APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN
AHS-NATIONAL App-33
Selection of sample areas App-33
Designation of sample housing units for the 1974 enumeration . App-33
Selection of the 1973 sample housing units App-33
Splitting of the sample App-34
Selection of supplemental sample housing units in rural areas App-34
1970 Census of Population and Housing App-34
ESTIMATION App-34
AHS-National sample App-34
1974 housing inventory App-34
1973-1974 lost units App-36
Ratio estimation procedure of the 1970 Census of Population and Housing App-36
RELIABILITY OF THE ESTIMATES App-36
Sampling errors for the 1974 AHS-National sample App-36
Illustrations of the use of the standard error tables.—Illustration I App-38
Illustration II App-38
Differences App-38
Illustration of the computation of the standard error of a difference App-39
Medians App-39
Illustration of the computation of the 95-percent confidence interval for a median App-39
Nonsampling errors App-39
1970 census App-39
Reinterview program App-40
Coverage errors App-41

SAMPLE DESIGN AHS-NATIONAL

The 1974 estimates are based on data collected in August through October 1974 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 71,300 sample housing units (both occupied and vacant) were eligible for interview in the 1974 AHS. Of this number, 2,200 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 71,300, there were also 7,000 sample units which were visited but found not to provide information relevant to the 1974 housing inventory. Of the 7,000, however, about 800 units were identified as units in existence in 1973 which were removed from the housing inventory that existed in 1974. This sample was used to obtain estimates of characteristics of units removed from the housing inventory between 1973 and 1974.

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 1974 enumeration.—The sample housing units designated to be interviewed in the 1974 enumeration consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1973 enumeration.
2. All 1973 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 noninterviews, see facsimile of 1974 AHS questionnaire, page 1.)
3. All sample housing units that were selected in the 1974 rural supplementation.
4. All sample housing units that were selected from the list of building permits issued since the 1973 enumeration. (This sample represents the housing units built since the 1973 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 a 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 subdivided into areas with an expected size of four housing units.

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). 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 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 series 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 1974 housing inventory and estimates of units removed from the housing inventory between 1973 and 1974, (i.e., 1973-1974 lost units). Each type of estimate employed a separate, though similar, estimation procedure, as described below.

1974 housing inventory.—The 1974 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 estimate 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 six categories of new construction units (i.e., four categories for conventional new construction units and two 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 so as 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 estimate 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 for conventional new construction units 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 estimate factor was then applied to the existing weight for each sample unit in each second-stage ratio estimation category.

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

The third-stage ratio estimate 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 estimate 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 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.

1973-1974 lost units.—The 1973-1974 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 1973-1974 lost units existed by definition in the 1973 housing inventory, there was a 1973 housing inventory weight associated with each 1973-1974 lost unit.

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 series report, HC(1)-B-1, **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 can be obtained in the 1970 census report, **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 for the 1974 AHS-National sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling 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 the 1974 housing inventory estimates in this report, and Table III presents the standard errors applicable to 1973-1974 lost housing unit estimates in this report. Table I shows the approximate standard errors applicable to all 1974 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.

TABLE I. **Standard Errors of Estimated Number of Housing Units: 1974 (Excludes estimates of housing units pertaining to Cooking Fuel, Households Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Source of Water, and Households with Head of Spanish Origin)**

(68 chances out of 100)

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

TABLE II. **Standard Errors of Estimated Numbers of Housing Units Pertaining to Cooking Fuel, Households Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Source of Water, and Households with Head of Spanish Origin: 1974**

(68 chances out of 100)

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

Tables IV, V, and VI present the standard errors of estimated percentages. Table IV shows the approximate standard errors of all estimated percentages of the 1974 housing inventory except those pertaining to the specified items in table V, and table VI 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, V, and VI.

TABLE III. **Standard Errors of Estimated Number of Lost Housing Units: 1973-1974**

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

TABLE IV. **Standard Errors of Estimated Percentages of Housing Units: 1974 (Excludes estimated percentages of housing units pertaining to Cooking Fuel, Households 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 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	0.1	0.1	0.2	0.2	0.3	0.3	0.4
50,000	0.1	0.1	0.1	0.2	0.2	0.2	0.3
75,000	0.0	0.1	0.1	0.1	0.2	0.2	0.2

TABLE V. **Standard Errors of Estimated Percentages of Housing Units Pertaining to Cooking Fuel, Households Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Source of Water, and Households with Head of Spanish Origin: 1974**

(68 chances out of 100)

Base of percentage (000)	Estimated percentage						
	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	0.1	0.1	0.2	0.3	0.3	0.4	0.4
50,000	0.1	0.1	0.1	0.2	0.2	0.3	0.3

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

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

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

Illustrations of the use of the standard error tables.—Illustration I.—Table A-1 of this report shows that inside SMSA's in the United States there were 8,826,000 owner-occupied housing units with two persons in 1974. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 105,000. Consequently, the 68-percent confidence interval, as shown by these data, is from 8,721,000 to 8,931,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples of 1974 owner-occupied housing units inside SMSA's with two persons, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples, lies within the interval from 8,658,000 to 8,994,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 8,616,000 to 9,036,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 8,826,000 owner-occupied housing units with two persons inside SMSA's, 3,522,000, or 39.9 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 39.3 to 40.5 percent; the 90-percent confidence interval is from 38.9 to 40.9 percent, and the 95-percent confidence interval is from 38.7 to 41.1 percent.

Illustration II.—Table A-8 of this report shows that inside SMSA's in the United States there were 179,000 owner-occupied housing units with two persons whose head was of Spanish origin in 1974. Interpolation in table II of this appendix shows that the standard error of an estimate of this size is approximately 18,000. Consequently, the 68-percent confidence interval, as shown by these data, is from 161,000 to 197,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples of 1974 owner-occupied housing units inside SMSA's with two persons whose head is of Spanish origin, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples, lies within the interval from 150,000 to 208,000 housing units with 90 percent confidence; and that the average estimate lies within the interval 143,000 to 215,000 housing units with 95 percent confidence.

Table A-8 also shows that of the 179,000 owner-occupied housing units with two persons inside SMSA's whose head is of Spanish origin, 101,000, or 56.4 percent, were in central cities. Interpolation in table V (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage is approximately 5.5 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 50.9 to 61.9 percent; the 90-percent confidence interval is from 47.6 to 65.2 percent; and the 95-percent confidence interval is from 45.4 to 67.4 percent.

TABLE VI. **Standard Errors of Estimated Percentages of Lost Housing Units: 1973-1974**

(68 chances out of 100)

Base of percentage (000)	Estimated percentage						
	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	6.2	8.7	13.5	18.6	22.1	26.8	31.0
10	4.4	6.1	9.5	13.1	15.6	19.0	21.9
25	2.8	3.9	6.0	8.3	9.9	12.0	13.9
50	1.9	2.7	4.3	5.9	7.0	8.5	9.8
100	1.4	1.9	3.0	4.2	4.9	6.0	6.9
250	0.9	1.2	1.9	2.6	3.1	3.8	4.4
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	0.9	1.2	1.4

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 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 5,368,000 owner-occupied housing units with three persons in 1974. Thus, the apparent difference between the number of 1974 owner-occupied housing units with two persons and those with three persons is 3,458,000. The standard error of 8,826,000 is approximately 105,000 as is shown above. Table I shows that the standard error on an estimate of 5,368,000 to be approximately 84,000. Therefore, the standard error of the estimated difference of 3,458,000 is about

$$134{,}000 = \sqrt{(105{,}000)^2 + (84{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 3,458,000 difference is from 3,324,000 to 3,592,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 3,244,000 to 3,672,000 housing units, and the 95-percent confidence interval is from 3,190,000 to 3,726,000. Thus, we can conclude with 95 percent confidence that the number of 1974 owner-occupied housing units inside SMSA's with two persons is greater than the number with three persons.

