,040 to 41,560. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval does not include zero or negative values.

Medians.—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From table II determine the standard error of a 50-percent characteristic on the base of the median;
2. add to and subtract from 50 percent the standard error determined in step 1; and
3. using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median.—Table 1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.3. The base of the distribution from which this median was determined is 131,200 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 131,200 is approximately 1.0 percentage points.
2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 48.0 and 52.0.
3. From table 1 of part A, it can be seen by cumulating the frequencies for the first two categories that 47,400 owner-occupied housing units, or 36.1 percent, had one or two persons (for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 23,500 owner-occupied housing units, or 17.9 percent, had three persons (i.e., 2.5 to 3.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{48.0-36.1}{17.9}\right) = 3.2$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{52.0-36.1}{17.9}\right) = 3.4$$

Thus, the 95-percent confidence interval ranges from 3.2 to 3.4 persons.

Kansas City 1975

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-41
Annual Housing Survey App-41
Selection of the sample. App-41
Building loss sample selection . . . App-42
1970 Census of Population and Housing. App-43
ESTIMATION. App-43
1975 housing inventory App-43
1970-1975 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing. App-44
RELIABILITY OF THE ESTIMATES. App-44
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-45
Rounding errors. App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences. App-47
Illustration of the computation of the standard error of a difference. App-47
Medians App-47
Illustration of the computation of the 95-percent confidence interval of a median App-48

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 21 SMSA's are based on data collected from the 1975 Annual Housing Survey (AHS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development. In each of the 21 SMSA's, the data were collected for the 12-month period from April 1975 through March 1976, with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and a third group of 20 SMSA's were collected from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated for the AHS every three years on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the second group (Year II SMSA's) are: Atlanta, Ga., Chicago, Ill., Philadelphia, Pa.-N.J., and San Francisco-Oakland, Calif.

The remaining SMSA's in the second group are: Cincinnati, Ohio-Ky.-Ind., Colorado Springs, Colo., Columbus, Ohio, Hartford, Conn., Kansas City, Mo.-Kans., Miami, Fla., Milwaukee, Wis., New Orleans, La., Newport News-Hampton, Va., Paterson-Clifton-Passaic, N.J., Portland, Oreg.-Wash., Rochester, N.Y., San Antonio, Tex., San Bernardino-Riverside-Ontario, Calif., San Diego, Calif., Springfield-Chicopee-Holyoke, Mass.-Conn., and Madison, Wis.

In this SMSA, 4,684 units were eligible for interview. Of these sample units, 210 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 472 units were visited but found not to be eligible for interview, because these units were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit issuing was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for those SMSA's which are *not* 100-percent permit issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). The following table indicates which SMSA's are 100-percent permit issuing and which contain a sample from the nonpermit universe.

100 Percent permit-issuing SMSA's	SMSA's with a sample from the nonpermit universe
Hartford, Conn.	Atlanta, Ga.
Miami, Fla.	Chicago, Ill.
Newport News-Hampton, Va.	Cincinnati, Ohio-Ky.-Ind.
Paterson-Clifton-Passaic, N.J.	Colorado Springs, Colo.
San Bernardino-Riverside-Ontario, Calif.	Columbus, Ohio
San Diego, Calif.	Kansas City, Mo.-Kans.
San Francisco-Oakland, Calif.	Madison, Wis.
	Milwaukee, Wis.
	New Orleans, La.
	Philadelphia, Pa.-N.J.
	Portland, Oreg.-Wash.
	Rochester, N.Y.
	San Antonio, Tex.
	Springfield-Chicopee-Holyoke, Mass.-Conn.

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of these sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and the balance of the SMSA, since the

sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—family size	Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000—$5,999 . .		
$6,000—$9,999 . .		
$10,000—$14,999		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. The housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued, and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \frac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Some tables in this report provide estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at enumeration time (i.e., the 1975 housing inventory), and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1975 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1975 housing inventory.—The AHS estimates of characteristics of the 1975 housing inventory employed a three-stage ratio estimation procedure. However, the third-stage ratio estimation procedure was employed in only three SMSA's. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 210 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units and one noninterview cell for the nonpermit universe (if applicable). Sample housing units from the nonpermit universe identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce the variation in sample size for strata used in the sample selection of the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes built after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or by other sources). This procedure was designed to adjust the AHS sample estimates of these types of units to independently derived current estimates. These independent estimates were considered to be the best estimates available. The adjustment was necessary to correct known deficiencies in the AHS sample in the representation of these units (see section on nonsampling error). The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1975 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimates of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units	
New mobile homes	
"Other additions"	

The numerators of the ratios were derived by applying the following factors:

1. For the "conventional new construction units" cell, either a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in the SMSA, or the missed conventional new construction was estimated from the Survey of Construction (SOC). (The table that follows indicates the procedure used to estimate the missing conventional new construction for each SMSA.)
2. For the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied.
3. For the "other additions" cell, the "other additions" rate from SCARF[1] was applied.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

SMSA's employing the SOC estimate of missed conventional new construction	SMSA's employing the national trend for missed conventional new construction
Atlanta, Ga.	Colorado Springs, Colo.
Chicago, Ill.	Hartford, Conn.
Cincinnati, Ohio-Ky.-Ind.	Madison, Wis.
Columbus, Ohio	New Orleans, La.
Kansas City, Mo.-Kans.	Newport News—Hampton, Va.
Miami, Fla.	Paterson-Clifton-Passaic, N.J.
Milwaukee, Wis.	Rochester, N.Y.
Philadelphia, Pa.-N.J.	San Antonio, Tex.
Portland, Oreg.-Wash.	San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.	Springfield-Chicopee-Holyoke, Mass.-Conn.
San Francisco-Oakland, Calif.	

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

For some Year II SMSA's (i.e., Colorado Springs, Colo., Miami, Fla., and San Diego, Calif.) a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1975 housing inventory to an independent estimate of the SMSA's October 1975 housing inventory. This ratio estimate factor equaled the following:

$$\frac{\text{Independent estimate of the October 1975 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using the 1970 census estimate of total housing units for the SMSA in conjunction with the estimate of the change in the housing inventory since the census. The estimate of change was based on either administrative records from utility companies, or, when utility data were not available, on estimates of new construction permits and post-census demolition data. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, are brought into agreement with known estimates of the SMSA housing population.

The third-stage ratio estimation procedure was not employed in all Year II SMSA's since the reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. The independent estimates were only employed in SMSA's where the estimated relative bias was thought to be low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results. A measure of the relative bias of the independent estimate was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1975 independent estimates. In some SMSA's, the measure of the relative bias of the independent estimate was less than 3 percent but the independent estimate was still not employed, since there was evidence that suggested that the procedure for generating the independent estimates was producing substantial overestimates for these SMSA's.

1970-1975 lost units.—The AHS estimate of characteristics of the 1970-1975 lost units employed a one-stage ratio estimation procedure which is similar to the first-stage ratio estimation procedure mentioned above. The 1970-1975 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units + Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for State, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and non-

sampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

Nonsampling errors.—Nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1975 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors associated with 1970 census estimates—"coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

AHS-SMSA.—For the 1975 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are not yet available, however, it is believed that they will be similar to the results of the Year I AHS-SMSA reinterview study which were presented in the following Census Bureau memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample: 1974."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year I SMSA's and not for any specific SMSA):

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."
2. The estimated indexes of inconsistency for most of the reinterview items were in the 20 to 50 range. The 20-50 interval is considered moderate indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview program because they had not been previously tested."
3. "Our bias indicator, the net difference rate, shows no signs of serious bias. For only one characteristic was there as much as a 2½ percent probability that the net difference rate is greater than 10 percent."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors.—With errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies in the representation of both conventional new construction and new mobile homes (and trailers) in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later, were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and

therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 4,700 conventional new construction units in this SMSA had permits issued prior to January 1970, and therefore, were missed by the 1975 AHS-SMSA survey.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. Although it is not known exactly, an estimated 1,200 new mobile homes in permit-issuing areas were missed by the 1975 AHS-SMSA survey in this SMSA. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS data still exists.

Rounding errors.—The rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling error, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in tables I and II are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1975 Housing Inventory and for Estimated Number of 1970-1975 Lost Units for the Kansas City, Mo.-Kans., SMSA

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100. . . .	110	10,000 .	1,040
200. . . .	150	25,000 .	1,620
500. . . .	230	50,000 .	2,230
700. . . .	280	100,000	3,000
1,000 . .	330	250,000	3,890
2,500 . .	520	400,000	3,520
5,000 . .	740		

Tables I and II present the standard errors applicable to estimates of characteristics of the 1975 housing inventory as well as estimates of characteristics of the 1970-1975 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in these tables.

For ratios, (100) (x/y), where x is not a subclass of y, table II underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be ob-

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the Kansas City, Mo.-Kans., SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
100	10.4	22.8	31.4	45.3	52.3
200	7.4	16.1	22.2	32.0	37.0
500	4.7	10.2	14.0	20.3	23.4
700	3.9	8.6	11.9	17.1	19.8
1,000	3.3	7.2	9.9	14.3	16.5
2,500	2.1	4.6	6.3	9.1	10.5
5,000	1.5	3.2	4.4	6.4	7.4
10,000	1.0	2.3	3.1	4.5	5.2
25,000	.7	1.4	2.0	2.9	3.3
50,000	.5	1.0	1.4	2.0	2.3
100,000	.3	.7	1.0	1.4	1.7
250,000	.2	.5	.6	.9	1.0
400,000	.2	.4	.5	.7	.8

tained by letting the standard error of the ratio be approximately equal to:

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

Let: x = the numerator of the ratio
y = the denominator of the ratio
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

Illustration of the use of the standard error tables.—Table 1 of part A of this report shows that in 1975 there were 276,500 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 3,820. Consequently, the 68-percent confidence interval, as shown by these data, is from 272,680 to 280,320 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 270,390 to 282,610 housing units with 90 percent confidence; and that the average estimate lies within the interval from 268,860 to 284,140 housing units with 95 percent confidence.

Table 1 of part A also shows that of the 276,500 owner-occupied housing units 78,100, or 28.2 percent, had two bedrooms. Interpolation in table II of this appendix shows that the standard error of the percent is approximately .9 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 27.3 to 29.1 percent; the 90-percent confidence interval is from 26.8 to 29.6 percent; and the 95-percent confidence interval is from 26.4 to 30.0 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics; the formula will overestimate the true standard error; while, if there is a high negative correlation, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table 1 of part A of this report shows that in 1975 there were 142,500 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 64,400. Table I shows the standard error of 78,100 is approximately 2,660 and the standard error of 142,500 is approximately 3,250. Therefore, the standard error of the estimated difference of 64,400 is about

$$4{,}200 = \sqrt{(2{,}660)^2 + (3{,}250)^2}$$

Consequently, the 68-percent confidence interval for the 64,400 difference is from 60,200 to 68,600 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 57,680 to 71,120 housing units, and the 95-percent confidence interval is from 56,000 to 72,800. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval does not include zero or negative values.

Medians.—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following

procedure may be used to estimate confidence limits of a median based on sample data:

1. From table II determine the standard error of a 50-percent characteristic on the base of the median,
2. add to and subtract from 50 percent, the standard error determined in step 1; and
3. using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median.—Table 1 of part A of this report shows the median number of persons for owner-occupied housing units is 2.8. The base of the distribution from which this median was determined is 276,500 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 276,500 is approximately 1.0 percentage points.
2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 48.0 and 52.0.
3. From table 1 of part A, it can be seen by cumulating the frequencies for the first two categories that 121,000 owner-occupied housing units, or 43.8 percent, had one or two persons (for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 54,500 owner-occupied housing units, or 19.7 percent, had three persons (i.e., 2.5 to 3.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{48.0-43.8}{19.7}\right) = 2.7$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{52.0-43.8}{19.7}\right) = 2.9$$

Thus, the 95-percent confidence interval ranges from 2.7 to 2.9 persons.

Madison 1975

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-41
Annual Housing Survey App-41
Selection of the sample. App-41
Building loss sample selection . . . App-42
1970 Census of Population and Housing. App-43
ESTIMATION. App-43
1975 housing inventory App-43
1970-1975 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing. App-44
RELIABILITY OF THE ESTIMATES. App-44
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-45
Rounding errors. App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences. App-47
Illustration of the computation of the standard error of a difference. App-47
Medians App-47
Illustration of the computation of the 95-percent confidence interval of a median App-48

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 21 SMSA's are based on data collected from the 1975 Annual Housing Survey (AHS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development. In each of the 21 SMSA's, the data were collected for the 12-month period from April 1975 through March 1976, with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and a third group of 20 SMSA's were collected from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated for the AHS every three years on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the second group (Year II SMSA's) are: Atlanta, Ga., Chicago, Ill., Philadelphia, Pa.-N.J., and San Francisco-Oakland, Calif.

The remaining SMSA's in the second group are: Cincinnati, Ohio-Ky.-Ind., Colorado Springs, Colo., Columbus, Ohio, Hartford, Conn., Kansas City, Mo.-Kans., Miami, Fla., Milwaukee, Wis., New Orleans, La., Newport News-Hampton, Va., Paterson-Clifton-Passaic, N.J., Portland, Oreg.-Wash., Rochester, N.Y., San Antonio, Tex., San Bernardino-Riverside-Ontario, Calif., San Diego, Calif., Springfield-Chicopee-Holyoke, Mass.-Conn., and Madison, Wis.

In this SMSA, 4,899 units were eligible for interview. Of these sample units, 129 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 495 units were visited but found not to be eligible for interview, because these units were found to be condemned, unfit, demolished, converted to group quarters use, etc..

Selection of the sample.—The sample for the SMSA's which are 100-percent permit issuing was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for those SMSA's which are *not* 100-percent permit issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). The following table indicates which SMSA's are 100-percent permit issuing and which contain a sample from the nonpermit universe.

100 Percent permit-issuing SMSA's	SMSA's with a sample from the nonpermit universe
Hartford, Conn.	Atlanta, Ga.
Miami, Fla.	Chicago, Ill.
Newport News-Hampton, Va.	Cincinnati, Ohio-Ky.-Ind.
Paterson-Clifton-Passaic, N.J.	Colorado Springs, Colo.
San Bernardino-Riverside-Ontario, Calif.	Columbus, Ohio
San Diego, Calif.	Kansas City, Mo.-Kans.
San Francisco-Oakland, Calif.	Madison, Wis.
	Milwaukee, Wis.
	New Orleans, La.
	Philadelphia, Pa.-N.J.
	Portland, Oreg.-Wash.
	Rochester, N.Y.
	San Antonio, Tex.
	Springfield-Chicopee-Holyoke, Mass.-Conn.

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of these sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and the balance of the SMSA, since the

sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—family size	Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000—$5,999 . .		
$6,000—$9,999 . .		
$10,000—$14,999		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. The housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued, and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \frac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Some tables in this report provide estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at enumeration time (i.e., the 1975 housing inventory), and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1975 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1975 housing inventory.—The AHS estimates of characteristics of the 1975 housing inventory employed a three-stage ratio estimation procedure. However, the third-stage ratio estimation procedure was employed in only three SMSA's. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 129 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units and one noninterview cell for the nonpermit universe (if applicable). Sample housing units from the nonpermit universe identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce the variation in sample size for strata used in the sample selection of the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes built after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or by other sources). This procedure was designed to adjust the AHS sample estimates of these types of units to independently derived current estimates. These independent estimates were considered to be the best estimates available. The adjustment was necessary to correct known deficiencies in the AHS sample in the representation of these units (see section on nonsampling error). The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1975 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimates of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units	
New mobile homes	
"Other additions"	

The numerators of the ratios were derived by applying the following factors:

1. For the "conventional new construction units" cell, either a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in the SMSA, or the missed conventional new construction was estimated from the Survey of Construction (SOC). (The table that follows indicates the procedure used to estimate the missing conventional new construction for each SMSA.)
2. For the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied.
3. For the "other additions" cell, the "other additions" rate from SCARF[1] was applied.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

SMSA's employing the SOC estimate of missed conventional new construction	SMSA's employing the national trend for missed conventional new construction
Atlanta, Ga.	Colorado Springs, Colo.
Chicago, Ill.	Hartford, Conn.
Cincinnati, Ohio-Ky.-Ind.	Madison, Wis.
Columbus, Ohio	New Orleans, La.
Kansas City, Mo.-Kans.	Newport News—Hampton, Va.
Miami, Fla.	Paterson-Clifton-Passaic, N.J.
Milwaukee, Wis.	Rochester, N.Y.
Philadelphia, Pa.-N.J.	San Antonio, Tex.
Portland, Oreg.-Wash.	San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.	Springfield-Chicopee-Holyoke, Mass.-Conn.
San Francisco-Oakland, Calif.	

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

For some Year II SMSA's (i.e., Colorado Springs, Colo., Miami, Fla., and San Diego, Calif.) a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1975 housing inventory to an independent estimate of the SMSA's October 1975 housing inventory. This ratio estimate factor equaled the following:

$$\frac{\text{Independent estimate of the October 1975 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using the 1970 census estimate of total housing units for the SMSA in conjunction with the estimate of the change in the housing inventory since the census. The estimate of change was based on either administrative records from utility companies, or, when utility data were not available, on estimates of new construction permits and post-census demolition data. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, are brought into agreement with known estimates of the SMSA housing population.

The third-stage ratio estimation procedure was not employed in all Year II SMSA's since the reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. The independent estimates were only employed in SMSA's where the estimated relative bias was thought to be low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results. A measure of the relative bias of the independent estimate was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1975 independent estimates. In some SMSA's, the measure of the relative bias of the independent estimate was less than 3 percent but the independent estimate was still not employed, since there was evidence that suggested that the procedure for generating the independent estimates was producing substantial overestimates for these SMSA's.

1970-1975 lost units.—The AHS estimate of characteristics of the 1970-1975 lost units employed a one-stage ratio estimation procedure which is similar to the first-stage ratio estimation procedure mentioned above. The 1970-1975 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units + Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for State, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and non-

sampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

Nonsampling errors.—Nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1975 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors associated with 1970 census estimates—"coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

AHS-SMSA.—For the 1975 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are not yet available, however, it is believed that they will be similar to the results of the Year I AHS-SMSA reinterview study which were presented in the following Census Bureau memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample: 1974."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year I SMSA's and not for any specific SMSA):

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. The estimated indexes of inconsistency for most of the reinterview items were in the 20 to 50 range. The 20-50 interval is considered moderate indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview program because they had not been previously tested.

3. "Our bias indicator, the net difference rate, shows no signs of serious bias. For only one characteristic was there as much as a 2½ percent probability that the net difference rate is greater than 10 percent."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors.—With errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies in the representation of both conventional new construction and new mobile homes (and trailers) in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later, were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and

therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 1,500 conventional new construction units in this SMSA had permits issued prior to January 1970, and therefore, were missed by the 1975 AHS-SMSA survey.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. Although it is not known exactly, an estimated 500 new mobile homes in permit-issuing areas were missed by the 1975 AHS-SMSA survey in this SMSA. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS data still exists.

Rounding errors.--The rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling error, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in tables I and II are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1975 Housing Inventory and for Estimated Number of 1970-1975 Lost Units for the Madison, Wis., SMSA

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100. . . .	50	5,000 . .	350
200. . . .	70	10,000. .	480
500. . . .	110	25,000. .	710
700. . . .	130	50,000. .	870
1,000 . .	160	100,000.	690
2,500 . .	250		

Tables I and II present the standard errors applicable to estimates of characteristics of the 1975 housing inventory as well as estimates of characteristics of the 1970-1975 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in these tables.

For ratios, (100) (x/y), where x is not a subclass of y, table II underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be ob-

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the Madison, Wis., SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
100	5.0	11.0	15.1	21.8	25.1
200	3.5	7.7	10.7	15.4	17.8
500	2.2	4.9	6.7	9.7	11.2
700	1.9	4.1	5.7	8.2	9.5
1,000	1.6	3.5	4.8	6.9	7.9
2,500	1.0	2.2	3.0	4.4	5.0
5,000	.7	1.5	2.1	3.1	3.6
10,000	.5	1.0	1.5	2.2	2.5
25,000	.3	.7	1.0	1.4	1.6
50,000	.2	.5	.7	1.0	1.1
100,000	.2	.3	.5	.7	.8

tained by letting the standard error of the ratio be approximately equal to:

$$(100)\ (x/y)\ \sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

Let: x = the numerator of the ratio
y = the denominator of the ratio
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

Illustration of the use of the standard error tables.—Table 1 of part A of this report shows that in 1975 there were 57,500 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 840. Consequently, the 68-percent confidence interval, as shown by these data, is from 56,660 to 58,340 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 56,160 to 58,840 housing units with 90 percent confidence; and that the average estimate lies within the interval from 55,820 to 59,180 housing units with 95 percent confidence.

