# Appendix B

## Source and Reliability of the Estimates


SAMPLE DESIGN ..... App-43
Selection of sample areas ..... App-43
Designation of sample housing units for the 1975 enumeration . App-43
Selection of the 1973 sample .... App-43
housing units ..... App-43
Splitting of the sample ..... App-44
Selection of supplemental sample housing units in rural areas ..... App-44
1970 Census of Population and Housing ..... App-44
ESTIMATION ..... App-44
AHS-National sample ..... App-44
1975 housing inventory ..... App-44
1973-1975 lost units ..... App-45
Ratio estimation procedure of the 1970 Census of Population and Housing ..... App-46
RELIABILITY OF THE ESTIMATES App-46
Sampling errors ..... App-46
Illustrations of the use of the standard error tables.—Illustration I ..... App-47
Illustration II ..... App-48
Differences ..... App-48
Illustration of the computation of the standard error of a difference ..... App-48
Medians ..... App-49
Illustration of the computation of the 95-percent confidence interval for a median ..... App-49
Nonsampling errors ..... App-49
1970 census ..... App-50
Reinterview program ..... App-50
Coverage errors ..... App-50
Rounding errors ..... App-51


### SAMPLE DESIGN

The 1975 estimates are based on data collected in October through December 1975 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 72,600 sample housing units (both occupied and vacant) were eligible for interview in the 1975 Annual Housing Survey. Of this number, 3,700 interviews were not obtained because, for occupied housing units, the occupants were not found at home after repeated calls or were unavailable for some other reason; or, for vacant housing units, no informed respondent could be found after repeated visits. In addition to the 72,600, there were also 7,300 sample units which were visited but found not to be eligible for interview for AHS in terms of collecting information relevant to the 1975 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). Then the PSU's were grouped into 376 strata, 156 of which consisted of only one PSU which was in sample with certainty. These 156 strata were mostly the larger SMSA's and were called self-representing (SR) since the sample from the sample area represented just that PSU. Each one of the other 220 strata consisted of a group of PSU's and were referred to as non-self-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSU's in the stratum as well.

One PSU was selected from each NSR stratum with probability proportionate to the 1970 census population of the PSU. (This resulted in 220 NSR sample PSU's.) In addition, the NSR strata were grouped into 110 pairs and one stratum was picked at random from each pair. From this stratum, an additional PSU was selected independent 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 1975 enumeration.**—The sample housing units designated to be interviewed in the 1975 enumeration consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1974 enumeration.
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 enumeration but which could become eligible in the future) in the 1974 enumeration. (For a list of type A and type B noninterview reasons, see facsimile of 1975 AHS questionnaire, page 1.)
3. All sample housing units that were selected from the list of building permits issued since the 1974 enumeration. (This sample represents the housing units built in permit-issuing areas, since the 1974 enumeration.)

**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 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 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 AHS, and one to be held in reserve for possible future use for AHS. The procedure used to split this sample

into half-samples is described in the next section.

The sample of 1970 census units was selected in several stages of sampling. The first step was the selection of a sample of census enumeration districts (ED's), administrative units used in the 1970 census. The probability of selection for an ED was proportionate to its 1970 population. 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. Those 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 for inclusion in the AHS at the sampling rate of 2 in 1,366. Housing units constructed since the 1970 census in areas which do not issue building permits were brought into the sample as a result of the area sample described above.

**Splitting of the sample.**—The sample selection procedure as described above produced 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). One can expect a minimum loss in precision for segments of size-four housing units in rural areas because of the heterogeneity of this type of housing unit. Segments of size-two, however, were considered to be more optimal for sampling within those areas where one could expect neighboring units to be very similar (e.g., urban areas and new construction units). It is felt that if one were to go to segments of size-four housing units in this type of area, a significant loss in precision would result. A splitting operation was then carried out for segments selected from the census address and new construction frames. This consisted of halving each segment of four housing units that was selected for the sample. Two housing units from each segment were to be included in the survey and two housing units were to be held in reserve. No splitting operation was carried out within the segments selected from the area sampling frame; every other area sample segment of four housing units was used for the survey and the remaining segments 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 segments, this meant that the other half of the segment (an expected two housing units) was reactivated in 1974 if the segment was rural. Similarly for area segments, this meant the entire reserve segment (an expected four housing units) was reactivated in 1974 if the segment 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.

**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-percent, 15-percent, 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 1975 housing inventory and estimates of units removed from the housing inventory between 1973 and 1975 (i.e., 1973-1975 lost units). Each type of estimate employed a separate, though similar, estimation procedure as described below.

**1975 housing inventory.**—The 1975 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 AHS. This noninterview adjustment was done separately for different categories of 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. The 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 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 only employed for AHS new construction sample units (i.e., sample units built April 1, 1970, or later). This procedure was designed to adjust the AHS sample estimates of new construction units to independently derived current estimates for nine categories of new construction units for each of the four regions (i.e., six categories for conventional new construction units and three for new construction mobile homes). These independent estimates were considered to be the best estimates available for the number of new construction units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of new construction units (see the section on nonsampling error).

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

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

The numerators of the ratios were derived from data based on the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. The numerators of the ratios for new construction mobile homes were derived from estimates of mobile home shipments adjusted to account for mobile homes shipped and actually occupied as primary residences.

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 (i.e., the estimates employing the noninterview and first-stage adjustments). The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in each second-stage ratio estimation category.

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

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

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

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

The second- and third-stage ratio estimation procedures were repeated in an iterative process in order to bring the AHS estimates into close agreement with both sets of independent estimates (i.e., the independent estimates employed for both the second-stage ratio estimation procedure as well as those employed for the third-stage ratio estimation procedure). 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 AHS. Therefore, through the use of the three-stage ratio estimation procedure one can expect the sample estimate to be improved substantially when the sample housing population is brought into close agreement with a known distribution of the entire housing population with respect to these basic housing characteristics.

**1973-1975 lost units.**—The 1973-1975 lost unit estimates employed the three-stage ratio estimation procedure used to produce the AHS-National estimates of the 1973 housing inventory, as was described in the 1973 Current Housing Report, series H-150-73A, **General Housing Characteristics for the United States and Regions.** Since the 1973-1975 lost units existed by definition in the 1973 housing inventory, there was a 1973 housing inventory weight associated with each 1973-1975 lost unit. This weight

was used to tabulate the estimates of the characteristics of the 1973-1975 lost units. Also, the general effect of this estimation procedure was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

**Ratio Estimation Procedure of the 1970 Census of Population and Housing.**—This report presents data on the housing characteristics of the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be obtained in the 1970 census report, HC(1)-B1, **Detailed Housing Characteristics, United States Summary.**

## RELIABILITY OF THE ESTIMATES

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

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

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling 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 such that the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

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

The average result of all possible samples may or may not be contained in any particular computed interval. 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 text tables are approximations to the standard errors of various estimates shown in this report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than the precise standard error for any specific item.

Tables I and II present the standard errors applicable to the 1975 housing inventory estimates in this report, and table III presents the standard errors applicable to 1973-1975 lost housing unit estimates in this report. Table I shows the approximate standard errors applicable to all 1975 housing inventory estimates except those pertaining to the specified items in the table.

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

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 IV and V present the standard errors of estimated percentages. Table IV shows the approximate standard errors of all estimated percentages of the 1975 housing inventory except those pertaining to the specified items in table V. Table V also shows the approximate standard errors of all estimated percentages of lost housing units.

The standard errors shown in table V should be used for those specified items. Two-way linear interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables IV and V.

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

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

where: x = the numerator of the ratio
y = the denominator of the ratio
$\sigma_x$ = the standard error of the numerator
$\sigma_y$ = the standard error of the denominator

**TABLE I. Standard Errors of Estimated Number of Housing Units and Vacant Housing Units: 1975 (Excluding estimates of housing units pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Source of Water, and Households with Head of Spanish Origin)**

(68 chances out of 100)

| Size of estimate (000) | Standard error: Total or White (000) | Standard error: Black (000) | Standard error: Vacant (000) | Size of estimate (000) | Standard error: Total or White (000) | Standard error: Black (000) | Standard error: Vacant (000) |
|---|---|---|---|---|---|---|---|
| 5 | 3 | 3 | 3 | 1,000 | 38 | 35 | 42 |
| 10 | 4 | 4 | 4 | 2,500 | 59 | 49 | 71 |
| 25 | 6 | 6 | 6 | 5,000 | 82 | 48 | 109 |
| 50 | 8 | 8 | 9 | 10,000 | 112 | — | — |
| 100 | 12 | 12 | 13 | 25,000 | 156 | — | — |
| 250 | 19 | 19 | 20 | 50,000 | 160 | — | — |
| 500 | 27 | 26 | 29 | 75,000 | 61 | — | — |

**TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Source of Water, and Households with Head of Spanish Origin: 1975**

(68 chances out of 100)

| Size of estimate (000) | Standard error: Total, White, or Spanish origin (000) | Standard error: Black (000) | Size of estimate (000) | Standard error: Total, White, or Spanish origin (000) | Standard error: Black (000) |
|---|---|---|---|---|---|
| 5 | 3 | 3 | 1,000 | 44 | 41 |
| 10 | 4 | 4 | 2,500 | 69 | 57 |
| 25 | 7 | 7 | 5,000 | 96 | 56 |
| 50 | 10 | 10 | 10,000 | 131 | — |
| 100 | 14 | 14 | 25,000 | 183 | — |
| 250 | 22 | 22 | 50,000 | 187 | — |
| 500 | 31 | 30 | 75,000 | 71 | — |

*Illustrations of the use of the standard error tables. Illustration I.*—Table A-1 of this report shows that inside SMSA's in the United States there were 9,188,000 owner-occupied housing units with two persons in 1975. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 107,000. Consequently, the 68-percent confidence interval, as shown by these data, is from 9,081,000 to 9,295,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units with two persons inside 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 9,017,000 to 9,359,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 8,974,000 to 9,402,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 9,188,000 owner-occupied housing units with two persons inside SMSA's, 3,582,000 or 39.0 percent, were in central cities. Interpolation in table IV (i.e., interpolation on both the base and the percent) of this appendix shows that the standard error of the above percentage is approximately .6 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 38.4 to 39.6 percent; the 90-percent confidence interval is from 38.0 to 40.0 percent, and the 95-percent

**TABLE III. Standard Errors of Estimated Number of Lost Housing Units: 1973-1975**

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

confidence interval is from 37.8 to 40.2 percent.

*Illustration II.*—Table A-1 of this report shows in 1975 there were 151,000 housing units in structures with four floors or more (see "Elevator in structure" item) that were outside of SMSA's in the United States. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 14,000. Consequently, the 68-percent confidence interval, as shown by these data, is from 137,000 to 165,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 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 129,000 to 173,000 housing units with 90 percent confidence; and that the average estimate lies within the interval 123,000 to 179,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 151,000 housing units in structures with four floors or more that were outside of SMSA's, 111,000, or 73.5 percent, were in structures that contained elevators. Interpolation in table IV (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage is approximately 4.6 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 68.9 to 78.1 percent; the 90-percent confidence interval is from 66.1 to 80.9 percent; and the 95-percent confidence interval is from 64.3 to 82.7 percent.

*Differences.*—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different areas or the difference between separate and uncorrelated characteristics in the same area. If, however, there is a high, positive correlation between the two characteristics, the formula will overestimate the true error. Also, if there is a high, negative correlation between 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 5,534,000 owner-occupied housing units with three persons in 1975. Thus, the apparent difference between the number of 1975 owner-occupied housing units with two persons and those with three persons is 3,654,000. The standard error of 9,188,000 is approximately 107,000 as is shown above. Table I shows the standard error of an estimate of 5,534,000 to be approximately 85,000. Therefore, the standard error of the estimated difference of 3,654,000 is about

$$137{,}000 = \sqrt{(107{,}000)^2 + (85{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 3,654,000 difference is from 3,517,000 to 3,791,000 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 3,435,000 to 3,873,000 housing units, and the 95-percent confidence interval is from 3,380,000 to 3,928,000. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied housing units inside SMSA's with two persons is greater than the number with three persons.

**TABLE IV. Standard Errors of Estimated Percentages of Housing Units: 1975 (Excluding estimated percentages of housing units pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Sewage, Water Supply, Households with Head of Spanish Origin, and Lost Housing Units: 1973-1975)**

(68 chances out of 100)

| Base of percentage (000) | Estimated percentage[1] | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 or 99 | 2 or 98 | 5 or 95 | 10 or 90 | 15 or 85 | 25 or 75 | 50 |
| 5 ..... | 5.3 | 7.5 | 11.7 | 16.1 | 19.2 | 23.3 | 26.9 |
| 10 ..... | 3.8 | 5.3 | 8.3 | 11.4 | 13.6 | 16.5 | 19.0 |
| 25 ..... | 2.4 | 3.4 | 5.2 | 7.2 | 8.6 | 10.4 | 12.0 |
| 50 ..... | 1.7 | 2.4 | 3.7 | 5.1 | 6.1 | 7.4 | 8.5 |
| 100 ..... | 1.2 | 1.7 | 2.6 | 3.6 | 4.3 | 5.2 | 6.0 |
| 250 ..... | 0.8 | 1.1 | 1.7 | 2.3 | 2.7 | 3.3 | 3.8 |
| 500 ..... | 0.5 | 0.8 | 1.2 | 1.6 | 1.9 | 2.3 | 2.7 |
| 1,000 ..... | 0.4 | 0.5 | 0.8 | 1.1 | 1.3 | 1.6 | 1.9 |
| 2,500 ..... | 0.2 | 0.3 | 0.5 | 0.7 | 0.8 | 1.0 | 1.2 |
| 5,000 ..... | 0.2 | 0.2 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 |
| 10,000 ..... | 0.1 | 0.2 | 0.3 | 0.4 | 0.4 | 0.5 | 0.6 |
| 25,000 ..... | 0.0 | 0.1 | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 |
| 50,000 ..... | 0.05 | 0.08 | 0.1 | 0.2 | 0.2 | 0.2 | 0.3 |
| 75,000 ..... | 0.04 | 0.06 | 0.1 | 0.1 | 0.2 | 0.2 | 0.2 |

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

**TABLE V. Standard Errors of Estimated Percentages of Housing Units Pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Water Supply, Households with Head of Spanish Origin, and Lost Housing Units: 1975**

(68 chances out of 100)

| Base of percentage (000) | Estimated percentage[1] | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 or 99 | 2 or 98 | 5 or 95 | 10 or 90 | 15 or 85 | 25 or 75 | 50 |
| 5 | 6.3 | 8.8 | 13.7 | 18.9 | 22.5 | 27.3 | 31.5 |
| 10 | 4.4 | 6.2 | 9.7 | 13.4 | 15.9 | 19.3 | 22.3 |
| 25 | 2.8 | 3.9 | 6.1 | 8.4 | 10.1 | 12.2 | 14.1 |
| 50 | 2.0 | 2.8 | 4.3 | 6.0 | 7.1 | 8.6 | 10.0 |
| 100 | 1.4 | 2.0 | 3.1 | 4.2 | 5.0 | 6.1 | 7.0 |
| 250 | 0.9 | 1.2 | 1.9 | 2.7 | 3.2 | 3.9 | 4.5 |
| 500 | 0.6 | 0.9 | 1.4 | 1.9 | 2.2 | 2.7 | 3.1 |
| 1,000 | 0.4 | 0.6 | 1.0 | 1.3 | 1.6 | 1.9 | 2.2 |
| 2,500 | 0.3 | 0.4 | 0.6 | 0.8 | 1.0 | 1.2 | 1.4 |
| 5,000 | 0.2 | 0.3 | 0.4 | 0.6 | 0.7 | 0.9 | 1.0 |
| 10,000 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 |
| 25,000 | 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 |

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

*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 such 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 an subtract from 50 percent the standard error determined in step 1; and
3. Using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

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

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

*Illustration of the computation of the 95-percent confidence interval for a median.*—Table A-1 of this report shows the median number of persons in owner-occupied housing units inside SMSA's was 2.9 in 1975. The base of the distribution, from which this median was determined, is 30,383,000 housing units.

