Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Selection of sample areas App-42
Designation of sample housing units for the 1976 survey App-42
Selection of the 1973 sample housing units App-42
Splitting of the sample App-43
Selection of supplemental sample housing units in rural areas App-43
Selection of sample housing units for the 1976 Coverage Improvement Program App-43
1970 Census of Population and Housing App-44
ESTIMATION
AHS-National sample App-44
1976 Housing inventory App-44
1973-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing App-45
RELIABILITY OF THE ESTIMATES App-45
Sampling errors App-45
Illustration of the use of the standard error tables. Illustration I App-48
Illustration II App-49
Differences App-49
Illustration of the computation of the standard error of a difference App-49
Medians App-49
Illustration of the computation of the 95-percent confidence interval for a median App-49
Nonsampling errors App-50
1970 census App-50
Reinterview program App-50
Coverage errors App-51
Rounding errors App-51

SAMPLE DESIGN

The 1976 estimates are based on data collected in October through December 1976 for the 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. The sample for this survey was spread over 461 sample areas (called primary sampling units), comprising 923 counties and independent cities with coverage in each of the 50 States and the District of Columbia.

Approximately 75,500 sample housing units (both occupied and vacant) were eligible for interview in the 1976 Annual Housing Survey. Of this number, 4,500 interviews were classified as "noninterview" for various reasons. For occupied housing units, the main reason was that the occupants were not found at home after repeated calls. For vacant housing units, interviews were not obtained because an informed respondent was not found after repeated visits. In addition to the 75,500, there were also 6,600 sample units which were visited but found not to be eligible for interview for the AHS in terms of collecting information relevant to the 1976 housing inventory.

Selection of sample areas.—The United States was divided into areas made up of counties and independent cities referred to as primary sampling units (PSU's). These PSU's were then grouped into 376 strata, 156 of which consisted of only one PSU which was in sample with certainty. These 156 strata were mostly the larger SMSA's and were called self-representing (SR) since the sample from the sample area represented just that PSU. Each one of the other 220 strata consisted of a group of PSU's and were referred to as non-self-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSU's in the stratum as well.

One PSU was selected from each NSR stratum with probability proportionate to the 1970 census population of the PSU. (This resulted in 220 NSR sample PSU's.) In addition, the NSR strata were grouped into 110 pairs and one stratum was picked at random from each pair. From this stratum, an additional PSU was selected independently of the other PSU selected from this stratum. Since the two PSU's were independently selected, it was possible for the same PSU to be selected twice. This occurred in 25 instances, producing an additional 85 NSR sample PSU's, thus giving a grand total of 461 PSU's.

Designation of sample housing units for the 1976 survey.—The sample housing units designated to be interviewed in the 1976 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1975 survey.
2. All sample housing units that were either type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of the survey but which could become eligible in the future) in the 1975 survey. (For a list of reasons for type A and type B noninterviews, see facsimile of the 1976 AHS questionnaire, page 1.)
3. All sample housing units that were selected from the list of building permits issued since the 1975 survey. (This sample represented the housing units built in permit-issuing areas, since the 1975 survey).
4. All sample housing units that were selected as part of the 1976 Coverage Improvement Program. (This sample represented most of the housing units which, until 1976, did not have a chance of selection).

Selection of the 1973 sample housing units.—The overall sampling rate used to select the sample for the 1973 AHS was about 1 in 1,366. The within-PSU sampling rate for the AHS was determined so that the overall probability of selection for each sample housing unit was the same (e.g., if the probability of selecting a NSR PSU was 1 in 10, then the within-PSU sampling rate would be 1 in 136.6).

Within the sample PSU's, a sample of the housing units enumerated in the 1970 Census of Population and Housing was selected for the AHS. In addition, a sample of new construction building permits was also selected to represent the units constructed since the 1970 census. These samples were selected at about twice the rate mentioned previously; i.e., at 2 in 1,366, thereby producing a sample twice as large as needed. This sample was split into two equal-sized samples—one to be used for the AHS, and one to be held in reserve for possible future use for the

AHS. The procedure used to split this sample into equal-sized samples is described in the next section.

The sample of 1970 census units was selected in several stages of sampling. Within the sample PSU's, the first step was the selection of a sample of census enumeration districts (ED's), administrative units used in the 1970 census. The probability of selection for an ED was proportional to the following 1970 census counts of housing units (HU's) and persons in group quarters, combined in the formula:

$$\frac{\text{Number of HU's in the ED} + \dfrac{\text{Number of group quarters persons in the ED}}{3}}{4}$$

The next step was to select an expected cluster of about four neighboring housing units within each sample ED. For most of the ED's, the selection was accomplished using the list of addresses for the ED as compiled in the 1970 census. However, in those ED's where addresses were incomplete or inadequate (mostly rural areas), the selection process was accomplished using area sampling methods. These ED's were 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) and a segment was selected. Those selected segments with an expected size which was a multiple of four were further subsampled at the time of the survey so that an expected four housing units were chosen for interview.

The sample of new construction units was selected from building permits issued since January 1970. Within each sample PSU, the building permits were chronologically ordered by month issued, and compact clusters of approximately four housing units were created. These clusters were then sampled for inclusion in the AHS at the sampling rate of 2 in 1,366. Housing units constructed since the 1970 census in areas which do not issue building permits were brought into the sample as a result of the area sample described.

Splitting of the sample.—The sample selection procedure as described above produced clusters (or segments) of size-four housing units for the sample taken from the census address frame, the new construction frame, and the area sampling frame (mainly rural areas). Clusters of this size should result in a minimum loss in precision for estimates of housing characteristics in rural areas because of the heterogeneity of neighboring units. However, clusters of size-two were considered to be more optimum within those areas where the housing characteristics of neighboring units tend to be very similar (i.e., urban areas and new construction units). A splitting operation was then carried out for clusters selected from the census address and the new construction frames. This consisted of halving each sample cluster from these frames. Thus, two housing units from each of these clusters were included in the survey and two housing units were held in reserve. No splitting operation was carried out within the clusters selected from the area sampling frame; every other area sample cluster of four housing units was used for the survey and the remaining clusters were assigned to the reserve sample.

Selection of supplemental sample housing units in rural areas.—In 1974, it was decided to increase the reliability of the AHS estimates of rural housing characteristics by doubling the number of sample housing units from rural areas. This was accomplished by reactivating the reserve sample, selected in the original sampling operations in 1973, from rural areas only. For the reserve sample selected in census address and new construction frames, this meant that the other half of each rural cluster (an expected two housing units) was reactivated in 1974. Similarly for the area sampling frame, this meant the entire reserve cluster (an expected four housing units) was reactivated in 1974 if the cluster was rural. This supplementation increased the overall probability of selection for sample housing units in rural areas to about 2 in 1,366; whereas, the overall probability of selection for sample housing units in urban areas remained at 1 in 1,366.

Selection of sample housing units for the 1976 Coverage Improvement Program.—The 1976 Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS National sample from the census address and new construction frames. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970.
2. Units converted to residential use in structures totally nonresidential at the time of the 1970 census.
3. Houses that have been moved onto their present site since the 1970 census.
4. Mobile homes placed in parks either missed in or established since the 1970 census.
5. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

A sample of new construction units whose permits were issued before January 1970 was selected in two stages. First, units whose permits were issued before January 1970, but which were completed after the census, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. These units were then sampled so that the overall probability of selection was about 1 in 1,320.

A sample of mobile homes placed in a park missed by the census or established after the census was also selected in two stages. During the first stage, a list of mobile home parks was obtained from commercial listings. This list was then supplemented by additional parks identified by a canvassing operation similar to that performed in ED's where area sampling methods are used. The second stage consisted of dividing the parks into clusters of an expected size of four sites. These clusters were then sampled so that the overall probability of selection was about 1 in 1,366.

For the remaining units (i.e., mobile homes placed outside parks since the 1970 census, mobile homes vacant at the time of the 1970 census, units converted from nonresidential to residential use since the 1970 census, and houses that had been moved onto their present site since the 1970 census), the sampling was done in three stages. First, a subsample of the regular AHS sample units from the census address frame was selected. Sec-

ond, succeeding structures that had been eligible to be selected from the census address frame were then listed until eight structures (excluding mobile home parks) were found. Finally, the intervening structures that had been listed which did not have a chance of selection in the AHS were identified and the units within these structures were interviewed.

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 can be obtained in the 1970 census report, HC(1)-B1, **Detailed Housing Characteristics, United States Summary.**

ESTIMATION

AHS-National sample.—The AHS-National sample produced estimates of two types: Estimates of the 1976 housing inventory, and estimates of units removed from the housing inventory between 1973 and 1976 (i.e., 1973-1976 lost units). Each type of estimate employed a separate, though similar, estimation procedure.

1976 housing inventory.—In 1976, the AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the type A noninterview housing units encountered in the AHS. This noninterview adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

$$\frac{\text{Interviewed housing units} + \text{Noninterviewed housing units}}{\text{Interviewed housing units}}$$

The first-stage ratio estimation procedure was employed for sample housing units from non-self-representing (NSR) PSU's only. This procedure was designed to reduce the contribution to the variance arising from the sampling of PSU's. The first-stage ratio estimation procedure takes into account the differences that existed at the time of the 1970 census in the distribution, by tenure and residence of the housing population, estimated from the sample NSR PSU's and that of the NSR housing population in each of the four census regions of the country.

The first-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{The 1970 census housing population in the residence-tenure category for all NSR strata in a census region}}{\text{Estimate of the housing population category using 1970 census housing counts for sample NSR PSU's in a census region}}$$

The numerators of the ratios were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR stratum and summing these counts across the NSR strata in each census region. The denominators were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR sample PSU, weighting these counts by the inverse of the probability of selecting that PSU, and summing these weighted counts across the NSR sample PSU's in each census region. The computed first-stage ratio estimation factor was then applied to the existing weight for each NSR sample unit in each first-stage ratio estimation category.

The second-stage ratio estimation procedure was designed to adjust the AHS sample estimate of one category of conventional new construction units; i.e., one category of sample units built April 1, 1970, or later, to an independently derived current estimate where a known deficiency in the AHS sample exists (see the section on nonsampling error) for each of the four regions. This estimate was considered to be the best estimate available for the number of conventional new construction units in this category. The second-stage ratio estimation factor was as follows:

$$\frac{\text{Current best estimate of new construction in the category}}{\text{AHS sample estimate of new construction units in the category}}$$

The numerators of the ratios were derived from data based on the Survey of Construction (SOC). The denominators of the ratios were obtained from the weighted estimates for the AHS sample units 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 each second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all the AHS sample units. This procedure was designed to adjust the AHS sample estimates of housing (i.e., the estimates employing the noninterview, first-stage, and second-stage adjustments) to independently derived current housing estimates for four categories of vacant housing units and for 24 categories for occupied housing units. Each of these 24 categories is a combination of the characteristics of residence, tenure, race of head, and sex of head. The third-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{Current independent estimate of housing units in the category}}{\text{AHS sample estimate of housing units in the category}}$$

The numerators of the ratios for occupied housing units were derived from data based on the Current Population Survey (CPS), a sample household survey conducted monthly by the Bureau of the Census. The numerators of the ratios for vacant housing units were derived from data based on the Housing Vacancy Survey (HVS), a quarterly vacancy survey also conducted by the Bureau of the Census.

The denominators of the ratios were obtained from the weighted estimates for the 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 in each third-stage ratio estimation category.

The second- and the third-stage ratio estimation procedures were repeated in an iterative process in order to bring the AHS estimates into close agreement with

both sets of "independent" estimates. The second-stage was modified so that the estimates for all 12 categories of new construction would be identical to the estimates before the third-stage. Hence, the repeated second-stage had the effect of controlling the AHS sample estimates of new construction units to the "unbiased" sample estimates for 11 categories of new construction units for each of the 4 regions (i.e., 7 categories for conventional new construction units and 4 for new construction mobile homes) and, as before, of adjusting the AHS sample estimate of I category of conventional new construction units to an independently derived current estimate.

The numerators were either the unbiased weighted estimates for the AHS sample units, using the existing weight after the first-stage ratio estimation procedure (i.e., the estimates employing the noninterview and first-stage adjustments) or the independent estimate derived from data based on the Survey of Construction (SOC).

The denominators of the ratios in this iterative process were obtained from the weighted estimates for the AHS sample units after the previous stage of ratio estimation. The factors resulting from this iterative process were then applied to the existing weight on the appropriate records, and the resulting product was used as the final weight for tabulation.

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. The distribution of the housing population selected for the sample differed somewhat, by chance, from that of the nation as a whole in such basic housing characteristics as tenure, vacancy status, residence, race of head, and sex of head. These characteristics are probably closely correlated with other housing characteristics measured for the AHS. Therefore, through the use of the three-stage ratio estimation procedure one can expect the sample estimate to be improved substantially.

1973-1976 lost units.—The 1973-1976 lost unit estimates employed the three-stage ratio estimation procedure used to produce the AHS-National estimates of the 1973 housing inventory, as was described in the 1973 Current Housing Report, series H-150-73A, **General Housing Characteristics for the United States and Regions.** Since the 1973-1976 lost units existed, by definition, in the 1973 housing inventory, there was a 1973 housing inventory weight associated with each 1973-1976 lost unit. This weight was used to tabulate the estimates of the characteristics of the 1973-1976 lost units. Also, the general effect of this 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.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of 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 the ratio estimation procedure employed for the 1970 census can be obtained in the 1970 census report, HC(1)-B1, **Detailed Housing Characteristics, United States Summary.**

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-National 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 appears in the 1970 census report, HC(1)-B1, **Detailed Housing Characteristics, United States Summary.** The sampling errors for 1970 census data are much smaller than for the AHS data. Therefore, in making comparisons between the two data sources, it can be safely assumed that the census data are subject to zero sampling errors.

Sampling errors.—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 interviewers were used, estimates from each of the different samples would differ from each other. The average deviation (from all possible samples) of an estimate from the average of the estimate (derived from 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 interviewer 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 errors, measured by the standard error, and biases and some additional nonsampling errors not measured by the standard error.

The procedures, as illustrated, provide a method to construct interval estimates so that a known proportion of the intervals would contain the average of all possible samples. For example, if all possible samples were selected, and each of these samples were surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated for 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;

TABLE I. Standard Errors of Estimated Number of Housing Units: 1976 (excluding estimates of housing units pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Source of Water, Households with Head of Spanish Origin, and Vacant Housing Units)

(68 chances out of 100)

Size of estimate (000)	Standard error: Total or White (000)	Standard error: Black (000)	Size of estimate (000)	Standard error: Total or White (000)	Standard error: Black (000)
0	1	1	1,000	38	36
5	3	3	2,500	60	50
10	4	4	5,000	83	50
25	6	6	10,000	113	–
50	9	9	25,000	158	–
100	12	12	50,000	164	–
250	19	19	75,000	76	–
500	27	26			

TABLE II. Standard Errors of Estimated Number of Housing Units Pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Source of Water, and Households with Head of Spanish Origin: 1976

(68 chances out of 100)

Size of estimate (000)	Standard error: Total, White, or Spanish origin (000)	Standard error: Black (000)	Size of estimate (000)	Standard error: Total, White, or Spanish origin (000)	Standard error: Black (000)
0	2	2	1,000	45	43
5	3	3	2,500	71	59
10	5	5	5,000	99	59
25	7	7	10,000	135	–
50	10	10	25,000	189	–
100	14	14	50,000	196	–
250	23	22	75,000	90	–
500	32	31			

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 either is or is not contained in any particular computed interval. However, for a particular sample one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables are approximations to the standard errors of various estimates shown in this 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 the precise standard error for any specific item.

Tables I through IV present the standard errors applicable to the 1976 housing inventory estimates in this report, and table V presents the standard errors applicable to 1973-1976 lost housing unit estimates in this report. Table I shows the approximate standard errors applicable to all 1976 housing inventory estimates except those pertaining to the specified items in the table.

The standard errors shown in tables II, III, and IV should be used for those specified items. Linear interpolation should be used to determine standard errors for levels of estimates not specifically shown in tables I through V.

The reliability of an estimated percentage, computed by using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more.

TABLE III. Standard Errors of Estimated Number of Year-Round Vacant Housing Units: 1976

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	1	250	15
5	2	500	21
10	3	1,000	28
25	5	2,500	34
50	7	3,500	31
100	10		

TABLE IV. Standard Errors of Estimated Number of Seasonal and Migratory Vacant Housing Units: 1976

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	19	100	42
5	19	250	64
10	19	500	82
25	22	1,000	87
50	30		

UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, D.C. 20233

Montfort

January 15, 1979

MEMORANDUM FOR Arthur F. Young, Chief
Housing Division

From: Charles D. Jones, Chief
Statistical Methods Division

Subject: AHS-National: Correction of the Year IV Report Standard Error Tables for Estimates of Seasonal and Migratory Vacant Housing Units

This memorandum brings to your attention corrections to the standard error tables for estimates of seasonal and migratory vacant housing units in the appendix B's for Year IV reports A and E. The standard error tables which were presented in these appendices were too conservative (i.e., too large), especially for small estimates. The following are the two corrected tables for the 1976 reports, corresponding to tables IV and IX in report A, and IV and X in report E.

Table I. Standard Errors of Estimated Number of Seasonal and Migratory Vacant Housing Units: 1976

(68 chances out of 100)

Size of Estimate (000)	Standard Error (000)	Size of Estimate (000)	Standard Error (000)
0	2	100	16
5	3	250	26
10	5	500	41
25	8	1,000	66
50	11	1,500	90

Table II. Standard Errors of Estimated Percentages of Seasonal and Migratory Vacant Housing Units: 1976

(68 chances out of 100)

Base of Percentage (000)	Estimated Percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	30.7	30.7	30.7	30.7	30.7	30.7	30.7	33.3
10	18.1	18.1	18.1	18.1	18.1	18.1	20.4	23.5
25	8.1	8.1	8.1	8.1	8.9	10.6	12.9	14.9
50	4.2	4.2	4.2	4.6	6.3	7.5	9.1	10.5
100	2.2	2.2	2.2	3.2	4.5	5.3	6.4	7.4
250	0.9	0.9	1.3	2.1	2.8	3.4	4.1	4.7
500	0.4	0.7	0.9	1.5	2.0	2.4	2.9	3.3
1,000	0.2	0.5	0.7	1.0	1.4	1.7	2.0	2.4
1,500	0.15	0.4	0.5	0.8	1.2	1.4	1.7	1.9

cc: L. Norry (HOUS)
E. Montfort "
J. Maynard "
P. Harple "
E. Knowles (DSD)
G. Shapiro (SMD)
D. Schwanz "
D. Luery "
M. Malmuth "

SMD/ALin/tn

TABLE V. Standard Errors of Estimated Number of Lost Housing Units: 1973-1976

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	100 . . .	14
5	3	250 . . .	22
10	4	500 . . .	32
25	7	1,000 . .	47
50	10	2,500 . .	82

Tables VI through IX present the standard errors of estimated percentages. Table VI shows the approximate standard errors of all estimated percentages of housing units except those pertaining to the specified items in tables II, III, and IV. The standard errors shown in tables VII, VIII, and IX should be used for those specified items. Table VII should also be used for estimated percentages of 1973-1976 lost housing units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables VI through IX.

Included in tables I through IX are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as overestimates of the true standard errors and should be used primarily for construction of confidence intervals for characteristics when an estimate of zero is obtained.

For ratios of the form (100) (x/y), where x is not a subclass of y, tables VI through IX 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:

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

where: 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

TABLE VI. Standard Errors of Estimated Percentages of Housing Units: 1976 (excludes estimated percentages of housing units pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Source of Water, Households with Head of Spanish Origin, and Vacant Housing Units)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage[1]							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	22.7	22.7	22.7	22.7	22.7	22.7	23.4	27.1
10	12.8	12.8	12.8	12.8	12.8	13.7	16.6	19.1
25	5.5	5.5	5.5	5.5	7.3	8.6	10.5	12.1
50	2.8	2.8	2.8	3.7	5.1	6.1	7.4	8.6
100	1.4	1.4	1.7	2.6	3.6	4.3	5.2	6.1
250	0.6	0.8	1.1	1.7	2.3	2.7	3.3	3.8
500	0.3	0.5	0.8	1.2	1.6	1.9	2.3	2.7
1,000	0.15	0.4	0.5	0.8	1.1	1.4	1.7	1.9
2,500	0.06	0.2	0.3	0.5	0.7	0.9	1.0	1.2
5,000	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.9
10,000 . . .	0.02	0.12	0.2	0.3	0.4	0.4	0.5	0.6
25,000 . . .	0.01	0.08	0.11	0.2	0.2	0.3	0.3	0.4
50,000 . . .	—	0.05	0.08	0.12	0.2	0.2	0.2	0.3
75,000 . . .	—	0.04	0.06	0.10	0.13	0.2	0.2	0.2

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than one-tenth of the percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point.

TABLE VII. Standard Errors of Estimated Percentages of 1976 Housing Units Pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Source of Water, Households with Head of Spanish Origin, and 1973-1976 Lost Housing Units

(68 chances out of 100)

Base of percentage (000)	Estimated percentage[1]							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	29.5	29.5	29.5	29.5	29.5	29.5	29.5	32.3
10	17.3	17.3	17.3	17.3	17.3	17.3	19.8	22.9
25	7.7	7.7	7.7	7.7	8.7	10.3	12.5	14.5
50	4.0	4.0	4.0	4.5	6.1	7.3	8.9	10.2
100	2.0	2.0	2.0	3.2	4.3	5.2	6.3	7.2
250	0.8	0.9	1.3	2.0	2.7	3.3	4.0	4.6
500	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
1,000	0.2	0.5	0.6	1.0	1.4	1.6	2.0	2.3
2,500	0.08	0.3	0.4	0.6	0.9	1.0	1.3	1.4
5,000	0.04	0.2	0.3	0.4	0.6	0.7	0.9	1.0
10,000 . . .	0.02	0.14	0.2	0.3	0.4	0.5	0.6	0.7
25,000 . . .	0.01	0.09	0.13	0.2	0.3	0.3	0.4	0.5
50,000 . . .	—	0.06	0.09	0.14	0.2	0.2	0.3	0.3
75,000 . . .	—	0.05	0.07	0.12	0.2	0.2	0.2	0.3

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than one-tenth of the percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point.