Medians.—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median 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 inside SMSA's was 3.0 in 1974. The base of the distribution, from which this median was determined is 29,708,000 housing units.

1. From table IV, the standard error of a 50-percent characteristic on the base of 29,708,000 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 12,352,000 owner-occupied housing units, or 41.6 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 5,368,000 owner-occupied housing units, or 18.1 percent, had three persons (i.e., 2.5 to 3.5 persons). By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$2.5 + (3.5-2.5)\left(\frac{49.2-41.6}{18.1}\right) = 2.9$$

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

$$2.5 + (3.5-2.5)\left(\frac{50.8-41.6}{18.1}\right) = 3.0$$

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 upper limit, it actually is a reflection of the rounding error associated with this median (see the paragraph on rounding error in the nonsampling 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 1974 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 the following:

1. *Space errors.*—Errors in which both the living quarters and its occupants were missed in the census or in which they were counted more than once

(overenumerated). Space errors are usually the largest component of housing coverage error.

2. *Definitional errors.*—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home, and consequently, receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one living quarters would have been counted where two existed.

3. *Occupancy errors.*—Errors of incorrect occupancy classification for enumerated units; i.e., vacant units that were improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on the census questionnaire. In these studies, content errors were measured by the following methods:

1. *Reinterviews.*—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.

2. *Record checks.*—The comparison of census data with data obtained from an independent record source.

3. *Other surveys.*—Comparisons of census data with that obtained from other sample 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 report 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.** Some of the results are:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent (after processing) . . . . For each 100 units that were finally enumerated in the census, an estimated 2.5 were missed . . . ." (PHC(E)-5, p.3)

2. "The occupied space missed rate for the total United States in 1970 is estimated at . . . 1.4 percent [after processing] . . . . About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures." (PHC(E)-5, p. 4)

3. "In 1970, the definitional underenumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error rate of 0.34 of 1 percent." (PHC(E)-5, p. 11)

4. "Most of the vacant units that were enumerated as occupied were procedural errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p. 16) "Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant." (PHC(E)-5, p. 15)

5. "Generally, owners report housing data more consistently than renters, responses for occupied units are more consistent than those for vacant units, and respondents in single-unit structures report more consistently than those in multiunit structures." (PHC(E)-10, p. 6) This statement applies to the following items:

a. Heating fuel
b. Renters paying extra for utilities and/or fuel
c. Bathtub or shower facilities
d. Flush toilet facilities
e. Telephone availability
f. Year structure built
g. Value of home
h. Seasonal vacancy status

6. "Heating fuel, year structure built, and value of home are more consistently reported for units with a non-Negro head while . . . bathtub or shower facilities, and flush toilet facilities are more consistently reported for units with a Negro head." (PHC(E)-10, p.8).

The results of these studies were based on sample data, so there is sampling error associated with the estimates of nonsampling error. The possibility of such errors should be taken into account when considering the results. A detailed description of the sample design and estimation procedure for each study is given in PHC(E)-5 and PHC(E)-10.

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 reinterview study are presented in the Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample: 1974." Twenty-five items were included in the reinterview, some for the first time. The estimated indexes of inconsistency of these items ranged from 4 to 35 with most items in the 20 to 30 range. (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 items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels.

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

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 1974 AHS sample missed about 6 percent (i.e., about 600,000 units) of all conventional new construction (i.e., in both permit-issuing and non-permit-issuing areas), because the permits for these 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. New mobile homes in area segments, however, do come into the AHS sample. In addition, new mobile homes in address segments 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 address 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 1974 AHS sample missed about one-fourth of all new mobile homes (i.e., about 400,000 units). 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.

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.

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-29
Selection of sample areas App-29
Designation of sample housing units for the 1974 enumeration. . App-29
Selection of the 1973 sample housing units App-29
Splitting of the sample App-30
Selection of supplemental sample housing units in rural areas App-30

ESTIMATION..... App-30

RELIABILITY OF THE ESTIMATES App-31
Sampling errors App-31
Illustration of the use of the standard error tables App-32
Differences App-33
Illustration of the computation of the standard error of a difference App-34
Nonsampling errors App-34
Reinterview program App-34
Coverage errors App-34

SAMPLE DESIGN

The estimates are based on data collected in August through October 1974 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 71,300 sample housing units (both occupied and vacant) were eligible for interview in the 1974 AHS. Of this number, 2,200 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 71,300, there were also 7,000 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 independently of the other PSU selected from this stratum. Since the two PSU's were independently selected, it was possible for the same PSU to be selected twice. This occurred in 25 instances, producing an additional 85 NSR sample PSU's.

Designation of sample housing units for the 1974 enumeration.—The sample housing units designated to be interviewed in the 1974 enumeration consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1973 enumeration.
2. All 1973 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 noninterviews, see facsimile of 1974 AHS questionnaire, page 2.)
3. All sample housing units that were selected in the 1974 rural supplementation.
4. All sample housing units that were selected from the list of building permits issued since the 1973 enumeration. (This sample represents the housing units built since the 1973 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 a 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 subdivided into areas with an expected size of four housing units.

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). 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 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.

ESTIMATION

The 1974 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 estimate 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 six categories of new construction units (i.e., four categories for conventional new construction units and two 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 so as 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 estimate 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 for conventional new construction units 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 estimate factor was then applied to the existing weight for each sample unit in each second-stage ratio estimation category.

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

The third-stage ratio estimate 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 estimate 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 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 stand-

ard 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 where x is not a subclass of y, the above tables; 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:

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

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

Illustration of the use of the standard error tables.—Illustration I.—Table A-2 of this report shows that in the United States there were 13,007,000 renter-occupied housing units with common

TABLE I. **Standard Errors of Estimated Number of Housing Units: 1974 (Excluding Estimates of Housing Units Pertaining to Garbage Collection Service, Water Supply, and Households with Head of Spanish Origin)**

(68 chances out of 100)

Size of Estimate (000)	Standard Error: White and Total (000)	Standard Error: Negro (000)	Size of Estimate (000)	Standard Error: White and Total (000)	Standard Error: Negro (000)
5	3	3	1,000	37	34
10	4	4	2,500	57	47
25	6	6	5,000	79	45
50	8	8	10,000	108	–
100	12	12	25,000	151	–
250	18	18	50,000	152	–
500	26	25	75,000	37	–

TABLE II. **Standard Errors of Estimated Numbers of Housing Units Pertaining to Garbage Collection Service, Water Supply, and Households with Head of Spanish Origin: 1974**

(68 chances out of 100)

Size of Estimate (000)	Standard Error (000)	Size of Estimate (000)	Standard Error (000)
5	3	1,000	42
10	4	2,500	66
25	7	5,000	92
50	10	10,000	125
100	13	25,000	176
250	21	50,000	183
500	30		

TABLE III. **Standard Errors of Estimated Percentages of Housing Units: 1974 (Excluding Estimated Percentages of Housing Units Pertaining to Garbage Collection Service, Water Supply, and Households with Head of Spanish Origin)**

