Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-44
Selection of sample areas App-44
Designation of sample housing units for the 1977 survey App-44
Selection of the 1973 sample housing units App-44
Splitting of the sample App-45
Selection of supplemental sample housing units in rural areas App-45
Selection of sample housing units for the 1976 Coverage Improvement Program App-45
1977 reduction..... App-46
1970 Census of Population and Housing App-46
ESTIMATION App-46
AHS national sample App-46
1977 housing inventory App-46
1973-1977 lost units App-47
Ratio estimation procedure of the 1970 Census of Population and Housing App-47
RELIABILITY OF THE ESTIMATES App-47
Sampling errors App-47
Standard errors of estimates of levels App-48
Standard errors of estimates of percentages App-49
Standard errors of ratios App-49
Illustration of the use of the standard error tables. Illustration I App-49
Illustration II App-50
Differences App-50
Illustration of the computation of the standard error of a difference App-50
Medians App-52
Illustration of the computation of the 95-percent confidence interval for a median. App-52
Nonsampling errors App-53
1970 census App-54
Reinterview program App-54
Coverage errors App-55
Rounding errors App-56

SAMPLE DESIGN

The 1977 estimates are based on data collected in October 1977 through January 1978 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 70,600 sample housing units (both occupied and vacant) were eligible for interview in the 1977 Annual Housing Survey. Of this number, 4,000 interviews were classified as "noninterview" for various reasons. Occupied housing units were classified as "noninterview," mainly, because the occupants refused to be interviewed 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 70,600, there were also 6,300 sample units which were visited but found not to be eligible for interview for the AHS in terms of collecting information relevant to the 1977 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 1 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 1 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 1977 survey.—The sample housing units designated to be interviewed in the 1977 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1976 survey (which included all sample housing units that were selected as part of the 1976 Coverage Improvement Program) and that were not part of the 1977 reduction.
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 1976 survey and that were not part of the 1977 reduction. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1977 AHS questionnaire, page 1.)
3. All sample housing units that were selected from the list of building permits issued since the 1976 survey. (This sample represented the housing units built in permit-issuing areas since the 1976 survey.)

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. 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 at the 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 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 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.

1977 reduction.—By 1977, the addition to the sample from primarily new construction and the coverage improvements had increased the total sample size (interviews plus noninterviews) to about 81,000. The sample was reduced by about 7 percent to approximately 75,000. However, this reduction did not include any CEN-SUP[1] units or units which were selected as part of the 1976 Coverage Improvement Program. Thus, the overall probability of selection for these latter units remained unchanged, and, for the rest of the units, their probability of selection was about 1 in 1,472, if they were urban, and about 1 in 736, if they were rural.

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.**

[1] CEN-SUP units resulted from a 1970 census evaluation study and represented units missed in the 1970 census.

ESTIMATION

AHS national sample.—The AHS national sample produced estimates of two types: Estimates of the 1977 housing inventory and estimates of units removed from the housing inventory between 1973 and 1977 (i.e., 1973-1977 lost units). Each type of estimate employed a separate, though similar, estimation procedure.

1977 housing inventory.—In 1977, 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 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- and third-stage ratio estimation procedures were iterated 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 15 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 14 categories of new construction units for each of the 4 regions (i.e., 9 categories for conventional new construction units and 5 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.

1973-1977 lost units.—The 1973-1977 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, H-150-73A, **General Housing Characteristics for the United States and Regions.** These 1973-1977 lost units do not include the HU's from the 1976 Coverage Improvement. Since the 1973-1977 lost units existed, by definition, in the 1973 housing inventory, there was a 1973 housing inventory weight associated with each 1973-1977 lost unit. This weight, adjusted for the 1977 reduction, was used to tabulate the estimates of the characteristics of the 1973-1977 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 variability between estimates from all possible samples is defined as sampling error. One common measure of sampling error is the standard error which measures the precision with which an estimate from a sample approximates the average results of all possible samples. In addition, the standard error, as calculated for this report, partially reflects the variation in the estimates due to some 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 sample estimate and its estimated standard error enable one to construct interval estimates so that the interval includes the average result of all possible samples with a known probability. 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 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.

Standard errors of estimates of levels.—Tables I, II, and III present the standard errors applicable to the 1977 national housing inventory estimates and tables IV and V present the standard errors applicable to 1973-1977 lost housing unit estimates in this report. Tables VIa, b, and c present the standard errors for each of the four regions; Northeast, North Cen-

TABLE I. Standard Errors of Estimated Numbers of Housing Units: 1977 (Excluding Estimates of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, New Construction, 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	2	2	1,000	40	38
5	3	3	2,500	63	53
10	4	4	5,000	88	55
25	6	6	10,000	120	—
50	9	9	25,000	168	—
100	13	13	50,000	177	—
250	20	20	75,000	95	—
500	29	28			

TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, New Construction, Source of Water, and Households with Head of Spanish Origin: 1977

(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	48	45
5	3	3	2,500	76	64
10	5	5	5,000	105	66
25	8	8	10,000	144	—
50	11	11	25,000	202	—
100	15	15	50,000	213	—
250	24	24	75,000	114	—
500	34	33			

tral, South, and West. Linear interpolation should be used to determine standard error for levels of estimates not specifically shown in tables I through VI.

Standard errors of estimates of percentages.—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 VII, VIII, IX, X, XI, and XIIa, b, and c present the standard errors of estimated percentages. Table VII shows the approximate standard errors of all national estimated percentages of housing units except those pertaining to the specified items in tables II and III. The standard errors shown in tables VIII and IX should be used for those specified items. Tables X and XI show the approximate standard errors of national estimated percentages of 1973-1977 lost housing units. Tables XIIa, b, and c show the approximate standard errors of all regional estimated percentages of housing units and 1973-1977 lost housing units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables VII through XII.

TABLE III. Standard Errors of Estimated Numbers of Vacant Housing Units: 1977

(68 chances out of 100)

Size of estimate (000)	Standard error: Year-round vacants (000)	Standard error: Seasonal and migratory vacants (000)
0	1	2
5	2	4
10	3	5
25	5	8
50	7	11
100	10	16
250	17	28
500	24	43
1,000	35	69
2,500	59	144
3,500	74	—

TABLE IV. Standard Errors of Estimated Numbers of Lost Housing Units: 1973-1977 (Excluding Estimates of Lost Housing Units Pertaining to Lacking Bedroom, Lacking Kitchen Facilities, Lacking Some or All Plumbing Facilities, and Other Vacant)

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	250	22
5	3	500	32
10	4	750	41
25	7	1,000	49
50	10	1,750	70
100	14	2,500	89

TABLE V. Standard Errors of Estimated Numbers of Lost Housing Units Pertaining to Lacking Bedroom, Lacking Kitchen Facilities, Lacking Some or All Plumbing Facilities, and Other Vacant: 1973-1977

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	250	27
5	3	500	42
10	5	750	56
25	8	1,000	69
50	11	1,750	107
100	16	2,500	145

Included in tables I through XII are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered to be 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.

Standard errors of ratios.—For ratios of the form (100) (x/y), where x is not a subclass of y, tables VII through XII 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:

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

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 inside SMSA's in the United States there were 9,631,000 owner-occupied housing units with 2 persons in 1977. Interpolation of the data in table I shows that the standard error of an estimate of this size is approximately 118,000. 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)
5,000	88
9,631	x
10,000	120

By vertically interpolating between 88 and 120, the entry for "x" is determined to be 118.

$$9{,}631-5{,}000 = 4{,}631$$
$$10{,}000-5{,}000 = 5{,}000$$

$$88+\frac{4{,}631}{5{,}000}(120-88) = 118$$

Consequently, the 68-percent confidence interval, as shown by these data, is from

9,513,000 to 9,749,000 housing units. Therefore, a conclusion that the average estimate of 1977 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 9,442,000 to 9,820,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 9,395,000 to 9,867,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 9,631,000 owner-occupied housing units with 2 persons inside SMSA's, 3,632,000, or 37.7 percent, were in central cities. Interpolation of the data in table VII (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.

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

Base of percentage (000)	Estimated percentage		
	25	37.7	50
5,000 . . .	0.8	a	0.9
9,631 . . .		p	
10,000 . .	0.6	b	0.6

1. By horizontal interpolation between 0.8 and 0.9, the entry for cell "a" is determined to be 0.9.

$$37.7-25.0 = 12.7$$
$$50.0-25.0 = 25.0$$

$$0.8 + \frac{12.7}{25.0}(0.9-0.8) = 0.9$$

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

$$37.7-25.0 = 12.7$$
$$50.0-25.0 = 25.0$$

$$0.6 + \frac{12.7}{25.0}(0.6-0.6) = 0.60$$

3. By vertical interpolation between 0.9 and 0.6, the entry for "p" is determined to be 0.6.

$$9{,}631-5{,}000 = 4{,}631$$
$$10{,}000-5{,}000 = 5{,}000$$

$$0.9 + \frac{4{,}631}{5{,}000}(0.6-0.9) = 0.6$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 37.1 to 38.3 percent; the 90-percent confidence interval is from 36.7 to 38.7 percent; and the 95-percent confidence interval is from 36.5 to 38.9 percent.

Illustration II.—Table A-1 of this report shows that in the United States in 1977 there were 112,000 housing units in structures with 4 floors or more (see "elevator in structure" item) that were outside of SMSA's. Interpolation of the data 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 98,000 to 126,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1977 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 90,000 to 134,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 84,000 to 140,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 112,000, 1977 housing units in structures with 4 floors or more that were outside SMSA's, 89,000, or 79.5 percent, were in structures that contained elevators. Interpolation of the data in table VII (i.e., interpolation on both the base and the percent) shows that the standard error of the percentage is 4.9 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 74.6 to 84.4 percent; the 90-percent confidence interval is from 71.7 to 87.3 percent; and the 95-percent confidence interval is from 69.7 to 89.3 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 characteristics in two different areas or the difference between separate and uncorrelated characteristics in the same area. If 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 5,766,000 owner-occupied housing units with 3 persons in 1977. Thus, the apparent difference between the number of 1977 owner-occupied housing units with 2 persons and those with 3 persons is 3,865,000.

TABLE VIa. Standard Errors of Estimated Numbers of Housing Units Pertaining to Each of the Regions; Northeast, North Central, South, and West: 1977 (Excluding Estimates of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, New Construction, Source of Water, Households with Head of Spanish Origin, and Vacant Housing Units)

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	1,000 . .	45
5	3	2,500 . .	70
10	4	5,000 . .	97
25	7	10,000 .	133
50	10	25,000 .	187
100	14	50,000 .	196
250	22	75,000 .	105
500	32		

Note: For standard errors of regional estimates of vacant housing units, use the national standard errors in table III multiplied by 1.1.

The standard error of 9,631,000 is approximately 118,000. Table I shows that the standard error on an estimate of 5,766,000 to be approximately 93,000. Therefore, the standard error of the estimated difference of 3,865,000 is about 150,000.

$$150{,}000 = \sqrt{(118{,}000)^2 + (93{,}000)^2}$$

TABLE VIb. Standard Errors of Estimated Numbers of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, New Construction, Source of Water, and Households with Head of Spanish Origin for Each of the Regions; Northeast, North Central, South and West: 1977

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	1,000	54
5	4	2,500	84
10	5	5,000	117
25	9	10,000	160
50	12	25,000	225
100	17	50,000	236
250	27	75,000	126
500	38		

TABLE VIc. Standard Errors of Estimated Numbers of Lost Housing Units for Each of the Regions; Northeast, North Central, South, and West: 1973-1977 (Excluding Estimates of Lost Housing Units Pertaining to Lacking Bedroom, Lacking Kitchen Facilities, Lacking Some or All Plumbing Facilities, and Other Vacant)

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	100	15
5	3	250	23
10	5	500	34
25	7	750	42
50	10	1,000	50

Note: For standard errors of regional estimates of lost housing units pertaining to these specified items, use the national standard errors in table V.

TABLE VII. Standard Errors of Estimated Percentages of Housing Units: 1977 (Excludes Estimated Percentages of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, New Construction, 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	24.6	24.6	24.6	24.6	24.6	24.6	24.8	28.6
10	14.0	14.0	14.0	14.0	14.0	14.0	17.5	20.2
25	6.1	6.1	6.1	6.1	7.7	9.1	11.1	12.8
50	3.2	3.2	3.2	3.9	5.4	6.5	7.8	9.0
100	1.6	1.6	1.8	2.8	3.8	4.6	5.5	6.4
250	0.6	0.8	1.1	1.8	2.4	2.9	3.5	4.0
500	0.3	0.6	0.8	1.2	1.7	2.0	2.5	2.9
1,000	0.2	0.4	0.6	0.9	1.2	1.4	1.8	2.0
2,500	0.07	0.3	0.4	0.6	0.8	0.9	1.1	1.3
5,000	0.03	0.2	0.3	0.4	0.5	0.6	0.8	0.9
10,000	0.02	0.13	0.2	0.3	0.4	0.5	0.6	0.6
25,000	0.01	0.08	0.11	0.2	0.2	0.3	0.4	0.4
50,000	—	0.06	0.08	0.12	0.2	0.2	0.2	0.3
75,000	—	0.05	0.07	0.10	0.14	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 VIII. Standard Errors of Estimated Percentages of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, New Construction, Source of Water, Households with Head of Spanish Origin, and Seasonal and Migratory Vacant Housing Units: 1977

(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	32.2	32.2	32.2	32.2	32.2	32.2	32.2	34.4
10	19.2	19.2	19.2	19.2	19.2	19.2	21.1	24.3
25	8.7	8.7	8.7	8.7	9.2	11.0	13.3	15.4
50	4.5	4.5	4.5	4.7	6.5	7.8	9.4	10.9
100	2.3	2.3	2.3	3.4	4.6	5.5	6.7	7.7
250	0.9	1.0	1.4	2.1	2.9	3.5	4.2	4.9
500	0.5	0.7	1.0	1.5	2.1	2.5	3.0	3.4
1,000	0.2	0.5	0.7	1.1	1.5	1.7	2.1	2.4
2,500	0.09	0.3	0.4	0.7	0.9	1.1	1.3	1.5
5,000	0.05	0.2	0.3	0.5	0.7	0.8	0.9	1.1
10,000	0.02	0.2	0.2	0.3	0.5	0.5	0.7	0.8
25,000	0.01	0.10	0.14	0.2	0.3	0.3	0.4	0.5
50,000	—	0.07	0.10	0.2	0.2	0.2	0.3	0.3
75,000	—	0.06	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.

Consequently, the 68-percent confidence interval for the 3,865,000 difference is from 3,715,000 to 4,015,000 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 3,625,000 to 4,105,000 housing units and the 95-percent confidence interval is from 3,565,000 to 4,165,000. Thus, we can conclude with 95 percent confidence that the number of 1977 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 appropriate standard error table, 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.