Illustration of the use of the standard error tables. Illustration I.—Table A-1 of this report shows that, inside SMSA's in the United States, there were 9,407,000 owner-occupied housing units with two persons in 1976. Interpolation in table I shows that the standard error of an estimate of this size is approximately 109,000. The following procedure was used in interpolating:

The information presented in the table below was extracted from table I. The entry for "x" is the one sought.

Size of estimate (000)	Standard error (000)
5,000	83
9,407	x
10,000	113

By vertically interpolating between 83 and 113, the entry for "x" is determined to be

$$9{,}407-5{,}000 = 4{,}407$$
$$10{,}000-5{,}000 = 5{,}000$$

$$83 + \frac{4{,}407}{5{,}000}(113-83) = 109$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 9,298,000 to 9,516,000 housing units. Therefore, a conclusion that the average estimate of 1976 housing units of this type lies 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 9,233,000 to 9,581,000 housing units with 90-percent confidence; and that the average estimate lies within the interval from 9,189,000 to 9,625,000 housing units with 95-percent confidence.

Table A-1 also shows that of the 9,407,000 owner-occupied housing units with two persons inside SMSA's, 3,601,000, or 38.3 percent, were in central cities. Interpolation in table VI (i.e., interpolation on both the base and percent) shows that the standard error of the above percentage is 0.6 percentage points. The following procedure was used in interpolating:

TABLE VIII. Standard Errors of Estimated Percentages of Year-Round Vacant Housing Units: 1976

(68 chances out of 100)

Base of percentage (000)	Estimated percentage[1]							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	16.1	16.1	16.1	16.1	16.1	16.1	19.0	21.9
10	8.7	8.7	8.7	8.7	9.3	11.1	13.4	15.5
25	3.7	3.7	3.7	4.3	5.9	7.0	8.5	9.8
50	1.9	1.9	1.9	3.0	4.2	4.9	6.0	6.9
100	0.9	1.0	1.4	2.1	2.9	3.5	4.2	4.9
250	0.4	0.6	0.9	1.3	1.9	2.2	2.7	3.1
500	0.2	0.4	0.6	1.0	1.3	1.6	1.9	2.2
1,000	0.10	0.3	0.4	0.7	0.9	1.1	1.3	1.5
2,500	0.04	0.2	0.3	0.4	0.6	0.7	0.8	1.0
3,500	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.8

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than one-tenth of the percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point.

TABLE IX. Standard Errors of Estimated Percentages of Seasonal and Migratory Vacant Housing Units: 1976

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	79.3	79.3	79.3	79.3	79.3	79.3	84.7	97.8
10	65.7	65.7	65.7	65.7	65.7	65.7	65.7	69.2
25	43.4	43.4	43.4	43.4	43.4	43.4	43.4	43.7
50	27.7	27.7	27.7	27.7	27.7	27.7	27.7	30.9
100	16.1	16.1	16.1	16.1	16.1	16.1	18.9	21.9
250	7.1	7.1	7.1	7.1	8.3	9.9	12.0	13.8
500	3.7	3.7	3.7	4.3	5.9	7.0	8.5	9.8
1,000	1.9	1.9	1.9	3.0	4.2	5.0	6.0	6.9

The information presented in the table below was extracted from table VI. The entry for "p" is the one sought.

Base of percentage (000)	Estimated percentage		
	25 or 75	38.3	50
5,000 . . .	0.7	a	0.9
9,407 . . .		p	
10,000 . .	0.5	b	0.6

1. By horizontal interpolation between 0.7 and 0.9, the entry for cell "a" is determined to be 0.81.

$$38.3-25.0 = 13.3$$
$$50.0-25.0 = 25.0$$

$$0.7 + \frac{13.3}{25.0}(0.9-0.7) = 0.81$$

2. By horizontal interpolation between 0.5 and 0.6, the entry for cell "b" is determined to be 0.55.

$$38.3-25.0 = 13.3$$
$$50.0-25.0 = 25.0$$

$$0.5 + \frac{13.3}{25.0}(0.6-0.5) = 0.55$$

3. By vertical interpolation between 0.81 and 0.55, the entry for "p" is determined to be 0.6.

$$9{,}407-5{,}000 = 4{,}407$$
$$10{,}000-5{,}000 = 5{,}000$$

$$0.81-\frac{4{,}407}{5{,}000}(0.81-0.55) = 0.6$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 37.7 to 38.9 percent; the 90-percent confidence interval is from 37.3 to 39.3 percent; and the 95-percent confidence interval is from 37.1 to 39.5 percent.

Illustration II.—Table A-1 shows that in the United States in 1976 there were 143,000 housing units in structures with four floors or more (see "Elevator in Structure" item) that were outside of SMSA's. Interpolation in table I shows that the standard error of an estimate of this size is approximately 14,000. Consequently, the 68-percent confidence interval is from 129,000 to 157,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 housing units in structures with four floors or more that were outside of SMSA's lies 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 121,000 to 165,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 115,000 to 171,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 143,000, 1976 housing units in structures with four floors or more that were outside SMSA's, 97,000, or 67.8 percent, were in structures that contained elevators. Interpolation in table VI (i.e., interpolation on both the base and the percent) shows that the standard error of the above percentage is 4.9 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 62.9 to 72.7 percent; the 90-percent confidence interval is from 60.0 to 75.6 percent; and the 95-percent confidence interval is from 58.0 to 77.6 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 errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different areas or the difference between separate and uncorrelated characteristics in the same area. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error. However, if there is a high negative correlation between the two characteristics, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table A-1 shows that inside SMSA's in the United States there were 5,727,000 owner-occupied housing units with three persons in 1976. Thus, the apparent difference between the number of 1976 owner-occupied housing units with two persons and those with three persons is 3,680,000. The standard error of 9,407,000 is approximately 109,000 as is shown above. Table I shows that the standard error on an estimate of 5,727,000 to be approximately 87,000. Therefore, the standard error of the estimated difference of 3,680,000 is about 139,000.

$$139{,}000 = \sqrt{(109{,}000)^2 + (87{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 3,680,000 difference is from 3,541,000 to 3,819,000 housing units. Therefore, a conclusion that the average estimate of this difference, derived from all possible samples, lies 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 3,458,000 to 3,902,000 housing units, and the 95-percent confidence interval is from 3,402,000 to 3,958,000. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied housing units inside SMSA's with two persons is greater than the number with three persons.

Medians.—For the 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 the tables, 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 shows the median number of persons in owner-occupied housing units inside SMSA's was 2.9 in 1976. The base of the distribution, from which this median was determined, is 30,895,000 housing units.

1. From table VI, the standard error of a 50-percent characteristic on the base of 30,895,000 is 0.4 percentage points.
2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 49.2 and 50.8.
3. From table A-1, it can be seen by cumulating the frequencies for the first two categories that 13,211,000 owner-occupied housing units, or 42.8 percent, had one and two persons (actually, 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 5,727,000 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:

$$2.5 + (3.5-2.5)\ \frac{49.2-42.8}{18.5} = 2.8$$

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

$$2.5 + (3.5-2.5)\ \frac{50.8-42.8}{18.5} = 2.9$$

Thus, the 95-percent confidence interval ranges from 2.8 to 2.9 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 this median (see the paragraph on rounding error in the nonsampling error section of this appendix).

Nonsampling errors.—In general, 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 this 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 1976 AHS National.

1970 census.—A number of 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 on coverage and content errors, as well as the methodology employed, can be found in the 1970 Census of Population and Housing Evaluation and Research Program series reports PHC(E)-5, **The Coverage of Housing in the 1970 Census;** and PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

Reinterview program.—For the AHS National 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 subsample 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 "content" error of these AHS estimates.

As part of the reinterview, an additional check was carried out for interviewer evaluation and quality control. This check was made at each of these households to determine if the following was done during the original interview.

1. The correct unit was visisted.
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.

The results of the reinterview study are presented in the following Census Bureau memorandum: "Reinterview Results for the Annual Housing Survey-National Sample, 1976." Some of these results are:

1. "Overall, the reinterview results showed moderate to high levels of inconsistency in response between the original interview and reinterview."
2. "One-third of the nonattitudinal items shown in this report showed high levels of inconsistency between the original interview and reinterview."
3. "For the attitudinal items which had enough data to compute reliable measures of response error, over one-third showed high levels of inconsistency, with the remaining items falling into the moderate range."

The range for evaluating inconsistency is from 0-100. The rule of thumb is that indices below 20 are low; indices from 20 to 50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

However, unlike past years, the respondent's answers in the reinterview were not reconciled to the original an-

swers given in the AHS interview; i.e., after the question is answered in the reinterview, the interviewer did not present the previous responses and then ask the respondent to decide upon the best answer. In the past, providing the reinterviewer with the original response had the effect of reducing the levels of inconsistency substantially.

The 1970 census reinterview results provide illustrations of possible nonsampling errors for some of the items which also appear in the AHS. For example, median value of homes was consistently underestimated by about 5 percent, and the average monthly costs of electricity and utility gas were consistently overestimated although the net effect on average gross rent was fairly small.

A possible explanation for the results of the AHS and census reinterview studies, as well as the surveys themselves, is that the data are based on the answers given by the respondents, who may lack precise information. Also, because the results of the reinterview studies are derived from sample surveys, 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 respect to errors of coverage and estimation for missing data, it was mentioned previously (in the section on estimation) that the AHS new construction sample had deficiencies with regard to the representation of conventional new construction. During the sampling of building permits, only those issued more than 5 months before the survey began were eligible to be selected to represent conventional new construction. Due to time constraints, it is not possible to sample units whose permits are issued less than 5 months in advance of the survey.

It is estimated that the 1976 AHS sample missed about 3 percent (i.e., about 300,000 units) of all conventional new construction (i.e., all conventional housing units built after April 1970) because the permits for these units, which were built before October 1976, were issued less than 5 months in advance of the survey.

In addition, the 1976 Coverage Improvement Program also had certain deficiencies. First, when the canvassing was done to identify mobile home parks that were not in the sample frame or not on the commercial lists, only 92 percent of the census address frame ED's were represented. Second, it appears that the listing procedure (used to find mobile homes placed outside parks, units converted from nonresidential to residential, and houses that had been moved onto their present site) was not very efficient for finding nonresidential conversions (which might be primarily in business districts), since the listing procedure started from a residential unit. (The sample estimate of this component was approximately 16,000 housing units with a standard error of 12,000.)

Finally, it is felt that deficiencies also exist in ED's where area sampling methods are used. As before, it had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent (i.e., as much as 400,000 units) of all housing units in ED's where area sampling methods are used because these units are not listed during the canvassing.

The third stage of ratio estimation corrects for these deficiencies as far as the count of total housing is concerned; i.e., it adjusts the estimate of the total housing inventory to the best available estimate. However, biases of subtotals would still remain.

Rounding errors.—With respect to errors associated with processing, 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 this survey.

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Selection of sample areas. App-42
Designation of sample housing units for the 1976 survey App-42
Selection of the 1973 sample housing units App-42
Splitting of the sample App-43
Selection of supplemental sample housing units in rural areas App-43
Selection of sample housing units for the 1976 Coverage Improvement Program App-43
ESTIMATION App-44
RELIABILITY OF THE ESTIMATES App-45
Sampling errors App-45
Illustration of the use of the standard error tables. Illustration I App-47
Illustration II App-48
Differences App-48
Illustration of the computation of the standard error of a difference App-48
Nonsampling errors App-49
Reinterview program App-49
Coverage errors App-49
Rounding errors App-50

SAMPLE DESIGN

The 1976 estimates are based on data collected in October through December 1976 for the 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. The sample for this survey was spread over 461 sample areas (called primary sampling units), comprising 923 counties and independent cities with coverage in each of the 50 States and the District of Columbia.

Approximately 75,500 sample housing units (both occupied and vacant) were eligible for interview in the 1976 Annual Housing Survey. Of this number, 4,500 interviews were classified as "noninterview" for various reasons. Occupied housing units were classified as "noninterview" mainly because the occupants were not found at home after repeated calls. For vacant housing units, interviews were not obtained because an informed respondent was not found after repeated visits. In addition to the 75,500, there were also 6,600 sample units which were visited but found not to be eligible for interview for the AHS in terms of collecting information relevant to the 1976 housing inventory.

Selection of sample areas.—The United States was divided into areas made up of counties and independent cities referred to as primary sampling units (PSU's). These PSU's were then grouped into 376 strata, 156 of which consisted of only one PSU which was in sample with certainty. These 156 strata were mostly the larger SMSA's and were called self-representing (SR) since the sample from the sample area represented just that PSU. Each one of the other 220 strata consisted of a group of PSU's and were referred to as non-self-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSU's in the stratum as well.

One PSU was selected from each NSR stratum with probability proportionate to the 1970 census population of the PSU. (This resulted in 220 NSR sample PSU's.) In addition, the NSR strata were grouped into 110 pairs, and one stratum was picked at random from each pair. From this stratum, an additional PSU was selected independently of the other PSU selected from this stratum. Since the two PSU's were independently selected, it was possible for the same PSU to be selected twice. This occurred in 25 instances, producing an additional 85 NSR sample PSU's, thus giving a grand total of 461 PSU's.

Designation of sample housing units for the 1976 survey.—The sample housing units designated to be interviewed in the 1976 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1975 survey.
2. All sample housing units that were either type A noninterviews; i.e., units eligible to be interviewed, or type B noninterviews; i.e., units not eligible for interview at the time of the survey but which could become eligible in the future, in the 1975 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1976 AHS questionnaire, page 1.)
3. All sample housing units that were selected from the list of building permits issued since the 1975 survey. (This sample represented the housing units built in permit-issuing areas since the 1975 survey).
4. All sample housing units that were selected as part of the 1976 Coverage Improvement Program. (This sample represented most of the housing units which, until 1976, did not have a chance of selection.)

Selection of the 1973 sample housing units.—The overall sampling rate used to select the sample for the 1973 AHS was about 1 in 1,366. The within-PSU sampling rate for the AHS was determined so that the overall probability of selection for each sample housing unit was the same; e.g., if the probability of selecting a NSR PSU was 1 in 10, then the within-PSU sampling rate would be 1 in 136.6.

Within the sample PSU's, a sample of the housing units enumerated in the 1970 Census of Population and Housing was selected for the AHS. In addition, a sample of new construction building permits was also selected to represent the units constructed since the 1970 census. These samples were selected at about twice the rate mentioned previously; i.e., at 2 in 1,366, thereby producing a sample twice as large as needed. This sample was split into two equal-sized samples—one to be used for the AHS, and one to be held in reserve for possible future use for the AHS. The procedure used to split this sample into equal-sized samples is described in the next section.

The sample of 1970 census units was selected in several stages of sampling. Within the sample PSU's, the first step was the selection of a sample of census enumeration districts (ED's), administrative units used in the 1970 census. The probability of selection for an ED was

proportionate to the 1970 census counts of housing units (HU's) and persons in group quarters in the formula:

$$\frac{\text{Number of HU's in the ED} + \dfrac{\text{Number of group quarters persons in the ED}}{3}}{4}$$

The next step was to select an expected cluster of about four neighboring housing units within each sample ED. For most of the ED's, the selection was accomplished using the list of addresses for the ED as compiled in the 1970 census. However, in those ED's where addresses were incomplete or inadequate (mostly rural areas), the selection process was accomplished using area sampling methods. These ED's were 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) and a segment was selected. Those selected segments with an expected size which was a multiple of four were further subsampled at the time of the survey so that an expected four housing units were chosen for interview.

The sample of new construction units was selected from building permits issued since January 1970. Within each sample PSU, the building permits were chronologically ordered by month issued, and compact clusters of approximately four housing units were created. These clusters were then sampled for inclusion in the AHS at the sampling rate of 2 in 1,366. Housing units constructed since the 1970 census in areas which do not issue building permits were brought into the sample as a result of the area sample described above.

Splitting of the sample.—The described sample selection procedure produced clusters (or segments) of size-four housing units for the sample taken from the census address frame, the new construction frame, and the area sampling frame (mainly rural areas). Clusters of this size should result in a minimum loss in precision for estimates of housing characteristics in rural areas because of the heterogeneity of neighboring units. However, clusters of size-two housing units were considered to be more optimum within those areas where the housing characteristics of neighboring units tend to be very similar (i.e., urban areas and new construction units). A splitting operation was then carried out for clusters selected from the census address and the new construction frames. This consisted of halving each sample cluster from these frames. Thus, two housing units from each of these clusters were included in the survey and two housing units were held in reserve. No splitting operation was carried out within the clusters selected from the area sampling frame; every other area sample cluster of four housing units was used for the survey and the remaining clusters were assigned to the reserve sample.

Selection of supplemental sample housing units in rural areas.—In 1974, it was decided to increase the reliability of the AHS estimates of rural housing characteristics by doubling the number of sample housing units from rural areas. This was accomplished by reactivating the reserve sample, selected in the original sampling operations in 1973, from rural areas only. For the reserve sample selected in census address and new construction frames, this meant that the other half of each rural cluster (an expected two housing units) was reactivated in 1974. Similarly for the area sampling frame, this meant the entire reserve cluster (an expected four housing units) was reactivated in 1974 if the cluster was rural. This supplementation increased the overall probability of selection for sample housing units in rural areas to about 2 in 1,366; whereas, the overall probability of selection for sample housing units in urban areas remained at 1 in 1,366.

Selection of sample housing units for the 1976 Coverage Improvement Program.—The 1976 Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS National sample from the census address and new construction frames. The coverage deficiencies included the following units:

1. New construction, from building permits issued prior to January 1970.
2. Units converted to residential use in structures totally nonresidential at the time of the 1970 census.
3. Houses that have been moved to their present site since the 1970 census.
4. Mobile homes placed in parks either missed in the 1970 census or established since the 1970 census.
5. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

A sample of new construction units whose permits were issued before January 1970 was selected in two stages. First, units whose permits were issued before January 1970, but which were completed after the census, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. These units were then sampled so that the overall probability of selection was about 1 in 1,320.

A sample of mobile homes placed in a park missed by the census or established after the census was also selected in two stages. During the first stage, a list of mobile home parks were obtained from commercial listings. This list was then supplemented by additional parks identified by a canvassing operation similar to that performed in ED's where area sampling methods are used. The second stage consisted of dividing the parks into clusters of an expected size of four sites. Then these clusters were sampled so that the overall probability of selection was about 1 in 1,366.

For the remaining units, (i.e., mobile homes placed outside parks since the 1970 census, mobile homes vacant at the time of the 1970 census, units converted from nonresidential to residential use since the 1970 census, and houses that had been moved onto their present site since the 1970 census), the sampling was done in three stages. First, a subsample of the regular AHS sample units from the census address frame was selected. Second, succeeding structures that had been eligible to be selected from the census address frame were then listed in a defined path of travel to the right from the sample unit until eight structures (excluding mobile home parks) were found. Finally, the intervening structures that had been

listed which did not have a chance of selection in the AHS were identified and the units within these structures were interviewed.

ESTIMATION

In 1976 the AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the type A noninterview housing units encountered in the AHS. This noninterview adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

$$\frac{\text{Interviewed housing units + noninterviewed housing units}}{\text{Interviewed housing units}}$$

The first-stage ratio estimation procedure was employed for sample housing units from non-self-representing (NSR) PSU's only. This procedure was designed to reduce the contribution to the variance arising from the sampling of PSU's. The first-stage ratio estimation procedure takes into account the differences that existed at the time of the 1970 census in the distribution by tenure and residence of the housing population estimated from the sample NSR PSU's and that of the NSR housing population in each of the four census regions of the country. The first-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{The 1970 census housing population in the residence-tenure category for all NSR strata in a census region}}{\text{Estimate of the housing population category using 1970 census housing counts for sample NSR PSU's in a census region}}$$

The numerators of the ratios were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR stratum and summing these counts across the NSR strata in each census region. The denominators were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR sample PSU, weighting these counts by the inverse of the probability of selecting that PSU, and summing these weighted counts across the NSR sample PSU's in each census region. The computed first-stage ratio estimation factor was then applied to the existing weight for each NSR sample unit in each first-stage ratio estimation category.

The second-stage ratio estimation procedure was designed to adjust the AHS sample estimate of one category of conventional new construction units (i.e., one category of sample units built April 1, 1970, or later) to an independently derived current estimate where a known deficiency in the AHS sample exists (see the section on non-sampling error) for each of the four regions. This estimate was considered to be the best estimate available for the number of conventional new construction units in this category. The second-stage ratio estimation factor was as follows:

$$\frac{\text{Current best estimate of new construction in the category}}{\text{AHS sample estimate of new construction units in the category}}$$

The numerators of the ratios were derived from data based on the Survey of Construction (SOC).

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units 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 each second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all the AHS sample units. This procedure was designed to adjust the AHS sample estimates of housing (i.e., the estimates employing the noninterview, first-stage, and second-stage adjustments) to independently derived current housing estimates for four categories of vacant housing units and for 24 categories for occupied housing units. Each of these 24 categories is a combination of the characteristics of residence, tenure, race of head, and sex of head. The third-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{Current independent estimate of housing units in the category}}{\text{AHS sample estimate of housing units in the category}}$$

The numerators of the ratios for occupied housing units were derived from data based on the Current Population Survey (CPS), a sample household survey conducted monthly by the Bureau of the Census. The numerators of the ratios for vacant housing units were derived from data based on the Housing Vacancy Survey (HVS), a quarterly vacancy survey also conducted by the Bureau of the Census.

The denominators of the ratios were obtained from the weighted estimates for the 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 in each third-stage ratio estimation category.

The second- and third-stage ratio estimation procedures were repeated in an iterative process in order to bring the AHS estimates into close agreement with both sets of "independent" estimates. The second stage was modified so that the estimates for all twelve categories of new construction would be identical to the estimates before the third stage. Hence, the repeated second stage had the effect of controlling the AHS sample estimates of new construction units to the "unbiased" sample estimates for eleven categories of new construction units for each of the four regions (i.e., seven categories for conventional new construction units and four for new construction mobile homes) and, as before, of adjusting the AHS sample estimate of one category of conventional new construction units to an independently derived current estimate.

The numerators were either the unbiased weighted estimates for the AHS sample units, using the existing weight after the first-stage ratio estimation procedure (i.e., the estimates employing the noninterview and first-stage adjustments) or the independent estimate derived from data based on the Survey of Construction (SOC).