Table 1 of part A also shows that of the 57,500 owner-occupied housing units 11,600, or 20.2 percent, had two bedrooms. Interpolation in table II of this appendix shows that the standard error of the percent is approximately .9 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 19.3 to 21.1 percent; the 90-percent confidence interval is from 18.8 to 21.6 percent; and the 95-percent confidence interval is from 18.4 to 22.0 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; while, if there is a high negative correlation, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table 1 of part A of this report shows that in 1975 there were 30,300 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 18,700. Table I shows the standard error of 11,600 is approximately 500 and the standard error of 30,300 is approximately 740. Therefore, the standard error of the estimated difference of 18,700 is about

$$890 = \sqrt{(500)^2 + (740)^2}$$

Consequently, the 68-percent confidence interval for the 18,700 difference is from 17,810 to 19,590 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 17,280 to 20,120 housing units, and the 95-percent confidence interval is from 16,920 to 20,480. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval does not include zero or negative values.

Medians.—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the

average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From table II determine the standard error of a 50-percent characteristic on the base of the median;
2. add to and subtract from 50 percent, the standard error determined in step 1; and
3. using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval of a median.—Table 1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.0. The base of the distribution from which this median was determined is 57,500 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 57,500 is approximately 1.1 percentage points.
2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 47.8 and 52.2.
3. From table 1 of part A, it can be seen by cumulating the frequencies for the first two categories that 23,600 owner-occupied housing units, or 41.0 percent, had one or two persons (for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 10,100 owner-occupied housing units, or 17.6 percent, had three persons (i.e., 2.5 to 3.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{47.8-41.0}{17.6}\right) = 2.9$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{52.2-41.0}{17.6}\right) = 3.1$$

Thus, the 95-percent confidence interval ranges from 2.9 to 3.1 persons.

Miami 1975

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-41
Annual Housing Survey App-41
Selection of the sample. App-41
Building loss sample selection . . . App-42
1970 Census of Population and Housing. App-43
ESTIMATION. App-43
1975 housing inventory App-43
1970-1975 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing. App-44
RELIABILITY OF THE ESTIMATES. App-44
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-45
Rounding errors. App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences. App-47
Illustration of the computation of the standard error of a difference. App-47
Medians App-47
Illustration of the computation of the 95-percent confidence interval of a median App-48

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 21 SMSA's are based on data collected from the 1975 Annual Housing Survey (AHS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development. In each of the 21 SMSA's, the data were collected for the 12-month period from April 1975 through March 1976, with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and a third group of 20 SMSA's were collected from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated for the AHS every three years on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the second group (Year II SMSA's) are: Atlanta, Ga., Chicago, Ill., Philadelphia, Pa.-N.J., and San Francisco-Oakland, Calif.

The remaining SMSA's in the second group are: Cincinnati, Ohio-Ky.-Ind., Colorado Springs, Colo., Columbus, Ohio, Hartford, Conn., Kansas City, Mo.-Kans., Miami, Fla., Milwaukee, Wis., New Orleans, La., Newport News-Hampton, Va., Paterson-Clifton-Passaic, N.J., Portland, Oreg.-Wash., Rochester, N.Y., San Antonio, Tex., San Bernardino-Riverside-Ontario, Calif., San Diego, Calif., Springfield-Chicopee-Holyoke, Mass.-Conn., and Madison, Wis.

In this SMSA, 4,537 units were eligible for interview. Of these sample units, 216 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 473 units were visited but found not to be eligible for interview, because these units were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit issuing was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for those SMSA's which are *not* 100-percent permit issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). The following table indicates which SMSA's are 100-percent permit issuing and which contain a sample from the nonpermit universe.

100 Percent permit-issuing SMSA's	SMSA's with a sample from the nonpermit universe
Hartford, Conn.	Atlanta, Ga.
Miami, Fla.	Chicago, Ill.
Newport News-Hampton, Va.	Cincinnati, Ohio-Ky.-Ind.
Paterson-Clifton-Passaic, N.J.	Colorado Springs, Colo.
San Bernardino-Riverside-Ontario, Calif.	Columbus, Ohio
San Diego, Calif.	Kansas City, Mo.-Kans.
San Francisco-Oakland, Calif.	Madison, Wis.
	Milwaukee, Wis.
	New Orleans, La.
	Philadelphia, Pa.-N.J.
	Portland, Oreg.-Wash.
	Rochester, N.Y.
	San Antonio, Tex.
	Springfield-Chicopee-Holyoke, Mass.-Conn.

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of these sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and the balance of the SMSA, since the

sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—family size	Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000—$5,999 . .		
$6,000—$9,999 . .		
$10,000—$14,999		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. The housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued, and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \frac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Some tables in this report provide estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at enumeration time (i.e., the 1975 housing inventory), and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1975 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1975 housing inventory.—The AHS estimates of characteristics of the 1975 housing inventory employed a three-stage ratio estimation procedure. However, the third-stage ratio estimation procedure was employed in only three SMSA's. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 216 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units and one noninterview cell for the nonpermit universe (if applicable). Sample housing units from the nonpermit universe identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce the variation in sample size for strata used in the sample selection of the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes built after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or by other sources). This procedure was designed to adjust the AHS sample estimates of these types of units to independently derived current estimates. These independent estimates were considered to be the best estimates available. The adjustment was necessary to correct known deficiencies in the AHS sample in the representation of these units (see section on nonsampling error). The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1975 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimates of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units.	
New mobile homes	
"Other additions".	

The numerators of the ratios were derived by applying the following factors:

1. For the "conventional new construction units" cell, either a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in the SMSA, or the missed conventional new construction was estimated from the Survey of Construction (SOC). (The table that follows indicates the procedure used to estimate the missing conventional new construction for each SMSA.)
2. For the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied.
3. For the "other additions" cell, the "other additions" rate from SCARF[1] was applied.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

SMSA's employing the SOC estimate of missed conventional new construction	SMSA's employing the national trend for missed conventional new construction
Atlanta, Ga.	Colorado Springs, Colo.
Chicago, Ill.	Hartford, Conn.
Cincinnati, Ohio-Ky.-Ind.	Madison, Wis.
Columbus, Ohio	New Orleans, La.
Kansas City, Mo.-Kans.	Newport News—Hampton, Va.
Miami, Fla.	Paterson-Clifton-Passaic, N.J.
Milwaukee, Wis.	Rochester, N.Y.
Philadelphia, Pa.-N.J.	San Antonio, Tex.
Portland, Oreg.-Wash.	San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.	Springfield-Chicopee-Holyoke, Mass.-Conn.
San Francisco-Oakland, Calif.	

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

For some Year II SMSA's (i.e., Colorado Springs, Colo., Miami, Fla., and San Diego, Calif.) a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1975 housing inventory to an independent estimate of the SMSA's October 1975 housing inventory. This ratio estimate factor equaled the following:

$$\frac{\text{Independent estimate of the October 1975 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using the 1970 census estimate of total housing units for the SMSA in conjunction with the estimate of the change in the housing inventory since the census. The estimate of change was based on either administrative records from utility companies, or, when utility data were not available, on estimates of new construction permits and post-census demolition data. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, are brought into agreement with known estimates of the SMSA housing population.

The third-stage ratio estimation procedure was not employed in all Year II SMSA's since the reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. The independent estimates were only employed in SMSA's where the estimated relative bias was thought to be low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results. A measure of the relative bias of the independent estimate was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1975 independent estimates. In some SMSA's, the measure of the relative bias of the independent estimate was less than 3 percent but the independent estimate was still not employed, since there was evidence that suggested that the procedure for generating the independent estimates was producing substantial overestimates for these SMSA's.

1970-1975 lost units.—The AHS estimate of characteristics of the 1970-1975 lost units employed a one-stage ratio estimation procedure which is similar to the first-stage ratio estimation procedure mentioned above. The 1970-1975 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units + Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for State, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and non-

sampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

Nonsampling errors.—Nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1975 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors associated with 1970 census estimates—"coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

AHS-SMSA.—For the 1975 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are not yet available, however, it is believed that they will be similar to the results of the Year I AHS-SMSA reinterview study which were presented in the following Census Bureau memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample: 1974."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year I SMSA's and not for any specific SMSA):

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. The estimated indexes of inconsistency for most of the reinterview items were in the 20 to 50 range. The 20-50 interval is considered moderate indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview program because they had not been previously tested.

3. "Our bias indicator, the net difference rate, shows no signs of serious bias. For only one characteristic was there as much as a 2½ percent probability that the net difference rate is greater than 10 percent."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors.—With errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies in the representation of both conventional new construction and new mobile homes (and trailers) in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later, were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and

therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 17,100 conventional new construction units in this SMSA had permits issued prior to January 1970, and therefore, were missed by the 1975 AHS-SMSA survey.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. Although it is not known exactly, an estimated 5,300 new mobile homes in permit-issuing areas were missed by the 1975 AHS-SMSA survey in this SMSA. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS data still exists.

Rounding errors.—The rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling error, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in tables I and II are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1975 Housing Inventory and for Estimated Number of 1970-1975 Lost Units for the Miami, Fla., SMSA

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100. . . .	120	10,000 .	1,240
200. . . .	180	25,000 .	1,930
500. . . .	280	50,000 .	2,670
700. . . .	330	100,000	3,590
1,000 . .	390	250,000	4,700
2,500 . .	620	400,000	4,380
5,000 . .	880	500,000	3,230

Tables I and II present the standard errors applicable to estimates of characteristics of the 1975 housing inventory as well as estimates of characteristics of the 1970-1975 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in these tables.

For ratios, (100) (x/y), where x is not a subclass of y, table II underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approxi-

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the Miami, Fla., SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
100	12.4	27.1	37.4	54.0	62.4
200	8.8	19.2	26.5	38.2	44.1
500	5.6	12.1	16.7	24.2	27.9
700	4.7	10.2	14.2	20.4	23.6
1,000	3.9	8.6	11.8	17.1	19.7
2,500	2.5	5.4	7.5	10.0	12.5
5,000	1.8	3.8	5.3	7.6	8.8
10,000	1.2	2.7	3.7	5.4	6.2
25,000	.8	1.7	2.4	3.4	3.9
50,000	.6	1.2	1.7	2.4	2.8
100,000	.4	.9	1.2	1.7	2.0
250,000	.2	.5	.7	1.1	1.2
400,000	.2	.4	.6	.9	1.0
500,000	.2	.4	.5	.8	.9

mation of the standard error may be obtained by letting the standard error of the ratio be approximately equal to:

$$(100)\,(x/y)\,\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

Let: x = the numerator of the ratio

y = the denominator of the ratio

σ_x = the standard error of the numerator

σ_y = the standard error of the denominator

Illustration of the use of the standard error tables.—Table 1 of part A of this report shows that in 1975 there were 282,700 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 4,630. Consequently, the 68-percent confidence interval, as shown by these data, is from 278,070 to 287,330 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 275,290 to 290,110 housing units with 90 percent confidence; and that the average estimate lies within the interval from 273,440 to 291,960 housing units with 95 percent confidence.

Table 1 of part A also shows that of the 282,700 owner-occupied housing units 88,000, or 31.1 percent, had two bedrooms. Interpolation in table II of this appendix shows that the standard error of the percent is approximately 1.1 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 30.0 to 32.2 percent; the 90-percent confidence interval is from 29.3 to 32.9 percent; and the 95-percent confidence interval is from 28.9 to 33.3 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; while, if there is a high negative correlation, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table 1 of part A of this report shows that in 1975 there were 132,600 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 44,600. Table I shows the standard error of 88,000 is approximately 3,370 and the standard error of 132,600 is approximately 3,830. Therefore, the standard error of the estimated difference of 44,600 is about

$$5{,}100 = \sqrt{(3{,}370)^2 + (3{,}830)^2}$$

Consequently, the 68-percent confidence interval for the 44,600 difference is from 39,500 to 49,700 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 36,440 to 52,760 housing units, and the 95-percent confidence interval is from 34,400 to 54,800. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval does not include zero or negative values.

Medians.—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following

procedure may be used to estimate confidence limits of a median based on sample data:

1. From table II determine the standard error of a 50-percent characteristic on the base of the median;
2. add to and subtract from 50 percent, the standard error determined in step 1; and
3. using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median.—Table 1 of part A of this report shows the median number of persons for owner-occupied housing units is 2.7. The base of the distribution from which this median was determined is 282,700 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 282,700 is approximately 1.2 percentage points.
2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 47.6 and 52.4.
3. From table 1 of part A, it can be seen by cumulating the frequencies for the first two categories that 131,200 owner-occupied housing units, or 46.4 percent, had one or two persons (for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 52,800 owner-occupied housing units, or 18.7 percent, had three persons (i.e., 2.5 to 3.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{47.6-46.4}{18.7}\right) = 2.6$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{52.4-46.4}{18.7}\right) = 2.8$$

Thus, the 95-percent confidence interval ranges from 2.6 to 2.8 persons.

Milwaukee 1975

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-41
Annual Housing Survey App-41
Selection of the sample. App-41
Building loss sample selection . . . App-42
1970 Census of Population and Housing. App-43
ESTIMATION. App-43
1975 housing inventory App-43
1970-1975 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing. App-44
RELIABILITY OF THE ESTIMATES. App-44
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-45
Rounding errors. App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences. App-47
Illustration of the computation of the standard error of a difference. App-47
Medians App-47
Illustration of the computation of the 95-percent confidence interval of a median App-48

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 21 SMSA's are based on data collected from the 1975 Annual Housing Survey (AHS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development. In each of the 21 SMSA's, the data were collected for the 12-month period from April 1975 through March 1976, with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and a third group of 20 SMSA's were collected from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated for the AHS every three years on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the second group (Year II SMSA's) are: Atlanta, Ga., Chicago, Ill., Philadelphia, Pa.-N.J., and San Francisco-Oakland, Calif.

The remaining SMSA's in the second group are: Cincinnati, Ohio-Ky.-Ind., Colorado Springs, Colo., Columbus, Ohio, Hartford, Conn., Kansas City, Mo.-Kans., Miami, Fla., Milwaukee, Wis., New Orleans, La., Newport News-Hampton, Va., Paterson-Clifton-Passaic, N.J., Portland, Oreg.-Wash., Rochester, N.Y., San Antonio, Tex., San Bernardino-Riverside-Ontario, Calif., San Diego, Calif., Springfield-Chicopee-Holyoke, Mass.-Conn., and Madison, Wis.

In this SMSA, 4,931 units were eligible for interview. Of these sample units, 204 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 298 units were visited but found not to be eligible for interview, because these units were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit issuing was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for those SMSA's which are *not* 100-percent permit issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). The following table indicates which SMSA's are 100-percent permit issuing and which contain a sample from the nonpermit universe.

100 Percent permit-issuing SMSA's	SMSA's with a sample from the nonpermit universe
Hartford, Conn.	Atlanta, Ga.
Miami, Fla.	Chicago, Ill.
Newport News-Hampton, Va.	Cincinnati, Ohio-Ky.-Ind.
Paterson-Clifton-Passaic, N.J.	Colorado Springs, Colo.
San Bernardino-Riverside-Ontario, Calif.	Columbus, Ohio
San Diego, Calif.	Kansas City, Mo.-Kans.
San Francisco-Oakland, Calif.	Madison, Wis.
	Milwaukee, Wis.
	New Orleans, La.
	Philadelphia, Pa.-N.J.
	Portland, Oreg.-Wash.
	Rochester, N.Y.
	San Antonio, Tex.
	Springfield-Chicopee-Holyoke, Mass.-Conn.

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of these sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and the balance of the SMSA, since the

sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—family size	Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000—$5,999 . .		
$6,000—$9,999 . .		
$10,000—$14,999		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. The housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued, and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED}}{4} + \frac{\text{Group quarters population in 1970 census ED}}{3}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in the sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Some tables in this report provide estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory), and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at enumeration time (i.e., the 1975 housing inventory), and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1975 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1975 housing inventory.—The AHS estimates of characteristics of the 1975 housing inventory employed a three-stage ratio estimation procedure. However, the third-stage ratio estimation procedure was employed in only three SMSA's. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 204 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units and one noninterview cell for the nonpermit universe (if applicable). Sample housing units from the nonpermit universe identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce the variation in sample size for strata used in the sample selection of the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes built after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or by other sources). This procedure was designed to adjust the AHS sample estimates of these types of units to independently derived current estimates. These independent estimates were considered to be the best estimates available. The adjustment was necessary to correct known deficiencies in the AHS sample in the representation of these units (see section on nonsampling error). The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1975 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimates of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units	
New mobile homes	
"Other additions"	

The numerators of the ratios were derived by applying the following factors:

1. For the "conventional new construction units" cell, either a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in the SMSA, or the missed conventional new construction was estimated from the Survey of Construction (SOC). (The table that follows indicates the procedure used to estimate the missing conventional new construction for each SMSA.)
2. For the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied.
3. For the "other additions" cell, the "other additions" rate from SCARF[1] was applied.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

SMSA's employing the SOC estimate of missed conventional new construction	SMSA's employing the national trend for missed conventional new construction
Atlanta, Ga.	Colorado Springs, Colo.
Chicago, Ill.	Hartford, Conn.
Cincinnati, Ohio-Ky.-Ind.	Madison, Wis.
Columbus, Ohio	New Orleans, La.
Kansas City, Mo.-Kans.	Newport News—Hampton, Va.
Miami, Fla.	Paterson-Clifton-Passaic, N.J.
Milwaukee, Wis.	Rochester, N.Y.
Philadelphia, Pa.-N.J.	San Antonio, Tex.
Portland, Oreg.-Wash.	San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.	Springfield-Chicopee-Holyoke, Mass.-Conn.
San Francisco-Oakland, Calif.	

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

For some Year II SMSA's (i.e., Colorado Springs, Colo., Miami, Fla., and San Diego, Calif.) a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1975 housing inventory to an independent estimate of the SMSA's October 1975 housing inventory. This ratio estimate factor equaled the following:

$$\frac{\text{Independent estimate of the October 1975 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using the 1970 census estimate of total housing units for the SMSA in conjunction with the estimate of the change in the housing inventory since the census. The estimate of change was based on either administrative records from utility companies, or, when utility data were not available, on estimates of new construction permits and post-census demolition data. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, are brought into agreement with known estimates of the SMSA housing population.

The third-stage ratio estimation procedure was not employed in all Year II SMSA's since the reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. The independent estimates were only employed in SMSA's where the estimated relative bias was thought to be low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results. A measure of the relative bias of the independent estimate was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1975 independent estimates. In some SMSA's, the measure of the relative bias of the independent estimate was less than 3 percent but the independent estimate was still not employed, since there was evidence that suggested that the procedure for generating the independent estimates was producing substantial overestimates for these SMSA's.

1970-1975 lost units.—The AHS estimate of characteristics of the 1970-1975 lost units employed a one-stage ratio estimation procedure which is similar to the first-stage ratio estimation procedure mentioned above. The 1970-1975 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for State, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and non-

sampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

Nonsampling errors.—Nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1975 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors associated with 1970 census estimates—"coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

AHS-SMSA.—For the 1975 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are not yet available, however, it is believed that they will be similar to the results of the Year I AHS-SMSA reinterview study which were presented in the following Census Bureau memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample: 1974."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year I SMSA's and not for any specific SMSA):

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. The estimated indexes of inconsistency for most of the reinterview items were in the 20 to 50 range. The 20-50 interval is considered moderate indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview program because they had not been previously tested.

3. "Our bias indicator, the net difference rate, shows no signs of serious bias. For only one characteristic was there as much as a 2½ percent probability that the net difference rate is greater than 10 percent."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors.—With errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies in the representation of both conventional new construction and new mobile homes (and trailers) in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later, were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and

therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 4,000 conventional new construction units in this SMSA had permits issued prior to January 1970, and therefore, were missed by the 1975 AHS-SMSA survey.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. Although it is not known exactly, an estimated 1,000 new mobile homes in permit-issuing areas were missed by the 1975 AHS-SMSA survey in this SMSA. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS data still exists.

Rounding errors.--The rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling error, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in tables I and II are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1975 Housing Inventory and for Estimated Number of 1970-1975 Lost Units for the Milwaukee, Wis., SMSA

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	110	10,000 . .	1,060
200	150	25,000 . .	1,650
500	240	50,000 . .	2,270
700	280	100,000. .	3,050
1,000 . .	340	250,000. .	3,920
2,500 . .	530	400,000. .	3,470
5,000 . .	750		

Tables I and II present the standard errors applicable to estimates of characteristics of the 1975 housing inventory as well as estimates of characteristics of the 1970-1975 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in these tables.

For ratios, (100) (x/y), where x is not a subclass of y, table II underestimates the standard error of the ratio when there is little or no correlation between x and

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the Milwaukee, Wis., SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
100	10.6	23.2	32.0	46.2	53.3
200	7.5	16.4	22.6	32.7	37.7
500	4.7	10.4	14.3	20.7	23.8
700	4.0	8.8	12.1	17.5	20.2
1,000	3.4	7.3	10.1	14.6	16.9
2,500	2.1	4.6	6.4	9.2	10.7
5,000	1.5	3.3	4.5	6.5	7.5
10,000	1.1	2.3	3.2	4.6	5.3
25,000	.7	1.5	2.0	2.9	3.4
50,000	.5	1.0	1.4	2.1	2.4
100,000	.3	.7	1.0	1.5	1.7
250,000	.2	.5	.6	.9	1.1
400,000	.2	.4	.5	.7	.8

y. For this type of ratio, a better approximation of the standard error may be obtained by letting the standard error of the ratio be approximately equal to:

$$(100)\ (x/y)\ \sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

Let: x = the numerator of the ratio
y = the denominator of the ratio
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

Illustration of the use of the standard error tables.—Table 1 of part A of this report shows that in 1975 there were 286,400 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 3,810. Consequently, the 68-percent confidence interval, as shown by these data, is from 282,590 to 290,210 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 280,300 to 292,500 housing units with 90 percent confidence; and that the average estimate lies within the interval from 278,780 to 294,020 housing units with 95 percent confidence.

