Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-36
Selection of sample areas. App-36
Designation of sample housing units for the 1977 survey. App-36
Selection of the 1973 sample housing units. App-36
Splitting of the sample App-37
Selection of supplemental sample housing units in rural areas App-37
Selection of sampling housing units for the 1976 Coverage Improvement Program App-37
1977 sample reduction. App-38
1970 Census of Population and Housing. App-38
ESTIMATION. App-38
AHS national sample App-38
1978 housing inventory App-38
1973-1978 lost units App-39
Ratio estimation procedure of the 1970 Census of Population and Housing App-39
RELIABILITY OF THE ESTIMATES. App-39
Sampling errors App-39
Standard errors of estimates of levels. App-40
Standard errors of estimates of percentages App-40
Standard errors of ratios App-41
Illustrations of the use of the standard error tables. Illustration I App-41
Illustration II App-42
Differences. App-43
Illustration of the computation of the standard error of a difference. App-44
Medians App-44
Illustration of the computation of the 95-percent confidence interval for a median App-45
Nonsampling errors. App-45
1970 census App-45
Reinterview program App-46
Coverage errors App-47
Rounding errors. App-47

SAMPLE DESIGN

The 1978 estimates are based on data collected in October 1978 through January 1979 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 72,000 sample housing units (both occupied and vacant) were eligible for interview in the 1978 Annual Housing Survey. Of this number, 4,400 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 72,000, there were also 5,900 sample units which were visited but were ineligible for interview for the AHS in terms of collecting information relevant to the 1978 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 in sample with certainty. These 156 strata, mostly the larger SMSA's, were called self-representing (SR) because 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 1978 survey—The sample housing units designated to be interviewed in the 1978 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1977 survey (which included all sample housing units that were selected as part of the 1976 Coverage Improvement Program).
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 1977 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1978 AHS questionnaire, page App-16.)
3. All sample housing units that were selected from the list of building permits issued since the 1977 survey. (This sample represented the housing units built in permit-issuing areas, since the 1977 survey.)
4. Units added as the result of the updated listings in selected areas which do not issue building permits.

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 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:

$$\frac{\text{Number of HU's in the ED} + \frac{\text{Number of Group Quarters Persons in the ED}}{3}}{4}$$

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, the other half of each rural cluster (an expected two housing units) was reactivated in 1974. Similarly, for the area sampling frame, 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, for which construction had not been completed at the time of the 1970 census.
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. In the second stage, 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 sample 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 in 1977. 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 1978 housing inventory and estimates of units removed from the housing inventory between 1973 and 1978 (i.e., 1973-1978 lost units). Each type of estimate employed a separate, though similar, estimation procedure.

1978 housing inventory—In 1978, the AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the type A noninterview housing units encountered in the AHS. This noninterview adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

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

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

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

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

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

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

The second-stage ratio estimation factor was as follows:

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

The numerators of the ratios were derived from data based on the Survey of Construction (SOC). The denominators of the ratios were obtained from the weighted estimates for the AHS sample units, using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in each second-stage ratio estimation category.

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

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 current vacant housing estimates for 4 categories of vacant housing units and to independently derived current housing estimates for 24 categories of occupied housing units. Each of these 24 categories is a combination of the characteristics of residence, tenure, race of head, and sex of head.

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

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

The numerators of the ratios for occupied housing units were derived from data based on the Current Population Survey (CPS), a sample household survey conducted monthly by the Bureau of the Census. The numerators of the ratios for vacant housing units were derived from data based on the Housing Vacancy Survey (HVS), a quarterly vacancy survey also conducted by the Bureau of the Census. The denominators of the ratios were obtained from the weighted estimates for the AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit in each third-stage ratio estimation category.

The second-stage and the third-stage ratio estimation procedures were 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 18 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 17 categories of new construction units for each of the 4 regions (i.e., 11 categories for conventional new construction units and 6 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.

1973-1978 lost units—The 1973-1978 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 1973-1978 lost units do not include the HU's from the 1976 Coverage Improvement. Since the 1973-1978 lost units existed, by definition, in the 1973 housing inventory, there was a 1973 housing inventory weight associated with each 1973-1978 lost unit. This weight, adjusted for the 1977 reduction, was used to tabulate the estimates of the characteristics of the 1973-1978 lost units. 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 questionnaires, 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 result 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 1978 national housing inventory estimates in this report, and tables III and IV present the standard errors applicable to 1973-1978 lost housing unit estimates in this report. Table V presents the standard errors applicable to estimates for the Northeast, North Central, South and West regions. Linear interpolation should be used to determine standard errors for levels of estimates not specifically shown in tables I through V.

TABLE I. Standard Errors of Estimated Numbers of Housing Units: 1978 (Excluding Estimates of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water, Individual Well, Lacking Complete Plumbing, Mobile Homes, and Households With Head of Spanish Origin)

(68 chances out of 100)

Size of estimate (000)	Standard error		Size of estimate (000)	Standard error	
	Total or White (000)	Black (000)		Total or White (000)	Black (000)
0	2	2	1,000 . . .	42	39
5	3	3	2,500 . . .	65	55
10	4	4	5,000 . . .	91	57
25	7	7	10,000. . .	124	–
50	9	9	25,000. . .	174	–
100.	13	13	50,000. . .	185	–
250.	21	21	75,000. . .	109	–
500.	29	29			

TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water, Individual Well, Lacking Complete Plumbing, Mobile Homes, and Households With Head of Spanish Origin: 1978

(68 chances out of 100)

Size of estimate (000)	Standard error		Size of estimate (000)	Standard error	
	Total or White (000)	Black (000)		Total or White (000)	Black (000)
0	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. . .	215	–
250.	24	24	75,000. . .	126	–
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 reliabile than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more.

Tables VI through X present the standard errors of estimated percentages. Table VI 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 VII should be used for those specified items. Table VIII and IX show the approximate standard errors of the estimated percentages of 1973-1978 lost housing units. Table X shows the approximate standard error of all regional estimated percentages of housing units and 1973-1978 lost housing units. Two-way interpolation should be used to determined 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.

TABLE III. Standard Errors of Estimated Numbers of Lost Housing Units: 1973-1978 (Excluding Estimates of Lost Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Some or All Plumbing, Mobile Homes, and Other Vacants)

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	250	21
5	3	500	31
10	4	1,000	47
25	6	2,500	88
50	9	5,000	150
100	13		

TABLE IV. Standard Errors of Estimated Numbers of Lost Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Some or All Plumbing, Mobile Homes, and Other Vacants: 1973-1978

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	100	16
5	4	250	26
10	5	500	38
25	8	1,000	55
50	12		

TABLE Va. Standard Errors of Estimated Numbers of Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1978 (Excluding Estimates of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and Excluding Source of Water, Individual Well, and Mobile Homes for Each of the Regions)

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	500	31
5	3	1,000	43
10	4	2,500	68
25	7	5,000	97
50	10	10,000	137
100	14	25,000	216
250	22		

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 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 10,072,000 owner-occupied housing units with two persons in 1978. Interpolation in standard error table I shows that the standard error of an estimate of this size is approximately 124,000. The following procedure was used in interpolation.

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

Size of estimate (000)	Standard error (000)
10,000	124
10,072	x
25,000	174

By vertically interpolating between 124 and 174, the entry for "x" is determined to be 124.

$$10{,}072-10{,}000 = 72$$
$$25{,}000-10{,}000 = 15{,}000$$
$$124 + \frac{72}{15{,}000}(174-124) = 124$$

TABLE Vb. Standard Errors of Estimated Numbers of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and to Source of Water, Individual Well, and Mobile Homes for the Northeast and North Central Regions: 1978

(68 chances out of 100. For estimates pertaining to Source of Water, Individual Well, Cooking Fuel, and Mobile Homes for the West Region, apply a factor of 1.66 to the standard errors)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	500	36
5	4	1,000	51
10	5	2,500	80
25	8	5,000	111
50	11	10,000	153
100	16	25,000	218
250	26		

TABLE Vc. Standard Errors of Estimated Numbers of Housing Units Pertaining to Source of Water, Individual Well, and Mobile Homes for the South Region: 1978

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	4	500	45
5	5	1,000	64
10	6	2,500	99
25	10	5,000	136
50	14	10,000	181
100	20	25,000	225
250	32		

Consequently, the 68-percent confidence interval, as shown by these data, is from 9,948,000 to 10,196,000 housing units. Therefore, a conclusion that the average estimate of 1978 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 9,874,000 to 10,270,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 9,824,000 to 10,320,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 10,072,000 owner-occupied housing units with two persons inside SMSA's, 3,733,000 or 37.1 percent, were in central cities. Interpolation in standard error table VI (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage is 0.6 percentage points. The following procedure was used in interpolating.

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

Base of percentage (000)	Estimated percentage		
	25	37.1	50
10,000	0.6	a	0.7
10,072		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.

$$37.1-25.0 = 12.1$$
$$50.0-25.0 = 25.0$$
$$0.6 + \frac{12.1}{25.0}(0.7-0.6) = 0.6$$

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

$$37.1-25.0 = 12.1$$
$$50.0-25.0 = 25.0$$
$$0.4 + \frac{12.1}{25.0}(0.4-0.4) = 0.4$$

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

$$10{,}072-10{,}000 = 72$$
$$25{,}000-10{,}000 = 15{,}000$$
$$0.6 + \frac{72}{15{,}000}(0.4-0.6) = 0.6$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 36.5 to 37.7 percent; the 90-percent confidence interval is from 36.1 to 38.1 percent; and the 95-percent confidence interval is from 35.9 to 38.3 percent.

Illustration II—Table A-1 of this report shows that in the United States in 1978 there were 130,000 housing units in structures with 4 floors or more (see "Elevator in Structure" item) that

were outside of SMSA's. Interpolation in standard error table I of this appendix shows that the standard error of an estimate of this size is approximately 15,000. Consequently, the 68-percent confidence interval is from 115,000 to 145,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1978 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 106,000 to 154,000 housing units with 90-percent confidence; and that the average estimate lies within the interval from 100,000 to 160,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 130,000 1978 housing units in structures with 4 floors or more that were outside SMSA's, 117,000, or 90.0 percent, were in structures that contained elevators. Interpolation in table VI (i.e., interpolation on both the base and the percent) of this appendix shows that the standard error of the above percentage is 3.7 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 86.3 to 93.7 percent; the 90-percent confidence interval is from 84.1 to 95.9 percent; and the 95-percent confidence interval is from 82.6 to 97.4 percent.

Differences—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the

TABLE Vd. Standard Errors of Estimated Numbers of Lost Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1973-1978 (Excluding Estimates of Lost Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Some or All Plumbing, Mobile Homes, and Other Vacants)

(68 chances out of 100)

Size of estimate (000)	Standard errors	
	Northeast or North Central	South or West
0	2	2
5	3	3
10	4	4
25	6	6
50	8	9
100	12	14
250	20	22
500	29	32
1,000	44	49
2,500	82	92
5,000	140	156

Note: For standard errors of regional estimates of lost housing units (1973-1978) pertaining to new construction, lacking complete kitchen facilities, no bedrooms, no bathrooms, lacking some or all plumbing, mobile homes, and other vacants, use the national standard errors presented in table IV.

TABLE VI. Standard Errors of Estimated Percentages of Housing Units: 1978 (Excluding Estimated Percentages of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water, Individual Well, Lacking Complete Plumbing, Mobile Homes, and Households With Head of Spanish Origin)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	25.9	25.9	25.9	25.9	25.9	25.9	25.9	29.5
10	14.8	14.8	14.8	14.8	14.8	14.9	18.1	20.9
25	6.5	6.5	6.5	6.5	7.9	9.4	11.4	13.2
50	3.4	3.4	3.4	4.1	5.6	6.7	8.1	9.3
100	1.7	1.7	1.8	2.9	4.0	4.7	5.7	6.6
250	0.7	0.8	1.2	1.8	2.5	3.0	3.6	4.2
500	0.3	0.6	0.8	1.3	1.8	2.1	2.6	3.0
1,000	0.2	0.4	0.6	0.9	1.3	1.5	1.8	2.1
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	0.6	0.7	0.8	0.9
10,000	0.02	0.13	0.2	0.3	0.4	0.5	0.6	0.7
25,000	0.01	0.08	0.12	0.2	0.3	0.3	0.4	0.4
50,000	–	0.06	0.08	0.13	0.2	0.2	0.3	0.3
75,000	–	0.05	0.07	0.11	0.14	0.2	0.2	0.2

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 of this report shows that inside SMSA's in the United States there were 5,878,000 owner-occupied housing units with three persons in 1978. Thus, the apparent difference between the number of 1978 owner-occupied housing units with two persons and those with three persons is 4,194,000. The standard error of 10,072,000 is approximately 124,000. Interpolation in standard error table I shows the standard error on an estimate of 5,878,000 to be approximately 97,000. Therefore, the standard error of the estimated difference of 4,194,000 is about 157,000.

$$157{,}000 = \sqrt{(124{,}000)^2 + (97{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 4,194,000 difference is from 4,037,000 to 4,351,000 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 3,943,000 to 4,445,000 housing units, and the 95-percent confidence interval is from 3,880,000 to 4,508,000. Thus, we can conclude with 95-percent confidence that the number of 1978 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 errror 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 VII. Standard Errors of Estimated Percentages of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water, Individual Well, Lacking Complete Plumbing, Mobile Homes, and Households With Head of Spanish Origin: 1978

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	32.0	32.0	32.0	32.0	32.0	32.0	32.0	34.3
10	19.1	19.1	19.1	19.1	19.1	19.1	21.0	24.3
25	8.6	8.6	8.6	8.6	9.2	11.0	13.3	15.3
50	4.5	4.5	4.5	4.7	6.5	7.7	9.4	10.8
100	2.3	2.3	2.3	3.3	4.6	5.5	6.6	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	2.4	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	0.15	0.2	0.2	0.3	0.3
75,000	–	0.06	0.08	0.12	0.2	0.2	0.2	0.3

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 inside SMSA's was 2.7 in 1978. The base of the distribution from which this median was determined is 32,390,000 housing units.