The denominators of the ratios in this iterative process were obtained from the weighted estimates for the AHS sample units after the previous stage of ratio estimation. The factors resulting from this iterative process were then applied to the existing weight on the appropriate records, and the resulting product was used as the final weight for tabulation.

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. The distribution of the housing population selected for the sample differed somewhat, by chance, from that of the nation as a whole in such basic housing characteristics as tenure, vacancy status, residence, race of head, and sex of head. These characteristics are probably closely correlated with other housing characteristics measured for the AHS. Therefore, through the use of the three-stage ratio estimation procedure, one can expect the sample estimate to be improved substantially.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS National sample.

Sampling errors.—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 interviewers were used, estimates from each of the different samples would differ from each other. The average deviation (from all possible samples) of a sample estimate from the average of the estimate (derived from 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 interviewer 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 errors, measured by the standard error, and biases and some additional nonsampling errors not measured by the standard error.

The procedure, as illustrated, provides a method to construct interval estimates so that a known proportion of the intervals would contain the average of all possible samples. For example, if all possible samples were selected, and each of these samples were surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated for 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 either is or is not contained in any particular computed interval. However, for a particular sample one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables are approximations to the standard errors of various estimates shown in this 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 the precise standard error for any specific item.

Tables I through IV present the standard errors applicable to the 1976 housing inventory estimates in this report. Table I shows the approximate standard errors applicable to all 1976 housing inventory

TABLE I. Standard Errors of Estimated Number of Housing Units: 1976 (Excluding estimates of housing units pertaining to Lacking Complete Plumbing Facilities, Households with Head of Spanish Origin, and Vacant Housing Units)

(68 chances out of 100)

Size of estimate (000)	Standard error		Size of estimate (000)	Standard error	
	Total or White (000)	Black (000)		Total or White (000)	Black (000)
0	1	1	1,000 ...	38	36
5	3	3	2,500 ...	60	50
10	4	4	5,000 ...	83	50
25	6	6	10,000 ..	113	–
50	9	9	25,000 ..	158	–
100	12	12	50,000 ..	164	–
250	19	19	75,000 ..	76	–
500	27	26			

TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to Lacking Complete Plumbing Facilities and Households with Head of Spanish Origin: 1976

(68 chances out of 100)

Size of estimate (000)	Standard error: Total, White, or Spanish origin (000)	Standard error: Black (000)	Size of estimate (000)	Standard error: Total, White, or Spanish origin (000)	Standard error: Black (000)
0	2	2	1,000	45	43
5	3	3	2,500	71	59
10	5	5	5,000	99	59
25	7	7	10,000	135	—
50	10	10	25,000	189	—
100	14	14	50,000	196	—
250	23	22	75,000	90	—
500	32	31			

estimates except those pertaining to the indicated specified items. The standard errors shown in tables II, III, and IV should be used for those specified items.

TABLE III. Standard Errors of Estimated Numbers of Year-Round Vacant Housing Units: 1976

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	1	250	15
5	2	500	21
10	3	1,000	28
25	5	2,500	34
50	7	3,500	31
100	10		

TABLE IV. Standard Errors of Estimated Numbers of Seasonal and Migratory Vacant Housing Units: 1976

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	19	100	42
5	19	250	64
10	19	500	82
25	22	1,000	87
50	30		

Linear interpolation should be used to determine standard errors for levels of estimates not specifically shown in tables I through IV.

The reliability of an estimated percentage, computed by using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more.

Tables V through VIII present the standard errors of estimated percentages. Table V shows the approximate standard errors of all estimated percentages of housing units except those pertaining to the specified items in tables II, III, and IV. The standard errors shown in tables VI, VII, and VIII should be used for those specified items. Two-way linear interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables V through VIII.

Included in tables I through VIII are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as overestimates of the true standard errors and should be used primarily for construction of confidence intervals for characteristics when an estimate of zero is obtained.

For ratios of the form (100) (x/y), where x is not a subclass of y, tables V

TABLE V. Standard Errors of Estimated Percentages of Housing Units: 1976 (Excludes estimated percentages of housing units pertaining to Lacking Complete Plumbing Facilities, Households with Head of Spanish Origin, and Vacant Housing Units)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage[1]: 0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	22.7	22.7	22.7	22.7	22.7	22.7	23.4	27.1
10	12.8	12.8	12.8	12.8	12.8	13.7	16.6	19.1
25	5.5	5.5	5.5	5.5	7.3	8.6	10.5	12.1
50	2.8	2.8	2.8	3.7	5.1	6.1	7.4	8.6
100	1.4	1.4	1.7	2.6	3.6	4.3	5.2	6.1
250	0.6	0.8	1.1	1.7	2.3	2.7	3.3	3.8
500	0.3	0.5	0.8	1.2	1.6	1.9	2.3	2.7
1,000	0.15	0.4	0.5	0.8	1.1	1.4	1.7	1.9
2,500	0.06	0.2	0.3	0.5	0.7	0.9	1.0	1.2
5,000	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.9
10,000	0.02	0.12	0.2	0.3	0.4	0.4	0.5	0.6
25,000	0.01	0.08	0.11	0.2	0.2	0.3	0.3	0.4
50,000	—	0.05	0.08	0.12	0.2	0.2	0.2	0.3
75,000	—	0.04	0.06	0.10	0.13	0.2	0.2	0.2

[1] Standard errors are presented to the nearest one-tenth of 1 percent except when the standard error is less than one-tenth of 1 percent; in those cases, the standard error is shown to the nearest one-hundredth of 1 percent.

TABLE VI. Standard Errors of Estimated Percentages of Housing Units Pertaining to Lacking Complete Plumbing Facilities and Households with Head of Spanish Origin: 1976

(68 chances out of 100)

Base of percentage (000)	Estimated percentage[1]							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	29.5	29.5	29.5	29.5	29.5	29.5	29.5	32.3
10	17.3	17.3	17.3	17.3	17.3	17.3	19.8	22.9
25	7.7	7.7	7.7	7.7	8.7	10.3	12.5	14.5
50	4.0	4.0	4.0	4.5	6.1	7.3	8.9	10.2
100	2.0	2.0	2.0	3.2	4.3	5.2	6.3	7.2
250	0.8	0.9	1.3	2.0	2.7	3.3	4.0	4.6
500	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
1,000	0.2	0.5	0.6	1.0	1.4	1.6	2.0	2.3
2,500	0.08	0.3	0.4	0.6	0.9	1.0	1.3	1.4
5,000	0.04	0.2	0.3	0.4	0.6	0.7	0.9	1.0
10,000 ...	0.02	0.14	0.2	0.3	0.4	0.5	0.6	0.7
25,000 ...	0.01	0.09	0.13	0.2	0.3	0.3	0.4	0.5
50,000 ...	–	0.06	0.09	0.14	0.2	0.2	0.3	0.3
75,000 ...	–	0.05	0.07	0.12	0.2	0.2	0.2	0.3

[1] Standard errors are presented to the nearest one-tenth of 1 percent except when the standard error is less than one-tenth of 1 percent; in those cases, the standard error is shown to the nearest one-hundredth of 1 percent.

through VIII 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:

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

where: 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. Illustration I.—Table A-2 of this report shows that in the United States there were 13,214,000 renter-occupied housing units with common stairways in 1976. Interpolation in table I above shows that the standard error of an estimate of this size is approximately 123,000. The following procedure was used in interpolating:

The information presented in the table below was extracted from table I. The entry for "x" is the one sought.

Size of estimate (000)	Standard error (000)
10,000	113
13,214	x
25,000	158

By vertically interpolating between 113 and 158, the entry for "x" is determined to be:

$$13{,}214 - 10{,}000 = 3{,}214$$
$$25{,}000 - 10000 = 15{,}000$$

$$113 + \frac{3{,}214}{15{,}000}\ (158 - 113) = 123$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 13,091,000 to 13,337,000 housing units. Therefore, a conclusion that the average estimate of 1976 housing units of this type lies 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 13,017,000 to 13,411,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 12,968,000 to 13,460,000 housing units with 95 percent confidence.

Table A-2 also shows that of the 13,214,000 renter-occupied housing units with common stairways, 11,790,000, or 89.2 percent, were located inside SMSA's. Interpolation in table V (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage is 0.4 percentage points. The following procedure was used in interpolating:

The information presented in the table below was extracted from table V. The entry for "p" is the one sought.

Base of percentage (000)	Estimated percentage		
	10 or 90	89.2	15 or 85
10,000 ..	0.4	a	0.4
13,214 ..		p	
25,000 ..	0.2	b	0.3

1. By horizontal interpolation between 0.4 and 0.4, the entry for cell "a" is determined to be 0.4.

$$\cancel{89.2 - 85.0 = 4.2}$$
$$90.0 - 85.0 = 5.0$$
$$0.4 + \frac{\cancel{4.2}\ .8}{5.0}\ (0.4 - 0.4) = 0.40$$

2. By horizontal interpolation between 0.2 and 0.3, the entry for cell "b" is determined to be 0.28.

$$\cancel{89.2 - 85.0 = 4.2}$$
$$90.0 - 85.0 = 5.0$$
$$0.2 + \frac{\cancel{4.2}}{5.0}\ (0.3 - 0.2) = 0.\cancel{28}\ .22$$

3. By vertical interpolation between 0.40 and 0.~~28~~, the entry for "p" is determined to be 0.4.

$$13{,}214 - 10{,}000 = 3{,}214$$
$$25{,}000 - 10{,}000 = 15{,}000$$
$$0.4 - \frac{3{,}214}{15{,}000}\ (0.40 - 0.\cancel{28})\ .22 = 0.4$$

TABLE VII. Standard Errors of Estimated Percentages of Year-Round Vacant Housing Units: 1976

(68 chances out of 100)

Base of percentage (000)	Estimated percentage[1]							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	16.1	16.1	16.1	16.1	16.1	16.1	19.0	21.9
10	8.7	8.7	8.7	8.7	9.3	11.1	13.4	15.5
25	3.7	3.7	3.7	4.3	5.9	7.0	8.5	9.8
50	1.9	1.9	1.9	3.0	4.2	4.9	6.0	6.9
100	0.9	1.0	1.4	2.1	2.9	3.5	4.2	4.9
250	0.4	0.6	0.9	1.3	1.9	2.2	2.7	3.1
500	0.2	0.4	0.6	1.0	1.3	1.6	1.9	2.2
1,000	0.10	0.3	0.4	0.7	0.9	1.1	1.3	1.5
2,500	0.04	0.2	0.3	0.4	0.6	0.7	0.8	1.0
3,500	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.8

[1] Standard errors are presented to the nearest one-tenth of 1 percent except when the standard error is less than one-tenth of 1 percent; in those cases, the standard error is shown to the nearest one-hundredth of 1 percent.

TABLE VIII. Standard Errors of Estimated Percentages of Seasonal and Migratory Vacant Housing Units: 1976

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	79.3	79.3	79.3	79.3	79.3	79.3	84.7	97.8
10	65.7	65.7	65.7	65.7	65.7	65.7	65.7	69.2
25	43.4	43.4	43.4	43.4	43.4	43.4	43.4	43.7
50	27.7	27.7	27.7	27.7	27.7	27.7	27.7	30.9
100	16.1	16.1	16.1	16.1	16.1	16.1	18.9	21.9
250	7.1	7.1	7.1	7.1	8.3	9.9	12.0	13.8
500	3.7	3.7	3.7	4.3	5.9	7.0	8.5	9.8
1,000	1.9	1.9	1.9	3.0	4.2	5.0	6.0	6.9

Consequently, the 68-percent confidence interval, as shown by these data, is from 88.8 to 89.6 percent; the 90-percent confidence interval is from 88.6 to 89.8 percent; and the 95-percent confidence interval is from 88.4 to 90.0 percent.

Illustration II.—Table A-3 of this report shows that in the United States in 1976 there were 434,000 owner-occupied housing units which had sewage disposal breakdowns. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 25,000. Consequently, the 68-percent confidence interval is from 409,000 to 459,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 owner-occupied housing units which had sewage disposal breakdowns lies 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 394,000 to 474,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 384,000 to 484,000 housing units with 95 percent confidence.

Table A-3 also shows that of the 434,000 owner-occupied housing units in 1976 which had sewage disposal breakdowns, 35,000, or 8.1 percent, had breakdowns three times or more. Interpolation in table V (i.e., interpolation on both the base and the percent) of this appendix shows that the standard error of the above percentage is 1.6 percentage points. Consequently, the 68-percent confidence interval, as shown by this data, is from 6.5 to 9.7 percent; the 90-percent confidence interval is from 5.5 to 10.7 percent; and the 95-percent confidence interval is from 4.9 to 11.3 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 errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different areas or the difference between separate and uncorrelated characteristics in the same area. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error. However, if there is a high negative correlation between the two characteristics, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table A-3 of this report shows that in the United States in 1976 there were 339,000 owner-occupied housing units, which had exactly one sewage disposal breakdown. Thus, the apparent difference between the number of 1976 owner-occupied housing units that had breakdowns three times or more and those that had breakdowns just one time, is 304,000. Interpolation in table I shows the standard error on an estimate of 35,000 to be

approximately 7,000 and the standard error on an estimate of 339,000 to be approximately 22,000. Therefore, the standard error of the estimated difference of 304,000 is about 23,000.

$$23{,}000 = \sqrt{(7{,}000)^2 + (22{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 304,000 difference is from 281,000 to 327,000 housing units. Therefore, a conclusion that the average estimate of this difference, derived from all possible samples, lies 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 267,000 to 341,000 housing units, and the 95-percent confidence interval is from 258,000 to 350,000. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied housing units which had three or more sewage disposal breakdowns is different than the number that had exactly one sewage disposal breakdown, since the 95-percent confidence interval of this difference does not include zero or negative values.

Nonsampling errors.—In general, 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.

Reinterview program.—For the AHS National 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 subsample 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 "content" error of these AHS estimates.

As part of the reinterview, an additional check was carried out for interviewer evaluation and quality control. This 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.

The results of the reinterview study are presented in the following Census Bureau memorandum: "Reinterview Results for Annual Housing Survey-National Sample, 1976." Some of these results are:

1. "Overall, the reinterview results showed moderate to high levels of inconsistency in response between the original interview and reinterview."
2. "One-third of the nonattitudinal items shown in this report showed high levels of inconsistency between the original interview and reinterview."
3. "For the attitudinal items which had enough data to compute reliable measures of response error, over one-third showed high levels of inconsistency, with the remaining items falling into the moderate range."

The range for evaluating inconsistency is from 0–100. The rule of thumb is that indices below 20 are low, indices from 20 to 50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

However, unlike past years, the respondent's answers in the reinterview were not reconciled to the original answers given in the AHS interview; i.e., after the question is answered in the reinterview, the interviewer did not present the previous responses and then ask the respondent to decide upon the best answer. In the past, providing the reinterviewer with the original response had the effect of reducing the levels of inconsistency substantially.

The 1970 census reinterview results provide illustrations of possible nonsampling errors for some of the items which also appear in the AHS. For example, median value of homes was consistently underestimated by about five percent, and the average monthly cost of electricity and utility gas were consistently overestimated although the net effect on average gross rent was fairly small.

A possible explanation for the results of the AHS and census reinterview studies, as well as the surveys themselves, is that the data are based on the answers given by the respondents who may lack precise information. Also, because the results of the reinterview studies are derived from sample surveys, 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 respect to errors of coverage and estimation for missing data, it was mentioned previously (in the section on estimation) that the AHS new construction sample had deficiencies with regard to the representation of conventional new construction. During the sampling of building permits, only those issued more than five months before the survey began were eligible to be selected to represent conventional new construction. Due to time constraints, it is not possible to sample units whose permits are issued less than five months in advance of the survey.

It is estimated that the 1976 AHS sample missed about three percent (i.e., about 300,000 units) of all conventional new construction (i.e., all conventional housing units built after April 1970), because the permits for these units, which were built before October 1976, were issued less than five months in advance of the survey.

In addition, the 1976 Coverage Improvement Program also had certain deficiencies. First, when the canvassing was done to identify mobile home parks that were not in the sample frame or not on the commercial lists, only 92 percent of the census address frame ED's were represented. Second, it appears that the listing procedure (used to find mobile homes placed outside parks, units converted from nonresidential to residential, and houses that had been moved onto their present site) was not very efficient for finding nonresidential conversions (which might be primarily in business districts), since the listing procedure started from a residential unit. (The sample estimate of this component was approximately 16,000 housing units with a standard error of 12,000.)

Finally, it is felt that deficiencies also exist in ED's where area sampling methods are used. As before, it had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as two percent (i.e., as much as 400,000 units) of all housing units in ED's where area sampling methods are used because these units are not listed during the canvassing. The third stage of ratio estimation corrects for these deficiencies as far as the count of total housing is concerned; i.e., it adjusts the estimate of the total housing inventory to the best available estimate. However, biases of subtotals would still remain.

Rounding errors.—With respect to errors associated with processing, 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 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 this survey.

☆ U. S. GOVERNMENT PRINTING OFFICE: 1978—261-239/1209

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Selection of sample areas App-42
Designation of sample housing units for the 1976 survey App-42
Selection of the 1973 sample housing units App-42
Splitting of the sample App-43
Selection of supplemental sample housing units in rural areas App-43
Selection of sample housing units for the 1976 Coverage Improvement Program App-43
ESTIMATION App-44
RELIABILITY OF THE ESTIMATES App-45
Sampling errors App-45
Illustration of the use of the standard error tables. Illustration I App-47
Illustration II App-48
Differences App-48
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 for a median . App-48
Nonsampling errors App-49
Reinterview program App-49
Coverage errors App-49
Rounding errors App-50

SAMPLE DESIGN

The 1976 estimates are based on data collected in October through December 1976 for the 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. The sample for this survey was spread over 461 sample areas (called primary sampling units), comprising 923 counties and independent cities with coverage in each of the 50 States and the District of Columbia.

Approximately 75,500 sample housing units (both occupied and vacant) were eligible for interview in the 1976 Annual Housing Survey. Of this number, 4,500 interviews were classified as "noninterview" for various reasons. Occupied housing units were classified as "noninterview" mainly because the occupants were not found at home after repeated calls. For vacant housing units, interviews were not obtained because an informed respondent was not found after repeated visits. In addition to the 75,500, there were also 6,600 sample units which were visited but found not to be eligible for interview for the AHS in terms of collecting information relevant to the 1976 housing inventory.

Selection of sample areas.—The United States was divided into areas made up of counties and independent cities referred to as primary sampling units (PSU's). These PSU's were then grouped into 376 strata, 156 of which consisted of only one PSU which was in sample with certainty. These 156 strata were mostly the larger SMSA's and were called self-representing (SR) since the sample from the sample area represented just that PSU. Each one of the other 220 strata consisted of a group of PSU's and were referred to as non-self-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSU's in the stratum as well.

One PSU was selected from each NSR stratum with probability proportionate to the 1970 census population of the PSU. (This resulted in 220 NSR sample PSU's.) In addition, the NSR strata were grouped into 110 pairs and one stratum was picked at random from each pair. From this stratum, an additional PSU was selected independently of the other PSU selected from this stratum. Since the two PSU's were independently selected, it was possible for the same PSU to be selected twice. This occurred in 25 instances, producing an additional 85 NSR sample PSU's, thus giving a grand total of 461 PSU's.

Designation of sample housing units for the 1976 survey.—The sample housing units designated to be interviewed in the 1976 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1975 survey.
2. All sample housing units that were either type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of the survey but which could become eligible in the future) in the 1975 survey. (For a list of reasons for type A and type B noninterviews, see facsimile of the 1976 AHS questionnaire, page 1.)
3. All sample housing units that were selected from the list of building permits issued since the 1975 survey. (This sample represented the housing units built in permit-issuing areas, since the 1975 survey.)
4. All sample housing units that were selected as part of the 1976 Coverage Improvement Program. (This sample represented most of the housing units which, until 1976, did not have a chance of selection.)

Selection of the 1973 sample housing units.—The overall sampling rate used to select the sample for the 1973 AHS was about 1 in 1,366. The within-PSU sampling rate for the AHS was determined so that the overall probability of selection for each sample housing unit was the same (e.g., if the probability of selecting a NSR PSU was 1 in 10, then the within-PSU sampling rate would be 1 in 136.6).

Within the sample PSU's, a sample of the housing units enumerated in the 1970 Census of Population and Housing was selected for the AHS. In addition, a sample of new construction building permits was also selected to represent the units constructed since the 1970 census. These samples were selected at about twice the rate mentioned previously (i.e., at 2 in 1,366), thereby producing a sample twice as large as needed. This sample was split into two equal-sized samples—one to be used for the AHS, and one to be held in reserve for possible future use for the AHS. The procedure used to split this sample into equal-sized samples is described in the next section.

The sample of 1970 census units was selected in several stages of sampling. Within the sample PSU's, the first step was the selection of a sample of census enumeration districts (ED's), administra-

tive units used in the 1970 census. The probability of selection for an ED was proportionate to the following 1970 census counts of housing units (HU's) and persons in group quarters, combined in the following formula:

$$\frac{\text{Number of HU's in the ED} + \dfrac{\text{Number of group quarters persons in the ED}}{3}}{4}$$

The next step was to select an expected cluster of about four neighboring housing units within each sample ED. For most of the ED's, the selection was accomplished using the list of addresses for the ED as compiled in the 1970 census. However, in those ED's where addresses were incomplete or inadequate (mostly rural areas), the selection process was accomplished using area sampling methods. These ED's were 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) and a segment was selected. Those selected segments with an expected size which was a multiple of four were further subsampled at the time of the survey so that an expected four housing units were chosen for interview.

The sample of new construction units was selected from building permits issued since January 1970. Within each sample PSU, the building permits were chronologically ordered by month issued, and compact clusters of approximately four housing units were created. These clusters were then sampled for inclusion in the AHS at the sampling rate of 2 in 1,366. Housing units constructed since the 1970 census in areas which do not issue building permits were brought into the sample as a result of the area sample described above.