TABLE IX. Standard Errors of Estimated Percentages of Year-Round Vacant Housing Units and Housing Vacancy Rates: 1977

(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	17.6	17.6	17.6	17.6	17.6	17.6	20.0	23.1
10	9.7	9.7	9.7	9.7	9.8	11.7	14.2	16.3
25	4.1	4.1	4.1	4.5	6.2	7.4	9.0	10.3
50	2.1	2.1	2.1	3.2	4.4	5.2	6.3	7.3
100	1.1	1.1	1.4	2.3	3.1	3.7	4.5	5.2
250	0.4	0.7	0.9	1.4	2.0	2.3	2.8	3.3
500	0.2	0.5	0.6	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	1.6
2,500	0.04	0.2	0.3	0.5	0.6	0.7	0.9	1.0
5,000	0.02	0.15	0.2	0.3	0.4	0.5	0.6	0.7
10,000	0.01	0.10	0.14	0.2	0.3	0.4	0.4	0.5
25,000	–	0.07	0.09	0.14	0.2	0.2	0.3	0.3
50,000	–	0.05	0.06	0.10	0.14	0.2	0.2	0.2
75,000	–	0.04	0.05	0.08	0.11	0.13	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 X. Standard Errors of Estimated Percentages of Lost Housing Units: 1973-1977 (Excluding Estimated Percentages of Lost Housing Units Pertaining to Lacking Bedroom, Lacking Kitchen Facilities, Lacking Some or All Plumbing Facilities, and Other Vacant)

(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	26.7	26.7	26.7	26.7	26.7	26.7	26.7	30.2
10	15.4	15.4	15.4	15.4	15.4	15.4	18.5	21.4
25	6.8	6.8	6.8	6.8	8.1	9.6	11.7	13.5
50	3.5	3.5	3.5	4.2	5.7	6.8	8.3	9.6
100	1.8	1.8	1.9	2.9	4.1	4.8	5.8	6.8
250	0.7	0.9	1.2	1.9	2.6	3.1	3.7	4.3
500	0.4	0.6	0.8	1.3	1.8	2.2	2.6	3.0
750	0.2	0.5	0.7	1.1	1.5	1.8	2.1	2.5
1,000	0.2	0.4	0.6	0.9	1.3	1.5	1.8	2.1
1,750	0.10	0.3	0.5	0.5	0.7	1.0	1.4	1.6
2,500	0.07	0.3	0.4	0.6	0.8	1.0	1.2	1.4

[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 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 2.8 in 1977. The base of the distribution, from which this median was determined, is 31,286,000 housing units.

1. From table VII, the standard error of a 50-percent characteristic on the base of 31,286,000 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 13,724,000, or 43.9 percent, owner-occupied housing units had 1 and 2 persons (actually, for purposes of calculating the median, the category of 2 persons is considered to be from 1.5 to 2.5 persons) and that an additional 10,490,000 owner-occupied housing units, or 18.4 percent, had 3 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:

$$2.5 + (3.5-2.5)\frac{49.8-43.9}{18.4} = 2.8$$

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

$$2.5 + (3.5-2.5)\frac{50.8-43.9}{18.4} = 2.9$$

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 lower limit, it actually is a reflection of the rounding error associated with the 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,

TABLE XI. Standard Errors of Estimated Percentages of Lost Housing Units Pertaining to Lacking Bedroom, Lacking Kitchen Facilities, Lacking Some or All Plumbing Facilities, and Other Vacant: 1973-1977

(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	31.3	31.3	31.3	31.3	31.3	31.3	31.3	33.8
10	18.6	18.6	18.6	18.6	18.6	18.6	20.7	23.9
25	8.4	8.4	8.4	8.4	9.1	10.8	13.1	15.1
50	4.4	4.4	4.4	4.7	6.4	7.6	9.3	10.7
100	2.2	2.2	2.2	3.3	4.5	5.4	6.5	7.6
250	0.9	1.0	1.3	2.1	2.9	3.4	4.1	4.8
500	0.5	0.7	0.9	1.5	2.0	2.4	2.9	3.4
750	0.3	0.5	0.8	1.2	1.7	2.0	2.4	2.8
1,000	0.2	0.5	0.7	1.0	1.4	1.7	2.1	2.4
1,750	0.13	0.4	0.5	0.8	1.1	1.3	1.6	1.8
2,500	0.09	0.3	0.4	0.7	0.9	1.1	1.3	1.5

[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 XIIa. Standard Errors of Estimated Percentages of Housing Units Pertaining to Each of Four Regions; Northeast, North Central, South, and West: 1977 (Excluding Estimated Percentages of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, New Construction, 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	28.7	28.7	28.7	28.7	28.7	28.7	28.7	31.8
10	16.8	16.8	16.8	16.8	16.8	16.8	19.4	22.5
25	7.5	7.5	7.5	7.5	8.5	10.1	12.3	14.2
50	3.9	3.9	3.9	4.4	6.0	7.2	8.7	10.0
100	2.0	2.0	2.0	3.1	4.3	5.1	6.2	7.1
250	0.8	0.9	1.3	2.0	2.7	3.2	3.9	4.5
500	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
1,000	0.2	0.4	0.6	1.0	1.3	1.6	1.9	2.2
2,500	0.08	0.3	0.4	0.6	0.9	1.0	1.2	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	0.4
50,000	—	0.06	0.09	0.14	0.2	0.2	0.3	0.3
75,000	—	0.05	0.07	0.11	0.2	0.2	0.2	0.3

Note: For standard errors of regional estimated percentages of seasonal and migratory vacant housing units and of year-round vacant housing units, use the national standard errors in tables VIII and IX, respectively, multiplied by 1.1.

[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.

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 the 1977 AHS national sample.

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 surveyed 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.

TABLE XIIb. Standard Errors of Estimated Percentages of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, New Construction, Source of Water, and Households with Head of Spanish Origin, for Each of the Four Regions; Northeast, North Central, South, and West: 1977

(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	36.9	36.9	36.9	36.9	36.9	36.9	36.9	38.2
10	22.6	22.6	22.6	22.6	22.6	22.6	23.4	27.0
25	10.5	10.5	10.5	10.5	10.5	12.2	14.8	17.1
50	5.5	5.5	5.5	5.5	7.2	8.6	10.5	12.1
100	2.8	2.8	2.8	3.7	5.1	6.1	7.4	8.5
250	1.2	1.2	1.5	2.4	3.2	3.9	4.7	5.4
500	0.6	0.8	1.1	1.7	2.3	2.7	3.3	3.8
1,000	0.3	0.5	0.8	1.2	1.6	1.9	2.3	2.7
2,500	0.12	0.3	0.5	0.7	1.0	1.2	1.5	1.7
5,000	0.06	0.2	0.3	0.5	0.7	0.9	1.0	1.2
10,000	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.9
25,000	0.01	0.11	0.2	0.2	0.3	0.4	0.5	0.5
50,000	0.01	0.08	0.11	0.2	0.2	0.3	0.3	0.4
75,000	–	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.

TABLE XIIc. Standard Errors of Estimated Percentages of Lost Housing Units for Each of the Four Regions; Northeast, North Central, South, and West: 1973-1977 (Excluding Estimated Percentages of Lost Housing Units Pertaining to Lacking Bedroom, Lacking Kitchen Facilities, Lacking Some or All Plumbing Facilities, and Other Vacant)

(68 chances out of 100)

Base of percentage (000)	Estimated percentages							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	29.2	29.2	29.2	29.2	29.2	29.2	29.2	32.1
10	17.1	17.1	17.1	17.1	17.1	17.1	19.7	22.7
25	7.6	7.6	7.6	7.6	8.6	10.3	12.4	14.4
50	4.0	4.0	4.0	4.4	6.1	7.3	8.8	10.2
100	2.0	2.0	2.0	3.1	4.3	5.1	6.2	7.2
250	0.8	0.9	1.3	2.0	2.7	3.2	3.9	4.5
500	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
750	0.3	0.5	0.7	1.1	1.6	1.9	2.3	2.6
1,000	0.2	0.5	0.6	1.0	1.4	1.6	2.0	2.3

Note: For standard errors of regional estimated percentages of lost housing units pertaining to these specified items, use the national standard errors in table XI.

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 1977 reinterview study were not available at the time of publication. However, it is expected that they will be similar to the results of the 1976 reinterview study which are presented in the Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample, 1976." Unlike the years prior to 1976, the respondent's answers in the reinterview were not reconciled to the original answers given in the AHS interview; i.e., after the question was answered in the reinterview, the interviewer did not present the previous responses and then ask the respondent to decide upon the best answer. Comparing the reinterview results of 1976 with the years prior to 1976, we found that the estimates of inconsistency of all items (nonattitudinal and attitudinal) increased substantially in the 1976 results. In other words, providing the interviewer with the original response had the effect of reducing the levels of inconsistency.

To summarize the results of the 1976 reinterview program: Overall, it showed moderate to high levels of inconsistency with about one-third of the nonattitudinal items and a high level of inconsistency with about one-third of the attitudinal items. Only one-eighth of the nonattitudinal items showed a low level of inconsistency.

The range for evaluating inconsistency is 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.

The following list shows the 1977 AHS-2 questionnaire numbers for those items asked in the 1976 reinterview: Nonattitudinal, sections IIIA and IIIB, 12*, 13*, 15a, 15b, 15c*, 36a, 36b*, 37a, 37b*, 49*, 50*, 51a, 51b, 52a*, 54a*, 54b, 55a, 55b, 55c*, 56a, 56b, 58a*, 58b*, 61a*; attitudinal, section IIIB, 102 all parts*, 103 all parts*, and 104 all parts*. Asterisks (*) indicate that the item had an estimated index of inconsistency greater than 40. Since cross tabulations involving these items may be subject to a large distortion because of the moderate to high response variance, they should be considered to be less reliable than comparable cross tabulations which do not involve these items. The cutoff at 40 was selected because (1) the shape of the distribution had a natural break before 40, (2) the large sampling errors on the estimated indices indicated little difference between those indices from 40 to 50 and those greater than 50, and (3) the break between moderate and high indices at 50 is arbitrary.

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 respondents 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 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 1977 AHS sample missed about 3.5 percent (i.e., about 375,000 units) of all conventional new construction built after April 1970, because the permits for these units, which were built before October 1977, 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 1977 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 were 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. However, biases of subtotals would still remain.

Rounding errors.—In 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. Also, since medians in this report were computed using unrounded data, instead of the shown published rounded data, they can differ from medians calculated directly from the published data.

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-44
- Selection of sample areas App-44
- Designation of sample housing units for the 1977 survey App-44
- Selection of the 1973 sample housing units App-44
- Splitting of the sample App-45
- Selection of supplemental sample housing units in rural areas App-45
- Selection of sample housing units for the 1976 Coverage Improvement Program App-45

1977 REDUCTION App-46

ESTIMATION App-46

RELIABILITY OF THE ESTIMATES App-47
- Sampling errors App-47
- Standard errors of estimates of levels App-47
- Standard errors of estimates of percentages App-47
- Standard errors of ratios App-48
  - Illustration of the use of the standard error tables. Illustration I App-49
  - Illustration II App-50
  - Differences App-51
  - Illustration of the computation of the standard error of a difference App-51
  - Medians App-52
  - Illustration of the computation of the 95-percent confidence interval for a median. App-52
- Nonsampling errors App-52
- Reinterview program App-52
- Coverage errors App-53
- Rounding errors App-53

SAMPLE DESIGN

The 1977 estimates are based on data collected in October 1977 through January 1978 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 70,600 sample housing units (both occupied and vacant) were eligible for interview in the 1977 Annual Housing Survey. Of this number, 4,000 interviews were classified as "noninterview" for various reasons. Occupied housing units were classified as "noninterview," mainly, because the occupants refused to be interviewed 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 70,600, there were also 6,300 sample units which were visited but found not to be eligible for interview for the AHS in terms of collecting information relevant to the 1977 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 1977 survey.—The sample housing units designated to be interviewed in the 1977 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1976 survey (which included all sample housing units that were selected as part of the 1976 Coverage Improvement Program) and that were not part of the 1977 reduction.
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 1976 survey and that were not part of the 1977 reduction. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1977 AHS questionnaire, page 1.)
3. All sample housing units that were selected from the list of building permits issued since the 1976 survey. (This sample represented the housing units built in permit-issuing areas since the 1976 survey.)

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. 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 enumeration 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 at the 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 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 have 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.

1977 REDUCTION

By 1977, the addition to the sample from primarily new construction and the coverage improvements had increased the total sample size (interviews plus noninterviews) to about 81,000. The sample was reduced by about 7 percent to approximately 75,000. However, this reduction did not include any CEN-SUP[1] units or units which were selected as part of the 1976 Coverage Improvement Program. Thus, the overall probability of selection for these latter units remained unchanged, and, for the rest of the units, the probability of selection was about 1 in 1,472 if they were urban and about 1 in 736 if they were rural.

ESTIMATION

In 1977, 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:

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

[1] CEN-SUP units resulted from a 1970 census evaluation study and represented units missed in the 1970 census.

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 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- and third-stage ratio estimation procedures were iterated 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 15 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 14 categories of new construction units for each of the 4 regions (i.e., 9 categories for conventional new construction units and 5 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 variability between estimates from all possible samples is defined as sampling error. One common measure of sampling error is the standard error which measures the precision with which an estimate from a sample approximates the average results of all possible samples. In addition, the standard error, as calculated for this report, partially reflects the variation in the estimates due to some 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 sample estimate and its estimated standard error enable one to construct interval estimates so that the interval includes the average result of all possible samples with a known probability. 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 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.

Standard errors of estimates of levels.—Tables I, II, and III present the standard errors applicable to the 1977 national housing inventory estimates in this report. Tables IV and V present the standard errors for each of the four regions; Northeast, North Central, South, and West. Linear interpolation should be used to determine standard errors for levels of estimates not specifically shown in tables I through V.

Standard errors of estimates of percentages.—The reliability of an estimated percentage, computed by using

TABLE I. Standard Errors of Estimated Numbers of Housing Units: 1977 (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: Total or White (000)	Standard error: Black (000)	Size of estimate (000)	Standard error: Total or White (000)	Standard error: Black (000)
0	2	2	1,000	40	38
5	3	3	2,500	63	53
10	4	4	5,000	88	55
25	6	6	10,000	120	—
50	9	9	25,000	168	—
100	13	13	50,000	177	—
250	20	20	75,000	95	—
500	29	28			

TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to Lacking Complete Plumbing Facilities and Households with Head of Spanish Origin: 1977

(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	48	45
5	3	3	2,500	76	64
10	5	5	5,000	105	66
25	8	8	10,000	144	—
50	11	11	25,000	202	—
100	15	15	50,000	213	—
250	24	24	75,000	114	—
500	34	33			

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, VII, VIII, IX, and X present the standard errors of estimated percentages. Table VI shows the approximate standard errors of all national estimated percentages of housing units except those pertaining to the specified items in tables II and III. The standard errors shown in tables VII and VIII should be used for those specified items. Table IX shows the approximate standard errors of all regional percentages of housing units except those pertaining to the specified items in table V. Table X should be used for those specified items for each of the four regions; Northeast, North Central, South, and West. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables VI through X.

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 to be 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.

Standard errors of ratios.—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

TABLE III. Standard Errors of Estimated Numbers of Year-Round Vacant Housing Units: 1977

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	1	250	17
5	2	500	24
10	3	1,000	35
25	5	2,500	59
50	7	3,500	74
100	10		

TABLE IV. Standard Errors of Estimated Numbers of Housing Units Pertaining to Each of the Four Regions; Northeast, North Central, South, and West: 1977 (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 (000)	Size of estimate (000)	Standard error (000)
0	2	1,000	45
5	3	2,500	70
10	4	5,000	97
25	7	10,000	133
50	10	25,000	187
100	14	50,000	196
250	22	75,000	105
500	32		

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 13,710,000 renter-occupied housing units with common stairways. Interpolation of the data in table I shows that the standard error of an estimate of this size is approximately 132,000. 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	120
13,710	x
25,000	168

By vertically interpolating between 120 and 168, the entry for "x" is determined to be:

$$13{,}710-10{,}000 = 3{,}710$$
$$25{,}000-10{,}000 = 15{,}000$$

$$120 + \frac{3{,}710}{15{,}000}(168-120) = 132$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 13,578,000 to 13,842,000 housing units. Therefore, a conclusion that the average estimate of 1977 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 13,499,000 to 13,921,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 13,446,000 to 13,974,000 housing units with 95 percent confidence.