1. From table VI, the standard error of a 50-percent characteristic on the base of 32,390,000 is 0.4 percentage points.
2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 49.2 and 50.8.
3. From table A-1, it can be seen by cumulating the frequencies for the first two categories that 14,804,000 owner-occupied housing units inside SMSA's, or 45.7 percent, had one and two persons (actually, for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 5,878,000 owner-occupied housing units, or 18.1 percent, had three persons (i.e., 2.5 to 3.5 persons). By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about 2.7.

$$2.5 + (3.5-2.5)\,\frac{(49.2-45.7)}{18.1} = 2.7$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about 2.8.

$$2.5 + (3.5-2.5)\,\frac{(50.8-45.7)}{18.1} = 2.8$$

Thus, the 95-percent confidence interval ranges from 2.7 to 2.8 persons. Although it appears that this confidence interval has the sample estimate as the upper limit, it actually is a reflection of the rounding error associated with the median (see the paragraph on rounding errors in the nonsampling errors section of this appendix).

Nonsampling errors—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases; definitional difficulties; differences in the interpretation of questions; inability or unwillingness to provide correct information on the part of respondents; mistakes in recording or coding the data; and other errors of collection, response, processing, coverage, and estimation for missing data. 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 1970 Census of Population and Housing and the 1978 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 the extent to which occupancy status was erroneously reported. 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.**

TABLE VIII. Standard Errors of Estimated Percentages of Lost Housing Units: 1973-1978 (Excluding Estimated Percentages of Lost Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Some or All Plumbing, Mobile Homes, and Other Vacants)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	24.5	24.5	24.5	24.5	24.5	24.5	24.7	28.5
10	14.0	14.0	14.0	14.0	14.0	14.4	17.5	20.2
25	6.1	6.1	6.1	6.1	7.6	9.1	11.0	12.7
50	3.1	3.1	3.1	3.9	5.4	6.4	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	1.7	2.0
2,500	0.06	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

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 1978 reinterview study were not available at the time of publication; however, it is expected that they will be similar to the results of the 1977 reinterview study which are presented in the Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample 1977."

The results of the 1977 reinterview study seem to indicate that there were some very slight improvements in the consistency of response from the 1976 to the 1977 survey. Overall, the 1977 reinterview program showed moderate to high levels of inconsistency with about 21 percent of the nonattitudinal and 56 percent of the attitudinal items showing high levels of inconsistency. But a large proportion (43 percent) of the nonattitudinal items showed a low level of inconsistency. Moderate levels indicate that there are some problems with inconsistent reporting and high levels indicate that improvements are needed in the data collection methods or that the category concepts themselves are ambiguous.

Cross-tabulations involving those items which are subject to substantial levels of inconsistency may be subject to a large distortion as a consequence of the associated high level of response variance, and thus, are considered to be less reliable than comparable cross-tabulations which do not involve these data. As a consequence, the tables, in this report, which contain such cross-tabulations have been footnoted with a cautionary statement.

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

A possible explanation for the results of the AHS and census reinterview studies, as well as the surveys themselves, is that the data are based on the answers given by the respondents, who may lack precise information. Also, because the results of the reinterview studies are derived from sample surveys, there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

TABLE IX. Standard Errors of Estimated Percentages of Lost Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Some or All Plumbing, Mobile Homes, and Other Vacants: 1973-1978

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	34.9	34.9	34.9	34.9	34.9	34.9	34.9	36.6
10	21.1	21.1	21.1	21.1	21.1	21.1	22.4	25.9
25	9.7	9.7	9.7	9.7	9.8	11.7	14.2	16.4
50	5.1	5.1	5.1	5.1	6.9	8.3	10.0	11.6
100	2.6	2.6	2.6	3.6	4.9	5.8	7.1	8.2
250	1.1	1.1	1.4	2.3	3.1	3.7	4.5	5.2
500	0.5	0.7	1.0	1.6	2.2	2.6	3.2	3.7
1,000	0.3	0.5	0.7	1.1	1.6	1.8	2.2	2.6
2,500	0.11	0.3	0.5	0.7	1.0	1.2	1.4	1.6
5,000	0.05	0.2	0.3	0.5	0.7	0.8	1.0	1.2

Coverage errors—Deficiencies in the representation of conventional new construction for the AHS new construction sample (mentioned previously in the section on estimation) is an example of coverage errors. 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 1978 AHS sample missed about 3.3 percent (i.e., about 375,000 units) of conventional housing units built after April 1970, because the permits for these units, which were built before October 1978, were issued less than 5 months in advance of the survey. The second stage of ratio estimation procedure was employed to reduce the effect of this deficiency although some bias in the AHS estimates of conventional new construction probably still exists.

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 1978 AHS sample missed as much as 2 percent (i.e., as much as 400,000 units) of all housing units in ED's where area sampling methods are used, because these units are not listed during the canvassing.

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

Rounding errors—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 published rounded data, they can differ from medians calculated directly from the published data.

TABLE Xa. Standard Errors of Estimated Percentages of Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1978 (Excluding Estimated Percentages Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing Facilities for the Northeast, North Central, and West Regions and Excluding Source of Water, Individual Well, and Mobile Homes for Each of the Regions)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	27.3	27.3	27.3	27.3	27.3	27.3	27.3	30.6
10	15.8	15.8	15.8	15.8	15.8	15.8	18.7	21.6
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	1.8	1.8	1.9	3.0	4.1	4.9	5.9	6.8
250	0.7	0.9	1.2	1.9	2.6	3.1	3.7	4.3
500	0.4	0.6	0.9	1.3	1.8	2.2	2.7	3.1
1,000	0.2	0.4	0.6	0.9	1.3	1.5	1.9	2.2
2,500	0.07	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

TABLE Xb. Standard Errors of Estimated Percentages of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing Facilities for the Northeast, North Central, and West Regions and to Source of Water, Individual Well, and Mobile Homes for the Northeast and North Central Regions: 1978

(68 chances out of 100. For estimated percentages pertaining to Source of Water, Individual Well, and Mobile Homes for the West Region, apply a factor of 1.66 to the standard errors)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	34.3	34.3	34.3	34.3	34.3	34.3	34.3	36.1
10	20.7	20.7	20.7	20.7	20.7	20.7	22.1	25.5
25	9.5	9.5	9.5	9.5	9.7	11.5	14.0	16.2
50	5.0	5.0	5.0	5.0	6.9	8.2	9.9	11.4
100	2.5	2.5	2.5	3.5	4.8	5.8	7.0	8.1
250	1.0	1.0	1.4	2.2	3.1	3.6	4.4	5.1
500	0.5	0.7	1.0	1.6	2.2	2.6	3.1	3.6
1,000	0.3	0.5	0.7	1.1	1.5	1.8	2.2	2.6
2,500	0.10	0.3	0.5	0.7	1.0	1.2	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	0.4	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

TABLE Xc. Standard Errors of Estimated Percentages of Housing Units Pertaining to Source of Water, Individual Well, and Mobile Homes for the South Region: 1978

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	45.2	45.2	45.2	45.2	45.2	45.2	45.2	45.4
10	29.2	29.2	29.2	29.2	29.2	29.2	29.2	32.1
25	14.1	14.1	14.1	14.1	14.1	14.5	17.6	20.3
50	7.6	7.6	7.6	7.6	8.6	10.2	12.4	14.4
100	4.0	4.0	4.0	4.4	6.1	7.2	8.8	10.1
250	1.6	1.6	1.8	2.8	3.9	4.6	5.6	6.4
500	0.8	0.9	1.3	2.0	2.7	3.2	3.9	4.5
1,000	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
2,500	0.2	0.4	0.6	0.9	1.2	1.4	1.8	2.0
5,000	0.08	0.3	0.4	0.6	0.9	1.0	1.2	1.4
10,000	0.04	0.2	0.3	0.4	0.6	0.7	0.9	1.0
25,000	0.02	0.13	0.2	0.3	0.4	0.5	0.6	0.6

TABLE Xd. Standard Errors of Estimated Percentages of Lost Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1973-1978 (Excluding Estimates of Lost Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Some or All Plumbing, Mobile Homes, and Other Vacants)

(68 chances out of 100)

Size of estimate (000)	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
	Estimated percentages for the Northeast or North Central Regions							
5	22.8	22.8	22.8	22.8	22.8	22.8	22.8	26.7
10	13.0	13.0	13.0	13.0	13.0	13.4	16.3	18.8
25	5.7	5.7	5.7	5.7	7.1	8.5	10.2	11.8
50	2.9	2.9	2.9	3.6	5.0	6.0	7.3	8.4
100	1.5	1.5	1.7	2.6	3.5	4.3	5.1	6.0
250	0.6	0.7	1.0	1.7	2.2	2.7	3.3	3.7
500	0.3	0.6	0.7	1.1	1.6	1.9	2.3	2.7
1,000	0.2	0.4	0.6	0.8	1.1	1.3	1.6	1.9
2,500	0.06	0.3	0.4	0.6	0.7	0.8	1.0	1.2
5,000	0.03	0.2	0.3	0.4	0.5	0.6	0.7	0.8
	Estimated percentages for the South or West Regions							
5	25.5	25.5	25.5	25.5	25.5	25.5	25.5	29.6
10	14.6	14.6	14.6	14.6	14.6	15.0	18.2	21.0
25	6.3	6.3	6.3	6.3	7.9	9.5	11.4	13.2
50	3.2	3.2	3.2	4.1	5.6	6.7	8.1	9.4
100	1.7	1.7	1.9	2.9	4.0	4.8	5.7	6.7
250	0.6	0.8	1.1	1.9	2.5	3.0	3.6	4.2
500	0.3	0.6	0.8	1.2	1.8	2.1	2.6	3.0
1,000	0.2	0.4	0.3	0.9	1.2	1.5	1.8	2.1
2,500	0.06	0.3	0.4	0.6	0.8	0.9	1.1	1.4
5,000	0.03	0.2	0.3	0.4	0.5	0.6	0.8	0.9

Note: For standard errors of regional estimated percentages of lost housing units (1973-1978) pertaining to new construction, lacking complete kitchen facilities, no bedrooms, no bathrooms, lacking some or all plumbing, mobile homes, and other vacants, use the national standard errors presented in table IX.

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-40
Selection of sample areas App-40
Designation of sample housing units for the 1978 survey App-40
Selection of the 1973 sample housing units App-40
Splitting of the sample App-41
Selection of supplemental sample housing units in rural areas App-41
Selection of sample housing units for the 1976 Coverage Improvement Program App-41
1977 SAMPLE REDUCTION App-42
ESTIMATION App-42
RELIABILITY OF THE ESTIMATES App-43
Sampling Errors App-43
Standard errors of estimates of levels App-43
Standard errors of estimates of percentages App-44
Standard errors of ratios App-44
Illustration of the use of the standard error tables .
Illustration I App-44
Illustration II App-45
Differences App-45
Illustration of the computation of the standard error of a difference App-46
Medians App-46
Illustration of the computation of the 95-percent confidence interval for a median App-46
Nonsampling errors App-47
Reinterview Program App-47
Coverage errors App-48
Rounding errors App-49

SAMPLE DESIGN

The 1978 estimates are based on data collected in October 1978 through January 1979 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 72,000 sample housing units (both occupied and vacant) were eligible for interview in the 1978 Annual Housing Survey. Of this number, 4,400 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 72,000, there were also 5,900 sample units which were visited but were ineligible for interview for the AHS in terms of collecting information relevant to the 1978 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 in sample with certainty. These 156 strata, mostly the larger SMSA's, were called self-representing (SR) because 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 1978 survey—The sample housing units designated to be interviewed in the 1978 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1977 survey (which included all sample housing units that were selected as part of the 1976 Coverage Improvement Program).
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 1977 survey and which were not part of the 1977 sample reduction. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1978 AHS questionnaire, page App-20.)
3. All sample housing units that were selected from the list of building permits issued since the 1977 survey. (This sample represented the housing units built in permit-issuing areas, since the 1977 survey.)
4. Units added as the result of the updated listings in selected areas which do not issue building permits.

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 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, the other half of each rural cluster (an expected two housing units) was reactivated in 1974. Similarly, for the area sampling frame, 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 for which construction had not been completed at the time of the 1970 census.
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. In the second stage, 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 SAMPLE 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 in 1977. 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.

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:

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

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.

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

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 current vacant housing estimates for 4 categories of vacant housing units and to independently derived current housing estimates for 24 categories of occupied housing units. Each of these 24 categories is a combination of the characteristics of residence, tenure, race of head, and sex of head.

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

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

The numerators of the ratios for occupied housing units were derived from data based on the Current Population Survey

(CPS), a sample household survey conducted monthly by the Bureau of the Census. The numerators of the ratios for vacant housing units were derived from data based on the Housing Vacancy Survey (HVS), a quarterly vacancy survey also conducted by the Bureau of the Census. The denominators of the ratios were obtained from the weighted estimates for the AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit in each third-stage ratio estimation category.

The second-stage and the third-stage ratio estimation procedures were 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 18 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 17 categories of new construction units for each of the 4 regions (i.e., 11 categories for conventional new construction units and 6 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 questionnaires, 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 result 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 were surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated for each sample, then:

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

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

The figures presented in the 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 1978 national housing inventory estimates in this report. Tables III and IV present the standard errors applicable to estimates for the Northeast, North Central, South, and West Regions. 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 through 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 IV. Table VIII should be used for those specified items for the Northeast, North Central, and West Regions. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables V through VIII.

TABLE I. Standard Errors of Estimated Numbers of Housing Units: 1978 (Excluding Estimates of Housing Units Pertaining to Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Complete Plumbing, and Households With Head of Spanish Origin)

(68 chances out of 100)

Size of estimate (000)	Standard error		Size of estimate (000)	Standard error	
	Total or White (000)	Black (000)		Total or White (000)	Black (000)
0	2	2	1,000 . . .	42	39
5	3	3	2,500 . . .	65	55
10	4	4	5,000 . . .	91	57
25	7	7	10,000. . .	124	—
50	9	9	25,000. . .	174	—
100.	13	13	50,000. . .	185	—
250.	21	21	75,000. . .	109	—
500.	29	29			

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.