1. From table IV, the standard error of a 50-percent characteristic on the base of 30,383,000 is .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 12,884,000 owner-occupied housing units, or 42.4 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 5,534,000 owner-occupied housing units, or 18.2 percent, had three persons (i.e., 2.5 to 3.5 persons). By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$2.5 + (3.5-2.5)\left(\frac{49.2-42.4}{18.2}\right) = 2.9$$

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

$$2.5 + (3.5-2.5)\left(\frac{50.8-42.4}{18.2}\right) = 3.0$$

Thus, the 95-percent confidence interval ranges from 2.9 to 3.0 persons. Although it appears that this confidence interval has the sample estimate as the lower limit, it actually is a reflection of the rounding error associated with this median (see the paragraph on rounding error in the nonsampling error section of this appendix).

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

Obtaining a measurement of the total nonsampling error associated with the

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

**1970 census.**—A number of studies were conducted to measure two types of general errors associated with 1970 census estimates: "Coverage" and "content" errors.

The "coverage" errors determined how completely housing units were counted in the census and included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

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

**Reinterview program.**—For the AHS-National sample a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and thus were the basis for the measurement of the "content" error of these AHS estimates.

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

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

This check was for interviewer evaluation and control.

The results of the 1975 reinterview study are not available; however, it is felt that they will be similar to the results of the 1974 reinterview study, which are presented in the following Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample: 1974." Twenty-five items were included in the 1974 reinterview study, some for the first time. The estimated indexes of inconsistency for these items ranged from 4 to 35 with most items in the 20 to 30 range. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview programs because they had not been previously tested. A 20-50 interval is considered moderate on a range of 0-100, with a high index associated with a high level of response variability.

The 1970 census reinterview results provide illustrations of possible nonsampling errors for some of the items which also appear in the AHS. For example, median value of homes was consistently underestimated by about five percent, and the average monthly costs of electricity and utility gas were consistently overestimated although the effect on the average gross rent figures 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, the results of the reinterview studies are derived from sample surveys so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

**Coverage errors.**—With respect to errors of coverage and estimation for missing data, it was mentioned previously in the section on estimation that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction in permit-issuing areas and new construction mobile homes. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and therefore would have been represented in the sample selected from 1970 census units. However, it has been estimated that the 1975 AHS sample misses about six percent (i.e., about 600,000 units) of all conventional new construction (i.e., all conventional housing units built after April 1970, in both permit-issuing and non-permit-issuing areas) because the permits for these missed units, which were built after April 1970, were issued before 1970.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new construction mobile homes. However, new mobile homes in segments where area sampling methods were used do come into the AHS sample. In addition, new mobile homes in segments sampled from the 1970 census list also come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. However, new mobile homes in these segments that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. It has been estimated that the 1975 AHS sample misses at least

200,000 new mobile homes. The second-stage ratio estimation was employed to reduce the effect of both these deficiencies although some bias in the AHS sample still exists. This is especially true for new mobile homes since it is believed that the corresponding independent estimate substantially understates the actual new mobile home inventory.

**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. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

# Appendix B

## Source and Reliability of the Estimates


SAMPLE DESIGN ..... App-43
Selection of sample areas ..... App-43
Designation of sample housing units for the 1975 enumeration App-43
Selection of the 1973 sample housing units ..... App-43
Splitting of the sample ..... App-44
Selection of supplemental sample housing units in rural areas .... App-44
ESTIMATION ..... App-44
RELIABILITY OF THE ESTIMATES App-45
Sampling errors ..... App-45
Illustrations of the use of the standard error tables ..... App-46
Differences ..... App-48
Illustration of the computation of the standard error of a difference ..... App-48
Nonsampling errors ..... App-48
Reinterview program ..... App-48
Coverage errors ..... App-49
Rounding errors ..... App-49


### SAMPLE DESIGN

The 1975 estimates are based on data collected in October through December 1975 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 72,600 sample housing units (both occupied and vacant) were eligible for interview in the 1975 Annual Housing Survey. Of this number, 3,700 interviews were not obtained because, for occupied housing units, the occupants were not found at home after repeated calls or were unavailable for some other reason; or, for vacant housing units, no informed respondent could be found after repeated visits. In addition to the 72,600, there were also 7,300 sample units which were visited but found not to be eligible for interview for AHS in terms of collecting information relevant to the 1975 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). Then the PSU's were grouped into 376 strata, 156 of which consisted of only one PSU which was in sample with certainty. These 156 strata were mostly the larger SMSA's and were called self-representing (SR) since the sample from the sample area represented just that PSU. Each one of the other 220 strata consisted of a group of PSU's and were referred to as non-self-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSU's in the stratum as well.

One PSU was selected from each NSR stratum with probability proportionate to the 1970 census population of the PSU. (This resulted in 220 NSR sample PSU's.) In addition, the NSR strata were grouped into 110 pairs and one stratum was picked at random from each pair. From this stratum, an additional PSU was selected independent 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 1975 enumeration.**—The sample housing units designated to be interviewed in the 1975 enumeration consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1974 enumeration.

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 enumeration but which could become eligible in the future) in the 1974 enumeration. (For a list of type A and type B noninterview reasons, see facsimile of 1975 AHS questionnaire, page 1.)

3. All sample housing units that were selected from the list of building permits issued since the 1974 enumeration. (This sample represents the housing units built in permit-issuing areas, since the 1974 enumeration.)

**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 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 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 AHS and one to be held in reserve for possible future use for AHS. The procedure used to split this sample into half-samples is described in the next section:

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

adequate (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. Those 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 for inclusion in the AHS at the sampling rate of 2 in 1,366. Housing units constructed since the 1970 census in areas which do not issue building permits were brought into the sample as a result of the area sample described above.

**Splitting of the sample.**—The sample selection procedure as described above produced 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). One can expect a minimum loss in precision for segments of size-four housing units in rural areas because of the heterogeneity of this type of housing unit. Segments of size two, however, were considered to be more optimal for sampling within those areas where one could expect neighboring units to be very similar (e.g., urban areas and new construction units). It is felt that if one were to go to segments of size-four housing units in this type of area, a significant loss in precision would result. A splitting operation was then carried out for segments selected from the census address and new construction frames. This consisted of halving each segment of four housing units that was selected for the sample. Two housing units from each segment were to be included in the survey and two housing units were to be held in reserve. No splitting operation was carried out within the segments selected from the area sampling frame; every other area sample segment of four housing units was used for the survey and the remaining segments 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 which was selected in the original sampling operations in 1973 from rural areas only. For the reserve sample selected in census address and new construction segments, this meant that the other half of the segment (an expected two housing units) was reactivated in 1974 if the segment was rural. Similarly for area segments, the entire reserve segment (an expected four housing units) was reactivated in 1974 if the segment 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.

## ESTIMATION

The 1975 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 AHS. This noninterview adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

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

The first-stage ratio estimation procedure was employed for sample housing units from non-self-representing (NSR) PSU's only. The 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 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 only employed for AHS new construction sample units (i.e., sample units built April 1, 1970, or later). This procedure was designed to adjust the AHS sample estimates of new construction units to independently derived current estimates for nine categories of new construction units for each of the four regions (i.e., six categories for conventional new construction units and three for new construction mobile homes). These independent estimates were considered to be the best estimates available for the number of new construction units. The adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of new construction units (see the section on nonsampling error).

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

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

The numerators of the ratios were derived from data based on the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. The numerators of the ratios for new construction mobile homes were derived from estimates of mobile home shipments adjusted to account for mobile homes shipped and actually occupied as primary residences.

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 (i.e., the estimates employing the noninterview and first-stage adjustments).

The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in each second-stage ratio estimation category.

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

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

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

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

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

The second- and third-stage ratio estimation procedures were repeated in an iterative process in order to bring the AHS estimates into close agreement with both sets of independent estimates (i.e., the independent estimates employed for both the second-stage ratio estimation process as well as those employed for the third-stage ratio estimation process). 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 AHS. Therefore, through the use of the three-stage ratio estimation procedure one can expect the sample estimate to be improved substantially when the sample housing population is brought into close agreement with a known distribution of the entire housing population with respect to these basic housing characteristics.

## 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 enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling 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 that include the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

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

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

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

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

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

Tables I and II present the standard errors applicable to housing unit estimates in this report. Table I shows the approximate standard errors applicable to all housing unit estimates except those pertaining to the specified items in the table.

**TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, and Households with Head of Spanish Origin: 1975**

(68 chances out of 100)

| Size of estimate (000) | Standard error: Total, White, or Spanish origin (000) | Standard error: Black (000) | Size of estimate (000) | Standard error: Total, White, or Spanish origin (000) | Standard error: Black (000) |
|---|---|---|---|---|---|
| 5 ..... | 3 | 3 | 1,000 .... | 44 | 41 |
| 10 ..... | 4 | 4 | 2,500 .... | 69 | 57 |
| 25 ..... | 7 | 7 | 5,000 .... | 96 | 56 |
| 50 ..... | 10 | 10 | 10,000 ... | 131 | – |
| 100 ..... | 14 | 14 | 25,000 ... | 183 | – |
| 250 ..... | 22 | 22 | 50,000 ... | 187 | – |
| 500 ..... | 31 | 30 | 75,000 ... | 71 | – |

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

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

Tables III and IV present the standard errors of estimated percentages. Table III shows the approximate standard errors of all estimated percentages of housing units except those pertaining to the specified items in table II.

The standard errors shown in table IV should be used for those specified items. Two-way linear interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables III and IV.

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

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

where: $x$ = the numerator of the ratio
$y$ = the denominator of the ratio
$\sigma_x$ = the standard error of the numerator
$\sigma_y$ = 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 13,254,000 renter-occupied housing units with common

**TABLE I. Standard Errors of Estimated Number of Housing Units: 1975 (Excluding estimates of housing units pertaining to Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, 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) |
|---|---|---|---|---|---|
| 5 ..... | 3 | 3 | 1,000 .... | 38 | 35 |
| 10 ..... | 4 | 4 | 2,500 .... | 59 | 49 |
| 25 ..... | 6 | 6 | 5,000 .... | 82 | 48 |
| 50 ..... | 8 | 8 | 10,000 ... | 112 | – |
| 100 ..... | 12 | 12 | 25,000 ... | 156 | – |
| 250 ..... | 19 | 19 | 50,000 ... | 160 | – |
| 500 ..... | 27 | 26 | 75,000 ... | 61 | – |

**TABLE III. Standard Errors of Estimated Percentages of Housing Units: 1975 (Excludes estimated percentages of housing units pertaining to Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, and Households with Head of Spanish Origin)**

(68 chances out of 100)

| Base of percentage (000) | Estimated percentage[1] | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 or 99 | 2 or 98 | 5 or 95 | 10 or 90 | 15 or 85 | 25 or 75 | 50 |
| 5 | 5.3 | 7.5 | 11.7 | 16.1 | 19.2 | 23.3 | 26.9 |
| 10 | 3.8 | 5.3 | 8.3 | 11.4 | 13.6 | 16.5 | 19.0 |
| 25 | 2.4 | 3.4 | 5.2 | 7.2 | 8.6 | 10.4 | 12.0 |
| 50 | 1.7 | 2.4 | 3.7 | 5.1 | 6.1 | 7.4 | 8.5 |
| 100 | 1.2 | 1.7 | 2.6 | 3.6 | 4.3 | 5.2 | 6.0 |
| 250 | 0.8 | 1.1 | 1.7 | 2.3 | 2.7 | 3.3 | 3.8 |
| 500 | 0.5 | 0.8 | 1.2 | 1.6 | 1.9 | 2.3 | 2.7 |
| 1,000 | 0.4 | 0.5 | 0.8 | 1.1 | 1.3 | 1.6 | 1.9 |
| 2,500 | 0.2 | 0.3 | 0.5 | 0.7 | 0.8 | 1.0 | 1.2 |
| 5,000 | 0.2 | 0.2 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 |
| 10,000 | 0.1 | 0.2 | 0.3 | 0.4 | 0.4 | 0.5 | 0.6 |
| 25,000 | 0.07 | 0.1 | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 |
| 50,000 | 0.05 | 0.08 | 0.1 | 0.2 | 0.2 | 0.2 | 0.3 |
| 75,000 | 0.04 | 0.06 | 0.1 | 0.1 | 0.2 | 0.2 | 0.2 |

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

**TABLE IV. Standard Errors of Estimated Percentages of Housing Units Pertaining to Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, and Households with Head of Spanish Origin: 1975**

(68 chances out of 100)

| Base of percentage (000) | Estimated percentage[1] | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 or 99 | 2 or 98 | 5 or 95 | 10 or 90 | 15 or 85 | 25 or 75 | 50 |
| 5 | 6.3 | 8.8 | 13.7 | 18.9 | 22.5 | 27.3 | 31.5 |
| 10 | 4.4 | 6.2 | 9.7 | 13.4 | 15.9 | 19.3 | 22.3 |
| 25 | 2.8 | 3.9 | 6.1 | 8.4 | 10.1 | 12.2 | 14.1 |
| 50 | 2.0 | 2.8 | 4.3 | 6.0 | 7.1 | 8.6 | 10.0 |
| 100 | 1.4 | 2.0 | 3.1 | 4.2 | 5.0 | 6.1 | 7.0 |
| 250 | 0.9 | 1.2 | 1.9 | 2.7 | 3.2 | 3.9 | 4.5 |
| 500 | 0.6 | 0.9 | 1.4 | 1.9 | 2.2 | 2.7 | 3.1 |
| 1,000 | 0.4 | 0.6 | 1.0 | 1.3 | 1.6 | 1.9 | 2.2 |
| 2,500 | 0.3 | 0.4 | 0.6 | 0.8 | 1.0 | 1.2 | 1.4 |
| 5,000 | 0.2 | 0.3 | 0.4 | 0.6 | 0.7 | 0.9 | 1.0 |
| 10,000 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 |
| 25,000 | 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 |

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

stairways. Interpolation in table I shows that the standard error of an estimate of this size is approximately 122,000. Consequently, the 68-percent confidence interval, as shown by these data, is from 13,132,000 to 13,376,000 housing units. Therefore, a conclusion that the average estimate of 1975 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 13,059,000 to 13,449,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 13,010,000 to 13,498,000 housing units with 95 percent confidence.