(68 chances out of 100)

Base of Percentage (000)	Estimated Percentage						
	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	5.2	7.3	11.3	15.6	18.6	22.5	26.0
10	3.7	5.1	8.0	11.0	13.1	15.9	18.4
25	2.3	3.3	5.1	7.0	8.3	10.1	11.6
50	1.6	2.3	3.6	5.0	5.9	7.1	8.2
100.	1.2	1.6	2.5	3.5	4.2	5.0	5.8
250.	0.7	1.0	1.6	2.2	2.6	3.2	3.7
500.	0.5	0.7	1.1	1.6	1.9	2.3	2.6
1,000	0.4	0.5	0.8	1.1	1.3	1.6	1.8
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 . . .	0.1	0.1	0.2	0.2	0.3	0.3	0.4
50,000 . . .	0.1	0.1	0.1	0.2	0.2	0.2	0.3
75,000 . . .	—	0.1	0.1	0.1	0.2	0.2	0.2

TABLE IV. **Standard Errors of Estimated Percentages of Housing Units Pertaining to Garbage Collection Service, Water Supply, and Households with Head of Spanish Origin: 1974**

(68 changes out of 100)

Base of Percentage (000)	Estimated Percentage						
	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	6.0	8.4	13.1	18.0	21.4	26.0	30.0
10	4.2	5.9	9.2	12.7	15.1	18.4	21.2
25	2.7	3.8	5.8	8.0	9.6	11.6	13.4
50	1.9	2.7	4.1	5.7	6.8	8.2	9.4
100.	1.3	1.9	2.9	4.0	4.8	5.8	6.7
250.	0.8	1.2	1.8	2.5	3.0	3.7	4.2
500.	0.6	0.8	1.3	1.8	2.1	2.6	3.0
1,000 . . .	0.4	0.6	0.9	1.3	1.5	1.8	2.1
2,500 . . .	0.3	0.4	0.6	0.8	1.0	1.2	1.3
5,000 . . .	0.2	0.3	0.4	0.6	0.7	0.8	0.9
10,000 . .	0.1	0.2	0.3	0.4	0.5	0.6	0.7
25,000 . .	0.1	0.1	0.2	0.3	0.3	0.4	0.4
50,000 . .	0.1	0.1	0.1	0.2	0.2	0.3	0.3

stairways. Interpolation in table I above shows that the standard error of an estimate of this size is approximately 117,000. Consequently, the 68-percent confidence interval, as shown by these data, is from 12,890,000 to 13,124,000 housing units. Therefore, a conclusion that the average estimate of 1974 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 12,820,000 to 13,194,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 12,773,000 to 13,241,000 housing units with 95 percent confidence.

Table A-2 also shows that of the 13,007,000 renter-occupied housing units with common stairways, 11,622,000 or 89.4 percent, were located inside SMSA's. Interpolation in table III (i.e., interpolation on both the base and percent) 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 89.0 to 89.8 percent; the 90-percent confidence interval is from 88.8 to 90.0 percent; and the 95-percent confidence interval is from 88.6 to 90.2 percent.

Illustration II.—Table A-11 of this report shows that in the United States in 1974 there were 135,000 owner-occupied housing units with head of Spanish origin which had fuse or switch blowouts. Interpolation in table II shows that the standard error of an estimate of this size is approximately 15,000. Consequently, the 68-percent confidence interval is from 120,000 to 150,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1974 owner-occupied housing units with head of Spanish origin which had fuse or switch blowouts 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 111,000 to 159,000 housing units with 90 percent confidence; and that average estimate lies within the interval from 105,000 to 165,000 housing units with 95 percent confidence.

Table A-11 also shows that of the 135,000 1974 owner-occupied housing units with head of Spanish origin which had fuse or switch blowouts, 40,000, or 29.6 percent had blowouts three times or more. Interpolation in table IV (i.e., interpolation on both the base and percent) shows that the standard error of the above percentage is 5.5 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 24.1 to 35.1 percent; the 90-percent confidence interval is from 20.8 to 38.4 percent; and the 95-percent confidence interval is from 18.6 to 40.6 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 standard error.

Illustration of the computation of the standard error of a difference.—Table A-11 of this report shows that of the Nation's owner-occupied housing units with head of Spanish origin which had fuse or switch blowouts, 71,000 had only one blowout. Thus, the apparent difference between the number of 1974 owner-occupied housing units with head of Spanish origin that had blowouts three times or more and the number that had blowouts just one time is 31,000.

Interpolation in table II shows the standard error on an estimate of 40,000 to be approximately 9,000 and the standard error on an estimate of 71,000 to be approximately 11,000. Therefore, the standard error of the estimated difference of 31,000 is about

$$14{,}000 = \sqrt{(9{,}000)^2 + (11{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 31,000 difference is from 17,000 to 45,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 9,000 to 53,000 housing units, and the 95-percent confidence interval is from 3,000 to 59,000. Thus, we can conclude with 95 percent confidence that the number of 1974 owner-occupied housing units with head of Spanish origin that had three or more fuse or switch blowouts is less than the number that had only one blowout, 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 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, i.e., tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items.

The results of the reinterview study are presented in the following Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey-National Sample: 1974." Some of the results are:

1. "Overall, the results showed a fairly consistent response between the original interview and the reinterview."
2. "For several of the basic survey items or categories within them, the measure of inconsistent response was in the 20-50 range which is moderate and indicates there is some problem with inconsistent reporting." (The range is from 0-100 with a high index being associated with a high level of response variability.)

The results of the 1974 study were based on sample data so there is sampling error associated with the estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of the 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 presentation 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 1974 AHS sample misses about 7 percent (i.e., about 600,000 units) of all conventional new construction (i.e., in both permit-issuing and non-permit-issuing areas) because the permits for these 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. New mobile homes in area segments, however, do come into the AHS sample. In addition, new mobile homes in address segments 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 address 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 1974 AHS sample missed about one-fourth of all new mobile homes (i.e., about 400,000 units). 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.

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.

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-30
Selection of sample areas. App-30
Designation of sample housing units for the 1974 enumeration . App-30
Selection of the 1973 sample housing units. App-30
Splitting of the sample App-31
Selection of supplemental sample housing units in rural areas. . . . App-31
ESTIMATION. App-31
RELIABILITY OF THE ESTIMATES App-32
Sampling errors App-32
Illustration of the use of the standard error tables App-33
Differences. App-34
Illustration of the computation of the standard error of a difference. App-34
Medians App-35
Illustration of the computation of the standard error of a median App-35
Nonsampling errors. App-35

SAMPLE DESIGN

The estimates are based on data collected in August through October 1974 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 71,300 sample housing units (both occupied and vacant) were eligible for interview in the 1974 AHS. Of this number, 2,200 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 71,300, there were also 7,000 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 1974 enumeration.—The sample housing units designated to be interviewed in the 1974 enumeration consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1973 enumeration.
2. All 1973 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 noninterviews, see facsimile of 1974 AHS questionnaire, page 1.)
3. All sample housing units that were selected in the 1974 rural supplementation.
4. All sample housing units that were selected from the list of building permits issued since the 1973 enumeration. (This sample represents the housing units built since the 1973 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 a 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 subdivided into areas with an expected size of four housing units.

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). One can except 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 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; wheras, the overall probability of selection for sample housing units in urban areas remained at 1 in 1,366.