Table A-2 also shows that of the 13,710,000 renter-occupied housing units with common stairways, 12,217,000, or 89.1 percent, were located inside SMSA's. Interpolation of the data in table VI (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 VI. The entry for "p" is the one sought.

Base of percentage (000)	Estimated percentage		
	15 or 85	89.1	10 or 90
10,000 ...	0.5	a	0.
13,710 ...		p	
25,000 ...	0.3	b	0.

TABLE V. Standard Errors of Estimated Numbers of Housing Units Pertaining to Lacking Complete Plumbing Facilities and Households with Head of Spanish Origin, for Each of the Four Regions; Northeast, North Central, South, and West: 1977

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	1,000 ..	54
5	4	2,500 ..	84
10	5	5,000 ..	117
25	9	10,000 .	160
50	12	25,000 .	225
100	17	50,000 .	236
250	27	75,000 .	126
500	38		

TABLE VI. Standard Errors of Estimated Percentages of Housing Units: 1977 (Excluding Estimate 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	24.6	24.6	24.6	24.6	24.6	24.6	24.8	28.6
10	14.0	14.0	14.0	14.0	14.0	14.0	17.5	20.2
25	6.1	6.1	6.1	6.1	7.7	9.1	11.1	12.8
50	3.2	3.2	3.2	3.9	5.4	6.5	7.8	9.0
100	1.6	1.6	1.8	2.8	3.8	4.6	5.5	6.4
250	0.6	0.8	1.1	1.8	2.4	2.9	3.5	4.0
500	0.3	0.6	0.8	1.2	1.7	2.0	2.5	2.9
1,000	0.2	0.4	0.6	0.9	1.2	1.4	1.8	2.0
2,500	0.07	0.3	0.4	0.6	0.8	0.9	1.1	1.3
5,000	0.03	0.2	0.3	0.4	0.5	0.6	0.8	0.9
10,000 ...	0.02	0.13	0.2	0.3	0.4	0.5	0.6	0.6
25,000 ...	0.01	0.08	0.11	0.2	0.2	0.3	0.4	0.4
50,000 ...	–	0.06	0.08	0.12	0.2	0.2	0.2	0.3
75,000 ...	–	0.05	0.07	0.10	0.14	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.

1. By horizontal interpolation between 0.5 and 0.4, the entry for cell "a" is determined to be 0.4.

$$89.1-85.0 = 4.1$$
$$90.0-85.0 = 5.0$$

$$0.5+\frac{4.1}{5.0}(0.4-0.5)= 0.4$$

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

$$89.1-85.0 = 4.1$$
$$90.0-85.0 = 5.0$$

$$0.3+\frac{4.1}{5.0}(0.2-0.3) = 0.2$$

3. By vertical interpolation between 0.4 and 0.2, the entry for "p" is determined to be 0.4.

$$13{,}710-10{,}000 = 3{,}710$$
$$25{,}000-10{,}000 = 15{,}000$$

$$0.4+\frac{3{,}710}{15{,}000}(0.2-0.4) = 0.4$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 88.7 to 89.5 percent; the 90-percent confidence interval is from 88.5 to 89.7 percent; and the 95-percent confidence interval is from 88.3 to 89.9 percent.

Illustration II.—Table A-3 of this report shows that in the United States in 1977 there were 562,000 owner-occupied housing units which had sewage disposal breakdowns. Interpolation of the data in table I of this appendix shows that the standard error of an estimate of this size is approximately 30,000. Consequently, the 68-percent confidence interval is from 532,000 to 592,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1977 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 514,000 to 610,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 502,000 to 622,000 housing units with 95 percent confidence.

Table A-3 also shows that of the 562,000, 1977 owner-occupied housing

TABLE VII. Standard Errors of Estimated Percentages of Housing Units Pertaining to Lacking Complete Plumbing Facilities and Households with Head of Spanish Origin: 1977

(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	32.2	32.2	32.2	32.2	32.2	32.2	32.2	34.4
10	19.2	19.2	19.2	19.2	19.2	19.2	21.1	24.3
25	8.7	8.7	8.7	8.7	9.2	11.0	13.3	15.4
50	4.5	4.5	4.5	4.7	6.5	7.8	9.4	10.9
100	2.3	2.3	2.3	3.4	4.6	5.5	6.7	7.7
250	0.9	1.0	1.4	2.1	2.9	3.5	4.2	4.9
500	0.5	0.7	1.0	1.5	2.1	2.5	3.0	3.4
1,000	0.2	0.5	0.7	1.1	1.5	1.7	2.1	2.4
2,500	0.09	0.3	0.4	0.7	0.9	1.1	1.3	1.5
5,000	0.05	0.2	0.3	0.5	0.7	0.8	0.9	1.1
10,000	0.02	0.2	0.2	0.3	0.5	0.5	0.7	0.8
25,000	0.01	0.10	0.14	0.2	0.3	0.3	0.4	0.5
50,000	–	0.07	0.10	0.2	0.2	0.2	0.3	0.3
75,000	–	0.06	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.

TABLE VIII. Standard Errors of Estimated Percentages of Year-Round Vacant Housing Units: 1977

(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	17.6	17.6	17.6	17.6	17.6	17.6	20.0	23.1
10	9.7	9.7	9.7	9.7	9.8	11.7	14.2	16.3
25	4.1	4.1	4.1	4.5	6.2	7.4	9.0	10.3
50	2.1	2.1	2.1	3.2	4.4	5.2	6.3	7.3
100	1.1	1.1	1.4	2.3	3.1	3.7	4.5	5.2
250	0.4	0.7	0.9	1.4	2.0	2.3	2.8	3.3
500	0.2	0.5	0.6	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	1.6
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	0.14	0.2	0.3	0.4	0.4	0.5
25,000	–	0.07	0.09	0.14	0.2	0.2	0.3	0.3
50,000	–	0.05	0.06	0.10	0.14	0.2	0.2	0.2
75,000	–	0.04	0.05	0.08	0.11	0.13	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.

units which had sewage disposal breakdowns, 55,000, or 9.8 percent, had breakdowns 3 times or more. Interpolation of the data in table VI (i.e., interpolation on both the base and the percent) shows that the standard error of the above percentage is 1.6 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 8.2 to 11.4 percent; the 90-percent confidence interval is from 7.2 to 12.4 percent; and the 95-percent confidence interval is from 6.6 to 13.0 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 characteristics in two different areas or the difference between separate and uncorrelated characteristics in the same area. If 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 shows that in the United States in 1977 there were 440,000 owner-occupied housing units which had exactly one sewage disposal breakdown. Thus, the apparent difference between the number of the 1977 owner-occupied housing units that had breakdowns 3 times or more and those that had breakdowns just 1 time is 385,000. Interpolation of the data in table I shows that the standard error on an estimate of 55,000 to be approximately 9,000 and the standard error on an estimate of 440,000 to be approximately 27,000. Therefore, the standard error of the estimated difference of 385,000 is about 28,000.

$$28{,}000 = \sqrt{(9{,}000)^2 + (27{,}000)^2}$$

TABLE IX. Standard Errors of Estimated Percentages of Housing Units Pertaining to Each of the Four Regions; Northeast, North Central, South, and West: 1977 (Excluding 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	28.7	28.7	28.7	28.7	28.7	28.7	28.7	31.8
10	16.8	16.8	16.8	16.8	16.8	16.8	19.4	22.5
25	7.5	7.5	7.5	7.5	8.5	10.1	12.3	14.2
50	3.9	3.9	3.9	4.4	6.0	7.2	8.7	10.0
100	2.0	2.0	2.0	3.1	4.3	5.1	6.2	7.1
250	0.8	0.9	1.3	2.0	2.7	3.2	3.9	4.5
500	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
1,000	0.2	0.4	0.6	1.0	1.3	1.6	1.9	2.2
2,500	0.08	0.3	0.4	0.6	0.9	1.0	1.2	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	0.4
50,000	–	0.06	0.09	0.14	0.2	0.2	0.3	0.3
75,000	–	0.05	0.07	0.11	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.

TABLE X. Standard Errors of Estimated Percentages of Housing Units Pertaining to Lacking Complete Plumbing Facilities and Households with Head of Spanish Origin, for Each of the Four Regions; Northeast, North Central, South, and West: 1977

(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	36.9	36.9	36.9	36.9	36.9	36.9	36.9	38.2
10	22.6	22.6	22.6	22.6	22.6	22.6	23.4	27.0
25	10.5	10.5	10.5	10.5	10.5	12.2	14.8	17.1
50	5.5	5.5	5.5	5.5	7.2	8.6	10.5	12.1
100	2.8	2.8	2.8	3.7	5.1	6.1	7.4	8.5
250	1.2	1.2	1.5	2.4	3.2	3.9	4.7	5.4
500	0.6	0.8	1.1	1.7	2.3	2.7	3.3	3.8
1,000	0.3	0.5	0.8	1.2	1.6	1.9	2.3	2.7
2,500	0.12	0.3	0.5	0.7	1.0	1.2	1.5	1.7
5,000	0.06	0.2	0.3	0.5	0.7	0.9	1.0	1.2
10,000	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.9
25,000	0.01	0.11	0.2	0.2	0.3	0.4	0.5	0.5
50,000	0.01	0.08	0.11	0.2	0.2	0.3	0.3	0.4
75,000	–	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.

Consequently, the 68-percent confidence interval for the 385,000 difference is from 357,000 to 413,000 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 340,000 to 430,000 housing units, and the 95-percent confidence interval is from 329,000 to 441,000. Thus, we can conclude with 95 percent confidence that the number of 1977 owner-occupied housing units, which had three or more sewage disposal breakdowns, is different from 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.

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 the 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-13 of this report shows the median asking price for specified vacant for sale housing units was $32,100 in 1977. The base of the distribution, from which this median was determined, is 445,000 housing units.

1. From table VIII, the standard error of a 50-percent characteristic on the base of 445,000 is 2.5 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 45.0 and 55.0.
3. From table A-13, it can be seen by cumulating the frequencies for the first 4 categories that 160,000 specified vacant for sale housing units, or 36.0 percent, were priced less than or equal to $24,999, and that an additional 88,000 specified vacant for sale housing units, or 19.8 percent, had a sale price of $25,000 to $34,999. By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$25{,}000 + (35{,}000-25{,}000)\,\frac{(45.0-36.0)}{19.8} = 29{,}500$$

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

$$25{,}000 + (35{,}000-25{,}000)\,\frac{(55.0-36.0)}{19.8} = 34{,}600$$

Thus, the 95-percent confidence interval ranges from $29,500 to $34,600.

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 the 1977 AHS national sample.

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 1977 reinterview study were not available at the time of publication. However, it is expected that they will be similar to the results of the 1976 reinterview study which are presented in the Census Bureau memo-

randum, "Reinterview Results for the Annual Housing Survey—National Sample, 1976." Unlike the years prior to 1976, the respondent's answers in the reinterview were not reconciled to the original answers given in the AHS interview; i.e., after the question was answered in the reinterview, the interviewer did not present the previous responses and then ask the respondent to decide upon the best answer. Comparing the reinterview results of 1976 with the years prior to 1976, we found that the estimates of inconsistency of all items (nonattitudinal and attitudinal) increased substantially in the 1976 results. In other words, providing the interviewer with the original response had the effect of reducing the levels of inconsistency.

To summarize the results of the 1976 reinterview program: Overall, it showed moderate to high levels of inconsistency with about one-third of the nonattitudinal items and a high level of inconsistency with about one-third of the attitudinal items. Only one-eighth of the nonattitudinal items showed a low level of inconsistency.

The range for evaluating inconsistency is 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.

The following list shows the 1977 AHS-2 questionnaire numbers for those items asked in the 1976 reinterview: Nonattitudinal, sections IIIA and IIIB, 12*, 13*, 15a, 15b, 15c*, 36a, 36b*, 37a, 37b*, 49*, 50*, 51a, 51b, 52a*, 54a*, 54b, 55a, 55b, 55c*, 56a, 56b, 58a*, 58b*, 61a*; Attitudinal, section IIIB, 102 all parts*, 103 all parts*, and 104 all parts*. Asterisks (*) indicate that the item had an estimated index of inconsistency greater than 40. Since cross tabulations involving these items may be subject to a large distortion because of the moderate to high response variance, they should be considered to be less reliable than comparable cross tabulations which do not involve these items. The cutoff at 40 was selected because (1) the shape of the distribution had a natural break before 40, (2) the large sampling errors on the estimated indices indicated little difference between those indices from 40 to 50 and those greater than 50, and (3) the break between moderate and high indices at 50 is arbitrary.

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 respondents 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 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 1977 AHS sample missed about 3.5 percent (i.e., about 375,000 units) of all conventional new construction built after April 1970, because the permits, for these units, which were built before October 1977, 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 1977 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 were 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. However, biases of subtotals would still remain.

Rounding errors.—In 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 or 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. Also, since medians in this report were computed using unrounded data, instead of the shown published rounded data, they can differ from medians calculated directly from the published data.

☆U.S. GOVERNMENT PRINTING OFFICE: 1979-281-049/1069

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-44
- Selection of sample areas App-44
- Designation of sample housing units for the 1977 survey App-44
- Selection of the 1973 sample housing units App-44
- Splitting of the sample App-45
- Selection of supplemental sample housing units in rural areas App-45
- Selection of sample housing units for the 1976 Coverage Improvement Program App-45

1977 REDUCTION App-46

ESTIMATION App-46

RELIABILITY OF THE ESTIMATES App-47
- Sampling errors App-47
- Standard errors of estimates of levels App-47
- Standard errors of estimates of percentages App-48
- Standard errors of ratios App-49
  - Illustration of the use of the standard error tables. Illustration I App-49
  - Illustration II App-50
  - Differences App-51
  - Illustration of the computation of the standard error of a difference App-51
  - Medians App-51
  - Illustration of the computation of the 95-percent confidence interval for a median. App-52
- Nonsampling errors App-52
- Reinterview program App-52
- Coverage errors App-53
- Rounding errors App-53

SAMPLE DESIGN

The 1977 estimates are based on data collected in October 1977 through January 1978 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 70,600 sample housing units (both occupied and vacant) were eligible for interview in the 1977 Annual Housing Survey. Of this number, 4,000 interviews were classified as "noninterview," for various reasons. Occupied housing units were classified as "noninterview," mainly, because the occupants refused to be interviewed 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 70,600, there were also 6,300 sample units which were visited but found not to be eligible for interview for the AHS in terms of collecting information relevant to the 1977 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 1977 survey.—The sample housing units designated to be interviewed in the 1977 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1976 survey (which included all sample housing units that were selected as part of the 1976 Coverage Improvement Program) and that were not part of the 1977 reduction.
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 1976 survey and that were not part of the 1977 reduction. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1977 AHS questionnaire, page 1.)
3. All sample housing units that were selected from the list of building permits issued since the 1976 survey. (This sample represented the housing units built in permit-issuing areas since the 1976 survey.)