Table A-2 also shows that of the 13,254,000 renter-occupied housing units with common stairways, 11,756,000, or 88.7 percent, were located inside SMSA's. Interpolation in table III (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage is .4 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 88.3 to 89.1 percent; the 90-percent confidence interval is from 88.1 to 89.3 percent; and the 95-percent confidence interval is from 87.9 to 89.5 percent.

*Illustration II.*—Table A-3 of this report shows that in the United States in 1975 there were 458,000 owner-occupied housing units which had sewage disposal breakdowns. Interpolation in table II of this appendix shows that the standard error of an estimate of this size is approximately 29,000. Consequently, the 68-percent confidence interval is from 429,000 to 487,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units which had sewage disposal breakdowns lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples, lies within the interval from 412,000 to 504,000 housing units with 90 percent

confidence; and that the average estimate lies within the interval from 400,000 to 516,000 housing units with 95 percent confidence.

Table A-3 also shows that of the 458,000 1975 owner-occupied housing units which had sewage disposal breakdowns, 43,000, or 9.4 percent, had breakdowns three times or more. Interpolation in table IV (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage is 1.9 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 7.5 to 11.3 percent; the 90-percent confidence interval is from 6.4 to 12.4 percent; and the 95-percent confidence interval is from 5.6 to 13.2 percent.

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

*Illustration of the computation of the standard error of a difference.*—Table A-3 of this report shows that in the United States in 1975 there were 351,000 owner-occupied housing units, which had exactly one sewage disposal breakdown. Thus, the apparent difference between the number of 1975 owner-occupied housing units that had breakdowns three times or more and that had breakdowns just one time, is 308,000. Interpolation in table II shows that the standard error on an estimate of 351,000 to be approximately 26,000 and the standard error on an estimate of 43,000 to be approximately 9,000. Therefore, the standard error of the estimated difference of 308,000 is about

$$28{,}000 = \sqrt{(26{,}000)^2 + (9{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 308,000 difference is from 280,000 to 336,000 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 263,000 to 353,000 housing units, and the 95-percent confidence interval is from 252,000 to 364,000. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied housing units which had three or more sewage disposal breakdowns is less than the number that had exactly one sewage disposal breakdown, since the 95-percent confidence interval of this difference does not include zero or negative values.

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

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error.

**Reinterview program.**—For the AHS-National sample a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and thus were the basis for the measurement of the "content" error of these AHS estimates.

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

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

This check was for interviewer evaluation and control.

The results of the 1975 reinterview study are not available; however, it is felt that they will be similar to the results of the 1974 reinterview study, which are presented in the following Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample: 1974."

Twenty-five items were included in the 1974 reinterview study, some for the first time. The estimated indexes of inconsistency for these items ranged from 4 to 35 with most items in the 20 to 30 range. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview programs because they had not been previously tested. A 20-50 interval is considered moderate on a range of 0-100, with a high index associated with a high level of response variability.

The 1970 census reinterview results provide illustrations of possible nonsampling errors for some of the items which also appear in the AHS. For example, median value of homes was

consistently underestimated by about five percent, and the average monthly costs of electricity and utility gas were consistently overestimated although the effect on the average gross rent figures 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, the results of the reinterview studies are derived from sample surveys so that there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

**Coverage Errors.**—With respect to errors of coverage and estimation for missing data, it was mentioned previously (in the section on estimation) that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction in permit-issuing areas and new construction mobile homes. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and, therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that the 1975 AHS sample misses about five percent (i.e., about 600,000 units) of all conventional new construction (i.e., all conventional housing units built after April 1970 in both permit-issuing and non-permit-issuing areas) because the permits for these missed units, which were built after April 1970, were issued before 1970.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new construction mobile homes. However, new mobile homes in segments where area sampling methods were used do come into the AHS sample. In addition, new mobile homes in segments sampled from the 1970 census list also come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. However, new mobile homes in these segments that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. It has been estimated that the 1975 AHS sample misses at least 200,000 new mobile homes. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies although some bias in the AHS sample still exists. This is especially true for new mobile homes, since it is believed that the corresponding independent estimate substantially understates the actual new mobile home inventory.

**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. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of this survey.

☆U.S. GOVERNMENT PRINTING OFFICE: 1977-240-869:1021

# Appendix B

## Source and Reliability of the Estimates


SAMPLE DESIGN . . . . . App-43
Selection of sample areas . . . . . App-43
Designation of sample housing units for the 1975 enumeration . App-43
Selection of the 1973 sample housing units . . . . . App-43
Splitting of the sample . . . . . App-44
Selection of supplemental sample housing units in rural areas. . . . . App-44
ESTIMATION. . . . . App-44
RELIABILITY OF THE ESTIMATES App-45
Sampling errors . . . . . App-45
Illustrations of the use of the standard error tables. Illustration I. . . . . App-46
Illustration II . . . . . App-47
Differences . . . . . App-48
Illustration of the computation of the standard error of a difference . . . . . App-48
Medians . . . . . App-48
Illustration of the computation of the 95-percent confidence interval for a median . . . . . App-48
Nonsampling errors . . . . . App-48
Reinterview program . . . . . App-49
Coverage errors . . . . . App-49
Rounding errors. . . . . App-49


### SAMPLE DESIGN

The 1975 estimates are based on data collected in October through December 1975 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 72,600 sample housing units (both occupied and vacant) were eligible for interview in the 1975 Annual Housing Survey. Of this number, 3,700 interviews were not obtained because, for occupied housing units, the occupants were not found at home after repeated calls or were unavailable for some other reason; or, for vacant housing units, no informed respondent could be found after repeated visits. In addition to the 72,600, there were also 7,300 sample units which were visited but found not to provide information relevant to the 1975 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). Then the PSU's were grouped into 376 strata, 156 of which consisted of only one PSU which was in sample with certainty. These 156 strata were mostly the larger SMSA's and were called self-representing (SR) since the sample from the sample area represented just that PSU. Each one of the other 220 strata consisted of a group of PSU's and were referred to as non-self-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSU's in the stratum as well.

One PSU was selected from each NSR stratum with probability proportionate to the 1970 census population of the PSU. (This resulted in 220 NSR sample PSU's). In addition, the NSR strata were grouped into 110 pairs and one stratum was picked at random from each pair. From this stratum, an additional PSU was selected independent 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 total of 461 PSU's.

**Designation of sample housing units for the 1975 enumeration.**—The sample housing units designated to be interviewed in the 1975 enumeration consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1974 enumeration.
2. All 1974 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 enumeration but which could become eligible in the future. (For a list of type A and type B noninterview reasons, see facsimile of 1975 AHS questionnaire, page 1.)
3. All sample housing units that were selected from the list of building permits issued since the 1974 enumeration. (This sample represents the housing units built in permit-issuing areas since the 1974 enumeration.)

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

The sample of 1970 census units was selected in several stages of sampling. The first step was selecting a sample of census enumeration districts (ED's), administrative units used in the 1970 census. The probability of selection for an ED was proportionate to its 1970 population. 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), area sampling methods were used in the selection process. 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. Those 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 for inclusion in the AHS at the sampling rate of 2 in 1,366. As a result of the area sampling methods described above, housing units constructed since the 1970 census in areas which do not issue building permits were brought into the sample.

**Splitting of the sample.**—The sample selection procedure as described above produced 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). A minimum loss in precision can be expected for segments of size-four housing units in rural areas because of the heterogeneity of this type of housing unit. Segments of size-two, however, were considered to be more optimal for sampling within those areas where neighboring units could be expected to be very similar (e.g., urban areas and new construction units). If segments of size-four housing units in this type of area were sampled, it is felt that a significant loss in precision would result. A splitting operation was then carried out for segments selected from the census address and new construction frames. Each segment of four housing units selected for the sample was halved. Two housing units from each segment were to be included in the survey and two housing units were to be held in reserve. No splitting operation was carried out within the segments selected from the area sampling frame; every other area sample segment of four housing units was used for the survey and the remaining segments were assigned to the reserve sample.

**Selection of supplemental sample housing units in rural areas.**—In 1974, the reliability of the AHS estimates of rural housing characteristics was increased 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 segments, the other half of the segment (an expected two housing units) was reactivated in 1974 if the segment was rural. Similarly for area segments, the entire reserve segment (an expected four housing units) was reactivated in 1974 if the segment 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.

## ESTIMATION

The 1975 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 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. The 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 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 only employed for AHS new construction sample units (i.e., sample units built April 1, 1970, or later). This procedure was designed to adjust the AHS sample estimates of new construction units to independently derived current estimates for nine categories of new construction units for each of the four regions (i.e., six categories for conventional, new construction units and three for new construction mobile homes). These independent estimates were considered to be the best estimates available for the number of new construction units. This adjustment was necessary to correct for known deficiencies in the

AHS sample with regard to representation of new construction units (see the section on nonsampling error).

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

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

The numerators of the ratios were derived from data based on the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. The numerators of the ratios for new construction mobile homes were derived from estimates of mobile home shipments adjusted to account for mobile homes shipped and actually occupied as primary residences.

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 (i.e., the estimates employing the noninterview and first-stage adjustments).

The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in each second-stage ratio estimation category.

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

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

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

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

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

The second-and third-stage ratio estimation procedures were repeated in an iterative process in order to bring the AHS estimates into close agreement with both sets of independent estimates (i.e., the independent estimates employed for both the second-stage ratio estimation process as well as those employed for the third-stage ratio estimation process). 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 AHS. Therefore, through the use of the three-stage ratio estimation procedure one can expect the sample estimate to be improved substantially when the sample housing population is brought into close agreement with a known distribution of the entire housing population with respect to these basic housing characteristics.

## 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 enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling 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 such that the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions and an estimate and its estimated standard error were calculated from each sample, then:

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

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

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

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

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

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

Tables III and IV present the standard errors of estimated percentages. Table III shows the approximate standard errors of all estimated percentages of housing units except those pertaining to the specified items in table II.

The standard errors shown in table IV should be used for those specified items. Two-way linear interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables III and IV.

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

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

Let: $x$ = the numerator
$y$ = the denominator
$\sigma_x$ = the standard error of the numerator
$\sigma_y$ = the standard error of the denominator

*Illustrations of the use of the standard error tables. Illustration I.*—Table A-2 of

**TABLE I. Standard Errors of Estimated Number of Housing Units: 1975 (Excluding estimates of housing units pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Source of Water, and Households with Head of Spanish Origin)**

(68 chances out of 100)

| Size of estimate (000) | Standard error: Total or White (000) | Standard error: Black (000) | Size of estimate (000) | Standard error: Total or White (000) | Standard error: Black (000) |
|---|---|---|---|---|---|
| 5 ..... | 3 | 3 | 1,000 .... | 38 | 35 |
| 10 ..... | 4 | 4 | 2,500 .... | 59 | 49 |
| 25 ..... | 6 | 6 | 5,000 .... | 82 | 48 |
| 50 ..... | 8 | 8 | 10,000 ... | 112 | – |
| 100 ..... | 12 | 12 | 25,000 ... | 156 | – |
| 250 ..... | 19 | 19 | 50,000 ... | 160 | – |
| 500 ..... | 27 | 26 | 75,000 ... | 61 | – |

**TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Source of Water, and Households with Head of Spanish Origin: 1975**

(68 chances out of 100)

| Size of estimate (000) | Standard error: Total, White, or Spanish origin (000) | Standard error: Black (000) | Size of estimate (000) | Standard error: Total, White or Spanish origin (000) | Standard error: Black (000) |
|---|---|---|---|---|---|
| 5 ..... | 3 | 3 | 1,000 .... | 44 | 41 |
| 10 ..... | 4 | 4 | 2,500 .... | 69 | 57 |
| 25 ..... | 7 | 7 | 5,000 .... | 96 | 56 |
| 50 ..... | 10 | 10 | 10,000 ... | 131 | – |
| 100 ..... | 14 | 14 | 25,000 ... | 183 | – |
| 250 ..... | 22 | 22 | 50,000 ... | 187 | – |
| 500 ..... | 31 | 30 | 75,000 ... | 71 | – |

**TABLE III. Standard Errors of Estimated Percentages of Housing Units: 1975 (Excludes estimated percentages of housing units pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Source of Water, and Households with Head of Spanish Origin)**

(68 chances out of 100)

| Base of percentage (000) | Estimated percentage[1] | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 or 99 | 2 or 98 | 5 or 95 | 10 or 90 | 15 or 85 | 25 or 75 | 50 |
| 5 | 5.3 | 7.5 | 11.7 | 16.1 | 19.2 | 23.3 | 26.9 |
| 10 | 3.8 | 5.3 | 8.3 | 11.4 | 13.6 | 16.5 | 19.0 |
| 25 | 2.4 | 3.4 | 5.2 | 7.2 | 8.6 | 10.4 | 12.0 |
| 50 | 1.7 | 2.4 | 3.7 | 5.1 | 6.1 | 7.4 | 8.5 |
| 100 | 1.2 | 1.7 | 2.6 | 3.6 | 4.3 | 5.2 | 6.0 |
| 250 | 0.8 | 1.1 | 1.7 | 2.3 | 2.7 | 3.3 | 3.8 |
| 500 | 0.5 | 0.8 | 1.2 | 1.6 | 1.9 | 2.3 | 2.7 |
| 1,000 | 0.4 | 0.5 | 0.8 | 1.1 | 1.3 | 1.6 | 1.9 |
| 2,500 | 0.2 | 0.3 | 0.5 | 0.7 | 0.8 | 1.0 | 1.2 |
| 5,000 | 0.2 | 0.2 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 |
| 10,000 | 0.1 | 0.2 | 0.3 | 0.4 | 0.4 | 0.5 | 0.6 |
| 25,000 | 0.07 | 0.1 | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 |
| 50,000 | 0.05 | 0.08 | 0.1 | 0.2 | 0.2 | 0.2 | 0.3 |
| 75,000 | 0.04 | 0.06 | 0.1 | 0.1 | 0.2 | 0.2 | 0.2 |