ESTIMATION

The 1974 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 estimate 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 six categories of new construction units (i.e., four categories for conventional new construction

units and two for new construction mobile homes). This adjustment was necessary so as to correct for known deficiencies in the AHS sample with regard to representation of new construction units (see the section on non-sampling error).

The second-stage ratio estimate 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 for conventional new construction 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 4 types of vacant housing units and for 24 residence-tenure-race of head-sex of head categories for occupied housing units.

The third-stage ratio estimate 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 estimate 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 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 erors), 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.

TABLE I. **Standard Errors of Estimated Number of Housing Units: 1974 (Excludes estimates of housing units pertaining to Cooking Fuel, Households Lacking Complete Plumbing Facilities, 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: Negro (000)
5	3	3
10	4	4
25	6	6
50	8	8
100	12	12
250	19	19
500	27	26
1,000	38	35
2,500	59	49
5,000	82	48
10,000	112	-
25,000	156	-
50,000	160	-
75,000	61	-

TABLE II. **Standard Errors of Estimated Numbers of Housing Units Pertaining to Cooking Fuel, Households Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Source of Water, and Households with Head of Spanish Origin: 1974**

(68 chances out of 100)

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

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, where x is not a subclass of y, the above tables; 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:

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

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

Illustration of the use of the standard error tables. Illustration 1.—Table A-2 of this report shows that in the United States there were 9,084,000 specified owner-occupied housing units with 2 bedrooms in 1974. 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 is from 8,977,000 to 9,191,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1974 owner-occupied housing units with 2 bedrooms lies within a range computed in this way would be correct for roughly

TABLE III. **Standard Errors of Estimated Percentages of Housing Units: 1974 (Excludes estimated percentages of housing units pertaining to Cooking Fuel, Households 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 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 . . .	0.1	0.1	0.2	0.2	0.3	0.3	0.4
50,000 . . .	0.1	0.1	0.1	0.2	0.2	0.2	0.3
75,000 . . .	0.0	0.1	0.1	0.1	0.2	0.2	0.2

TABLE IV. **Standard Errors of Estimated Percentages of Housing Units Pertaining to Cooking Fuel, Households Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Source of Water, and Households with Head of Spanish Origin: 1974**

(68 chances out of 100)

Base of percentage (000)	Estimated percentage						
	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 . . .	0.1	0.1	0.2	0.3	0.3	0.4	0.4
50,000 . . .	0.1	0.1	0.1	0.2	0.2	0.3	0.3

68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples, lies within the interval from 8,913,000 to 9,255,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 8,870,000 to 9,298,000 housing units with 95 percent confidence.

Table A-2 also shows that of the 9,084,000 specified owner-occupied housing units with two bedrooms, 1,622,000 or 17.9 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 .5 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 17.4 to 18.4 percent; the 90-percent confidence interval is from 17.2 to 18.7 percent; and the 95-percent confidence interval is from 16.9 to 18.9 percent.

Illustration 2.—Table A-9 shows that in the Nation there were 147,000 specified renter-occupied housing units with Spanish head with five persons. Interpolation in table II shows that the standard error of an estimate of this size is approximately 17,000 (table II should be used since this estimate pertains to households with Spanish head). Consequently, we could conclude with 68 percent confidence that the average estimate of 1974 Spanish renter units with five persons lies within the interval from 130,000 to 164,000 housing units. Similarly, the 90-percent confidence interval is from 120,000 to 174,000 housing units; and the 95-percent confidence interval is from 113,000 to 181,000 housing units.

Table A-9 also shows that of the 147,000 specified Spanish renters with five persons, 58,000 or 39.5 percent paid between $100 to $149 a month for rent. 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 5.9 percentage points. Consequently, the 68-percent confidence interval is from 33.6 to 45.4 percent; the 90-percent confidence interval is from 30.1 to 48.9 percent; and the 95-percent confidence interval is from 27.7 to 51.3 percent.

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

Illustration of the computation of the standard error of a difference.—Table A-2 of this report shows that in the United States there were 1,575,000 owner-occupied housing units, with two bed-

rooms, valued between $10,000 and $14,999 in 1974. Thus the apparent difference between the number of 1974 owner-occupied housing units, with two bedrooms, valued between $10,000 and $14,999 and those valued between $15,000 and $19,999 is 47,000. Table I shows the standard error of 1,622,000 is approximately 47,000 and the standard error of 1,575,000 is approximately 46,000. Therefore, the standard error of the estimated difference of 47,000 is about

$$66{,}000 = \sqrt{(47{,}000)^2 + (46{,}000)^2}.$$

Consequently, the 68-percent confidence interval for the 47,000 difference is from −19,000 to 113,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples of this difference, 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 −59,000 to 153,000 housing units, and the 95-percent confidence interval is from −85,000 to 179,000. Thus, we *cannot* conclude with 95 percent confidence that the number of 1974 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 includes zero or negative values. We also *cannot* conclude with 90 percent confidence, or even 68 percent confidence, that the 1,622,000 is greater than the 1,575,000 since both the 90-percent and 68-percent confidence intervals 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 standard error of a median.—Table A-2 of this report shows the median value of owner-occupied housing units with two bedrooms was $19,900 in 1974. The base of the distribution, from which this median was determined is 9,084,000 housing units.

1. Table III shows that the standard error of 50 percent on a base of 9,084,000 is approximately .6 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 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 2,950,000 owner-occupied housing units with two bedrooms, or 32.5 percent, had a value less than $15,000, and an additional 1,622,000 owner-occupied housing units with two bedrooms, or 17.9 percent, had a value between $15,000 and $19,999, and an additional 1,379,000, or 15.2 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

$$\$15{,}000 + (\$20{,}000 - \$15{,}000)\left(\frac{48.8 - 32.5}{17.9}\right) = \$19{,}600$$

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

$$\$20{,}000 + (\$25{,}000 - \$20{,}000)\left(\frac{51.2 - 32.5 - 17.9}{15.2}\right) = \$20{,}300$$

Thus the 95-percent confidence interval ranges from $19,600 to $20,300.

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. However, for the 1974 AHS-National sample, a study was conducted to obtain a measurement of some of the nonsampling errors 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 two observations (i.e., the original interview and the reinterview) were assumed to be two independent readings and thus were the basis for the measurement of the accuracy of the AHS estimates.

As part of the reinterview, a check was made at each of these households to deter-

mine 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 reinterview study are presented in the Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample: 1974." Twenty-five items were included in the reinterview, some for the first time. The estimated indexes of inconsistency of these items ranged from 4 to 35 with most items in the 20 to 30 range. (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 items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have higher inconsistency levels.

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

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 1974 AHS sample missed about 7 percent (i.e., about 600,000 units) of all conventional new construction (i.e., in both permit-issuing and non-permit-issuing areas) because the permits for these 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. New mobile homes in area segments, however, do come into the AHS sample. In addition, new mobile homes in address segments 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 address 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 1974 AHS sample missed about one-fourth of all new mobile homes (i.e., about 400,000 units). 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.

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.