Table 1 of part A also shows that of the 286,400 owner-occupied housing units, 64,000, or 22.3 percent, had two bedrooms. Interpolation in table II of this appendix shows that the standard error of the percent is approximately .8 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 21.5 to 23.1 percent; the 90-percent confidence interval is from 21.0 to 23.6 percent; and the 95-percent confidence interval is from 20.7 to 23.9 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; while, if there is a high negative correlation, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table 1 of part A of this report shows that in 1975 there were 157,900 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 93,900. Table I shows the standard error of 64,000 is approximately 2,490 and the standard error of 157,900 is approximately 3,390. Therefore, the standard error of the estimated difference of 93,900 is about

$$4{,}210 = \sqrt{(2{,}490)^2 + (3{,}390)^2}$$

Consequently, the 68-percent confidence interval for the 93,900 difference is from 89,690 to 98,110 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lying within a range computed in this way would be correct for roughly 68 percent of all possible samples: Similarly, the 90-percent confidence interval is from 87,160 to 100,640 housing units, and the 95-percent confidence interval is from 85,480 to 102,320. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval does not include zero or negative values.

Medians.—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from

all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From table II determine the standard error of a 50-percent characteristic on the base of the median;
2. add to and subtract from 50 percent, the standard error determined in step 1; and
3. using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median.—Table 1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.1. The base of the distribution from which this median was determined is 286,400 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 286,400 is approximately 1.0 percentage points.
2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 48.0 and 52.0.
3. From table 1 of part A, it can be seen by cumulating the frequencies for the first two categories that 111,800 owner-occupied housing units or 39.0 percent, had one or two persons (for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 49,100 owner-occupied housing units, or 17.1 percent, had three persons (i.e., 2.5 to 3.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{48.0-39.0}{17.1}\right) = 3.0$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{52.0-39.0}{17.1}\right) = 3.3$$

Thus, the 95-percent confidence interval ranges from 3.0 to 3.3 persons.

New Orleans 1975

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-41
Annual Housing Survey App-41
Selection of the sample. App-41
Building loss sample selection . . . App-42
1970 Census of Population and Housing. App-43
ESTIMATION. App-43
1975 housing inventory App-43
1970-1975 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing. App-44
RELIABILITY OF THE ESTIMATES. App-44
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-45
Rounding errors. App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences. App-47
Illustration of the computation of the standard error of a difference. App-47
Medians App-47
Illustration of the computation of the 95-percent confidence interval of a median App-48

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 21 SMSA's are based on data collected from the 1975 Annual Housing Survey (AHS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development. In each of the 21 SMSA's, the data were collected for the 12-month period from April 1975 through March 1976, with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and a third group of 20 SMSA's were collected from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated for the AHS every three years on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the second group (Year II SMSA's) are: Atlanta, Ga., Chicago, Ill., Philadelphia, Pa.-N.J., and San Francisco-Oakland, Calif.

The remaining SMSA's in the second group are: Cincinnati, Ohio-Ky.-Ind., Colorado Springs, Colo., Columbus, Ohio, Hartford, Conn., Kansas City, Mo.-Kans., Miami, Fla., Milwaukee, Wis., New Orleans, La., Newport News-Hampton, Va., Paterson-Clifton-Passaic, N.J., Portland, Oreg.-Wash., Rochester, N.Y., San Antonio, Tex., San Bernardino-Riverside-Ontario, Calif., San Diego, Calif., Springfield-Chicopee-Holyoke, Mass.-Conn., and Madison, Wis.

In this SMSA, 4,746 units were eligible for interview. Of these sample units, 231 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 422 units were visited but found not to be eligible for interview, because these units were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit issuing was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for those SMSA's which are *not* 100-percent permit issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). The following table indicates which SMSA's are 100-percent permit issuing and which contain a sample from the nonpermit universe.

100 Percent permit-issuing SMSA's	SMSA's with a sample from the nonpermit universe
Hartford, Conn.	Atlanta, Ga.
Miami, Fla.	Chicago, Ill.
Newport News-Hampton, Va.	Cincinnati, Ohio-Ky.-Ind.
Paterson-Clifton-Passaic, N.J.	Colorado Springs, Colo.
San Bernardino-Riverside-Ontario, Calif.	Columbus, Ohio
San Diego, Calif.	Kansas City, Mo.-Kans.
San Francisco-Oakland, Calif.	Madison, Wis.
	Milwaukee, Wis.
	New Orleans, La.
	Philadelphia, Pa.-N.J.
	Portland, Oreg.-Wash.
	Rochester, N.Y.
	San Antonio, Tex.
	Springfield-Chicopee-Holyoke, Mass.-Conn.

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of these sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and the balance of the SMSA, since the

sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—family size	Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000—$5,999		
$6,000—$9,999		
$10,000—$14,999		
$15,000 and over		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. The housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued, and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \frac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four or a multiple of four housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Some tables in this report provide estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at enumeration time (i.e., the 1975 housing inventory), and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1975 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1975 housing inventory.—The AHS estimates of characteristics of the 1975 housing inventory employed a three-stage ratio estimation procedure. However, the third-stage ratio estimation procedure was employed in only three SMSA's. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 231 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units and one noninterview cell for the nonpermit universe (if applicable). Sample housing units from the nonpermit universe identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce the variation in sample size for strata used in the sample selection of the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes built after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or by other sources). This procedure was designed to adjust the AHS sample estimates of these types of units to independently derived current estimates. These independent estimates were considered to be the best estimates available. The adjustment was necessary to correct known deficiencies in the AHS sample in the representation of these units (see section on nonsampling error). The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1975 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimates of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units	
New mobile homes	
"Other additions"	

The numerators of the ratios were derived by applying the following factors:

1. For the "conventional new construction units" cell, either a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in the SMSA, or the missed conventional new construction was estimated from the Survey of Construction (SOC). (The table that follows indicates the procedure used to estimate the missing conventional new construction for each SMSA.)
2. For the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied.
3. For the "other additions" cell, the "other additions" rate from SCARF[1] was applied.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

SMSA's employing the SOC estimate of missed conventional new construction	SMSA's employing the national trend for missed conventional new construction
Atlanta, Ga.	Colorado Springs, Colo.
Chicago, Ill.	Hartford, Conn.
Cincinnati, Ohio-Ky.-Ind.	Madison, Wis.
Columbus, Ohio	New Orleans, La.
Kansas City, Mo.-Kans.	Newport News—Hampton, Va.
Miami, Fla.	Paterson-Clifton-Passaic, N.J.
Milwaukee, Wis.	Rochester, N.Y.
Philadelphia, Pa.-N.J.	San Antonio, Tex.
Portland, Oreg.-Wash.	San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.	Springfield-Chicopee-Holyoke, Mass.-Conn.
San Francisco-Oakland, Calif.	

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

For some Year II SMSA's (i.e., Colorado Springs, Colo., Miami, Fla., and San Diego, Calif.) a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1975 housing inventory to an independent estimate of the SMSA's October 1975 housing inventory. This ratio estimate factor equaled the following:

$$\frac{\text{Independent estimate of the October 1975 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using the 1970 census estimate of total housing units for the SMSA in conjunction with the estimate of the change in the housing inventory since the census. The estimate of change was based on either administrative records from utility companies, or, when utility data were not available, on estimates of new construction permits and post-census demolition data. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, are brought into agreement with known estimates of the SMSA housing population.

The third-stage ratio estimation procedure was not employed in all Year II SMSA's since the reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. The independent estimates were only employed in SMSA's where the estimated relative bias was thought to be low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results. A measure of the relative bias of the independent estimate was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1975 independent estimates. In some SMSA's, the measure of the relative bias of the independent estimate was less than 3 percent but the independent estimate was still not employed, since there was evidence that suggested that the procedure for generating the independent estimates was producing substantial overestimates for these SMSA's.

1970-1975 lost units.—The AHS estimate of characteristics of the 1970-1975 lost units employed a one-stage ratio estimation procedure which is similar to the first-stage ratio estimation procedure mentioned above. The 1970-1975 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units + Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and non-

sampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

Nonsampling errors.—Nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1975 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors associated with 1970 census estimates—"coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

AHS-SMSA.—For the 1975 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are not yet available, however, it is believed that they will be similar to the results of the Year I AHS-SMSA reinterview study which were presented in the following Census Bureau memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample: 1974."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year I SMSA's and not for any specific SMSA):

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. The estimated indexes of inconsistency for most of the reinterview items were in the 20 to 50 range. The 20-50 interval is considered moderate indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview program because they had not been previously tested.

3. "Our bias indicator, the net difference rate, shows no signs of serious bias. For only one characteristic was there as much as a 2½ percent probability that the net difference rate is greater than 10 percent."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors.—With errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies in the representation of both conventional new construction and new mobile homes (and trailers) in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later, were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and

therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 3,500 conventional new construction units in this SMSA had permits issued prior to January 1970, and therefore, were missed by the 1975 AHS-SMSA survey.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. Although it is not known exactly, an estimated 1,800 new mobile homes in permit-issuing areas were missed by the 1975 AHS-SMSA survey in this SMSA. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies; although some bias in the AHS data still exists.

Rounding errors.—The rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling error, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in tables I and II are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1975 Housing Inventory and for Estimated Number of 1970-1975 Lost Units for the New Orleans, La., SMSA

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	100	10,000 . .	960
200	140	25,000 . .	1,490
500	220	50,000 . .	2,050
700	260	100,000 .	2,710
1,000 . .	310	250,000 .	3,250
2,500 . .	490	400,000 .	2,090
5,000 . .	680		

Tables I and II present the standard errors applicable to estimates of characteristics of the 1975 housing inventory as well as estimates of characteristics of the 1970-1975 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in these tables.

For ratios, (100) (x/y), where x is not a subclass of y, table II underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approxi-

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the New Orleans, La., SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
100	9.7	21.2	29.2	42.1	48.6
200	6.8	15.0	20.6	29.8	34.4
500	4.3	9.5	13.0	18.8	21.7
700	3.7	8.0	11.0	15.9	18.4
1,000	3.1	6.7	9.2	13.3	15.4
2,500	1.9	4.2	5.8	8.4	9.7
5,000	1.4	3.0	4.1	6.0	6.9
10,000	1.0	2.1	2.9	4.2	4.9
25,000	.6	1.3	1.8	2.7	3.1
50,000	.4	.9	1.3	1.9	2.2
100,000	.3	.7	.9	1.3	1.5
250,000	.2	.4	.6	.8	1.0
400,000	.2	.3	.5	.7	.8

mation of the standard error may be obtained by letting the standard error of the ratio be approximately equal to:

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2+\left(\frac{\sigma_y}{y}\right)^2}$$

Let: x = the numerator of the ratio
y = the denominator of the ratio
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

Illustration of the use of the standard error tables.—Table 1 of part A of this report shows that in 1975 there were 186,200 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 3,020. Consequently, the 68-percent confidence interval, as shown by these data, is from 183,180 to 189,220 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 181,370 to 191,030 housing units with 90 percent confidence; and that the average estimate lies within the interval from 180,160 to 192,240 housing units with 95 percent confidence.

Table 1 of part A also shows that of the 186,200 owner-occupied housing units 48,100, or 25.8 percent, had two bedrooms. Interpolation in table II of this appendix shows that the standard error of the percent is approximately 1.0 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 24.8 to 26.8 percent; the 90-percent confidence interval is from 24.2 to 27.4 percent; and the 95-percent confidence interval is from 23.8 to 27.8 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; while, if there is a high negative correlation, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table 1 of part A of this report shows that in 1975 there were 94,500 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 46,400. Table I shows the standard error of 48,100 is approximately 2,010 and the standard error of 94,500 is approximately 2,640. Therefore, the standard error of the estimated difference of 46,400 is about

$$3{,}320 = \sqrt{(2{,}010)^2 + (2{,}640)^2}$$

Consequently, the 68-percent confidence interval for the 46,400 difference is from 43,080 to 49,720 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 41,090 to 51,710 housing units, and the 95-percent confidence interval is from 39,760 to 53,040. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval does not include zero or negative values.

Medians.—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from

all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From table II determine the standard error of a 50-percent characteristic on the base of the median,
2. add to and subtract from 50 percent the standard error determined in step 1; and
3. using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median.—Table 1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.1. The base of the distribution from which this median was determined is 186,200 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 186,200 is approximately 1.2 percentage points.
2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 47.6 and 52.4.
3. From table 1 of part A, it can be seen by cumulating the frequencies for the first two categories that 70,700 owner-occupied housing units, or 38.0 percent, had one or two persons (for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 35,600 owner-occupied housing units, or 19.1 percent, had three persons (i.e., 2.5 to 3.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{47.6-38.0}{19.1}\right) = 3.0$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{52.4-38.0}{19.1}\right) = 3.3$$

Thus, the 95-percent confidence interval ranges from 3.0 to 3.3 persons.

Newport News - Hampton 1975

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-41
Annual Housing Survey App-41
Selection of the sample. App-41
Building loss sample selection . . . App-42
1970 Census of Population and Housing. App-43
ESTIMATION. App-43
1975 housing inventory App-43
1970-1975 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing. App-44
RELIABILITY OF THE ESTIMATES. App-44
Nonsampling errors. App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-45
Rounding errors. App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences. App-47
Illustration of the computation of the standard error of a difference. App-47
Medians App-47
Illustration of the computation of the 95-percent confidence interval of a median App-48

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 21 SMSA's are based on data collected from the 1975 Annual Housing Survey (AHS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development. In each of the 21 SMSA's, the data were collected for the 12-month period from April 1975 through March 1976, with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and a third group of 20 SMSA's were collected from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated for the AHS every three years on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the second group (Year II SMSA's) are: Atlanta, Ga., Chicago, Ill., Philadelphia, Pa.-N.J., and San Francisco-Oakland, Calif.

The remaining SMSA's in the second group are: Cincinnati, Ohio-Ky.-Ind., Colorado Springs, Colo., Columbus, Ohio, Hartford, Conn., Kansas City, Mo.-Kans., Miami, Fla., Milwaukee, Wis., New Orleans, La., Newport News-Hampton, Va., Paterson-Clifton-Passaic, N.J., Portland, Oreg.-Wash., Rochester, N.Y., San Antonio, Tex., San Bernardino-Riverside-Ontario, Calif., San Diego, Calif., Springfield-Chicopee-Holyoke, Mass.-Conn., and Madison, Wis.

In this SMSA, 4,105 units were eligible for interview. Of these sample units, 118 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 492 units were visited but found not to be eligible for interview, because these units were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit issuing was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for those SMSA's which are *not* 100-percent permit issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). The following table indicates which SMSA's are 100-percent permit issuing and which contain a sample from the nonpermit universe.

100 Percent permit-issuing SMSA's	SMSA's with a sample from the nonpermit universe
Hartford, Conn.	Atlanta, Ga.
Miami, Fla.	Chicago, Ill.
Newport News-Hampton, Va.	Cincinnati, Ohio-Ky.-Ind.
Paterson-Clifton-Passaic, N.J.	Colorado Springs, Colo.
San Bernardino-Riverside-Ontario, Calif.	Columbus, Ohio
San Diego, Calif.	Kansas City, Mo.-Kans.
San Francisco-Oakland, Calif.	Madison, Wis.
	Milwaukee, Wis.
	New Orleans, La.
	Philadelphia, Pa.-N.J.
	Portland, Oreg.-Wash.
	Rochester, N.Y.
	San Antonio, Tex.
	Springfield-Chicopee-Holyoke, Mass.-Conn.

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of these sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and the balance of the SMSA, since the

sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—family size	Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000—$5,999 . .		
$6,000—$9,999 . .		
$10,000—$14,999		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. The housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued, and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \frac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Some tables in this report provide estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at enumeration time (i.e., the 1975 housing inventory), and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1975 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1975 housing inventory.—The AHS estimates of characteristics of the 1975 housing inventory employed a three-stage ratio estimation procedure. However, the third-stage ratio estimation procedure was employed in only three SMSA's. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 118 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units and one noninterview cell for the nonpermit universe (if applicable). Sample housing units from the nonpermit universe identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce the variation in sample size for strata used in the sample selection of the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes built after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or by other sources). This procedure was designed to adjust the AHS sample estimates of these types of units to independently derived current estimates. These independent estimates were considered to be the best estimates available. The adjustment was necessary to correct known deficiencies in the AHS sample in the representation of these units (see section on nonsampling error). The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1975 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimates of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units	
New mobile homes	
"Other additions"	

The numerators of the ratios were derived by applying the following factors:

1. For the "conventional new construction units" cell, either a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in the SMSA, or the missed conventional new construction was estimated from the Survey of Construction (SOC). (The table that follows indicates the procedure used to estimate the missing conventional new construction for each SMSA.)
2. For the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied.
3. For the "other additions" cell, the "other additions" rate from SCARF[1] was applied.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

SMSA's employing the SOC estimate of missed conventional new construction	SMSA's employing the national trend for missed conventional new construction
Atlanta, Ga.	Colorado Springs, Colo.
Chicago, Ill.	Hartford, Conn.
Cincinnati, Ohio-Ky.-Ind.	Madison, Wis.
Columbus, Ohio	New Orleans, La.
Kansas City, Mo.-Kans.	Newport News—Hampton, Va.
Miami, Fla.	Paterson-Clifton-Passaic, N.J.
Milwaukee, Wis.	Rochester, N.Y.
Philadelphia, Pa.-N.J.	San Antonio, Tex.
Portland, Oreg.-Wash.	San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.	Springfield-Chicopee-Holyoke, Mass.-Conn.
San Francisco-Oakland, Calif.	

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

For some Year II SMSA's (i.e., Colorado Springs, Colo., Miami, Fla., and San Diego, Calif.) a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1975 housing inventory to an independent estimate of the SMSA's October 1975 housing inventory. This ratio estimate factor equaled the following:

$$\frac{\text{Independent estimate of the October 1975 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using the 1970 census estimate of total housing units for the SMSA in conjunction with the estimate of the change in the housing inventory since the census. The estimate of change was based on either administrative records from utility companies, or, when utility data were not available, on estimates of new construction permits and post-census demolition data. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, are brought into agreement with known estimates of the SMSA housing population.

The third-stage ratio estimation procedure was not employed in all Year II SMSA's since the reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. The independent estimates were only employed in SMSA's where the estimated relative bias was thought to be low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results. A measure of the relative bias of the independent estimate was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1975 independent estimates. In some SMSA's, the measure of the relative bias of the independent estimate was less than 3 percent but the independent estimate was still not employed, since there was evidence that suggested that the procedure for generating the independent estimates was producing substantial overestimates for these SMSA's.

1970-1975 lost units.—The AHS estimate of characteristics of the 1970-1975 lost units employed a one-stage ratio estimation procedure which is similar to the first-stage ratio estimation procedure mentioned above. The 1970-1975 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units + Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for State, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and non-

sampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

Nonsampling errors.—Nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1975 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors associated with 1970 census estimates—"coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

AHS-SMSA.—For the 1975 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are not yet available, however, it is believed that they will be similar to the results of the Year I AHS-SMSA reinterview study which were presented in the following Census Bureau memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample: 1974."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year I SMSA's and not for any specific SMSA):

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. The estimated indexes of inconsistency for most of the reinterview items were in the 20 to 50 range. The 20-50 interval is considered moderate indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview program because they had not been previously tested.

3. "Our bias indicator, the net difference rate, shows no signs of serious bias. For only one characteristic was there as much as a 2½ percent probability that the net difference rate is greater than 10 percent."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors.—With errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies in the representation of both conventional new construction and new mobile homes (and trailers) in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later, were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and

therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 1,400 conventional new construction units in this SMSA had permits issued prior to January 1970, and therefore, were missed by the 1975 AHS-SMSA survey.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. Although it is not known exactly, an estimated 800 new mobile homes in permit-issuing areas were missed by the 1975 AHS-SMSA survey in this SMSA. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS data still exists.

Rounding errors.—The rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling error, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in tables I and II are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1975 Housing Inventory and for Estimated Number of 1970-1975 Lost Units for the Newport News-Hampton, Va., SMSA

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100. . . .	50	5,000. . .	370
200. . . .	80	10,000. .	510
500. . . .	120	25,000. .	760
700. . . .	140	50,000. .	950
1,000 . .	170	100,000 .	900
2,500 . .	260		

Tables I and II present the standard errors applicable to estimates of characteristics of the 1975 housing inventory as well as estimates of characteristics of the 1970-1975 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in these tables.