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:

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

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-2 (section I) of this report shows that in the United States there were 14,111,000 renter-occupied housing units with common stairways. Interpolation in standard error table I shows that the standard error of an estimate of this size is approximately 138,000. The following procedure was used in interpolating:

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

Size of estimate (000)	Standard error (000)
10,000	124
14,111	x
25,000	174

By vertically interpolating between 124 and 174, the entry for "x" is determined to be 138.

$$14{,}111 - 10{,}000 = 4{,}111$$
$$25{,}000 - 10{,}000 = 15{,}000$$

$$124 + \frac{4{,}111}{15{,}000}(174 - 124) = 138$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 13,973,000 to 14,249,000 housing units. Therefore, a conclusion that the average estimate of 1978 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 13,890,000 to 14,332,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 13,835,000 to 14,387,000 housing units with 95 percent confidence.

Table A-2 (section I) also shows that of the 14,111,000 renter-occupied housing units with common stairways, 12,505,000 or 88.6 percent, were located inside SMSA's.

Interpolation in table V (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage is 0.4 percentage points. The following procedure was used in interpolating.

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

Base of percentage (000)	Estimated percentage		
	85	88.6	90
10,000	0.5	a	0.4
14,111		p	
25,000	0.3	b	0.3

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

$$88.6-85.0 = 3.6$$
$$90.0-85.0 = 5.0$$
$$0.5 + \frac{3.6}{5.0}(0.4-0.5) = 0.4$$

2. Horizontal interpolation between 0.3 and 0.3 is not necessary.

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

$$14{,}111-10{,}000 = 4{,}111$$
$$25{,}000-10{,}000 = 15{,}000$$
$$0.4 + \frac{4{,}111}{15{,}000}(0.3-0.4) = 0.4$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 88.2 to 89.0 percent; the 90-percent confidence interval is from 88.0 to 89.2 percent; and the 95-percent confidence interval is from 87.8 to 89.4 percent.

Illustration II—Table A-3 (section I) of this report shows that in the United States in 1978 there were 1,312,000 owner-occupied housing units which had water supply breakdowns. Interpolation in standard error table I shows that the standard error of an estimate of this size is approximately 47,000. Consequently, the 68-percent confidence interval is from 1,265,000 to 1,359,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1978 owner-occupied housing units which had water supply breakdowns lies within a range computed in this way and would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples, lies within the interval from 1,237,000 to 1,387,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 1,218,000 to 1,406,000 housing units with 95 percent confidence.

Table A-3 (section I) also shows that of the 1,312,000 1978 owner-occupied housing units which had water supply breakdowns, 220,000 or 16.8 percent, had breakdowns 3 times or more. Interpolation in standard error table V (i.e., interpolation on both the base and the percent) shows that the standard error of the above percentage is 1.4 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 15.4 to 18.2 percent; the 90-percent confidence interval is from 14.6 to 19.0 percent; and the 95-percent confidence interval is from 14.0 to 19.6 percent.

Differences—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately

TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Complete Plumbing, and Households With Head of Spanish Origin: 1978

(68 chances out of 100)

Size of estimate (000)	Standard error		Size of estimate (000)	Standard error	
	Total, White or Spanish origin (000)	Black (000)		Total, White or Spanish origin (000)	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. .	215	–
250. . . .	24	24	75,000. .	126	–
500. . . .	34	33			

TABLE III. Standard Errors of Estimated Numbers of Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1978 (Excluding Estimates of Housing Units Pertaining to Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions)

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	500.	31
5	3	1,000	43
10	4	2,500	68
25	7	5,000	97
50	10	10,000.	137
100.	14	25,000.	216
250.	22		

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.

TABLE IV. Standard Errors of Estimated Numbers of Housing Units Pertaining to Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions: 1978

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	500	36
5	4	1,000	51
10	5	2,500	80
25	8	5,000	111
50	11	10,000	153
100	16	25,000	218
250	26		

Illustration of the computation of the standard error of a difference—Table A-3 (section I) of this report shows that in the United States in 1978 there were 913,000 owner-occupied housing units, which had exactly 1 water supply breakdown and 220,000 owner-occupied housing units which had 3 or more water supply breakdowns. Thus, the apparent difference between the number of 1978 owner-occupied housing units that had breakdowns 3 times or more and that had breakdowns just 1 time, is 693,000. Interpolation in standard error table I shows that the standard error on an estimate of 913,000 to be approximately 40,000 and the standard error on an estimate of 220,000 to be approximately 19,000. Therefore, the standard error of the estimated difference of 693,000 is about 44,000.

$$44{,}000 = \sqrt{(40{,}000)^2 + (19{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 693,000 difference is from 649,000 to 737,000 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 and would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 623,000 to 763,000 housing units, and the 95-percent confidence interval is from 605,000 to 781,000. Thus, we can conclude with 95 percent confidence that the number of 1978 owner-occupied housing units, which had three or more water supply breakdowns, is different than the number that had exactly one water supply 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 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 our 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-3 (section II) of this report shows the median value of specified owner-occupied housing units occupied less than 3 months was $18,500 in 1978. The base of the distribution, from which this median was determined is 1,540,000 housing units.

1. From standard error table V, the standard error of a 50-percent characteristic on the base of 1,540,000 is 1.8 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 46.4 and 53.6.
3. From table A-3 (section II), it can be seen by cumulating the frequencies for the first 4 categories that 566,000 specified owner-occupied housing units or 36.8 percent, had a value less than $15,000 and that an additional 288,000 specified owner-occupied housing units, or 18.7 percent, had a value of $15,000 to $19,999. By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about $17,600.

$$\$14{,}999 + (\$19{,}999 - \$14{,}999)\ \frac{(46.4-36.8)}{18.7} = \$17{,}600$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about $19,500.

$$\$14{,}999 + (\$19{,}999 - \$14{,}999)\ \frac{(53.6 - 36.8)}{18.7} = \$19{,}500$$

Thus, the 95-percent confidence interval ranges from $17,600 to $19,500.

Nonsampling errors—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, and estimation for missing data. 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 1978 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 1978 reinterview study were not available at the time of publication; however, it is expected that they will be similar to the results of the 1977 reinterview study which are presented in the Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample, 1977."

The results of the 1977 reinterview study seem to indicate that there are some very slight improvements in the consistency of response from 1976 to 1977. Overall, the 1977 reinterview program showed moderate to high levels of inconsistency with about 21 percent of the nonattitudinal and 56 percent of the attitudinal items showing high levels of inconsistency. But a large proportion (43 percent) of the nonattitudinal items showed a low level of inconsistency. Moderate levels indicate that there are some problems with inconsistent reporting and

TABLE V. Standard Errors of Estimated Percentages of Housing Units: 1978 (Excludes Estimated Percentages of Housing Units Pertaining to Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Complete Plumbing, and Households With Head of Spanish Origin)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	25.9	25.9	25.9	25.9	25.9	25.9	25.9	29.5
10	14.8	14.8	14.8	14.8	14.8	14.9	18.1	20.9
25	6.5	6.5	6.5	6.5	7.9	9.4	11.4	13.2
50	3.4	3.4	3.4	4.1	5.6	6.7	8.1	9.3
100	1.7	1.7	1.8	2.9	4.0	4.7	5.7	6.6
250	0.7	0.8	1.2	1.8	2.5	3.0	3.6	4.2
500	0.3	0.6	0.8	1.3	1.8	2.1	2.6	3.0
1,000	0.2	0.4	0.6	0.9	1.3	1.5	1.8	2.1
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	0.6	0.7	0.8	0.9
10,000	0.02	0.13	0.2	0.3	0.4	0.5	0.6	0.7
25,000	0.01	0.08	0.12	0.2	0.3	0.3	0.4	0.4
50,000	–	0.06	0.08	0.13	0.2	0.2	0.3	0.3
75,000	–	0.05	0.07	0.11	0.14	0.2	0.2	0.2

high levels indicate that improvements are needed in the data collection methods or that the category concepts themselves are ambiguous.

Cross-tabulations involving those items, which are subject to substantial levels of inconsistency, may be subject to a large distortion as a consequence of the associated high level of response variance and, thus, are considered to be less reliable than comparable cross-tabulations which do not involve these data. As a consequence, the tables in this report which contain such cross-tabulations have been footnoted with a cautionary statement.

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

A possible explanation for the results of the AHS and census reinterview studies, as well as the surveys themselves, is that the 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—Deficiences in the representation of conventional new construction for the AHS new construction sample (mentioned previously in the section on estimation) is an example of coverage errors. 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 1978 AHS sample missed about 3.3 percent (i.e., about 375,000 units) of conventional housing units built after April 1970, because the permits for these units, which were built before October 1978, were issued less than 5 months in advance of the survey. The second stage of the ratio estimation procedure was employed to reduce the effect of this deficiency although some bias in the AHS estimates of conventional new construction probably still exists.

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 1978 AHS sample missed as much as 2 percent (i.e., as much as 400,000 units) of all housing units in ED's where area sampling methods are used because these units are not listed during the canvassing.

TABLE VI. Standard Errors of Estimated Percentages of Housing Units Pertaining to Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Complete Plumbing, and Households With Head of Spanish Origin: 1978

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	32.0	32.0	32.0	32.0	32.0	32.0	32.0	34.3
10	19.1	19.1	19.1	19.1	19.1	19.1	21.0	24.3
25	8.6	8.6	8.6	8.6	9.2	11.0	13.3	15.3
50	4.5	4.5	4.5	4.7	6.5	7.7	9.4	10.8
100	2.3	2.3	2.3	3.3	4.6	5.5	6.6	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	2.4	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	0.15	0.2	0.2	0.3	0.3
75,000	–	0.06	0.08	0.12	0.2	0.2	0.2	0.3

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

Rounding errors—With respect to errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of this survey. Also, since medians in this report were computed using unrounded data, instead of the published rounded data, they can differ from medians calculated directly from the published data.

TABLE VII. Standard Errors of Estimated Percentages of Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1978 (Excluding Estimated Percentages Pertaining to Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	27.3	27.3	27.3	27.3	27.3	27.3	27.3	30.6
10	15.8	15.8	15.8	15.8	15.8	15.8	18.7	21.6
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	1.8	1.8	1.9	3.0	4.1	4.9	5.9	6.8
250	0.7	0.9	1.2	1.9	2.6	3.1	3.7	4.3
500	0.4	0.6	0.9	1.3	1.8	2.2	2.7	3.1
1,000	0.2	0.4	0.6	0.9	1.3	1.5	1.9	2.2
2,500	0.07	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

TABLE VIII. Standard Errors of Estimated Percentages of Housing Units Pertaining to Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions: 1978

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	34.3	34.3	34.3	34.3	34.3	34.3	34.3	36.1
10	20.7	20.7	20.7	20.7	20.7	20.7	22.1	25.5
25	9.5	9.5	9.5	9.5	9.7	11.5	14.0	16.2
50	5.0	5.0	5.0	5.0	6.9	8.2	9.9	11.4
100	2.5	2.5	2.5	3.5	4.8	5.8	7.0	8.1
250	1.0	1.0	1.4	2.2	3.1	3.6	4.4	5.1
500	0.5	0.7	1.0	1.6	2.2	2.6	3.1	3.6
1,000	0.3	0.5	0.7	1.1	1.5	1.8	2.2	2.6
2,500	0.10	0.3	0.5	0.7	1.0	1.2	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	0.4	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

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-36
Selection of sample areas. App-36
Designation of sample housing units for the 1978 survey. App-36
Selection of the 1973 sample housing units. App-36
Splitting of the sample App-37
Selection of supplemental sample housing units in rural areas App-37
Selection of sample housing units for the 1976 Coverage Improvement Program App-37
1977 SAMPLE REDUCTION App-38
ESTIMATION. App-38
RELIABILITY OF THE ESTIMATES. App-39
Sampling errors App-39
Standard errors of estimates of levels App-39
Standard errors of estimates of percentages. App-40
Standard errors of ratios App-41
Illustrations of the use of the standard error tables. Illustration I App-41
Illustration II App-42
Differences. App-42
Illustration of the computation of the standard error of a difference. App-43
Medians App-43
Illustration of the computation of the 95-percent confidence interval for a median . App-44
Nonsampling errors App-44
Reinterview program App-44
Coverage errors App-45
Rounding errors. App-46

SAMPLE DESIGN

The 1978 estimates are based on data collected in October 1978 through January 1979 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 72,000 sample housing units (both occupied and vacant) were eligible for interview in the 1978 Annual Housing Survey. Of this number, 4,400 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 72,000 eligible housing units, there were also 5,900 sample units which were visited but were ineligible for interview for the AHS in terms of collecting information relevant to the 1978 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 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 1978 survey—The sample housing units designated to be interviewed in the 1978 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1977 survey (which included all sample housing units that were selected as part of the 1976 Coverage Improvement Program).
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 1977 survey. (For a list of reasons for type A and type B noninterviews, see the facsimilie of the 1978 AHS questionnaire, page App-16.)
3. All sample housing units that were selected from the list of building permits issued since the 1977 survey. (This sample represented the housing units built in permit-issuing areas since the 1977 survey.)
4. Units added as the result of the updated listings in selected areas which do not issue building permits.

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, the other half of each rural cluster (an expected two housing units) was reactived in 1974. Similarly for the area sampling frame, 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 for which construction had not been completed at the time of the 1970 census.
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. In the second stage, 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 SAMPLE 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 in 1977. 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.

ESTIMATION

In 1978, the AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the type A noninterview housing units encountered in the AHS. This noninterview adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

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

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

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

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 deficience 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 current vacant housing estimates for 4 categories of vacant housing units and to independently derived current housing estimates for 24 categories of occupied housing units. Each of these 24 categories is a combination of the characteristics of residence, tenure, race of head, and sex of head.

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

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

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

housing units were derived from data based on the Housing Vacancy Survey (HVS), a quarterly vacancy survey also conducted by the Bureau of the Census. The denominators of the ratios were obtained from the weighted estimates for the AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit in each third-stage ratio estimation category.

The second-stage and the third-stage ratio estimation procedures were 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 18 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 17 categories of new construction units for each of the 4 regions (i.e., 11 categories for conventional new construction units and 6 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 third-stage ratio estimation procedure, as well as the overall estimation procedure, reduced 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 result of all possible samples. In addition, the standard error, as calculated for this report, also 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 1978 national housing inventory estimates in this report. Tables III and IV present the standard errors applicable to estimates for the Northeast, North Central and West regions, and tables III and V present the standard errors applicable to estimates for the South Region. 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 sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more.