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

**TABLE IV. Standard Errors of Estimated Percentages of Housing Units Pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Source of Water, and Households with Head of Spanish Origin: 1975**

(68 chances out of 100)

| Base of percentage (000) | Estimated percentage[1] | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 or 99 | 2 or 98 | 5 or 95 | 10 or 90 | 15 or 85 | 25 or 75 | 50 |
| 5 | 6.3 | 8.8 | 13.7 | 18.9 | 22.5 | 27.3 | 31.5 |
| 10 | 4.4 | 6.2 | 9.7 | 13.4 | 15.9 | 19.3 | 22.3 |
| 25 | 2.8 | 3.9 | 6.1 | 8.4 | 10.1 | 12.2 | 14.1 |
| 50 | 2.0 | 2.8 | 4.3 | 6.0 | 7.1 | 8.6 | 10.0 |
| 100 | 1.4 | 2.0 | 3.1 | 4.2 | 5.0 | 6.1 | 7.0 |
| 250 | 0.9 | 1.2 | 1.9 | 2.7 | 3.2 | 3.9 | 4.5 |
| 500 | 0.6 | 0.9 | 1.4 | 1.9 | 2.2 | 2.7 | 3.1 |
| 1,000 | 0.4 | 0.6 | 1.0 | 1.3 | 1.6 | 1.9 | 2.2 |
| 2,500 | 0.3 | 0.4 | 0.6 | 0.8 | 1.0 | 1.2 | 1.4 |
| 5,000 | 0.2 | 0.3 | 0.4 | 0.6 | 0.7 | 0.9 | 1.0 |
| 10,000 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 |
| 25,000 | 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 |

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

this report shows that in the United States there were 9,517,000 specified owner-occupied housing units with two bedrooms in 1975. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 109,000. Consequently, the 68-percent confidence interval is from 9,408,000 to 9,626,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units with two bedrooms lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples, lies within the interval from 9,343,000 to 9,691,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 9,299,000 to 9,735,000 housing units with 95 percent confidence.

Table A-2 also shows that of the 9,517,000 specified owner-occupied housing units with two bedrooms, 1,672,000, or 17.6 percent, were of value between $15,000 and $19,999. Interpolation in table III (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage is approximately .4 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 17.2 to 18.0 percent; the 90-percent confidence interval is from 17.0 to 18.2 percent; and the 95-percent confidence interval is from 16.8 to 18.4 percent.

*Illustration II.*—Table A-2 shows that in the nation there were 63,000 specified owner-occupied housing units whose cooking fuel was wood. Interpolation in table II shows that the standard error of an estimate of this size is approximately 11,000 (table II should be used since this estimate pertains to cooking fuel).

Consequently, we could conclude with 68 percent confidence that the average estimate, derived from all possible samples, of owner-occupied housing units whose cooking fuel is wood lies within the interval from 52,000 to 74,000 housing units. Similarly, the 90-percent confi-

dence interval is from 45,000 to 81,000 housing units; and the 95-percent confidence interval is from 41,000 to 85,000 housing units.

Table A-2 also shows that of the 63,000 specified owner-occupied housing units whose cooking fuel is wood, 27,000, or 42.9 percent, were valued at less than $10,000. Interpolation in table IV (i.e., interpolation on both the base and the percent) of this appendix shows that the standard error of the above percentage is approximately 8.9 percentage points. Consequently, the 68-percent confidence interval is from 34.0 to 51.8 percent; the 90-percent confidence interval is from 28.7 to 57.1 percent; and the 95-percent confidence interval is from 25.1 to 60.7 percent.

*Differences.*—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different areas or the difference between separate and uncorrelated characteristics in the same area. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

*Illustration of the computation of the standard error of a difference.*—Table A-2 of this report shows that in the United States there were 1,388,000 specified owner-occupied housing units with two bedrooms valued between $10,000 and $14,999 in 1975. Thus the apparent difference between the number of 1975 specified owner-occupied housing units with two bedrooms valued between $10,000 and $14,999 and those valued between $15,000 and $19,999 is 284,000. Table I shows the standard error of an estimate of 1,672,000 to be approximately 47,000 and the standard error of an estimate of 1,388,000 to be approximately 43,000. Therefore, the standard error of an estimated difference of 284,000 is about

$$64{,}000 = \sqrt{(47{,}000)^2 + (43{,}000)^2}$$

Consequently, the 68-percent confidence interval for the difference is from 220,000 to 348,000 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 182,000 to 386,000 housing units, and the 95-percent confidence interval is from 156,000 to 412,000 housing units. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied housing units with two bedrooms valued between $15,000 and $19,999 is greater than the number valued between $10,000 and $14,999 since the 95-percent confidence interval 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 such that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

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

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

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

*Illustration of the computation of the 95-percent confidence interval for a median.*—Table A-2 of this report shows the median value of specified owner-occupied housing units with two bedrooms was $21,900 in 1975. The base of the distribution, from which this median was determined is 9,517,000 housing units.

1. Table III shows that the standard error of 50 percent on a base of 9,517,000 is approximately .6 percent.
2. To obtain a two-standard error confidence interval on the estimated median initially, add to and subtract from 50 percent twice the standard error determined in 1. This yields percentage limits of 48.8 and 51.2.
3. From table A-2, it can be seen by cumulating the frequencies for the first three categories that 4,241,000 owner-occupied housing units with two bedrooms, or 44.6 percent, had a value less than $20,000, and an additional 1,355,000 owner-occupied housing units with two bedrooms, or 14.2 percent, had a value between $20,000 and $24,999.

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

$$\$20{,}000 + (\$5{,}000)\left(\frac{48.8-44.6}{14.2}\right) = \$21{,}500$$

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

$$\$20{,}000 + (\$5{,}000)\left(\frac{51.2-44.6}{14.2}\right) = \$22{,}300$$

Thus, the 95-percent confidence interval ranges from $21,500 to $22,300.

**Nonsampling errors.**—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or

coding the data, and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error.

**Reinterview program.**—For the AHS-National sample a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and thus were the basis for the measurement of the "content" error of these AHS estimates.

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

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

This check was for interviewer evaluation and control.

The results of the 1975 reinterview study are not available; however, it is felt that they will be similar to the results of the 1974 reinterview study, which are presented in the following Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample: 1974." Twenty-five items were included in the 1974 reinterview study, some for the first time. The estimated indexes of inconsistency for these items ranged from 4 to 35 with most items in the 20 to 30 range. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview programs because they had not been previously tested. A 20-50 interval is considered moderate on a range of 0-100, with a high index associated with a high level of response variability.

The 1970 census reinterview results provide illustrations of possible nonsampling errors for some of the items which also appear in the AHS. For example, median value of homes was consistently underestimated by about five percent, and the average monthly costs of electricity and utility gas were consistently overestimated although the effect on the average gross rent figures 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, the results of the reinterview studies are derived from sample surveys so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

**Coverage errors.**—It was mentioned previously, in the section on estimation, that the AHS new construction sample had deficiencies in the representation of both conventional new construction in permit-issuing areas and new construction mobile homes. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and therefore would have been represented in the sample selected from 1970 census units. However, it has been estimated that the 1975 AHS sample misses about five percent (i.e., about 600,000 units) of all conventional new construction (i.e., all conventional housing units built after April 1970 in both permit-issuing and nonpermit-issuing areas) because the permits for these missed units, which were built after April 1970, were issued before 1970.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new construction mobile homes. However, new mobile homes in segments where area sampling methods were used do come into the AHS sample. In addition, new mobile homes in segments sampled from the 1970 census list also come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. However, new mobile homes in these segments that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. It has been estimated that the 1975 AHS sample misses at least 200,000 new mobile homes. The second-stage ratio estimation was employed to reduce the effect of both these deficiencies although some bias in the AHS sample still exists. This is especially true for new mobile homes, since it is belived that the corresponding independent estimate substantially understates the actual new mobile home inventory.

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

# Appendix B

## Source and Reliability of the Estimates


SAMPLE DESIGN ..... App-43
Selection of sample areas ..... App-43
Designation of sample housing units for the 1975 enumeration . App-43
Selection of the 1973 sample housing units ..... App-43
Splitting of the sample ..... App-44
Selection of supplemental sample housing units in rural areas ..... App-44
ESTIMATION ..... App-44
RELIABILITY OF THE ESTIMATES ..... App-45
Sampling errors ..... App-45
Illustrations of the use of the standard error tables. Illustration I ..... App-46
Illustration II ..... App-47
Differences ..... App-48
Illustration of the computation of the standard error of a difference App-48
Medians ..... App-48
Illustration of the computation of 95-percent confidence interval for a median ..... App-48
Nonsampling errors ..... App-48
Reinterview program ..... App-49
Coverage errors ..... App-49
Rounding errors ..... App-49


### SAMPLE DESIGN

The 1975 estimates are based on data collected in October through December 1975 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 72,600 sample housing units (both occupied and vacant) were eligible for interview in the 1975 Annual Housing Survey. Of this number, 3,700 interviews were not obtained because, for occupied housing units, the occupants were not found at home after repeated calls or were unavailable for some other reason; or, for vacant housing units, no informed respondent could be found after repeated visits. In addition to the 72,600, there were also 7,300 sample units which were visited but found not to provide information relevant to the 1974 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). Then the PSU's were grouped into 376 strata, 156 of which consisted of only one PSU which was in sample with certainty. These 156 strata were mostly the larger SMSA's and were called self-representing (SR) since the sample from the sample area represented just that PSU. Each one of the other 220 strata consisted of a group of PSU's and were referred to as non-self-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSU's in the stratum as well.

One PSU was selected from each NSR stratum with probability proportionate to the 1970 census population of the PSU. (This resulted in 220 NSR sample PSU's.) In addition, the NSR strata were grouped into 110 pairs and one stratum was picked at random from each pair. From this stratum, an additional PSU was selected independent 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.

**Designation of sample housing units for the 1975 enumeration.**—The sample housing units designated to be interviewed in the 1975 enumeration consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1974 enumeration.
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 enumeration but which could become eligible in the future) in the 1974 enumeration. (For a list of type A and type B noninterview reasons, see facsimile of 1975 AHS questionnaire, page 1.)
3. All sample housing units that were selected from the list of building permits issued since the 1974 enumeration. (This sample represents the housing units built in permit-issuing areas since the 1974 enumeration.)

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

The sample of 1970 census units was selected in several stages of sampling. The first step was the selection of a sample of census enumeration districts (ED's), administrative units used in the 1970 census. The probability of selection for an ED was proportionate to its 1970 population. The next step was to select an expected cluster of about four neigh-

boring 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. Those 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 for inclusion in the AHS at the sampling rate of 2 in 1,366. Housing units constructed since the 1970 census in areas which do not issue building permits were brought into the sample as a result of the area sampling technique described above.

**Splitting of the sample.**—The sample selection procedure as described above produced 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). One can expect a minimum loss in precision for segments of size-four housing units in rural areas because of the heterogeneity of this type of housing unit. Segments of size-two, however, were considered to be more optimal for sampling within those areas where one could expect neighboring units to be very similar (e.g., urban areas and new construction units). It is felt that if one were to go to segments of size-four housing units in this type of area, a significant loss in precision would result. A splitting operation was then carried out for segments selected from the census address and new construction frames. This consisted of halving each segment of four housing units that was selected for the sample. Two housing units from each segment were to be included in the survey and two housing units were to be held in reserve. No splitting operation was carried out within the segments selected from the area sampling frame; every other area sample segment of four housing units was used for the survey and the remaining segments 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. The other half of the segment (an expected two housing units) for the reserve sample selected in census address and new construction segments was reactivated in 1974 if the segment was rural. Similarly for area segments, the entire reserve segment (an expected four housing units) was reactivated in 1974 if the segment was rural. This supplementation increased the overall probability of selection for sample housing units in rural areas to about 2 in 1,366; but the overall probability of selection for sample housing units in urban areas remained at 1 in 1,366.

## ESTIMATION

The 1975 AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the type A noninterview housing units encountered in the AHS. This noninterview adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

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

The first-stage ratio estimation procedure was employed for sample housing units from non-self-representing (NSR) PSU's only. The 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 PSU's in each census region. The computed first-stage ratio estimate 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 only employed for AHS new construction sample units (i.e., sample units built April 1, 1970, or later). This procedure was designed to adjust the AHS sample estimates of new construction units to independently derived current estimates for nine categories of new construction units for each of the four regions (i.e., six categories for conventional new construction units and three for new construction mobile homes). These independent estimates were considered to be the best estimates available

for the number of new construction units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of new construction units (see the section on nonsampling error).

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

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

The numerators of the ratios were derived from data based on the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. The numerators of the ratios for new construction mobile homes were derived from estimates of mobile home shipments adjusted to account for mobile homes shipped and actually occupied as primary residences.

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 (i.e., the estimates employing the noninterview and first-stage adjustments). The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in each second-stage ratio estimation category.

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

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

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

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

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

The second- and third-stage ratio estimation procedures were repeated in an iterative process in order to bring the AHS estimates into close agreement with both sets of independent estimates (i.e., the independent estimates employed for both the second-stage ratio estimation process as well as those employed for the third-stage ratio estimation process). 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 AHS. Therefore, through the use of the three-stage ratio estimation procedure, one can expect the sample estimate to be improved substantially when the sample housing population is brought into close agreement with a known distribution of the entire housing population with respect to these basic housing characteristics.

## 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 enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling 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 such that the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error be-

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

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

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

Tables I and II present the standard errors applicable to housing unit estimates in this report. Table I shows the approximate standard errors applicable to all housing unit estimates except those pertaining to the specified items in the table.