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-26
Selection of sample areas App-26
Designation of sample housing units for the 1974 enumeration .. App-26
Selection of the 1973 sample housing units App-26
Splitting of the sample App-27
Selection of supplemental sample housing units in rural areas App-27
ESTIMATION App-27
RELIABILITY OF THE ESTIMATES App-28
Sampling errors App-28
Illustrations of the use of the standard error tables App-29
Differences App-30
Illustration of the computation of the standard error of a difference . App-30
Medians App-31
Illustration of the computation of the standard error of a median ... App-31
Nonsampling errors App-31

SAMPLE DESIGN

The estimates are based on data collected in August through October 1974 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 71,300 sample housing units (both occupied and vacant) were eligible for interview in the 1974 AHS. Of this number, 2,200 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 71,300, there were also 7,000 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 1974 enumeration.—The sample housing units designated to be interviewed in the 1974 enumeration consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1973 enumeration.

2. All 1973 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 the Type A and Type B noninterviews, see facsimile of 1974 AHS questionnaire, page 1.)

3. All sample housing units that were selected in the 1974 rural supplementation.

4. All sample housing units that were selected from the list of building permits issued since the 1973 enumeration. (This sample represents the housing units built since the 1973 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 a 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 subdivided into areas with an expected size of four housing units.

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). 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 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.

ESTIMATION

The 1974 AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the Type A noninterview housing units encountered in the AHS. This noninterview adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

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

The first-stage ratio estimation procedure was employed for sample housing units from non-self-representing (NSR) PSU's only. 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 six categories of new construction units (i.e., four categories for conventional new construction units and two for new construction mobile homes). This adjustment was necessary so

as to correct for known deficiencies in the AHS sample with regard to representation of new construction units (see the section on non-sampling error).

The second-stage ratio estimate 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 for conventional new construction 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 4 types of vacant housing units and for 24 residence-tenure-race of head-sex of head categories for occupied housing units.

The third-stage ratio estimate 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 estimate 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 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 measure 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 estimates not specifically shown in tables I and II.

TABLE I. **Standard Errors of Estimated Number of Housing Units: 1974 (Excludes estimates of housing units pertaining to Cooking Fuel, Households 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: Negro (000)
5	3	3
10	4	4
25	6	6
50	8	8
100	12	12
250	19	19
500	27	26
1,000	38	35
2,500	59	49
5,000	82	48
10,000	112	–
25,000	156	–
50,000	160	–
75,000	61	–

TABLE II. **Standard Errors of Estimated Numbers of Housing Units Pertaining to Cooking Fuel, Households Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Source of Water, and Households With Head of Spanish Origin: 1974**

(68 chances out of 100)

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

The reliability of an estimated percentage depends upon 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, where x is not a subclass of y, the following tables; 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:

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

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

Illustrations of the use of the standard error tables:—Illustration 1.—Table A-1 of this report shows that in the United States there were 4,482,000 owner-occupied housing units occupied by recent movers in 1974. Interpolation in table I of this appendix shows the standard error on an estimate of this size is approximately 77,000. Consequently, the 68-percent confidence interval, as shown by these data, is from 4,405,000 to 4,559,000 housing units. Therefore, a conclusion that the average estimate,

TABLE III. **Standard Errors of Estimated Percentages of Housing Units: 1974 (Excludes estimated percentages of housing units pertaining to Cooking Fuel, Households 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 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 ...	0.1	0.1	0.2	0.2	0.3	0.3	0.4
50,000 ...	0.1	0.1	0.1	0.2	0.2	0.2	0.3
75,000 ...	0.0	0.1	0.1	0.1	0.2	0.2	0.2

TABLE IV. **Standard Errors of Estimated Percentages of Housing Units Pertaining to Cooking Fuel, Households Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Source of Water, and Households With Head of Spanish Origin: 1974**

(68 chances out of 100)

Base of percentage (000)	Estimated percentage						
	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 ...	0.1	0.1	0.2	0.3	0.3	0.4	0.4
50,000 ...	0.1	0.1	0.1	0.2	0.2	0.3	0.3

derived from all possible samples, of 1974 housing units of this type lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples, lies within the interval from 4,359,000 to 4,605,000 housing units with 90 percent confidence and that the average estimate lies within the interval from 4,328,000 to 4,636,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 4,482,000 1974 owner-occupied housing units occupied by recent movers in the United States, 366,000 or 8.2 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 7.7 to 8.7 percent; the 90-percent confidence interval is from 7.4 to 9.0 percent; and the 95-percent confidence interval is from 7.2 to 9.2 percent.

Illustration 2.—Table A-19 shows that in the Nation there were 125,000 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 15,000 (table II should be used since this estimate pertains to households with Spanish head). Consequently, we could conclude with 68 percent confidence that the average estimate of 1974 homeowner units occupied by Spanish recent movers lies within the interval from 110,000 to 140,000 housing units. Similarly, the 90-percent confidence interval is from 101,000 to 149,000 housing units; and the 95-percent confidence interval is from 95,000 to 155,000 housing units.

Table A-19 also shows that of the 125,000 homeowner units occupied by Spanish recent movers 38,000 or 30.4 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 this data, is from 24.5 to 36.3 percent; the 90-percent confidence interval is from 21.0 to 39.8 percent; and the 95-percent confidence interval is from 18.6 to 42.2 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 standard 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 457,000 owner-

occupied housing units occupied by recent movers with five persons in 1974. 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 91,000. Table I shows that the standard error on an estimate of 366,000 to be approximately 23,000 and that the standard error on an estimate of 457,000 to be approximately 26,000. Therefore, the standard error of the estimated difference of 91,000 is about

$$35{,}000 = \sqrt{(23{,}000)^2 + (26{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 91,000 difference is from 56,000 to 126,000 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 35,000 to 147,000 housing units and the 95-percent confidence interval is from 21,000 to 161,000 housing units. 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 1974.

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 I, II, III, and IV, in conjunction with the formula for the standard error of a percentage, 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 standard error of 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 1974. The base of the distribution, from which this median was determined, is 4,482,000 housing units.

1. Table III shows that the standard error of 50 percent on a base of 4,482,000 is .9.

2. To obtain a two-standard error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 48.2 and 51.8.

3. From table A-1, it can be seen by cumulating the frequencies for the first two categories that 1,721,000 owner-occupied housing units occupied by recent movers, or 38.4 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 960,000 housing units, or 21.4 percent, had three persons (i.e., 2.5 to 3.5 persons). By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$2.5 + (3.5-2.5)\left(\frac{48.2 + 38.4}{21.4}\right) = 3.0$$

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

$$2.5 + (3.5-2.5)\left(\frac{51.8 + 38.4}{21.4}\right) = 3.1$$

Thus, the 95-percent confidence interval ranges from 3.0 to 3.1 persons. Although it appears that this confidence interval has the sample estimate as the lower limit, it actually is a reflection of the rounding error associated with the median (see the paragraph on rounding error in the nonsampling error section).

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, for the 1974 AHS-National sample, a study was conducted to obtain a measurement of some of the nonsampling errors 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 two observations (i.e., the original interview and the reinterview) were assumed to be two independent readings and thus were the basis for the measurement of the accuracy of the AHS 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 reinterview study are presented in the Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample: 1974." Twenty-five items were included in the reinterview, some for the first time. The estimated indexes of inconsistency of these items ranged from 4 to 35 with most items in the 20 to 30 range. (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 items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have higher inconsistency levels.

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

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 1974 AHS sample missed about 7 percent (i.e., about 600,000 units) of all conventional new construction (i.e., in both permit-issuing and non-permit-issuing areas) because the permits for these 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. New mobile homes in area segments, however, do come into the AHS sample. In addition, new mobile homes in address segments 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 address 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 1974 AHS sample missed about one-fourth of all new mobile homes (i.e., about 400,000 units). 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.