TABLE I. Standard Errors of Estimated Numbers of Housing Units: 1978 (Excluding Estimates of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water, Individual Well, Lacking Complete Plumbing, 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	42	39
5	3	3	2,500	65	55
10	4	4	5,000	91	57
25	7	7	10,000	124	—
50	9	9	25,000	174	—
100	13	13	50,000	185	—
250	21	21	75,000	109	—
500	29	29			

TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water, Individual Well, Lacking Complete Plumbing, Mobile Homes, and Households with Head of Spanish Origin: 1978

(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	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	215	—
250	24	24	75,000	126	—
500	34	33			

TABLE III. Standard Errors of Estimated Numbers of Housing Units Pertaining to the Northeast, North Central, South and West Regions: 1978 (Excluding Estimates of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and Excluding Source of Water, Individual Well, and Mobile Homes for Each of the Regions)

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	500	31
5	3	1,000	43
10	4	2,500	68
25	7	5,000	97
50	10	10,000	137
100	14	25,000	216
250	22		

TABLE IV. Standard Errors of Estimated Numbers of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing for the Northeast, North Central and West Regions and to Source of Water, Individual Well, and Mobile Homes for the Northeast and North Central Regions: 1978

(68 chances out of 100. For estimates pertaining to source of water, individual well, cooking fuel, and mobile homes for the West Region, apply a factor of 1.66 to the standard errors listed below)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	500	36
5	4	1,000	51
10	5	2,500	80
25	8	5,000	111
50	11	10,000	153
100	16	25,000	218
250	26		

Tables VI through 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 table II. The standard errors shown in table VII should be used for those specified items. Table VIII shows the approximate standard errors of the estimated percentages of housing units for the Northeast, North Central, South, and West regions except for those percentages pertaining to the specified items in table IV. Table IX should

be used for those specified items for the Northeast, North Central, and West regions and table X for the South Region. 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) $(\frac{x}{y})$, where x is not a subclass of y, tables VI through X underestimate the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be obtained by letting the standard error of the ratio be approximately equal to:

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

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

Illustrations of the use of the standard error tables. Illustration I—Table A-2 of this report shows that in the United States there were 9,446,000 specified owner-occupied housing units with two bedrooms in 1978. Interpolation of the standard error in table I shows that the standard error of an estimate of this size is approximately 120,000. The following procedure was used in interpolating.

TABLE V. Standard Errors of Estimated Numbers of Housing Units Pertaining to Source of Water, Individual Well, and Mobile Homes for the South Region: 1978

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	4	500	45
5	5	1,000	64
10	6	2,500	99
25	10	5,000	136
50	14	10,000	181
100	20	25,000	225
250	32		

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	91
9,446	x
10,000	124

By vertically interpolating between 91 and 124, the entry for "x" is determined to be 120.

$$9{,}446-5{,}000 = 4{,}446$$
$$10{,}000-5{,}000 = 5{,}000$$

$$91+\frac{4{,}446}{5{,}000}(124-91) = 120$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 9,326,000 to 9,566,000 housing units. Therefore, a conclusion that the average estimate of 1978 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 9,254,000 to 9,638,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 9,206,000 to 9,686,000 housing units with 95 percent confidence.

Table A-2 also shows that of the 9,446,000 specified owner-occupied housing units with two bedrooms in 1978, 1,703,000, or 18.0 percent, were valued between $10,000 and $19,999. Interpolation of the standard error in table VI (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 VI. The entry for "p" is the one sought.

Base of percentage (000)	Estimated percentage		
	15	18	25
5,000	0.7	a	0.8
9,446		p	
10,000	0.5	b	0.6

1. By horizontal interpolation between 0.7 and 0.8, the entry for cell "a" is determined to be 0.7.

$$18.0-15.0 = 3.0$$
$$25.0-15.0 = 10.0$$

$$0.7+\frac{3.0}{10.0}(0.8-0.7) = 0.7$$

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

$$18.0-15.0 = 3.0$$
$$25.0-15.0 = 10.0$$

$$0.5 + \frac{3.0}{10.0}(0.6-0.5) = 0.5$$

3. By vertical interpolation between 0.7 and 0.5, the entry for "p" is determined to be 0.5.

$$9{,}446-5{,}000 = 4{,}446$$
$$10{,}000-5{,}000 = 5{,}000$$

$$0.7 + \frac{4{,}446}{5{,}000}(0.5-0.7) = 0.5$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 17.5 to 18.5 percent; the 90-percent confidence interval is from 17.2 to 18.8 percent; and the 95-percent confidence interval is from 17.0 to 19.0 percent.

Illustration II—Table A-2 of this report shows that in the United States in 1978 there were 5,684,000 specified owner-occupied housing units whose source of water was an individual well. Interpolation of the data in table II shows that the standard error of an estimate of this size is approximately 110,000. Consequently, the 68-percent confidence interval is from 5,574,000 to 5,794,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1978 specified owner-occupied housing units whose source of water was an individual well, 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 5,508,000 to 5,860,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 5,464,000 to 5,904,000 housing units with 95 percent confidence.

Table A-2 also shows that of the 5,684,000 specified owner-occupied housing units in 1978 whose source of water was an individual well, 253,000, or 4.5 percent, were valued at less than $10,000. Interpolation in table VII (i.e., interpolation on both the base and the percent) shows that the standard error of the above percentage is 0.5 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 4.0 to 5.0 percent; the 90-percent confidence interval is from 3.7 to 5.3 percent; and the 95-percent confidence interval is from 3.5 to 5.5 percent.

Differences—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard 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

TABLE VI. Standard Errors of Estimated Percentages of Housing Units: 1978 (Excluding Estimated Percentages of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water, Individual Well, Lacking Complete Plumbing, Mobile Homes, and Households with Head of Spanish Origin)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	25.9	25.9	25.9	25.9	25.9	25.9	25.9	29.5
10	14.8	14.8	14.8	14.8	14.8	14.9	18.1	20.9
25	6.5	6.5	6.5	6.5	7.9	9.4	11.4	13.2
50	3.4	3.4	3.4	4.1	5.6	6.7	8.1	9.3
100	1.7	1.7	1.8	2.9	4.0	4.7	5.7	6.6
250	0.7	0.8	1.2	1.8	2.5	3.0	3.6	4.2
500	0.3	0.6	0.8	1.3	1.8	2.1	2.6	3.0
1,000	0.2	0.4	0.6	0.9	1.3	1.5	1.8	2.1
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	0.6	0.7	0.8	0.9
10,000	0.02	0.13	0.2	0.3	0.4	0.5	0.6	0.7
25,000	0.01	0.08	0.12	0.2	0.3	0.3	0.4	0.4
50,000	–	0.06	0.08	0.13	0.2	0.2	0.3	0.3
75,000	–	0.05	0.07	0.11	0.14	0.2	0.2	0.2

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 shows that in the United States in 1978 there were 1,703,000 specified owner-occupied housing units with two bedrooms valued between $10,000 and $19,999. It also shows that in the United States in 1978 there were 2,128,000 specified owner-occupied housing units with two bedrooms valued between $20,000 and $29,999. Thus, the apparent difference between the number of 1978 specified owner-occupied housing units with two bedrooms valued between $10,000 and $19,999 and those valued between $20,000 and $29,999 is 425,000. Interpolation of the data in table I shows the standard error on an estimate of 1,703,000 to be approximately 53,000 and the standard error on an estimate of 2,128,000 to be approximately 59,000. Therefore the standard error of the estimated difference of 425,000 is about 79,000.

$$79{,}000 = \sqrt{(53{,}000)^2 + (59{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 425,000 difference is from 346,000 to 504,000 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 299,000 to 551,000 housing units, and the 95-percent confidence interval is from 267,000 to 583,000. Thus, we can conclude with 95 percent confidence that the number of 1978 specified owner-occupied housing units with two bedrooms valued between $20,000 and $29,999 is greater than the number valued between $10,000 and $19,999 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 appropriate 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.

TABLE VII. Standard Errors of Estimated Percentages of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water, Individual Well, Lacking Complete Plumbing, Mobile Homes, and Households with Head of Spanish Origin: 1978

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	32.0	32.0	32.0	32.0	32.0	32.0	32.0	34.3
10	19.1	19.1	19.1	19.1	19.1	19.1	21.0	24.3
25	8.6	8.6	8.6	8.6	9.2	11.0	13.3	15.3
50	4.5	4.5	4.5	4.7	6.5	7.7	9.4	10.8
100	2.3	2.3	2.3	3.3	4.6	5.5	6.6	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	2.4	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	0.15	0.2	0.2	0.3	0.3
75,000	–	0.06	0.08	0.12	0.2	0.2	0.2	0.3

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 $31,400 in 1978. The base of the distribution, from which this median was determined is 9,446,000 housing units.

1. From table VI, the standard error of a 50-percent characteristic on the base of 9,446,000 is 0.7 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 48.6 and 51.4.
3. From table A-2, it can be seen by cumulating the frequencies for the first three categories that 4,438,000 owner-occupied housing units with two bedrooms, or 47.0 percent, had a value less than $30,000 and an additional 1,983,000 owner-occupied housing units with two bedrooms, or 21.0 percent, had a value between $30,000 and $40,000. By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$\$30{,}000 + (\$40{,}000 - \$30{,}000)\left(\frac{48.6 - 47.0}{21.0}\right) = \$30{,}800$$

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

$$\$30{,}000 + (\$40{,}000 - \$30{,}000)\left(\frac{51.4 - 47.0}{21.0}\right) = \$32{,}100$$

Thus, the 95-percent confidence interval ranges from $30,800 to $32,100.

Nonsampling errors--In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the 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 1978 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.

TABLE VIII. Standard Errors of Estimated Percentages of Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1978 (Excluding Estimated Percentages Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and Excluding Source of Water, Individual Well, and Mobile Homes for Each of the Regions)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	27.3	27.3	27.3	27.3	27.3	27.3	27.3	30.6
10	15.8	15.8	15.8	15.8	15.8	15.8	18.7	21.6
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	1.8	1.8	1.9	3.0	4.1	4.9	5.9	6.8
250	0.7	0.9	1.2	1.9	2.6	3.1	3.7	4.3
500	0.4	0.6	0.9	1.3	1.8	2.2	2.7	3.1
1,000	0.2	0.4	0.6	0.9	1.3	1.5	1.9	2.2
2,500	0.07	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

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 1978 reinterview study were not available at the time of publication; however, it is expected that they will be similar to the results of the 1977 reinterview study which are presented in the Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample, 1977."

The results of the 1977 reinterview study seem to indicate that there are some very slight improvements in the consistency of response from 1976 to 1977. Overall, the 1977 reinterview program showed moderate to high levels of inconsistency with about 21 percent of the nonattitudinal and 56 percent of the attitudinal items showing high levels of inconsistency. But a large proportion (43 percent) of the nonattitudinal items showed a low level of inconsistency. Moderate levels indicate that there are some problems with inconsistent reporting and high levels indicate that improvements are needed in the data collection methods or that the category concepts themselves are ambiguous.

Cross-tabulations involving those items which are subject to substantial levels of inconsistency may be subject to a large distortion as a consequence of the associated high level of response variance, and thus, are considered to be less reliable than comparable cross-tabulations which do not involve these data. As a consequence, the tables in this report which contain such cross-tabulations have been footnoted with a cautionary statement.

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

A possible explanation for the results of the AHS and census reinterview studies, as well as the surveys themselves, is that the 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—A deficiency in the representation of conventional new construction for the AHS new construction sample (mentioned previously in the section on estimation) is an example of coverage errors. 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 1978 AHS sample missed about 3.3 percent (i.e., about 375,000 units) of conventional housing

TABLE IX. Standard Errors of Estimated Percentages of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and to Source of Water, Individual Well, and Mobile Homes for the Northeast and North Central Regions: 1978

(68 chances out of 100. For estimated percentages pertaining to source of water, individual well, cooking fuel, and mobile homes for the West Region, apply a factor of 1.66 to the standard errors listed below)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	34.3	34.3	34.3	34.3	34.3	34.3	34.3	36.1
10	20.7	20.7	20.7	20.7	20.7	20.7	22.1	25.5
25	9.5	9.5	9.5	9.5	9.7	11.5	14.0	16.2
50	5.0	5.0	5.0	5.0	6.9	8.2	9.9	11.4
100	2.5	2.5	2.5	3.5	4.8	5.8	7.0	8.1
250	1.0	1.0	1.4	2.2	3.1	3.6	4.4	5.1
500	0.5	0.7	1.0	1.6	2.2	2.6	3.1	3.6
1,000	0.3	0.5	0.7	1.1	1.5	1.8	2.2	2.6
2,500	0.10	0.3	0.5	0.7	1.0	1.2	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	0.4	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

units built after April 1970 because the permits for these units, which were built before October 1978, were issued less than 5 months in advance of the survey. The second stage of ratio estimation procedure was employed to reduce the effect of this deficiency although some bias in the AHS estimates of conventional new construction probably still exists.

In addition, the 1976 Coverage Improvement Program 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 1978 AHS sample missed as much as 2 percent (i.e., as much as 400,000 units) of all housing units in ED's where area sampling methods are used because these units are not listed during the canvassing.

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

Rounding errors—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 published rounded data, they can differ from medians calculated directly from the published data.