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

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

Tables III and IV present the standard errors of estimated percentages. Table III shows the approximate standard errors of all estimated percentages of housing units except those pertaining to the specified items in table II.

The standard errors shown in table IV should be used for those specified items. Two-way linear interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables III and IV.

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

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

where: x = the numerator of the ratio
y = the denominator of the ratio
$\sigma_x$ = the standard error of the numerator
$\sigma_y$ = the standard error of the denominator

*Illustrations of the use of the standard error tables. Illustration I.*—Table A-1 of this report shows that in the United

**TABLE I. Standard Errors of Estimated Number of Housing Units: 1975 (Excluding estimates of housing units pertaining to Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, 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) |
|---|---|---|---|---|---|
| 5 | 3 | 3 | 1,000 | 38 | 35 |
| 10 | 4 | 4 | 2,500 | 59 | 49 |
| 25 | 6 | 6 | 5,000 | 82 | 48 |
| 50 | 8 | 8 | 10,000 | 112 | — |
| 100 | 12 | 12 | 25,000 | 156 | — |
| 250 | 19 | 19 | 50,000 | 160 | — |
| 500 | 27 | 26 | 75,000 | 61 | — |

**TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, and Households with Head of Spanish Origin: 1975**

(68 chances out of 100)

| Size of estimate (000) | Standard error: Total, White or Spanish (000) | Standard error: Black (000) | Size of estimate (000) | Standard error: Total, White or Spanish (000) | Standard error: Black (000) |
|---|---|---|---|---|---|
| 5 | 3 | 3 | 1,000 | 44 | 41 |
| 10 | 4 | 4 | 2,500 | 69 | 57 |
| 25 | 7 | 7 | 5,000 | 96 | 56 |
| 50 | 10 | 10 | 10,000 | 131 | — |
| 100 | 14 | 14 | 25,000 | 183 | — |
| 250 | 22 | 22 | 50,000 | 187 | — |
| 500 | 31 | 30 | 75,000 | 71 | — |

**TABLE III. Standard Errors of Estimated Percentages of Housing Units: 1975 (Excludes estimated percentages of housing units pertaining to Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, and Households with Head of Spanish Origin)**

(68 chances out of 100)

| Base of percentage (000) | Estimated percentage[1] | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 or 99 | 2 or 98 | 5 or 95 | 10 or 90 | 15 or 85 | 25 or 75 | 50 |
| 5 | 5.3 | 7.5 | 11.7 | 16.1 | 19.2 | 23.3 | 26.9 |
| 10 | 3.8 | 5.3 | 8.3 | 11.4 | 13.6 | 16.5 | 19.0 |
| 25 | 2.4 | 3.4 | 5.2 | 7.2 | 8.6 | 10.4 | 12.0 |
| 50 | 1.7 | 2.4 | 3.7 | 5.1 | 6.1 | 7.4 | 8.5 |
| 100 | 1.2 | 1.7 | 2.6 | 3.6 | 4.3 | 5.2 | 6.0 |
| 250 | 0.8 | 1.1 | 1.7 | 2.3 | 2.7 | 3.3 | 3.8 |
| 500 | 0.5 | 0.8 | 1.2 | 1.6 | 1.9 | 2.3 | 2.7 |
| 1,000 | 0.4 | 0.5 | 0.8 | 1.1 | 1.3 | 1.6 | 1.9 |
| 2,500 | 0.2 | 0.3 | 0.5 | 0.7 | 0.8 | 1.0 | 1.2 |
| 5,000 | 0.2 | 0.2 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 |
| 10,000 | 0.1 | 0.2 | 0.3 | 0.4 | 0.4 | 0.5 | 0.6 |
| 25,000 | 0.07 | 0.1 | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 |
| 50,000 | 0.05 | 0.08 | 0.1 | 0.2 | 0.2 | 0.2 | 0.3 |
| 75,000 | 0.04 | 0.06 | 0.1 | 0.1 | 0.2 | 0.2 | 0.2 |

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

**TABLE IV. Standard Errors of Estimated Percentages of Housing Units Pertaining to Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, and Households with Head of Spanish Origin: 1975**

(68 chances out of 100)

| Base of percentage (000) | Estimated percentage[1] | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 or 99 | 2 or 98 | 5 or 95 | 10 or 90 | 15 or 85 | 25 or 75 | 50 |
| 5 | 6.3 | 8.8 | 13.7 | 18.9 | 22.5 | 27.3 | 31.5 |
| 10 | 4.4 | 6.2 | 9.7 | 13.4 | 15.9 | 19.3 | 22.3 |
| 25 | 2.8 | 3.9 | 6.1 | 8.4 | 10.1 | 12.2 | 14.1 |
| 50 | 2.0 | 2.8 | 4.3 | 6.0 | 7.1 | 8.6 | 10.0 |
| 100 | 1.4 | 2.0 | 3.1 | 4.2 | 5.0 | 6.1 | 7.0 |
| 250 | 0.9 | 1.2 | 1.9 | 2.7 | 3.2 | 3.9 | 4.5 |
| 500 | 0.6 | 0.9 | 1.4 | 1.9 | 2.2 | 2.7 | 3.1 |
| 1,000 | 0.4 | 0.6 | 1.0 | 1.3 | 1.6 | 1.9 | 2.2 |
| 2,500 | 0.3 | 0.4 | 0.6 | 0.8 | 1.0 | 1.2 | 1.4 |
| 5,000 | 0.2 | 0.3 | 0.4 | 0.6 | 0.7 | 0.9 | 1.0 |
| 10,000 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 |
| 25,000 | 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 |

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

States there were 3,999,000 owner-occupied housing units occupied by recent movers in 1975. Interpolation in table I of this appendix shows the standard error on an estimate of this size is approximately 73,000. Consequently, the 68-percent confidence interval as shown by these data is from 3,926,000 to 4,072,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 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 3,882,000 to 4,116,000 housing units with 90 percent confidence and that the average estimate lies within the interval from 3,853,000 to 4,145,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 3,999,000 1975 owner-occupied housing units containing recent movers in the United States, 307,000, or 7.7 percent, had six persons or more. Interpolation in table III (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage is approximately .5 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 7.2 to 8.2 percent; the 90-percent confidence interval is from 6.9 to 8.5 percent; and the 95-percent confidence interval is from 6.7 to 8.7 percent.

*Illustration II.*—Table A-19 shows that in the United States there were 133,000 owner-occupied housing units occupied by Spanish recent movers. Table II above shows that the standard error of an estimate of this size is approximately 16,000. (Table II should be used since this estimate pertains to households with head of Spanish origin.) Consequently, we could conclude with 68 percent confidence that the average estimate, derived from all possible samples, of 1975 homeowner units occupied by Spanish recent movers lies within the interval from 117,000 to 149,000 housing units. Similarly, the 90-percent confidence interval is from 107,000 to 159,000 housing units; and

the 95-percent confidence interval is from 101,000 to 165,000 housing units.

Table A-19 also shows that of the 133,000 owner-occupied housing units occupied by Spanish recent movers, 46,000, or 34.6 percent, had two bedrooms. Interpolation in table IV (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage is approximately 5.9 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 28.7 to 40.5 percent; the 90-percent confidence interval is from 25.2 to 44.0 percent; and the 95-percent confidence interval is from 22.8 to 46.4 percent.

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

*Illustration of the computation of the standard error of a difference.*—Table A-1 of this report also shows that in the United States there were 428,000 owner-occupied housing units occupied by recent movers with five persons in 1975. Thus, the apparent difference between the number of owner-occupied housing units occupied by recent movers with five persons and the number with six persons or more is 121,000. Table I shows that the standard error on an estimate of 307,000 to be approximately 21,000 and that the standard error on an estimate of 428,000 to be approximately 25,000. Therefore, the standard error of the estimated difference of 120,000 is about

$$33{,}000 = \sqrt{(21{,}000)^2 + (25{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 121,000 difference is from 88,000 to 154,000 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 68,000 to 174,000 housing units and the 95-percent confidence interval is from 55,000 to 187,000. Thus, we can conclude with 95 percent confidence that the number of owner-occupied housing units occupied by recent movers with five persons is greater than the number with six or more persons in the United States in 1975.

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

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

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

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

*Illustration of the computation of the 95-percent confidence interval for a median.*—Table A-1 of this report shows that the median number of persons in owner-occupied housing units occupied by recent movers in the United States was 3.0 in 1975. The base of the distribution, from which this median was determined is 3,999,000.

1. Table III shows that the standard error of 50 percent on a base of 3,999,000 is 1.0 percent.
2. To obtain a two-standard-error confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 48.0 and 52.0.
3. From table A-1, it can be seen by cumulating the frequencies for the first two categories that 1,569,000 owner-occupied housing units occupied by recent movers, or 39.2 percent, had one and two persons (actually, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 841,000 housing units, or 21.0 percent, had three persons (i.e., 2.5 to 3.5 persons). By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$2.5 + (3.5-2.5)\left(\frac{48.0-39.2}{21.0}\right) = 2.9$$

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

$$2.5 + (3.5-2.5)\left(\frac{52.0-39.2}{21.0}\right) = 3.1$$

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

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

and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error.

**Reinterview program.**—For the AHS-National sample a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and thus were the basis for the measurement of the "content" error of these AHS estimates.

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

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

This check was for interviewer evaluation and control.

The results of the 1975 reinterview study are not available, however, it is felt that they will be similar to the results of the 1974 reinterview study, which are presented in the following Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample: 1974." Twenty-five items were included in the 1974 reinterview study, some for the first time. The estimated indexes of inconsistency for these items ranged from 4 to 35 with most items in the 20 to 30 range. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview programs because they had not been previously tested. A 20-50 interval is considered moderate on a range of 0-100 with a high index associated with a high level of response variability.

The 1970 census reinterview results provide illustrations of possible nonsampling errors for some of the items which also appear in the AHS. For example, median value of homes was consistently underestimated by about five percent, and the average monthly costs of electricity and utility gas were consistently overestimated although the effect on the average gross rent figures 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, the results of the reinterview studies are derived from sample surveys so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

**Coverage Errors.**—With respect to errors of coverage and estimation for missing data, it was mentioned previously in the section on estimation that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction in permit-issuing areas and new construction mobile homes. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and therefore would have been represented in the sample selected from 1970 census units. However, it has been estimated that the 1975 AHS sample misses about six percent (i.e., about 600,000 units) of all conventional new construction (i.e., all conventional housing units built after April 1970 in both permit-issuing and non-permit-issuing areas) because the permits for these missed units, which were built after April 1970, were issued before 1970.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new construction mobile homes. However, new mobile homes in segments where area sampling methods were used do come into the AHS sample. In addition, new mobile homes in segments sampled from the 1970 census list also come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. However, new mobile homes in these segments that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. It has been estimated that the 1975 AHS sample misses at least 200,000 new mobile homes. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS sample still exists. This is especially true for new mobile homes since it is believed that the corresponding independent estimate substantially understates the actual new mobile home inventory.

**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 the survey.

☆U.S. GOVERNMENT PRINTING OFFICE: 1977-240-869/1051

# Appendix B

## Source and Reliability of the Estimates


SAMPLE DESIGN . . . . . App-43
Selection of sample areas. . . . . App-43
Designation of sample housing units for the 1975 enumeration . App-43
Selection of the 1973 sample housing units. . . . . App-43
Splitting of the sample . . . . . App-44
Selection of supplemental sample housing units in rural areas. . . . . App-44
1970 Census of Population and Housing. . . . . App-44
ESTIMATION. . . . . App-44
AHS-National sample. . . . . App-44
1975 housing inventory . . . . . App-44
1973-1975 lost units . . . . . App-45
Ratio estimation procedure of the 1970 Census of Population and Housing . . . . . App-46
RELIABILITY OF THE ESTIMATES. App-46
Sampling errors . . . . . App-46
Illustrations of the use of the standard error tables. Illustration I . . . . . App-48
Illustration II . . . . . App-49
Differences. . . . . App-50
Illustration of the computation of the standard error of a difference. . . . . App-50
Medians . . . . . App-51
Illustration of the computation of the 95-percent confidence interval for a median . . . . . App-51
Nonsampling errors. . . . . App-51
1970 census . . . . . App-51
Reinterview program . . . . . App-52
Coverage errors . . . . . App-52
Rounding errors in AHS . . . . . App-52


### SAMPLE DESIGN

The 1975 estimates are based on data collected in October through December 1975 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 72,600 sample housing units (both occupied and vacant) were eligible for interview in the 1975 Annual Housing Survey. Of this number, 3,700 interviews were not obtained because, for occupied housing units, the occupants were not found at home after repeated calls or were unavailable for some other reason; or, for vacant housing units, no informed respondent could be found after repeated visits. In addition to the 72,600, there were also 7,300 sample units which were visited but found not to provide information relevant to the 1974 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). Then the PSU's were grouped into 376 strata, 156 of which consisted of only one PSU which was in sample with certainty. These 156 strata were mostly the larger SMSA's and were called self-representing (SR) since the sample from the sample area represented just that PSU. Each one of the other 220 strata consisted of a group of PSU's and were referred to as non-self-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSU's in the stratum as well.

One PSU was selected from each NSR stratum with probability proportionate to the 1970 census population of the PSU. (This resulted in 220 NSR sample PSU's.) In addition, the NSR strata were grouped into 110 pairs and one stratum was picked at random from each pair. From this stratum, an additional PSU was selected independent 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, and giving a grand total of 461 PSU's.

**Designation of sample housing units for the 1975 enumeration.**—The sample housing units designated to be interviewed in the 1975 enumeration consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1974 enumeration.
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 enumeration but which could become eligible in the future) in the 1974 enumeration. (For a list of type A and type B noninterview reasons, see facsimile of 1975 AHS questionnaire, page 1.)
3. All sample housing units that were selected from the list of building permits issued since the 1974 enumeration. (This sample represents the housing units built in permit-issuing areas, since the 1974 enumeration.)

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

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

first step was the selection of a sample of census enumeration districts (ED's), administrative units used in the 1970 census. The probability of selection for an ED was proportionate to its 1970 population. 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. Those 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 for inclusion in the AHS at the sampling rate of 2 in 1,366. Housing units constructed since the 1970 census in areas which do not issue building permits were brought into the sample as a result of the area sample described above.