Splitting of the sample.—The described sample selection procedure produced clusters (or segments) of size-four housing units for the sample taken from the census address frame, the new construction frame, and the area sampling frame (mainly rural areas). Clusters of this size should result in a minimum loss in precision for estimates of housing characteristics in rural areas because of the heterogeneity of neighboring units. However, clusters of size-two housing units were considered to be more optimum within those areas where the housing characteristics of neighboring units tend to be very similar (i.e., urban areas and new construction units). A splitting operation was then carried out for clusters selected from the census address and the new construction frames. This consisted of halving each sample cluster from these frames. Thus, two housing units from each of these clusters were included in the survey and two housing units were held in reserve. No splitting operation was carried out within the clusters selected from the area sampling frame; every other area sample cluster of four housing units was used for the survey and the remaining clusters were assigned to the reserve sample.

Selection of supplemental sample housing units in rural areas.—In 1974, it was decided to increase the reliability of the AHS estimates of rural housing characteristics by doubling the number of sample housing units from rural areas. This was accomplished by reactivating the reserve sample, selected in the original sampling operations in 1973, from rural areas only. For the reserve sample selected in census address and new construction frames, this meant that the other half of each rural cluster (an expected two housing units) was reactivated in 1974. Similarly for the area sampling frame, this meant the entire reserve cluster (an expected four housing units) was reactivated in 1974 if the cluster was rural. This supplementation increased the overall probability of selection for sample housing units in rural areas to about 2 in 1,366; whereas, the overall probability of selection for sample housing units in urban areas remained at 1 in 1,366.

Selection of sample housing units for the 1976 Coverage Improvement Program.—The 1976 Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS National sample from the census address and new construction frames. The coverage deficiencies included the following units:

1. New construction, from building permits issued prior to January 1970.
2. Units converted to residential use in structures totally nonresidential at the time of the 1970 census.
3. Houses that have been moved onto their present site since the 1970 census.
4. Mobile homes placed in parks either missed in the 1970 census or established since the 1970 census.
5. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

A sample of new construction units whose permits were issued before January 1970 was selected in two stages. First, units whose permits were issued before January 1970, but which were completed after the census, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. These units were then sampled so that the overall probability of selection was about 1 in 1,320.

A sample of mobile homes placed in a park missed by the census or established after the census was also selected in two stages. During the first stage, a list of mobile home parks was obtained from commercial listings. This list was then supplemented by additional parks identified by a canvassing operation similar to that performed in ED's where area sampling methods are used. The second stage consisted of dividing the parks into clusters of an expected size of four sites. These clusters were then sampled so that the overall probability of selection was about 1 in 1,366.

For the remaining units (i.e., mobile homes placed outside parks since the 1970 census, mobile homes vacant at the time of the 1970 census, units converted from nonresidential to residential use since the 1970 census, and houses that had been moved onto their present site since the 1970 census), the sampling was done in three stages. First, a subsample of the regular AHS sample units from the census address frame was selected. Second, succeeding structures that had been eligible to be selected from the census address frame were then listed in a defined path of travel to the right from

the sample unit until eight structures (excluding mobile home parks) were found. Finally, the intervening structures that had been listed, which did not have a chance of selection in the AHS, were identified and the units within these structures were interviewed.

ESTIMATION

In 1976, the AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the type A noninterview housing units encountered in the AHS. This noninterview adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

$$\frac{\text{Interviewed housing units} + \text{Noninterviewed housing units}}{\text{Interviewed housing units}}$$

The first-stage ratio estimation procedure was employed for sample housing units from non-self-representing (NSR) PSU's only. This procedure was designed to reduce the contribution to the variance arising from the sampling of PSU's. The first-stage ratio estimation procedure takes into account the differences that existed at the time of the 1970 census in the distribution by tenure and residence of the housing population estimated from the sample NSR PSU's and that of the NSR housing population in each of the four census regions of the country. The first-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{The 1970 census housing population in the residence-tenure category for all NSR strata in a census region}}{\text{Estimate of the housing population category using 1970 census housing counts for sample NSR PSU's in a census region}}$$

The numerators of the ratios were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR stratum and summing these counts across the NSR strata in each census region. The denominators were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR sample PSU, weighting these counts by the inverse of the probability of selecting that PSU and summing these weighted counts across the NSR sample PSU's in each census region. The computed first-stage ratio estimation factor was then applied to the existing weight for each NSR sample unit in each first-stage ratio estimation category.

The second-stage ratio estimation procedure was designed to adjust the AHS sample estimate of one category of conventional new construction units; i.e., one category of sample units built April 1, 1970, or later to an independently derived current estimate where a known deficiency in the AHS sample exists (see the section on nonsampling error) for each of the four regions. This estimate was considered to be the best estimate available for the number of conventional new construction units in this category. The second-stage ratio estimation factor was as follows:

$$\frac{\text{Current best estimate of new construction in the category}}{\text{AHS sample estimate of new construction units in the category}}$$

The numerators of the ratios were derived from data based on the Survey of Construction (SOC). The denominators of the ratios were obtained from the weighted estimates for the AHS sample units as before, 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 each second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all the AHS sample units. This procedure was designed to adjust the AHS sample estimates of housing (i.e., the estimates employing the noninterview, first-stage, and second-stage adjustments) to independently derived current housing estimates for four categories of vacant housing units and for 24 categories of occupied housing units. Each of these 24 categories is a combination of the characteristics of residence, tenure, race of head, and sex of head. The third-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{Current independent estimate of housing units in the category}}{\text{AHS sample estimate of housing units in the category}}$$

The numerators of the ratios for occupied housing units were derived from data based on the Current Population Survey (CPS), a sample household survey conducted monthly by the Bureau of the Census. The numerators of the ratios for vacant housing units were derived from data based on the Housing Vacancy Survey (HVS), a quarterly vacancy survey also conducted by the Bureau of the Census.

The denominators of the ratios were obtained from the weighted estimates for the 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 in each third-stage ratio estimation category.

The second-stage and the third-stage ratio estimation procedures were repeated in an iterative process in order to bring the AHS estimates into close agreement with both sets of "independent" estimates. The second stage was modified so that the estimates for all 12 categories of new construction would be identical to the estimates before the third stage. Hence, the repeated second stage had the effect of controlling the AHS sample estimates of new construction units to the "unbiased" sample estimates for 11 categories of new construction units for each of the 4 regions (i.e., 7 categories for conventional new construction units and 4 for new construction mobile homes) and, as before, of adjusting the AHS sample estimate of 1 category of conventional new construction units to an independently derived current estimate.

The numerators were either the unbiased weighted estimates for the AHS sample units, using the existing weight

after the first-stage ratio estimation procedure (i.e., the estimates employing the noninterview and first-stage adjustments) or the independent estimate derived from data based on the Survey of Construction (SOC).

The denominators of the ratios in this iterative process were obtained from the weighted estimates for the AHS sample units after the previous stage of ratio estimation. The factors resulting from this iterative process were then applied to the existing weight on the appropriate records, and the resulting product was used as the final weight for tabulation.

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. The distribution of the housing population selected for the sample differed somewhat, by chance, from that of the nation as a whole in such basic housing characteristics as tenure, vacancy status, residence, race of head, and sex of head. These characteristics are probably closely correlated with other housing characteristics measured for the AHS. Therefore, through the use of the three-stage ratio estimation procedure, one can expect the sample estimate to be improved substantially.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS National sample.

Sampling errors.—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 interviewers were used, estimates from each of the different samples would differ from each other. The average deviation (from all possible samples) of a sample estimate from the average of the estimate (derived from 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 interviewer 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 errors, measured by the standard error, and biases and some additional nonsampling errors not measured by the standard error.

The procedure as illustrated provides a method to construct interval estimates so that a known proportion of the intervals would contain the average of all possible samples. For example, if all possible samples were selected and each of these samples was surveyed under essentially the same general conditions and an estimate and its estimated standard error were calculated for 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 either is or is not contained in any particular computed interval. However, for a particular sample one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables are approximations to the standard errors of various estimates shown in this 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 the precise standard error for any specific item.

TABLE I. Standard Errors of Estimated Number of Housing Units: 1976 (Excluding estimates of housing units pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Source of Water, and Households with Head of Spanish Origin)

(68 chances out of 100)

Size of estimate (000)	Standard error: Total or White (000)	Standard error: Black (000)	Size of estimate (000)	Standard error: Total or White (000)	Standard error: Black (000)
0	1	1	1,000	38	36
5	3	3	2,500	60	50
10	4	4	5,000	83	50
25	6	6	10,000	113	–
50	9	9	25,000	158	–
100	12	12	50,000	164	–
250	19	19	75,000	76	–
500	27	26			

TABLE II. Standard Errors of Estimated Number of Housing Units Pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Source of Water, and Households with Head of Spanish Origin: 1976

(68 chances out of 100)

Size of estimate (000)	Standard error: Total, White, or Spanish origin (000)	Standard error: Black (000)	Size of estimate (000)	Standard error: Total, White, or Spanish orign (000)	Standard error: Black (000)
0	2	2	1,000	45	43
5	3	3	2,500	71	59
10	5	5	5,000	99	59
25	7	7	10,000	135	—
50	10	10	25,000	189	—
100	14	14	50,000	196	—
250	23	22	75,000	90	—
500	32	31			

Tables I and II present the standard errors applicable to the 1976 housing inventory estimates in this report. Table I shows the approximate standard errors applicable to all 1976 housing inventory estimates except those pertaining to the indicated specified items. The standard errors shown in table II should be used for those specified items. Linear interpolation should be used to determine standard errors for levels of estimates not specifically shown in tables I and II.

The reliability of an estimated percentage, computed by using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more.

Tables III and IV present the standard errors of estimated percentages. Table III shows the approximate standard errors of all estimated percentages of housing units except those pertaining to the specified items in table IV. The standard errors shown in table IV should be used for those specified items. Two-way linear interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables III and IV.

Included in these tables; i.e., tables I through IV, are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as overestimates of the true standard errors, and should be used primarily for construction of confidence intervals for characteristics when an estimate of zero is obtained.

For ratios of the form (100) (x/y), where x is not a subclass of y, tables III 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:

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

where: 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

TABLE III. Standard Errors of Estimated Percentages of Housing Units: 1976 (Excludes estimated percentages of housing units pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Source of Water, and Households with Head of Spanish Origin)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage[1]: 0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	22.7	22.7	22.7	22.7	22.7	22.7	23.4	27.1
10	12.8	12.8	12.8	12.8	12.8	13.7	16.6	19.1
25	5.5	5.5	5.5	5.5	7.3	8.6	10.5	12.1
50	2.8	2.8	2.8	3.7	5.1	6.1	7.4	8.6
100	1.4	1.4	1.7	2.6	3.6	4.3	5.2	6.1
250	0.6	0.8	1.1	1.7	2.3	2.7	3.3	3.8
500	0.3	0.5	0.8	1.2	1.6	1.9	2.3	2.7
1,000	0.15	0.4	0.5	0.8	1.1	1.4	1.7	1.9
2,500	0.06	0.2	0.3	0.5	0.7	0.9	1.0	1.2
5,000	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.9
10,000	0.02	0.12	0.2	0.3	0.4	0.4	0.5	0.6
25,000	0.01	0.08	0.11	0.2	0.2	0.3	0.3	0.4
50,000	—	0.05	0.08	0.12	0.2	0.2	0.2	0.3
75,000	—	0.04	0.06	0.10	0.13	0.2	0.2	0.2

[1] Standard errors are presented to the nearest one-tenth of 1 percent except when the standard error is less than one-tenth of 1 percent; in those cases, the standard error is shown to the nearest one-hundredth of 1 percent.

TABLE IV. Standard Errors of Estimated Percentages of Housing Units Pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Source of Water, and Households with Head of Spanish Origin: 1976

(68 chances out of 100)

Base of percentage (000)	Estimated percentage[1]							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	29.5	29.5	29.5	29.5	29.5	29.5	29.5	32.3
10	17.3	17.3	17.3	17.3	17.3	17.3	19.8	22.9
25	7.7	7.7	7.7	7.7	8.7	10.3	12.5	14.5
50	4.0	4.0	4.0	4.5	6.1	7.3	8.9	10.2
100.	2.0	2.0	2.0	3.2	4.3	5.2	6.3	7.2
250.	0.8	0.9	1.3	2.0	2.7	3.3	4.0	4.6
500.	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
1,000	0.2	0.5	0.6	1.0	1.4	1.6	2.0	2.3
2,500	0.08	0.3	0.4	0.6	0.9	1.0	1.3	1.4
5,000	0.04	0.2	0.3	0.4	0.6	0.7	0.9	1.0
10,000 . . .	0.02	0.14	0.2	0.3	0.4	0.5	0.6	0.7
25,000 . . .	0.01	0.09	0.13	0.2	0.3	0.3	0.4	0.5
50,000 . . .	–	0.06	0.09	0.14	0.2	0.2	0.3	0.3
75,000 . . .	–	0.05	0.07	0.12	0.2	0.2	0.2	0.3

[1] Standard errors are presented to the nearest one-tenth of 1 percent except when the standard error is less than one-tenth of 1 percent; in those cases, the standard error is shown to the nearest one-hundredth of 1 percent.

Illustration of the use of the standard error tables. Illustration I.—Table A-2 of this report shows that in the United States there were 9,329,000 specified owner-occupied housing units with two bedrooms in 1976. Interpolation in table I shows that the standard error of an estimate of this size is approximately 110,000. The following procedure was used in interpolating:

The information presented in the table below was extracted from table I. The entry for "x" is the one sought.

Size of estimate (000)	Standard error (000)
5,000	83
9,329	X
10,000	113

By vertically interpolating between 83 and 113, the entry for "x" is determined to be:

$$9{,}329-5{,}000 = 4{,}329$$
$$10{,}000-5{,}000 = 5{,}000$$
$$83 + \frac{4{,}329}{5{,}000}(113-83) = 110$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 9,219,000 to 9,439,000 housing units. Therefore, a conclusion that the average estimate of 1976 housing units of this type lies 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 9,153,000 to 9,505,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 9,109,000 to 9,549,000 housing units with 95 percent confidence.

Table A-2 also shows that of the 9,329,000 specified owner-occupied housing units with two bedrooms, 1,463,000, or 15.7 percent, were of value between $15,000 and $19,999. Interpolation in table III (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage is 0.4 percentage points. The following procedure was used in interpolating:

The information presented in the following table was extracted from table III. The entry for "p" is the one sought.

Base of percentage (000)	Estimated percentage		
	15 or 85	15.7	25 or 75
5,000 . . .	0.6	a	0.7
9,329 . . .		p	
10,000 . .	0.4	b	0.5

1. By horizontal interpolation between 0.6 and 0.7, the entry for cell "a" is determined to be 0.61.

$$15.7-15.0 = 0.7$$
$$25.0-15.0 = 10.0$$
$$0.6 + \frac{0.7}{10.0}(0.7-0.6) = 0.61$$

2. By horizontal interpolation between 0.4 and 0.5, the entry for cell "b" is determined to be 0.41.

$$15.7-15.0 = 0.7$$
$$25.0-15.0 = 10.0$$
$$0.4 + \frac{0.7}{10.0}(0.5-0.4) = 0.41$$

3. By vertical interpolation between 0.61 and 0.41, the entry for "p" is determined to be 0.4.

$$9{,}329-5{,}000 = 4{,}329$$
$$10{,}000-5{,}000 = 5{,}000$$
$$0.61-\frac{4{,}329}{5{,}000}(0.61-0.41) = 0.4$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 15.3 to 16.1 percent; the 90-percent confidence interval is from 15.1 to 16.3 percent; and the 95-percent confidence interval is from 14.9 to 16.5 percent.

Illustration II.—Table A-2 of this report shows that in the United States in 1976 there were 53,000 specified owner-occupied housing units whose cooking fuel was wood. Interpolation in table II of this appendix shows that the standard error of an estimate of this size is approximately 10,000. Consequently, the 68-percent confidence interval is from 43,000 to 63,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 specified owner-occupied housing units whose cooking fuel was wood lies 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 37,000 to 69,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 33,000 to 73,000 housing units with 95 percent confidence.

Table A-2 also shows that of the 53,000 specified owner-occupied housing units whose cooking fuel was wood, 26,000, or 49.1 percent, were valued at less than $10,000. Interpolation in table IV (i.e., interpolation on both the base and the percent) of this appendix shows that the standard error of the above percentage is 10.0 percentage points. Consequently, the 68-percent confidence interval, as shown by this data, is from 39.1 to 59.1 percent; the 90-percent confidence interval is from 33.1 to 65.1 percent; and the 95-percent confidence interval is from 29.1 to 69.1 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 errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different areas or the difference between separate and uncorrelated characteristics in the same area. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error. However, if there is a high negative correlation between the two characteristics, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table A-2 of this report shows that in the United States in 1976 there were 1,194,000 specified owner-occupied housing units with two bedrooms valued between $10,000 and $14,999. Thus, the apparent difference between the number of 1976 specified owner-occupied housing units with two bedrooms valued between $10,000 and $14,999 and those valued between $15,000 and $19,999 is 269,000. Interpolation in table I shows the standard error on an estimate of 1,194,000 to be approximately 41,000 and the standard error on an estimate of 1,463,000 to be approximately 45,000. Therefore, the standard error of the estimated difference of 269,000 is about 61,000.

$$61{,}000 = \sqrt{(41{,}000)^2 + (45{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 269,000 difference is from 208,000 to 330,000 housing units. Therefore, a conclusion that the average estimate of this difference, derived from all possible samples, lies 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 171,000 to 367,000 housing units, and the 95-percent confidence interval is from 147,000 to 391,000. Thus, we can conclude with 95 percent confidence that the number of 1976 specified owner-occupied housing units with two bedrooms valued between $15,000 and $19,999 is greater than the number valued between $10,000 and $14,999, since the 95-percent confidence interval of this difference does not include zero or negative values.

Medians.—For the 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 the tables, 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-2 of this report shows the median value of specified owner-occupied housing units with two bedrooms was $24,100 in 1976. The base of the distribution, from which this median was determined is 9,329,000 housing units.

1. Table III shows that the standard error of 50 percent on a base of 9,329,000 is approximately 0.6 percent.
2. To obtain a two-standard-error confidence interval on the estimated median, initally add to and subtract from 50 percent twice the standard error determined in 1. This yields percentage limits of 48.8 and 51.2.
3. From table A-2, it can be seen by cumulating the frequencies for the first three categories that 3,624,000 owner-occupied housing units with two bedrooms, or 38.8 percent, had a value less than $20,000 and an addi-

tional 1,274,000 owner-occupied housing units with two bedrooms, or 13.7 percent, had a value between $20,000 and $24,999.

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

$$\$20{,}000 + (\$5{,}000)\left(\frac{48.8-38.9}{13.7}\right) = \$23{,}600$$

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

$$\$20{,}000 + (\$5{,}000)\left(\frac{51.2-38.9}{13.7}\right) = \$24{,}500$$

Thus, the 95-percent confidence interval ranges from $23,600 to $24,500.

Nonsampling errors.—In general, 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.

Reinterview program.—For the AHS National 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 subsample 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 "content" error of these AHS estimates.

As part of the reinterview an additional check was carried out for interviewer evaluation and quality control. This 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.

The results of the reinterview study are presented in the following Census Bureau memorandum: "Reinterview Results for the Annual Housing Survey-National Sample, 1976." Some of these results are:

1. "Overall, the reinterview results showed moderate to high levels of inconsistency in response between the original interview and reinterview."
2. "One-third of the nonattitudinal items shown in this report showed high levels of inconsistency between the original interview and reinterview."
3. "For the attitudinal items which had enough data to compute reliable measures of response error, over one-third showed high levels of inconsistency, with the remaining items falling into the moderate range."

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low, indices from 20 to 50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

However, unlike past years, the respondent's answers in the reinterview were not reconciled to the original answers given in the AHS interviews; i.e., after the question is answered in the reinterview, the interviewer did not present the previous responses and then ask the respondent to decide upon the best answer. In the past, providing the reinterviewer with the original response had the effect of reducing the levels of inconsistency substantially.

The 1970 census reinterview results provide illustrations of possible nonsampling errors for some of the items which also appear in the AHS. For example, median value of homes was consistently underestimated by about five percent, and the average monthly costs of electricity and utility gas were consistently overestimated although the net effect on average gross rent was fairly small.

A possible explanation for the results of the AHS and census reinterview studies, as well as the surveys themselves, is that the data are based on the answers given by the respondents, who may lack precise information. Also, because the results of the reinterview studies are derived from sample surveys, 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 respect to errors of coverage and estimation for missing data, it was mentioned previously (in the section on estimation) that the AHS new construction sample had deficiencies with regard to the representation of conventional new construction. During the sampling of building permits, only those issued more than five months before the survey began were eligible to be selected to represent conventional new construction. Due to time constraints, it is not possible to sample units whose permits are issued less than five months in advance of the survey.

It is estimated that the 1976 AHS sample missed about three percent (i.e., about 300,000 units) of all conventional new construction (i.e., all conventional housing units built after April 1970) because the permits for these units, which were built before October 1976, were issued less than five months in advance of the survey.

In addition, the 1976 Coverage Improvement Program also had certain deficiencies. First, when the canvassing was done to identify mobile home parks that were not in the sample frame or not on the commercial lists, only 92 percent of the census address frame ED's were represented. Second, it appears that the listing procedure (used to find mobile homes placed outside parks, units converted from nonresidential to residential, and houses that had been moved onto their present site) was not very efficient for finding nonresidential conversions (which might be primarily in business districts), since the listing procedure started from a residential unit. (The sample estimate of this component was approximately 16,000 housing units with a standard error of 12,000.)

Finally, it is felt that deficiencies also exist in ED's where area sampling methods are used. As before, it had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent (i.e., as much as 400,000 units) of all housing units in ED's where area sampling methods are used because these units are not listed during the canvassing.

The third stage of ratio estimation corrects for these deficiencies as far as the count of total housing is concerned; i.e., as stated above, it adjusts the estimate of the total housing inventory to the best available estimate. However, biases of subtotals would still remain.

Rounding errors.—With respect to errors associated with processing, 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 this survey.

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Selection of sample areas App-42
Designation of sample housing units for the 1976 survey App-42
Selection of the 1973 sample housing units App-42
Splitting of the sample App-43
Selection of supplemental sample housing units in rural areas App-43
Selection of sample housing units for the 1976 Coverage Improvement Program App-43
ESTIMATION App-44
RELIABILITY OF THE ESTIMATES App-45
Sampling errors App-45
Illustrations of the use of the standard error tables. Illustration I App-46
Illustration II App-47
Differences App-48
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 for a median App-48
Nonsampling errors App-48
Reinterview program App-49
Coverage errors App-49
Rounding errors App-50

SAMPLE DESIGN

The 1976 estimates are based on data collected in October through December 1976 for the 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. The sample for this survey was spread over 461 sample areas (called primary sampling units), comprising 923 counties and independent cities with coverage in each of the 50 States and the District of Columbia.

