# APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN
Annual Housing Survey-National sample ..... App-31
Selection of sample areas ..... App-31
Selection of sample housing units App-31
Splitting of the sample ..... App-32
1970 Census of Population and Housing ..... App-32
ESTIMATION
AHS-National sample ..... App-32
1973 housing inventory ..... App-32
1970-1973 lost units ..... App-33
Ratio estimation procedure of the 1970 Census of Population and Housing ..... App-33
Reliability of the estimates ..... App-33
Sampling errors for AHS-National sample ..... App-33
Illustration of the use of standard error tables ..... App-34
Differences ..... App-35
Illustration of the computation of the standard error of a difference App-35
Medians ..... App-35
Illustration of the computation of the standard error of a median .. App-35
Nonsampling errors ..... App-35
1970 census ..... App-36
Reinterview program ..... App-36

## SAMPLE DESIGN

**Annual Housing Survey-National sample.**—The estimates for 1973 are based on data collected in August through December 1973 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 53,800 sample housing units (both occupied and vacant) were eligible for interview in the 1973 AHS. Of this number, 1,500 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 53,800, there were also 5,500 sample units which were visited but found not to provide information relevant to the 1973 housing inventory. Of this number, however, about 1,800 units were identified as being in whole structures removed from the housing inventory since 1970. This sample was used to obtain estimates of units in whole structures removed from the housing inventory between 1970 and 1973. The estimates pertain to lost units in whole structure losses only (i.e., structures in which all housing units were removed from the inventory); not obtained were estimates of lost units in part structure losses (i.e., structures in which some, but not all, of the units were removed from the 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 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 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.

**Selection of sample housing units.**—The overall sampling rate used in 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 subdivided into segments, i.e., small land areas with well-defined boundaries having an expected size of about four housing units.

The sample of new construction units was selected from building permits issued since the time of the 1970 census. Within each sample PSU, the building permits were chronologically ordered by month issued, and clusters of housing units with an expected size of four were created. These clusters were then sampled for inclusion in AHS at the sampling rate of 2 in 1,366. As a result of the area sampling methods mentioned 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.

**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 employed for the 1970 census 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 1973 housing inventory and estimates of units in whole structures removed from the housing inventory between 1970 and 1973. Each type of estimate employed a separate, though similar, estimation procedure as described below.

**1973 housing inventorv.**– The 1973 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 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. This procedure was designed to reduce the contribution to the variance arising from the sampling of PSU's. The first-stage ratio estimation procedure takes into account the differences that existed at the time of the 1970 census in the distribution by tenure and residence of the housing population estimated from the sample NSR PSU's and that of the NSR housing population in each of the four census regions of the country.

The first-stage ratio 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 three categories of new construction units (i.e., two categories for conventional new construction units and one for new construction mobile homes). These independent estimates were considered to be the best estimates available for current estimates 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.

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 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 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 four types of vacant housing units and for 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.

**1970-1973 lost units.**—Statistics pertaining to units in whole structures removed from the housing inventory between 1970 and 1973 employed a one-stage ratio estimation procedure similar to the first-stage ratio estimation procedure employed in AHS for the 1973 housing inventory statistics. In fact, the same categories and factors were utilized for this estimation procedure. The computed first-stage ratio estimate factor for each residence-tenure category was applied to the weight for each NSR PSU sample unit, in a whole structure removed from the inventory, in that residence-tenure category.

**Ratio estimation procedure of the 1970 Census of Population and Housing.**—This report presents data on the housing characteristics of the 1970 housing inventory from 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 in **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 "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. 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 is in **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 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 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, the average result would be included in the interval from one standard error below and above the sample estimate for about 68 out of 100 possible samples. For about 90 out of 100 possible samples, the interval from 1.6 standard errors below and above the sample estimate would include the average result; for about 95 out of 100 possible samples, the interval from two standard errors below and above the sample estimate would include the average result.

All statements of comparison appearing in this report are significant at the 1.6 standard error level or higher, and most are significant at a level of more than 2.0 standard errors. This means that for most differences cited in the text, the estimated difference is greater than twice the standard error of the difference. Statements of comparison qualified in some way (e.g., by the use of the phrase, "some evidence") have a level of significance between 1.6 and 2.0 standard errors.