**Splitting of the sample.**—The sample selection procedure as described above produced 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). One can expect a minimum loss in precision for segments of size-four housing units in rural areas because of the heterogeneity of this type of housing unit. Segments of size-two, however, were considered to be more optimal for sampling within those areas where one could expect neighboring units to be very similar (e.g., urban areas and new construction units). It is felt that if one were to go to segments of size-four housing units in this type of area, a significant loss in precision would result. A splitting operation was then carried out for segments selected from the census address and new construction frames. This consisted of halving each segment of four housing units that was selected for the sample. Two housing units from each segment were to be included in the survey and two housing units were to be held in reserve. No splitting operation was carried out within the segments selected from the area sampling frame; every other area sample segment of four housing units was used for the survey and the remaining segments 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 segments, this meant that the other half of the segment (an expected two housing units) was reactivated in 1974 if the segment was rural. Similarly for area segments, this meant the entire reserve segment (an expected four housing units) was reactivated in 1974 if the segment 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.

**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-percent, 15-percent, 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 1975 housing inventory and estimates of units removed from the housing inventory between 1973 and 1975 (i.e., 1973-1975 lost units). Each type of estimate employed a separate, though similar, estimation procedure as described below.

**1975 housing inventory.**—The 1975 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 AHS. This noninterview adjustment was done separately for different categories of 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. The procedure was designed to reduce the contribution to the variance arising from the sampling of NSR 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 calcuated 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 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 only employed for AHS new construction sample units (i.e., sample units built April 1, 1970, or later). This procedure was designed to adjust the AHS sample estimates of new construction units to independently derived current estimates for nine categories of new construction units for each of the four regions (i.e., six categories for conventional new construction units and three for new construction mobile homes). These independent estimates were considered to be the best estimates available for the number of new construction units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of new construction units (see the section on nonsampling error).

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

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

The numerators of the ratios were derived from data based on the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. The numerators of the ratios for new construction mobile homes were derived from estimates of mobile home shipments adjusted to account for mobile homes shipped and actually occupied as primary residences.

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 (i.e., the estimates employing the noninterview and first-stage adjustments). The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in each second-stage ratio estimation category.

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

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

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

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

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

The second- and third-stage ratio estimation procedures were repeated in an iterative process in order to bring the AHS estimates into close agreement with both sets of independent estimates (i.e., the independent estimates employed for both the second-stage ratio estimation process as well as those employed for the third-stage ratio estimation process). 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 AHS. Therefore, through the use of the three-stage ratio estimation procedure one can expect the sample estimate to be improved substantially when the sample housing population is brought into close agreement with a known distribution of the entire housing population with respect to these basic housing characteristics.

**1973-1975 lost units.**—The 1973-1975 lost unit estimates employed the three-stage ratio estimation procedure used to produce the AHS-National estimates of the 1973 housing inventory, as was described in the 1973 Current Housing Report, series H-150-73A, **General Housing Characteristics for the United States and Regions.** Since the 1973-1975 lost units existed by definition in the 1973 housing inventory, there was a 1973 housing inventory weight associated with each 1973-1975 lost unit. This weight was used to tabulate the estimates of the characteristics of the 1973-1975 lost units. Also, the general effect of this estimation procedure was to reduce the sampling error for most statistics below what would have been obtained by simply

weighting the results of the sample by the inverse of the probability of selection.

**Ratio Estimation Procedure of the 1970 Census of Population and Housing.**—This report presents data on the housing characteristics of the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be obtained in the 1970 census report, HC(1)-B1, **Detailed Housing Characteristics, United States Summary.**

## RELIABILITY OF THE ESTIMATES

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

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

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling 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 such that the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

TABLE I. **Standard Errors of Estimated Number of Total Housing Units: 1975 (Excluding estimates of housing units pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Source of Water, and Households with Head of Spanish Origin)**

(68 chances out of 100)

| Size of estimate (000) | Standard error: Total or White (000) | Standard error: Black (000) | Size of estimate (000) | Standard error: Total or White (000) | Standard error: Black (000) |
|---|---|---|---|---|---|
| 5 | 3 | 3 | 1,000 | 38 | 35 |
| 10 | 4 | 4 | 2,500 | 59 | 49 |
| 25 | 6 | 6 | 5,000 | 82 | 48 |
| 50 | 8 | 8 | 10,000 | 112 | – |
| 100 | 12 | 12 | 25,000 | 156 | – |
| 250 | 19 | 19 | 50,000 | 160 | – |
| 500 | 27 | 26 | 75,000 | 61 | – |

TABLE II. **Standard Errors of Estimated Numbers of Total Housing Units Pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Source of Water, and Households with Head of Spanish Origin: 1975**

(68 chances out of 100)

| Size of estimate (000) | Standard error: Total, White, or Spanish origin (000) | Standard error: Black (000) | Size of estimate (000) | Standard error: Total, White, or Spanish origin (000) | Standard error: Black (000) |
|---|---|---|---|---|---|
| 5 | 3 | 3 | 1,000 | 44 | 41 |
| 10 | 4 | 4 | 2,500 | 69 | 57 |
| 25 | 7 | 7 | 5,000 | 96 | 56 |
| 50 | 10 | 10 | 10,000 | 131 | – |
| 100 | 14 | 14 | 25,000 | 183 | – |
| 250 | 22 | 22 | 50,000 | 187 | – |
| 500 | 31 | 30 | 75,000 | 71 | – |

TABLE III. **Standard Errors of Estimated Number of Urban Housing Units: 1975**

(68 chances out of 100)

| Size of estimate | Standard error | | | | Size of estimate | Standard error | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Urban housing units (except those in the next column) | | Mobile homes, lacking plumbing, cooking fuel, or sewage disposal | | | Urban housing units (except those in the next column) | | Mobile homes, lacking plumbing, cooking fuel, or sewage disposal | |
| (000) | Total, White, or Spanish origin (000) | Black (000) | Total, White, or Spanish origin (000) | Black (000) | (000) | Total, White, or Spanish origin (000) | Black (000) | Total, White, or Spanish origin (000) | Black (000) |
| 5 . . . . . | 3 | 3 | 3 | 3 | 1,000 . . . | 40 | 37 | 46 | 43 |
| 10 . . . . . | 4 | 4 | 5 | 5 | 2,500 . . . | 62 | 51 | 72 | 59 |
| 25 . . . . . | 6 | 6 | 7 | 7 | 5,000 . . . | 86 | 50 | 101 | 58 |
| 50 . . . . . | 9 | 9 | 10 | 10 | 10,000. . . | 116 | — | 136 | — |
| 100. . . . . | 13 | 13 | 15 | 15 | 25,000. . . | 157 | — | 183 | — |
| 250. . . . . | 20 | 20 | 23 | 23 | 50,000. . . | 135 | — | 158 | — |
| 500. . . . . | 28 | 27 | 33 | 32 | | | | | |

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

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

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

Tables I through V present the standard errors applicable to the 1975 housing inventory estimates in this report, and table VI presents the standard errors applicable to 1973-1975 lost housing unit estimates in this report. Table I shows the approximate standard errors applicable to all 1975 housing inventory estimates except those pertaining to the specified items in the table, those pertaining to urban or rural items, and those pertaining to estimates of vacant housing units.

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

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 nu-

TABLE IV. **Standard Errors of Estimated Number of Rural Housing Units: 1975**

(68 chances out of 100)

| Size of estimate | Standard error | |
|---|---|---|
| | Rural housing units (except those in the next column) | Source of water, mobile homes, sewage disposal, or Spanish origin |
| (000) | (000) | (000) |
| 5 . . . . . | 2 | 3 |
| 10 . . . . . | 3 | 4 |
| 25 . . . . . | 5 | 6 |
| 50 . . . . . | 8 | 9 |
| 100. . . . . | 11 | 13 |
| 250. . . . . | 17 | 20 |
| 500. . . . . | 24 | 28 |
| 1,000 . . . . | 34 | 40 |
| 2,500 . . . | 54 | 63 |
| 5,000 . . . | 75 | 88 |
| 10,000 . . . | 104 | 121 |
| 20,000 . . | 139 | 162 |

TABLE V. **Standard Errors of Estimated Number of Vacant Housing Units: 1975**

(68 chances out of 100)

| Size of estimate (000) | Standard error | | |
|---|---|---|---|
| | Total (000) | Urban (000) | Rural (000) |
| 5 | 3 | 3 | 2 |
| 10 | 4 | 4 | 3 |
| 25 | 6 | 7 | 5 |
| 50 | 9 | 9 | 8 |
| 100 | 13 | 13 | 11 |
| 250 | 20 | 21 | 18 |
| 500 | 29 | 30 | 27 |
| 1,000 | 42 | 43 | 41 |
| 2,500 | 71 | 73 | — |
| 5,000 | 109 | — | — |

TABLE VI. **Standard Errors of Estimated Number of Lost Housing Units: 1973-1975**

(68 chances out of 100)

| Size of estimate (000) | Standard error | | |
|---|---|---|---|
| | Total (000) | Urban (000) | Rural (000) |
| 5 | 3 | 3 | 3 |
| 10 | 4 | 5 | 4 |
| 25 | 7 | 7 | 6 |
| 50 | 10 | 10 | 9 |
| 100 | 14 | 15 | 13 |
| 250 | 22 | 24 | 20 |
| 500 | 32 | 34 | 29 |
| 1,000 | 47 | — | — |

merators of the percentages, particularly if the percentages are 50 percent or more.

Tables VII through XI present the standard errors of estimated percentages. Table VII shows the approximate standard errors of all estimated percentages of the 1975 housing inventory except those pertaining to the specified items in table VIII. In addition to those specified items, table VIII also shows the approximate standard errors of urban, vacant housing units and of total 1973-1975 lost housing units. Table IX shows the approximate standard errors of all estimated percentages of the 1975 rural housing inventory except those pertaining to the specified items in table X. Table X also shows the approximate standard errors of vacant housing units and of the 1975 urban housing inventory except those pertaining to the specified items in table XI. Table XI also shows the approximate standard errors of urban 1973-1975 lost housing units. Two-way linear interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables VII through XI.

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

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

where: $x$ = the numerator of the ratio

$y$ = the denominator of the ratio

$\sigma_x$ = the standard error of the numerator

$\sigma_y$ = the standard error of the denominator

*Illustrations of the use of the standard error tables. Illustration I.*—Table A-1 of this report shows that in urban areas of the United States there were 31,107,000 owner-occupied housing units in 1975. Interpolation in column 1 of table III of this appendix shows that the standard error of an estimate of this size is approximately 151,000. Consequently, the 68-percent confidence interval, as shown by these data, is from 30,956,000 to

TABLE VII. **Standard Errors of Estimated Percentages of Housing Units: 1975 (Excluding estimated percentages of housing units pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Water Supply, Households with Head of Spanish Origin, 1973-1975 Lost Housing Units, and Vacants)**

(68 chances out of 100)

| Base of percentage (000) | Estimated percentage[1] | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 or 99 | 2 or 98 | 5 or 95 | 10 or 90 | 15 or 85 | 25 or 75 | 50 |
| 5 | 5.3 | 7.5 | 11.7 | 16.1 | 19.2 | 23.3 | 26.9 |
| 10 | 3.8 | 5.3 | 8.3 | 11.4 | 13.6 | 16.5 | 19.0 |
| 25 | 2.4 | 3.4 | 5.2 | 7.2 | 8.6 | 10.4 | 12.0 |
| 50 | 1.7 | 2.4 | 3.7 | 5.1 | 6.1 | 7.4 | 8.5 |
| 100 | 1.2 | 1.7 | 2.6 | 3.6 | 4.3 | 5.2 | 6.0 |
| 250 | 0.8 | 1.1 | 1.7 | 2.3 | 2.7 | 3.3 | 3.8 |
| 500 | 0.5 | 0.8 | 1.2 | 1.6 | 1.9 | 2.3 | 2.7 |
| 1,000 | 0.4 | 0.5 | 0.8 | 1.1 | 1.3 | 1.6 | 1.9 |
| 2,500 | 0.2 | 0.3 | 0.5 | 0.7 | 0.8 | 1.0 | 1.2 |
| 5,000 | 0.2 | 0.2 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 |
| 10,000 | 0.1 | 0.2 | 0.3 | 0.4 | 0.4 | 0.5 | 0.6 |
| 25,000 | 0.07 | 0.1 | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 |
| 50,000 | 0.05 | 0.08 | 0.1 | 0.2 | 0.2 | 0.2 | 0.3 |
| 75,000 | 0.04 | 0.06 | 0.1 | 0.1 | 0.2 | 0.2 | 0.2 |

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

31,258,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 owner-occupied housing units in urban areas lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples, lies within the interval from 30,865,000 to 31,349,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 30,805,000 to 31,409,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 31,107,000 owner-occupied housing units in urban areas, 9,691,000, or 31.1 percent, were occupied by two persons. Interpolation in table X (i.e., interpolation on both the base and the percent) of this appendix shows that the standard error of the above percentage is approximately .3 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 30.8 to 31.4 percent, the 90-percent confidence interval is from 30.6 to 31.6 percent, and the 95-percent confidence interval is from 30.5 to 31.7 percent.