Approximately 75,500 sample housing units (both occupied and vacant) were eligible for interview in the 1976 Annual Housing Survey. Of this number, 4,500 interviews were classified as "noninterview" for various reasons. For occupied housing units, the main reason was that the occupants were not found at home after repeated calls. For vacant housing units, interviews were not obtained because an informed respondent was not found after repeated visits. In addition to the 75,500, there were also 6,600 sample units which were visited but found not to be eligible for interview for the AHS in terms of collecting information relevant to the 1976 housing inventory.

Selection of sample areas.—The United States was divided into areas made up of counties and independent cities referred to as primary sampling units (PSU's). These PSU's were then grouped into 376 strata, 156 of which consisted of only one PSU which was in sample with certainty. These 156 strata were mostly the larger SMSA's and were called self-representing (SR) since the sample from the sample area represented just that PSU. Each one of the other 220 strata consisted of a group of PSU's and were referred to as non-self-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSU's in the stratum as well.

One PSU was selected from each NSR stratum with probability proportionate to the 1970 census population of the PSU. (This resulted in 220 NSR sample PSU's.) In addition, the NSR strata were grouped into 110 pairs and one stratum was picked at random from each pair. From this stratum, an additional PSU was selected independently of the other PSU selected from this stratum. Since the two PSU's were independently selected, it was possible for the same PSU to be selected twice. This occurred in 25 instances, producing an additional 85 NSR sample PSU's, thus giving a grand total of 461 PSU's.

Designation of sample housing units for the 1976 survey.—The sample housing units designated to be interviewed in the 1976 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1975 survey.
2. All sample housing units that were either type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of the survey but which could become eligible in the future) in the 1975 survey. (For a list of reasons for type A and type B noninterviews, see facsimile of 1976 AHS questionnaire, page 1.)
3. All sample housing units that were selected from the list of building permits issued since the 1975 survey. (This sample represented the housing units built in permit-issuing areas since the 1975 survey).
4. All sample housing units that were selected as part of the 1976 Coverage Improvement Program. (This sample represented most of the housing units which, until 1976, did not have a chance of selection.)

Selection of the 1973 sample housing units.—The overall sampling rate used to select the sample for the 1973 AHS was about 1 in 1,366. The within-PSU sampling rate for the AHS was determined so that the overall probability of selection for each sample housing unit was the same (e.g., if the probability of selecting a NSR PSU was 1 in 10, then the within-PSU sampling rate would be 1 in 136.6).

Within the sample PSU's, a sample of the housing units enumerated in the 1970 Census of Population and Housing was selected for the AHS. In addition, a sample of new construction building permits was also selected to represent the units constructed since the 1970 census. These samples were selected at about twice the rate mentioned previously (i.e., at 2 in 1,366), thereby producing a sample twice as large as needed. This sample was split into two equal-sized samples—one to be used for the AHS, and one to be held in reserve for possible future use for the AHS. The procedure used to split this sample into equal-sized samples is described in the next section.

The sample of 1970 census units was selected in several stages of sampling. Within the sample PSU's, the first step was the selection of a sample of census enumeration districts (ED's), adminis-

trative units used in the 1970 census. The probability of selection for an ED was proportionate to the 1970 census counts of housing units (HU's) and persons in group quarters in the formula:

$$\frac{\text{Number of HU'S in the ED} + \dfrac{\text{Number of group quarters persons in the ED}}{3}}{4}$$

The next step was to select an expected cluster of about four neighboring housing units within each sample ED. For most of the ED's, the selection was accomplished using the list of addresses for the ED as compiled in the 1970 census. However, in those ED's where addresses were incomplete or inadequate (mostly rural areas), the selection process was accomplished using area sampling methods. These ED's were 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) and a segment was selected. Those selected segments with an expected size which was a multiple of four were further subsampled at the time of the survey so that an expected four housing units were chosen for interview.

The sample of new construction units was selected from building permits issued since January 1970. Within each sample PSU, the building permits were chronologically ordered by month issued, and compact clusters of approximately four housing units were created. These clusters were then sampled for inclusion in the AHS at the sampling rate of 2 in 1,366. Housing units constructed since the 1970 census in areas which do not issue building permits were brought into the sample as a result of the area sample described above.

Splitting of the sample.—The described sample selection procedure produced clusters (or segments) of size-four housing units for the sample taken from the census address frame, the new construction frame, and the area sampling frame (mainly rural areas). Clusters of this size should result in a minimum loss in precision for estimates of housing characteristics in rural areas because of the heterogeneity of neighboring units. However, clusters of size-two housing units were considered to be more optimum within those areas where the housing characteristics of neighboring units tend to be very similar (i.e., urban areas and new construction units). A splitting operation was then carried out for clusters selected from the census address and the new construction frames. This consisted of halving each sample cluster from these frames. Thus, two housing units from each of these clusters were included in the survey and two housing units were held in reserve. No splitting operation was carried out within the clusters selected from the area sampling frame; every other area sample cluster of four housing units was used for the survey and the remaining clusters were assigned to the reserve sample.

Selection of supplemental sample housing units in rural areas.—In 1974, it was decided to increase the reliability of the AHS estimates of rural housing characteristics, by doubling the number of sample housing units from rural areas. This was accomplished by reactivating the reserve sample, selected in the original sampling operations in 1973, from rural areas only. For the reserve sample selected in census address and new construction frames, this meant that the other half of each rural cluster (an expected two housing units) was reactivated in 1974. Similarly for the area sampling frame, this meant the entire reserve cluster (an expected four housing units) was reactivated in 1974 if the cluster was rural. This supplementation increased the overall probability of selection for sample housing units in rural areas to about 2 in 1,366; whereas, the overall probability of selection for sample housing units in urban areas remained at 1 in 1,366.

Selection of sample housing units for the 1976 Coverage Improvement Program.—The 1976 Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS National sample from the census address and new construction frames. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970.
2. Units converted to residential use in structures totally nonresidential at the time of the 1970 census.
3. Houses that have been moved to their present site since the 1970 census.
4. Mobile homes placed in parks either missed in the 1970 census or established since the 1970 census.
5. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

A sample of new construction units whose permits were issued before January 1970 was selected in two stages. First, units whose permits were issued before January 1970, but which were completed after the census, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. These units were then sampled so that the overall probability of selection was about 1 in 1,320.

A sample of mobile homes placed in a park missed by the census or established after the census was also selected in two stages. During the first stage, a list of mobile home parks was obtained from commercial listings. This list was then supplemented by additional parks identified by a canvassing operation similar to that performed in ED's where area sampling methods are used. The second stage consisted of dividing the parks into clusters of an expected size of four sites. Then these clusters were sampled so that the overall probability of selection was about 1 in 1,366.

For the remaining units, (i.e., mobile homes placed outside parks since the 1970 census, mobile homes vacant at the time of the 1970 census, units converted from nonresidential to residential use since the 1970 census, and houses that had been moved to their present site since the 1970 census), the sampling was done in three stages. First, a subsample of the regular AHS sample units from the census address frame was selected. Second, succeeding structures that had been eligible to be selected from the census address frame were then listed in a defined path of travel to the right from the sample unit

until eight structures (excluding mobile home parks) were found. Finally, the intervening structures that had been listed which did not have a chance of selection in the AHS were identified and the units within these structures were interviewed.

ESTIMATION

In 1976 the AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the type A noninterview housing units encountered in the AHS. This noninterview adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

$$\frac{\text{Interviewed housing units + Noninterviewed housing units}}{\text{Interviewed housing units}}$$

The first-stage ratio estimation procedure was employed for sample housing units from non-self-representing (NSR) PSU's only. This procedure was designed to reduce the contribution to the variance arising from the sampling of PSU's. The first-stage ratio estimation procedure takes into account the differences that existed at the time of the 1970 census in the distribution by tenure and residence of the housing population estimated from the sample NSR PSU's and that of the NSR housing population in each of the four census regions of the country. The first-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{The 1970 census housing population in the residence-tenure category for all NSR strata in a census region}}{\text{Estimate of the housing population category using 1970 census housing counts for sample NSR PSU's in a census region}}$$

The numerators of the ratios were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR stratum and summing these counts across the NSR strata in each census region. The denominators were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR sample PSU, weighting these counts by the inverse of the probability of selecting that PSU and summing these weighted counts across the NSR sample PSU's in each census region. The computed first-stage ratio estimation factor was then applied to the existing weight for each NSR sample unit in each first-stage ratio estimation category.

The second-stage ratio estimation procedure was designed to adjust the AHS sample estimate of one category of conventional new construction units; i.e., one category of sample units built April 1, 1970, or later to an independently derived current estimate where a known deficiency in the AHS sample exists (see the section on nonsampling error) for each of the four regions. This estimate was considered to be the best estimate available for the number of conventional new construction units in this category. The second-stage ratio estimation factor was as follows:

$$\frac{\text{Current best estimate of new construction in the category}}{\text{AHS sample estimate of new construction units in the category}}$$

The numerators of the ratios were derived from data based on the Survey of Construction (SOC). The denominators of the ratios were obtained from the weighted estimates for the AHS sample units 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 each second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all the AHS sample units. This procedure was designed to adjust the AHS sample estimates of housing (i.e., the estimates employing the noninterview, first-stage, and second-stage adjustments) to independently derived current housing estimates for 4 categories of vacant housing units and for 24 categories of occupied housing units. Each of these 24 categories is a combination of the characteristics of residence, tenure, race of head, and sex of head. The third-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{Current independent estimate of housing units in the category}}{\text{AHS sample estimate of housing units in the category}}$$

The numerators of the ratios for occupied housing units were derived from data based on the Current Population Survey (CPS), a sample household survey conducted monthly by the Bureau of the Census. The numerators of the ratios for vacant housing units were derived from data based on the Housing Vacancy Survey (HVS), a quarterly vacancy survey also conducted by the Bureau of the Census.

The denominators of the ratios were obtained from the weighted estimates for the 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 in each third-stage ratio estimation category.

The second-stage and the third-stage ratio estimation procedures were repeated in an iterative process in order to bring the AHS estimates into close agreement with both sets of "independent" estimates. The second stage was modified so that the estimates for all twelve categories of new construction would be identical to the estimates before the third stage. Hence, the repeated second stage had the effect of controlling the AHS sample estimates of new construction units to the "unbiased" sample estimates for eleven categories of new construction units for each of the four regions (i.e., seven categories for conventional new construction units and four for new construction mobile homes) and, as before, of adjusting the AHS sample estimate of one category of conventional new construction units to an independently derived current estimate.

The numerators were either the unbiased weighted estimates for the AHS sample units, using the existing weight after the first-stage ratio estimation procedure (i.e., the estimates employing the noninterview and first-stage adjustments) or the independent estimate derived from data based on the Survey of Construction (SOC).

The denominators of the ratios in this iterative process were obtained from the weighted estimates for the AHS sample units after the previous stage of ratio estimation. The factors resulting from this iterative process were then applied to the existing weight on the appropriate records, and the resulting product was used as the final weight for tabulation.

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. The distribution of the housing population selected for the sample differed somewhat, by chance, from that of the nation as a whole in such basic housing characteristics as tenure, vacancy status, residence, race of head, and sex of head. These characteristics are probably closely correlated with other housing characteristics measured for the AHS. Therefore, through the use of the three-stage ratio estimation procedure one can expect the sample estimate to be improved substantially.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS National sample.

Sampling errors.—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 interviewers were used, estimates from each of the different samples would differ from each other. The average deviation (from all possible samples) of a sample estimate from the average of the estimate (derived from 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 interviewer 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 errors, measured by the standard error, and biases and some additional nonsampling errors not measured by the standard error.

The procedure as illustrated provides a method to construct interval estimates so that a known proportion of the intervals would contain the average of all possible samples. For example, if all possible samples were selected, and each of these samples were surveyed under essentially the same general conditions and an estimate and its estimated standard error were calculated for 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 either is or is not contained in any particular computed interval. However, for a particular sample one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables are approximations to the standard errors of various estimates shown in this 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 the precise standard error for any specific item.

Tables I and II present the standard errors applicable to the 1976 housing inventory estimates in this report. Table I shows the approximate standard errors applicable to all 1976 housing inventory estimates except those pertaining to the indicated specified items. The standard errors shown in table II should be used for those specified items. Linear inter-

TABLE I. Standard Errors of Estimated Number of Housing Units: 1976 (Excluding estimates of housing units pertaining to Lacking Complete Plumbing Facilities, Mobile Homes, and Households with Head of Spanish Origin)

(68 chances out of 100)

Size of estimate (000)	Standard error		Size of estimate (000)	Standard error	
	Total or White (000)	Black (000)		Total or White (000)	Black (000)
0	1	1	1,000	38	36
5	3	3	2,500	60	50
10	4	4	5,000	83	50
25	6	6	10,000	113	—
50	9	9	25,000	158	—
100	12	12	50,000	164	—
250	19	19	75,000	76	—
500	27	26			

TABLE II. Standard Errors of Estimated Number of Housing Units Pertaining to Lacking Complete Plumbing Facilities, Mobile Homes, and Households with Head of Spanish Origin: 1976

(68 chances out of 100)

Size of estimate	Standard error		Size of estimate	Standard error	
	Total, White or Spanish origin (000)	Black (000)		Total, White or Spanish origin (000)	Black (000)
0	2	2	1,000	45	43
5	3	3	2,500	71	59
10	5	5	5,000	99	59
25	7	7	10,000	135	–
50	10	10	25,000	189	–
100	14	14	50,000	196	–
250	23	22	75,000	90	–
500	32	31			

polation should be used to determine standard errors for levels of estimates not specifically shown in table I.

The reliability of an estimated percentage, computed by using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more.

Tables III and IV present the standard errors of estimated percentages. Table III shows the approximate standard errors of all estimated percentages of housing units except those pertaining to the specified items in table IV. The standard errors shown in table IV should be used for those specified items. Two-way linear interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables III and IV.

Included in these tables; i.e., tables I through IV, are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as overestimates of the true standard errors, and should be used primarily for construction of confidence intervals for characteristics when an estimate of zero is obtained.

For ratios of the form (100) (x/y), where x is not a subclass of y, tables III 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:

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

where: 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. Illustration I.—Table A-1 of this report shows that in the United States there were 4,665,000 owner-occupied housing units occupied by recent movers in 1976. Interpolation in table I shows that the standard error of an estimate of this size is approximately 80,000. The following procedure was used in interpolating:

TABLE III. Standard Errors of Estimated Percentages of Housing Units: 1976 (Excludes estimated percentages of housing units pertaining to Lacking Complete Plumbing Facilities, Mobile Homes, and Households with Head of Spanish Origin)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage[1]							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	22.7	22.7	22.7	22.7	22.7	22.7	23.4	27.1
10	12.8	12.8	12.8	12.8	12.8	13.7	16.6	19.1
25	5.5	5.5	5.5	5.5	7.3	8.6	10.5	12.1
50	2.8	2.8	2.8	3.7	5.1	6.1	7.4	8.6
100	1.4	1.4	1.7	2.6	3.6	4.3	5.2	6.1
250	0.6	0.8	1.1	1.7	2.3	2.7	3.3	3.8
500	0.3	0.5	0.8	1.2	1.6	1.9	2.3	2.7
1,000	0.15	0.4	0.5	0.8	1.1	1.4	1.7	1.9
2,500	0.06	0.2	0.3	0.5	0.7	0.9	1.0	1.2
5,000	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.9
10,000	0.02	0.12	0.2	0.3	0.4	0.4	0.5	0.6
25,000	0.01	0.08	0.11	0.2	0.2	0.3	0.3	0.4
50,000	–	0.05	0.08	0.12	0.2	0.2	0.2	0.3
75,000	–	0.04	0.06	0.10	0.13	0.2	0.2	0.2

[1] Standard errors are presented to the nearest one-tenth of 1 percent except when the standard error is less than one-tenth of 1 percent; in those cases, the standard error is shown to the nearest one-hundredth of 1 percent.

TABLE IV. Standard Errors of Estimated Percentages of Housing Units Pertaining to Lacking Complete Plumbing Facilities, Mobile Homes, and Households with Head of Spanish Origin: 1976

(68 chances out of 100)

Base of percentage (000)	Estimated percentage[1]							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	29.5	29.5	29.5	29.5	29.5	29.5	29.5	32.3
10	17.3	17.3	17.3	17.3	17.3	17.3	19.8	22.9
25	7.7	7.7	7.7	7.7	8.7	10.3	12.5	14.5
50	4.0	4.0	4.0	4.5	6.1	7.3	8.9	10.2
100.	2.0	2.0	2.0	3.2	4.3	5.2	6.3	7.2
250.	0.8	0.9	1.3	2.0	2.7	3.3	4.0	4.6
500.	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
1,000	0.2	0.5	0.6	1.0	1.4	1.6	2.0	2.3
2,500	0.08	0.3	0.4	0.6	0.9	1.0	1.3	1.4
5,000	0.04	0.2	0.3	0.4	0.6	0.7	0.9	1.0
10,000 . . .	0.02	0.14	0.2	0.3	0.4	0.5	0.6	0.7
25,000 . . .	0.01	0.09	0.13	0.2	0.3	0.3	0.4	0.5
50,000 . . .	–	0.06	0.09	0.14	0.2	0.2	0.3	0.3
75,000 . . .	–	0.05	0.07	0.12	0.2	0.2	0.2	0.3

[1]Standard errors are presented to the nearest one-tenth of 1 percent except when the standard error is less than one-tenth of 1 percent; in those cases, the standard error is shown to the nearest one-hundredth of 1 percent.

The information presented in the table below was extracted from table I. The entry for "x" is the one sought.

Size of estimate (000)	Standard error (000)
2,500	60
4,665	x
5,000	83

By vertically interpolating between 60 and 83 the entry for "x" is determined to be:

$$4{,}665-2{,}500 = 2{,}165$$
$$5{,}000-2{,}500 = 2{,}500$$
$$60 + \frac{2{,}165}{2{,}500}(83-60) = 80$$

Consequently, the 68-percent confidence interval, as shown by this data, is from 4,585,000 to 4,745,000 housing units. Therefore, a conclusion that the average estimate of 1976 housing units of this type lies 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 4,537,000 to 4,793,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 4,505,000 to 4,825,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 4,665,000 owner-occupied housing units occupied by recent movers in 1976, 300,000, or 6.4 percent, had six persons or more. Interpolation in table III (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage is 0.4 percentage points. The following procedure was used in interpolating:

The information presented in the table below was extracted from table III. The entry for "p" is the one sought.

Base of percentage (000)	Estimated percentage		
	5 or 95	6.4	10 or 90
2,500 . . .	0.5	a	0.7
4,665 . . .		p	
5,000 . . .	0.4	b	0.5

1. By horizontal interpolation between 0.5 and 0.7, the entry for cell "a" is determined to be 0.56.

$$6.4-5.0 = 1.4$$
$$10.0-5.0 = 5.0$$
$$0.5 + \frac{1.4}{5.0}(0.7-0.5) = 0.56$$

2. By horizontal interpolation between 0.4 and 0.5, the entry for cell "b" is determined to be 0.43.

$$6.4-5.0 = 1.4$$
$$10.0-5.0 = 5.0$$
$$0.4 + \frac{1.4}{5.0}(0.5-0.4) = 0.43$$

3. By vertical interpolation between 0.56 and 0.43, the entry for "p" is determined to be 0.4.

$$4{,}665-2{,}500 = 2{,}165$$
$$5{,}000-2{,}500 = 2{,}500$$
$$0.56-\frac{2{,}165}{2{,}500}(0.56-0.43) = 0.4$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 6.0 to 6.8 percent; the 90-percent confidence interval is from 5.8 to 7.0 percent; and the 95-percent confidence interval is from 5.6 to 7.2 percent.

Illustration II.—Table A-19 of this report shows that in the United States in 1976 there were 155,000 owner-occupied housing units occupied by Spanish recent movers. Interpolation in table II of this appendix shows that the standard error of an estimate of this size is approximately 17,000. Consequently, the 68-percent confidence interval is from 138,000 to 172,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 owner-occupied housing units which were occupied by Spanish recent movers lies 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 128,000 to 182,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 121,000 to 189,000 housing units with 95 percent confidence.

Table A-19 also shows that of the 155,000 owner-occupied housing units occupied by Spanish recent movers, 44,000, or 28.4 percent, had two bedrooms. Interpolation in table IV (i.e., interpolation on both the base and the percent) of this appendix shows that the standard error of the above percentage is 5.6 percentage points. Consequently, the 68-percent confidence interval, as shown by this data, is from 22.8 to 34.0 percent; the 90-percent confidence interval is from 19.4 to 37.4 percent; and the 95-percent confidence interval is from 17.2 to 39.6 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 errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different areas or the difference between separate and uncorrelated characteristics in the same area. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error. However, if there is a high negative correlation between the two characteristics, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table A-1 of this report shows that in the United States there were 498,000 owner-occupied housing units occupied by recent movers with five persons in 1976. Thus, the apparent difference between the number of owner-occupied housing units occupied by recent movers with five persons and the number with six persons or more is 198,000. Interpolation in table I shows that the standard error on an estimate of 498,000 to be approximately 27,000 and the standard error on an estimate of 300,000 to be approximately 21,000. Therefore, the standard error of the estimated difference of 198,000 is about 34,000.

$$34{,}000 = \sqrt{(27{,}000)^2 + (21{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 198,000 difference is from 164,000 to 232,000 housing units. Therefore, a conclusion that the average estimate of this difference, derived from all possible samples, lies 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 144,000 to 252,000 housing units, and the 95-percent confidence interval is from 130,000 to 266,000. Thus, we can conclude with 95 percent confidence that the number of owner-occupied units occupied by recent movers with five persons is different than the number of owner-occupied units occupied by recent movers with six persons or more, since the 95-percent confidence interval of this difference does not include zero or negative values.