For ratios, (100) (x/y), where x is not a subclass of y, table II underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be ob-

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for Newport News-Hampton, Va., SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
100	5.3	11.6	16.0	23.0	26.6
200	3.7	8.2	11.3	16.3	18.8
500	2.4	5.2	7.1	10.3	11.9
700	2.0	4.4	6.0	8.7	10.0
1,000	1.7	3.7	5.0	7.3	8.4
2,500	1.1	2.3	3.2	4.6	5.3
5,000	.7	1.6	2.3	3.3	3.8
10,000	.5	1.2	1.6	2.3	2.7
25,000	.3	.7	1.0	1.5	1.7
50,000	.2	.5	.7	1.0	1.2
100,000	.2	.4	.5	.7	.8

tained by letting the standard error of the ratio be approximately equal to:

$$(100)\,(x/y)\,\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

Let: x = the numerator of the ratio

y = the denominator of the ratio

σ_x = the standard error of the numerator

σ_y = the standard error of the denominator

Illustration of the use of the standard error tables.—Table 1 of part A of this report shows that in 1975 there were 54,500 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 950. Consequently, the 68-percent confidence interval, as shown by these data, is from 53,550 to 55,450 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 52,980 to 56,020 housing units with 90 percent confidence; and that the average estimate lies within the interval from 52,600 to 56,400 housing units with 95 percent confidence.

Table 1 of part A also shows that of the 54,500 owner-occupied housing units 10,300, or 18.9 percent, had two bedrooms. Interpolation in table II of this appendix shows that the standard error of the percent is approximately .9 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 18.0 to 19.8 percent; the 90-percent confidence interval is from 17.5 to 20.3 percent; and the 95-percent confidence interval is from 17.1 to 20.7 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; while, if there is a high negative correlation, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table 1 of part A of this report shows that in 1975 there were 31,600 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 21,300. Table I shows the standard error of 10,300 is approximately 520 and the standard error of 31,600 is approximately 810. Therefore, the standard error of the estimated difference of 21,300 is about

$$960 = \sqrt{(520)^2 + (810)^2}$$

Consequently, the 68-percent confidence interval for the 21,300 difference is from 20,340 to 22,260 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 19,760 to 22,840 housing units, and the 95-percent confidence interval is from 19,380 to 23,220. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval does not include zero or negative values.

Medians.—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From table II determine the standard error of a 50-percent characteristic on the base of the median;
2. add to and subtract from 50 percent, the standard error determined in step 1; and
3. using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For abour 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median.—Table 1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.0. The base of the distribution from which this median was determined is 54,500 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 54,500 is approximately 1.2 percentage points.
2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 47.6 and 52.4.
3. From table 1 of part A, it can be seen by cumulating the frequencies for the first two categories that 22,000 owner-occupied housing units, or 40.4 percent, had one or two persons (for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 11,200 owner-occupied, housing units, or 20.6 percent, had three persons (i.e., 2.5 to 3.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{47.6-40.4}{20.6}\right) = 2.8$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{52.4-40.4}{20.6}\right) = 3.1$$

Thus, the 95-percent confidence interval ranges from 2.8 to 3.1 persons.

Patterson-Clifton-Passaic 1975

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-41
Annual Housing Survey App-41
Selection of the sample. App-41
Building loss sample selection . . . App-42
1970 Census of Population and Housing. App-43
ESTIMATION. App-43
1975 housing inventory App-43
1970-1975 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing. App-44
RELIABILITY OF THE ESTIMATES. App-44
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-45
Rounding errors. App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences. App-47
Illustration of the computation of the standard error of a difference. App-47
Medians App-47
Illustration of the computation of the 95-percent confidence interval of a median App-48

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 21 SMSA's are based on data collected from the 1975 Annual Housing Survey (AHS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development. In each of the 21 SMSA's, the data were collected for the 12-month period from April 1975 through March 1976, with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and a third group of 20 SMSA's were collected from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated for the AHS every three years on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the second group (Year II SMSA's) are: Atlanta, Ga., Chicago, Ill., Philadelphia, Pa.-N.J., and San Francisco-Oakland, Calif.

The remaining SMSA's in the second group are: Cincinnati, Ohio-Ky.-Ind., Colorado Springs, Colo., Columbus, Ohio, Hartford, Conn., Kansas City, Mo.-Kans., Miami, Fla., Milwaukee, Wis., New Orleans, La., Newport News-Hampton, Va., Paterson-Clifton-Passaic, N.J., Portland, Oreg.-Wash., Rochester, N.Y., San Antonio, Tex., San Bernardino-Riverside-Ontario, Calif., San Diego, Calif., Springfield-Chicopee-Holyoke, Mass.-Conn., and Madison, Wis.

In this SMSA, 4,981 units were eligible for interview. Of these sample units, 253 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 193 units were visited but found not to be eligible for interview, because these units were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit issuing was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for those SMSA's which are *not* 100-percent permit issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). The following table indicates which SMSA's are 100-percent permit issuing and which contain a sample from the nonpermit universe.

100 Percent permit-issuing SMSA's	SMSA's with a sample from the nonpermit universe
Hartford, Conn.	Atlanta, Ga.
Miami, Fla.	Chicago, Ill.
Newport News-Hampton, Va.	Cincinnati, Ohio-Ky.-Ind.
Paterson-Clifton-Passaic, N.J.	Colorado Springs, Colo.
San Bernardino-Riverside-Ontario, Calif.	Columbus, Ohio
San Diego, Calif.	Kansas City, Mo.-Kans.
San Francisco-Oakland, Calif.	Madison, Wis.
	Milwaukee, Wis.
	New Orleans, La.
	Philadelphia, Pa.-N.J.
	Portland, Oreg.-Wash.
	Rochester, N.Y.
	San Antonio, Tex.
	Springfield-Chicopee-Holyoke, Mass.-Conn.

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of these sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and the balance of the SMSA, since the

sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—family size	Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000—$5,999 . .		
$6,000—$9,999 . .		
$10,000—$14,999		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. The housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued, and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \frac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Some tables in this report provide estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at enumeration time (i.e., the 1975 housing inventory), and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1975 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1975 housing inventory.—The AHS estimates of characteristics of the 1975 housing inventory employed a three-stage ratio estimation procedure. However, the third-stage ratio estimation procedure was employed in only three SMSA's. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 253 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units and one noninterview cell for the nonpermit universe (if applicable). Sample housing units from the nonpermit universe identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce the variation in sample size for strata used in the sample selection of the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes built after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or by other sources). This procedure was designed to adjust the AHS sample estimates of these types of units to independently derived current estimates. These independent estimates were considered to be the best estimates available. The adjustment was necessary to correct known deficiencies in the AHS sample in the representation of these units (see section on nonsampling error). The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1975 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimates of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units.	
New mobile homes	
"Other additions".	

The numerators of the ratios were derived by applying the following factors:

1. For the "conventional new construction units" cell, either a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in the SMSA, or the missed conventional new construction was estimated from the Survey of Construction (SOC). (The table that follows indicates the procedure used to estimate the missing conventional new construction for each SMSA.)
2. For the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied.
3. For the "other additions" cell, the "other additions" rate from SCARF[1] was applied.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

SMSA's employing the SOC estimate of missed conventional new construction	SMSA's employing the national trend for missed conventional new construction
Atlanta, Ga.	Colorado Springs, Colo.
Chicago, Ill.	Hartford, Conn.
Cincinnati, Ohio-Ky.-Ind.	Madison, Wis.
Columbus, Ohio	New Orleans, La.
Kansas City, Mo.-Kans.	Newport News—Hampton, Va.
Miami, Fla.	Paterson-Clifton-Passaic, N.J.
Milwaukee, Wis.	Rochester, N.Y.
Philadelphia, Pa.-N.J.	San Antonio, Tex.
Portland, Oreg.-Wash.	San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.	Springfield-Chicopee-Holyoke, Mass.-Conn.
San Francisco-Oakland, Calif.	

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

For some Year II SMSA's (i.e., Colorado Springs, Colo., Miami, Fla., and San Diego, Calif.) a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1975 housing inventory to an independent estimate of the SMSA's October 1975 housing inventory. This ratio estimate factor equaled the following:

$$\frac{\text{Independent estimate of the October 1975 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using the 1970 census estimate of total housing units for the SMSA in conjunction with the estimate of the change in the housing inventory since the census. The estimate of change was based on either administrative records from utility companies, or, when utility data were not available, on estimates of new construction permits and post-census demolition data. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, are brought into agreement with known estimates of the SMSA housing population.

The third-stage ratio estimation procedure was not employed in all Year II SMSA's since the reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. The independent estimates were only employed in SMSA's where the estimated relative bias was thought to be low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results. A measure of the relative bias of the independent estimate was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1975 independent estimates. In some SMSA's, the measure of the relative bias of the independent estimate was less than 3 percent but the independent estimate was still not employed, since there was evidence that suggested that the procedure for generating the independent estimates was producing substantial overestimates for these SMSA's.

1970-1975 lost units.—The AHS estimate of characteristics of the 1970-1975 lost units employed a one-stage ratio estimation procedure which is similar to the first-stage ratio estimation procedure mentioned above. The 1970-1975 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for State, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and non-

sampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Nonsampling errors.—Nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1975 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors associated with 1970 census estimates—"coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census,** and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

AHS-SMSA.—For the 1975 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are not yet available, however, it is believed that they will be similar to the results of the Year I AHS-SMSA reinterview study which were presented in the following Census Bureau memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample: 1974."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year I SMSA's and not for any specific SMSA):

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. The estimated indexes of inconsistency for most of the reinterview items were in the 20 to 50 range. The 20-50 interval is considered moderate indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview program because they had not been previously tested.

3. "Our bias indicator, the net difference rate, shows no signs of serious bias. For only one characteristic was there as much as a 2½ percent probability that the net difference rate is greater than 10 percent."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors.—With errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies in the representation of both conventional new construction and new mobile homes (and trailers) in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later, were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and

therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 2,100 conventional new construction units in this SMSA had permits issued prior to January 1970, and therefore, were missed by the 1975 AHS-SMSA survey.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. Although it is not known exactly, it was estimated that no new mobile homes in permit-issuing areas were missed by the 1975 AHS-SMSA survey in this SMSA. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS data still exists.

Rounding errors.—The rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling error, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in tables I and II are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1975 Housing Inventory and for Estimated Number of 1970-1975 Lost Units for the Paterson-Clifton-Passaic, N.J., SMSA

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100. . . .	100	10,000 . .	960
200. . . .	140	25,000 . .	1,500
500. . . .	220	50,000 . .	2,060
700. . . .	260	100,000 .	2,740
1,000 . .	310	250,000 .	3,370
2,500 . .	490	400,000 .	2,520
5,000 . .	690		

Tables I and II present the standard errors applicable to estimates of characteristics of the 1975 housing inventory as well as estimates of characteristics of the 1970-1975 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in these tables.

For ratios, (100) (x/y), where x is not a subclass of y, table II underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the Paterson-Clifton-Passaic, N.J., SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
100	9.7	21.2	29.2	42.1	48.7
200	6.8	15.0	20.6	29.8	34.4
500	4.3	9.5	13.1	18.8	21.8
700	3.7	8.0	11.0	15.9	18.4
1,000	3.1	6.7	9.2	13.3	15.4
2,500	1.9	4.2	5.8	8.4	9.7
5,000	1.4	3.0	4.1	6.0	6.9
10,000	1.0	2.1	2.9	4.2	4.9
25,000	.6	1.3	1.8	2.7	3.1
50,000	.4	.9	1.3	1.9	2.2
100,000	.3	.7	.9	1.3	1.5
250,000	.2	.4	.6	.8	1.0
400,000	.2	.3	.5	.7	.8

obtained by letting the standard error of the ratio be approximately equal to:

$$(100)\ (x/y)\ \sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

Let: x = the numerator of the ratio

y = the denominator of the ratio

σ_x = the standard error of the numerator

σ_y = the standard error of the denominator

Illustration of the use of the standard eror tables.—Table 1 of part A of this report shows that in 1975 there were 270,500 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 3,250. Consequently, the 68-percent confidence interval, as shown by these data, is from 267,250 to 273,750 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 265,300 to 275,700 housing units with 90 percent confidence; and that the average estimate lies within the interval from 264,000 to 277,000 housing units with 95 percent confidence.

Table 1 of part A also shows that of the 270,500 owner-occupied housing units 63,800, or 23.6 percent, had two bedrooms. Interpolation in table II of this appendix shows that the standard error of the percent is approximately .8 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 22.8 to 24.4 percent; the 90-percent confidence interval is from 22.3 to 24.9 percent; and the 95-percent confidence interval is from 22.0 to 25.2 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics; the formula will overestimate the true standard error; while, if there is a high negative correlation, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table 1 of part A of this report shows that in 1975 there were 132,200 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 68,400. Table I shows the standard error of 63,800 is approximately 2,250 and the standard error of 132,200 is approximately 2,880. Therefore, the standard error of the estimated difference of 68,400 is about

$$3{,}650 = \sqrt{(2{,}250)^2 + (2{,}880)^2}$$

Consequently, the 68-percent confidence interval for the 68,400 difference is from 64,750 to 72,050 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 62,560 to 74,240 housing units, and the 95-percent confidence interval is from 61,100 to 75,700. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval does not include zero or negative values.

Medians.—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies

within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From table II determine the standard error of a 50-percent characteristic on the base of the median;
2. add to and subtract from 50 percent, the standard error determined in step 1; and
3. using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median.—Table 1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.2. The base of the distribution from which this median was determined is 270,500 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 270,500 is approximately 1.0 percentage points.
2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 48.0 and 52.0.
3. From table 1 of part A, it can be seen by cumulating the frequencies for the first two categories that 97,700 owner-occupied housing units, or 36.1 percent, had one or two persons (for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 52,600 owner-occupied housing units, or 19.4 percent, had three persons (i.e., 2.5 to 3.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{48.0-36.1}{19.4}\right) = 3.1$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{52.0-36.1}{19.4}\right) = 3.3$$

Thus, the 95-percent confidence interval ranges from 3.1 to 3.3 persons.

Philadelphia 1975

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-41
Annual Housing Survey App-41
Selection of the sample. App-41
Building loss sample selection . . . App-42
1970 Census of Population and Housing. App-43
ESTIMATION. App-43
1975 housing inventory App-43
1970-1975 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing. App-44
RELIABILITY OF THE ESTIMATES. App-44
Nonsampling errors. App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-45
Rounding errors. App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences. App-47
Illustration of the computation of the standard error of a difference. App-47
Medians App-48
Illustration of the computation of the 95-percent confidence interval of a median App-48

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 21 SMSA's are based on data collected from the 1975 Annual Housing Survey (AHS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development. In each of the 21 SMSA's, the data were collected for the 12-month period from April 1975 through March 1976, with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and a third group of 20 SMSA's were collected from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated for the AHS every three years on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the second group (Year II SMSA's) are: Atlanta, Ga., Chicago, Ill., Philadelphia, Pa.-N.J., and San Francisco-Oakland, Calif.

The remaining SMSA's in the second group are: Cincinnati, Ohio-Ky.-Ind., Colorado Springs, Colo., Columbus, Ohio, Hartford, Conn., Kansas City, Mo.-Kans., Miami, Fla., Milwaukee, Wis., New Orleans, La., Newport News-Hampton, Va., Paterson-Clifton-Passaic, N.J., Portland, Oreg.-Wash., Rochester, N.Y., San Antonio, Tex., San Bernardino-Riverside-Ontario, Calif., San Diego, Calif., Springfield-Chicopee-Holyoke, Mass.-Conn., and Madison, Wis.

In this SMSA, 14,447 units were eligible for interview. Of these sample units, 863 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 1,203 units were visited but found not to be eligible for interview, because these units were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit issuing was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for those SMSA's which are *not* 100-percent permit issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). The following table indicates which SMSA's are 100-percent permit issuing and which contain a sample from the nonpermit universe.

100 Percent permit-issuing SMSA's	SMSA's with a sample from the nonpermit universe
Hartford, Conn.	Atlanta, Ga.
Miami, Fla.	Chicago, Ill.
Newport News-Hampton, Va.	Cincinnati, Ohio-Ky.-Ind.
Paterson-Clifton-Passaic, N.J.	Colorado Springs, Colo.
San Bernardino-Riverside-Ontario, Calif.	Columbus, Ohio
San Diego, Calif.	Kansas City, Mo.-Kans.
San Francisco-Oakland, Calif.	Madison, Wis.
	Milwaukee, Wis.
	New Orleans, La.
	Philadelphia, Pa.-N.J.
	Portland, Oreg.-Wash.
	Rochester, N.Y.
	San Antonio, Tex.
	Springfield-Chicopee-Holyoke, Mass.-Conn.

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of these sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and the balance of the SMSA, since the

sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—family size	Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000—$5,999 . .		
$6,000—$9,999 . .		
$10,000—$14,999		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. The housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued, and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \frac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Some tables in this report provide estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at enumeration time (i.e., the 1975 housing inventory), and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1975 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1975 housing inventory.—The AHS estimates of characteristics of the 1975 housing inventory employed a three-stage ratio estimation procedure. However, the third-stage ratio estimation procedure was employed in only three SMSA's. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 863 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units and one noninterview cell for the nonpermit universe (if applicable). Sample housing units from the nonpermit universe identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce the variation in sample size for strata used in the sample selection of the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes built after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or by other sources). This procedure was designed to adjust the AHS sample estimates of these types of units to independently derived current estimates. These independent estimates were considered to be the best estimates available. The adjustment was necessary to correct known deficiencies in the AHS sample in the representation of these units (see section on nonsampling error). The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1975 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimates of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units	
New mobile homes	
"Other additions"	

The numerators of the ratios were derived by applying the following factors:

1. For the "conventional new construction units" cell, either a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in the SMSA, or the missed conventional new construction was estimated from the Survey of Construction (SOC). (The table that follows indicates the procedure used to estimate the missing conventional new construction for each SMSA.)
2. For the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied.
3. For the "other additions" cell, the "other additions" rate from SCARF[1] was applied.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

SMSA's employing the SOC estimate of missed conventional new construction	SMSA's employing the national trend for missed conventional new construction
Atlanta, Ga.	Colorado Springs, Colo.
Chicago, Ill.	Hartford, Conn.
Cincinnati, Ohio-Ky.-Ind.	Madison, Wis.
Columbus, Ohio	New Orleans, La.
Kansas City, Mo.-Kans.	Newport News—Hampton, Va.
Miami, Fla.	Paterson-Clifton-Passaic, N.J.
Milwaukee, Wis.	Rochester, N.Y.
Philadelphia, Pa.-N.J.	San Antonio, Tex.
Portland, Oreg.-Wash.	San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.	Springfield-Chicopee-Holyoke, Mass.-Conn.
San Francisco-Oakland, Calif.	

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

For some Year II SMSA's (i.e., Colorado Springs, Colo., Miami, Fla., and San Diego, Calif.) a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1975 housing inventory to an independent estimate of the SMSA's October 1975 housing inventory. This ratio estimate factor equaled the following:

$$\frac{\text{Independent estimate of the October 1975 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using the 1970 census estimate of total housing units for the SMSA in conjunction with the estimate of the change in the housing inventory since the census. The estimate of change was based on either administrative records from utility companies, or, when utility data were not available, on estimates of new construction permits and post-census demolition data. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, are brought into agreement with known estimates of the SMSA housing population.

The third-stage ratio estimation procedure was not employed in all Year II SMSA's since the reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. The independent estimates were only employed in SMSA's where the estimated relative bias was thought to be low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results. A measure of the relative bias of the independent estimate was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1975 independent estimates. In some SMSA's, the measure of the relative bias of the independent estimate was less than 3 percent but the independent estimate was still not employed, since there was evidence that suggested that the procedure for generating the independent estimates was producing substantial overestimates for these SMSA's.

1970-1975 lost units.—The AHS estimate of characteristics of the 1970-1975 lost units employed a one-stage ratio estimation procedure which is similar to the first-stage ratio estimation procedure mentioned above. The 1970-1975 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units + Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for State, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and non-

sampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Nonsampling errors.—Nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1975 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors associated with 1970 census estimates—"coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census,** and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

AHS-SMSA.—For the 1975 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are not yet available, however, it is believed that they will be similar to the results of the Year I AHS-SMSA reinterview study which were presented in the following Census Bureau memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample: 1974."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year I SMSA's and not for any specific SMSA):

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. The estimated indexes of inconsistency for most of the reinterview items were in the 20 to 50 range. The 20-50 interval is considered moderate indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview program because they had not been previously tested.

3. "Our bias indicator, the net difference rate, shows no signs of serious bias. For only one characteristic was there as much as a 2½ percent probability that the net difference rate is greater than 10 percent."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors.—With errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies in the representation of both conventional new construction and new mobile homes (and trailers) in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later, were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and

therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 18,800 conventional new construction units in this SMSA had permits issued prior to January 1970, and therefore, were missed by the 1975 AHS-SMSA survey.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. Although it is not known exactly, an estimated 2,200 new mobile homes in permit-issuing areas were missed by the 1975 AHS-SMSA survey in this SMSA. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS data still exists.