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: 1976 210-802/1147 1-3

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-38
Selection of sample areas App-38
Designation of sample housing units for the 1974 enumeration . App-38
Selection of the 1973 sample housing units App-38
Splitting of the sample App-39
Selection of supplemental sample housing units in rural areas App-39
ESTIMATION App-39
RELIABILITY OF THE ESTIMATES App-40
Sampling errors App-40
Illustrations of the use of the standard error tables App-41
Differences App-43
Illustration of the computation of the standard error of a difference App-44
Medians App-44
Illustration of the computation of the 95-percent confidence interval for a median App-45
Nonsampling errors App-45
1970 Census App-45

SAMPLE DESIGN

The estimates are based on data collected in August through October 1974 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 71,300 sample housing units (both occupied and vacant) were eligible for interview in the 1974 AHS. Of this number, 2,200 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 71,300, there were also 7,000 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 1974 enumeration.—The sample housing units designated to be interviewed in the 1974 enumeration consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1973 enumeration.

2. All 1973 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 the Type A and Type B noninterviews, see facsimile of 1974 AHS questionnaire, page 1.)

3. All sample housing units that were selected in the 1974 rural supplementation.

4. All sample housing units that were selected from the list of building permits issued since the 1973 enumeration. (This sample represents the housing units built since the 1973 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 a 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 accom-

plished 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 subdivided into areas with an expected size of four housing units.

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). 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 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.

ESTIMATION

The 1974 AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the Type A noninterview housing units encountered in the AHS. This noninterview adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

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

The first-stage ratio estimation procedure was employed for sample housing units from NSR PUS'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 six categories of new construction units (i.e., four categories for conventional new construction units and two for new construction mobile homes). This adjustment was necessary so as to correct for known deficiencies in the AHS sample with regard to represen-

tation of new construction units (see the section on nonsampling error).

The second-stage ratio estimate 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 for conventional new construction 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 4 types of vacant housing units and for 24 residence-tenure-race of head-sex of head categories for occupied housing units.

The third-stage ratio estimate factor for each specified catgory 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 estimate 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 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 measure 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.

Tables I-V present the standard errors applicable to the 1974 housing inventory estimates in this report, and table VI presents the standard errors applicable to 1973-1974 lost housing unit estimates in this report. Table I shows the approximate standard errors applicable to all 1974 housing inventory estimates except those pertaining to the specified items in the table, to urban or rural items, and to estimates of vacant housing units.

The standard errors shown in table II should be used for the specified items in table I, in tables III for urban and rural items, and in table V for estimates on vacant housing units. Linear interpolation should be used to determine standard errors for levels of estimates not specifically shown in tables I-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 numerators of the percentages, particularly if the percentages are 50 percent or more.

Tables VII-XI present the standard errors of estimated percentages. Table VII shows the approximate standard errors of all estimated percentages of the 1974 housing inventory except those pertaining to the specified items in table VIII. In addition to specified items, table VIII shows the approximate standard errors of urban vacant housing units and of total 1973-1974 lost housing units. Table IX shows the approximate standard errors of all estimated percentages of the 1974 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 1974 urban housing inventory except those pertaining to the specified items in table XI. Table XI also shows the approximate standard errors of urban 1973-1974 lost housing units. Two-way linear interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables VII-XI.

For ratios 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. A better approximation of the standard error of this type of ratio may be obtained by letting the standard eror of the ratio be approximately equal to:

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

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

Illustration of the use of the standard error tables. Illustration I. Table A-1 of this report shows that in urban areas of the United States there were 30,517,000

TABLE I. **Standard Errors of Estimated Number of Total Housing Units: 1974 (Excluding estimates of housing units pertaining to Cooking Fuel, Households 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: Negro (000)
5	3	3
10	4	4
25	6	6
50	8	8
100	12	12
250	19	19
500	27	26
1,000	38	35
2,500	59	49
5,000	82	48
10,000	112	–
25,000	156	–
50,000	160	–
75,000	61	–

TABLE II. **Standard Errors of Estimated Number of Total Housing Units Pertaining to Cooking Fuel, Households Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Source of Water, and Households With Head of Spanish Origin: 1974**

(68 chances out of 100)

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

TABLE III. **Standard Errors of Estimated Number of Urban Housing Units: 1974**

(68 chances out of 100)

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

TABLE IV. **Standard Errors of Estimated Number of Rural Housing Units: 1974**

(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

owner-occupied housing units in 1974. Interpolation in column 1 of table III of this appendix shows that the standard error of an estimate of this size is approximately 152,000. Consequently, the 68 percent confidence interval, as shown by these data, is from 30,365,000 to 30,669,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1974 owner-occupied housing units in urban areas 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 30,274,000 to 30,760,000 housing units with 90-percent confidence; and that the average estimate lies within the interval from 30,213,000 to 30,821,000 housing units with 95-percent confidence.

Table A-1 also shows that of the 30,517,000 owner-occupied housing units in urban areas, 9,221,000, or 30.2 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 29.9 to 30.5 percent, the 90-percent confidence interval is from 29.7 to 30.7 percent, and the 95-percent confidence interval is from 29.6 to 30.8 percent.

TABLE V. **Standard Errors of Estimated Number of Vacant Housing Units: 1974**

(68 chances out of 100)

Size of estimate	Standard error		
	Total	Urban	Rural
(000)	(000)	(000)	(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: 1973-1974**

(68 chances out of 100)

Size of estimate	Standard error		
	Total	Urban	Rural
(000)	(000)	(000)	(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	–	–

TABLE VII. **Standard Errors of Estimated Percentages of Housing Units: 1974 (Excluding estimated percentages of housing units pertaining to Cooking Fuel, Households Lacking Complete Plumbing Facilities, Mobile Homes, Sewage, Water Supply, Households With Head of Spanish Origin, and Vacant Units)**

(68 chances out of 100)

Base of percentage (000)	Estimated percentage						
	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.....	0.1	0.1	0.2	0.2	0.3	0.3	0.4
50,000.....	0.1	0.1	0.1	0.2	0.2	0.2	0.3
75,000.....	0.1	0.1	0.1	0.1	0.2	0.2	0.2

TABLE VIII. **Standard Errors of Estimated Percentages of Total Housing Units Pertaining to Cooking Fuel, Households Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Water Supply, Households With Head of Spanish Origin, or Urban Housing Units Pertaining to Vacants, and Lost Housing Units: 1973-1974**

(68 chances out of 100)

Base of percentage (000)	Estimated percentage						
	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.....	0.1	0.1	0.2	0.3	0.3	0.4	0.4
50,000.....	0.1	0.1	0.1	0.2	0.2	0.3	0.3

Illustration II. Table A-13 of this report shows that in the rural areas of the United States there were 170,000 owner-occupied housing units whose head was of Spanish origin in 1974. Interpolation in column 1 of table IV 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 156,000 to 184,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1974 owner-occupied housing units in rural areas whose head is of Spanish origin lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples lies within the interval from 148,000 to 192,000 housing units with 90-percent confidence; and that the average estimate lies within the interval 142,000 to 198,000 housing units with 95-percent confidence.