Medians.—For the 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 the tables, 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 this report shows that the median number of persons in owner-occupied housing units occupied by recent movers in the United States was 3.0 in 1976. The base of the distribution from which this median was determined is 4,665,000.

1. Table III shows that the standard error of 50 percent on a base of 4,665,000 is 0.9 percent.
2. To obtain a two-standard-error confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in 1. This yields percentage limits of 48.2 and 51.8.
3. From table A-1, it can be seen by cumulating the frequencies for the first two categories that 1,857,000 owner-occupied housing units occupied by recent movers, or 39.8 percent, had one or two persons (actually, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 1,000,000 housing units, or 21.4 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

$$2.5 + (3.5-2.5)\left(\frac{48.2-39.8}{21.4}\right) = 2.9$$

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

$$2.5 + (3.5-2.5)\left(\frac{51.8-39.8}{21.4}\right) = 3.1$$

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

Nonsampling errors.—In general, 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.

Reinterview program.—For the AHS National 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 subsample 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 "content" error of these AHS estimates.

As part of the reinterview, an additional check was carried out for interviewer evaluation and quality control. This 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.

The results of the reinterview study are presented in the following Census Bureau memorandum: "Reinterview Results for the Annual Housing Survey-National Sample, 1976." Some of these results are:

1. "Overall, the reinterview results showed moderate to high levels of inconsistency in response between the original interview and reinterview."
2. "One-third of the nonattitudinal items shown in this report showed high levels of inconsistency between the original interview and reinterview."
3. "For the attitudinal items which had enough data to compute reliable measures of response error, over one-third showed high levels of inconsistency, with the remaining items falling into the moderate range."

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low, indices from 20 to 50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

However, unlike past years, the respondent's answers in the reinterview were not reconciled to the original answers given in the AHS interview; i.e., after the question is answered in the reinterview, the interviewer did not present the previous responses and then ask the respondent to decide upon the best answer. In the past, providing the reinterviewer with the original response had the effect of reducing the levels of inconsistency substantially.

The 1970 census reinterview results provide illustrations of possible nonsampling errors for some of the items which also appear in the AHS. For example, median value of homes was consistently underestimated by about five percent, and the average monthly costs of electricity and utility gas were consistently overestimated although the net effect on average gross rent was fairly small.

A possible explanation for the results of the AHS and census reinterview studies, as well as the surveys themselves, is that the data are based on the answers given by the respondents, who may lack precise information. Also, because the results of the reinterview studies are derived from sample surveys, 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 respect to errors of coverage and estimation for missing data, it was mentioned previously (in the section on estimation) that the AHS new construction sample had deficiencies with regard to the representation of conventional new construction. During the sampling of building permits, only those issued more than five months before the survey began were eligible to be selected to represent conventional new construction. Due to time constraints, it is not possible to sample units whose permits are issued less than five months in advance of the survey.

It is estimated that the 1976 AHS sample missed about 3 percent (i.e., about 300,000 units) of all conventional new construction (i.e., all conventional housing units built after April 1970), because the permits for these units, which were built before October 1976, were issued less than 5 months in advance of the survey.

In addition, the 1976 Coverage Improvement Program also had certain deficiencies. First, when the canvassing was done to identify mobile home parks that were not in the sample frame or not on the commercial lists, only 92 percent of the census address frame ED's were represented. Second, it appears that the listing procedure (used to find mobile homes placed outside parks, units converted from nonresidential to residential, and houses that had been moved onto their present site) was not very efficient for finding nonresidential conversions (which might be primarily in business districts), since the listing procedure started from a residential unit. (The sample estimate of this component was approximately 16,000 housing units with a standard error of 12,000.)

Finally, it is felt that deficiences also exist in ED's where area sampling methods are used. As before, it had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2

percent (i.e., as much as 400,000 units) of all housing units in ED's where area sampling methods are used because these units are not listed during the canvassing.

The third stage of ratio estimation corrects for these deficiences as far as the count of total housing is concerned; i.e., as stated above, it adjusts the estimate of the total housing inventory to the best available estimate. However, biases of subtotals would still remain.

Rounding Errors.—With respect to errors associated with processing, 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. This means that when these figures are derived from relatively large bases 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.

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Selection of sample areas App-42
Designation of sample housing units for the 1976 survey App-42
Selection of the 1973 sample housing units App-42
Splitting of the sample App-43
Selection of supplemental sample housing units in rural areas App-43
Selection of sample housing units for the 1976 Coverage Improvement Program App-43
1970 Census of Population and Housing App-44
ESTIMATION App-44
AHS national sample App-44
1976 housing inventory App-44
1973-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing App-45
RELIABILITY OF THE ESTIMATES App-45
Sampling errors App-45
Illustrations of the use of the standard error tables. Illustration I App-48
Illustration II App-49
Differences App-50
Illustration of the computation of the standard error of a difference App-50
Medians App-50
Illustration of the computation of the 95-percent confidence interval for a median App-51
Nonsampling errors App-51
1970 Census App-51
Reinterview program App-51
Coverage errors App-52
Rounding errors App-52

SAMPLE DESIGN

The 1976 estimates are based on data collected in October through December 1976 for the 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. The sample for this survey was spread over 461 sample areas (called primary sampling units), comprising 923 counties and independent cities with coverage in each of the 50 States and the District of Columbia.

Approximately 75,500 sample housing units (both occupied and vacant) were eligible for interview in the 1976 Annual Housing Survey. Of this number, 4,500 interviews were classified as "noninterview" for various reasons. For occupied housing units, the main reason was that the occupants were not found at home after repeated calls. For vacant housing units, interviews were not obtained because an informed respondent was not found after repeated visits. In addition to the 75,500, there were also 6,600 sample units which were visited but which were found not to be eligible for interview for the AHS in terms of collecting information relevant to the 1976 housing inventory.

Selection of sample areas.—The United States was divided into areas made up of counties and independent cities referred to as primary sampling units (PSU's). These PSU's were then grouped into 376 strata, 156 of which consisted of only one PSU which was in sample with certainty. These 156 strata were mostly the larger SMSA's and were called self-representing (SR) since the sample from the sample area represented just that PSU. Each one of the other 220 strata consisted of a group of PSU's and were referred to as non-self-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSU's in the stratum as well.

One PSU was selected from each NSR stratum with probability proportionate to the 1970 census population of the PSU. (This resulted in 220 NSR sample PSU's.) In addition, the NSR strata were grouped into 110 pairs and one stratum was picked at random from each pair. From this stratum, an additional PSU was selected independently of the other PSU selected from this stratum. Since the two PSU's were independently selected, it was possible for the same PSU to be selected twice. This occurred in 25 instances producing an additional 85 NSR sample PSU's, thus giving a grand total of 461 PSU's.

Designation of sample housing units for the 1976 survey.—The sample housing units designated to be interviewed in the 1976 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1975 survey.
2. All sample housing units that were either type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of the survey but which could become eligible in the future) in the 1975 survey. (For a list of reasons for type A and type B noninterviews, see facsimile of 1976 AHS questionnaire, page 1.)
3. All sample housing units that were selected from the list of building permits issued since the 1975 survey. (This sample represented the housing units built in permit-issuing areas since the 1975 survey.)
4. All sample housing units that were selected as part of the 1976 Coverage Improvement Program. (This sample represented most of the housing units which, until 1976, did not have a chance of selection.)

Selection of the 1973 sample housing units.—The overall sampling rate used to select the sample for the 1973 AHS was about 1 in 1,366. The within-PSU sampling rate for the AHS was determined so that the overall probability of selection for each sample housing unit was the same (e.g., if the probability of selecting a NSR PSU was 1 in 10, then the within-PSU sampling rate would be 1 in 136.6).

Within the sample PSU's, a sample of the housing units enumerated in the 1970 Census of Population and Housing was selected for the AHS. In addition, a sample of new construction building permits was also selected to represent the units constructed since the 1970 census. These samples were selected at about twice the rate mentioned previously (i.e., at 2 in 1,366), thereby producing a sample twice as large as needed. This sample was split into two equal-sized

samples—one to be used for the AHS, and one to be held in reserve for possible future use for the AHS. The procedure used to split this sample into equal-sized samples is described in the next section.

The sample of 1970 census units was selected in several stages of sampling. Within the sample PSU's, the first step was the selection of a sample of census enumeration districts (ED's), administrative units used in the 1970 census. The probability of selection for an ED was proportional to the following 1970 census counts of housing units (HU's) and persons in group quarters, combined in the following formula:

$$\frac{\text{Number of HU's in the ED} + \dfrac{\text{Number of group quarters persons in the ED}}{3}}{4}$$

The next step was to select an expected cluster of about four neighboring housing units within each sample ED. For most of the ED's, the selection was accomplished using the list of addresses for the ED as compiled in the 1970 census. However, in those ED's where addresses were incomplete or inadequate (mostly rural areas), the selection process was accomplished using area sampling methods. These ED's were 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) and a segment was selected. Those selected segments with an expected size which was a multiple of four were further subsampled at the time of the survey so that an expected four housing units were chosen for interview.

The sample of new construction units was selected from building permits issued since January 1970. Within each sample PSU, the building permits were chronologically ordered by month issued, and compact clusters of approximately four housing units were created. These clusters were then sampled for inclusion in the AHS at the sampling rate of 2 in 1,366. Housing units constructed since the 1970 census in areas which do not issue building permits were brought into the sample as a result of the area sample described above.

Splitting of the sample.—The described sample selection procedure produced clusters (or segments) of size-four housing units for the sample taken from the census address frame, the new construction frame, and the area sampling frame (mainly rural areas). Clusters of this size should result in a minimum loss in precision for estimates of housing characteristics in rural areas because of the heterogeneity of neighboring units. However, clusters of size-two, were considered to be more optimum within those areas where the housing characteristics of neighboring units tend to be very similar (i.e., urban areas and new construction units). A splitting operation was then carried out for clusters selected from the census address and the new construction frames. This consisted of halving each sample cluster from these frames. Thus, two housing units from each of these clusters were included in the survey and two housing units were held in reserve. No splitting operation was carried out within the clusters selected from the area sampling frame; every other area sample cluster of four housing units was used for the survey and the remaining clusters were assigned to the reserve sample.

Selection of supplemental sample housing units in rural areas.—In 1974, it was decided to increase the reliability of the AHS estimates of rural housing characteristics by doubling the number of sample housing units from rural areas. This was accomplished by reactivating the reserve sample, selected in the original sampling operations in 1973 from rural areas only. For the reserve sample selected in census address and new construction frames, this meant that the other half of each rural cluster (an expected two housing units) was reactivated in 1974. Similarly for the area sampling frame, this meant the entire reserve cluster (an expected four housing units) was reactivated in 1974 if the cluster was rural. This supplementation increased the overall probability of selection for sample housing units in rural areas to about 2 in 1,366; whereas, the overall probability of selection for sample housing units in urban areas remained at 1 in 1,366.

Selection of sample housing units for the 1976 Coverage Improvement Program.—The 1976 Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS National sample from the census address and new construction frames. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970.
2. Units converted to residential use in structures totally nonresidential at the time of the 1970 census.
3. Houses that have been moved onto their present site since the 1970 census.
4. Mobile homes placed in parks either missed in or established since the 1970 census.
5. Mobile homes placed outside of parks since the 1970 census or vacant at the time of the 1970 census.

A sample of new construction units whose permits were issued before January 1970 was selected in two stages. First, units whose permits were issued before January 1970, but which were completed after the census, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. These units were then sampled so that the overall probability of selection was about 1 in 1,320.

A sample of mobile homes placed in a park missed by the census or established after the census was also selected in two stages. During the first stage, a list of mobile home parks was obtained from commercial listings. This list was then supplemented by additional parks identified by a canvassing operation similar to that performed in ED's where area sampling methods are used. The second stage consisted of dividing the parks into clusters of an expected size of four sites. These clusters were then sampled so that the overall probability of selection was about 1 in 1,366.

For the remaining units, (i.e., mobile homes placed outside parks since the 1970 census, mobile homes vacant at the time of the 1970 census, units converted from nonresidential to residential use since the 1970 census, and houses that

had been moved onto their present site since the 1970 census), the sampling was done in three stages. First, a subsample of the regular AHS sample units from the census address frame was selected. Second, succeeding structures that had been eligible to be selected from the census address frame were then listed until eight structures (excluding mobile home parks) were found. Finally, the intervening structures, that had been listed, which did not have a chance of selection in the AHS were identified and the units within these structures were interviewed.

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 can be obtained in the 1970 census report, HC(1)-B1, **Detailed Housing Characteristics, United States Summary.**

ESTIMATION

AHS National sample.—The AHS National sample produced estimates of two types: Estimates of the 1976 housing inventory and estimates of units removed from the housing inventory between 1973 and 1976 (i.e., 1973-1976 lost units). Each type of estimate employed a separate, though similar, estimation procedure.

1976 housing inventory.—In 1976 the AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the type A noninterview housing units encountered in the AHS. This noninterview adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

$$\frac{\text{Interviewed housing units} + \text{Noninterviewed housing units}}{\text{Interviewed housing units}}$$

The first-stage ratio estimation procedure was employed for sample housing units from non-self-representing (NSR) PSU's only. This procedure was designed to reduce the contribution to the variance arising from the sampling of PSU's. The first-stage ratio estimation procedure takes into account the differences that existed at the time of the 1970 census in the distribution by tenure and residence of the housing population estimated from the sample NSR PSU's and that of the NSR housing population in each of the four census regions of the country. The first-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{The 1970 census housing population in the residence-tenure category for all NSR strata in a census region}}{\text{Estimate of the housing population category using 1970 census housing counts for sample NSR PSU's in a census region}}$$

The numerators of the ratios were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR stratum and summing these counts across the NSR strata in each census region. The denominators were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR sample PSU, weighting these counts by the inverse of the probability of selecting that PSU, and summing these weighted counts across the NSR sample PSU's in each census region. The computed first-stage ratio estimation factor was then applied to the existing weight for each NSR sample unit in each first-stage ratio estimation category.

The second-stage ratio estimation procedure was designed to adjust the AHS sample estimate of one category of conventional new construction units; i.e., one category of sample units built April 1, 1970, or later to an independently derived current estimate where a known deficiency in the AHS sample exists (see the section on nonsampling error) for each of the four regions. This estimate was considered to be the best estimate available for the number of conventional, new construction units in this category. The second-stage ratio estimation factor was as follows:

$$\frac{\text{Current best estimate of new construction in the category}}{\text{AHS sample estimate of new construction units in the category}}$$

The numerators of the ratios were derived from data based on the Survey of Construction (SOC). The denominators of the ratios were obtained from the weighted estimates for the AHS sample units 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 each second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all the AHS sample units. This procedure was designed to adjust the AHS sample estimates of housing (i.e., the estimates employing the noninterview, first-stage, and second-stage adjustments) to independently derived current housing estimates for four categories of vacant housing units and for 24 categories of occupied housing units. Each of these 24 categories is a combination of the characteristics of residence, tenure, race of head, and sex of head. The third-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{Current independent estimate of housing units in the category}}{\text{AHS sample estimate of housing units in the category}}$$

The numerators of the ratios for occupied housing units were derived from data based on the Current Population Survey (CPS), a sample household survey conducted monthly by the Bureau of the Census. The numerators of the ratios for vacant housing units were derived from data based on the Housing Vacancy Survey (HVS), a quarterly vacancy survey also conducted by the Bureau of the Census.

The denominators of the ratios were obtained from the weighted estimates for the 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 in each third-stage ratio estimation category.

The second-stage and the third-stage ratio estimation procedures were repeated in an iterative process in order to bring the AHS estimates into close agreement with both sets of "independent" estimates. The second stage was modified so that the estimates for all twelve categories of new construction would be identical to the estimates before the third stage. Hence, the repeated second stage had the effect of controlling the AHS sample estimates of new construction units to the "unbiased" sample estimates for eleven categories of new construction units for each of the four regions (i.e., seven categories for conventional, new construction units and four for new construction mobile homes) and, as before, of adjusting the AHS sample estimate of one category of conventional, new construction units to an independently derived current estimate.

The numerators were either the unbiased weighted estimates for the AHS sample units, using the existing weight after the first-stage ratio estimation procedure (i.e., the estimates employing the noninterview and first-stage adjustments) or the independent estimate derived from data based on the Survey of Construction (SOC).

The denominators of the ratios in this iterative process were obtained from the weighted estimates for the AHS sample units after the previous stage of ratio estimation. The factors resulting from this iterative process were then applied to the existing weight on the appropriate records, and the resulting product was used as the final weight for tabulation.

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. The distribution of the housing population selected for the sample differed somewhat, by chance, from that of the nation as a whole in such basic housing characteristics as tenure, vacancy status, residence, race of head, and sex of head. These characteristics are probably closely correlated with other housing characteristics measured for the AHS. Therefore, through the use of the three-stage ratio estimation procedure, one can expect the sample estimate to be improved substantially.

1973-1976 lost units.—The 1973-1976 lost unit estimates employed the three-stage ratio estimation procedure used to produce the AHS National estimates of the 1973 housing inventory described in the 1973 Current Housing Report, series H-150-73A, **General Housing Characteristics for the United States and Regions.** Since the 1973-1976 lost units existed, by definition, in the 1973 housing inventory, there was a 1973 housing inventory weight associated with each 1973-1976 lost unit. This weight was used to tabulate the estimates of the characteristics of the 1973-1976 lost units. Also, the general effect of this 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.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of 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 the ratio estimation procedure employed for the 1970 census can be obtained in the 1970 census report, HC(1)-B1, **Detailed Housing Characteristics, United States Summary.**

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS National 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 appears in the 1970 census report, HC(1)-B1, **Detailed Housing characteristics, United States Summary.** The sampling errors for 1970 census data are much smaller than for the AHS data. Therefore, in making comparisons between the two data sources, it can be safely assumed that the census data are subject to zero sampling errors.

Sampling errors.—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 interviewers were used, estimates from each of the different samples would differ from each other. The average deviation (from all possible samples) of an estimate from the average of all estimates (derived from 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 interviewer 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 errors measured by the standard error, and biases and some additional nonsampling errors not measured by the standard error.

The procedure, as illustrated, provides a method to construct interval estimates so that a known proportion of the intervals would contain the average of all possible samples. For example, if all possible samples were selected and each of these samples was surveyed under essentially the same general conditions and an estimate and its estimated stand-

ard error were calculated for 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 intevals 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 either is or is not contained in any particular computed interval. However, for a particular sample, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the following tables are approximations to the standard errors of various estimates shown in this 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 the precise standard error for any specific item.

In this report, tables I through IV present the standard errors applicable to the 1976 housing inventory estimates and table V presents the standard errors applicable to 1973-1976 lost housing unit estimates. Table I shows the approximate standard errors applicable to all 1976 housing inventory estimates except those pertaining to the specified items in the table, those pertaining to urban and rural items, and those pertaining to estimates of vacant units.

The standard errors shown in table II should be used for those specified items in table I pertaining to the total inventory, the standard errors shown in table III should be used for urban and rural items, and the standard errors shown in table IV should be used for estimates pertaining to vacant housing units. Linear interpolation should be used to determine standard errors for levels of estimates not specifically shown in tables I through V.

The reliability of an estimated percentage, computed by using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more.

Tables VI through X present the standard errors of estimated percentages for the 1976 housing inventory and for the 1973-1976 lost housing units. Two-way linear interpolation should be used to determine the standard error for estimated percentages not specifically shown in the tables. An index to these tables,

TABLE I. Standard Errors of Estimated Number of Total Housing Units: 1976 (Excluding estimates of total housing units pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Source of Water, Households with Head of Spanish Origin, Urban Housing Units, Rural Housing Units, and Vacant Housing Units)

(68 chances out of 100)

Size of estimate (000)	Standard error: Total or White (000)	Standard error: Black (000)	Size of estimate (000)	Standard error: Total or White (000)	Standard error: Black (000)
0	1	1	1,000	38	36
5	3	3	2,500	60	50
10	4	4	5,000	83	50
25	6	6	10,000	113	–
50	9	9	25,000	158	–
100	12	12	50,000	164	–
250	19	19	75,000	76	–
500	27	26			

TABLE II. Standard Errors of Estimated Number of Housing Units Pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Source of Water, and Households with Head of Spanish Origin: 1976

(68 chances out of 100)

Size of estimate (000)	Standard error: Total, White or Spanish origin (000)	Standard error: Black (000)	Size of estimate (000)	Standard error: Total, White or Spanish origin (000)	Standard error: Black (000)
0	2	2	1,000	45	43
5	3	3	2,500	71	59
10	5	5	5,000	99	59
25	7	7	10,000	135	–
50	10	10	25,000	189	–
100	14	14	50,000	196	–
250	23	22	75,000	90	–
500	32	31			

TABLE III. Standard Errors of Estimated Number of Urban or Rural Housing Units: 1976

(68 chances out of 100)

Size of estimate (000)	Standard error: Rural housing units (except those in next column)		Rural housing units pertaining to cooking fuel, mobile homes, new construction, source of water, households with head of Spanish origin, and urban housing units (except those in next column)		Urban housing units pertaining to cooking fuel, lacking complete plumbing facilities, mobile homes, new construction, and source of water	
	Total or White (000)	Black (000)	Total, White, or Spanish origin (000)	Black (000)	Total, White, or Spanish origin (000)	Black (000)
0	1	1	2	2	2	2
5	2	2	3	3	3	3
10	3	3	4	4	5	5
25	5	5	7	6	7	7
50	7	7	9	9	11	11
100	10	10	13	13	15	15
250	17	16	21	20	24	23
500	24	23	29	28	33	32
1,000	33	31	41	38	47	44
2,500	52	43	64	53	74	61
5,000	72	–	89	54	103	62
10,000	98	–	122	–	140	–
25,000	137	–	170	–	196	–
50,000	–	–	177	–	203	–

TABLE IV. Standard Errors of Estimated Number of Vacant Housing Units: 1976

(68 chances out of 100)

Size of estimate (000)	Standard error: Total year-round vacants (000)	Urban year-round vacants (000)	Rural year-round vacants (000)	Total, urban or rural seasonal and migratory vacants (000)
0	1	1	1	19
5	2	3	2	19
10	3	4	3	19
25	5	6	5	22
50	7	8	7	30
100	10	11	10	42
250	15	18	16	64
500	21	25	22	82
1,000	28	33	29	87
2,500	34	42	36	–
3,500	31	38	–	–

table XI, can be found on the last page of this appendix.