TABLE X. Standard Errors of Estimated Percentages of Housing Units Pertaining to Source of Water, Individual Well, and Mobile Homes for the South Region: 1978

(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	45.2	45.2	45.2	45.2	45.2	45.2	45.2	45.4
10	29.2	29.2	29.2	29.2	29.2	29.2	29.2	32.1
25	14.1	14.1	14.1	14.1	14.1	14.5	17.6	20.3
50	7.6	7.6	7.6	7.6	8.6	10.2	12.4	14.4
100	4.0	4.0	4.0	4.4	6.1	7.2	8.8	10.1
250	1.6	1.6	1.8	2.8	3.9	4.6	5.6	6.4
500	0.8	0.9	1.3	2.0	2.7	3.2	3.9	4.5
1,000	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
2,500	0.2	0.4	0.6	0.9	1.2	1.4	1.8	2.0
5,000	0.08	0.3	0.4	0.6	0.9	1.0	1.2	1.4
10,000	0.04	0.2	0.3	0.4	0.6	0.7	0.9	1.0
25,000	0.02	0.13	0.2	0.3	0.4	0.5	0.6	0.6

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-40
Selection of sample areas. App-40
Designation of sample housing units for the 1978 enumeration . App-40
Selection of the 1973 sample housing units. App-40
Splitting of the sample . . App-41
Selection of supplemental sample housing units in rural areas. App-41
Selection of sample housing units for the 1976 Coverage Improvement Program . App-41
1977 SAMPLE REDUCTION App-42
ESTIMATION. . App-42
RELIABILITY OF THE ESTIMATES. . App-43
Sampling errors . App-43
Standard errors of estimates of levels . App-43
Standard errors of estimates of percentages. App-44
Standard errors of ratios . App-44
Illustration of the use of the standard error tables. Illustration I . App-44
Illustration II . App-46
Differences. App-46
Illustration of the computation of the standard error of a difference. App-46
Medians . App-47
Illustration of the computation of the 95-percent confidence interval for a median. App-47
Nonsampling errors. App-48
Reinterview Program . App-48
Coverage errors . App-50
Rounding errors. App-50

SAMPLE DESIGN

The 1978 estimates are based on data collected in October 1978 through January 1979 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 72,000 sample housing units (both occupied and vacant) were eligible for interview in the 1978 Annual Housing Survey. Of this number, 4,400 interviews were classified as "noninterview" for various reasons. Occupied housing units were calssified 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 72,000 housing units which were eligible for interview, there were also 5,900 sample units which were visited but were ineligible for interview for the AHS in terms of collecting information relevant to the 1978 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 in sample with certainty. These 156 strata, mostly the larger SMSA's, 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 1978 survey—The sample housing units designated to be interviewed in the 1978 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1977 survey (which included all sample housing units that were selected as part of the 1976 Coverage Improvement Program).
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 1977 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1978 AHS questionnaire, page App-20.)
3. All sample housing units that were selected from the list of building permits issued since the 1977 survey. (This sample represented the housing units built in permit-issuing areas, since the 1977 survey.)
4. Units added as the result of the updated listings in selected areas which do not issue building permits.

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 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 interview 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, the other half of each rural cluster (an expected two housing units) was reactivated in 1974. Similarly, for the area sampling frame, 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 deficiences 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, for which construction had not been completed at the time of the 1970 census.
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. In the second stage, 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 SAMPLE 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 in 1977. 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.

ESTIMATION

In 1978, the AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the type A noninterview housing units encountered in the AHS. This noninterview adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

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

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

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

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 current vacant housing estimates for 4 categories of vacant housing units and to independently derived current housing estimates for 24 categories of occupied housing units. Each of these 24 categories is a combination of the characteristics of residence, tenure, race of head, and sex of head.

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

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

The numerators of the ratios for occupied housing units were derived from data based on the Current Population Survey

(CPS), a sample household survey conducted monthly by the Bureau of the Census. The numerators of the ratios for vacant housing units were derived from data based on the Housing Vacancy Survey (HVS), a quarterly vacancy survey also conducted by the Bureau of the Census. The denominators of the ratios were obtained from the weighted estimates for the AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit in each third-stage ratio estimation category.

The second-stage and the third-stage ratio estimation procedures were 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 18 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 17 categories of new construction units for each of the 4 regions (i.e., 11 categories for conventional new construction units and 6 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 adjustment) 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 third-stage ratio estimation procedure, as well as the overall estimation procedure, reduced 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 questionnaires, 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 result 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 1978 national housing inventory estimates in this report, and tables III, IV, and V present the standard errors applicable to estimates for the Northeast, North Central, South, and West Regions. Linear interpolation should be used to determine standard errors for levels of estimates not specifically shown in tables I through V.

TABLE I. Standard Errors of Estimated Numbers of Housing Units: 1978 (Excluding Estimates of Housing Units Pertaining to New Construction, No Bedrooms, Lacking Complete Plumbing, Mobile Homes, and Households With Head of Spanish Origin)

(68 chances out of 100)

Size of estimate (000)	Standard error		Size of estimate (000)	Standard error	
	Total or White (000)	Black (000)		Total or White (000)	Black (000)
0	2	2	1,000 . . .	42	39
5	3	3	2,500 . . .	65	55
10	4	4	5,000 . . .	91	57
25	7	7	10,000. . .	124	–
50	9	9	25,000. . .	174	–
100.	13	13	50,000. . .	185	–
250.	21	21	75,000. . .	109	–
500.	29	29			

TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to New Construction, No Bedrooms, Lacking Complete Plumbing, Mobile Homes, and Households With Head of Spanish Origin: 1978

(68 chances out of 100)

Size of estimate (000)	Standard error		Size of estimate (000)	Standard error	
	Total or White (000)	Black (000)		Total or White (000)	Black (000)
0	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. . .	215	–
250.	24	24	75,000. . .	126	–
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 VI through 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 table II. The standard errors shown in table VII should be used for those specified items. Table VIII shows the approximate standard errors of all regional percentages of housing units except those pertaining to the specified items in tables IV and V. Tables IX and X should be used for those specified items. 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 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 in the United States there were 5,387,000 owner-occupied housing units occupied by recent movers in 1978. Interpolation in table I shows that the standard error of an estimate of this size is approximately 94. The following procedure was used in interpolating:

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

Size of estimate (000)	Standard error (000)
5,000	91
5,387	x
10,000	124

By vertically interpolating between 91 and 124, the entry for "x" is determined to be 94.

$$5{,}387-5{,}000 = 387$$
$$10{,}000-5{,}000 = 5{,}000$$

$$91+\frac{387}{5{,}000}(124-91) = 94$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 5,293,000 to 5,481,000 housing units. Therefore, a conclusion that the average estimate of 1978 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 5,237,000 to 5,537,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 5,199,000 to 5,575,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 5,387,000 owner-occupied housing units occupied by recent movers in 1978, 1,123,000, or 20.8 percent had 4 persons. Interpolation in standard error table VI (i.e., interpolation on both the base and percent) shows that the standard error of the percentage is 0.8 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	20.8	25
5,000	0.7	a	0.8
5,387		p	
10,000	0.5	b	0.6

1. By horizontal interpolation between 0.7 and 0.8, the entry for cell "a" is determined to be 0.8.

$$20.8-15.0 = 5.8$$
$$25.0-15.0 = 10.0$$
$$0.7 + \frac{5.8}{10.0}(0.8-0.7) = 0.8$$

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

$$20.8-15.0 = 5.8$$
$$25.0-15.0 = 10.0$$
$$0.5 + \frac{5.8}{10.0}(0.6-0.5) = 0.6$$

3. By vertical interpolation between 0.8 and 0.6, the entry for "p" is determined to be 0.8.

$$5{,}387-5{,}000 = 387$$
$$10{,}000-5{,}000 = 5{,}000$$
$$0.8 - \frac{387}{5{,}000}(0.8-0.6) = 0.8$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 20.0 to 21.6 percent; the 90-percent confidence interval is from 19.5 to 22.1 percent; and the 95-percent confidence interval is from 19.2 to 22.4 percent.

TABLE III. Standard Errors of Estimated Numbers of Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1978 (Excluding Estimates of Housing Units Pertaining to New Construction, No Bedrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and Excluding Mobile Homes for Each of the Regions)

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	500.	31
5	3	1,000	43
10	4	2,500	68
25	7	5,000	97
50	10	10,000.	137
100.	14	25,000.	216
250.	22		

TABLE IV. Standard Errors of Estimated Numbers of Housing Units Pertaining to New Construction, No Bedrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and Mobile Homes for the Northeast and North Central Regions: 1978

(68 chances out of 100. For estimates pertaining to mobile homes for the West Region, apply a factor of 1.66 to the standard errors)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	500.	36
5	4	1,000	51
10	5	2,500	80
25	8	5,000	111
50	11	10,000.	153
100.	16	25,000.	218
250.	26		

TABLE V. Standard Errors of Estimated Numbers of Housing Units Pertaining to Mobile Homes for the South Region: 1978

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	4	100.	20
5	5	250.	32
10	6	500.	45
25	10	1,000	64
50	14	2,500	99

Illustration II—Table A-21 of this report shows that in the United States in 1978 there were 213,000 owner-occupied housing units having a recent mover household head of Spanish origin. Interpolation in standard error 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 191,000 to 235,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1978 owner-occupied housing units having a recent mover household head of Spanish origin, 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 178,000 to 248,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 169,000 to 257,000 housing units with 95 percent confidence.

Table A-21 also shows that of the 213,000 1978 owner-occupied housing units having a recent mover household head of Spanish origin, 111,000, or 52.1 percent, had 3 bedrooms. Interpolation in table VII (i.e., interpolation on both the base and the percent) shows that the standard error of the above percentage is 5.5 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 46.6 to 57.6 percent; the 90-percent confidence interval is from 43.3 to 60.9 percent; and the 95-percent confidence interval is from 41.1 to 63.1 percent.

Differences—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same 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 of this report shows that in the United States in 1978 there were 511,000 owner-occupied housing units occupied by recent movers with 5 persons. Thus the apparent difference between the number of 1978 owner-occupied housing units occupied by recent movers with 5 persons and the number with 4 persons is 612,000. Interpolation in standard error table I shows that the standard error on an estimate of 1,123,000 to be approximately 44,000 and the standard error on an estimate of 511,000 to be approximately 29,000. Therefore, the standard error of the estimated difference of 612,000 is about 53,000.

$$53{,}000 = \sqrt{(44{,}000)^2 + (29{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 612,000 difference is from 559,000 to 665,000 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

TABLE VI. Standard Errors of Estimated Percentages of Housing Units: 1978 (Excluding Estimated Percentages of Housing Units Pertaining to New Construction, No Bedrooms, Lacking Complete Plumbing, Mobile Homes, and Households With Head of Spanish Origin)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	25.9	25.9	25.9	25.9	25.9	25.9	25.9	25.9
10	14.8	14.8	14.8	14.8	14.8	14.9	18.1	20.9
25	6.5	6.5	6.5	6.5	7.9	9.4	11.4	13.2
50	3.4	3.4	3.4	4.1	5.6	6.7	8.1	9.3
100	1.7	1.7	1.8	2.9	4.0	4.7	5.7	6.6
250	0.7	0.8	1.2	1.8	2.5	3.0	3.6	4.2
500	0.3	0.6	0.8	1.3	1.8	2.1	2.6	3.0
1,000	0.2	0.4	0.6	0.9	1.3	1.5	1.8	2.1
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	0.6	0.7	0.8	0.9
10,000	0.02	0.13	0.2	0.3	0.4	0.5	0.6	0.7
25,000	0.01	0.08	0.12	0.2	0.3	0.3	0.4	0.4
50,000	–	0.06	0.08	0.13	0.2	0.2	0.3	0.3
75,000	–	0.05	0.07	0.11	0.14	0.2	0.2	0.2

samples. Similarly, the 90-percent confidence interval is from 527,000 to 697,000 housing units, and the 95-percent confidence interval is from 506,000 to 718,000. Thus, we can conclude with 95 percent confidence that the number of 1978 owner-occupied housing units occupied by recent movers with four persons, is different than the number of owner-occupied units occupied by recent movers with five persons since the 95-percent confidence interval of this difference does not include zero or negative values.

Medians—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median base 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 occupied by recent movers in the United States was 2.8 in 1978. The base of the distribution, from which this median was determined is 5,387,000 housing units.

1. From standard error table VI, the standard error of a 50-percent characteristic on the base of 5,387,000 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 2,353,000 owner-occupied housing units occupied by recent movers, or 43.7 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 1,145,000 housing units, or 21.3 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:

$$2.5 + (3.5-2.5)\,\frac{(48.2-43.7)}{21.3} = 2.7$$

TABLE VII. Standard Errors of Estimated Percentages of Housing Units Pertaining to New Construction, No Bedrooms, Lacking Complete Plumbing, Mobile Homes, and Households With Head of Spanish Origin: 1978

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	32.0	32.0	32.0	32.0	32.0	32.0	32.0	34.3
10	19.1	19.1	19.1	19.1	19.1	19.1	21.0	24.3
25	8.6	8.6	8.6	8.6	9.2	11.0	13.3	15.3
50	4.5	4.5	4.5	4.7	6.5	7.7	9.4	10.8
100	2.3	2.3	2.3	3.3	4.6	5.5	6.6	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	2.4	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	0.15	0.2	0.2	0.3	0.3
75,000	—	0.06	0.08	0.12	0.2	0.2	0.2	0.3

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

$$2.5 + (3.5-2.5)\frac{(51.8-43.7)}{21.3} = 2.9$$

Thus, the 95-percent confidence interval ranges from 2.7 to 2.9 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. 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 1978 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 1978 reinterview study were not available at the time of publication; however, it is expected that they will be similar to the results of the 1977 reinterview study which are presented in the Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample 1977."

The results of the 1977 reinterview study seem to indicate that there are some very slight improvements in the consistency of response from the 1976 to the 1977 survey. Overall, the 1977 reinterview program showed moderate to high levels of inconsistency with about 21 percent of the nonattitudinal and 56 percent of the attitudinal items showing high levels of inconsistency. But a large proportion (43 percent) of the nonattitu-

TABLE VIII. Standard Errors of Estimated Percentages of Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1978 (Excluding Estimated Percentages Pertaining to New Construction, No Bedrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and Excluding Mobile Homes for Each of the Regions)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	27.3	27.3	27.3	27.3	27.3	27.3	27.3	30.6
10	15.8	15.8	15.8	15.8	15.8	15.8	18.7	21.6
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	1.8	1.8	1.9	3.0	4.1	4.9	5.9	6.8
250	0.7	0.9	1.2	1.9	2.6	3.1	3.7	4.3
500	0.4	0.6	0.9	1.3	1.8	2.2	2.7	3.1
1,000	0.2	0.4	0.6	0.9	1.3	1.5	1.9	2.2
2,500	0.07	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

dinal items showed a low level of inconsistency. Moderate levels indicate that there are some problems with inconsistent reporting and high levels indicate that improvements are needed in the data collection methods or that the category concepts themselves are ambiguous.

Cross-tabulations involving those items, which are subject to substantial levels of inconsistency, may be subject to a large distortion as a consequence of the associated high level of response variance, and thus, are considered to be less reliable than comparable cross-tabulations which do not involve these data. As a consequence, the tables in this report, which contain such cross-tabulations have been footnoted with a cautionary statement.