Rounding errors.—The rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling error, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in tables I and II are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1975 Housing Inventory and for Estimated Number of 1970-1975 Lost Units for the Philadelphia, Pa.-N.J., SMSA, for the Central City and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central city	Not in central city
200	170	150	180
500	270	240	280
700	320	290	340
1,000	380	340	400
2,500	600	540	640
5,000	840	760	900
10,000	1,190	1,080	1,270
25,000	1,880	1,680	2,000
50,000	2,650	2,340	2,810
100,000	3,720	3,200	3,920
250,000	5,760	4,480	5,960
400,000	7,130	4,830	7,210
500,000	7,850	4,660	7,810
1,000,000	10,200	—	9,060
1,500,000	11,300	—	—

Tables I through IV present the standard errors applicable to estimates of

characteristics of the 1975 housing inventory as well as estimates of characteristics of the 1970-1975 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in these tables.

For ratios, (100) (x/y), where x is not a subclass of y, tables II through IV underestimate the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be obtained by letting the standard error of the ratio be approximately equal to:

$$(100)\,(x/y)\,\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

Let: x = the numerator of the ratio
y = the denominator of the ratio
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

Illustration of the use of the standard error tables.—Table A-1 of part A of this report shows that in 1975 there were 1,030,300 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 10,270. Consequently, the 68-percent confidence interval, as shown by these data, is from 1,020,030 to 1,040,570 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 1,013,870 to 1,046,730 housing units with 90 percent confidence; and that the average estimate lies within the interval from 1,009,760 to 1,050,840 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 1,030,300 owner-occupied housing units 149,800, or 14.5 percent, had two bedrooms. Interpolation in table II of this appendix shows that the standard error of the percent is approximately .4 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 14.1 to 14.9 percent; the 90-percent confidence interval is from 13.9 to 15.1 percent; and the 95-percent confidence interval is from 13.7 to 15.3 percent.

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the Philadelphia, Pa.-N.J., SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
200	8.4	18.4	25.3	36.5	42.2
500	5.3	11.6	16.0	23.1	26.7
700	4.5	9.8	13.5	19.5	22.5
1,000	3.8	8.2	11.3	16.3	18.9
2,500	2.4	5.2	7.2	10.3	11.9
5,000	1.7	3.7	5.1	7.3	8.4
10,000	1.2	2.6	3.6	5.2	6.0
25,000	.8	1.6	2.3	3.3	3.8
50,000	.5	1.2	1.6	2.3	2.7
100,000	.4	.8	1.1	1.6	1.9
250,000	.2	.5	.7	1.0	1.2
400,000	.2	.4	.6	.8	.9
500,000	.2	.4	.5	.7	.8
1,000,000	.12	.3	.4	.5	.6
1,500,000	.10	.2	.3	.4	.5

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; while, if there is a high negative correlation, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table A-1 of part A of this report shows that in 1975 there were 599,200 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 449,400. Table I shows the standard error of 149,800 is approximately 4,400 and the standard error of 599,200 is approximately 8,320. Therefore, the standard error of the estimated difference of 449,400 is about

$$9{,}410 = \sqrt{(4{,}400)^2 + (8{,}320)^2}$$

Consequently, the 68-percent confidence interval for the 449,400 difference is from 439,990 to 458,810 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 434,340 to 464,460 housing units, and the 95-percent confidence interval is from 430,580 to 468,220. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied units

TABLE III. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the Central City

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
200	7.6	16.7	23.0	33.1	38.3
500	4.8	10.5	14.5	21.0	24.2
700	4.1	8.9	12.3	17.7	20.4
1,000	3.4	7.5	10.3	14.8	17.1
2,500	2.2	4.7	6.5	9.4	10.8
5,000	1.5	3.3	4.6	6.6	7.7
10,000	1.1	2.4	3.2	4.7	5.4
25,000	.7	1.5	2.1	3.0	3.4
50,000	.5	1.1	1.5	2.1	2.4
100,000	.3	.7	1.0	1.5	1.7
250,000	.2	.5	.6	.9	1.1
400,000	.2	.4	.5	.7	.9
500,000	.2	.3	.5	.7	.8

with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval does not include zero or negative values.

Medians.—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From table II determine the standard error of a 50-percent characteristic on the base of the median;
2. add to and subtract from 50 percent, the standard error determined in step 1; and
3. using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.1. The base of the distribution from which this median was determined is 1,030,300 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 1,030,300 is approximately .6 percentage points.
2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 48.8 and 51.2.
3. From table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 406,700 owner-occupied housing units, or 39.5 percent, had one or two persons (for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 193,200 owner-occupied housing units, or 18.8 percent, had three persons (i.e., 2.5 to 3.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

TABLE IV. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the Balance of the SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
100	12.7	27.7	38.2	55.1	63.6
200	9.0	19.6	27.0	39.0	45.0
500	5.7	12.4	17.1	24.7	28.5
700	4.8	10.5	14.4	20.8	24.1
1,000	4.0	8.8	12.1	17.4	20.1
2,500	2.5	5.5	7.6	11.0	12.7
5,000	1.8	3.9	5.4	7.8	9.0
10,000	1.3	2.8	3.8	5.5	6.4
25,000	.8	1.8	2.4	3.5	4.0
50,000	.6	1.2	1.7	2.5	2.8
100,000	.4	.9	1.2	1.7	2.0
250,000	.3	.6	.8	1.1	1.3
500,000	.2	.4	.5	.8	.9
1,000,000	.13	.3	.4	.6	.6

$$2.5 + (3.5-2.5)\left(\frac{48.8-39.5}{18.8}\right) = 3.0$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{51.2-39.5}{18.8}\right) = 3.1$$

Thus, the 95-percent confidence interval ranges from 3.0 to 3.1 persons. Although it appears that this confidence interval has the sample estimate as the upper limit, it actually is a reflection of the rounding error associated with the median (see the paragraph on rounding errors in the nonsampling errors section).

Portland 1975

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-41
Annual Housing Survey App-41
Selection of the sample. App-41
Building loss sample selection . . . App-42
1970 Census of Population and Housing. App-43
ESTIMATION. App-43
1975 housing inventory App-43
1970-1975 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing. App-44
RELIABILITY OF THE ESTIMATES. App-44
Nonsampling errors. App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-45
Rounding errors. App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences. App-47
Illustration of the computation of the standard error of a difference. App-47
Medians App-47
Illustration of the computation of the 95-percent confidence interval of a median App-48

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 21 SMSA's are based on data collected from the 1975 Annual Housing Survey (AHS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development. In each of the 21 SMSA's, the data were collected for the 12-month period from April 1975 through March 1976, with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and a third group of 20 SMSA's were collected from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated for the AHS every three years on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the second group (Year II SMSA's) are: Atlanta, Ga., Chicago, Ill., Philadelphia, Pa.-N.J., and San Francisco-Oakland, Calif.

The remaining SMSA's in the second group are: Cincinnati, Ohio-Ky.-Ind., Colorado Springs, Colo., Columbus, Ohio, Hartford, Conn., Kansas City, Mo.-Kans., Miami, Fla., Milwaukee, Wis., New Orleans, La., Newport News-Hampton, Va., Paterson-Clifton-Passaic, N.J., Portland, Oreg.-Wash., Rochester, N.Y., San Antonio, Tex., San Bernardino-Riverside-Ontario, Calif., San Diego, Calif., Springfield-Chicopee-Holyoke, Mass.-Conn., and Madison, Wis.

In this SMSA, 4,875 units were eligible for interview. Of these sample units, 157 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 363 units were visited but found not to be eligible for interview, because these units were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit issuing was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for those SMSA's which are *not* 100-percent permit issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). The following table indicates which SMSA's are 100-percent permit issuing and which contain a sample from the nonpermit universe.

100 Percent permit-issuing SMSA's	SMSA's with a sample from the nonpermit universe
Hartford, Conn.	Atlanta, Ga.
Miami, Fla.	Chicago, Ill.
Newport News-Hampton, Va.	Cincinnati, Ohio-Ky.-Ind.
Paterson-Clifton-Passaic, N.J.	Colorado Springs, Colo.
San Bernardino-Riverside-Ontario, Calif.	Columbus, Ohio
San Diego, Calif.	Kansas City, Mo.-Kans.
San Francisco-Oakland, Calif.	Madison, Wis.
	Milwaukee, Wis.
	New Orleans, La.
	Philadelphia, Pa.-N.J.
	Portland, Oreg.-Wash.
	Rochester, N.Y.
	San Antonio, Tex.
	Springfield-Chicopee-Holyoke, Mass.-Conn.

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of these sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and the balance of the SMSA, since the

sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—family size	Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000—$5,999 . .		
$6,000—$9,999 . .		
$10,000—$14,999		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. The housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued, and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \frac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Some tables in this report provide estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at enumeration time (i.e., the 1975 housing inventory), and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1975 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1975 housing inventory.—The AHS estimates of characteristics of the 1975 housing inventory employed a three-stage ratio estimation procedure. However, the third-stage ratio estimation procedure was employed in only three SMSA's. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 157 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units and one noninterview cell for the nonpermit universe (if applicable). Sample housing units from the nonpermit universe identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce the variation in sample size for strata used in the sample selection of the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes built after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or by other sources). This procedure was designed to adjust the AHS sample estimates of these types of units to independently derived current estimates. These independent estimates were considered to be the best estimates available. The adjustment was necessary to correct known deficiencies in the AHS sample in the representation of these units (see section on nonsampling error). The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1975 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimates of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units	
New mobile homes	
"Other additions"	

The numerators of the ratios were derived by applying the following factors:

1. For the "conventional new construction units" cell, either a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in the SMSA, or the missed conventional new construction was estimated from the Survey of Construction (SOC). (The table that follows indicates the procedure used to estimate the missing conventional new construction for each SMSA.)
2. For the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied.
3. For the "other additions" cell, the "other additions" rate from SCARF[1] was applied.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

SMSA's employing the SOC estimate of missed conventional new construction	SMSA's employing the national trend for missed conventional new construction
Atlanta, Ga.	Colorado Springs, Colo.
Chicago, Ill.	Hartford, Conn.
Cincinnati, Ohio-Ky.-Ind.	Madison, Wis.
Columbus, Ohio	New Orleans, La.
Kansas City, Mo.-Kans.	Newport News—Hampton, Va.
Miami, Fla.	Paterson-Clifton-Passaic, N.J.
Milwaukee, Wis.	Rochester, N.Y.
Philadelphia, Pa.-N.J.	San Antonio, Tex.
Portland, Oreg.-Wash.	San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.	Springfield-Chicopee-Holyoke, Mass.-Conn.
San Francisco-Oakland, Calif.	

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

For some Year II SMSA's (i.e., Colorado Springs, Colo., Miami, Fla., and San Diego, Calif.) a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1975 housing inventory to an independent estimate of the SMSA's October 1975 housing inventory. This ratio estimate factor equaled the following:

$$\frac{\text{Independent estimate of the October 1975 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using the 1970 census estimate of total housing units for the SMSA in conjunction with the estimate of the change in the housing inventory since the census. The estimate of change was based on either administrative records from utility companies, or, when utility data were not available, on estimates of new construction permits and post-census demolition data. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, are brought into agreement with known estimates of the SMSA housing population.

The third-stage ratio estimation procedure was not employed in all Year II SMSA's since the reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. The independent estimates were only employed in SMSA's where the estimated relative bias was thought to be low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results. A measure of the relative bias of the independent estimate was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1975 independent estimates. In some SMSA's, the measure of the relative bias of the independent estimate was less than 3 percent but the independent estimate was still not employed, since there was evidence that suggested that the procedure for generating the independent estimates was producing substantial overestimates for these SMSA's.

1970-1975 lost units.—The AHS estimate of characteristics of the 1970-1975 lost units employed a one-stage ratio estimation procedure which is similar to the first-stage ratio estimation procedure mentioned above. The 1970-1975 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units + Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for State, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and non-

sampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Nonsampling errors.—Nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1975 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors associated with 1970 census estimates—"coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census,** and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

AHS-SMSA.—For the 1975 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are not yet available, however, it is believed that they will be similar to the results of the Year I AHS-SMSA reinterview study which were presented in the following Census Bureau memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample: 1974."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year I SMSA's and not for any specific SMSA):

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. The estimated indexes of inconsistency for most of the reinterview items were in the 20 to 50 range. The 20-50 interval is considered moderate indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview program because they had not been previously tested.

3. "Our bias indicator, the net difference rate, shows no signs of serious bias. For only one characteristic was there as much as a 2½ percent probability that the net difference rate is greater than 10 percent."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors.—With errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies in the representation of both conventional new construction and new mobile homes (and trailers) in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later, were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and

therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 5,100 conventional new construction units in this SMSA had permits issued prior to January 1970, and therefore, were missed by the 1975 AHS-SMSA survey.

—In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. Although it is not known exactly, an estimated 4,400 new mobile homes in permit-issuing areas were missed by the 1975 AHS-SMSA survey in this SMSA. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS data still exists.

Rounding errors.—The rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling error, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in tables I and II are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1975 Housing Inventory and for Estimated Number of 1970-1975 Lost Units for the Portland, Oreg.-Wash., SMSA

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100. . . .	100	10,000 .	980
200. . . .	140	25,000 .	1,540
500. . . .	220	50,000 .	2,130
700. . . .	260	100,000	2,870
1,000 . .	310	250,000	3,820
2,500 . .	500	400,000	3,700
5,000 . .	700		

Tables I and II present the standard errors applicable to estimates of characteristics of the 1975 housing inventory as well as estimates of characteristics of the 1970-1975 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in these tables.

For ratios, (100) (x/y), where x is not a subclass of y, table II underestimates the standard error of the ratio when there is little or no correlation between x and

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-75 Lost Units for the Portland, Oreg.-Wash., SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
100	9.9	21.6	29.8	43.0	49.6
200	7.0	15.3	21.0	30.4	35.1
500	4.4	9.7	13.3	19.2	22.2
700	3.7	8.2	11.3	16.2	18.8
1,000	3.1	6.8	9.4	13.6	15.7
2,500	2.0	4.3	6.0	8.6	9.9
5,000	1.4	3.1	4.2	6.1	7.0
10,000	1.0	2.2	3.0	4.3	5.0
25,000	.6	1.4	1.9	2.7	3.1
50,000	.4	1.0	1.3	1.9	2.2
100,000	.3	.7	.9	1.4	1.6
250,000	.2	.4	.6	.9	1.0
400,000	.2	.3	.5	.7	.8

y. For this type of ratio, a better approximation of the standard error may be obtained by letting the standard error of the ratio be approximately equal to:

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2+\left(\frac{\sigma_y}{y}\right)^2}$$

Let: x = the numerator of the ratio
y = The denominator of the ratio
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

Illustration of the use of the standard error tables.—Table 1 of part A of this report shows that in 1975 there were 251,300 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 3,820. Consequently, the 68-percent confidence interval, as shown by these data, is from 247,480 to 255,120 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 245,190 to 257,410 housing units with 90 percent confidence; and that the average estimate lies within the interval from 243,660 to 258,940 housing units with 95 percent confidence.

Table 1 of part A also shows that of the 251,300 owner-occupied housing units 67,000, or 26.7 percent, had two bedrooms. Interpolation in table II of this appendix shows that the standard error of the percent is approximately .9 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 25.8 to 27.6 percent; the 90-percent confidence interval is from 25.3 to 28.1 percent; and the 95-percent confidence interval is from 24.9 to 28.5 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximatley equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; while, if there is a high negative correlation, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table 1 of part A of this report shows that in 1975 there were 116,000 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 49,000. Table I shows the standard error of 67,000 is approximately 2,380 and the standard error of 116,000 is approximately 2,970. Therefore, the standard error of the estimated difference of 49,000 is about

$$3{,}810 = \sqrt{(2{,}380)^2 + (2{,}970)^2}$$

Consequently, the 68-percent confidence interval for the 49,000 difference is from 45,190 to 52,810 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 42,900 to 55,100 housing units, and the 95-percent confidence interval is from 41,380 to 56,620. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval does not include zero or negative values.

Medians.—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for

measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From table II determine the standard error of a 50-percent characteristic on the base of the median;
2. add to an subtract from 50 percent, the standard error determined in step 1; and
3. using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median.— Table 1 of part A of this report shows the median number of persons for owner-occupied housing units is 2.5. The base of the distribution from which this median was determined is 251,300 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 251,300 is approximately 1.0 percentage points.
2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 48.0 and 52.0.
3. From table 1 of part A, it can be seen that 37,200 owner-occupied housing units, or 14.8 percent, had one person (for purposes of calculating the median, the category of one person is considered to be from .5 to 1.5 persons); that an additional 86,100 owner-occupied housing units, or 34.3 percent, had two persons (i.e., 1.5 to 2.5 persons); and that 41,500 or 16.5 percent, had three persons (i.e., 2.5 to 3.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$1.5 + (2.5-1.5)\left(\frac{48.0-14.8}{34.3}\right) = 2.5$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{52.0-14.8-34.3}{16.5}\right) = 2.7$$

Thus, the 95-percent confidence interval ranges from 2.5 to 2.7 persons. Although it appears that this confidence interval has the sample estimate as the lower limit, it actually is a reflection of the rounding error associated with the median (see the paragraph on rounding errors in the nonsampling errors section).

Riverside-San Bernadino-Ontario 1975

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-41
Annual Housing Survey App-41
Selection of the sample. App-41
Building loss sample selection . . . App-42
1970 Census of Population and Housing. App-43
ESTIMATION. App-43
1975 housing inventory App-43
1970-1975 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing. App-44
RELIABILITY OF THE ESTIMATES. App-44
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-45
Rounding errors. App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences. App-47
Illustration of the computation of the standard error of a difference. App-47
Medians App-47
Illustration of the computation of the 95-percent confidence interval of a median App-48

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 21 SMSA's are based on data collected from the 1975 Annual Housing Survey (AHS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development. In each of the 21 SMSA's, the data were collected for the 12-month period from April 1975 through March 1976, with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and a third group of 20 SMSA's were collected from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated for the AHS every three years on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the second group (Year II SMSA's) are: Atlanta, Ga., Chicago, Ill., Philadelphia, Pa.-N.J., and San Francisco-Oakland, Calif.

The remaining SMSA's in the second group are: Cincinnati, Ohio-Ky.-Ind., Colorado Springs, Colo., Columbus, Ohio, Hartford, Conn., Kansas City, Mo.-Kans., Miami, Fla., Milwaukee, Wis., New Orleans, La., Newport News-Hampton, Va., Paterson-Clifton-Passaic, N.J., Portland, Oreg.-Wash., Rochester, N.Y., San Antonio, Tex., San Bernardino-Riverside-Ontario, Calif., San Diego, Calif., Springfield-Chicopee-Holyoke, Mass.-Conn., and Madison, Wis.

In this SMSA, 4,963 units were eligible for interview. Of these sample units 241 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 347 units were visited but found not to be eligible for interview, because these units were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit issuing was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for those SMSA's which are *not* 100-percent permit issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). The following table indicates which SMSA's are 100-percent permit issuing and which contain a sample from the nonpermit universe.

100 Percent permit-issuing SMSA's	SMSA's with a sample from the nonpermit universe
Hartford, Conn.	Atlanta, Ga.
Miami, Fla.	Chicago, Ill.
Newport News-Hampton, Va.	Cincinnati, Ohio-Ky.-Ind.
Paterson-Clifton-Passaic, N.J.	Colorado Springs, Colo.
San Bernardino-Riverside-Ontario, Calif.	Columbus, Ohio
San Diego, Calif.	Kansas City, Mo.-Kans.
San Francisco-Oakland, Calif.	Madison, Wis.
	Milwaukee, Wis.
	New Orleans, La.
	Philadelphia, Pa.-N.J.
	Portland, Oreg.-Wash.
	Rochester, N.Y.
	San Antonio, Tex.
	Springfield-Chicopee-Holyoke, Mass.-Conn.

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of these sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and the balance of the SMSA, since the

sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—family size	Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000—$5,999 . .		
$6,000—$9,999 . .		
$10,000—$14,999		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. The housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued, and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \frac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Some tables in this report provide estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at enumeration time (i.e., the 1975 housing inventory), and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1975 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1975 housing inventory.—The AHS estimates of characteristics of the 1975 housing inventory employed a three-stage ratio estimation procedure. However, the third-stage ratio estimation procedure was employed in only three SMSA's. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 241 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units and one noninterview cell for the nonpermit universe (if applicable). Sample housing units from the nonpermit universe identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce the variation in sample size for strata used in the sample selection of the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes built after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or by other sources). This procedure was designed to adjust the AHS sample estimates of these types of units to independently derived current estimates. These independent estimates were considered to be the best estimates available. The adjustment was necessary to correct known deficiencies in the AHS sample in the representation of these units (see section on nonsampling error). The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1975 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimates of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units.	
New mobile homes	
"Other additions".	

The numerators of the ratios were derived by applying the following factors:

1. For the "conventional new construction units" cell, either a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in the SMSA, or the missed conventional new construction was estimated from the Survey of Construction (SOC). (The table that follows indicates the procedure used to estimate the missing conventional new construction for each SMSA.)
2. For the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied.
3. For the "other additions" cell, the "other additions" rate from SCARF[1] was applied.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

SMSA's employing the SOC estimate of missed conventional new construction	SMSA's employing the national trend for missed conventional new construction
Atlanta, Ga.	Colorado Springs, Colo.
Chicago, Ill.	Hartford, Conn.
Cincinnati, Ohio-Ky.-Ind.	Madison, Wis.
Columbus, Ohio	New Orleans, La.
Kansas City, Mo.-Kans.	Newport News—Hampton, Va.
Miami, Fla.	Paterson-Clifton-Passaic, N.J.
Milwaukee, Wis.	Rochester, N.Y.
Philadelphia, Pa.-N.J.	San Antonio, Tex.
Portland, Oreg.-Wash.	San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.	Springfield-Chicopee-Holyoke, Mass.-Conn.
San Francisco-Oakland, Calif.	