Included in tables I through X are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as overestimates of the true standard errors, and should be used primarily for construction of confidence intervals for characteristics when an estimate of zero is obtained.

TABLE V. Standard Errors of Estimated Number of 1973-1976 Lost Housing Units

(68 chances out of 100)

Size of estimate (000)	Standard error (000)
0	2
5	3
10	4
25	7
50	10
100	14
250	22
500	32
1,000	47
2,500	82

For ratios of the form (100) (x/y), where x is not a subclass of y, tables VI through X 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:

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

where: 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

TABLE VI. Standard Errors of Estimated Percentages of Total Housing Units (excludes estimated percentages of housing units pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Source of Water, and Households with Head of Spanish Origin), and of Urban Year-Round Vacant Housing Units: 1976

(68 chances out of 100)

Base of percentage (000)	Estimated percentage[1]							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	22.7	22.7	22.7	22.7	22.7	22.7	23.4	27.1
10	12.8	12.8	12.8	12.8	12.8	13.7	16.6	19.1
25	5.5	5.5	5.5	5.5	7.3	8.6	10.5	12.1
50	2.8	2.8	2.8	3.7	5.1	6.1	7.4	8.6
100	1.4	1.4	1.7	2.6	3.6	4.3	5.2	6.1
250	0.6	0.8	1.1	1.7	2.3	2.7	3.3	3.8
500	0.3	0.5	0.8	1.2	1.6	1.9	2.3	2.7
1,000	0.15	0.4	0.5	0.8	1.1	1.4	1.7	1.9
2,500	0.06	0.2	0.3	0.5	0.7	0.9	1.0	1.2
5,000	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.9
10,000	0.02	0.12	0.2	0.3	0.4	0.4	0.5	0.6
25,000	0.01	0.08	0.11	0.2	0.2	0.3	0.3	0.4
50,000	–	0.05	0.08	0.12	0.2	0.2	0.2	0.3
75,000	–	0.04	0.06	0.10	0.13	0.2	0.2	0.2

[1] Standard errors are presented to the nearest one-tenth of 1 percent except when the standard error is less than one-tenth of 1 percent; in those cases, the standard error is shown to the nearest one-hundredth of 1 percent.

TABLE VII. Standard Errors of Estimated Percentages of Total Housing Units Pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Source of Water, and Households with Head of Spanish Origin, and of Urban Housing Units Pertaining to Cooking Fuel, Mobile Homes, Lacking Complete Plumbing Facilities, and New Construction: 1976

(68 chances out of 100)

Base of percentage (000)	Estimated percentage[1]							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	31.0	31.0	31.0	31.0	31.0	31.0	31.0	33.5
10	18.4	18.4	18.4	18.4	18.4	18.4	20.5	23.7
25	8.3	8.3	8.3	8.3	9.0	10.7	13.0	15.0
50	4.3	4.3	4.3	4.6	6.4	7.6	9.2	10.6
100	2.2	2.2	2.2	3.3	4.5	5.4	6.5	7.5
250	0.9	0.9	1.3	2.1	2.8	3.4	4.1	4.7
500	0.5	0.7	0.9	1.5	2.0	2.4	2.9	3.4
1,000	0.2	0.5	0.7	1.0	1.4	1.7	2.1	2.4
2,500	0.10	0.3	0.4	0.7	0.9	1.1	1.3	1.5
5,000	0.05	0.2	0.3	0.5	0.6	0.8	0.9	1.1
10,000	0.02	0.15	0.2	0.3	0.4	0.5	0.6	0.7
25,000	0.01	0.09	0.13	0.2	0.3	0.3	0.4	0.5
50,000	–	0.07	0.09	0.15	0.2	0.2	0.3	0.3
75,000	–	0.05	0.08	0.12	0.2	0.2	0.2	0.3

[1] Standard errors are presented to the nearest one-tenth of 1 percent except when the standard error is less than one-tenth of 1 percent; in those cases, the standard error is shown to the nearest one-hundredth of 1 percent.

Illustrations of the use of the standard error tables. Illustration I.—Table A-1 of this report shows that in urban areas of the United States there were 21,459,000 renter-occupied housing units in 1976. Interpolation in table III shows that the standard error of an estimate of this size is approximately 159,000. The following procedure was used in interpolating:

The information presented in the table below was extracted from table III. The entry for "x" is the one sought.

Size of estimate (000)	Standard error (000)
10,000	122
21,459	x
25,000	170

By vertically interpolating between 122 and 170, the entry for "x" is determined to be:

$$21{,}459 - 10{,}000 = 11{,}459$$
$$25{,}000 - 10{,}000 = 15{,}000$$
$$122 + \frac{11{,}459}{15{,}000}(170 - 122) = 159$$

Consequently, the 68-percent confidence interval, as shown by these data is from 21,300,000 to 21,618,000 housing units. Therefore, a conclusion that the average estimate of 1976 housing units of this type lies 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 21,205,000 to 21,713,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 21,141,000 to 21,777,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 21,459,000 renter-occupied housing units in urban areas, 6,558,000, or 30.6 percent, were occupied by two persons. Interpolation in table IX (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage is 0.5 percentage points. The following procedure was used in interpolating.

TABLE VIII. Standard Errors of Estimated Percentages of Total Year-Round Vacant Housing Units, Rural Housing Units (excludes estimated percentages of rural housing units pertaining to Mobile Homes, New Construction, Source of Water, and Households with Head of Spanish Origin), and Rural Year-Round Vacant Housing Units: 1976

(68 chances out of 100)

Base of percentage (000)	Estimated percentage[1]							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	18.1	18.1	18.1	18.1	18.1	18.1	20.3	23.5
10	9.9	9.9	9.9	9.9	10.0	11.9	14.4	16.6
25	4.2	4.2	4.2	4.6	6.3	7.5	9.1	10.5
50	2.2	2.2	2.2	3.2	4.5	5.3	6.4	7.4
100	1.1	1.1	1.5	2.3	3.1	3.7	4.5	5.2
250	0.4	0.7	0.9	1.4	2.0	2.4	2.9	3.3
500	0.2	0.5	0.7	1.0	1.4	1.7	2.0	2.3
1,000	0.11	0.3	0.5	0.7	1.0	1.2	1.4	1.7
2,500	0.04	0.2	0.3	0.5	0.6	0.7	0.9	1.0
5,000	0.02	0.14	0.2	0.3	0.4	0.5	0.6	0.7
10,000 ...	0.01	0.10	0.15	0.2	0.3	0.4	0.5	0.5
25,000 ...	–	0.07	0.09	0.14	0.2	0.2	0.3	0.3

[1] Standard errors are presented to the nearest one-tenth of 1 percent except when the standard error is less than one-tenth of 1 percent; in those cases, the standard error is shown to the nearest one-hundredth of 1 percent.

TABLE IX. Standard Errors of Estimated Percentages of 1976 Urban Housing Units (excludes estimated percentages of urban housing units pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, and New Construction), of 1976 Rural Housing Units Pertaining to Mobile Homes, New Construction, Source of Water, and Households with Head of Spanish Origin, and of the 1973-1976 Lost Housing Units.

(68 chances out of 100)

Base of percentage (000)	Estimated percentage[1]							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	25.3	25.3	25.3	25.3	25.3	25.3	26.8	31.0
10	14.5	14.5	14.5	14.5	14.5	15.6	19.0	21.9
25	6.3	6.3	6.3	6.3	8.3	9.9	12.0	13.9
50	3.2	3.2	3.2	4.3	5.9	7.0	8.5	9.8
100	1.7	1.7	1.9	3.0	4.2	4.9	6.0	6.9
250	0.7	0.9	1.2	1.9	2.6	3.1	3.8	4.4
500	0.4	0.6	0.9	1.4	1.9	2.2	2.7	3.1
1,000	0.2	0.4	0.6	1.0	1.3	1.6	1.9	2.2
2,500	0.06	0.3	0.4	0.6	0.8	1.0	1.2	1.4
5,000	0.03	0.2	0.3	0.4	0.6	0.7	0.8	1.0
10,000 ...	0.02	0.14	0.2	0.3	0.4	0.5	0.6	0.7
25,000 ...	0.01	0.09	0.12	0.2	0.3	0.3	0.4	0.4
50,000 ...	–	0.06	0.09	0.14	0.2	0.2	0.3	0.3

[1] Standard errors are presented to the nearest one-tenth of 1 percent except when the standard error is less than one-tenth of 1 percent; in those cases, the standard error is shown to the nearest one-hundredth of 1 percent.

The information presented in the following table was extracted from table IX. The entry for "p" is the one sought.

Base of percentage (000)	Estimated percentage		
	25 or 75	30.6	50
10,000 ..	0.6	a	0.7
21,459 ..		p	
25,000 ..	0.4	b	0.4

1. By horizontal interpolation between 0.6 and 0.7, the entry for cell "a" is determined to be 0.62.

$$30.6-25.0 = 5.6$$
$$50.0-25.0 = 25.0$$

$$0.6 + \frac{5.6}{25.0}(0.7-0.6) = 0.62$$

2. By horizontal interpolation between 0.4 and 0.4, the entry for cell "b" is determined to be 0.40.

$$30.6-25.0 = 5.6$$
$$50.0-25.0 = 25.0$$

$$0.4 + \frac{5.6}{25.0}(0.4-0.4) = 0.40$$

3. By vertical interpolation between 0.62 and 0.40, the entry for "p" is determined to be 0.5.

$$21{,}459-10{,}000 = 11{,}459$$
$$25{,}000-10{,}000 = 15{,}000$$

$$0.62 - \frac{11{,}459}{15{,}000}(0.62-0.40) = 0.5$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 30.1 to 31.1 percent; the 90-percent confidence interval is from 29.8 to 31.4 percent; and the 95-percent confidence interval is from 29.6 to 31.6 percent.

Illustration II.—Table A-2 of this report shows in the rural areas of the United States there were 130,000 specified owner-occupied farm housing units in 1976. Interpolation in table III of this appendix shows that the standard error of an

TABLE X. Standard Errors of Estimated Percentages of Seasonal and Migratory Vacant Housing Units: 1976

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	79.3	79.3	79.3	79.3	79.3	79.3	84.7	97.8
10	65.7	65.7	65.7	65.7	65.7	65.7	65.7	69.2
25	43.4	43.4	43.4	43.4	43.4	43.4	43.4	43.7
50	27.7	27.7	27.7	27.7	27.7	27.7	27.7	30.9
100	16.1	16.1	16.1	16.1	16.1	16.1	18.9	21.9
250	7.1	7.1	7.1	7.1	8.3	9.9	12.0	13.8
500	3.7	3.7	3.7	4.3	5.9	7.0	8.5	9.8
1,000	1.9	1.9	1.9	3.0	4.2	5.0	6.0	6.9

estimate of this size is approximately 11,000.

Consequently, the 68-percent confidence interval, as shown by these data, is from 119,000 to 141,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 specified owner-occupied farm housing units lies 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 112,000 to 148,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 108,000 to 152,000 housing units with 95 percent confidence.

Table A-2 also shows that of the 130,000 specified owner-occupied farm housing units in rural areas, 71,000, or 54.6 percent, were owned free and clear. Interpolation in table VIII (i.e., interpolation on both the base and percent) shows that the standard error of the above percentage is approximately 4.7 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 49.9 to 59.3 percent; the 90-percent confidence interval is from 47.1 to 62.1 percent; and the 95-percent confidence interval is from 45.2 to 64.0 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 errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different areas or the difference between separate and uncorrelated characteristics in the same area. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error. Also, if there is a high negative correlation between the two characteristics, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table A-1 of this report shows that in urban areas of the United States there were 3,212,000 renter-occupied housing units with three persons in 1976. Thus, the apparent difference between the number of 1976 renter-occupied housing units in urban areas with two persons and those with three persons is 3,346,000. Table III shows that the standard error of 6,558,000 is approximately 99,000 and that the standard error of 3,212,000 is approximately 71,000. Therefore, the standard error of the estimated difference of 3,346,000 is about:

$$122{,}000 = \sqrt{(99{,}000)^2 + (71{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 3,346,000 difference is from 3,224,000 to 3,468,000 housing units. Therefore, a conclusion that the average estimate of this difference, derived from all possible samples, lies 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 3,151,000 to 3,541,000 and the 95-percent confidence interval is from 3,102,000 to 3,590,000. Thus, we can conclude with 95 percent confidence that the number of 1976 renter-occupied housing units in urban areas with two persons is greater than the number with three persons since the 95-percent confidence interval of this difference does not include zero or negative values.

Medians.—For the 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 the tables, 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 this report shows the median number of persons in owner-occupied housing units in urban areas was 2.8 in 1976. The base of the distribution from which this median was determined is 31,615,000.

1. From table IX, the standard error of a 50-percent characteristic on the base of 31,615,000 is 0.4 percentage points.
2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 49.2 and 50.8.
3. From table A-1, it can be seen by cumulating the frequencies for the first two categories that 14,212,000 housing units, or 45.0 percent, had one and two persons (actually, 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 5,690,000 owner-occupied housing units, or 18.0 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

$$2.5 + (3.5-2.5)\left(\frac{49.2-45.0}{18.0}\right) = 2.7$$

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

$$2.5 + (3.5-2.5)\left(\frac{50.8-45.0}{18.0}\right) = 2.8$$

Thus, the 95-percent confidence interval ranges from 2.7 to 2.8 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 this median (see the paragraph on rounding errors in this appendix).

Nonsampling errors.—In general, 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.

1970 Census.—A number of 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 interviewed housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies on coverage and content errors, as well as the methodology employed, can be found in the 1970 Census of Population and Housing Evaluation and Research Program reports, PHC(E)-5, **The Coverage of Housing in the 1970 Census;** and PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.**

Reinterview program.—For the AHS National 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 subsample 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 "content" error of these AHS estimates.

As part of the reinterview, an additional check was carried out for interviewer evaluation and quality control. This 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 interveiwed 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.

The results of the reinterview study are presented in the following Census Bureau memorandum: "Reinterview Results for Annual Housing Survey-National Sample, 1976." Some of these results are:

1. "Overall, the reinterview results showed moderate to high levels of inconsistency in response between the original interview and reinterview."
2. "One-third of the nonattitudinal items shown in this report showed high levels of inconsistency between the original interview and reinterview."
3. "For the attitudinal items which had enough data to compute reliable measures of response error, over one-third showed high levels of inconsistency, with the remaining items falling into the moderate range."

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low; indices from 20 to 50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

However, unlike past years, the respondent's answers in the reinterview were not reconciled to the original

answers given in the AHS interview; i.e., after the question is answered in the reinterview, the interviewer did not present the previous responses and then ask the respondent to decide upon the best answer. In the past, providing the reinterviewer with the original response had the effect of reducing the levels of inconsistency substantially.

The 1970 census reinterview results provide illustrations of possible nonsampling errors for some of the items which also appear in the AHS. For example, median value of homes was consistently underestimated by about five percent, and the average monthly costs of electricity and utility gas were consistently overestimated although the net effect on average gross rent was fairly small.

A possible explanation for the results of the AHS and census reinterview studies, as well as the surveys themselves, is that the data are based on the answers given by the respondents who may lack precise information. Also, because the results of the reinterview studies are derived from sample surveys, 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 respect to errors of coverage and estimation for missing data, it was mentioned previously (in the section on estimation) that the AHS new construction sample had deficiencies in the representation of conventional new construction. During the sampling of building permits, only those issued more than five months before the survey began were eligible to be selected to represent coventional new construction. Due to time constraints, it is not possible to sample units whose permits are issued less than five months in advance of the survey.

It is estimated that the 1976 AHS sample missed about three percent (i.e., about 300,000 units) of all conventional new construction (i.e., all conventional housing units built after April 1970) because the permits for these units, which were built before October 1976, were issued less than five months in advance of the survey.

In addition, the 1976 Coverage Improvement Program also had certain deficiencies. First, when the canvassing was done to identify mobile home parks that were not in the sample frame or not on the commercial lists, only 92 percent of the census address frame ED's were represented. Second, it appears that the listing procedure (used to find mobile homes placed outside parks, units converted from nonresidential to residential, and houses that had been moved onto their present site) was not very efficient for finding nonresidential conversions (which might be primarily in business districts), since the listing procedure started from a residential unit. (The sample estimate of this component was approximately 16,000 housing units with a standard error of 12,000.)

Finally, it is felt that deficiencies also exist in ED's where area sampling methods are used. As before, it had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as two percent (i.e., as much as 400,000 units) of all housing units in ED's where area sampling methods are used because these units are not listed during the canvassing.

The third stage of ratio estimation corrects for these deficiencies as far as the count of total housing is concerned; i.e., it adjusts the estimate of the total housing inventory to the best available estimate. However, biases of subtotals would still remain.

Rounding errors.—With respect to errors associated with processing, 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 this survey.

TABLE XI. Index for Percentage Tables

Type of housing units	Table number: Occupied and vacant/ occupied only	Table number: Vacant only
Year-round housing units:		
Total housing units—		
Cooking fuel, lacking complete plumbing facilities, mobile homes, source of water, and households with head of Spanish origin	VII	VIII
Other than above	VI	
Urban housing units—		
Cooking fuel, lacking complete plumbing facilities, mobile homes, and new construction	VII	VI
Other than above	IX	
Rural housing units—		
Mobile homes, new construction, source of water, and households with head of Spanish origin	IX	VIII
Other than above	VIII	
Seasonal and Migratory Vacants:		
Total, Urban, Rural	...	X
1973-1976 Lost Housing Units:		
Total, Urban, Rural	...	IX

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Selection of sample areas App-42
Designation of sample housing units for the 1976 survey App-42
Selection of the 1973 sample housing units App-42
Splitting of the sample App-43
Selection of supplemental sample housing units in rural areas App-43
Selection of sample housing units for the 1976 Coverage Improvement Program App-43
ESTIMATION App-44
RELIABILITY OF THE ESTIMATES App-45
Sampling errors App-45
Illustrations of the use of the standard error tables. Illustration I App-46
Illustration II App-47
Differences App-48
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 the median. App-48
Nonsampling errors App-49
Reinterview program App-49
Coverage errors App-49
Rounding errors App-50

SAMPLE DESIGN

The 1976 estimates are based on data collected in October through December 1976 for the 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. The sample for this survey was spread over 461 sample areas (called primary sampling units), comprising 923 counties and independent cities with coverage in each of the 50 States and the District of Columbia.

Approximately 75,500 sample housing units (both occupied and vacant) were eligible for interview in the 1976 Annual Housing Survey. Of this number, 4,500 interviews were classified as "noninterview" for various reasons. Occupied housing units were classified as "noninterview" mainly because the occupants were not found at home after repeated calls. For vacant housing units, interviews were not obtained because an informed respondent was not found after repeated visits. In addition to the 75,500, there were also 6,600 sample units which were visited but found not to be eligible for interview for the AHS in terms of collecting information relevant to the 1976 housing inventory.

Selection of sample areas.—The United States was divided into areas made up of counties and independent cities referred to as primary sampling units (PSU's). These PSU's were then grouped into 376 strata, 156 of which consisted of only one PSU which was in sample with certainty. These 156 strata were mostly the larger SMSA's and were called self-representing (SR) since the sample from the sample area represented just that PSU. Each one of the other 220 strata consisted of a group of PSU's and were referred to as non-self-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSU's in the stratum as well.

One PSU was selected from each NSR stratum with probability proportionate to the 1970 census population of the PSU. (This resulted in 220 NSR sample PSU's.) In addition, the NSR strata were grouped into 110 pairs, and one stratum was picked at random from each pair. From this stratum, an additional PSU was selected independently of the other PSU selected from this stratum. Since the two PSU's were independently selected, it was possible for the same PSU to be selected twice. This occurred in 25 instances, producing an additional 85 NSR sample PSU's, thus giving a grand total of 461 PSU's.

Designation of sample housing units for the 1976 survey.—The sample housing units designated to be interviewed in the 1976 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1975 survey.
2. All sample housing units that were either type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of the survey but which could become eligible in the future) in the 1975 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1976 AHS questionnaire, page 1.)
3. All sample housing units that were selected from the list of building permits issued since the 1975 survey. (This sample represented the housing units built in permit-issuing areas since the 1975 survey).
4. All sample housing units that were selected as part of the 1976 Coverage Improvement Program. (This sample represented most of the housing units which, until 1976, did not have a chance of selection.)

Selection of the 1973 sample housing units.—The overall sampling rate used to select the sample for the 1973 AHS was about 1 in 1,366. The within-PSU sampling rate for the AHS was determined so that the overall probability of selection for each sample housing unit was the same (e.g., if the probability of selecting a NSR PSU was 1 in 10, then the within-PSU sampling rate would be 1 in 136.6).

Within the sample PSU's, a sample of the housing units surveyed in the 1970 Census of Population and Housing was selected for the AHS. In addition, a sample of new construction building permits was also selected to represent the units constructed since the 1970 census. These samples were selected at about twice the rate mentioned previously (i.e., at 2 in 1,366), thereby producing a sample twice as large as needed. This sample was split into two equal-sized samples—one to be used for the AHS, and one to be held in reserve for possible future use for the AHS. The procedure used to split this sample into equal-sized samples is described in the next section.

The sample of 1970 census units was selected in several stages of sampling.

Within the sample PSU's, the first step was the selection of a sample of census enumeration districts (ED's), administrative units used in the 1970 census. The probability of selection for an ED was proportional to the following 1970 census counts of housing units and persons in group quarters, combined in the following formula:

$$\frac{\text{Number of HU's in the ED} + \dfrac{\text{Number of group quarters persons in the ED}}{3}}{4}$$

The next step was to select an expected cluster of about four neighboring housing units within each sample ED. For most of the ED's, the selection was accomplished using the list of addresses for the ED as compiled in the 1970 census. However, in those ED's where addresses were incomplete or inadequate (mostly rural areas), the selection process was accomplished using area sampling methods. These ED's were 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), and a segment was selected. Those selected segments with an expected size which was a multiple of four were further subsampled at the time of the survey so that an expected four housing units were chosen for interview.

The sample of new construction units was selected from building permits issued since January 1970. Within each sample PSU, the building permits were chronologically ordered by month issued, and compact clusters of approximately four housing units were created. These clusters were then sampled for inclusion in the AHS at the sampling rate of 2 in 1,366. Housing units constructed since the 1970 census in areas which do not issue building permits were brought into the sample as a result of the area sample described.