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

A possible explanation for the results of the AHS and census reinterview studies, as well as the surveys themselves, is that the

TABLE IX. Standard Errors of Estimated Percentages of Housing Units Pertaining to New Construction, No Bedrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and to Mobile Homes for the Northeast and North Central Regions: 1978

(68 chances out of 100. For estimated percentages pertaining to mobile homes for the West Region, apply a factor of 1.66 to the standard errors)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	34.3	34.3	34.3	34.3	34.3	34.3	34.3	36.1
10	20.7	20.7	20.7	20.7	20.7	20.7	22.1	25.5
25	9.5	9.5	9.5	9.5	9.7	11.5	14.0	16.2
50	5.0	5.0	5.0	5.0	6.9	8.2	9.9	11.4
100	2.5	2.5	2.5	3.5	4.8	5.8	7.0	8.1
250	1.0	1.0	1.4	2.2	3.1	3.6	4.4	5.1
500	0.5	0.7	1.0	1.6	2.2	2.6	3.1	3.6
1,000	0.3	0.5	0.7	1.1	1.5	1.8	2.2	2.6
2,500	0.10	0.3	0.5	0.7	1.0	1.2	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	0.4	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

TABLE X. Standard Errors of Estimated Percentages of Housing Units Pertaining to Mobile Homes for the South Region: 1978

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	45.2	45.2	45.2	45.2	45.2	45.2	45.2	45.4
10	29.2	29.2	29.2	29.2	29.2	29.2	29.2	32.1
25	14.1	14.1	14.1	14.1	14.1	14.5	17.6	20.3
50	7.6	7.6	7.6	7.6	8.6	10.2	12.4	14.4
100	4.0	4.0	4.0	4.4	6.1	7.2	8.8	10.1
250	1.6	1.6	1.8	2.8	3.9	4.6	5.6	6.4
500	0.8	0.9	1.3	2.0	2.7	3.2	3.9	4.5
1,000	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
2,500	0.2	0.4	0.6	0.9	1.2	1.4	1.8	2.0
5,000	0.08	0.3	0.4	0.6	0.9	1.0	1.2	1.4
10,000	0.04	0.2	0.3	0.4	0.6	0.7	0.9	1.0
25,000	0.02	0.13	0.2	0.3	0.4	0.5	0.6	0.6

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—Deficiencies in the representation of conventional new construction for the AHS new construction sample (mentioned previously in the section on estimation) is an example of coverage errors. 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 1978 AHS sample missed about 3.3 percent (i.e., about 375,000 units) of conventional housing units built after April 1970, because the permits for these units, which were built before October 1978, were issued less than 5 months in advance of the survey. The second stage of the ratio estimation procedure was employed to reduce the effect of this deficiency although some bias in the AHS estimates of conventional new construction probably still exists.

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 1978 AHS sample missed as much as 2 percent (i.e., as much as 400,000 units) of all housing units in ED's where area sampling methods are used because these units are not listed during the canvassing.

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

Rounding errors—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 published rounded data, they can differ from medians calculated directly from the published data.

☆U.S. GOVERNMENT PRINTING OFFICE: 1981-340-997/1656

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-40
Selection of sample areas. . . App-40
Designation of sample housing units for the 1978 survey. . . App-40
Selection of the 1973 sample housing units. . . App-41
Splitting of the sample . . . App-41
Selection of supplemental sample housing units in rural areas . . . App-41
Selection of sample housing units for the 1976 Coverage Improvement Program . . . App-41
1977 sample reduction. . . App-42
1970 Census of Population and Housing . . . App-42

ESTIMATION. . . App-42
AHS national sample . . . App-42
1978 housing inventory . . . App-42
1973-1978 lost units . . . App-43
Ratio estimation procedure of the 1970 Census of Population and Housing . . . App-43

RELIABILITY OF THE ESTIMATES. . . App-43
Sampling errors . . . App-43
Standard errors of estimates of levels. . . App-44
Standard errors of estimates of percentages . . . App-44
Standard errors of ratios . . . App-45
Illustration of the use of the standard error tables. Illustration I . . . App-45
Illustration II . . . App-46
Differences. . . App-46
Illustration of the computation of the standard error of a difference. . . App-47
Medians. . . App-48
Illustration of the computation of the 95-percent confidence interval for a median. . App-48

Nonsampling errors. . . App-48
1970 census . . . App-48
Reinterview program . . . App-50
Coverage errors . . . App-51
Rounding errors. . . App-52

SAMPLE DESIGN

The 1978 estimates are based on data collected in October 1978 through January 1979 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 72,000 sample housing units (both occupied and vacant) were eligible for interview in the 1978 Annual Housing Survey. Of this number, 4,400 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 72,000 eligible housing units, there were also 5,900 sample units which were visited but were ineligible for interview for the AHS in terms of collecting information relevant to the 1978 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 in sample with certainty. These 156 strata, mostly the larger SMSA's, were called self-representing (SR) because 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 1978 survey—The sample housing units designated to be interviewed in the 1978 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1977 survey (which included all sample housing units that were selected as part of the 1976 Coverage Improvement Program).
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 1977 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1978 AHS questionnaire, page App-20.)
3. All sample housing units that were selected from the list of building permits issued since the 1977 survey. (This sample represented the housing units built in permit-issuing areas, since the 1977 survey.)
4. Units added as the result of the updated listings in selected areas which do not issue building permits.

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 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 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, the other half of each rural cluster (an expected two housing units) was reactivated in 1974. Similarly, for the area sampling frame, 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 for which construction had not been completed at the time of the 1970 census.
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. In the second stage, 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 sample reduction—By 1977, the addition to the sample (primarily from 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 in 1977. 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 1978 housing inventory and estimates of units removed from the housing inventory between 1973 and 1978 (i.e., 1973-1978 lost units). Each type of estimate employed a separate, though similar, estimation procedure.

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

1978 housing inventory—In 1978, the AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the type A noninterview housing units encountered in the AHS. This noninterview adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

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

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

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

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

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

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

The second-stage ratio estimation factor was as follows:

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

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

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

The third-stage ratio estimation procedure was employed for all 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 current vacant housing estimates for 4 categories of vacant housing units and to independently derived current housing estimates for 24 categories of occupied housing units. Each of these 24 categories is a combination of the characteristics of residence, tenure, race of head, and sex of head.

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

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

The numerators of the ratios for occupied housing units were derived from data based on the Current Population Survey (CPS), a sample household survey conducted monthly by the Bureau of the Census. The numerators of the ratios for vacant housing units were derived from data based on the Housing Vacancy Survey (HVS), a quarterly vacancy survey also conducted by the Bureau of the Census. The denominators of the ratios were obtained from the weighted estimates for the AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit in each third-stage ratio estimation category.

The second-stage and the third-stage ratio estimation procedures were 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 18 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 17 categories of new construction units for each of the 4 regions (i.e., 11 categories for conventional new construction units and 6 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 third-stage ratio estimation procedure, as well as the overall estimation procedure, reduced 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.

1973-1978 lost units—The 1973-1978 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 1973-1978 lost units do not include the housing units from the 1976 Coverage Improvement. Since the 1973-1978 lost units existed, by definition, in the 1973 housing inventory, there was a 1973 housing inventory weight associated with each 1973-1978 lost unit. This weight, adjusted for the 1977 reduction, was used to tabulate the estimates of the characteristics of the 1973-1978 lost units. 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 questionnaires, 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 result 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 1978 national housing inventory estimates in this report, and tables IV and V present the standard errors applicable to 1973-1978 lost housing unit estimates in this report. Table VI presents the standard errors applicable for the Northeast, North Central, South, and West Regions. Linear interpolation should be used to determine standard errors 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.

TABLE I. Standard Errors of Estimated Numbers of Housing Units: 1978 (Excluding Estimates of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water-Individual Well, Cooking Fuel, Lacking Complete Plumbing, Mobile Homes, and Households With Head of Spanish Origin)

(68 chances out of 100)

Size of estimate (000)	Standard error		Size of estimate (000)	Standard error	
	Total or White (000)	Black (000)		Total or White (000)	Black (000)
0	2	2	1,000 . . .	42	39
5	3	3	2,500 . . .	65	55
10	4	4	5,000 . . .	91	57
25	7	7	10,000. . .	124	–
50	9	9	25,000. . .	174	–
100.	13	13	50,000. . .	185	–
250.	21	21	75,000. . .	109	–
500.	29	29			

TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water-Individual Well, Cooking Fuel, Lacking Complete Plumbing, Mobile Homes, and Households With Head of Spanish Origin: 1978

(68 chances out of 100)

Size of estimate (000)	Standard error		Size of estimate (000)	Standard errors	
	Total or White (000)	Black (000)		Total or White (000)	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. . .	215	–
250.	24	24	75,000. . .	126	–
500.	34	33			

Tables VII through XI present the standard errors of estimated percentages. Tables VII and VIII show the approximate standard errors of all national estimated percentages of housing units. Tables IX and X show the approximate standard errors of the estimated percentages of 1973-1978 lost housing units. Table XI shows the approximate standard error 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 VII through XI.

Included in tables I through XI 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 XI 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 in urban areas of the United States there were 22,148,000 renter-occupied housing units in 1978. Interpolation in standard error 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 following table was extracted from standard error table III. The entry for "x" is the one sought.

Size of estimate (000)	Standard error (000)
10,000	126
22,148	x
25,000	180

By vertically interpolating between 126,000 and 180,000, the entry for "x" is determined to be 170,000.

$$22{,}148-10{,}000 = 12{,}148$$
$$25{,}000-10{,}000 = 15{,}000$$

$$126 + \frac{12{,}148}{15{,}000}(180-126) = 170$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 21,978,000 to 22,318,000 housing units. Therefore, a conclusion that the average estimate of 1978 housing units of this type lies within a range computed in this way and 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 21,876,000 to 22,420,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 21,808,000 to 22,488,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 22,148,000 renter-occupied housing units in urban areas, 6,420,000, or 29.0 percent, were occupied by 2 persons. Interpolation in standard error table VII (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage is 0.5. The following procedure was used in interpolating.

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

Base of percentage (000)	Estimated percentage		
	25	29.0	50
10,000	0.6	a	0.7
22,148		p	
25,000	0.4	b	0.4

1. By horizontal interpolation between 0.7 and 0.8, the entry for cell "a" is determined to be 0.7.

$$29.0-25.0 = 4.0$$
$$50.0-25.0 = 25.0$$

$$0.6 + \frac{4.0}{25.0}(0.7-0.6) = 0.6$$

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

$$29.0-25.0 = 4.0$$
$$50.0-25.0 = 25.0$$

$$0.4 + \frac{4.0}{25.0}(0.4-0.4) = 0.4$$

3. By vertical interpolation between 0.7 and 0.4, the entry for "p" is determined to be 0.5.

$$22{,}148-10{,}000 = 12{,}148$$
$$25{,}000-10{,}000 = 15{,}000$$

$$0.7 + \frac{12{,}148}{15{,}000}(0.4-0.6) = 0.5$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 28.5 to 29.5 percent; the 90-percent confidence interval is from 28.2 to 29.8 percent; and the 95-percent confidence interval is from 28.0 to 30.0 percent.

Illustriation II—Table A-2 of this report shows that in the rural areas of the United States in 1978 there were 11,851,000 specified owner-occupied housing units. Interpolation in standard error table III of this appendix shows that the standard error of an estimate of this size is approximately 142,000. Consequently, the 68-percent confidence interval is from 11,709,000 to 11,993,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1978 specified owner-occupied housing units lies within a range computed in this way and 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 11,624,000 to 12,078,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 11,567,000 to 12,135,000 housing units with 95 percent confidence.

Table A-2 also shows that of the 11,851,000 specified owner-occupied housing units in rural areas, 4,774,000, or 40.3 percent, had no mortgage. Table VII gives instructions to multiply 0.86 times the standard errors in the table to produce the applicable standard errors. Interpolation in table VII with the factor 0.86 applied (i.e., interpolation on both the base and the percent) shows that the standard error of the above percentage is 0.6 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 39.7 to 40.9 percent; the 90-percent confidence interval is from 39.3 to 41.3 percent; and the 95-percent confidence interval is from 39.1 to 41.5 percent.

Differences—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard 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.

TABLE III. Standard Errors of Estimated Numbers of Urban or Rural Housing Units: 1978

(68 chances out of 100)

Size of estimate (000)	Rural housing units (except those in the next column) Total, White, Black, or Spanish origin (000)	Rural housing units pertaining to new construction, no bedrooms, source of water-individual well, cooking fuel, lacking some or all plumbing, and mobile homes Total, White, or Black (000)	Urban housing units (except those in the next column)		Urban housing units pertaining to new construction, lacking complete kitchen facilities, no bedrooms, no bathrooms, source of water-individual well, cooking fuel, lacking some or all plumbing, mobile homes, and household head of Spanish origin Total, White, Black, or Spanish origin (000)
			Total or White (000)	Black (000)	
0	1	2	2	2	2
5	3	3	3	3	3
10	4	4	4	4	5
25	6	7	7	7	7
50	8	10	9	9	11
100	11	14	13	13	15
250	18	23	21	21	24
500	26	33	30	29	35
1,000	36	51	42	40	52
2,500	59	95	66	58	91
5,000	86	164	92	63	148
10,000	129	298	126	–	256
25,000	235	–	180	–	–
50,000	–	–	200	–	–
60,000	–	–	190	–	–

Illustration of the computation of the standard error of a difference—Table A-1 of this report shows that in urban areas of the United States there were 6,420,000 renter-occupied housing units with 2 persons and 3,459,000 renter-occupied housing units with 3 persons in 1978. Thus, the apparent difference between the number of 1978 renter-occupied housing units in urban areas with 2 persons and those with 3 persons is 2,961,000. Interpolation in standard error table III shows the standard error of 6,420,000 is approximately 102,000 and that the standard error on an estimate of 3,459,000 is approximately 76,000. Therefore, the standard error of the estimated difference of 2,961,000 is about 127,000.

$$127{,}000 = \sqrt{(102{,}000)^2 + (76{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 2,961,000 difference is from 2,834,000 to 3,088,000 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 and would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 2,758,000 to 3,164,000 housing units, and the 95-percent confidence interval is from 2,707,000 to 3,215,000. Thus, we can conclude with 95 percent confidence that the number of 1978 renter-occupied housing units in urban areas with two persons is greater than the number with three persons.