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

For some Year II SMSA's (i.e., Colorado Springs, Colo., Miami, Fla., and San Diego, Calif.) a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1975 housing inventory to an independent estimate of the SMSA's October 1975 housing inventory. This ratio estimate factor equaled the following:

$$\frac{\text{Independent estimate of the October 1975 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using the 1970 census estimate of total housing units for the SMSA in conjunction with the estimate of the change in the housing inventory since the census. The estimate of change was based on either administrative records from utility companies, or, when utility data were not available, on estimates of new construction permits and post-census demolition data. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, are brought into agreement with known estimates of the SMSA housing population.

The third-stage ratio estimation procedure was not employed in all Year II SMSA's since the reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. The independent estimates were only employed in SMSA's where the estimated relative bias was thought to be low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results. A measure of the relative bias of the independent estimate was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1975 independent estimates. In some SMSA's, the measure of the relative bias of the independent estimate was less than 3 percent but the independent estimate was still not employed, since there was evidence that suggested that the procedure for generating the independent estimates was producing substantial overestimates for these SMSA's.

1970-1975 lost units.—The AHS estimate of characteristics of the 1970-1975 lost units employed a one-stage ratio estimation procedure which is similar to the first-stage ratio estimation procedure mentioned above. The 1970-1975 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units + Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for State, Cities, and Counties,** Part 1.[3]

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and non-

sampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

Nonsampling errors.—Nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1975 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors associated with 1970 census estimates—"coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

AHS-SMSA.—For the 1975 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are not yet available, however, it is believed that they will be similar to the results of the Year I AHS-SMSA reinterview study which were presented in the following Census Bureau memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample: 1974."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year I SMSA's and not for any specific SMSA):

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. The estimated indexes of inconsistency for most of the reinterview items were in the 20 to 50 range. The 20-50 interval is considered moderate indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview program because they had not been previously tested.

3. "Our bias indicator, the net difference rate, shows no signs of serious bias. For only one characteristic was there as much as a 2½ percent probability that the net difference rate is greater than 10 percent."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors.—With errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies in the representation of both conventional new construction and new mobile homes (and trailers) in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later, were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and

therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 3,100 conventional new construction units in this SMSA had permits issued prior to January 1970, and therefore, were missed by the 1975 AHS-SMSA survey.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. Although it is not known exactly, an estimated 10,300 new mobile homes in permit-issuing areas were missed by the 1975 AHS-SMSA survey in this SMSA. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS data still exists.

Rounding errors.—The rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling error, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in tables I and II are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1975 Housing Inventory and for Estimated Number of 1970-1975 Lost Units for the San Bernardino-Riverside-Ontario, Calif., SMSA

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	110	10,000 . .	1,060
200	150	25,000 . .	1,650
500	240	50,000 . .	2,290
700	280	100,000 .	3,080
1,000 . .	340	250,000 .	4,060
2,500 . .	530	400,000 .	3,840
5,000 . .	750	500,000 .	2,950

Tables I and II present the standard errors applicable to estimates of characteristics of the 1975 housing inventory as well as estimates of characteristics of the 1970-1975 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in these tables.

For ratios, (100) (x/y), where x is not a subclass of y, table II underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the San Bernardino-Riverside-Ontario, Calif., SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
100	10.6	23.3	32.1	46.3	53.5
200	7.5	16.5	22.7	32.7	37.8
500	4.8	10.4	14.3	20.7	23.9
700	4.0	8.8	12.1	17.5	20.2
1,000	3.4	7.4	10.1	14.6	16.9
2,500	2.1	4.7	6.4	9.3	10.7
5,000	1.5	3.3	4.5	6.5	7.6
10,000	1.1	2.3	3.2	4.6	5.3
25,000	.7	1.5	2.0	2.9	3.4
50,000	.5	1.0	1.4	2.1	2.4
100,000	.3	.7	1.0	1.5	1.7
250,000	.2	.5	.6	.9	1.1
400,000	.2	.4	.5	.7	.8
500,000	.2	.3	.5	.7	.8

obtained by letting the standard error of the ratio be approximately equal to:

$$(100)\ (x/y)\ \sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

Let: x = the numerator of the ratio
y = the denominator of the ratio
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

Illustration of the use of the standard error tables.—Table 1 of part A of this report shows that in 1975 there were 282,300 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 4,010. Consequently, the 68-percent confidence interval, as shown by these data, is from 278,290 to 286,310 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 275,880 to 288,720 housing units with 90 percent confidence; and that the average estimate lies within the interval from 274,280 to 290,320 housing units with 95 percent confidence.

Table 1 of part A also shows that of the 282,300 owner-occupied housing units, 87,500, or 31.0 percent, had two bedrooms. Interpolation in table II of this appendix shows that the standard error of the percent is approximately .9 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 30.1 to 31.9 percent; the 90-percent confidence interval is from 29.6 to 32.4 percent; and the 95-percent confidence interval is from 29.2 to 32.8 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; while, if there is a high negative correlation, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table 1 of part A of this report shows that in 1975 there were 127,100 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 39,600. Table I shows the standard error of 87,500 is approximately 2,880 and the standard error of 127,100 is approximately 3,260. Therefore, the standard error of the estimated difference of 39,600 is about

$$4{,}350 = \sqrt{(2{,}880)^2 + (3{,}260)^2}$$

Consequently, the 68-percent confidence interval for the 39,600 difference is from 35,250 to 43,950 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 32,640 to 46,560 housing units, and the 95-percent confidence interval is from 30,900 to 48,300. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval does not include zero or negative values.

Medians.—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies within the inter-

val. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From table II determine the standard error of a 50-percent characteristic on the base of the median;
2. add to and subtract from 50 percent, the standard error determined in step 1; and
3. using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median.—Table 1 of part A of this report shows the median number of persons for owner-occupied housing units is 2.5. The base of the distribution from which this median was determined is 282,300 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 282,300 is approximately 1.0 percentage points.
2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 48.0 and 52.0.
3. From table I of part A, it can be seen that 39,400 owner-occupied housing units, or 14.0 percent, had one person (for purposes of calculating the median, the category of one person is considered to be from .5 to 1.5 persons); that an additional 100,100 owner-occupied housing units, or 35.5 percent, had two persons (i.e., 1.5 to 2.5 persons); and that 47,800, or 16.9 percent, had three persons (i.e., 2.5 to 3.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$1.5 + (2.5-1.5)\left(\frac{48.0-14.0}{35.5}\right) = 2.5$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{52.0-14.0-35.5}{16.9}\right) = 2.6$$

Thus, the 95-percent confidence interval ranges from 2.5 to 2.6 persons. Although it appears that this confidence interval has the sample estimate as the lower limit, it actually is a reflection of the rounding error associated with the median (see the paragraph on rounding errors in the nonsampling errors section).

Rochester 1975

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-41
Annual Housing Survey App-41
Selection of the sample. App-41
Building loss sample selection . . . App-42
1970 Census of Population and Housing. App-43
ESTIMATION. App-43
1975 housing inventory App-43
1970-1975 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing. App-44
RELIABILITY OF THE ESTIMATES. App-44
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-45
Rounding errors. App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences. App-47
Illustration of the computation of the standard error of a difference. App-47
Medians App-47
Illustration of the computation of the 95-percent confidence interval of a median App-48

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 21 SMSA's are based on data collected from the 1975 Annual Housing Survey (AHS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development. In each of the 21 SMSA's, the data were collected for the 12-month period from April 1975 through March 1976, with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and a third group of 20 SMSA's were collected from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated for the AHS every three years on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the second group (Year II SMSA's) are: Atlanta, Ga., Chicago, Ill., Philadelphia, Pa.-N.J., and San Francisco-Oakland, Calif.

The remaining SMSA's in the second group are: Cincinnati, Ohio-Ky.-Ind., Colorado Springs, Colo., Columbus, Ohio, Hartford, Conn., Kansas City, Mo.-Kans., Miami, Fla., Milwaukee, Wis., New Orleans, La., Newport News-Hampton, Va., Paterson-Clifton-Passaic, N.J., Portland, Oreg.-Wash., Rochester, N.Y., San Antonio, Tex., San Bernardino-Riverside-Ontario, Calif., San Diego, Calif., Springfield-Chicopee-Holyoke, Mass.-Conn., and Madison, Wis.

In this SMSA, 4,845 units were eligible for interview. Of these sample units, 205 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 508 units were visited but found not to be eligible for interview, because these units were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit issuing was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for those SMSA's which are *not* 100-percent permit issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). The following table indicates which SMSA's are 100-percent permit issuing and which contain a sample from the nonpermit universe.

100 Percent permit-issuing SMSA's	SMSA's with a sample from the nonpermit universe
Hartford, Conn.	Atlanta, Ga.
Miami, Fla.	Chicago, Ill.
Newport News-Hampton, Va.	Cincinnati, Ohio-Ky.-Ind.
Paterson-Clifton-Passaic, N.J.	Colorado Springs, Colo.
San Bernardino-Riverside-Ontario, Calif.	Columbus, Ohio
San Diego, Calif.	Kansas City, Mo.-Kans.
San Francisco-Oakland, Calif.	Madison, Wis.
	Milwaukee, Wis.
	New Orleans, La.
	Philadelphia, Pa.-N.J.
	Portland, Oreg.-Wash.
	Rochester, N.Y.
	San Antonio, Tex.
	Springfield-Chicopee-Holyoke, Mass.-Conn.

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of these sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and the balance of the SMSA, since the

sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—family size	Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000–$5,999 . .		
$6,000–$9,999 . .		
$10,000–$14,999		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. The housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued, and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \frac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Some tables in this report provide estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at enumeration time (i.e., the 1975 housing inventory), and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1975 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1975 housing inventory.—The AHS estimates of characteristics of the 1975 housing inventory employed a three-stage ratio estimation procedure. However, the third-stage ratio estimation procedure was employed in only three SMSA's. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 205 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units and one noninterview cell for the nonpermit universe (if applicable). Sample housing units from the nonpermit universe identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce the variation in sample size for strata used in the sample selection of the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes built after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or by other sources). This procedure was designed to adjust the AHS sample estimates of these types of units to independently derived current estimates. These independent estimates were considered to be the best estimates available. The adjustment was necessary to correct known deficiencies in the AHS sample in the representation of these units (see section on nonsampling error). The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1975 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimates of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units	
New mobile homes	
"Other additions"	

The numerators of the ratios were derived by applying the following factors:

1. For the "conventional new construction units" cell, either a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in the SMSA, or the missed conventional new construction was estimated from the Survey of Construction (SOC). (The table that follows indicates the procedure used to estimate the missing conventional new construction for each SMSA.)
2. For the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied.
3. For the "other additions" cell, the "other additions" rate from SCARF[1] was applied.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

SMSA's employing the SOC estimate of missed conventional new construction	SMSA's employing the national trend for missed conventional new construction
Atlanta, Ga.	Colorado Springs, Colo.
Chicago, Ill.	Hartford, Conn.
Cincinnati, Ohio-Ky.-Ind.	Madison, Wis.
Columbus, Ohio	New Orleans, La.
Kansas City, Mo.-Kans.	Newport News—Hampton, Va.
Miami, Fla.	Paterson-Clifton-Passaic, N.J.
Milwaukee, Wis.	Rochester, N.Y.
Philadelphia, Pa.-N.J.	San Antonio, Tex.
Portland, Oreg.-Wash.	San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.	Springfield-Chicopee-Holyoke, Mass.-Conn.
San Francisco-Oakland, Calif.	

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

For some Year II SMSA's (i.e., Colorado Springs, Colo., Miami, Fla., and San Diego, Calif.) a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1975 housing inventory to an independent estimate of the SMSA's October 1975 housing inventory. This ratio estimate factor equaled the following:

$$\frac{\text{Independent estimate of the October 1975 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using the 1970 census estimate of total housing units for the SMSA in conjunction with the estimate of the change in the housing inventory since the census. The estimate of change was based on either administrative records from utility companies, or, when utility data were not available, on estimates of new construction permits and post-census demolition data. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, are brought into agreement with known estimates of the SMSA housing population.

The third-stage ratio estimation procedure was not employed in all Year II SMSA's since the reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. The independent estimates were only employed in SMSA's where the estimated relative bias was thought to be low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results. A measure of the relative bias of the independent estimate was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1975 independent estimates. In some SMSA's, the measure of the relative bias of the independent estimate was less than 3 percent but the independent estimate was still not employed, since there was evidence that suggested that the procedure for generating the independent estimates was producing substantial overestimates for these SMSA's.

1970-1975 lost units.—The AHS estimate of characteristics of the 1970-1975 lost units employed a one-stage ratio estimation procedure which is similar to the first-stage ratio estimation procedure mentioned above. The 1970-1975 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units + Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for State, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and non-

sampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

Nonsampling errors.—Nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1975 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors associated with 1970 census estimates—"coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

AHS-SMSA.—For the 1975 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are not yet available, however, it is believed that they will be similar to the results of the Year I AHS-SMSA reinterview study which were presented in the following Census Bureau memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample: 1974."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year I SMSA's and not for any specific SMSA):

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. The estimated indexes of inconsistency for most of the reinterview items were in the 20 to 50 range. The 20-50 interval is considered moderate indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview program because they had not been previously tested.

3. "Our bias indicator, the net difference rate, shows no signs of serious bias. For only one characteristic was there as much as a 2½ percent probability that the net difference rate is greater than 10 percent."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors.—With errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies in the representation of both conventional new construction and new mobile homes (and trailers) in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later, were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and

therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 2,800 conventional new construction units in this SMSA had permits issued prior to January 1970, and therefore, were missed by the 1975 AHS-SMSA survey.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. Although it is not known exactly, an estimated 700 new mobile homes in permit-issuing areas were missed by the 1975 AHS-SMSA survey in this SMSA. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS data still exists.

Rounding errors.—The rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling error, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in tables I and II are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1975 Housing Inventory and for Estimated Number of 1970-1975 Lost Units for the Rochester, N.Y., SMSA

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100. . . .	80	5,000. .	590
200. . . .	120	10,000.	830
500. . . .	190	25,000.	1,290
700. . . .	220	50,000.	1,770
1,000 . .	270	100,000	2,320
2,500 . .	420	250,000	2,660

Tables I and II present the standard errors applicable to estimates of characteristics of the 1975 housing inventory as well as estimates of characteristics of the 1970-1975 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in these tables.

For ratios, (100) (x/y), where x is not a subclass of y, table II underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be ob-

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the Rochester, N.Y., SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
100	8.4	18.4	25.3	36.5	42.1
200	5.9	13.0	17.9	25.8	29.8
500	3.7	8.2	11.3	16.3	18.8
700	3.2	6.9	9.6	13.8	15.9
1,000	2.7	5.8	8.0	11.5	13.3
2,500	1.7	3.7	5.1	7.3	8.4
5,000	1.2	2.6	3.6	5.2	6.0
10,000	.8	1.8	2.5	3.6	4.2
25,000	.5	1.2	1.6	2.3	2.7
50,000	.4	.8	1.1	1.6	1.9
100,000	.3	.6	.8	1.2	1.3
250,000	.2	.4	.5	.7	.8

tained by letting the standard error of the ratio be approximately equal to:

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2+\left(\frac{\sigma_y}{y}\right)^2}$$

Let: x = the numerator of the ratio

y = the denominator of the ratio

σ_x = the standard error of the numerator

σ_y = the standard error of the denominator

Illustration of the use of the standard error tables.—Table 1 of part A of this report shows that in 1975 there were 188,000 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 2,520. Consequently, the 68-percent confidence interval, as shown by these data, is from 185,480 to 190,520 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 183,970 to 192,030 housing units with 90 percent confidence; and that the average estimate lies within the interval from 182,960 to 193,040 housing units with 95 percent confidence.

Table 1 of part A also shows that of the 188,000 owner-occupied housing units 33,000, or 17.6 percent, had two bedrooms. Interpolation in table II of this appendix shows that the standard error of the percent is approximately .8 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 16.8 to 18.4 percent; the 90-percent confidence interval is from 16.3 to 18.9 percent; and the 95-percent confidence interval is from 16.0 to 19.2 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; while, if there is a high negative correlation, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table 1 of part A of this report shows that in 1975 there were 96,200 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 63,200. Table I shows the standard error of 33,000 is approximately 1,440 and the standard error of 96,200 is approximately 2,280. Therefore, the standard error of the estimated difference of 63,200 is about

$$2{,}700 = \sqrt{(1{,}440)^2 + (2{,}280)^2}$$

Consequently, the 68-percent confidence interval for the 63,200 difference is from 60,500 to 65,900 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 58,880 to 67,520 housing units, and the 95-percent confidence interval is from 57,800 to 68,600. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval does not include zero or negative values.

Medians.—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From table II determine the standard error of a 50-percent characteristic on the base of the median;
2. add to and subtract from 50 percent, the standard error determined in step 1; and
3. using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median.—Table 1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.1. The base of the distribution from which this median was determined is 188,000 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 188,000 is approximately 1.0 percentage points.
2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 48.0 and 52.0.
3. From table 1 of part A, it can be seen by cumulating the frequencies for the first two categories that 75,600 owner-occupied housing units, or 40.2 percent, had one or two persons (for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 30,700 owner-occupied housing units, or 16.3 percent, had three persons (i.e., 2.5 to 3.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{48.0-40.2}{16.3}\right) = 3.0$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about.

$$2.5 + (3.5-2.5)\left(\frac{52.0-40.2}{16.3}\right) = 3.2$$

Thus, the 95-percent confidence interval ranges from 3.0 to 3.2 persons.

San Antonio 1975

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-41
Annual Housing Survey App-41
Selection of the sample. App-41
Building loss sample selection . . . App-42
1970 Census of Population and Housing. App-43
ESTIMATION. App-43
1975 housing inventory App-43
1970-1975 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing. App-44
RELIABILITY OF THE ESTIMATES. App-44
Nonsampling errors. App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-45
Rounding errors. App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences. App-47
Illustration of the computation of the standard error of a difference. App-47
Medians App-47
Illustration of the computation of the 95-percent confidence interval of a median App-48

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 21 SMSA's are based on data collected from the 1975 Annual Housing Survey (AHS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development. In each of the 21 SMSA's, the data were collected for the 12-month period from April 1975 through March 1976, with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and a third group of 20 SMSA's were collected from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated for the AHS every three years on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the second group (Year II SMSA's) are: Atlanta, Ga., Chicago, Ill., Philadelphia, Pa.-N.J., and San Francisco-Oakland, Calif.

The remaining SMSA's in the second group are: Cincinnati, Ohio-Ky.-Ind., Colorado Springs, Colo., Columbus, Ohio, Hartford, Conn., Kansas City, Mo.-Kans., Miami, Fla., Milwaukee, Wis., New Orleans, La., Newport News-Hampton, Va., Paterson-Clifton-Passaic, N.J., Portland, Oreg.-Wash., Rochester, N.Y., San Antonio, Tex., San Bernardino-Riverside-Ontario, Calif., San Diego, Calif., Springfield-Chicopee-Holyoke, Mass.-Conn., and Madison, Wis.

In this SMSA, 4,883 units were eligible for interview. Of these sample units, 170 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 421 units were visited but found not to be eligible for interview, because these units were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit issuing was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for those SMSA's which are *not* 100-percent permit issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). The following table indicates which SMSA's are 100-percent permit issuing and which contain a sample from the nonpermit universe.

100 Percent permit-issuing SMSA's	SMSA's with a sample from the nonpermit universe
Hartford, Conn.	Atlanta, Ga.
Miami, Fla.	Chicago, Ill.
Newport News-Hampton, Va.	Cincinnati, Ohio-Ky.-Ind.
Paterson-Clifton-Passaic, N.J.	Colorado Springs, Colo.
San Bernardino-Riverside-Ontario, Calif.	Columbus, Ohio
San Diego, Calif.	Kansas City, Mo.-Kans.
San Francisco-Oakland, Calif.	Madison, Wis.
	Milwaukee, Wis.
	New Orleans, La.
	Philadelphia, Pa.-N.J.
	Portland, Oreg.-Wash.
	Rochester, N.Y.
	San Antonio, Tex.
	Springfield-Chicopee-Holyoke, Mass.-Conn.