Splitting of the sample.—The described sample selection procedure produced clusters (or segments) of size-four housing units for the sample taken from the census address frame, the new construction frame, and the area sampling frame (mainly rural areas). Clusters of this size should result in a minimum loss in precision for estimates of housing characteristics in rural areas because of the heterogeneity of neighboring units. However, clusters of size-two housing units were considered to be more optimum within those areas where the housing characteristics of neighboring units tend to be very similar (i.e., urban areas and new construction units). A splitting operation was then carried out for clusters selected from the census address and the new construction frames. This consisted of halving each sample cluster from these frames. Thus, two housing units from each of these clusters were included in the survey and two housing units were held in reserve. No splitting operation was carried out within the clusters selected from the area sampling frame; every other area sample cluster of four housing units was used for the survey and the remaining clusters were assigned to the reserve sample.

Selection of supplemental sample housing units in rural areas.—In 1974, it was decided to increase the reliability of the AHS estimates of rural housing characteristics by doubling the number of sample housing units from rural areas. This was accomplished by reactivating the reserve sample, selected in the original sampling operations in 1973, from rural areas only. For the reserve sample selected in census address and new construction frames, this meant that the other half of each rural cluster (an expected two housing units) was reactivated in 1974. Similarly for the area sampling frame, this meant the entire reserve cluster (an expected four housing units) was reactivated in 1974 if the cluster was rural. This supplementation increased the overall probability of selection for sample housing units in rural areas to about 2 in 1,366; whereas, the overall probability of selection for sample housing units in urban areas remained at 1 in 1,366.

Selection of sample housing units for the 1976 Coverage Improvement Program.—The 1976 Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS National sample from the census address and new construction frames. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970.
2. Units converted to residential use in structures totally nonresidential at the time of the 1970 census.
3. Houses that have been moved onto their present site since the 1970 census.
4. Mobile homes placed in parks either missed in or established since the 1970 census.
5. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

A sample of new construction units whose permits were issued before January 1970 was selected in two stages. First, units whose permits were issued before January 1970, but which were completed after the census, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. These units were then sampled so that the overall probability of selection was about 1 in 1,320.

A sample of mobile homes placed in a park, missed by the census, or established after the census was also selected in two stages. During the first stage, a list of mobile home parks was obtained from commercial listings. This list was then supplemented by additional parks identified by a canvassing operation similar to that performed in ED's where area sampling methods are used. The second stage consisted of dividing the parks into clusters of an expected size of four sites. These clusters were then sampled so that the overall probability of selection was about 1 in 1,366.

For the remaining units (i.e., mobile homes placed outside parks since the 1970 census, mobile homes vacant at the time of the 1970 census, units converted from nonresidential to residential use since the 1970 census, and houses that had been moved onto their present site since the 1970 census), the sampling was done in three stages. First, a subsample of the regular AHS sample units from the census address frame was selected. Second, succeeding structures that had been eligible to be selected from the

census address frame were then listed until eight structures (excluding mobile home parks) were found. Finally, the intervening structures that had been listed which did not have a chance of selection in the AHS were identified and the units within these structures were interviewed.

ESTIMATION

In 1976 the AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the type A noninterview housing units encountered in the AHS. This noninterview adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

$$\frac{\text{Interviewed housing units} + \text{Noninterviewed housing units}}{\text{Interviewed housing units}}$$

The first-stage ratio estimation procedure was employed for sample housing units from non-self-representing (NSR) PSU's only. This procedure was designed to reduce the contribution to the variance arising from the sampling of PSU's. The first-stage ratio estimation procedure takes into account the differences that existed at the time of the 1970 census in the distribution by tenure and residence of the housing population estimated from the sample NSR PSU's and that of the NSR housing population in each of the four census regions of the country. The first-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{The 1970 census housing population in the residence-tenure category for all NSR strata in a census region}}{\text{Estimate of the housing population category using 1970 census housing counts for sample NSR PSU's in a census region}}$$

The numerators of the ratios were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR stratum and summing these counts across the NSR strata in each census region. The denominators were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR sample PSU, weighting these counts by the inverse of the probability of selecting that PSU, and summing these weighted counts across the NSR sample PSU's in each census region. The computed first-stage ratio estimation factor was then applied to the existing weight for each NSR sample unit in each first-stage ratio estimation category.

The second-stage ratio estimation procedure was designed to adjust the AHS sample estimate of one category of conventional new construction units; i.e., one category of sample units built April 1, 1970, or later to an independently derived current estimate where a known deficiency in the AHS sample exists (see the section on nonsampling error) for each of the four regions. This estimate was considered to be the best estimate available for the number of conventional new construction units in this category. The second-stage ratio estimation factor was as follows:

$$\frac{\text{Current best estimate of new construction in the category}}{\text{AHS sample estimate of new construction units in the category}}$$

The numerators of the ratios were derived from data based on the Survey of Construction. The denominators of the ratios were obtained from the weighted estimates for the AHS sample units 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 each second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all the AHS sample units. This procedure was designed to adjust the AHS sample estimates of housing (i.e., the estimates employing the noninterview, first-stage, and second-stage adjustments) to independently derived current housing estimates for 4 categories of vacant housing units and for 24 categories of occupied housing units. Each of these 24 categories is a combination of the characteristics of residence, tenure, race of head, and sex of head.

The third-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{Current independent estimate of housing units in the category}}{\text{AHS sample estimate of housing units in the category}}$$

The numerators of the ratios for occupied housing units were derived from data based on the Current Population Survey (CPS), a sample household survey conducted monthly by the Bureau of the Census. The numerators of the ratios for vacant housing units were derived from data based on the Housing Vacancy Survey (HVS), a quarterly vacancy survey also conducted by the Bureau of the Census.

The denominators of the ratios were obtained from the weighted estimates for the 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 in each third-stage ratio estimation category.

The second-stage and the third-stage ratio estimation procedures were repeated in an iterative process in order to bring the AHS estimates into close agreement with both sets of "independent" estimates. The second stage was modified so that the estimates for all 12 categories of new construction would be identical to the estimates before the third stage. Hence, the repeated second stage had the effect of controlling the AHS sample estimates of new construction units to the "unbiased" sample estimates for 11 categories of new construction units for each of the 4 regions (i.e., 7 categories for conventional new construction units and 4 for new construction mobile homes) and, as before, of adjusting the AHS sample estimate of 1 category of conventional new construction units to an independently derived current estimate.

The numerators were either the unbiased weighted estimates for the AHS sample units, using the existing weight after the first-stage ratio estimation procedure (i.e., the estimates employing the noninterview and first-stage adjustments) or the independent estimate derived from data based on the Survey of Construction.

The denominators of the ratios in this iterative process were obtained from the weighted estimates for the AHS sample units after the previous stage of ratio estimation. The factors resulting from this iterative process were then applied to the existing weight on the appropriate records, and the resulting product was used as the final weight for tabulation.

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. The distribution of the housing population selected for the sample differed somewhat, by chance, from that of the nation as a whole in such basic housing characteristics as tenure, vacancy status, residence, race of head, and sex of head. These characteristics are probably closely correlated with other housing characteristics measured for the AHS. Therefore, through the use of the three-stage ratio estimation procedure, one can expect the sample estimate to be improved substantially.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS National sample.

Sampling errors.—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 interviewers were used, estimates from each of the different samples would differ from each other. The average deviation (from all possible samples) of a sample estimate from the average of the estimate (derived from 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 interviewer 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 errors measured by the standard error and biases and some additional nonsampling errors not measured by the standard error.

The procedure, as illustrated, provides a method to construct interval estimates so that a known proportion of the intervals would contain the average of all possible samples. For example, if all possible samples were selected, each of these samples were surveyed under 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 either is or is not contained in any particular computed interval. However, for a particular sample one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables are approximations to the standard errors of various estimates shown in this 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 the precise standard error for any specific item.

Tables I and II present the standard errors applicable to the 1976 housing inventory estimates in this report. Table I shows the approximate standard errors applicable to all 1976 housing inventory estimates except those pertaining to the indicated specified items. The standard errors shown in table II should be used for those specified items. Linear inter-

TABLE I. Standard Errors of Estimated Number of Housing Units: 1976 (Excluding estimates of housing units pertaining to Lacking Complete Plumbing Facilities and Households with Head of Spanish Origin)

(68 chances out of 100)

Size of estimate (000)	Standard error: Total or White (000)	Standard error: Black (000)	Size of estimate (000)	Standard error: Total or White (000)	Standard error: Black (000)
0	1	1	1,000	38	36
5	3	3	2,500	60	50
10	4	4	5,000	83	50
25	6	6	10,000	113	–
50	9	9	25,000	158	–
100	12	12	50,000	164	–
250	19	19	75,000	76	–
500	27	26			

TABLE II. Standard Errors of Estimated Number of Housing Units Pertaining to Lacking Complete Plumbing Facilities and Households with Head of Spanish Origin: 1976

(68 chances out of 100)

Size of estimate (000)	Standard error: Total, White, or Spanish origin (000)	Standard error: Black (000)	Size of estimate (000)	Standard error: Total, White, or Spanish origin (000)	Standard error: Black (000)
0	2	2	1,000	45	43
5	3	3	2,500	71	59
10	5	5	5,000	99	59
25	7	7	10,000	135	—
50	10	10	25,000	189	—
100	14	14	50,000	196	—
250	23	22	75,000	90	—
500	32	31			

TABLE III. Standard Errors of Estimated Percentages of Housing Units: 1976 (Excludes estimated percentages of housing units pertaining to Lacking Complete Plumbing Facilities and Households with Head of Spanish Origin)

(68 chances out of 100)

Base of Percentage (000)	Estimated percentage[1]: 0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	22.7	22.7	22.7	22.7	22.7	22.7	23.4	27.1
10	12.8	12.8	12.8	12.8	12.8	13.7	16.6	19.1
25	5.5	5.5	5.5	5.5	7.3	8.6	10.5	12.1
50	2.8	2.8	2.8	3.7	5.1	6.1	7.4	8.6
100	1.4	1.4	1.7	2.6	3.6	4.3	5.2	6.1
250	0.6	0.8	1.1	1.7	2.3	2.7	3.3	3.8
500	0.3	0.5	0.8	1.2	1.6	1.9	2.3	2.7
1,000	0.15	0.4	0.5	0.8	1.1	1.4	1.7	1.9
2,500	0.06	0.2	0.3	0.5	0.7	0.9	1.0	1.2
5,000	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.9
10,000	0.02	0.12	0.2	0.3	0.4	0.4	0.5	0.6
25,000	0.01	0.08	0.11	0.2	0.2	0.3	0.3	0.4
50,000	—	0.05	0.08	0.12	0.2	0.2	0.2	0.3
75,000	—	0.04	0.06	0.10	0.13	0.2	0.2	0.2

[1] Standard errors are presented to the nearest one-tenth of 1 percent except when the standard error is less than one-tenth of 1 percent; in those cases, the standard error is shown to the nearest one-hundredth of 1 percent.

polation should be used to determine standard errors for levels of estimates not specifically shown in tables I and II.

The reliability of an estimated percentage, computed by using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more.

Tables III and IV present the standard errors of estimated percentages. Table III shows the approximate standard errors of all estimated percentages of housing units except those pertaining to the specified items in table IV. The standard errors shown in table IV should be used for those specified items. Two-way linear interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables III and IV.

Included in tables I through IV are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as overestimates of the true standard errors, and should be used primarily for construction of confidence intervals for characteristics when an estimate of zero is obtained.

For ratios of the form (100) (x/y), where x is not a subclass of y, tables III 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:

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

where: 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

Illustrations of the use of the standard error tables. Illustration I.—Table A-1 of this report shows that in the United States there were 21,989,000 owner-occupied housing units in 1976 with garbage collection service once a week. Interpolation in table I shows that the standard error of an estimate of this size is approximately 149,000. The following procedure was used in interpolating:

The information presented in the following table was extracted from table I. The entry for "x" is the one sought.

Size of estimate (000)	Standard error (000)
10,000	113
21,989	x
25,000	158

By vertically interpolating between 113,000 and 158,000, the entry for "x" is determined to be

$$21{,}989-10{,}000 = 11{,}989$$
$$25{,}000-10{,}000 = 15{,}000$$
$$113 + \frac{11{,}989}{15{,}000}(158-113) = 149$$

Consequently, the 68-percent confidence interval, as shown by these data is from 21,840,000 to 22,138,000 housing units. Therefore, a conclusion that the average estimate of 1976 housing units of this type lies 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 21,751,000 to 22,227,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 21,691,000 to 22,287,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 21,989,000, 1976 owner-occupied housing units with garbage collection service once a week, 1,471,000, or 6.7 percent, had a family and primary individual income between $5,000 and $6,999. Interpolation in table III (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the percentage is 0.2 percentage points. The following procedure was used in interpolating:

The information presented in the following table was extracted from table III. The entry for "p" is the one sought.

Base of percentage (000)	Estimated percentage		
	5 or 95	6.7	10 or 90
10,000 . .	0.3	a	0.4
21,989 . .		p	
25,000 . .	0.2	b	0.2

1. By horizontal interpolation between 0.3 and 0.4, the entry for cell "a" is determined to be 0.33.

$$6.7-5.0 = 1.7$$
$$10.0-5.0 = 5.0$$
$$0.3 + \frac{1.7}{5.0}(0.4-0.3) = 0.33$$

2. By horizontal interpolation between 0.2 and 0.2, the entry for cell "b" is determined to be 0.20.

$$6.7-5.0 = 1.7$$
$$10.0-5.0 = 5.0$$
$$0.2 + \frac{1.7}{5.0}(0.2-0.2) = 0.20$$

3. By vertical interpolation between 0.33 and 0.20, the entry for "p" is determined to be 0.2.

$$21{,}989-10{,}000 = 11{,}989$$
$$25{,}000-10{,}000 = 15{,}000$$
$$0.33-\frac{11{,}989}{15{,}000}(0.33-0.20) = 0.2$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 6.5 to 6.9 percent; the 90-percent confidence interval is from 6.4 to 7.0 percent; and the 95-percent confidence interval is from 6.3 to 7.1 percent.

Illustration II.—Table A-3 of this report shows that in the United States in 1976 there were 718,000 owner-occupied housing units lacking some or all plumbing facilities that were occupied three months or longer. Interpolation in table II of this appendix shows that the standard error of an estimate of this size is approximately 38,000. Consequently, the 68-percent confidence interval is from 680,000 to 756,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 owner-occupied housing units lacking some or all plumbing facilities that were occupied three months or longer lies 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 657,000 to 779,000 housing units with 90 percent confidence; and that average estimate lies within the interval from 642,000 to

TABLE IV. Standard Errors of Estimated Percentages of Housing Units Pertaining to Lacking Complete Plumbing Facilities and Households with Head of Spanish Origin: 1976

(68 chances out of 100)

Base of percentage (000)	Estimated percentage[1]							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	29.5	29.5	29.5	29.5	29.5	29.5	29.5	32.3
10	17.3	17.3	17.3	17.3	17.3	17.3	19.8	22.9
25	7.7	7.7	7.7	7.7	8.7	10.3	12.5	14.5
50	4.0	4.0	4.0	4.5	6.1	7.3	8.9	10.2
100	2.0	2.0	2.0	3.2	4.3	5.2	6.3	7.2
250	0.8	0.9	1.3	2.0	2.7	3.3	4.0	4.6
500	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
1,000	0.2	0.5	0.6	1.0	1.4	1.6	2.0	2.3
2,500	0.08	0.3	0.4	0.6	0.9	1.0	1.3	1.4
5,000	0.04	0.2	0.3	0.4	0.6	0.7	0.9	1.0
10,000 ...	0.02	0.14	0.2	0.3	0.4	0.5	0.6	0.7
25,000 ...	0.01	0.09	0.13	0.2	0.3	0.3	0.4	0.5
50,000 ...	–	0.06	0.09	0.14	0.2	0.2	0.3	0.3
75,000 ...	–	0.05	0.07	0.12	0.2	0.2	0.2	0.3

[1] Standard errors are presented to the nearest one-tenth of 1 percent except when the standard error is less than one-tenth of 1 percent; in those cases, the standard error is shown to the nearest one-hundredth of 1 percent.

794,000 housing units with 95 percent confidence.

Table A-3 also shows that of the 718,000, 1976 owner-occupied housing units lacking some or all plumbing facilities that were occupied three months or longer, 84,000, or 11.7 percent, had a family and primary individual income of between $5,000 and $6,999. Interpolation in table IV (i.e., interpolation on both the base and the percent) of this appendix shows that the standard error of the above percentage is 1.8 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 9.9 to 13.5 percent; the 90-percent confidence interval is from 8.8 to 14.6 percent; and the 95-percent confidence interval is from 8.1 to 15.3 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 errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different areas or the difference between separate and uncorrelated characteristics in the same area. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error. However, if there is a high negative correlation between the two characteristics, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference.—Table 1 of this report shows that in the United States there were 4,384,000 owner-occupied housing units in 1976 with garbage collection service once a week that had a family and primary individual income of between $10,000 and $14,999. Thus the apparent difference between the number of owner-occupied housing units with garbage collection service once a week that had a family and primary individual income of between $5,000 and $6,999 and the number that had a family and primary individual income of between $10,000 and $14,999 is 2,913,000. Table I shows the standard error on an estimate of 4,384,000 to be approximately 77,000 and table I also shows the standard error on an estimate of 1,471,000 to be approximately 45,000. Therefore, the standard error of the estimated difference of 2,913,000 is about

$$89{,}000 = \sqrt{(77{,}000)^2 + (45{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 2,913,000 difference is from 2,824,000 to 3,002,000 housing units. Therefore, a conclusion that the average estimate of this difference, derived from all possible samples, lies 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 2,771,000 to 3,055,000 housing units, and the 95-percent confidence interval is from 2,735,000 to 3,091,000. Thus we can conclude with 95 percent confidence that of the 1976 owner-occupied housing units with garbage collection service once a week, the number with an income between $10,000 and $14,999 was greater than the number with an income between $5,000 and $6,999.

Medians.—For the 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 the tables, 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 the median.—Table 1 of this report shows that the median family and primary individual income of owner-occupied housing units in the United States with garbage collection service once a week was $15,600 in 1976. The base of the distribution, from which this median was determined, is 21,989,000 housing units.

1. Table III shows that the standard error of 50 percent on a base of 21,989,000 is 0.4 percentage points.
2. To obtain a two-standard error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 49.2 and 50.8.
3. From table A-1, it can be seen by cumulating the frequencies for the first five categories that 10,485,000 housing units, or 47.7 percent, had a family and primary individual income of up to $14,999, and that an additional 4,045,000, or 18.4 percent, had a family and primary individual income of between $15,000 and $19,999. By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$\$15{,}000 + (\$5{,}000)\left(\frac{49.2-47.7}{18.4}\right) = \$15{,}400$$

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

$$\$15{,}000 + (\$5{,}000)\left(\frac{50.8-47.7}{18.4}\right) = \$15{,}800$$

Thus, the 95-percent confidence interval ranges from $15,400 to $15,800.

Nonsampling errors.—In general, 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.

Reinterview program.—For the AHS National 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 subsample 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 "content" error of these AHS estimates.

As part of the reinterview, an additional check was carried out for interviewer evaluation and quality control. This 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.

The results of the reinterview study are presented in the following Census Bureau memorandum: "Reinterview Results for Annual Housing Survey-National Sample, 1976." Some of these results are:

1. "Overall, the reinterview results showed moderate to high levels of inconsistency in response between the original interview and reinterview."
2. "One-third of the nonattitudinal items shown in this report showed high levels of inconsistency between the original interview and reinterview."
3. "For the attitudinal items which had enough data to compute reliable measures of response error, over one-third showed high levels of inconsistency, with the remaining items falling into the moderate range."

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low; indices from 20 to 50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

However, unlike past years, the respondent's answers in the reinterview were not reconciled to the original answers given in the AHS interview; i.e., after the question is answered in the reinterview, the interviewer did not present the previous responses and then ask the respondent to decide upon the best answer. In the past, providing the reinterviewer with the original response had the effect of reducing the levels of inconsistency substantially.

The 1970 census reinterview results provide illustrations of possible nonsampling errors for some of the items which also appear in the AHS. For example, median value of homes was consistently underestimated by about 5 percent, and the average monthly costs of electricity and utility gas were consistently overestimated although the net effect on average gross rent was fairly small.

A possible explanation for the results of the AHS and census reinterview studies, as well as the surveys themselves, is that the data are based on the answers given by the respondents, who may lack precise information. Also, because the results of the reinterview studies are derived from sample surveys, 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 respect to errors of coverage and estimation for missing data, it was mentioned previously (in the section on estimation) that the AHS new construction sample had deficiencies in representing conventional new construction. During the sampling of building permits, only those issued more than 5 months before the survey began were eligible to be selected to represent conventional new construction. Due to time constraints, it is not possible to sample units whose permits are issued less than 5 months in advance of the survey.

It is estimated that the 1976 AHS sample missed about 3 percent (i.e., about 300,000 units) of all conventional new construction (i.e., all conventional housing units built after April 1970) because the permits for these units, which were built before October 1976, were issued less than 5 months in advance of the survey.

In addition, the 1976 Coverage Improvement Program also had certain deficiencies. First, when the canvassing was done to identify mobile home parks that were not in the sample frame or not on the commercial lists, only 92 percent of the census address frame ED's were represented. Second, it appears that the listing procedure (used to find mobile homes placed outside parks, units converted from nonresidential to residential, and houses that had been moved onto their present site) was not very efficient for finding nonresidential conversions, (which might be primarily in business districts), since the listing procedure started from a residential unit. (The sample estimate of this component was approximately 16,000 housing units with a standard error of 12,000.)

Finally, it is felt that deficiencies also exist in ED's where area sampling

methods are used. As before, it had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent (i.e., as much as 400,000 units) of all housing units in ED's where area sampling methods are used because these units are not listed during the canvassing.

The third stage of ratio estimation corrects for these deficiencies as far as the count of total housing is concerned; i.e., as stated before, it adjusts the estimate of the total housing inventory to the best available estimate. However, biases of subtotals would still remain.

Rounding errors.—With respect to errors associated with processing, 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 results of this survey.