TABLE IV. Standard Errors of Estimated Numbers of Lost Housing Units and of Urban and Rural Lost Housing Units: 1973-1978 (Excluding Estimates of Lost Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Some or All Plumbing, Mobile Homes, Other Vacants, and Rural Vacants for Rent)

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	250	21
5	3	500	31
10	4	1,000	47
25	6	2,500	88
50	9	5,000	150
100	13		

TABLE V. Standard Errors of Estimated Numbers of Lost Housing Units and of Urban and Rural Lost Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Some or All Plumbing, Mobile Homes, Other Vacants, and Rural Vacants for Rent: 1973-1978

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	100	16
5	4	250	26
10	5	500	38
25	8	1,000	55
50	12		

TABLE VIa. Standard Errors of Estimated Numbers of Housing Units and of Urban Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1978, Excluding Estimates of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and Excluding Source of Water-Individual Well, Cooking Fuel, and Mobile Homes for Each of the Regions

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	500	31
5	3	1,000	43
10	4	2,500	68
25	7	5,000	97
50	10	10,000	137
100	14	25,000	216
250	22		

TABLE VIb. Standard Errors of Estimated Numbers of Housing Units and of Urban Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and to Source of Water-Individual Well, Cooking Fuel, and Mobile Homes for the Northeast and North Central Regions: 1978

(68 chances out of 100. For estimates pertaining to source of water-individual well, cooking fuel, and mobile homes for the West Region, apply a factor of 1.66 to the standard errors)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	500	36
5	4	1,000	51
10	5	2,500	80
25	8	5,000	111
50	11	10,000	153
100	16	25,000	218
250	26		

TABLE VIc. Standard Errors of Estimated Numbers of Housing Units and of Urban Housing Units Pertaining to Source of Water-Individual Well, Cooking Fuel, and Mobile Homes for the South Region: 1978

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	4	500	45
5	5	1,000	64
10	6	2,500	99
25	10	5,000	136
50	14	10,000	181
100	20	25,000	225
250	32		

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 2.6 in 1978. The base of the distribution, from which this median was determined, is 32,729,000 housing units.

1. From standard error table VII, the standard error of a 50-percent characteristic on the base of 32,729,000 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 2 categories that 15,645,000 owner-occupied housing units, or 47.8 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 5,785,000 owner-occupied housing units, or 17.7 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 2.6.

$$2.5 + (3.5-2.5)\,\frac{(49.2-47.8)}{17.7} = 2.6$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about 2.7.

$$2.5 + (3.5-2.5)\,\frac{(50.8-47.8)}{17.7} = 2.7$$

Thus, the 95-percent confidence interval ranges from 2.6 to 2.7 persons. Although it appears that this confidence interval has the sample estimate as the upper limit, it actually is a reflection of the rounding error associated with the median (see the paragraph on rounding errors in the nonsampling errors section of this appendix).

Nonsampling errors—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data; and other errors of collection, response, processing, coverage, and estimation for missing data. 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 1970 Census of Population and Housing and the 1978 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 the extent to which occupancy status was erroneously reported. 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.**

TABLE VId. Standard Errors of Estimated Numbers of Rural Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1978

(68 chances out of 100)

Size of estimate (000)	Rural housing units (except in the following columns) for the Northeast, North Central, South, and West Regions (000)	Rural housing units pertaining to new construction, no bedrooms, lacking complete plumbing for the Northeast, North Central, and West Regions and to source of water-individual well, cooking fuel, and mobile homes for the Northeast and North Central Regions (000)	Rural housing units pertaining to source of water-individual well, cooking fuel, and mobile homes for the West Region (000)	Rural housing units pertaining to source of water-individual well, cooking fuel, and mobile homes for the South Region (000)
0	2	3	4	4
5	3	4	6	4
10	3	4	7	5
25	6	7	12	9
50	9	10	16	12
100	12	14	24	18
250	19	23	38	28
500	27	32	53	40
1,000	37	45	75	57
2,500	58	71	118	88
5,000	83	99	164	121
10,000	118	136	226	161

TABLE VII. Standard Errors of Estimated Percentages of Housing Units and of Urban Housing Units: 1978, Excluding Estimated Percentages of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water-Individual Well, Cooking Fuel, Lacking Complete Plumbing, Mobile Homes, and Households With Head of Spanish Origin

(68 chances out of 100. For standard errors of rural housing units, excluding estimated percentages of rural housing units pertaining to new construction, no bedrooms, source of water-individual well, cooking fuel, lacking some or all plumbing, and mobile homes, multiply the standard errors by 0.86)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	25.9	25.9	25.9	25.9	25.9	25.9	25.9	29.5
10	14.8	14.8	14.8	14.8	14.8	14.9	18.1	20.9
25	6.5	6.5	6.5	6.5	7.9	9.4	11.4	13.2
50	3.4	3.4	3.4	4.1	5.6	6.7	8.1	9.3
100	1.7	1.7	1.8	2.9	4.0	4.7	5.7	6.6
250	0.7	0.8	1.2	1.8	2.5	3.0	3.6	4.2
500	0.3	0.6	0.8	1.3	1.8	2.1	2.6	3.0
1,000	0.2	0.4	0.6	0.9	1.3	1.5	1.8	2.1
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	0.6	0.7	0.8	0.9
10,000	0.02	0.1	0.2	0.3	0.4	0.5	0.6	0.7
25,000	0.01	0.08	0.1	0.2	0.3	0.3	0.4	0.4
50,000	–	0.06	0.08	0.1	0.2	0.2	0.3	0.3
75,000	–	0.05	0.07	0.1	0.1	0.2	0.2	0.2

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.

TABLE VIII. Standard Errors of Estimated Percentages of Housing Units and of Urban Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water-Individual Well, Cooking Fuel, Lacking Complete Plumbing, Mobile Homes, and Households With Head of Spanish Origin: 1978

(68 chances out of 100. For standard errors of rural housing units pertaining to new construction, no bedrooms, source of water-individual well, cooking fuel, lacking some or all plumbing, and mobile homes, use the standard errors in table XIa.

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	32.0	32.0	32.0	32.0	32.0	32.0	32.0	34.3
10	19.1	19.1	19.1	19.1	19.1	19.1	21.0	24.3
25	8.6	8.6	8.6	8.6	9.2	11.0	13.3	15.3
50	4.5	4.5	4.5	4.7	6.5	7.7	9.4	10.8
100	2.3	2.3	2.3	3.3	4.6	5.5	6.6	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	2.4	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	0.1	0.1	0.2	0.3	0.3	0.4	0.5
50,000	–	0.07	0.1	0.1	0.2	0.2	0.3	0.3
75,000	–	0.06	0.08	0.1	0.2	0.2	0.2	0.3

TABLE IX. Standard Errors of Estimated Percentages of Lost Housing Units and of Urban and Rural Lost Housing Units: 1973-1978, Excluding Estimated Percentages of Lost Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Some or All Plumbing, Mobile Homes, Other Vacants, and Rural Vacants for Rent

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	24.5	24.5	24.5	24.5	24.5	24.5	24.7	28.5
10	14.0	14.0	14.0	14.0	14.0	14.4	17.5	20.2
25	6.1	6.1	6.1	6.1	7.6	9.1	11.0	12.7
50	3.1	3.1	3.1	3.9	5.4	6.4	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	1.7	2.0
2,500	0.06	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

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 1978 reinterview study were not available at the time of publication; however, it is expected that they will be similar to the results of the 1977 reinterview study which are presented in the Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample 1977."

The results of the 1977 reinterview study seem to indicate that there were some very slight improvements in the consistency of response from the 1976 to 1977 survey. Overall, the 1977 reinterview program showed moderate to high levels of inconsistency with about 21 percent of the nonattitudinal and 56 percent of the attitudinal items showing high levels of inconsistency. But a large proportion (43 percent) of the nonattitudinal items showed a low level of inconsistency. Moderate levels indicate that there are some problems with inconsistent reporting and high levels indicate that improvements are needed in the data collection methods or that the category concepts themselves are ambiguous.

Cross-tabulations involving those items, which are subject to substantial levels of inconsistency, may be subject to a large distortion as a consequence of the associated high level of response variance, and thus, are considered to be less reliable than comparable cross tabulations which do not involve these data. As a consequence, the tables in this report, which contain such cross-tabulations have been footnoted with a cautionary statement.

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

A possible explanation for the results of the AHS and census reinterview studies, as well as the surveys themselves, is that the data are based on the answers given by the respondents, who may lack precise information. Also, because the results of the reinterview studies are derived from sample surveys, there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

Coverage errors—Deficiencies in the representation of conventional new construction for the AHS new construction sample (mentioned previously in the section on estimation) is an example of coverage errors. 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 1978 AHS sample missed about 3.3 percent (i.e., about 375,000 units) of conventional housing units built after April 1970, because the permits for these units, which were built before October 1978, were issued less than 5 months in advance of the survey. The second stage of ratio

TABLE X. Standard Errors of Estimated Percentages of Lost Housing Units and of Urban and Rural Lost Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Some or All Plumbing, Mobile Homes, Other Vacants, and Rural Vacants for Rent: 1973-1978

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	34.9	34.9	34.9	34.9	34.9	34.9	34.9	36.6
10	21.1	21.1	21.1	21.1	21.1	21.1	22.4	25.9
25	9.7	9.7	9.7	9.7	9.8	11.7	14.2	16.4
50	5.1	5.1	5.1	5.1	6.9	8.3	10.0	11.6
100	2.6	2.6	2.6	3.6	4.9	5.8	7.1	8.2
250	1.1	1.1	1.4	2.3	3.1	3.7	4.5	5.2
500	0.5	0.7	1.0	1.6	2.2	2.6	3.2	3.7
1,000	0.3	0.5	0.7	1.1	1.6	1.8	2.2	2.6
2,500	0.1	0.3	0.5	0.7	1.0	1.2	1.4	1.6
5,000	0.05	0.2	0.3	0.5	0.7	0.8	1.0	1.2

estimation procedure was employed to reduce the effect of this deficiency although some bias in the AHS estimates of conventional new construction probably still exists.

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 use; 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 1978 AHS sample missed as much as 2 percent (i.e., as much as 400,000 units) of all housing units in ED's where area sampling methods are used because these units are not listed during the canvassing.

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

Rounding errors—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 published rounded data, they can differ from medians calculated directly from the published data.

TABLE XIa. Standard Errors of Estimated Percentages of Housing Units and of Urban Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1978, Excluding Estimated Percentages Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing Facilities for the Northeast, North Central, and West Regions and Excluding Source of Water-Individual Well, Cooking Fuel, and Mobile Homes for Each of the Regions

(68 chances out of 100. For standard errors of estimated percentages of rural housing units pertaining to the Northeast, North Central, South, and West Regions, excluding estimates of rural housing units pertaining to new construction, no bedrooms, and lacking complete plumbing for the Northeast, North Central, and West Regions, and excluding source of water-individual well, cooking fuel, and mobile homes for each of the regions, multiply the standard errors by 0.86)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	27.3	27.3	27.3	27.3	27.3	27.3	27.3	30.6
10	15.8	15.8	15.8	15.8	15.8	15.8	18.7	21.6
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	1.8	1.8	1.9	3.0	4.1	4.9	5.9	6.8
250	0.7	0.9	1.2	1.9	2.6	3.1	3.7	4.3
500	0.4	0.6	0.9	1.3	1.8	2.2	2.7	3.1
1,000	0.2	0.4	0.6	0.9	1.3	1.5	1.9	2.2
2,500	0.07	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	0.1	0.2	0.3	0.4	0.5	0.6	0.7
25,000	0.01	0.09	0.1	0.2	0.3	0.3	0.4	0.4

TABLE XIb. Standard Errors of Estimated Percentages of Housing Units and of Urban Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing Facilities for the Northeast, North Central, and West Regions and to Source of Water-Individual Well, Cooking Fuel, and Mobile Homes for the Northeast and North Central Regions: 1978

(68 chances out of 100. For estimated percentages pertaining to source of water-individual well, cooking fuel, and mobile homes for the West Region, apply a factor of 1.66 to the standard errors. For standard errors of regional rural estimates pertaining to new construction, no bedrooms, lacking complete plumbing, source of water-individual well, cooking fuel, and mobile homes for the Northeast, North Central, and West Regions, multiply the standard errors by 0.89, except for estimates for the West Region pertaining to source of water-individual well, cooking fuel, and mobile homes, multiply the standard errors by 1.48)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	34.3	34.3	34.3	34.3	34.3	34.3	34.3	36.1
10	20.7	20.7	20.7	20.7	20.7	20.7	22.1	25.5
25	9.5	9.5	9.5	9.5	9.7	11.5	14.0	16.2
50	5.0	5.0	5.0	5.0	6.9	8.2	9.9	11.4
100	2.5	2.5	2.5	3.5	4.8	5.8	7.0	8.1
250	1.0	1.0	1.4	2.2	3.1	3.6	4.4	5.1
500	0.5	0.7	1.0	1.6	2.2	2.6	3.1	3.6
1,000	0.3	0.5	0.7	1.1	1.5	1.8	2.2	2.6
2,500	0.1	0.3	0.5	0.7	1.0	1.2	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	0.4	0.5	0.6	0.7	0.8
25,000	0.01	0.1	0.1	0.2	0.3	0.4	0.4	0.5

TABLE XIc. Standard Errors of Estimated Percentages of Housing Units and of Urban Housing Units Pertaining to Source of Water-Individual Well, Cooking Fuel, and Mobile Homes for the South Region: 1978

(68 chances out of 100. For standard errors of estimated percentages of rural housing units pertaining to source of water-individual well, cooking fuel, and mobile homes for the South Region, multiply the standard errors by 0.89)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	45.2	45.2	45.2	45.2	45.2	45.2	45.2	45.4
10	29.2	29.2	29.2	29.2	29.2	29.2	29.2	32.1
25	14.1	14.1	14.1	14.1	14.1	14.5	17.6	20.3
50	7.6	7.6	7.6	7.6	8.6	10.2	12.4	14.4
100	4.0	4.0	4.0	4.4	6.1	7.2	8.8	10.1
250	1.6	1.6	1.8	2.8	3.9	4.6	5.6	6.4
500	0.8	0.9	1.3	2.0	2.7	3.2	3.9	4.5
1,000	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
2,500	0.2	0.4	0.6	0.9	1.2	1.4	1.8	2.0
5,000	0.08	0.3	0.4	0.6	0.9	1.0	1.2	1.4
10,000	0.04	0.2	0.3	0.4	0.6	0.7	0.9	1.0
25,000	0.02	0.1	0.2	0.3	0.4	0.5	0.6	0.6