*Illustration II.*—Table A-2 of this report shows that in the rural areas of the United States there were 127,000 specified owner-occupied, farm housing units in 1975. Interpolation in column 1 of table IV of this appendix shows that the standard error of an estimate of this size is approximately 12,000. Consequently, the 68-percent confidence interval, as shown by these data, is from 115,000 to 139,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 specified owner-occupied, farm housing units lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples, lies within the interval from 108,000 to 146,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 103,000 to

TABLE VIII. **Standard Errors of Estimated Percentages of Housing Units Pertaining to Cooking Fuel, Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Water Supply, Households with Head of Spanish Origin, 1973-1975 Lost Housing Units, and Vacant Urban Housing Units: 1975**

(68 chances out of 100)

| Base of percentage (000) | Estimated percentage[1] | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 or 99 | 2 or 98 | 5 or 95 | 10 or 90 | 15 or 85 | 25 or 75 | 50 |
| 5 | 6.3 | 8.8 | 13.7 | 18.9 | 22.5 | 27.3 | 31.5 |
| 10 | 4.4 | 6.2 | 9.7 | 13.4 | 15.9 | 19.3 | 22.3 |
| 25 | 2.8 | 3.9 | 6.1 | 8.4 | 10.1 | 12.2 | 14.1 |
| 50 | 2.0 | 2.8 | 4.3 | 6.0 | 7.1 | 8.6 | 10.0 |
| 100 | 1.4 | 2.0 | 3.1 | 4.2 | 5.0 | 6.1 | 7.0 |
| 250 | 0.9 | 1.2 | 1.9 | 2.7 | 3.2 | 3.9 | 4.5 |
| 500 | 0.6 | 0.9 | 1.4 | 1.9 | 2.2 | 2.7 | 3.1 |
| 1,000 | 0.4 | 0.6 | 1.0 | 1.3 | 1.6 | 1.9 | 2.2 |
| 2,500 | 0.3 | 0.4 | 0.6 | 0.8 | 1.0 | 1.2 | 1.4 |
| 5,000 | 0.2 | 0.3 | 0.4 | 0.6 | 0.7 | 0.9 | 1.0 |
| 10,000 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 |
| 25,000 | 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 |

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

TABLE IX. **Standard Errors of Estimated Percentages of Rural Housing Units: 1975 (Excluding estimated percentages of rural housing units pertaining to Mobile Homes, Source of Water, Sewage Disposal, Households with Head of Spanish Origin, 1973-1975 Lost Units, and Vacant Rural Housing Units)**

(68 chances out of 100)

| Base of percentage (000) | Estimated percentage[1] | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 or 99 | 2 or 98 | 5 or 95 | 10 or 90 | 15 or 85 | 25 or 75 | 50 |
| 5 | 4.9 | 7.9 | 10.7 | 14.7 | 17.5 | 21.3 | 24.5 |
| 10 | 3.5 | 4.9 | 7.6 | 10.4 | 12.4 | 15.0 | 17.4 |
| 25 | 2.2 | 3.1 | 4.8 | 6.6 | 7.8 | 9.5 | 11.0 |
| 50 | 1.5 | 2.2 | 3.4 | 4.7 | 5.5 | 6.7 | 7.8 |
| 100 | 1.1 | 1.5 | 2.4 | 3.3 | 3.9 | 4.8 | 5.5 |
| 250 | 0.7 | 1.0 | 1.5 | 2.1 | 2.5 | 3.0 | 3.5 |
| 500 | 0.5 | 0.7 | 1.1 | 1.5 | 1.8 | 2.1 | 2.5 |
| 1,000 | 0.3 | 0.5 | 0.8 | 1.0 | 1.2 | 1.5 | 1.7 |
| 2,500 | 0.2 | 0.3 | 0.5 | 0.7 | 0.8 | 1.0 | 1.1 |
| 5,000 | 0.2 | 0.2 | 0.3 | 0.5 | 0.6 | 0.7 | 0.8 |
| 10,000 | 0.1 | 0.2 | 0.2 | 0.3 | 0.4 | 0.5 | 0.5 |
| 20,000 | 0.08 | 0.1 | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 |

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

TABLE X. **Standard Errors of Estimated Percentages of Rural Housing Units Pertaining to Mobile Homes, Source of Water, Sewage Disposal, Households with Head of Spanish Origin, and 1973-1975 Lost Units; of Total Vacant Housing Units; and of Urban Housing Units: 1975 (Excluding estimated percentages of urban housing units pertaining to Mobile Homes, Lacking Complete Plumbing Facilities, Cooking Fuel, Sewage Disposal, and 1973-1975 Lost Housing Units)**

(68 chances out of 100)

| Base of percentage (000) | Estimated percentage[1] | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 or 99 | 2 or 98 | 5 or 95 | 10 or 90 | 15 or 85 | 25 or 75 | 50 |
| 5 | 5.7 | 8.0 | 12.4 | 17.1 | 20.4 | 24.7 | 28.5 |
| 10 | 4.0 | 5.6 | 8.8 | 12.1 | 14.4 | 17.5 | 20.2 |
| 25 | 2.5 | 3.6 | 5.6 | 7.6 | 9.1 | 11.0 | 12.7 |
| 50 | 1.8 | 2.5 | 3.9 | 5.4 | 6.4 | 7.8 | 9.0 |
| 100 | 1.3 | 1.8 | 2.8 | 3.8 | 4.6 | 5.5 | 6.4 |
| 250 | 0.8 | 1.1 | 1.8 | 2.4 | 2.9 | 3.5 | 4.0 |
| 500 | 0.6 | 0.8 | 1.2 | 1.7 | 2.0 | 2.5 | 2.9 |
| 1,000 | 0.4 | 0.6 | 0.9 | 1.2 | 1.4 | 1.7 | 2.0 |
| 2,500 | 0.3 | 0.4 | 0.6 | 0.8 | 0.9 | 1.1 | 1.3 |
| 5,000 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.8 | 0.9 |
| 10,000 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.6 |
| 25,000 | 0.08 | 0.1 | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 |
| 50,000 | 0.06 | 0.08 | 0.1 | 0.2 | 0.2 | 0.2 | 0.3 |

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

TABLE XI. **Standard Errors of Estimated Percentages of Urban Housing Units Pertaining to Cooking Fuel, Mobile Homes, Sewage Disposal, Lacking Complete Plumbing Facilities, and 1973-1975 Lost Housing Units: 1975**

(68 chances out of 100)

| Base of percentage (000) | Estimated percentage[1] | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 or 99 | 2 or 98 | 5 or 95 | 10 or 90 | 15 or 85 | 25 or 75 | 50 |
| 5 | 6.6 | 9.3 | 14.4 | 19.9 | 23.7 | 28.7 | 33.1 |
| 10 | 4.7 | 6.6 | 10.2 | 14.1 | 16.7 | 20.3 | 23.4 |
| 25 | 2.9 | 4.1 | 6.5 | 8.9 | 10.6 | 12.9 | 14.8 |
| 50 | 2.1 | 2.9 | 4.6 | 6.3 | 7.5 | 9.1 | 10.4 |
| 100 | 1.5 | 2.1 | 3.2 | 4.4 | 5.3 | 6.4 | 7.4 |
| 250 | 0.9 | 1.3 | 2.0 | 2.8 | 3.3 | 4.1 | 4.7 |
| 500 | 0.7 | 0.9 | 1.4 | 2.0 | 2.4 | 2.9 | 3.3 |
| 1,000 | 0.5 | 0.7 | 1.0 | 1.4 | 1.7 | 2.0 | 2.3 |
| 2,500 | 0.3 | 0.4 | 0.6 | 0.9 | 1.1 | 1.3 | 1.5 |
| 5,000 | 0.2 | 0.3 | 0.5 | 0.6 | 0.7 | 0.9 | 1.0 |
| 10,000 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 |
| 25,000 | 0.09 | 0.1 | 0.2 | 0.3 | 0.3 | 0.4 | 0.5 |
| 50,000 | 0.07 | 0.09 | 0.1 | 0.2 | 0.2 | 0.3 | 0.3 |

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

151,000 housing units with 95 percent confidence.

Table A-2 also shows that of the 127,000 specified owner-occupied, farm housing units in rural areas, 64,000, or 50.4 percent, were owned free and clear. Interpolation in table IX (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage is approximately 5.1 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 45.3 to 55.5 percent; the 90-percent confidence interval is from 42.2 to 58.6 percent; and the 95-percent confidence interval is from 40.2 to 60.6 percent.

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

*Illustration of the computation of the standard error of a difference.*—Table A-1 of this report shows that in urban areas of the United States there were 5,537,000 owner-occupied housing units with three persons in 1975. Thus, the apparent difference between the number of 1975 owner-occupied housing units in urban areas with two persons and those with three persons is 4,154,000. Column 1 of table III shows that the standard error of 9,691,000 is approximately 114,000 and that the standard error of 5,537,000 is approximately 89,000. Therefore, the standard error

of the estimated difference of 4,154,000 is about

$$145{,}000 = \sqrt{(114{,}000)^2 + (89{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 4,154,000 difference is from 4,009,000 to 4,299,000 housing units. Therefore, a conclusion that the average estimate of this difference, derived from all possible samples, lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 3,922,000 to 4,386,000; and the 95-percent confidence interval is from 3,864,000 to 4,444,000. Thus, we can conclude with 95 percent confidence that the number of 1975 owner-occupied housing units inside SMSA's with two persons is greater than the number with three persons since the 95-percent confidence interval of this difference does not include zero or negative values.

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

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

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

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

*Illustration of the computation of the 95-percent confidence interval for a median.*—Table A-1 of this report shows the median number of persons in owner-occupied housing units in urban areas was 2.8 in 1975. The base of the distribution from which this median was determined is 31,107,000.

1. From table X, the standard error of a 50-percent characteristic on the base of 31,107,000 is .4 percentage points.
2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 49.2 and 50.8.
3. From table A-1, it can be seen by cumulating the frequencies for the first two categories that 13,885,000 housing units, or 44.5 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 5,537,000 owner-occupied housing units, or 17.8 percent, had three persons (i.e., 2.5 to 3.5 persons). By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$2.5 + (3.5 - 2.5)\left(\frac{49.2 - 44.5}{17.8}\right) = 2.8$$

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

$$2.5 + (3.5 - 2.5)\left(\frac{50.8 - 44.5}{17.8}\right) = 2.9$$

Thus, the 95-percent confidence interval ranges from 2.8 to 2.9 persons. Although it appears that this confidence interval has the sample estimate as the lower limit, it actually is a reflection of the rounding error associated with this median (see the paragraph on rounding error 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. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

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

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

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

**Reinterview program.**—For the AHS-National sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and thus were the basis for the measurement of the "content" error of these AHS estimates.

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

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

This check was for interviewer evaluation and control.

The results of the 1975 reinterview study are not available; however, it is felt that they will be similar to the results of the 1974 reinterview study which are presented in the following Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample: 1974."

Twenty-five items were included in the 1974 reinterview study, some for the first time. The estimated indexes of inconsistency for these items ranged from 4 to 35 with most items in the 20 to 30 range. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview programs because they had not been previously tested. A 20-50 interval is considered moderate on a range of 0-100 with a high index associated with a high level of response variability.

The 1970 census reinterview results provide illustrations of possible nonsampling errors for some of the items which also appear in the AHS. For example, median value of homes was consistently underestimated by about five percent, and the average monthly costs of electricity and utility gas were consistently overestimated although the effect on the 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, the results of the reinterview studies are derived from sample surveys, so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

**Coverage Errors.**—With respect to errors of coverage and estimation for missing data, it was mentioned previously, in the section on estimation that the AHS new construction sample had deficiencies in representing both conventional new construction in permit-issuing areas and new construction mobile homes. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and therefore would have been represented in the sample selected from 1970 census units. However, it has been estimated that the 1975 AHS sample misses about six percent (i.e., about 600,000 units) of *all* conventional new construction (i.e., all conventional housing units built after April 1970 in both permit-issuing and non-permit-issuing areas) because the permits for these missed units, which were built after April 1970, were issued before 1970.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new construction mobile homes. However, new mobile homes in segments where area sampling methods were used do come into the AHS sample. In addition, new mobile homes in segments sampled from the 1970 census list also come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. However, new mobile homes in these segments that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970 census have no chance of coming into the AHS sample. It has been estimated that the 1975 AHS sample misses at least 200,000 new mobile homes. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies although some bias in the AHS sample still exists. This is especially true for new mobile homes, since it is believed that the corresponding, independent estimate substantially understates the actual new mobile home inventory.

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

# Appendix B

## Source and Reliability of the Estimates


SAMPLE DESIGN . . . . . App-43
Selection of sample areas. . . . . App-43
Designation of sample housing units for the 1975 enumeration . App-43
Selection of the 1973 sample housing units. . . . . App-43
Splitting of the sample . . . . . App-44
Selection of supplemental sample housing units in rural areas . . . . App-44
ESTIMATION. . . . . App-44
RELIABILITY OF THE ESTIMATES. App-45
Sampling errors . . . . . App-45
Illustrations of the use of the standard error tables
Illustration I . . . . . App-46
Illustration II . . . . . App-47
Differences. . . . . App-48
Illustration of the computation of the standard error of a difference. . . . . App-48
Medians . . . . . App-48
Illustration of the computation of the 95-percent confidence interval of the median. . . . . App-48
Nonsampling errors. . . . . App-49
Reinterview program . . . . . App-49
Coverage errors . . . . . App-49
Rounding errors in AHS . . . . . App-50


### SAMPLE DESIGN

The 1975 estimates are based on data collected in October through December 1975 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 72,600 sample housing units (both occupied and vacant) were eligible for interview in the 1975 Annual Housing Survey. Of this number, 3,700 interviews were not obtained because, for occupied housing units, the occupants were not found at home after repeated calls or were unavailable for some other reason; or, for vacant housing units, no informed respondent could be found after repeated visits. In addition to the 72,600, there were also 7,300 sample units which were visited but found not to provide information relevant to the 1975 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). Then the PSU's were grouped into 376 strata, 156 of which consisted of only one PSU which was in sample with certainty. These 156 strata were mostly the larger SMSA's and were called self-representing (SR) since the sample from the sample area represented just that PSU. Each one of the other 220 strata consisted of a group of PSU's and were referred to as non-self-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSU's in the stratum as well.

One PSU was selected from each NSR stratum with probability proportionate to the 1970 census population of the PSU. (This resulted in 220 NSR sample PSU's.) In addition, the NSR strata were grouped into 110 pairs and one stratum was picked at random from each pair. From this stratum, an additional PSU was selected independent 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 1975 enumeration.**—The sample housing units designated to be interviewed in the 1975 enumeration consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1974 enumeration.
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 enumeration but which could become eligible in the future) in the 1974 enumeration. (For a list of type A and type B noninterview reasons, see facsimile of the 1975 AHS questionnaire, page 1.)
3. All sample housing units that were selected from the list of building permits issued since the 1974 enumeration. (This sample represents the housing units built in permit-issuing areas, since the 1974 enumeration.)

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

The sample of 1970 census units was selected in several stages of sampling. The first step was the selection of a sample of census enumeration districts (ED's), administrative units used in the 1970 census. The probability of selection for an ED was proportionate to its 1970 population. 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. Those 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 for inclusion in the AHS at the sampling rate of 2 in 1,366. As a result of the area sample described above, housing units constructed since the 1970 census in areas which do not issue building permits were brought into the sample.

**Splitting of the sample.**—The sample selection procedure as described above produced 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). One can expect a minimum loss in precision for segments of size-four housing units in rural areas because of the heterogeneity of this type of housing unit. Segments of size-two, however, were considered to be more optimal for sampling within those areas where one could expect neighboring units to be very similar (e.g., urban areas and new construction units). It is felt that if one were to go to segments of size-four housing units in this type of area, a significant loss in precision would result. A splitting operation was then carried out for segments selected from the census address and new construction frames. This consisted of halving each segment of four housing units that was selected for the sample. Two housing units from each segment were to be included in the survey and two housing units were to be held in reserve. No splitting operation was carried out within the segments selected from the area sampling frame; every other area sample segment of four housing units was used for the survey and the remaining segments 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 segments, this meant that the other half of the segment (an expected two housing units) was reactivated in 1974 if the segment was rural. Similarly for area segments, this meant the entire reserve segment (an expected four housing units) was reactivated in 1974 if the segment was rural. This supplementation increased the overall probability of selection for sample housing units in rural areas to about 2 in 1,366; while the overall probability of selection for sample housing units in urban areas remained at 1 in 1,366.