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of these sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and the balance of the SMSA, since the

sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—family size	Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000–$5,999 . .		
$6,000–$9,999 . .		
$10,000–$14,999		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. The housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued, and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \frac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Some tables in this report provide estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listings of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at enumeration time (i.e., the 1975 housing inventory), and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1975 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1975 housing inventory.—The AHS estimates of characteristics of the 1975 housing inventory employed a three-stage ratio estimation procedure. However, the third-stage ratio estimation procedure was employed in only three SMSA's. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 170 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units and one noninterview cell for the nonpermit universe (if applicable). Sample housing units from the nonpermit universe identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce the variation in sample size for strata used in the sample selection of the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes built after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or by other sources). This procedure was designed to adjust the AHS sample estimates of these types of units to independently derived current estimates. These independent estimates were considered to be the best estimates available. The adjustment was necessary to correct known deficiencies in the AHS sample in the representation of these units (see section on nonsampling error). The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1975 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimates of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units	
New mobile homes	
"Other additions"	

The numerators of the ratios were derived by applying the following factors:

1. For the "conventional new construction units" cell, either a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in the SMSA, or the missed conventional new construction was estimated from the Survey of Construction (SOC). (The table that follows indicates the procedure used to estimate the missing conventional new construction for each SMSA.)
2. For the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied.
3. For the "other additions" cell, the "other additions" rate from SCARF[1] was applied.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

SMSA's employing the SOC estimate of missed conventional new construction	SMSA's employing the national trend for missed conventional new construction
Atlanta, Ga.	Colorado Springs, Colo.
Chicago, Ill.	Hartford, Conn.
Cincinnati, Ohio-Ky.-Ind.	Madison, Wis.
Columbus, Ohio	New Orleans, La.
Kansas City, Mo.-Kans.	Newport News–Hampton, Va.
Miami, Fla.	Paterson-Clifton-Passaic, N.J.
Milwaukee, Wis.	Rochester, N.Y.
Philadelphia, Pa.-N.J.	San Antonio, Tex.
Portland, Oreg.-Wash.	San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.	Springfield-Chicopee-Holyoke, Mass.-Conn.
San Francisco-Oakland, Calif.	

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

For some Year II SMSA's (i.e., Colorado Springs, Colo., Miami, Fla., and San Diego, Calif.) a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1975 housing inventory to an independent estimate of the SMSA's October 1975 housing inventory. This ratio estimate factor equaled the following:

$$\frac{\text{Independent estimate of the October 1975 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using the 1970 census estimate of total housing units for the SMSA in conjunction with the estimate of the change in the housing inventory since the census. The estimate of change was based on either administrative records from utility companies, or, when utility data were not available, on estimates of new construction permits and post-census demolition data. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, are brought into agreement with known estimates of the SMSA housing population.

The third-stage ratio estimation procedure was not employed in all Year II SMSA's since the reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. The independent estimates were only employed in SMSA's where the estimated relative bias was thought to be low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results. A measure of the relative bias of the independent estimate was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1975 independent estimates. In some SMSA's, the measure of the relative bias of the independent estimate was less than 3 percent but the independent estimate was still not employed, since there was evidence that suggested that the procedure for generating the independent estimates was producing substantial overestimates for these SMSA's.

1970-1975 lost units.—The AHS estimate of characteristics of the 1970-1975 lost units employed a one-stage ratio estimation procedure which is similar to the first-stage ratio estimation procedure mentioned above. The 1970-1975 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units + Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for State, Cities, and Counties**, Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and non-

sampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Nonsampling errors:—Nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1975 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors associated with 1970 census estimates—"coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census,** and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

AHS-SMSA.—For the 1975 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are not yet available, however, it is believed that they will be similar to the results of the Year I AHS-SMSA reinterview study which were presented in the following Census Bureau memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample: 1974."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year I SMSA's and not for any specific SMSA):

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. The estimated indexes of inconsistency for most of the reinterview items were in the 20 to 50 range. The 20-50 interval is considered moderate indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview program because they had not been previously tested.

3. "Our bias indicator, the net difference rate, shows no signs of serious bias. For only one characteristic was there as much as a 2½ percent probability that the net difference rate is greater than 10 percent."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors.—With errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies in the representation of both conventional new construction and new mobile homes (and trailers) in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later, were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and

therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 2,900 conventional new construction units in this SMSA had permits issued prior to January 1970, and therefore, were missed by the 1975 AHS-SMSA survey.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. Although it is not known exactly, an estimated 400 new mobile homes in permit-issuing areas were missed by the 1975 AHS-SMSA survey in this SMSA. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS data still exists.

Rounding errors.—The rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling error, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in tables I and II are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1975 Housing Inventory and for Estimated Number of 1970-1975 Lost Units for the San Antonio, Texas, SMSA

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100. . . .	80	5,000. .	600
200. . . .	120	10,000.	840
500. . . .	190	25,000.	1,310
700. . . .	220	50,000.	1,810
1,000 . .	270	100,000	2,420
2,500 . .	420	250,000	3,090

Tables I and II present the standard errors applicable to estimates of characteristics of the 1975 housing inventory as well as estimates of characteristics of the 1970-1975 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in these tables.

For ratios, (100) (x/y), where x is not a subclass of y, table II underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be obtained by letting the

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the San Antonio, Texas, SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
100	8.4	18.5	25.5	36.8	42.4
200	6.0	13.1	18.0	26.0	30.0
500	3.8	8.3	11.4	16.4	19.0
700	3.2	7.0	9.6	13.9	16.0
1,000	2.7	5.9	8.1	11.6	13.4
2,500	1.7	3.7	5.1	7.4	8.5
5,000	1.2	2.6	3.6	5.2	6.0
10,000	.8	1.9	2.5	3.7	4.2
25,000	.5	1.2	1.6	2.3	2.7
50,000	.4	.8	1.1	1.6	1.9
100,000	.3	.6	.8	1.2	1.3
250,000	.2	.4	.5	.7	.8

standard error of the ratio be approximately equal to:

$$(100)\,(x/y)\,\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

Let: x = the numerator of the ratio
y = the denominator of the ratio
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

Illustration of the use of the standard error tables.—Table 1 of part A of this report shows that in 1975 there were 172,600 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 2,740. Consequently, the 68-percent confidence interval, as shown by these data, is from 169,860 to 175,340 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval 168,220 to 176,980 housing units with 90 percent confidence; and that the average estimate lies within the interval 167,120 to 178,080 housing units with 95 percent confidence.

Table 1 of part A also shows that of the 172,600 owner-occupied housing units, 54,500, or 31.6 percent, had two bedrooms. Interpolation in table II of this appendix shows that the standard error of the percent is approximately 1.0 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 30.6 to 32.6 percent; the 90-percent confidence interval is from 30.0 to 33.2 percent; and the 95-percent confidence interval is from 29.6 to 33.6 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics; the formula will overestimate the true standard error; while, if there is a high negative correlation, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table 1 of part A of this report shows that in 1975 there were 89,100 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 34,600. Table I shows the standard error of 54,500 is approximately 1,860 and the standard error of 89,100 is approximately 2,290. Therefore, the standard error of the estimated difference of 34,600 is about

$$2{,}950 = \sqrt{(1{,}860)^2 + (2{,}290)^2}$$

Consequently, the 68-percent confidence interval for the 34,600 difference is from 31,650 to 37,550 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 29,880 to 39,320 housing units, and the 95-percent confidence interval is from 28,700 to 40,500. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval does not include zero or negative values.

Medians.—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From table II determine the standard error of a 50-percent characteristic on the base of the median;
2. add to and subtract from 50 percent, the standard error determined in step 1; and
3. using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1: For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval of a median.— Table 1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.0. The base of the distribution from which this median was determined is 172,600 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 172,600 is approximately 1.1 percentage points.
2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 47.8 and 52.2.
3. From table 1 of part A, it can be seen by cumulating the frequencies for the first two categories that 68,600 owner-occupied housing units, or 39.7 percent, had one or two persons (for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 32,000 owner-occupied housing units, or 18.5 percent had three persons (i.e., 2.5 to 3.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{47.8-39.7}{18.5}\right) = 2.9$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{52.2-39.7}{18.5}\right) = 3.2$$

Thus, the 95-percent confidence interval ranges from 2.9 to 3.2 persons.

San Diego 1975

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-41
Annual Housing Survey App-41
Selection of the sample. App-41
Building loss sample selection . . . App-42
1970 Census of Population and Housing. App-43
ESTIMATION. App-43
1975 housing inventory App-43
1970-1975 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing. App-44
RELIABILITY OF THE ESTIMATES. App-44
Nonsampling errors. App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-45
Rounding errors. App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences. App-47
Illustration of the computation of the standard error of a difference. App-47
Medians App-47
Illustration of the computation of the 95-percent confidence interval of a median App-48

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 21 SMSA's are based on data collected from the 1975 Annual Housing Survey (AHS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development. In each of the 21 SMSA's, the data were collected for the 12-month period from April 1975 through March 1976, with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and a third group of 20 SMSA's were collected from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated for the AHS every three years on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the second group (Year II SMSA's) are: Atlanta, Ga., Chicago, Ill., Philadelphia, Pa.-N.J., and San Francisco-Oakland, Calif.

The remaining SMSA's in the second group are: Cincinnati, Ohio-Ky.-Ind., Colorado Springs, Colo., Columbus, Ohio, Hartford, Conn., Kansas City, Mo.-Kans., Miami, Fla., Milwaukee, Wis., New Orleans, La., Newport News-Hampton, Va., Paterson-Clifton-Passaic, N.J., Portland, Oreg.-Wash., Rochester, N.Y., San Antonio, Tex., San Bernardino-Riverside-Ontario, Calif., San Diego, Calif., Springfield-Chicopee-Holyoke, Mass.-Conn., and Madison, Wis.

In this SMSA, 4,556 units were eligible for interview. Of these sample units, 212 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 352 units were visited but found not to be eligible for interview, because these units were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit issuing was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for those SMSA's which are *not* 100-percent permit issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). The following table indicates which SMSA's are 100-percent permit issuing and which contain a sample from the nonpermit universe.

100 Percent permit-issuing SMSA's	SMSA's with a sample from the nonpermit universe
Hartford, Conn.	Atlanta, Ga.
Miami, Fla.	Chicago, Ill.
Newport News-Hampton, Va.	Cincinnati, Ohio-Ky.-Ind.
Paterson-Clifton-Passaic, N.J.	Colorado Springs, Colo.
San Bernardino-Riverside-Ontario, Calif.	Columbus, Ohio
San Diego, Calif.	Kansas City, Mo.-Kans.
San Francisco-Oakland, Calif.	Madison, Wis.
	Milwaukee, Wis.
	New Orleans, La.
	Philadelphia, Pa.-N.J.
	Portland, Oreg.-Wash.
	Rochester, N.Y.
	San Antonio, Tex.
	Springfield-Chicopee-Holyoke, Mass.-Conn.

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of these sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and the balance of the SMSA, since the

sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—family size	Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000—$5,999 . .		
$6,000—$9,999 . .		
$10,000—$14,999		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. The housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued, and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \frac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Some tables in this report provide estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at enumeration time (i.e., the 1975 housing inventory), and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1975 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1975 housing inventory.—The AHS estimates of characteristics of the 1975 housing inventory employed a three-stage ratio estimation procedure. However, the third-stage ratio estimation procedure was employed in only three SMSA's. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 212 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units and one noninterview cell for the nonpermit universe (if applicable). Sample housing units from the nonpermit universe identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce the variation in sample size for strata used in the sample selection of the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes built after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or by other sources). This procedure was designed to adjust the AHS sample estimates of these types of units to independently derived current estimates. These independent estimates were considered to be the best estimates available. The adjustment was necessary to correct known deficiencies in the AHS sample in the representation of these units (see section on nonsampling error). The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1975 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimates of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units	
New mobile homes	
"Other additions"	

The numerators of the ratios were derived by applying the following factors:

1. For the "conventional new construction units" cell, either a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in the SMSA, or the missed conventional new construction was estimated from the Survey of Construction (SOC). (The table that follows indicates the procedure used to estimate the missing conventional new construction for each SMSA.)
2. For the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied.
3. For the "other additions" cell, the "other additions" rate from SCARF[1] was applied.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

SMSA's employing the SOC estimate of missed conventional new construction	SMSA's employing the national trend for missed conventional new construction
Atlanta, Ga.	Colorado Springs, Colo.
Chicago, Ill.	Hartford, Conn.
Cincinnati, Ohio-Ky.-Ind.	Madison, Wis.
Columbus, Ohio	New Orleans, La.
Kansas City, Mo.-Kans.	Newport News—Hampton, Va.
Miami, Fla.	Paterson-Clifton-Passaic, N.J.
Milwaukee, Wis.	Rochester, N.Y.
Philadelphia, Pa.-N.J.	San Antonio, Tex.
Portland, Oreg.-Wash.	San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.	Springfield-Chicopee-Holyoke, Mass.-Conn.
San Francisco-Oakland, Calif.	

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

For some Year II SMSA's (i.e., Colorado Springs, Colo., Miami, Fla., and San Diego, Calif.) a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1975 housing inventory to an independent estimate of the SMSA's October 1975 housing inventory. This ratio estimate factor equaled the following:

$$\frac{\text{Independent estimate of the October 1975 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using the 1970 census estimate of total housing units for the SMSA in conjunction with the estimate of the change in the housing inventory since the census. The estimate of change was based on either administrative records from utility companies, or, when utility data were not available, on estimates of new construction permits and post-census demolition data. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, are brought into agreement with known estimates of the SMSA housing population.

The third-stage ratio estimation procedure was not employed in all Year II SMSA's since the reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. The independent estimates were only employed in SMSA's where the estimated relative bias was thought to be low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results. A measure of the relative bias of the independent estimate was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1975 independent estimates. In some SMSA's, the measure of the relative bias of the independent estimate was less than 3 percent but the independent estimate was still not employed, since there was evidence that suggested that the procedure for generating the independent estimates was producing substantial overestimates for these SMSA's.

1970-1975 lost units.—The AHS estimate of characteristics of the 1970-1975 lost units employed a one-stage ratio estimation procedure which is similar to the first-stage ratio estimation procedure mentioned above. The 1970-1975 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units + Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for State, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and non-

sampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Nonsampling errors.—Nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1975 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors associated with 1970 census estimates—"coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census,** and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

AHS-SMSA.—For the 1975 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are not yet available, however, it is believed that they will be similar to the results of the Year I AHS-SMSA reinterview study which were presented in the following Census Bureau memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample: 1974."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year I SMSA's and not for any specific SMSA):

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. The estimated indexes of inconsistency for most of the reinterview items were in the 20 to 50 range. The 20-50 interval is considered moderate indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview program because they had not been previously tested.

3. "Our bias indicator, the net difference rate, shows no signs of serious bias. For only one characteristic was there as much as a 2½ percent probability that the net difference rate is greater than 10 percent."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors.—With errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies in the representation of both conventional new construction and new mobile homes (and trailers) in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later, were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and

therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 8,900 conventional new construction units in this SMSA had permits issued prior to January 1970, and therefore, were missed by the 1975 AHS-SMSA survey.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. Although it is not known exactly, an estimated 13,400 new mobile homes in permit-issuing areas were missed by the 1975 AHS-SMSA survey in this SMSA. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS data still exists.

Rounding errors.—The rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling error, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in tables I and II are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1975 Housing Inventory and for Estimated Number of 1970-1975 Lost Units for the San Diego, Calif., SMSA

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	120	10,000	1,220
200	170	25,000	1,900
500	270	50,000	2,630
700	320	100,000	3,530
1,000	390	250,000	4,630
2,500	610	400,000	4,320
5,000	860	500,000	3,210

Tables I and II present the standard errors applicable to estimates of characteristics of the 1975 housing inventory as well as estimates of characteristics of the 1970-1975 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in these tables.

For ratios, (100) (x/y), where x is not a subclass of y, table II underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the San Diego, Calif., SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
100	12.2	26.8	36.9	53.2	61.4
200	8.6	18.9	26.1	37.6	43.4
500	5.5	12.0	16.5	23.8	27.5
700	4.6	10.1	13.9	20.1	23.2
1,000	3.9	8.5	11.7	16.8	19.4
2,500	2.4	5.4	7.4	10.6	12.3
5,000	1.7	3.8	5.2	7.5	8.7
10,000	1.2	2.7	3.7	5.3	6.1
25,000	.8	1.7	2.3	3.4	3.9
50,000	.5	1.2	1.6	2.4	2.7
100,000	.4	.8	1.2	1.7	1.9
250,000	.2	.5	.7	1.1	1.2
400,000	.2	.4	.6	.8	1.0
500,000	.2	.4	.5	.8	.9

be obtained by letting the standard error of the ratio be approximately equal to:

$$(100)\ (x/y)\ \sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

Let: x = the numerator of the ratio
y = the denominator of the ratio
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

Illustration of the use of the standard error tables.—Table 1 of part A of this report shows that in 1975 there were 304,500 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 4,520. Consequently, the 68-percent confidence interval, as shown by these data, is from 299,980 to 309,020 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 297,270 to 311,730 housing units with 90 percent confidence; and that the average estimate lies within the interval from 295,460 to 313,540 housing units with 95 percent confidence.

Table 1 of part A also shows that of the 304,500 owner-occupied housing units 86,600, or 28.4 percent, had two bedrooms. Interpolation in table II of this appendix shows that the standard error of the percent is approximately 1.0 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 27.4 to 29.4 percent; the 90-percent confidence interval is from 26.8 to 30.0 percent; and the 95-percent confidence interval is from 26.4 to 30.4 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; while, if there is a high negative correlation, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table 1 of part A of this report shows that in 1975 there were 135,600 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 49,000. Table I shows the standard error of 86,600 is approximately 3,290 and the standard error of 135,600 is approximately 3,790. Therefore, the standard error of the estimated difference of 49,000 is about

$$5{,}020 = \sqrt{(3{,}290)^2 + (3{,}790)^2}$$

Consequently, the 68-percent confidence interval for the 49,000 difference is from 43,980 to 54,020 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 40,970 to 57,030 housing units, and the 95-percent confidence interval is from 38,960 to 59,040. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval does not include zero or negative values.

Medians.—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies

within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From table II determine the standard error of a 50-percent characteristic on the base of the median;
2. add to and subtract from 50 percent, the standard error determined in step 1; and
3. using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval of a median.—Table 1 of part A of this report shows the median number of persons for owner-occupied housing units is 2.5. The base of the distribution from which this median was determined is 304,500 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 304,500 is approximately 1.1 percentage points.
2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 47.8 and 52.2.
3. From table 1 of part A, it can be seen that 42,000 owner-occupied housing units, or 13.8 percent, had one person (for purposes of calculating the median, the category of one person is considered to be from .5 to 1.5 persons); that an additional 110,700 owner-occupied housing units, or 36.4 percent, had two persons (i.e., 1.5 to 2.5 persons); and that 46,000, or 15.1 percent, had three persons (i.e., 2.5 to 3.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$1.5 + (2.5-1.5)\left(\frac{47.8-13.8}{36.4}\right) = 2.4$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{52.2-13.8-36.4}{15.1}\right) = 2.6$$

Thus, the 95-percent confidence interval ranges from 2.4 to 2.6 persons.

San Francisco- Oakland 1975

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-41
Annual Housing Survey App-41
Selection of the sample. App-41
Building loss sample selection . . . App-42
1970 Census of Population and Housing. App-43
ESTIMATION. App-43
1975 housing inventory App-43
1970-1975 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing. App-44
RELIABILITY OF THE ESTIMATES. App-44
Nonsampling errors. App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-45
Rounding errors. App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences. App-47
Illustration of the computation of the standard error of a difference. App-48
Medians App-48
Illustration of the computation of the 95-percent confidence interval of a median App-48

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 21 SMSA's are based on data collected from the 1975 Annual Housing Survey (AHS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development. In each of the 21 SMSA's, the data were collected for the 12-month period from April 1975 through March 1976, with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and a third group of 20 SMSA's were collected from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated for the AHS every three years on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the second group (Year II SMSA's) are: Atlanta, Ga., Chicago, Ill., Philadelphia, Pa.-N.J., and San Francisco-Oakland, Calif.

The remaining SMSA's in the second group are: Cincinnati, Ohio-Ky.-Ind., Colorado Springs, Colo., Columbus, Ohio, Hartford, Conn., Kansas City, Mo.-Kans., Miami, Fla., Milwaukee, Wis., New Orleans, La., Newport News-Hampton, Va., Paterson-Clifton-Passaic, N.J., Portland, Oreg.-Wash., Rochester, N.Y., San Antonio, Tex., San Bernardino-Riverside-Ontario, Calif., San Diego, Calif., Springfield-Chicopee-Holyoke, Mass.-Conn., and Madison, Wis.

In this SMSA, 14,523 units were eligible for interview. Of these sample units, 861 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 934 units were visited but found not to be eligible for interview, because these units were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit issuing was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for those SMSA's which are *not* 100-percent permit issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). The following table indicates which SMSA's are 100-percent permit issuing and which contain a sample from the nonpermit universe.

100 Percent permit-issuing SMSA's	SMSA's with a sample from the nonpermit universe
Hartford, Conn.	Atlanta, Ga.
Miami, Fla.	Chicago, Ill.
Newport News-Hampton, Va.	Cincinnati, Ohio-Ky.-Ind.
Paterson-Clifton-Passaic, N.J.	Colorado Springs, Colo.
San Bernardino-Riverside-Ontario, Calif.	Columbus, Ohio
San Diego, Calif.	Kansas City, Mo.-Kans.
San Francisco-Oakland, Calif.	Madison, Wis.
	Milwaukee, Wis.
	New Orleans, La.
	Philadelphia, Pa.-N.J.
	Portland, Oreg.-Wash.
	Rochester, N.Y.
	San Antonio, Tex.
	Springfield-Chicopee-Holyoke, Mass.-Conn.