Table A-13 also shows that of the 170,000 owner-occupied housing units in rural areas whose head is of Spanish origin, 38,000 or 22.4 percent were occupied by two persons. Interpolation in table IX (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage is approximately 3.8 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 18.6 to 26.2 percent; the 90-percent confidence interval is from 16.3 to 28.5 percent; and the 95-percent confidence interval is from 14.8 to 30.0 percent.

Differences.—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard 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-1 of this report shows that in urban areas of the United States there were 5,410,000 owner-occupied housing units with three persons in 1974. Thus, the apparent difference between the number of 1974 owner-occupied housing units in urban areas with two persons and those with three persons is 3,811,000. Column 1 of table III shows that the standard error of 9,221,000 is approximately 111,000 and that the standard error of 5,410,000 is approximately 88,000. Therefore, the standard error of the estimated difference of 3,811,000 is about

$$142{,}000 = \sqrt{(111{,}000)^2 + (88{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 3,811,000 difference is from 3,669,000 to 3,953,000 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 3,584,000 to 4,038,000; and the 95-percent confidence interval is from 3,527,000 to 4,095,000. Thus, we can conclude with 95-percent confidence that the number of 1974 owner-occupied housing units inside SMSA's with two persons is greater than the number with three persons.

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

TABLE IX. **Standard Errors of Estimated Percentages of Rural Housing Units: 1974 (Excluding estimated percentages of rural housing units pertaining to Mobile Homes, Source of Water, Sewage Disposal, and 1973-74 Lost Units)**

(68 chances out of 100)

Base of percentage (000)	Estimated percentage						
	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	0.1	0.1	0.2	0.2	0.3	0.3	0.4

TABLE X. **Standard Errors of Estimated Percentages of Rural Housing Units Pertaining to Mobile Homes, Source of Water and Sewage Disposal; of Total Housing Units Pertaining to Vacants; and of Urban Housing Units: 1974 (Excluding Estimated Percentages of Urban Housing Units Pertaining to Mobile Homes, Households Lacking Complete Plumbing Facilities, Cooking Fuel, Sewage Disposal, and Households With Head of Spanish Origin)**

(68 chances out of 100)

Base of percentage (000)	Estimated percentage						
	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	0.1	0.1	0.2	0.2	0.3	0.3	0.4
50,000	0.1	0.1	0.1	0.2	0.2	0.2	0.3

TABLE XI. **Standard Errors of Estimated Percentages of Urban Housing Units Pertaining to Cooking Fuel, Mobile Homes, Sewage Disposal, Households Lacking Complete Plumbing Facilities, Households With Head of Spanish Origin, and Lost Housing Units: 1973-1974**

(68 chances out of 100)

Base of percentage (000)	Estimated percentage						
	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.....	0.1	0.1	0.2	0.3	0.3	0.4	0.5
50,000.....	0.1	0.1	0.1	0.2	0.2	0.3	0.3

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 2.9 in 1974. The base of the distribution from which this median was determined is 30,517,000.

1. From table X, the standard error of a 50 percent characteristic on the base of 30,517,000 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 13,215,000 housing units, or 43.3 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 5,410,000 owner-occupied housing units or 17.7 percent, had three persons (i.e., 2.5 to 3.5 persons). By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$2.5 + (3.5 - 2.5)\left(\frac{49.2 - 43.3}{17.7}\right) = 2.8$$

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

$$2.5 + (3.5 - 2.5)\left(\frac{50.8 - 43.3}{17.7}\right) = 2.9$$

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 upper limit, it actually is a reflection of the rounding error associated with this median (see the paragraph on rounding error in the nonsampling 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 1974 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 in the following:

1. Space errors.—Errors in which both the living quarters and its occupants were missed in the census or in which they were counted more than once (overenumerated). Space errors are usually the largest component of housing coverage error.

2. Definitional errors.—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home and consequently receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one living quarters would have been counted where two existed.

3. Occupancy errors.—Errors of incorrect occupancy classification for enumerated units, i.e., vacant units that were improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on the census questionnaire. In these studies, content errors were measured by the following methods:

1. Reinterviews.—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.

2. Record checks.—The comparison of census data with data obtained from an independent record source.

3. Other surveys.—Comparisons of census data with that obtained from other sample 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.** Some of the results are:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent (after processing). . . . For each 100 units that were finally enumerated in the census, an estimated 2.5 were missed. . . ." (PHC(E)-5, p.3)

2. "The occupied space missed rate for the total United States in 1970 is estimated at . . . 1.4 percent (after processing). . . .About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures." (PHC(E)-5, p.4)

3. "In 1970, the definitional underenumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error rate of 0.34 of 1 percent." (PHC(E)-5, p.11)

4. "Most of the vacant units that were enumerated as occupied were procedural errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p.16) "Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant." (PHC(E)-5, p.15)

5. "Generally, owners report housing data more consistently than renters, responses for occupied units are more consistent than those for vacant units, and respondents in single-unit structures report more consistently than those in multiunit structures." (PHC(E)-10, p.6) This statement applies to the following items:

a. Heating fuel
b. Renters paying extra for utilities and/or fuel
c. Bathtub or shower facilities
d. Flush toilet facilities
e. Telephone availability
f. Year structure built
g. Value of home
h. Seasonal vacancy status

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 reinterview study are presented in the Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample: 1974." Twenty-five items were included in the reinterview, some for the first time. The estimated indexes of inconsistency of these items ranged from 4 to 35 with most items in the 20 to 30 range. (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 items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have higher inconsistency levels.

The 1970 census reinterview results provided illustrations of possible nonsampling errors for some of the items which also appear in the AHS. For

example, median value of homes was consistently underestimated by about 5 percent, and the average monthly costs of electricity and utility gas were consistently overestimated although the 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.

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 1974 AHS sample missed about 7 percent (i.e., about 600,000 units) of all conventional new construction (i.e., in both permit-issuing and non-permit-issuing areas) because the permits for these 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. New mobile homes in area segments, however, do come into the AHS sample. In addition, new mobile homes in address segments 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 address 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 1974 AHS sample missed about one-fourth of all new mobile homes (i.e., about 400,000 units). 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.

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.

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-28
Selection of sample areas App-28
Designation of sample housing units for the 1974 enumeration . App-28
Selection of the 1973 sample housing units App-28
Splitting of the sample App-29
Selection of supplemental sample housing units in rural areas App-29
ESTIMATION..... App-29
RELIABILITY OF THE ESTIMATES App-30
Sampling errors App-30
Illustrations of the use of the standard error tables App-31
Differences App-32
Illustration of the computation of the standard error of a difference App-33
Medians App-33
Illustration of the computation of the 95-percent confidence interval of the median App-33
Nonsampling errors App-33
Reinterview program App-33
Coverage errors App-34
Rounding errors App-34

SAMPLE DESIGN

The estimates are based on data collected in August through October 1974 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 71,300 sample housing units (both occupied and vacant) were eligible for interview in the 1974 AHS. Of this number, 2,200 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 71,300, there were also 7,000 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 independently of the other PSU selected from this stratum. Since the two PSU's were independently selected, it was possible for the same PSU to be selected twice. This occurred in 25 instances, producing an additional 85 NSR sample PSU's.

Designation of sample housing units for the 1974 enumeration.—The sample housing units designated to be interviewed in the 1974 enumeration consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1973 enumeration.
2. All 1973 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 noninterviews, see facsimile of 1974 AHS questionnaire, page 2.)
3. All sample housing units that were selected in the 1974 rural supplementation.
4. All sample housing units that were selected from the list of building permits issued since the 1973 enumeration. (This sample represents the housing units built since the 1973 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 a 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 subdivided into areas with an expected size of four housing units.