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. 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 enumeration 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 at the 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 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 have 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.

1977 REDUCTION

By 1977, the addition to the sample from primarily new construction and the coverage improvements had increased the total sample size (interviews plus noninterviews) to about 81,000. The sample was reduced by about 7 percent to approximately 75,000. However, this reduction did not include any CEN-SUP[1] units or units which were selected as part of the 1976 Coverage Improvement Program. Thus, the overall probability of selection for these latter units remained unchanged, and, for the rest of the units, the probability of selection was about 1 in 1,472 if they were urban and about 1 in 736 if they were rural.

ESTIMATION

In 1977, 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 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 Sur-

[1] CEN-SUP units resulted from a 1970 census evaluation study and represented units missed in the 1970 census.

vey (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 iterated 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 15 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 14 categories of new construction units for each of the 4 regions (i.e., 9 categories for conventional new construction units and 5 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 variability between estimates from all possible samples is defined as sampling error. One common measure of sampling error is the standard error which measures the precision with which an estimate from a sample approximates the average results of all possible samples. In addition, the standard error, as calculated for this report, partially reflects the variation in the estimates due to some 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 sample estimate and its estimated standard error enable one to construct interval estimates so that the interval includes the average result of all possible samples with a known probability. 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 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.

Standard errors of estimates of levels.—Tables I and II present the standard errors applicable to the 1977 national housing inventory estimates in this report. Tables III and IV present the standard errors for each of the four regions; Northeast, North Central, South, and West. Linear interpolation should be used to determine standard errors for levels of estimates not specifically shown in tables I through IV.

TABLE I. Standard Errors of Estimated Numbers of Housing Units: 1977 (Excluding Estimates of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, 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	2	2	1,000	40	38
5	3	3	2,500	63	53
10	4	4	5.000	88	55
25	6	6	10,000	120	—
50	9	9	25,000	168	—
100	13	13	50,000	177	—
250	20	20	75,000	95	—
500	29	28			

TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, Source of Water, and Households with Head of Spanish Origin: 1977

(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	48	45
5	3	3	2,500	76	64
10	5	5	5,000	105	66
25	8	8	10,000	144	—
50	11	11	25,000	202	—
100	15	15	50,000	213	—
250	24	24	75,000	114	—
500	34	33			

Standard errors of estimates of percentages.—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, VI, VII, and VIII present the standard errors of estimated percentages. Table V shows the approximate standard errors of all national estimated percentages of housing units except those pertaining to the specified items in table II. The standard errors shown in table VI should be used for those specified items. Table VII shows the approximate standard errors of all regional percentages of housing units except those pertaining to the specified items in table II. Table VIII should be used for those specified items for each of the four regions; Northeast, North Central, South, and West. Two-way 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

TABLE III. Standard Errors of Estimated Numbers of Housing Units Pertaining to Each of the Regions; Northeast, North Central, South, and West: 1977 (Excluding Estimates of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, and Source of Water)

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	1,000	45
5	3	2,500	70
10	4	5,000	97
25	7	10,000	133
50	10	25,000	187
100	14	50,000	196
250	22	75,000	105
500	32		

TABLE IV. Standard Errors of Estimated Numbers of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, and Source of Water, for Each of the Regions; Northeast, North Central, South and West: 1977

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	1,000	54
5	4	2,500	84
10	5	5,000	117
25	9	10,000	160
50	12	25,000	225
100	17	50,000	236
250	27	75,000	126
500	38		

of zero and zero percent. These estimates of standard errors are considered to be 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.

Standard errors of ratios.—For ratios of the form (100) (x/y) where x is not a subclass of y, tables V 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 9,187,000 specified owner-occupied housing units with 2 bedrooms in 1977. Interpolation of the data in table I shows that the standard error of an estimate of this size is approximately 115,000. 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	88
9,187	x
10,000	120

By vertically interpolating between 88 and 120, the entry for "x" is determined to be 115.

TABLE V. Standard Errors of Estimated Percentages of Housing Units: 1977 (Excludes Estimated Percentages of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, 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	24.6	24.6	24.6	24.6	24.6	24.6	24.8	28.6
10	14.0	14.0	14.0	14.0	14.0	14.0	17.5	20.2
25	6.1	6.1	6.1	6.1	7.7	9.1	11.1	12.8
50	3.2	3.2	3.2	3.9	5.4	6.5	7.8	9.0
100	1.6	1.6	1.8	2.8	3.8	4.6	5.5	6.4
250	0.6	0.8	1.1	1.8	2.4	2.9	3.5	4.0
500	0.3	0.6	0.8	1.2	1.7	2.0	2.5	2.9
1,000	0.2	0.4	0.6	0.9	1.2	1.4	1.8	2.0
2,500	0.07	0.3	0.4	0.6	0.8	0.9	1.1	1.3
5,000	0.03	0.2	0.3	0.4	0.5	0.6	0.8	0.9
10,000 ...	0.02	0.13	0.2	0.3	0.4	0.5	0.6	0.6
25,000 ...	0.01	0.08	0.11	0.2	0.2	0.3	0.4	0.4
50,000 ...	–	0.06	0.08	0.12	0.2	0.2	0.2	0.3
75,000 ...	–	0.05	0.07	0.10	0.14	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 Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, Source of Water, and Households with Head of Spanish Origin: 1977

(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	32.2	32.2	32.2	32.2	32.2	32.2	32.2	34.4
10	19.2	19.2	19.2	19.2	19.2	19.2	21.1	24.3
25	8.7	8.7	8.7	8.7	9.2	11.0	13.3	15.4
50	4.5	4.5	4.5	4.7	6.5	7.8	9.4	10.9
100	2.3	2.3	2.3	3.4	4.6	5.5	6.7	7.7
250	0.9	1.0	1.4	2.1	2.9	3.5	4.2	4.9
500	0.5	0.7	1.0	1.5	2.1	2.5	3.0	3.4
1,000	0.2	0.5	0.7	1.1	1.5	1.7	2.1	2.4
2,500	0.09	0.3	0.4	0.7	0.9	1.1	1.3	1.5
5,000	0.05	0.2	0.3	0.5	0.7	0.8	0.9	1.1
10,000 ...	0.02	0.2	0.2	0.3	0.5	0.5	0.7	0.8
25,000 ...	0.01	0.10	0.14	0.2	0.3	0.3	0.4	0.5
50,000 ...	–	0.07	0.10	0.2	0.2	0.2	0.3	0.3
75,000 ...	–	0.06	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.

$$9{,}187-5{,}000 = 4{,}187$$
$$10{,}000-5{,}000 = 5{,}000$$

$$88 + \frac{4{,}187}{5{,}000}(120-88) = 115$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 9,072,000 to 9,302,000 housing units. Therefore, a conclusion that the average estimate of 1977 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 9,003,000 to 9,371,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 8,957,000 to 9,417,000 housing units with 95 percent confidence.

Table A-2 also shows that of the 9,187,000 specified owner-occupied housing units with 2 bedrooms, 1,229,000, or 13.4 percent, were valued between $15,000 and $19,999. Interpolation of the data in table V (i.e., interpolation on both the base and percent) shows that the standard error of the above percentage is 0.5 percentage points. The following procedure was used in interpolating:

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

Base of percentage (000)	Estimated percentage		
	10	13.4	15
5,000	0.5	a	0.6
9,187		p	
10,000 ...	0.4	b	0.5

1. By horitontal interpolation between 0.5 and 0.6, the entry for cell "a" is determined to be 0.6.

$$13.4-10.0 = 3.4$$
$$15.0-10.0 = 5.0$$

$$0.5 + \frac{3.4}{5.0}(0.6-0.5) = 0.6$$

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

$$13.4-10.0 = 3.4$$
$$15.0-10.0 = 5.0$$

$$0.4 + \frac{3.4}{5.0}(0.5-0.4) = 0.5$$

3. By vertical interpolation between 0.6 and 0.5, the entry for "p" is determined to be 0.5.

$$9{,}187-5{,}000 = 4{,}187$$
$$10{,}000-5{,}000 = 5{,}000$$

$$0.6 + \frac{4{,}187}{5{,}000}(0.5-0.6) = 0.5$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 12.9 to 13.9 percent; the 90-percent confidence interval is from 12.6 to 14.2 percent; and the 95-percent confidence interval is from 12.4 to 14.4 percent.

Illustration II.—Table A-2 of this report shows that in the United States in 1977, there were 55,000 specified owner-occupied housing units whose cooking fuel was wood. Interpolation of the data in table II shows that the standard error of an estimate of this size is approximately 11,000. Consequently, the 68-percent confidence interval is from 44,000 to 66,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1977 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 73,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 33,000 to 77,000 housing units with 95 percent confidence.

TABLE VII. Standard Errors of Estimated Percentages of Housing Units Pertaining to Each of the Four Regions; Northeast, North Central, South, and West: 1977 (Excluding Estimated Percentages of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, and Source of Water)

(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	28.7	28.7	28.7	28.7	28.7	28.7	28.7	31.8
10	16.8	16.8	16.8	16.8	16.8	16.8	19.4	22.5
25	7.5	7.5	7.5	7.5	8.5	10.1	12.3	14.2
50	3.9	3.9	3.9	4.4	6.0	7.2	8.7	10.0
100	2.0	2.0	2.0	3.1	4.3	5.1	6.2	7.1
250	0.8	0.9	1.3	2.0	2.7	3.2	3.9	4.5
500	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
1,000	0.2	0.4	0.6	1.0	1.3	1.6	1.9	2.2
2,500	0.08	0.3	0.4	0.6	0.9	1.0	1.2	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	0.4
50,000 ...	—	0.06	0.09	0.14	0.2	0.2	0.3	0.3
75,000 ...	—	0.05	0.07	0.11	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.

TABLE VIII. Standard Errors of Estimated Percentages of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, and Source of Water, for Each of the Four Regions; Northeast, North Central, South, and West: 1977

(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	36.9	36.9	36.9	36.9	36.9	36.9	36.9	38.2
10	22.6	22.6	22.6	22.6	22.6	22.6	23.4	27.0
25	10.5	10.5	10.5	10.5	10.5	12.2	14.8	17.1
50	5.5	5.5	5.5	5.5	7.2	8.6	10.5	12.1
100	2.8	2.8	2.8	3.7	5.1	6.1	7.4	8.5
250	1.2	1.2	1.5	2.4	3.2	3.9	4.7	5.4
500	0.6	0.8	1.1	1.7	2.3	2.7	3.3	3.8
1,000	0.3	0.5	0.8	1.2	1.6	1.9	2.3	2.7
2,500	0.12	0.3	0.5	0.7	1.0	1.2	1.5	1.7
5,000	0.06	0.2	0.3	0.5	0.7	0.9	1.0	1.2
10,000	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.9
25,000	0.01	0.1	0.2	0.2	0.3	0.4	0.5	0.5
50,000	0.01	0.08	0.11	0.2	0.2	0.3	0.3	0.4
75,000	—	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.

Table A-2 also shows that of the 55,000 specified owner-occupied housing units whose cooking fuel is wood, 17,000, or 30.9 percent, were valued at less than $10,000. Interpolation of the data in table VI (i.e., interpolation on both the base and the percent) shows that the standard error of the above percentage is 9.5 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 21.4 to 40.4 percent; the 90-percent confidence interval is from 15.7 to 46.1 percent; and the 95-percent confidence interval is from 11.9 to 49.9 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristics in two different areas or the difference between separate and uncorrelated characteristics in the same area. If 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 1977 there were 882,000 specified owner-occupied housing units with two bedrooms valued between $10,000 and $14,999. Thus, the apparent difference between the number of 1977 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 347,000. Interpolation of the data in table I shows that the standard error on an estimate of 882,000 to be approximately 37,000 and the standard error on an estimate 1,229,000 to be approximately 44,000. Therefore, the standard error of the estimated difference of 347,000 is about 57,000.

$$57{,}000 = \sqrt{(37{,}000)^2 + (44{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 347,000 difference is from 290,000 to 404,000 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 256,000 to 438,000 housing units, and the 95-percent confidence interval is from 233,000 to 461,000. Thus, we can conclude with 95 percent confidence that the number of 1977 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 in the United States was $27,500 in 1977. The base of the distribution, from which this median was determined, is 9,187,000 housing units.

1. From table V, the standard error of a 50-percent characteristic on the base of 9,187,000 is 0.6 percentage points.

2. To obtain a two-standard-error confidence interval on the estimated median, add to an subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 48.8 and 51.2.

3. From table A-2, it can be seen by cumulating the frequencies for the first 4 categories that 4,009,000 owner-occupied housing units with 2 bedrooms, or 43.6 percent, had a value less than $25,000 and an additional 1,160,000 owner-occupied housing units with 2 bedrooms, or 12.6 percent, had a value between $25,000 and $30,000. By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$\$25{,}000 + (30{,}000-25{,}000)\frac{(48.8-43.6)}{12.6}$$

$$= \$27{,}100$$

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

$$\$25{,}000 + (30{,}000-25{,}000)\frac{(51.2-43.6)}{12.6}$$

$$= \$28{,}000$$

Thus, the 95-percent confidence interval ranges from $27,100 to $28,000.

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 the 1977 AHS national sample.

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 1977 reinterview study were not available at the time of publication. However, it is expected that they will be similar to the results of the 1976 reinterview study which are presented in the Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample, 1976." Unlike the years prior to 1976, the respondent's answers in the reinterview were not reconciled to the original answers given in the AHS interview; i.e., after the question was answered in the reinterview, the interviewer did not present the previous responses and then ask the respondent to decide upon the best answer. Comparing the reinterview results of 1976 with the years prior to 1976, we found that the estimates of inconsistency of all items (nonattitudinal and attitudinal) increased substantially in the 1976 results. In other words, providing the interviewer with the original response had the effect of reducing the levels of inconsistency.

To summarize the results of the 1976 reinterview program: Overall, it showed moderate to high levels of inconsistency with about one-third of the nonattitudinal items and a high level of inconsistency with about one-third of the attitudinal items. Only one-eighth of the nonattitudinal items showed a low level of inconsistency.

The range for evaluating inconsistency is 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.

The following list shows the 1977 AHS-2 questionnaire numbers for those items asked in the 1976 reinterview: Nonattitudinal, sections IIIA and IIIB, 12*, 13*, 15a, 15b, 15c*, 36a, 36b*, 37a, 37b*, 49*, 50*, 51a, 51b, 52a*,

54a*, 54b, 55a, 55b, 55c*, 56a, 56b, 58a*, 58b*, 61a*; Attitudinal, section IIIB, 102 all parts*, 103 all parts*, and 104 all parts*. Asterisks (*) indicate that the item had an estimated index of inconsistency greater than 40. Since cross tabulations involving these items may be subject to a large distortion because of the moderate to high response variance, they should be considered to be less reliable than comparable cross tabulations which do not involve these items. The cutoff at 40 was selected because (1) the shape of the distribution had a natural break before 40, (2) the large sampling errors on the estimated indices indicated little difference between those indices from 40 to 50 and those greater than 50, and (3) the break between moderate and high indices at 50 is arbitrary.

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 respondents 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 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 1977 AHS sample missed about 3.5 percent (i.e., about 375,000 housing units) of all conventional new construction built after April 1970, because the permits for these units, which were built before October 1977, 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 1977 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 were 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. However, biases of subtotals would still remain.