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-40
Selection of sample areas. App-40
Designation of sample housing units for the 1978 survey. App-40
Selection of the 1973 sample housing units. App-40
Splitting of the sample App-41
Selection of supplemental sample housing units in rural areas App-41
Selection of sample housing units for the 1976 Coverage Improvement Program App-41
1977 SAMPLE REDUCTION App-42
ESTIMATION. App-42
RELIABILITY OF THE ESTIMATES. App-43
Sampling errors App-43
Standard errors of estimates of levels. App-43
Standard errors of estimates of percentages App-43
Standard errors of ratios App-44
Illustration of the use of the standard error tables. Illustration I App-44
Illustration II App-45
Differences. App-46
Illustration of the computation of the standard error of a difference App-46
Medians App-46
Illustration of the computation of the 95-percent confidence interval for a median App-47
Nonsampling errors App-47
Reinterview program App-47
Coverage errors App-48
Rounding errors. App-48

SAMPLE DESIGN

The 1978 estimates are based on data collected in October 1978 through January 1979 for the Annual Housing Survey (AHS)—national sample, which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. Under the sponsorship of the Department of Energy (DOE), the 1978 AHS—national sample questionnaire included a supplementary group of questions pertaining to energy conservation. 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 72,000 sample housing units (both occupied and vacant) were eligible for interview in the 1978 Annual Housing Survey. Of this number, 4,400 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 72,000 eligible units, there were also 5,900 sample units which were visited but were ineligible for interview for the AHS in terms of collecting information relevant to the 1978 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 in sample with certainty. These 156 strata, mostly the larger SMSA's, were called self-representing (SR) because 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 1978 survey—The sample housing units designated to be interviewed in the 1978 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1977 survey (which included all sample housing units that were selected as part of the 1976 Coverage Improvement Program).
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 1977 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1978 AHS questionnaire, page App-20.)
3. All sample housing units that were selected from the list of building permits issued since the 1977 survey. (This sample represented the housing units built in permit-issuing areas, since the 1977 survey.)
4. Units added as the result of the updated listings in selected areas which do not issue building permits.

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 selected to represent housing 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, the other half of each rural cluster (an expected two housing units) was reactivated in 1974. Similarly, for the area sampling frame, 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, for which construction had not been completed at the time of the 1970 census.
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. In the second stage, 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 SAMPLE REDUCTION

By 1977, the addition to the sample (primarily from 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.

ESTIMATION

In 1978, the AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the type A noninterview housing units encountered in the AHS. This noninterview adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

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

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

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

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 housing 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 ratio were derived from data based on the Survey of Construction (SOC). The denominators of the ratio 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 current vacant housing estimates for 4 categories of vacant housing units and to independently derived current housing estimates for 24 categories of occupied housing units. Each of these 24 categories is a combination of the characteristics of residence, tenure, race of head, and sex of head.

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

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

The numerators of the ratios for occupied housing units were derived from data based on the Current Population Survey

(CPS), a sample household survey conducted monthly by the Bureau of the Census. The numerators of the ratios for vacant housing units were derived from data based on the Housing Vacancy Survey (HVS), a quarterly vacancy survey also conducted by the Bureau of the Census. The denominators of the ratios were obtained from the weighted estimates for the AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit in each third-stage ratio estimation category.

The second-stage and the third-stage ratio estimation procedures were 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 18 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 17 categories of new construction units for each of the 4 regions (i.e., 11 categories for conventional new construction units and 6 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 for the repeated second-stage ratio estimation procedure 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 third-stage ratio estimation procedure, as well as the overall estimation procedure, reduced 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 estimates.

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 questionnaires, 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 the sampling error. One common measure of this sampling error is the standard error which measures the precision with which an estimate from a sample approximates the average result of all possible samples. In addition, the standard error, as calculated for these tabulations, 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 estimates in this report.

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 III and IV present the standard errors for estimated percentages.

TABLE I. Standard Errors of Estimated Numbers of National and Regional Housing Units: 1978 (Excluding Estimates of Housing Units for the West Region and of Households With Head of Spanish Origin)

(68 chances out of 100)

Size of estimate (000)	Standard error		Size of estimate (000)	Standard error	
	National (000)	Regional (000)		National (000)	Regional (000)
0	2	2	1,000 . . .	44	43
5	3	3	2,500 . . .	69	70
10	4	4	5,000 . . .	96	102
25	7	7	10,000. . .	131	151
50	10	10	25,000. . .	184	271
100.	14	14	50,000. . .	196	—
250.	22	22	75,000. . .	115	—
500.	31	31			

TABLE II. Standard Errors of Estimated Numbers of Housing Units for the West Region and of Households With Head of Spanish Origin: 1978

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	500.	35
5	3	1,000	50
10	5	2,500	85
25	8	5,000	131
50	11	10,000.	213
100.	15	25,000.	446
250.	24		

Included in tables I through IV are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered 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 III and IV underestimate the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be obtained by letting the standard error of the ratio be approximately equal to:

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

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 there were 8,936,000 renter-occupied, 1-unit structures (including mobile homes and trailers), in the United States in 1978. Interpolation in standard error table I of this appendix shows that the standard error of an estimate of this size is approximately 124,000. The following procedure was used in interpolating:

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

Size of estimate (000)	Standard error (000)
5,000	96
8,936	x
10,000	131

By vertically interpolating between 96 and 131, the entry for "x" is determined to be 124.

$$8{,}936-5{,}000 = 3{,}936$$
$$10{,}000-5{,}000 = 5{,}000$$

$$96+\frac{3{,}936}{5{,}000}(131-96) = 124$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 8,812,000 to 9,060,000 housing units. Therefore, a conclusion that the average estimate of 1978 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 8,738,000 to 9,134,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 8,688,000 to 9,184,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 8,936,000 renter-occupied housing units in 1-unit structures (including mobile homes and trailers), 2,385,000 or 26.7 percent, have all doors covered by storm doors. Interpolation in standard error table III (i.e., interpolation on both the base and percent) of this appendix

shows that the standard error of the above percentage is 0.7 percentage points. The following procedure was used in interpolating.

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

Base of percentage (000)	Estimated percentage		
	25	26.7	50
5,000	0.9	a	1.0
8,936		p	
10,000	0.6	b	0.7

1. By horizontal interpolation between 0.9 and 1.0, the entry for cell "a" is determined to be 0.9.

$$26.7-25.0 = 1.7$$
$$50.0-25.0 = 25.0$$

$$0.9 + \frac{1.7}{25.0}(1.0-0.9) = 0.9$$

2. By horizontal interpolation between 0.6 and 0.7, using the same procedure as step 1, the entry for cell "b" is determined to be 0.6.

3. By vertical interpolation between 0.9 and 0.6, using the same procedure as step 1, the entry for "p" is determined to be 0.7.

Consequently, the 68-percent confidence interval, as shown by these data, is from 26.0 to 27.4 percent; the 90-percent confidence interval is from 25.6 to 27.8 percent; and the 95-percent confidence interval is from 25.3 to 28.1 percent.

Illustration II—Table E-1 of this report shows that in the West Region there were 9,091,000 owner-occupied housing units in 1978. Interpolation in standard error table II of this appendix shows that the standard error of an estimate of this size is approximately 198,000. Consequently, the 68-percent confidence interval is from 8,893,000 to 9,289,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1978 owner-occupied housing units in the West Region lies within a range computed in this way and 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 8,774,000 to 9,408,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 8,695,000 to 9,487,000 housing units with 95 percent confidence.

Table E-1 also shows that of the 9,091,000 owner-occupied housing units in the West Region, 2,096,000 or 23.1 percent, have a central air-conditioning system. Interpolation in standard error table IV (i.e., interpolation on both the base and the percent) of this appendix shows that the standard error of the above percentage is 0.7 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 22.4 to 23.8 percent; the 90-percent confidence interval is from 22.0 to 24.2 percent; and the 95-percent confidence interval is from 21.7 to 24.5 percent.

TABLE III. Standard Errors of Estimated Percentages of Housing Units: 1978 (Excluding Estimated Percentages of Housing Units for the West Region and of Households With Head of Spanish Origin)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	28.0	28.0	28.0	28.0	28.0	28.0	28.0	31.2
10	16.3	16.3	16.3	16.3	16.3	16.3	19.1	22.1
25	7.2	7.2	7.2	7.2	8.4	10.0	12.1	14.0
50	3.8	3.8	3.8	4.3	5.9	7.0	8.5	9.9
100.	1.9	1.9	2.0	3.0	4.2	5.0	6.0	7.0
250.	0.8	0.9	1.2	1.9	2.6	3.2	3.8	4.4
500.	0.4	0.6	0.9	1.4	1.9	2.2	2.7	3.1
1,000	0.2	0.4	0.6	1.0	1.3	1.6	1.9	2.2
2,500	0.08	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	0.9	1.0
10,000	0.02	0.1	0.2	0.3	0.4	0.5	0.6	0.7
25,000	0.01	0.09	0.1	0.2	0.3	0.3	0.4	0.4
50,000	–	0.06	0.09	0.1	0.2	0.2	0.3	0.3
75,000	–	0.05	0.07	0.1	0.2	0.2	0.2	0.3

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 there were 5,150,000 renter-occupied, 1-unit structures (including mobile homes and trailers), with no doors covered by storm doors in the United States in 1978. Thus, the apparent difference between the number of 1978 renter-occupied 1-unit structures (including mobile homes and trailers) with no doors covered and those with all doors covered is 2,765,000. Interpolation in standard error table I of this appendix shows that the standard error of 2,385,000 is approximately 67,000 and that the standard error on an estimate of 5,150,000 is approximately 97,000. Therefore, the standard error of the estimated difference of 2,765,000 is about:

$$118{,}000 = \sqrt{(67{,}000)^2 + (97{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 2,765,000 difference is from 2,647,000 to 2,883,000 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 and would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 2,576,000 to 2,954,000 housing units, and the 95-percent confidence interval is from 2,529,000 to 3,001,000. Thus, we can conclude with 95 percent confidence that the number of 1978 renter-occupied, 1-unit structures (including mobile homes and trailers), with no doors covered is greater than the number with all doors covered.

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 IV. Standard Errors of Estimated Percentages of Housing Units for the West Region and of Households With Head of Spanish Origin: 1978

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	31.7	31.7	31.7	31.7	31.7	31.7	31.7	34.0
10	18.8	18.8	18.8	18.8	18.8	18.8	20.8	24.1
25	8.5	8.5	8.5	8.5	9.1	10.9	13.2	15.2
50	4.4	4.4	4.4	4.7	6.5	7.7	9.3	10.8
100	2.3	2.3	2.3	3.3	4.6	5.4	6.6	7.6
250	0.9	1.0	1.3	2.1	2.9	3.4	4.2	4.8
500	0.5	0.7	1.0	1.5	2.0	2.4	2.9	3.4
1,000	0.2	0.5	0.7	1.0	1.4	1.7	2.1	2.4
2,500	0.09	0.3	0.4	0.7	0.9	1.1	1.3	1.5
5,000	0.05	0.2	0.3	0.5	0.6	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	0.1	0.2	0.3	0.3	0.4	0.5

Illustration of the computation of the 95-percent confidence interval for a median—Table A-2 of this report shows the median income of families and primary individuals in specified owner-occupied housing units was $18,000 in 1978. The base of the distribution, from which this median was determined, is 40,054,000 housing units.

1. From standard error table III, the standard error of a 50-percent characteristic on the base of 40,054,000 is 0.3 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 49.4 and 50.6.
3. From table A-2, it can be seen by cumulating the frequencies for the first 4 categories that 16,216,000 specified owner-occupied housing units, or 40.5 percent, had an income of less than $15,000 and that an additional 6,273,000 specified owner-occupied housing units, or 15.7 percent, had an income from $15,000 to less than $20,000. By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$\$15{,}000 + (\$20{,}000 - \$15{,}000)\left(\frac{49.4 - 40.5}{15.7}\right) = \$17{,}800$$

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

$$\$15{,}000 + (\$20{,}000 - \$15{,}000)\left(\frac{50.6 - 40.5}{15.7}\right) = \$18{,}200$$

Thus, the 95-percent confidence interval ranges from $17,800 to $18,200.

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. 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. No attempt was made to measure the nonsampling errors associated with the estimates for the 1978 Energy Conservation Supplement. However, this was done for the 1977 Energy Conservation Supplement.

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 nonattitudinal and attitudinal questions on the AHS questionnaire were obtained again.

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 1978 reinterview study were not available at the time of publication; however, it is expected that they will be similar to the results of the 1977 reinterview study which are presented in the Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample 1977."

The results of the 1977 reinterview study seem to indicate that there are some very slight improvements in the consistency of response from 1976 to 1977. 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 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 reinterviewer with the original response had the effect of reducing the levels of inconsistency.

To summarize the results of the 1977 reinterview program; 43 percent of nonattitudinal items showed low levels of inconsistency, 36 percent showed moderate levels of inconsistency, which indicated that there were some problems with inconsistent reporting, and 21 percent showed high levels of inconsistency, which indicated that improvements were needed in the data collection methods or that the category concepts themselves were ambiguous.

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

A possible explanation for the results of the AHS and census reinterview studies, as well as the surveys themselves, is that the respondents may lack precise or complete 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—Deficiencies in the representation of conventional new construction for the AHS new construction sample (mentioned previously in the section on estimation) is an example of coverage errors. 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 1978 AHS sample missed about 3.3 percent (i.e., about 375,000 units) of conventional housing units built after April 1970, because the permits for these units, which were built before October 1978, were issued less than 5 months in advance of the survey. The second stage of ratio estimation procedure was employed to reduce the effect of this deficiency although some bias in the AHS estimates of conventional new construction probably still exists.

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 1978 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, i.e., it adjusts the estimate of the total housing inventory to the best available estimate. However, biases of subtotals would still remain.

Rounding errors—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 these tabulations were computed using unrounded data, instead of the published rounded data, they can differ from medians calculated directly from the published data.