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of these sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and the balance of the SMSA, since the

sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—family size	Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000—$5,999 . .		
$6,000—$9,999 . .		
$10,000—$14,999		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. The housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued, and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \frac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Some tables in this report provide estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at enumeration time (i.e., the 1975 housing inventory), and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1975 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1975 housing inventory.—The AHS estimates of characteristics of the 1975 housing inventory employed a three-stage ratio estimation procedure. However, the third-stage ratio estimation procedure was employed in only three SMSA's. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 861 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units and one noninterview cell for the nonpermit universe (if applicable). Sample housing units from the nonpermit universe identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce the variation in sample size for strata used in the sample selection of the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes built after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or by other sources). This procedure was designed to adjust the AHS sample estimates of these types of units to independently derived current estimates. These independent estimates were considered to be the best estimates available. The adjustment was necessary to correct known deficiencies in the AHS sample in the representation of these units (see section on nonsampling error). The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1975 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimates of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units	
New mobile homes	
"Other additions"	

The numerators of the ratios were derived by applying the following factors:

1. For the "conventional new construction units" cell, either a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in the SMSA, or the missed conventional new construction was estimated from the Survey of Construction (SOC). (The table that follows indicates the procedure used to estimate the missing conventional new construction for each SMSA.)
2. For the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied.
3. For the "other additions" cell, the "other additions" rate from SCARF[1] was applied.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

SMSA's employing the SOC estimate of missed conventional new construction	SMSA's employing the national trend for missed conventional new construction
Atlanta, Ga.	Colorado Springs, Colo.
Chicago, Ill.	Hartford, Conn.
Cincinnati, Ohio-Ky.-Ind.	Madison, Wis.
Columbus, Ohio	New Orleans, La.
Kansas City, Mo.-Kans.	Newport News—Hampton, Va.
Miami, Fla.	Paterson-Clifton-Passaic, N.J.
Milwaukee, Wis.	Rochester, N.Y.
Philadelphia, Pa.-N.J.	San Antonio, Tex.
Portland, Oreg.-Wash.	San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.	Springfield-Chicopee-Holyoke, Mass.-Conn.
San Francisco-Oakland, Calif.	

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

For some Year II SMSA's (i.e., Colorado Springs, Colo., Miami, Fla., and San Diego, Calif.) a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1975 housing inventory to an independent estimate of the SMSA's October 1975 housing inventory. This ratio estimate factor equaled the following:

$$\frac{\text{Independent estimate of the October 1975 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using the 1970 census estimate of total housing units for the SMSA in conjunction with the estimate of the change in the housing inventory since the census. The estimate of change was based on either administrative records from utility companies, or, when utility data were not available, on estimates of new construction permits and post-census demolition data. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, are brought into agreement with known estimates of the SMSA housing population.

The third-stage ratio estimation procedure was not employed in all Year II SMSA's since the reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. The independent estimates were only employed in SMSA's where the estimated relative bias was thought to be low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results. A measure of the relative bias of the independent estimate was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1975 independent estimates. In some SMSA's, the measure of the relative bias of the independent estimate was less than 3 percent but the independent estimate was still not employed, since there was evidence that suggested that the procedure for generating the independent estimates was producing substantial overestimates for these SMSA's.

1970-1975 lost units.—The AHS estimate of characteristics of the 1970-1975 lost units employed a one-stage ratio estimation procedure which is similar to the first-stage ratio estimation procedure mentioned above. The 1970-1975 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units + Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for State, Cities, and Counties**, Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and non-

sampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Nonsampling errors.—Nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1975 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors associated with 1970 census estimates—"coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census,** and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

AHS-SMSA.—For the 1975 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are not yet available, however, it is believed that they will be similar to the results of the Year I AHS-SMSA reinterview study which were presented in the following Census Bureau memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample: 1974."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year I SMSA's and not for any specific SMSA):

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. The estimated indexes of inconsistency for most of the reinterview items were in the 20 to 50 range. The 20-50 interval is considered moderate indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview program because they had not been previously tested.

3. "Our bias indicator, the net difference rate, shows no signs of serious bias. For only one characteristic was there as much as a 2½ percent probability that the net difference rate is greater than 10 percent."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors.—With errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies in the representation of both conventional new construction and new mobile homes (and trailers) in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later, were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and

therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 7,200 conventional new construction units in this SMSA had permits issued prior to January 1970, and therefore, were missed by the 1975 AHS-SMSA survey.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. Although it is not known exactly, an estimated 2,300 new mobile homes in permit-issuing areas were missed by the 1975 AHS-SMSA survey in this SMSA. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS data still exists.

Rounding errors.—The rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling error, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in tables I and II are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1975 Housing Inventory and for Estimated Number of 1970-1975 Lost Units for the San Francisco-Oakland, Calif., SMSA, for the Central Cities and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central cities	Not in central cities
200.	140	130	160
500.	230	200	250
700.	270	240	290
1,000	320	280	350
2,500	500	440	550
5,000	710	630	780
10,000 . . .	1,010	880	1,100
25,000 . . .	1,580	1,370	1,730
50,000 . . .	2,220	1,900	2,420
100,000. . .	3,080	2,540	3,320
250,000. . .	4,620	3,270	4,790
400,000. . .	5,500	2,900	5,410
500,000. . .	5,870	—	5,510
700,000. . .	6,260	—	5,020
1,000,000 .	6,020	—	—

Tables I through IV present the standard errors applicable to estimates of characteristics of the 1975 housing in-

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the San Francisco-Oakland, Calif., SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
200	7.1	15.5	21.4	30.9	35.6
500	4.5	9.8	13.5	19.5	22.5
700	3.8	8.3	11.4	16.5	19.1
1,000	3.2	6.9	9.6	13.8	15.9
2,500	2.0	4.4	6.0	8.7	10.1
5,000	1.4	3.1	4.3	6.2	7.1
10,000	1.0	2.2	3.0	4.4	5.0
25,000	.6	1.4	1.9	2.8	3.2
50,000	.4	1.0	1.4	2.0	2.3
100,000	.3	.7	1.0	1.4	1.6
250,000	.2	.4	.6	.9	1.0
400,000	.2	.3	.5	.7	.8
500,000	.14	.3	.4	.6	.7
700,000	.12	.3	.4	.5	.6
1,000,000	.10	.2	.3	.4	.5

ventory as well as estimates of characteristics of the 1970-1975 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in these tables.

For ratios, (100) (x/y), where x is not a subclass of y, tables II through IV underestimate the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be obtained by letting the standard error of the ratio be approximately equal to:

$$(100)\,(x/y)\,\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

Let: x = the numerator of the ratio

y = the denominator of the ratio

σ_x = the standard error of the numerator

σ_y = the standard error of the denominator

Illustration of the use of the standard error tables.—Table A-1 of part A of this report shows that in 1975 there were 620,000 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 6,100. Consequently, the 68-percent confidence interval, as shown by these data, is from 613,900 to 626,100 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 610,240 to 629,760 housing units with 90 percent confidence; and that the average estimate lies within the interval from 607,800 to 632,200 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 620,000 owner-occupied housing units, 175,600, or 28.3 percent, had two bedrooms. Interpolation in table II of this appendix shows that the standard error of the percent is approximately .6 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 27.7 to 28.9 percent; the 90-percent confidence interval is from 27.3 to 29.3 percent; and the 95-percent confidence interval is from 27.1 to 29.5 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation

TABLE III. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the Central Cities

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
200	6.3	13.7	18.9	27.3	31.5
500	4.0	8.7	11.9	17.2	19.9
700	3.3	7.3	10.1	14.6	16.8
1,000	2.8	6.1	8.4	12.2	14.1
2,500	1.8	3.9	5.3	7.7	8.9
5,000	1.3	2.7	3.8	5.5	6.3
10,000	.9	1.9	2.7	3.9	4.5
25,000	.6	1.2	1.7	2.4	2.8
50,000	.4	.9	1.2	1.7	2.0
100,000	.3	.6	.8	1.2	1.4
250,000	.2	.4	.5	.8	.9
400,000	.14	.3	.4	.6	.7

TABLE IV. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the Balance of the SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
200.	7.8	17.1	23.5	33.9	39.2
500.	4.9	10.8	14.9	21.5	24.8
700.	4.2	9.1	12.6	18.1	21.0
1,000	3.5	7.6	10.5	15.2	17.5
2,500	2.2	4.8	6.7	9.6	11.1
5,000	1.6	3.4	4.7	6.8	7.8
10,000	1.1	2.4	3.3	4.8	5.5
25,000	.7	1.5	2.1	3.0	3.5
50,000	.5	1.1	1.5	2.1	2.5
100,000.	.3	.8	1.1	1.5	1.8
250,000.	.2	.5	.7	1.0	1.1
400,000.	.2	.4	.5	.8	.9
500,000.	.2	.3	.5	.7	.8
700,000.	.13	.3	.4	.6	.7

between the two characteristics, the formula will overestimate the true standard error; while, if there is a high negative correlation, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table A-1 of part A of this report shows that in 1975 there were 291,500 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 115,900. Table I shows the standard error of 175,600 is approximately 3,860 and the standard error of 291,500 is approximately 4,860. Therefore, the standard error of the estimated difference of 115,900 is about

$$6{,}210 = \sqrt{(3{,}860)^2 + (4{,}860)^2}$$

Consequently, the 68-percent confidence interval for the 115,900 difference is from 109,690 to 122,110 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 105,960 to 125,840 housing units, and the 95-percent confidence interval is from 103,480 to 128,320. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval does not include zero or negative values.

Medians.—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From table II determine the standard error of a 50-percent characteristic on the base of the median;
2. add to and subtract from 50 percent, the standard error determined in step 1; and
3. using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 2.7. The base of the distribution from which this median was determined is 620,000 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 620,000 is approximately .6 percentage points.
2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 48.8 and 51.2.
3. From table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 285,100 owner-occupied housing units, or 46.0 percent, had one or two persons (for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 116,700 owner-occupied housing units, or 18.8 percent, had three persons (i.e., 2.5 to 3.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{48.8-46.0}{18.8}\right) = 2.6$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{51.2-46.0}{18.8}\right) = 2.8$$

Thus, the 95-percent confidence interval ranges from 2.6 to 2.8 persons.

Springfield-Chicopee-Holyoke 1975

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-41
Annual Housing Survey App-41
Selection of the sample. App-41
Building loss sample selection . . . App-42
1970 Census of Population and Housing. App-43
ESTIMATION. App-43
1975 housing inventory App-43
1970-1975 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing. App-44
RELIABILITY OF THE ESTIMATES. App-44
Nonsampling errors. App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-45
Rounding errors. App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences. App-47
Illustration of the computation of the standard error of a difference. App-47
Medians App-47
Illustration of the computation of the 95-percent confidence interval of a median App-48

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 21 SMSA's are based on data collected from the 1975 Annual Housing Survey (AHS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development. In each of the 21 SMSA's, the data were collected for the 12-month period from April 1975 through March 1976, with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and a third group of 20 SMSA's were collected from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated for the AHS every three years on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the second group (Year II SMSA's) are: Atlanta, Ga., Chicago, Ill., Philadelphia, Pa.-N.J., and San Francisco-Oakland, Calif.

The remaining SMSA's in the second group are: Cincinnati, Ohio-Ky.-Ind., Colorado Springs, Colo., Columbus, Ohio, Hartford, Conn., Kansas City, Mo.-Kans., Miami, Fla., Milwaukee, Wis., New Orleans, La., Newport News-Hampton, Va., Paterson-Clifton-Passaic, N.J., Portland, Oreg.-Wash., Rochester, N.Y., San Antonio, Tex., San Bernardino-Riverside-Ontario, Calif., San Diego, Calif., Springfield-Chicopee-Holyoke, Mass.-Conn., and Madison, Wis.

In this SMSA, 5,239 units were eligible for interview. Of these sample units, 167 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 504 units were visited but found not to be eligible for interview, because these units were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit issuing was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for those SMSA's which are *not* 100-percent permit issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). The following table indicates which SMSA's are 100-percent permit issuing and which contain a sample from the nonpermit universe.

100 Percent permit-issuing SMSA's	SMSA's with a sample from the nonpermit universe
Hartford, Conn.	Atlanta, Ga.
Miami, Fla.	Chicago, Ill.
Newport News-Hampton, Va.	Cincinnati, Ohio-Ky.-Ind.
Paterson-Clifton-Passaic, N.J.	Colorado Springs, Colo.
San Bernardino-Riverside-Ontario, Calif.	Columbus, Ohio
San Diego, Calif.	Kansas City, Mo.-Kans.
San Francisco-Oakland, Calif.	Madison, Wis.
	Milwaukee, Wis.
	New Orleans, La.
	Philadelphia, Pa.-N.J.
	Portland, Oreg.-Wash.
	Rochester, N.Y.
	San Antonio, Tex.
	Springfield-Chicopee-Holyoke, Mass.-Conn.

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of these sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and the balance of the SMSA, since the

sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—family size	Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000–$5,999 . .		
$6,000–$9,999 . .		
$10,000–$14,999		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. The housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued, and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Some tables in this report provide estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at enumeration time (i.e., the 1975 housing inventory), and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1975 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1975 housing inventory.—The AHS estimates of characteristics of the 1975 housing inventory employed a three-stage ratio estimation procedure. However, the third-stage ratio estimation procedure was employed in only three SMSA's. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 167 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units and one noninterview cell for the nonpermit universe (if applicable). Sample housing units from the nonpermit universe identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce the variation in sample size for strata used in the sample selection of the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes built after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or by other sources). This procedure was designed to adjust the AHS sample estimates of these types of units to independently derived current estimates. These independent estimates were considered to be the best estimates available. The adjustment was necessary to correct known deficiencies in the AHS sample in the representation of these units (see section on nonsampling error). The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1975 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimates of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units.	
New mobile homes	
"Other additions".	

The numerators of the ratios were derived by applying the following factors:

1. For the "conventional new construction units" cell, either a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in the SMSA, or the missed conventional new construction was estimated from the Survey of Construction (SOC). (The table that follows indicates the procedure used to estimate the missing conventional new construction for each SMSA.)
2. For the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied.
3. For the "other additions" cell, the "other additions" rate from SCARF[1] was applied.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

SMSA's employing the SOC estimate of missed conventional new construction	SMSA's employing the national trend for missed conventional new construction
Atlanta, Ga.	Colorado Springs, Colo.
Chicago, Ill.	Hartford, Conn.
Cincinnati, Ohio-Ky.-Ind.	Madison, Wis.
Columbus, Ohio	New Orleans, La.
Kansas City, Mo.-Kans.	Newport News—Hampton, Va.
Miami, Fla.	Paterson-Clifton-Passaic, N.J.
Milwaukee, Wis.	Rochester, N.Y.
Philadelphia, Pa.-N.J.	San Antonio, Tex.
Portland, Oreg.-Wash.	San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.	Springfield-Chicopee-Holyoke, Mass.-Conn.
San Francisco-Oakland, Calif.	

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

For some Year II SMSA's (i.e., Colorado Springs, Colo., Miami, Fla., and San Diego, Calif.) a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1975 housing inventory to an independent estimate of the SMSA's October 1975 housing inventory. This ratio estimate factor equaled the following:

$$\frac{\text{Independent estimate of the October 1975 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using the 1970 census estimate of total housing units for the SMSA in conjunction with the estimate of the change in the housing inventory since the census. The estimate of change was based on either administrative records from utility companies, or, when utility data were not available, on estimates of new construction permits and post-census demolition data. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, are brought into agreement with known estimates of the SMSA housing population.

The third-stage ratio estimation procedure was not employed in all Year II SMSA's since the reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. The independent estimates were only employed in SMSA's where the estimated relative bias was thought to be low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results. A measure of the relative bias of the independent estimate was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1975 independent estimates. In some SMSA's, the measure of the relative bias of the independent estimate was less than 3 percent but the independent estimate was still not employed, since there was evidence that suggested that the procedure for generating the independent estimates was producing substantial overestimates for these SMSA's.

1970-1975 lost units.—The AHS estimate of characteristics of the 1970-1975 lost units employed a one-stage ratio estimation procedure which is similar to the first-stage ratio estimation procedure mentioned above. The 1970-1975 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for State, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and non-

sampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Nonsampling errors.—Nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1975 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors associated with 1970 census estimates—"coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census,** and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

AHS-SMSA.—For the 1975 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are not yet available, however, it is believed that they will be similar to the results of the Year I AHS-SMSA reinterview study which were presented in the following Census Bureau memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample: 1974."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year I SMSA's and not for any specific SMSA):

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. The estimated indexes of inconsistency for most of the reinterview items were in the 20 to 50 range. The 20-50 interval is considered moderate indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview program because they had not been previously tested.

3. "Our bias indicator, the net difference rate, shows no signs of serious bias. For only one characteristic was there as much as a 2½ percent probability that the net difference rate is greater than 10 percent."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors.—With errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies in the representation of both conventional new construction and new mobile homes (and trailers) in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later, were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and

therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 1,500 conventional new construction units in this SMSA had permits issued prior to January 1970, and therefore, were missed by the 1975 AHS-SMSA survey.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. Although it is not known exactly, an estimated 300 new mobile homes in permit-issuing areas were missed by the 1975 AHS-SMSA survey in this SMSA. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS data still exists.

Rounding errors.—The rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling error, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in tables I and II are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1975 Housing Inventory and for Estimated Number of 1970-1975 Lost Units for the Springfield-Chicopee-Holyoke, Mass.-Conn., SMSA

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	60	5,000 . . .	440
200	90	10,000 . .	620
500	140	25,000 . .	950
700	170	50,000 . .	1,280
1,000 . .	200	100,000 .	1,590
2,500 . .	320	150,000 .	1,620

Tables I and II present the standard errors applicable to estimates of characteristics of the 1975 housing inventory as well as estimates of characteristics of the 1970-1975 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in these tables.

For ratios, (100) (x/y), where x is not a subclass of y, table II underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1975 Housing Inventory and for Estimated Percentages of 1970-1975 Lost Units for the Springfield-Chicopee-Holyoke, Mass.-Conn., SMSA

Base of percentage	Estimated percentage				
	1 or 99	5 or 95	10 or 90	25 or 75	50
100	6.3	13.8	19.0	27.5	31.7
200	4.5	9.8	13.5	19.4	22.4
500	2.8	6.2	8.5	12.3	14.2
700	2.4	5.2	7.2	10.4	12.0
1,000	2.0	4.4	6.0	8.7	10.0
2,500	1.3	2.8	3.8	5.5	6.3
5,000	.9	2.0	2.7	3.9	4.5
10,000	.6	1.4	1.9	2.7	3.2
25,000	.4	.9	1.2	1.7	2.0
50,000	.3	.6	.9	1.2	1.4
100,000	.2	.4	.6	.9	1.0
150,000	.2	.4	.5	.7	.8

obtained by letting the standard error of the ratio be approximately equal to:

$$(100)\ (x/y)\ \sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

Let: x = the numerator of the ratio
y = the denominator of the ratio
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

Illustration of the use of the standard error tables.—Table 1 of part A of this report shows that in 1975 there were 103,100 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 1,590. Consequently, the 68-percent confidence interval, as shown by these data, is from 101,510 to 104,690 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 100,560 to 105,640 housing units with 90 percent confidence; and that the average estimate lies within the interval from 99,920 to 106,280 housing units with 95 percent confidence.

Table 1 of part A also shows that of the 103,100 owner-occupied housing units 28,700, or 27.8 percent, had two bedrooms. Interpolation in table II of this appendix shows that the standard error of the percent is approximately .9 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 26.9 to 28.7 percent; the 90-percent confidence interval is from 26.4 to 29.2 percent; and the 95-percent confidence interval is from 26.0 to 29.6 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; while, if there is a high negative correlation, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table 1 of part A of this report shows that in 1975 there were 50,600 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 21,900. Table I shows the standard error of 28,700 is approximately 1,000 and the standard error of 50,600 is approximately 1,280. Therefore, the standard error of the estimated difference of 21,900 is about

$$1{,}620 = \sqrt{(1{,}000)^2 + (1{,}280)^2}$$

Consequently, the 68-percent confidence interval for the 21,900 difference is from 20,280 to 23,520 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 19,310 to 24,490 housing units, and the 95-percent confidence interval is from 18,660 to 25,140. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval does not include zero or negative values.

Medians.—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From table II determine the standard error of a 50-percent characteristic on the base of the median;

2. add to and subtract from 50 percent, the standard error determined in step 1; and

3. using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median.—Table 1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.0. The base of the distribution from which this median was determined is 103,100 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 103,100 is approximately 1.0 percentage points.

2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 48.0 and 52.0.

3. From table 1 of part A, it can be seen by cumulating the frequencies for the first two categories that 42,100 owner-occupied housing units, or 40.8 percent, had one or two persons (for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 19,100 owner-occupied housing units, or 18.5 percent, had three persons (i.e., 2.5 to 3.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{48.0-40.8}{18.5}\right) = 2.9$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{52.0-40.8}{18.5}\right) = 3.1$$

Thus, the 95-percent confidence interval ranges from 2.9 to 3.1 persons.