Rounding errors.—In 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. Also, since medians in this report were computed using unrounded data, instead of the shown published rounded data, they can differ from medians calculated directly from the published data.

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-44
Selection of sample areas App-44
Designation of sample housing units for the 1977 survey App-44
Selection of the 1973 sample housing units App-44
Splitting of the sample App-45
Selection of supplemental sample housing units in rural areas App-45
Selection of sample housing units for the 1976 Coverage Improvement Program App-45
1977 REDUCTION App-46
ESTIMATION App-46
RELIABILITY OF THE ESTIMATES App-47
Sampling errors App-47
Standard errors of estimates of levels App-47
Standard errors of estimates of percentages App-48
Standard errors of ratios App-48
Illustration 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-51
Illustration of the computation of the 95-percent confidence interval for a median. App-52
Nonsampling errors App-52
Reinterview program App-52
Coverage errors App-53
Rounding errors App-53

SAMPLE DESIGN

The 1977 estimates are based on data collected in October 1977 through January 1978 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 70,600 sample housing units (both occupied and vacant) were eligible for interview in the 1977 Annual Housing Survey. Of this number, 4,000 interviews were classified as "noninterview," for various reasons. Occupied housing units were classified as "noninterview," mainly, because the occupants refused to be interviewed 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 70,600, there were also 6,300 sample units which were visited but found not to be eligible for interview for the AHS in terms of collecting information relevant to the 1977 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 1 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 1977 survey.—The sample housing units designated to be interviewed in the 1977 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1976 survey (which included all sample housing units that were selected as part of the 1976 Coverage Improvement Program) and that were not part of the 1977 reduction.

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 1976 survey and that were not part of the 1977 reduction. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1977 AHS questionnaire, page 1.)

3. All sample housing units that were selected from the list of building permits issued since the 1976 survey. (This sample represented the housing units built in permit-issuing areas since the 1976 survey.)

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. 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 at the 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 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 have 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.

1977 REDUCTION

By 1977, the addition to the sample from primarily new construction and the coverage improvements had increased the total sample size (interviews plus noninterviews) to about 81,000. The sample was reduced by about 7 percent to approximately 75,000. However, this reduction did not include any CEN-SUP[1] units or units which were selected as part of the 1976 Coverage Improvement Program. Thus, the overall probability of selection for these latter units remained unchanged, and, for the rest of the units, the probability of selection was about 1 in 1,472 if they were urban and about 1 in 736 if they were rural.

ESTIMATION

In 1977, 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 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

[1] CEN-SUP units resulted from a 1970 census evaluation study and represented units missed in the 1970 census.

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 iterated 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 15 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 14 categories of new construction units for each of the 4 regions (i.e., 9 categories for conventional new construction units and 5 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 variability between estimates from all possible samples is defined as sampling error. One common measure of sampling error is the standard error which measures the precision with which an estimate from a sample approximates the average results of all possible samples. In addition, the standard error, as calculated for this report, partially reflects the variation in the estimates due to some 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 sample estimate and its estimated standard error enable one to construct interval estimates so that the interval includes the average result of all possible samples with a known probability. 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 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.

Standard errors of estimates of levels.—Tables I and II present the standard errors applicable to the 1977 national housing inventory estimates in this report. Tables III and IV present the standard errors for

each of the four regions; Northeast, North Central, South, and West. Linear interpolation should be used to determine standard errors for levels of estimates not specifically shown in tables I through IV.

Standard errors of estimates of percentages.—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, VI, VII, and VIII present the standard errors of estimated percentages. Table V shows the approximate standard errors of all national estimated percentages of housing units except those pertaining to the specified items in table II. The standard errors shown in table VI should be used for those specified items. Table VII shows the approximate standard errors of all regional percentages of housing units except those pertaining to the specified items in table II. Table VIII should be used for those specified items for each of the four regions; Northeast, North Central, South, and West. Two-way 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 to be 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 I. Standard Errors of Estimated Numbers of Housing Units: 1977 (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: Total or White (000)	Standard error: Black (000)	Size of estimate (000)	Standard error: Total or White (000)	Standard error: Black (000)
0	2	2	1,000	40	38
5	3	3	2,500	63	53
10	4	4	5,000	88	55
25	6	6	10,000	120	–
50	9	9	25,000	168	–
100	13	13	50,000	177	–
250	20	20	75,000	95	–
500	29	28			

TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to Lacking Complete Plumbing Facilities, Mobile Homes, and Households with Head of Spanish Origin: 1977

(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	48	45
5	3	3	2,500	76	64
10	5	5	5,000	105	66
25	8	8	10,000	144	–
50	11	11	25,000	202	–
100	15	15	50,000	213	–
250	24	24	75,000	114	–
500	34	33			

Standard errors of ratios.—For ratios of the form (100) (x/y), where x is not a subclass of y, tables V 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-1 of this report shows that in the United States there were 5,489,000 owner-occupied housing units occupied by recent movers in 1977. Interpolation of the data in table I shows that the standard error of an estimate of this size is approximately 91,000. 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)
5,000	88
5,489	x
10,000	120

By vertically interpolating between 88 and 120, the entry for "x" is determined to be 91.

$$5{,}489-5{,}000 = 489$$
$$10{,}000-5{,}000 = 5{,}000$$

$$88 + \frac{489}{5{,}000}(120-88) = 91$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 5,398,000 to 5,580,000 housing units. Therefore, a conclusion that the average estimate of 1977 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 5,343,000 to 5,635,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 5,307,000 to 5,671,000 housing units with 95 percent confidence.

TABLE III. Standard Errors of Estimated Numbers of Housing Units Pertaining to Each of the Regions; Northeast, North Central, South, and West: 1977 (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 (000)	Size of estimate (000)	Standard error (000)
0	2	1,000 ..	45
5	3	2,500 ..	70
10	4	5,000 ..	97
25	7	10,000 .	133
50	10	25,000 .	187
100	14	50,000 .	196
250	22	75,000 .	105
500	32		

TABLE IV. Standard Errors of Estimated Numbers of Housing Units Pertaining to Lacking Complete Plumbing Facilities, Mobile Homes, and Households with Head of Spanish Origin, for Each of the Regions: Northeast, North Central, South, and West: 1977

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	1,000 ..	54
5	4	2,500 ..	84
10	5	5,000 ..	117
25	9	10,000 .	160
50	12	25,000 .	225
100	17	50,000 .	236
250	27	75,000 .	126
500	38		

Table A-1 also shows that of the 5,489,000 owner-occupied housing units occupied by recent movers in 1977, 340,000, or 6.2 percent, had 6 persons or more. Interpolation of the data in table V (i.e., interpolation on both the base and percent) shows that the standard error of the 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		
	5	6.2	10
5,000 ...	0.4	a	0.5
5,489 ...		p	
10,000 ..	0.3	b	0.4

1. By horizontal interpolation between 0.4 and 0.5, the entry for cell "a" is determined to be 0.4.

$$6.2-5.0 = 1.2$$
$$10.0-5.0 = 5.0$$

$$0.4 + \frac{1.2}{5.0}(0.5-0.4) = 0.4$$

2. By horizontal interpolation between 0.3 and 0.4, the entry for cell "b" is determined to be 0.3.

$$6.2-5.0 = 1.2$$
$$10.0-5.0 = 5.0$$

$$0.3 + \frac{1.2}{5.0}(0.4-0.3) = 0.3$$

3. By vertical interpolation between 0.4 and 0.3, the entry for "p" is determined to be 0.4.

$$5{,}489-5{,}000 = 489$$
$$10{,}000-5{,}000 = 5{,}000$$

$$0.4 + \frac{489}{5{,}000}(0.3-0.4) = 0.4$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 5.8 to 6.6 percent; the 90-percent confidence interval is from 5.6 to 6.8 percent; and the 95-percent confidence interval is from 5.4 to 7.0 percent.

Illustration II.—Table A-19 of this report shows that in the United States in 1977 there were 223,000 owner-occupied housing units occupied by Spanish recent movers. Interpolation of the data in table II shows that the standard error of an estimate of this size is approximately 22,000. Consequently, the 68-percent confidence interval is from 201,000 to 245,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1977 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 188,000 to 258,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 179,000 to 267,000 housing units with 95 percent confidence.

Table A-19 also shows that of the 223,000, 1977 owner-occupied housing units occupied by Spanish recent movers, 78,000, or 35.0 percent, had two bedrooms. Interpolation of the data in table

VI (i.e., interpolation on both the base and the percent) shows that the standard error of the above percentage is 5.0 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 30.0 to 40.0 percent; the 90-percent confidence interval is from 27.0 to 43.0 percent; and the 95-percent confidence interval is from 25.0 to 45.0 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 characteristics 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 in the United States in 1977 there were 512,000 owner-occupied housing units occupied by recent movers with 5 persons. Thus, the apparent difference between the number of 1977 owner-occupied housing units occupied by recent movers with 5 persons and the number with 6 persons or more is 172,000. Interpolation of the data in table I shows that the standard error on an estimate of 512,000 to be approximately 29,000 and the standard error on an estimate of 340,000 to be approximately 23,000. Therefore, the standard error of the estimated difference of 172,000 is about 37,000.

$$37{,}000 = \sqrt{(29{,}000)^2 + (23{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 172,000 difference is

TABLE V. Standard Errors of Estimated Percentages of Housing Units: 1977 (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	24.6	24.6	24.6	24.6	24.6	24.6	24.8	28.6
10	14.0	14.0	14.0	14.0	14.0	14.0	17.5	20.2
25	6.1	6.1	6.1	6.1	7.7	9.1	11.1	12.8
50	3.2	3.2	3.2	3.9	5.4	6.5	7.8	9.0
100	1.6	1.6	1.8	2.8	3.8	4.6	5.5	6.4
250	0.6	0.8	1.1	1.8	2.4	2.9	3.5	4.0
500	0.3	0.6	0.8	1.2	1.7	2.0	2.5	2.9
1,000	0.2	0.4	0.6	0.9	1.2	1.4	1.8	2.0
2,500	0.07	0.3	0.4	0.6	0.8	0.9	1.1	1.3
5,000	0.03	0.2	0.3	0.4	0.5	0.6	0.8	0.9
10,000	0.02	0.13	0.2	0.3	0.4	0.5	0.6	0.6
25,000	0.01	0.08	0.11	0.2	0.2	0.3	0.4	0.4
50,000	–	0.06	0.08	0.12	0.2	0.2	0.2	0.3
75,000	–	0.05	0.07	0.10	0.14	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, Mobile Homes, and Households with Head of Spanish Origin: 1977

(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	32.2	32.2	32.2	32.2	32.2	32.2	32.2	34.4
10	19.2	19.2	19.2	19.2	19.2	19.2	21.1	24.3
25	8.7	8.7	8.7	8.7	9.2	11.0	13.3	15.4
50	4.5	4.5	4.5	4.7	6.5	7.8	9.4	10.9
100	2.3	2.3	2.3	3.4	4.6	5.5	6.7	7.7
250	0.9	1.0	1.4	2.1	2.9	3.5	4.2	4.9
500	0.5	0.7	1.0	1.5	2.1	2.5	3.0	3.4
1,000	0.2	0.5	0.7	1.1	1.5	1.7	2.1	2.4
2,500	0.09	0.3	0.4	0.7	0.9	1.1	1.3	1.5
5,000	0.05	0.2	0.3	0.5	0.7	0.8	0.9	1.1
10,000	0.02	0.2	0.2	0.3	0.5	0.5	0.7	0.8
25,000	0.01	0.10	0.14	0.2	0.3	0.3	0.4	0.5
50,000	–	0.07	0.10	0.2	0.2	0.2	0.3	0.3
75,000	–	0.06	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.

from 135,000 to 209,000 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 113,000 to 231,000 housing units, and the 95-percent confidence interval is from 98,000 to 246,000. Thus, we can conclude with 95 percent confidence that the number of owner-occupied housing 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 base 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 de-

TABLE VII. Standard Errors of Estimated Percentages of Housing Units Pertaining to Each of the Four Regions; Northeast, North Central, South, and West: 1977 (Excluding 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	28.7	28.7	28.7	28.7	28.7	28.7	28.7	31.8
10	16.8	16.8	16.8	16.8	16.8	16.8	19.4	22.5
25	7.5	7.5	7.5	7.5	8.5	10.1	12.3	14.2
50	3.9	3.9	3.9	4.4	6.0	7.2	8.7	10.0
100	2.0	2.0	2.0	3.1	4.3	5.1	6.2	7.1
250	0.8	0.9	1.3	2.0	2.7	3.2	3.9	4.5
500	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
1,000	0.2	0.4	0.6	1.0	1.3	1.6	1.9	2.2
2,500	0.08	0.3	0.4	0.6	0.9	1.0	1.2	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	0.4
50,000	–	0.06	0.09	0.14	0.2	0.2	0.3	0.3
75,000	–	0.05	0.07	0.11	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.

TABLE VIII. Standard Errors of Estimated Percentages of Housing Units Pertaining to Lacking Complete Plumbing Facilities, Mobile Homes, and Households with Head of Spanish Origin, for Each of the Four Regions; Northeast, North Central, South, and West: 1977

(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	36.9	36.9	36.9	36.9	36.9	36.9	36.9	38.2
10	22.6	22.6	22.6	22.6	22.6	22.6	23.4	27.0
25	10.5	10.5	10.5	10.5	10.5	12.2	14.8	17.1
50	5.5	5.5	5.5	5.5	7.2	8.6	10.5	12.1
100	2.8	2.8	2.8	3.7	5.1	6.1	7.4	8.5
250	1.2	1.2	1.5	2.4	3.2	3.9	4.7	5.4
500	0.6	0.8	1.1	1.7	2.3	2.7	3.3	3.8
1,000	0.3	0.5	0.8	1.2	1.6	1.9	2.3	2.7
2,500	0.12	0.3	0.5	0.7	1.0	1.2	1.5	1.7
5,000	0.06	0.2	0.3	0.5	0.7	0.9	1.0	1.2
10,000	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.9
25,000	0.01	0.11	0.2	0.2	0.3	0.4	0.5	0.5
50,000	0.01	0.08	0.11	0.2	0.2	0.3	0.3	0.4
75,000	–	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.

termined 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 occupied by recent movers in the United States was 2.9 in 1977. The base of the distribution from which this median was determined is 5,489,000 housing units.

1. From table V, the standard error of a 50-percent characteristic on the base of 5,489,000 is 0.9 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 48.2 and 51.8.

3. From table A-1, it can be seen by cumulating the frequencies for the first 2 categories that 2,288,000 owner-occupied housing units occupied by recent movers, or 41.7 percent, had 1 or 2 persons (actually, the category of 2 persons is considered to be from 1.5 to 2.5 persons) and that an additional 1,108,000 housing units, or 20.2 percent, had 3 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 2.8.

$$2.5 + (3.5-2.5)\,\frac{(48.2-41.7)}{20.2} = 2.8$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about 3.0.

$$2.5 + (3.5-2.5)\,\frac{(51.8-41.7)}{20.2} = 3.0$$

Thus, the 95-percent confidence interval ranges from 2.8 to 3.0 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 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 the 1977 AHS national sample.