## ESTIMATION

The 1975 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 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. The 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 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 only employed for AHS new construction sample units (i.e., sample units built April 1, 1970, or later). This procedure was designed to adjust the AHS sample estimates for new construction units to independently derived current estimates for nine categories of new construction units for each of the four regions (i.e., six categories for conventional new construction units and three for new construction mobile homes). These independent estimates were considered to be the best estimates available for the number of new construction units. This adjustment was necessary to correct known deficiencies in the AHS sample with regard to representation of new construction units (see the section on nonsampling error).

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

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

The numerators of the ratios were derived from data based on the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. The numerators of the ratios for new construction mobile homes were derived from estimates of mobile home shipments adjusted to account for mobile homes shipped and actually occupied as primary residences.

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 (i.e., the estimates employing the noninterview and first-stage adjustments).

The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in each second-stage ratio estimation category.

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

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

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

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

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

The second- and third-stage ratio estimation procedures were repeated in an iterative process in order to bring the AHS estimates into close agreement with both sets of independent estimates (i.e., the independent estimates employed for both the second-stage ratio estimation process as well as those employed for the third-stage ratio estimation process). 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 AHS. Therefore, through the use of the three-stage ratio estimation procedure one can expect the sample estimate to be improved substantially when the sample housing population is brought into close agreement with a known distribution of the entire housing population with respect to these basic housing characteristics.

## 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 enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling 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 conclusion that the interval includes the average result of all possible samples would be correct with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

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

low the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples; 3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

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

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

Tables I and II present the standard errors applicable to housing unit estimates in this report. Table I shows the approximate standard errors applicable to all housing unit estimates except those pertaining to the specified items in the table.

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

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

Tables III and IV present the standard errors of estimated percentages. Table III shows the approximate standard errors of all estimated percentages of housing units except those pertaining to the specified items in table II.

The standard errors shown in table IV should be used for those specified items. Two-way linear interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables III and IV.

Tables III and IV are not appropriate for percentage ratios where the numerator x is not a subset of the denominator y; i.e., tables III and IV underestimate the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be obtained by letting the standard error of the ratio be approximately equal to:

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

where: x = the numerator of the ratio
y = the denominator of the ratio
$\sigma_x$ = the standard error of the numerator
$\sigma_y$ = the standard error of the denominator

*Illustrations of the use of the standard error tables. Illustration I.*—Table A-1 of this report shows that in the United States there were 21,294,000 owner-occupied housing units in 1975 with garbage collection service once a week. Interpola-

**TABLE I. Standard Errors of Estimated Number of Housing Units: 1975 (Excluding estimates of housing units pertaining to Lacking Complete Plumbing Facilities, Sewage Disposal, 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) |
|---|---|---|---|---|---|
| 5 | 3 | 3 | 1,000 | 38 | 35 |
| 10 | 4 | 4 | 2,500 | 59 | 49 |
| 25 | 6 | 6 | 5,000 | 82 | 48 |
| 50 | 8 | 8 | 10,000 | 112 | – |
| 100 | 12 | 12 | 25,000 | 156 | – |
| 250 | 19 | 19 | 50,000 | 160 | – |
| 500 | 27 | 26 | 75,000 | 61 | – |

**TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to Lacking Complete Plumbing Facilities, Sewage Disposal, and Households with Head of Spanish Origin: 1975**

(68 chances out of 100)

| Size of estimate (000) | Standard error: Total, White, or Spanish origin (000) | Standard error: Black (000) | Size of estimate (000) | Standard error: Total, White, or Spanish origin (000) | Standard error: Black (000) |
|---|---|---|---|---|---|
| 5 | 3 | 3 | 1,000 | 44 | 41 |
| 10 | 4 | 4 | 2,500 | 69 | 57 |
| 25 | 7 | 7 | 5,000 | 96 | 56 |
| 50 | 10 | 10 | 10,000 | 131 | – |
| 100 | 14 | 14 | 25,000 | 183 | – |
| 250 | 22 | 22 | 50,000 | 187 | – |
| 500 | 31 | 30 | 75,000 | 71 | – |

**TABLE III. Standard Errors of Estimated Percentages of Housing Units: 1975 (Excludes estimated percentages of housing units pertaining to Lacking Complete Plumbing Facilities, Sewage Disposal, and Households with Head of Spanish Origin)**

(68 chances of out 100)

| Base of percentage (000) | Estimated percentage[1] | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 or 99 | 2 or 98 | 5 or 95 | 10 or 90 | 15 or 85 | 25 or 75 | 50 |
| 5 | 5.3 | 7.5 | 11.7 | 16.1 | 19.2 | 23.3 | 26.9 |
| 10 | 3.8 | 5.3 | 8.3 | 11.4 | 13.6 | 16.5 | 19.0 |
| 25 | 2.4 | 3.4 | 5.2 | 7.2 | 8.6 | 10.4 | 12.0 |
| 50 | 1.7 | 2.4 | 3.7 | 5.1 | 6.1 | 7.4 | 8.5 |
| 100 | 1.2 | 1.7 | 2.6 | 3.6 | 4.3 | 5.2 | 6.0 |
| 250 | 0.8 | 1.1 | 1.7 | 2.3 | 2.7 | 3.3 | 3.8 |
| 500 | 0.5 | 0.8 | 1.2 | 1.6 | 1.9 | 2.3 | 2.7 |
| 1,000 | 0.4 | 0.5 | 0.8 | 1.1 | 1.3 | 1.6 | 1.9 |
| 2,500 | 0.2 | 0.3 | 0.5 | 0.7 | 0.8 | 1.0 | 1.2 |
| 5,000 | 0.2 | 0.2 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 |
| 10,000 | 0.1 | 0.2 | 0.3 | 0.4 | 0.4 | 0.5 | 0.6 |
| 25,000 | 0.07 | 0.1 | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 |
| 50,000 | 0.05 | 0.08 | 0.1 | 0.2 | 0.2 | 0.2 | 0.3 |
| 75,000 | 0.04 | 0.06 | 0.1 | 0.1 | 0.2 | 0.2 | 0.2 |

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

**TABLE IV. Standard Errors of Estimated Percentages of Housing Units Pertaining to Lacking Complete Plumbing Facilities, Sewage Disposal, and Households with Head of Spanish Origin: 1975**

(68 chances out of 100)

| Base of percentage (000) | Estimated percentage[1] | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 or 99 | 2 or 98 | 5 or 95 | 10 or 90 | 15 or 85 | 25 or 75 | 50 |
| 5 | 6.3 | 8.8 | 13.7 | 18.9 | 22.5 | 27.3 | 31.5 |
| 10 | 4.4 | 6.2 | 9.7 | 13.4 | 15.9 | 19.3 | 22.3 |
| 25 | 2.8 | 3.9 | 6.1 | 8.4 | 10.1 | 12.2 | 14.1 |
| 50 | 2.0 | 2.8 | 4.3 | 6.0 | 7.1 | 8.6 | 10.0 |
| 100 | 1.4 | 2.0 | 3.1 | 4.2 | 5.0 | 6.1 | 7.0 |
| 250 | 0.9 | 1.2 | 1.9 | 2.7 | 3.2 | 3.9 | 4.5 |
| 500 | 0.6 | 0.9 | 1.4 | 1.9 | 2.2 | 2.7 | 3.1 |
| 1,000 | 0.4 | 0.6 | 1.0 | 1.3 | 1.6 | 1.9 | 2.2 |
| 2,500 | 0.3 | 0.4 | 0.6 | 0.8 | 1.0 | 1.2 | 1.4 |
| 5,000 | 0.2 | 0.3 | 0.4 | 0.6 | 0.7 | 0.9 | 1.0 |
| 10,000 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 |
| 25,000 | 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 |

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

tion in table I of this appendix shows that the standard error on an estimate of this size is approximately 145,000. Consequently, the 68-percent confidence interval as shown by these data is from 21,149,000 to 21,439,000. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 housing units of this type lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples, lies within the interval from 21,062,000 to 21,526,000 housing units with 90 percent confidence and that the average estimate lies within the interval from 21,004,000 to 21,584,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 21,294,000 1975 owner-occupied housing units with garbage collection service once a week, 1,478,000, or 6.9 percent, had a family and primary individual income of between $5,000 and $6,999. Interpolation in table III (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage, is approximately .2 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 6.7 to 7.1 percent; the 90-percent confidence interval is from 6.6 to 7.2 percent; and the 95-percent confidence interval is from 6.5 to 7.3 percent.

*Illustration II.*—Table A-3 of this report shows that in the United States there were 823,000 owner-occupied housing units in 1975 lacking some or all plumbing facilities that were occupied three months or longer. Interpolation in table II of this appendix shows that the standard error on an estimate of this size is approximately 39,000. (Table II should be used since this estimate pertains to lacking some or all plumbing facilities.) Consequently, the 68-percent confidence interval as shown by these data is from 784,000 to 862,000. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1975 housing units of this type lying within a range

computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples, lies, within the interval from 761,000 to 885,000 housing units with 90 percent confidence and that the average estimate lies within the interval from 745,000 to 901,000 housing units with 95 percent confidence.

Table A-3 also shows that of the 823,000 1975 owner-occupied housing units lacking some or all plumbing facilities that were occupied three months or longer, 105,000, or 12.8 percent, had a family and primary individual income of between $5,000 and $6,999. Interpolation in table IV (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage is approximately 1.7 percentage points. Consequently, the 68-percent confidence interval, as shown by this data, is from 11.1 to 14.5 percent; the 90-percent confidence interval is from 10.1 to 15.5 percent; and the 95-percent confidence interval is from 9.4 to 16.2 percent.

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

*Illustration of the computation of the standard error of a difference.*—Table A-1 of this report shows that in the United States there were 4,569,000 owner-occupied housing units in 1975 with garbage collection service once a week that had a family and primary individual income of between $10,000 and $14,999. Thus, the apparent difference between the number of owner-occupied housing units with garbage collection service once a week that had a family and primary individual income of between $5,000 and $6,999 and the number that had a family and primary individual income of between $10,000 and $14,999 is 3,091,000. Table I shows that the standard error on an estimate of 1,478,000 to be approximately 45,000, and table I also shows that the standard error on an estimate of 4,569,000 to be approximately 78,000. Therefore, the standard error of the estimated difference of 3,091,000 is about

$$90{,}000 = \sqrt{(45{,}000)^2 + (78{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 3,091,000 difference is from 3,001,000 to 3,181,000 housing units. Therefore, a conclusion that the average estimate of this difference, derived from all possible samples, lying within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 2,947,000 to 3,235,000 housing units, and the 95-percent confidence interval is from 2,911,000 to 3,271,000. Thus we can conclude with 95 percent confidence that of the 1975 owner-occupied housing units with garbage collection service once a week, the number with an income between $10,000 and $14,999 was greater than the number with an income between $5,000 and $6,999.

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

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

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

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

*Illustration of the computation of the 95-percent confidence interval of the median.*—Table A-1 of this report shows that the median family and primary individual income of owner-occupied housing units in the United States with garbage collection service once a week was $14,600 in 1975. The base of the distribution from which this median was determined is 21,294,000 housing units.

1. Table III shows that the standard error of 50 percent on a base of 21,294,000 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 four categories that 6,466,000 housing units, or 30.4 percent, had a family and primary individual income of up to $9,999 and that an additional 4,569,000, or 21.5 percent, had a family and primary individual income of between $10,000 and $14,999. By linear interpolation, the lower limit

of the 95-percent confidence interval is found to be about

$$\$9{,}999 + (\$5{,}000)\left(\frac{49.2 - 30.4}{21.5}\right) = \$14{,}400$$

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

$$\$9{,}999 + (\$5{,}000)\left(\frac{50.8 - 30.4}{21.5}\right) = \$14{,}700$$

Thus, the 95-percent confidence interval ranges from $14,400 to $14,700.

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

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error.

**Reinterview program.**—For the AHS-National sample a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and thus were the basis for the measurement of the "content" error of these AHS estimates.

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

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

This check was for interviewer evaluation and control.

The results of the 1975 reinterview study are not available, however, it is felt that they will be similar to the results of the 1974 reinterview study, which are presented in the following Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample: 1974."

Twenty-five items were included in the 1974 reinterview study, with some being selected for the first time. The estimated indexes of inconsistency for these items ranged from 4 to 35 with most items in the 20 to 30 range. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels and were included in the AHS reinterview programs because they had not been previously tested. A 20-50 interval is considered moderate, indicating that there is some problem with inconsistent reporting on a range of 0-100, with a high index associated with a high level of response variability.

The 1970 census reinterview results provide illustrations of possible nonsampling errors for some of the items which also appear in the AHS. For example, median value of homes was consistently underestimated by about five percent, and the average monthly costs of electricity and utility gas were consistently overestimated although the effect on the average gross rent figures 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, the results of the reinterview studies are derived from sample surveys so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of this study.

**Coverage Errors.**—With respect to errors of coverage and estimation for missing data, it was mentioned previously in the section on estimation that the AHS new construction sample had deficiencies in the representation of both conventional new construction in permit-issuing areas and new construction mobile homes. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970), and therefore would have been represented in the sample selected from 1970 census units. However, it has been estimated that the 1975 AHS sample misses about six percent (i.e., about 600,000 units) of all conventional new construction (i.e., all conventional housing units built after April 1970, in both permit-issuing and non-permit-issuing areas) because the permits for these missed units, which were built after April 1970, were issued before 1970.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new construction mobile homes. However, new mobile homes in segments where area sampling methods were used do come into the AHS sample. In addition, new mobile homes in segments sampled from the 1970 census list also come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. However, new mobile homes in these segments that are located in mobile home parks not in existence at the time of the 1970 census or not identified as such in the 1970

census have no chance of coming into the AHS sample. It has been estimated that the 1975 AHS sample misses at least 200,000 new mobile homes. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies although some bias in the AHS sample still exists. This is especially true for new mobile homes since it is believed that the corresponding independent estimate substantially understates the actual new mobile home inventory.

**Rounding errors in AHS.**—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 the survey.

☆ U. S. GOVERNMENT PRINTING OFFICE : 1977--240-869/1075