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). 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 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.

ESTIMATION

The 1974 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 estimate 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 six categories of new construction units (i.e., four categories for conventional new construction units and two 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 so as 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 estimate 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 for conventional new construction units 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 estimate factor was then applied to the existing weight for each sample unit in each second-stage ratio estimation category.

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

The third-stage ratio estimate 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 estimate 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 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 stand-

ard 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 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
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

Illustrations of the use of the standard error tables.—Illustration I.—Table 1 of this report shows that in the United States there were 20,242,000 owner-occupied housing units in 1974 with garbage collection service once a week. Interpolation in table I of this appendix shows that the standard error on an estimate of this size is approximately 142,000. Consequently, the 68-percent

TABLE I. Standard Errors of Estimated Number of Housing Units: 1974 (Excluding estimates of housing units pertaining to Cooking Fuel, Households 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: Negro (000)
5	3	3
10	4	4
25	6	6
50	8	8
100	12	12
250	19	19
500	27	26
1,000	38	35
2,500	59	49
5,000	82	48
10,000	112	–
25,000	156	–
50,000	160	–
75,000	61	–

TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to Cooking Fuel, Households Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Source of Water, and Households with Head of Spanish Origin: 1974

(68 chances out of 100)

Size of estimate (000)	Standard Error: Total or white (000)	Standard Error: Negro (000)
5	3	3
10	4	4
25	7	7
50	10	10
100	14	14
250	22	22
500	31	30
1,000	44	41
2,500	69	57
5,000	96	56
10,000	131	–
25,000	183	–
50,000	187	–
75,000	71	–

TABLE III. Standard Errors of Estimated Percentages of Housing Units: 1974 (Excluding estimated percentages of housing units pertaining to Cooking Fuel, Households 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 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	0.1	0.1	0.2	0.2	0.3	0.3	0.4
50,000	0.1	0.1	0.1	0.2	0.2	0.2	0.3
75,000	0.0	0.1	0.1	0.1	0.2	0.2	0.2

TABLE IV. Standard Errors of Estimated Percentages of Housing Units Pertaining to Cooking Fuel, Households Lacking Complete Plumbing Facilities, Mobile Homes, Sewage Disposal, Source of Water, and Households with Head of Spanish Origin: 1974

(68 chances out of 100)

Base of percentage (000)	Estimated percentage						
	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	0.1	0.1	0.2	0.3	0.3	0.4	0.4
50,000	0.1	0.1	0.1	0.2	0.2	0.3	0.3

confidence interval as shown by these data is from 20,100,000 to 20,384,000. Therefore, a conclusion that the average estimate (derived from all possible samples) of 1974 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 20,015,000 to 20,469,000 housing units with 90 percent confidence and that the average estimate lies within the interval from 19,958,000 to 20,526,000 housing units with 95 percent confidence.

Table 1 also shows that of the 20,242,000, 1974 owner-occupied housing units with garbage collection service once a week, 1,482,000, or 7.3 percent, had a family or primary individual income between $5,000 and $6,999. Interpolation in table III (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage is approximately .2 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 7.1 to 7.5 percent; the 90-percent confidence interval is from 7.0 to 7.6 percent; and the 95-percent confidence interval is from 6.9 to 7.7 percent.

Illustration II.—Table 25 of this report shows that in the United States there were 450,000 owner-occupied housing units with head of Spanish origin in 1974 that had garbage collection service once a week. Interpolation in table II of this appendix shows that the standard error on an estimate of this size is approximately 29,000 (table II should be used since this estimate pertains to households with Spanish head). Consequently, the 68-percent confidence interval as shown by these data is from 421,000 to 479,000. Therefore, a conclusion that the average estimate (derived from all possible samples) of 1974 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 404,000 to 496,000 housing units with 90 percent confidence and that the average estimate lies within the interval from 392,000 to 508,000 housing units with 95 percent confidence.

Table 25 also shows that of the 450,000, 1974 owner-occupied housing units with head of Spanish origin that had garbage collection service once a week, 30,000, or 6.7 percent, had a family or primary individual income 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 5.0 to 8.4 percent; the 90-percent confidence interval is from 4.0 to 9.4 percent; and the 95-percent confidence interval is from 3.3 to 10.1 percent.

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

accurate for the difference between estimates of the same characteristic in two different areas or the difference between separate and uncorrelated characteristics in the same area. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

Illustration of the computation of the standard error of a difference.—Table 1 of this report shows that in the United States there were 4,873,000 owner-occupied housing units in 1974 with garbage collection service once a week that had a family or primary individual income 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 or primary individual income between $5,000 and $6,999 and the number that had a family or primary individual income between $10,000 and $14,999 is 3,391,000. Table I shows that the standard error on an estimate of 1,482,000 to be approximately 45,000 and table I also shows that the standard error on an estimate of 4,873,000 to be approximately 81,000. Therefore, the standard error of the estimated difference of 3,391,000 is about

$$93{,}000 = \sqrt{(45{,}000)^2 + (81{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 3,391,000 difference is from 3,298,000 to 3,484,000 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 95-percent confidence interval is from 3,205,000 to 3,577,000 housing units, and, thus, we can conclude with 95 percent confidence that of the 1974 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 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 of the median.—Table 1 of this report shows that the median family or primary individual income of owner-occupied housing units in the United States with garbage collection service once a week was $13,700 in 1974. The base of the distribution, from which this median was determined, is 20,242,000 housing units.

1. Table III shows that the standard error of 50 percent on a base of 20,242,000 is .5 percentage points.

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 49.0 and 51.0.

3. From table 1, it can be seen by cumulating the frequencies for the first four categories that 6,548,000 housing units, or 32.3 percent, had a family or primary individual income up to $9,999 and that an additional 4,873,000 or 24.1 percent, had a family or primary individual income between $10,000 and $14,999. By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$\$9{,}999 + (\$15{,}000 - \$10{,}000)\left(\frac{49.0 - 32.3}{24.1}\right) = \$13{,}500$$

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

$$\$9{,}999 + (\$15{,}000 - \$10{,}000)\left(\frac{51.0 - 32.3}{24.1}\right) = \$13{,}900$$

Thus, the 95-percent confidence interval ranges from $13,500 to $13,900.

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 1974 AHS-National sample, a study was conducted to obtain a measurement of some of the nonsampling errors 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 two observations (i.e., the original interview and the reinterview) were assumed to be two independent readings and thus were the basis

for the measurement of the accuracy of the AHS 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 reinterview study are presented in the Census Bureau memorandum, "Reinterview Results for the Annual Housing Survey—National Sample: 1974." Twenty-five items were included in the reinterview, some for the first time. The estimated indexes of inconsistency of these items ranged from 4 to 35 with most items in the 20 to 30 range. (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 items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have higher inconsistency levels.

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

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 presentation 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 1974 AHS sample misses about 6 percent (i.e., about 600,000 units) of all conventional new construction (i.e., in both permit-issuing and non-permit-issuing areas) because the permits for these 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. New mobile homes in area segments, however, do come into the AHS sample. In addition, new mobile homes in address segments 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 address 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 1974 AHS sample missed about one-fourth of all new mobile homes (i.e., about 400,000 units). 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.

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: 1976 240-869/1006 1-3