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 1977 reinterview study were not available at the time of publication. However, it is expected that they will be similar to the results of the 1976 reinterview study which are presented in the Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample, 1976." Unlike the years prior to 1976, the respondent's answers in the reinterview were not reconciled to the original answers given in the AHS interview; i.e., after the question was answered in the reinterview, the interviewer did not present the previous responses and then ask the respondent to decide upon the best answer. Comparing the reinterview results of 1976 with the years prior to 1976, we found that the estimates of inconsistency of all items (nonattitudinal and attitudinal) increased substantially in the 1976 results. In other words, providing the interviewer with the original response had the effect of reducing the levels of inconsistency.

To summarize the results of the 1976 reinterview program: Overall, it showed moderate to high levels of inconsistency with about one-third of the nonattitudinal items and a high level of inconsistency with about one-third of the attitudinal items. Only one-eighth of the nonattitudinal items showed a low level of inconsistency.

The range for evaluating inconsistency is 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.

The following list shows the 1977 AHS-2 questionnaire numbers for those items asked in the 1976 reinterview: Nonattitudinal, sections IIIA and IIIB, 12*, 13*, 15a, 15b, 15c*, 36a, 36b*, 37a, 37b*, 49*, 50*, 51a, 51b, 52a*, 54a*, 54b, 55a, 55b, 55c*, 56a, 56b, 58a*, 58b*, 61a*; Attitudinal, section IIIB, 102 all parts*, 103 all parts*, and 104 all parts*. Asterisks (*) indicate that the item had an estimated index of

inconsistency greater than 40. Since cross tabulations involving these items may be subject to a large distortion because of the moderate to high response variance, they should be considered to be less reliable than comparable cross tabulations which do not involve these items. The cutoff at 40 was selected because (1) the shape of the distribution had a natural break before 40, (2) the large sampling errors on the estimated indices indicated little difference between those indicies from 40 to 50 and those greater than 50, and (3) the break between moderate and high indices at 50 is arbitrary.

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 respondents 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 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 1977 AHS sample missed about 3.5 percent (i.e., about 375,000 units) of all conventional new construction built after April 1970, because the permits for these units, which were built before October 1977, 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 1977 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 were 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. However, biases of subtotals would still remain.

Rounding errors.—In 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. Also, since medians in this report were computed using unrounded data, instead of the shown published rounded data, they can differ from medians calculated directly from the published data.

*U.S. GOVERNMENT PRINTING OFFICE: 1979 O-281-049/1074

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-44
Selection of sample areas App-44
Designation of sample housing units for the 1977 survey App-44
Selection of the 1973 sample housing units App-44
Splitting of the sample App-45
Selection of supplemental sample housing units in rural areas App-45
Selection of sample housing units for the 1976 Coverage Improvement Program App-45
1977 reduction App-46
1970 Census of Population and Housing App-46
ESTIMATION App-46
AHS national sample App-46
1977 housing inventory App-46
1973-1977 lost units App-47
Ratio estimation procedure of the 1970 Census of Population and Housing App-47
RELIABILITY OF THE ESTIMATES App-47
Sampling errors App-48
Standard errors of estimates of levels App-48
Standard errors of estimates of percentages App-49
Standard errors of ratios App-49
Illustration of the use of the standard error tables. Illustration I App-50
Illustration II App-51
Differences App-51
Illustration of the computation of the standard error of a difference App-52
Medians App-53
Illustration of the computation of the 95-percent confidence interval for a median. App-53
Nonsampling errors App-54
1970 census App-54
Reinterview program App-54
Coverage errors App-56
Rounding errors App-56

SAMPLE DESIGN

The 1977 estimates are based on data collected in October 1977 through January 1978 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 70,600 sample housing units (both occupied and vacant) were eligible for interview in the 1977 Annual Housing Survey. Of this number, 4,000 interviews were classified as "noninterview" for various reasons. Occupied housing units were classified as "noninterview," mainly, because the occupants refused to be interviewed 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 70,600, there were also 6,300 sample units which were visited but found not to be eligible for interview for the AHS in terms of collecting information relevant to the 1977 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 1 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 1 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 2 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 1977 survey.—The sample housing units designated to be interviewed in the 1977 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1976 survey (which included all sample housing units that were selected as part of the 1976 Coverage Improvement Program) and that were not part of the 1977 reduction.
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 1976 survey and that were not part of the 1977 reduction. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1977 AHS questionnaire, page 1.)
3. All sample housing units that were selected from the list of building permits issued since the 1976 survey. (This sample represented the housing units built in permit-issuing areas since the 1976 survey.)

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. 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 at the 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 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 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.

1977 reduction.—By 1977, the addition to the sample from primarily new construction and the coverage improvements had increased the total sample size (interviews plus noninterviews) to about 81,000. The sample was reduced by about 7 percent to approximately 75,000. However, this reduction did not include any CEN-SUP[1] units or units which were selected as part of the 1976 Coverage Improvement Program. Thus, the overall probability of selection for these latter units remained unchanged, and, for the rest of the units, their probability of selection was about 1 in 1,472, if they were urban, and about 1 in 736, if they were rural.

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 1977 housing inventory and estimates of units removed from the housing inventory between 1973 and 1977 (i.e., 1973-1977 lost units). Each type of estimate employed a separate, though similar, estimation procedure.

1977 housing inventory.—In 1977, 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

[1] CEN-SUP units resulted from a 1970 census evaluation study and represented units missed in the 1970 census.

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 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- and third-stage ratio estimation procedures were iterated 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 15 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 14 categories of new construction units for each of the 4 regions (i.e., 9 categories for conventional new construction units and 5 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.

1973-1977 lost units.—The 1973-1977 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-150-73, **General Housing Characteristics for the United States and Regions.** These 1973-1977 lost units do not include the housing units from the 1976 Coverage Improvement. Since the 1973-1977 lost units existed, by definition, in the 1973 housing inventory, there was a 1973 housing inventory weight associated with each 1973-1977 lost unit. This weight, adjusted for the 1977 reduction, was used to tabulate the estimates of the characteristics of the 1973-1977 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 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 variability between estimates from all possible samples is defined as sampling error. One common measure of sampling error is the standard error which measures the precision with which an estimate from a sample approximates the average results of all possible samples. In addition, the standard error, as calculated for this report, partially reflects the variation in the estimates due to some 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 sample estimate and its estimated standard error enable one to construct interval estimates so that the interval includes the average result of all possible samples with a known probability. 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 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.

Standard errors of estimates of levels.—Tables I, II, III, and IV present the standard errors applicable to the 1977 national housing inventory estimates in

TABLE I. Standard Errors of Estimated Numbers of Housing Units: 1977 (Excluding Estimates of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, New Construction, Rural Housing Units, Source of Water, Households with Head of Spanish Origin, Urban 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	2	2	1,000	40	38
5	3	3	2,500	63	53
10	4	4	5,000	88	55
25	6	6	10,000	120	–
50	9	9	25,000	168	–
100	13	13	50,000	177	–
250	20	20	75,000	95	–
500	29	28			

TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, New Construction, Source of Water, and Households with Head of Spanish Origin: 1977

(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	48	45
5	3	3	2,500	76	64
10	5	5	5,000	105	66
25	8	8	10,000	144	–
50	11	11	25,000	202	–
100	15	15	50,000	213	–
250	24	24	75,000	114	–
500	34	33			

this report and tables V and VI present the standard errors applicable to 1973-1977 lost housing unit estimates in this report. Table VII presents the standard errors for each of the four regions; Northeast, North Central, South, and West. Linear interpolation should be used to determine standard error for levels of estimates not specifically shown in tables I through VII.

Standard errors of estimates of percentages.—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 VIII, IX, X, XI, XII, XIII, and XIV present the standard errors of estimated percentages. Tables VIII, IX, X, and XI show the approximate standard errors of all national estimated percentages of housing units. Tables XII and XIII show the approximate standard errors of national estimated percentages of 1973-1977 lost housing units. Table XIV shows the approximate standard errors of all regional estimated percentages of housing units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables VIII through XIV.

Included in tables I through XIV are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered to be 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.

Standard errors of ratios.—For ratios of the form (100) (x/y), where x is not a subclass of y, tables VIII through XIV 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

TABLE III. Standard Errors of Estimated Numbers of Urban or Rural Housing Units: 1977

(68 chances out of 100)

Size of estimate	Standard error					
	Rural housing units (except those in next column)		Rural housing units pertaining to cooking fuel, heating fuel-electricity only, 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, heating fuel-electricity only, lacking complete plumbing facilities, mobile homes, new construction, and source of water	
(000)	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	3	3
5	2	2	3	3	4	4
10	4	4	4	4	5	5
25	6	6	7	7	8	8
50	8	8	10	10	11	11
100	11	11	14	14	16	16
250	18	17	22	21	25	25
500	25	24	31	30	35	34
1,000	35	33	43	41	50	47
2,500	55	46	68	57	78	66
5,000	76	48	94	59	108	68
10,000	104	–	129	–	148	–
25,000	146	–	181	–	208	–
50,000	153	–	190	–	219	–

TABLE IV. Standard Errors of Estimated Numbers of Vacant Housing Units: 1977

(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	2	1	2
5	2	3	2	4
10	3	4	3	5
25	5	6	5	8
50	7	9	8	11
100	10	12	11	16
250	17	20	18	28
500	24	28	27	43
1,000	35	40	42	69
2,500	59	66	83	144
3,500	74	79	–	–

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 urban areas of the United States there were 21,809,000 renter-occupied housing units in 1977. Interpolation of the data in table III shows that the standard error of an estimate of this size is approximately 170,000. 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	129
21,809	x
25,000	181

By vertically interpolating between 129 and 181, the entry for "x" is determined to be:

$$21{,}809-10{,}000 = 11{,}809$$
$$25{,}000-10{,}000 = 15{,}000$$

$$129 + \frac{11{,}809}{15{,}000}(181-129) = 170$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 21,639,000 to 21,979,000 housing units. Therefore, a conclusion that the average estimate of 1977 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 21,537,000 to 22,081,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 21,469,000 to 22,149,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 21,809,000 renter-occupied housing units in urban areas, 6,447,000, or 29.6 percent, were occupied by two persons. Interpolation of the data in table XI (i.e., interpolation on both the base and percent) shows that the standard error of the above percentage is 0.4. The following procedure was used in interpolating.

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

Base of percentage (000)	Estimated percentage		
	25	29.6	50
10,000	0.6	a	0.7
21,809		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 0.6.

$$29.6-25.0 = 4.6$$
$$50.0-25.0 = 25.0$$

$$0.6 + \frac{4.6}{25.0}(0.7-0.6) = 0.6$$

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

$$29.6-25.0 = 4.6$$
$$50.0-25.0 = 25.0$$

$$0.4 + \frac{4.6}{25.0}(0.4-0.4) = 0.4$$

3. By vertical interpolation between 0.6 and 0.4, the entry for "p" is determined to be 0.4.

$$21{,}809-10{,}000 = 11{,}809$$
$$25{,}000-10{,}000 = 15{,}000$$

$$0.6 + \frac{11{,}809}{15{,}000}(0.4-0.6) = 0.4$$

TABLE V. Standard Errors of Estimated Numbers of Lost Housing Units: 1973-1977 (Excluding Estimates of Lost Housing Units Pertaining to Lacking Bedroom, Lacking Kitchen Facilities, Lacking Some or All Plumbing Facilities, and Other Vacant)

(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
750	41
1,000	49
1,750	70
2,500	89

TABLE VI. Standard Errors of Estimated Numbers of Lost Housing Units Pertaining to Lacking Bedroom, Lacking Kitchen Facilities, Lacking Some or All Plumbing Facilities, and Other Vacant: 1973-1977

(68 chances out of 100)

Size of estimate (000)	Standard error (000)
0	2
5	3
10	5
25	8
50	11
100	16
250	27
500	42
750	56
1,000	69
1,750	107
2,500	145

TABLE VIIa. Standard Errors of Estimated Numbers of Housing Units Pertaining to Each of the Regions; Northeast, North Central, South, and West: 1977 (Excluding Estimates of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, New Construction, Rural Housing Units, Source of Water, Households with Head of Spanish Origin, Urban Housing Units, and Vacant Housing Units)

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	1,000	45
5	3	2,500	70
10	4	5,000	97
25	7	10,000	133
50	10	25,000	187
100	14	50,000	196
250	22	75,000	105
500	32		

Note: For standard errors of regional estimates of rural housing units or urban housing units and of vacant housing units, use the national standard errors in tables III and IV, respectively, multiplied by 1.1.

TABLE VIIb. Standard Errors of Estimated Numbers of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, New Construction, Source of Water, and Households with Head of Spanish Origin, for Each of the Regions; Northeast, North Central, South, and West: 1977

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	1,000	54
5	4	2,500	84
10	5	5,000	117
25	9	10,000	160
50	12	25,000	225
100	17	50,000	236
250	27	75,000	126
500	38		

TABLE VIII. Standard Errors of Estimated Percentages of Urban Year-Round Vacant Housing Units and of Total Housing Units (Excluding Estimated Percentages of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, New Construction, Rural Housing Units, Source of Water, Households with Head of Spanish Origin, Urban Occupied Housing Units, and Total Vacant Year-Round Housing Units): 1977

(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	24.6	24.6	24.6	24.6	24.6	24.6	24.8	28.6
10	14.0	14.0	14.0	14.0	14.0	14.0	17.5	20.2
25	6.1	6.1	6.1	6.1	7.7	9.1	11.1	12.8
50	3.2	3.2	3.2	3.9	5.4	6.5	7.8	9.0
100	1.6	1.6	1.8	2.8	3.8	4.6	5.5	6.4
250	0.6	0.8	1.1	1.8	2.4	2.9	3.5	4.0
500	0.3	0.6	0.8	1.2	1.7	2.0	2.5	2.9
1,000	0.2	0.4	0.6	0.9	1.2	1.4	1.8	2.0
2,500	0.07	0.3	0.4	0.6	0.8	0.9	1.1	1.3
5,000	0.03	0.2	0.3	0.4	0.5	0.6	0.8	0.9
10,000	0.02	0.13	0.2	0.3	0.4	0.5	0.6	0.6
25,000	0.01	0.08	0.11	0.2	0.2	0.3	0.4	0.4
50,000	—	0.06	0.08	0.12	0.2	0.2	0.2	0.3
75,000	—	0.05	0.07	0.10	0.14	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.

Consequently, the 68-percent confidence interval, as shown by these data, is from 29.2 to 30.0 percent; the 90-percent confidence interval is from 29.0 to 30.2 percent; and the 95-percent confidence interval is from 28.8 to 30.4 percent.

Illustration II.—Table A-2 of this report shows that in the rural areas of the United States in 1977 there were 127,000 specified owner-occupied farm housing units. Interpolation of the data in table III shows that the standard error of an estimate of this size is approximately 12,000. Consequently, the 68-percent confidence interval is from 115,000 to 139,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1977 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 108,000 to 146,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 103,000 to 151,000 housing units with 95 percent confidence.

Table A-2 also shows that of the 127,000 specified owner-occupied farm housing units in rural areas, 54,000, or 42.5 percent, were owned free and clear. Interpolation of the data in table X (i.e., interpolation on both the base and the percent) shows that the standard error of the above percentage is 5.0 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 37.5 to 47.5 percent; the 90-percent confidence interval is from 34.5 to 50.5 percent; and the 95-percent confidence interval is from 32.5 to 52.5 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The stand-

ard 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 characteristics in two different areas or the difference between separate and uncorrelated characteristics in the same area. If 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 in urban areas of the United States there were 3,359,000 renter-occupied housing units with three persons in 1977. Thus, the apparent difference between the number of 1977 renter-occupied housing units in urban areas with 2 persons and those with 3 persons is 3,088,000. Interpolation of the data in table III shows that the standard error of an estimate of 6,447,000 is approximately 104,000, and that the standard error of an estimate of 3,359,000 is approximately 77,000. Therefore, the standard error of the estimated difference of 3,088,000 is about:

$$129{,}000 = \sqrt{(104{,}000)^2 + (77{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 3,088,000 difference is from 2,959,000 to 3,217,000 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 2,882,000 to 3,294,000 housing units, and the 95-percent confidence interval is from 2,830,000 to 3,346,000. Thus, we can conclude with 95 percent confidence that the number of 1977 renter-occupied housing units in urban areas with two persons is greater than the number with three persons, since the 95-percent con-

TABLE IX. Standard Errors of Estimated Percentages of Total Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, New Construction, Source of Water, and Households with Head of Spanish Origin; of Seasonal and Migratory Vacant Housing Units; and of Urban Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Mobile Homes, Lacking Complete Plumbing Facilities, New Construction, and Source of Water: 1977

(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	33.4	33.4	33.4	33.4	33.4	33.4	33.4	35.4
10	20.1	20.1	20.1	20.1	20.1	20.1	21.7	25.0
25	9.1	9.1	9.1	9.1	9.5	11.3	13.7	15.8
50	4.8	4.8	4.8	4.9	6.7	8.0	9.7	11.2
100	2.4	2.4	2.4	3.5	4.8	5.7	6.9	7.9
250	1.0	1.0	1.4	2.2	3.0	3.6	4.3	5.0
500	0.5	0.7	1.0	1.5	2.1	2.5	3.1	3.5
1,000	0.3	0.5	0.7	1.1	1.5	1.8	2.2	2.5
2,500	0.10	0.3	0.4	0.7	1.0	1.1	1.4	1.6
5,000	0.05	0.2	0.3	0.5	0.7	0.8	1.0	1.1
10,000	0.03	0.2	0.2	0.3	0.5	0.6	0.7	0.8
25,000	0.01	0.10	0.14	0.2	0.3	0.4	0.4	0.5
50,000	0.01	0.07	0.10	0.2	0.2	0.3	0.3	0.4
75,000	–	0.06	0.08	0.13	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 X. Standard Errors of Estimated Percentages of Total Year-Round Vacant Housing Units, of Year-Round Rural Vacant Housing Units, and of Rural Housing Units (Excluding Estimated Percentages of Rural Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Mobile Homes, New Construction, Source of Water, and Households with Head of Spanish Origin: 1977

(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	19.7	19.7	19.7	19.7	19.7	19.7	21.5	24.8
10	10.9	10.9	10.9	10.9	10.9	12.5	15.2	17.5
25	4.7	4.7	4.7	4.8	6.7	7.9	9.6	11.1
50	2.4	2.4	2.4	3.4	4.7	5.6	6.8	7.8
100	1.2	1.2	1.6	2.4	3.3	4.0	4.8	5.5
250	0.5	0.7	1.0	1.5	2.1	2.5	3.0	3.5
500	0.2	0.5	0.7	1.1	1.5	1.8	2.1	2.5
1,000	0.12	0.3	0.5	0.8	1.1	1.3	1.5	1.8
2,500	0.05	0.2	0.3	0.5	0.7	0.8	1.0	1.1
5,000	0.02	0.2	0.2	0.3	0.5	0.6	0.7	0.8
10,000	0.01	0.11	0.2	0.2	0.3	0.4	0.5	0.6
25,000	–	0.07	0.10	0.2	0.2	0.3	0.3	0.4

[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.

fidence 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 appropriate standard error table, 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 2.7 in 1977. The base of the distribution, from which this median was determined, is 31,890,000 housing units.

1. From table XI, the standard error of a 50-percent characteristic on the base of 31,890,000 is 0.4 percentage points.

TABLE XI. Standard Errors of Estimated Percentages of Rural Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Mobile Homes, New Construction, Source of Water, and Households with Head of Spanish Origin, and of Urban Housing Units (Excluding Estimated Percentages of Urban Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, New Construction, and Source of Water): 1977

(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	27.4	27.4	27.4	27.4	27.4	27.4	27.4	30.7
10	15.9	15.9	15.9	15.9	15.9	15.9	18.8	21.7
25	7.0	7.0	7.0	7.0	8.2	9.8	11.9	13.7
50	3.6	3.6	3.6	4.2	5.8	6.9	8.4	9.7
100	1.9	1.9	1.9	3.0	4.1	4.9	6.0	6.9
250	0.8	0.9	1.2	1.9	2.6	3.1	3.8	4.3
500	0.4	0.6	0.9	1.3	1.8	2.2	2.7	3.1
1,000	0.2	0.4	0.6	0.9	1.3	1.6	1.9	2.2
2,500	0.08	0.3	0.4	0.6	0.8	1.0	1.2	1.4
5,000	0.04	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	0.13	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 XII. Standard Errors of Estimated Percentages of Lost Housing Units: 1973-1977 (Excluding Estimated Percentages of Lost Housing Units Pertaining to Lacking Bedroom, Lacking Kitchen Facilities, Lacking Some or All Plumbing Facilities, and Other Vacant)

(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	26.7	26.7	26.7	26.7	26.7	26.7	26.7	30.2
10	15.4	15.4	15.4	15.4	15.4	15.4	18.5	21.4
25	6.8	6.8	6.8	6.8	8.1	9.6	11.7	13.5
50	3.5	3.5	3.5	4.2	5.7	6.8	8.3	9.6
100	1.8	1.8	1.9	2.9	4.1	4.8	5.8	6.8
250	0.7	0.9	1.2	1.9	2.6	3.1	3.7	4.3
500	0.4	0.6	0.8	1.3	1.8	2.2	2.6	3.0
750	0.2	0.5	0.7	1.1	1.5	1.8	2.1	2.5
1,000	0.2	0.4	0.6	0.9	1.3	1.5	1.8	2.1
1,750	0.10	0.3	0.5	0.5	0.7	1.0	1.4	1.6
2,500	0.07	0.3	0.4	0.6	0.8	1.0	1.2	1.4

[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.

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 2 categories that 14,705,000 owner-occupied housing units, or 46.1 percent, had 1 and 2 persons (actually, for purposes of calculating the median, the category of 2 persons is considered to be from 1.5 to 2.5 persons) and that an additional 5,679,000 owner-occupied housing units, or 17.8 percent, had 3 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:

$$2.5 + (3.5-2.5)\,\frac{49.2-46.1}{17.8} = 2.7$$

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

$$2.5 + (3.5-2.5)\,\frac{50.8-46.1}{17.8} = 2.8$$

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 lower limit, it actually is a reflection of the rounding error associated with the 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 this list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

TABLE XIII. Standard Errors of Estimated Percentages of Lost Housing Units Pertaining to Lacking Bedroom, Lacking Kitchen Facilities, Lacking Some or All Plumbing Facilities, and Other Vacant: 1973-1977

(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	31.3	31.3	31.3	31.3	31.3	31.3	31.3	33.8
10	18.6	18.6	18.6	18.6	18.6	18.6	20.7	23.9
25	8.4	8.4	8.4	8.4	9.1	10.8	13.1	15.1
50	4.4	4.4	4.4	4.7	6.4	7.6	9.3	10.7
100	2.2	2.2	2.2	3.3	4.5	5.4	6.5	7.6
250	0.9	1.0	1.3	2.1	2.9	3.4	4.1	4.8
500	0.5	0.7	0.9	1.5	2.0	2.4	2.9	3.4
750	0.3	0.5	0.8	1.2	1.7	2.0	2.4	2.8
1,000	0.2	0.5	0.7	1.0	1.4	1.7	2.1	2.4
1,750	0.13	0.4	0.5	0.8	1.1	1.3	1.6	1.8
2,500	0.09	0.3	0.4	0.7	0.9	1.1	1.3	1.5

[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.

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 1977 AHS national sample.

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 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 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 1977 reinterview study were not available at the time of publication. However, it is expected that they will be similar to the results of the 1976 reinterview study which are presented in the Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample, 1976." Unlike the years prior to 1976, the respondent's answers in the reinterview were not reconciled to the original answers given in the AHS interview; i.e., after the question was answered in the reinterview, the interviewer did not present the previous responses and then ask the respondent to decide upon the best answer. Comparing the reinterview results of 1976 with the years prior to 1976, we found that the estimates of inconsistency of all items (nonattitudinal and attitudinal) increased substantially in the 1976 results. In other words, providing the interviewer with the original response had the effect of reducing the levels of inconsistency.

To summarize the results of the 1976 reinterview program: Overall, it showed moderate to high levels of inconsistency with about one-third of the nonattitudinal items and a high level of inconsistency with about one-third of the attitudinal items. Only one-eighth of the nonattitudinal items showed a low level of inconsistency.

The range for evaluating inconsistency is 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.

TABLE XIVa. Standard Errors of Estimated Percentages of Housing Units Pertaining to Each of the Four Regions; Northeast, North Central, South, and West: 1977 (Excluding Estimated Percentages of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, New Construction, Rural Housing Units, Source of Water, Households With Head of Spanish Origin, Urban Housing Units, 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	28.7	28.7	28.7	28.7	28.7	28.7	28.7	31.8
10	16.8	16.8	16.8	16.8	16.8	16.8	19.4	22.5
25	7.5	7.5	7.5	7.5	8.5	10.1	12.3	14.2
50	3.9	3.9	3.9	4.4	6.0	7.2	8.7	10.0
100	2.0	2.0	2.0	3.1	4.3	5.1	6.2	7.1
250	0.8	0.9	1.3	2.0	2.7	3.2	3.9	4.5
500	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
1,000	0.2	0.4	0.6	1.0	1.3	1.6	1.9	2.2
2,500	0.08	0.3	0.4	0.6	0.9	1.0	1.2	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	0.4
50,000	—	0.06	0.09	0.14	0.2	0.2	0.3	0.3
75,000	—	0.05	0.07	0.11	0.2	0.2	0.2	0.3

Note: For standard errors of regional estimated percentages of rural housing units, use national standard errors in tables X or XI, multiplied by 1.1. For regional estimated percentages of urban housing units, use national standard errors in tables IX or XI, multiplied by 1.1. For regional estimated percentages of vacant housing units, use national standard errors in tables VIII, IX or X, multiplied by 1.1.

[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 following list shows the 1977 AHS-2 questionnaire numbers for those items asked in the 1976 reinterview: Nonattitudinal, sections IIIA and IIIB, 12*, 13*, 15a, 15b, 15c*, 36a, 36b*, 37a, 37b*, 49*, 50*, 51a, 51b, 52a*, 54a*, 54b, 55a, 55b, 55c*, 56a, 56b, 58a*, 58b*, 61a*; Attitudinal, section IIIB, 102 all parts*, 103 all parts*, and 104 all parts*. Asterisks (*) indicate that the item had an estimated index of inconsistency greater than 40. Since cross tabulations involving these items may be subject to a large distortion because of the moderate to high response variance, they should be considered to be less reliable than comparable cross tabulations which do not involve these items. The cutoff at 40 was selected because (1) the shape of the distribution had a natural break before 40, (2) the large sampling errors on the estimated indices indicated little difference between those indices from 40 to 50 and those greater than 50, and (3) the break between moderate and high indices at 50 is arbitrary.

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 respondents 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 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 1977 AHS sample missed about 3.5 percent (i.e., about 375,000 units) of all conventional new construction built after April 1970, because the permits for these units, which were built before October 1977, 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 1977 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 were 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. However, biases of subtotals would still remain.

TABLE XIVb. Standard Errors of Estimated Percentages of Housing Units Pertaining to Cooking Fuel, Heating Fuel-Electricity Only, Lacking Complete Plumbing Facilities, Mobile Homes, New Construction, Source of Water, and Households with Head of Spanish Origin, for Each of the Four Regions; Northeast, North Central, South, and West: 1977

(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	36.9	36.9	36.9	36.9	36.9	36.9	36.9	38.2
10	22.6	22.6	22.6	22.6	22.6	22.6	23.4	27.0
25	10.5	10.5	10.5	10.5	10.5	12.2	14.8	17.1
50	5.5	5.5	5.5	5.5	7.2	8.6	10.5	12.1
100	2.8	2.8	2.8	3.7	5.1	6.1	7.4	8.5
250	1.2	1.2	1.5	2.4	3.2	3.9	4.7	5.4
500	0.6	0.8	1.1	1.7	2.3	2.7	3.3	3.8
1,000	0.3	0.5	0.8	1.2	1.6	1.9	2.3	2.7
2,500	0.12	0.3	0.5	0.7	1.0	1.2	1.5	1.7
5,000	0.06	0.2	0.3	0.5	0.7	0.9	1.0	1.2
10,000	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.9
25,000	0.01	0.11	0.2	0.2	0.3	0.4	0.5	0.5
50,000	0.01	0.08	0.11	0.2	0.2	0.3	0.3	0.4
75,000	–	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.

Rounding errors.—In 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. Also, since medians in this report were computed using unrounded data, instead of the shown published rounded data, they can differ from medians calculated directly from the published data.

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-44
Selection of sample areas App-44
Designation of sample housing units for the 1977 survey App-44
Selection of the 1973 sample housing units App-44
Splitting of the sample App-45
Selection of supplemental sample housing units in rural areas App-45
Selection of sample housing units for the 1976 Coverage Improvement Program App-45
1977 REDUCTION App-46
ESTIMATION App-46
RELIABILITY OF THE ESTIMATES App-47
Sampling errors App-47
Standard errors of estimates of levels App-47
Standard errors of estimates of percentages App-48
Standard errors of ratios App-49
Illustration of the use of the standard error tables. Illustration I App-49
Illustration II App-49
Differences App-50
Illustration of the computation of the standard error of a difference App-50
Medians App-51
Illustration of the computation of the 95-percent confidence interval for a median. App-51
Nonsampling errors App-51
Reinterview program App-51
Coverage errors App-53
Rounding errors App-53

SAMPLE DESIGN

The 1977 estimates are based on data collected in October 1977 through January 1978 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 70,600 sample housing units (both occupied and vacant) were eligible for interview in the 1977 Annual Housing Survey. Of this number, 4,000 interviews were classified as "noninterview" for various reasons. Occupied housing units were classified as "noninterview," mainly, because the occupants refused to be interviewed 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 70,600, there were also 6,300 sample units which were visited but found not to be eligible for interview for the AHS in terms of collecting information relevant to the 1977 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 1 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 1977 survey.—The sample housing units designated to be interviewed in the 1977 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1976 survey (which included all sample housing units that were selected as part of the 1976 Coverage Improvement Program) and that were not part of the 1977 reduction.
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 1976 survey and that were not part of the 1977 reduction. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1977 AHS questionnaire, page 1.)
3. All sample housing units that were selected from the list of building permits issued since the 1976 survey. (This sample represented the housing units built in permit-issuing areas since the 1976 survey.)

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. 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 at the 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 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 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.

1977 REDUCTION

By 1977, the addition to the sample from primarily new construction and the coverage improvements had increased the total sample size (interviews plus noninterviews) to about 81,000. The sample was reduced by about 7 percent to approximately 75,000. However, this reduction did not include any CEN-SUP[1] units or units which were selected as part of the 1976 Coverage Improvement Program. Thus, the overall probability of selection for these latter units remained unchanged, and, for the rest of the units, the probability of selection was about 1 in 1,472 if they were urban and about 1 in 736 if they were rural.

ESTIMATION

In 1977, 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 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

[1] CEN-SUP units resulted from a 1970 census evaluation study and represented units missed in the 1970 census.

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 iterated 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 15 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 14 categories of new construction units for each of the 4 regions (i.e., 9 categories for conventional new construction units and 5 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 variability between estimates from all possible samples is defined as sampling error. One common measure of sampling error is the standard error which measures the precision with which an estimate from a sample approximates the average results of all possible samples. In addition, the standard error, as calculated for this report, partially reflects the variation in the estimates due to some 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 sample estimate and its estimated standard error enable one to construct interval estimates so that the interval includes the average result of all possible samples with a known probability. 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 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.

Standard errors of estimates of levels.—Tables I and II present the standard errors applicable to the 1977 national housing

inventory estimates in this report. Tables III and IV present the standard errors for each of the four regions; Northeast, North Central, South, and West. Linear interpolation should be used to determine standard errors for levels of estimates not specifically shown in tables I through IV.

Standard errors of estimates of percentages.—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, VI, VII, and VIII present the standard errors of estimated percentages. Table V shows the approximate standard errors of all national estimated percentages of housing units except those pertaining to the specified items in table II. The standard errors shown in table VI should be used for those specified items. Table VII shows the approximate standard errors of all regional percentages of housing units except those pertaining to the specified item in table IV. Table VIII should be used for that specified item for each of the four regions; Northeast, North Central, South, and West. Two-way 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 to be

TABLE I. Standard Errors of Estimated Numbers of Housing Units: 1977 (Excluding Estimates of Housing Units Pertaining to Lacking Some or All 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	2	2	1,000	40	38
5	3	3	2,500	63	53
10	4	4	5,000	88	55
25	6	6	10,000	120	—
50	9	9	25,000	168	—
100	13	13	50,000	177	—
250	20	20	75,000	95	—
500	29	28			

TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to Lacking Some or All Plumbing Facilities and Households with Head of Spanish Origin: 1977

(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	48	45
5	3	3	2,500	76	64
10	5	5	5,000	105	66
25	8	8	10,000	144	—
50	11	11	25,000	202	—
100	15	15	50,000	213	—
250	24	24	75,000	114	—
500	34	33			

TABLE III. Standard Errors of Estimated Numbers of Housing Units Pertaining to Each of the Regions; Northeast, North Central, South, And West: 1977 (Excluding Estimates of Housing Units Pertaining to Lacking Some or All Plumbing Facilities)

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	1,000	45
5	3	2,500	70
10	4	5,000	97
25	7	10,000	133
50	10	25,000	187
100	14	50,000	196
250	22	75,000	105
500	32		

TABLE IV. Standard Errors of Estimated Numbers of Housing Units Pertaining to Lacking Some or All Plumbing Facilities for Each of the Regions; Northeast, North Central, South, and West: 1977

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	1,000	54
5	4	2,500	84
10	5	5,000	117
25	9	10,000	160
50	12	25,000	225
100	17	50,000	236
250	27	75,000	126
500	38		

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.

Standard errors of ratios.—For ratios of the form (100) (x/y), where x is not a subclass of y, tables V 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-1 of this report shows that in the United States there were 22,346,000 owner-occupied housing units with garbage collection service once a week in 1977. Interpolation of the data in table I shows that the standard error of an estimate of this size is approximately 160,000. 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	120
22,346	x
25,000	168

By vertically interpolating between 120 and 168, the entry for "x" is determined to be 160.

$$22{,}346-10{,}000 = 12{,}346$$
$$25{,}000-10{,}000 = 15{,}000$$

$$120+\frac{12{,}346}{15{,}000}(168-120) = 160$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 22,186,000 to 22,506,000 housing units. Therefore, a conclusion that the average estimate of 1977 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 22,090,000 to 22,602,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 22,026,000 to 22,666,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 22,346,000 owner-occupied housing units with garbage collection service once a week in 1977, 1,431,000, or 6.4 percent, had a family or primary individual income between $5,000 and $6,999. Interpolation of the data in table V (i.e., interpolation on both the base and percent) shows that the standard error of the above percentage is 0.2 percentage points. The following procedure was used in interpolating.

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

Base of percentage (000)	Estimated percentage		
	5	6.4	10
10,000	0.3	a	0.4
22,346		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 0.3.

$$6.4-5.0 = 1.4$$
$$10.0-5.0 = 5.0$$

$$0.3+\frac{1.4}{5.0}(0.4-0.3) = 0.3$$

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

$$6.4-5.0 = 1.4$$
$$10.0-5.0 = 5.0$$

$$0.2+\frac{1.4}{5.0}(0.2-0.2) = 0.2$$

3. By vertical interpolation between 0.3 and 0.2, the entry for "p" is determined to be 0.2.

$$22{,}346-10{,}000 = 12{,}346$$
$$25{,}000-10{,}000 = 15{,}000$$

$$0.3+\frac{12{,}346}{15{,}000}(0.2-0.3) = 0.2$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 6.2 to 6.6 percent; the 90-percent confidence interval is from 6.1 to 6.7 percent; and the 95-percent confidence interval is from 6.0 to 6.8 percent.

Illustration II.—Table A-3 of this report shows that in the United States in 1977 there were 664,000 owner-occupied housing units lacking some or all plumbing facilities that were occupied 3 months or longer. Interpolation of the data in table II shows that the standard error of an estimate of this size is approximately 39,000. Consequently, the 68-percent confidence interval is from 625,000 to 703,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1977 owner-occupied housing units lacking some or all plumbing facilities that were occupied 3 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 possible samples, lies within the interval from 602,000 to 726,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 586,000 to 742,000 housing units with 95 percent confidence.

Table A-3 also shows that of the 664,000, 1977 owner-occupied housing units lacking some or all plumbing facilities, that were occupied 3 months or

longer, 80,000, or 12.0 percent, had a family or primary individual income between $5,000 and $6,999. Interpolation of the data in table VI (i.e., interpolation on both the base and the percent) shows that the standard error of the above percentage is 2.1 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 9.9 to 14.1 percent; the 90-percent confidence interval is from 8.6 to 15.4 percent; and the 95-percent confidence interval is from 7.8 to 16.2 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristics in two different areas or the difference between separate and uncorrelated characteristics in the same area. If 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 in the United States in 1977 there were 2,077,000 owner-occupied housing units with garbage collection service once a week that had a family or primary individual income between $10,000 and $12,499. Thus, the apparent difference between the number of owner-occupied housing units with garbage collection service once a week that had a family or primary individual income between $5,000 and $6,999 and the number that had a family or primary individual income between $10,000 and $12,499 is 646,000. Interpolation of the data in table I shows that the standard error on an estimate of 2,077,000 to be approximately 57,000 and the standard error on an estimate of 1,431,000 to be approximately 47,000. Therefore, the

TABLE V. Standard Errors of Estimated Percentages of Housing Units: 1977 (Excludes Estimated Percentages of Housing Units Pertaining to Lacking Some or All 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	24.6	24.6	24.6	24.6	24.6	24.6	24.8	28.6
10	14.0	14.0	14.0	14.0	14.0	14.0	17.5	20.2
25	6.1	6.1	6.1	6.1	7.7	9.1	11.1	12.8
50	3.2	3.2	3.2	3.9	5.4	6.5	7.8	9.0
100	1.6	1.6	1.8	2.8	3.8	4.6	5.5	6.4
250	0.6	0.8	1.1	1.8	2.4	2.9	3.5	4.0
500	0.3	0.6	0.8	1.2	1.7	2.0	2.5	2.9
1,000	0.2	0.4	0.6	0.9	1.2	1.4	1.8	2.0
2,500	0.07	0.3	0.4	0.6	0.8	0.9	1.1	1.3
5,000	0.03	0.2	0.3	0.4	0.5	0.6	0.8	0.9
10,000	0.02	0.13	0.2	0.3	0.4	0.5	0.6	0.6
25,000	0.01	0.08	0.11	0.2	0.2	0.3	0.4	0.4
50,000	–	0.06	0.08	0.12	0.2	0.2	0.2	0.3
75,000	–	0.05	0.07	0.10	0.14	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 Some or All Plumbing Facilities and Households with Head of Spanish Origin: 1977

(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	32.2	32.2	32.2	32.2	32.2	32.2	32.2	34.4
10	19.2	19.2	19.2	19.2	19.2	19.2	21.1	24.3
25	8.7	8.7	8.7	8.7	9.2	11.0	13.3	15.4
50	4.5	4.5	4.5	4.7	6.5	7.8	9.4	10.9
100	2.3	2.3	2.3	3.4	4.6	5.5	6.7	7.7
250	0.9	1.0	1.4	2.1	2.9	3.5	4.2	4.9
500	0.5	0.7	1.0	1.5	2.1	2.5	3.0	3.4
1,000	0.2	0.5	0.7	1.1	1.5	1.7	2.1	2.4
2,500	0.09	0.3	0.4	0.7	0.9	1.1	1.3	1.5
5,000	0.05	0.2	0.3	0.5	0.7	0.8	0.9	1.1
10,000	0.02	0.2	0.2	0.3	0.5	0.5	0.7	0.8
25,000	0.01	0.10	0.14	0.2	0.3	0.3	0.4	0.5
50,000	–	0.07	0.10	0.2	0.2	0.2	0.3	0.3
75,000	–	0.06	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.

standard error of the estimated difference of 646,000 is about 74,000.

$$74{,}000 = \sqrt{(57{,}000)^2 + (47{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 646,000 difference is from 572,000 to 720,000 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 528,000 to 764,000 housing units, and the 95-percent confidence interval is from 498,000 to 794,000. Thus, we can conclude with 95 percent confidence that the number of 1977 owner-occupied housing units with garbage collection service once a week that had a family or primary individual income between $10,000 and $12,499, is greater than the number with an income between $5,000 and $6,999, 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 medians 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 family or primary individual income of owner-occupied housing units with garbage collection service once a week in the United States was $17,100 in 1977. The base of the distribution, from which this median was determined, is 22,346,000 housing units.

1. From table V, the standard error of a 50-percent characteristic on the base of 22,346,000 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 6 categories that 9,527,000 housing units, or 42.6 percent, had a family or primary individual income up to $14,999 and an additional 3,872,000 housing units, or 17.3 percent, had a family or primary individual income between $15,000 and $19,999. By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$\$15{,}000 + (\$20{,}000 - \$15{,}000)\frac{(49.2 - 42.6)}{17.3}$$
$$= \$16{,}900$$

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

$$\$15{,}000 + (\$20{,}000 - \$15{,}000)\frac{(50.8 - 42.6)}{17.3}$$
$$= \$17{,}400$$

Thus, the 95-percent confidence interval ranges from $16,900 to $17,400.

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 the 1977 AHS national sample.

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 1977 reinterview study were not available at the time of publication. However, it is expected that they will be similar to the results of the 1976 reinterview study which are presented in the Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample, 1976." Unlike the years prior to 1976, the respondent's answers in the reinterview were not reconciled to the original answers given in the AHS interview; i.e., after the question was answered in the reinterview, the interviewer did not present the previous responses and then ask the respondent to decide upon the best answer. Comparing the reinterview results of 1976 with the years prior to 1976, we found that the estimates of inconsistency of all items (nonattitudinal and attitudinal) increased substantially in the 1976 results. In other words, providing the interviewer with the original response had the effect of reducing the levels of inconsistency.

To summarize the results of the 1976 reinterview program: Overall, it showed moderate to high levels of inconsistency with about one-third of the nonattitudinal items and a high level of inconsistency with about one-third of the attitudinal items. Only one-eighth of the nonattitudinal items showed a low level of inconsistency.

The range for evaluating inconsistency is 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.

The following list shows the 1977 AHS-2 questionnaire numbers for those items asked in the 1976 reinterview: Nonattitudinal, sections IIIA and IIIB, 12*, 13*, 15a, 15b, 15c*, 36a, 36b*,

TABLE VII. Standard Errors of Estimated Percentages of Housing Units Pertaining to Each of the Four Regions; Northeast, North Central, South, and West: 1977 (Excluding Estimated Percentages of Housing Units Pertaining to Lacking Some or All Plumbing Facilities)

(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	28.7	28.7	28.7	28.7	28.7	28.7	28.7	31.8
10	16.8	16.8	16.8	16.8	16.8	16.8	19.4	22.5
25	7.5	7.5	7.5	7.5	8.5	10.1	12.3	14.2
50	3.9	3.9	3.9	4.4	6.0	7.2	8.7	10.0
100	2.0	2.0	2.0	3.1	4.3	5.1	6.2	7.1
250	0.8	0.9	1.3	2.0	2.7	3.2	3.9	4.5
500	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
1,000	0.2	0.4	0.6	1.0	1.3	1.6	1.9	2.2
2,500	0.08	0.3	0.4	0.6	0.9	1.0	1.2	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	0.4
50,000	–	0.06	0.09	0.14	0.2	0.2	0.3	0.3
75,000	–	0.05	0.07	0.11	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.

TABLE VIII. Standard Errors of Estimated Percentages of Housing Units Pertaining to Lacking Some or All Plumbing Facilities for Each of the Four Regions; Northeast, North Central, South, and West: 1977

(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	36.9	36.9	36.9	36.9	36.9	36.9	36.9	38.2
10	22.6	22.6	22.6	22.6	22.6	22.6	23.4	27.0
25	10.5	10.5	10.5	10.5	10.5	12.2	14.8	17.1
50	5.5	5.5	5.5	5.5	7.2	8.6	10.5	12.1
100	2.8	2.8	2.8	3.7	5.1	6.1	7.4	8.5
250	1.2	1.2	1.5	2.4	3.2	3.9	4.7	5.4
500	0.6	0.8	1.1	1.7	2.3	2.7	3.3	3.8
1,000	0.3	0.5	0.8	1.2	1.6	1.9	2.3	2.7
2,500	0.12	0.3	0.5	0.7	1.0	1.2	1.5	1.7
5,000	0.06	0.2	0.3	0.5	0.7	0.9	1.0	1.2
10,000	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.9
25,000	0.01	0.11	0.2	0.2	0.3	0.4	0.5	0.5
50,000	0.01	0.08	0.11	0.2	0.2	0.3	0.3	0.4
75,000	–	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.

37a, 37b*, 49*, 50*, 51a, 51b, 52a*, 54a*, 54b, 55a, 55b, 55c*, 56a, 56b, 58a*, 58b*, 61a*; Attitudinal, section IIIB, 102 all parts*, 103 all parts*, and 104 all parts*. Asterisks (*) indicate that the item had an estimated index of inconsistency greater than 40. Since cross tabulations involving these items may be subject to a large distortion because of the moderate to high response variance, they should be considered to be less reliable than comparable cross tabulations which do not involve these items. The cutoff at 40 was selected because (1) the shape of the distribution had a natural break before 40, (2) the large sampling errors on the estimated indices indicated little difference between those indices from 40 to 50 and those greater than 50, and (3) the break between moderate and high indices at 50 is arbitrary.

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 respondents 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 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 1977 AHS sample missed about 3.5 percent (i.e., about 375,000 housing units) of all conventional new construction built after April 1970, because the permits for these units, which were built before October 1977, 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 1977 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 were 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. However, biases of subtotals would still remain.

Rounding errors.—In 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. Also, since medians in this report were computed using unrounded data, instead of the shown published rounded data, they can differ from medians calculated directly from the published data.