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 estimates from the 1973 Annual Housing Survey. Table I shows the approximate standard errors applicable to all 1973 estimates of housing units except those pertaining to selected items in the tables. The standard errors shown in table II should be used for those selected 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: 1973 (Excludes estimates of Housing Units pertaining to Basement, Source of Water, Sewage Disposal, Households With Negro Head, or Households With Head of Spanish Origin)**

(68 chances out of 100)

| Size of estimate (000) | Standard error (000) | Size of estimate (000) | Standard error (000) |
|---|---|---|---|
| 25 | 5 | 2,500 | 50 |
| 50 | 7 | 5,000 | 70 |
| 100 | 10 | 10,000 | 96 |
| 250 | 16 | 25,000 | 134 |
| 500 | 23 | 50,000 | 138 |
| 1,000 | 32 | | |

TABLE II. **Standard Errors of Estimated Number of Housing Units Pertaining to Basement, Source of Water, Sewage Disposal, Households With Negro Head, or Households With Head of Spanish Origin: 1973**

(68 chances out of 100)

| Size of Estimate (000) | Standard Error (000) | Size of Estimate (000) | Standard Error (000) |
|---|---|---|---|
| 25 | 6 | 1,000 | 40 |
| 50 | 9 | 2,500 | 63 |
| 100 | 13 | 5,000 | 87 |
| 250 | 20 | 10,000 | 119 |
| 500 | 28 | 25,000 | 165 |

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. An approximation to the standard error of a percentage may be obtained by using the following formula:

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

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

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

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula underestimates 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 changing the sign in the formula from minus (-) to plus (+).

**Illustration of the use of standard error tables.**—Table A-1 of this report shows that in this United States in 1973, there were 8,280,000 year-round housing units with three rooms. Interpolation in table I above shows that the standard error of an estimate of this size is approximately 87,000. Consequently, the 68-percent confidence interval, as shown by these data, is from 8,193,000 to 8,367,000 housing units. Therefore, a conclusion that the average estimate of 1973 year-round housing units with three rooms 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 of 1973 year-round housing units with three rooms, derived from all possible samples, lies within the interval from 8,141,000 to 8,419,000 housing units with 90 percent confidence. We could also conclude that the average estimate of 1973 year-round housing units with three rooms lies within the interval from 8,106,000 to 8,454,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 8,280,000 year-round housing units with three rooms, 6,166,000, or 74.5 percent, were located inside SMSA's in 1973. Table I above shows that the standard error for 6,166,000 is approximately 76,000. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph following table II above), the standard error of the 74.5 percent is approximately .5 percentage points:

$$.5 = (100)\left(\frac{6,166,000}{8,280,000}\right)\sqrt{\left(\frac{76,000}{6,166,000}\right)^2 - \left(\frac{87,000}{8,280,000}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 74.0 to 75.0 percent; the 90-percent confidence interval is from 73.7 to 75.3 percent; and the 95-percent confidence interval is from 73.5 to 75.5 percent.

**Differences.**—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same 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 1973, there were 15,621,000 year-round housing units with four rooms in the United States. Thus, the apparent difference between the number of 1973 year-round housing units in the United States with three rooms and with four rooms is 7,341,000. The standard error of 8,280,000 is 87,000 as is shown above. Table I shows that the standard error on an estimate of 15,621,000 is approximately 110,000. Therefore, the standard error of the estimated difference of 7,341,000 is about

$$140,000 = \sqrt{(87,000)^2 + (110,000)^2}$$

Consequently, the 68-percent confidence interval for the 7,341,000 difference is from 7,201,000 to 7,481,000 housing units. Therefore, a conclusion that the average estimate of the 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 7,061,000 to 7,621,000 housing units, and thus we can conclude with 95 percent confidence that the number of 1973 year-round housing units in the United States with four rooms is actually greater than the number of 1973 year-round housing units in the United States with three rooms.

**Medians.**—For the medians presented in certain tables, the sampling error depends on the size of the base and 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 tables I and II 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 rooms in year-round housing units inside central cities of SMSA's was 4.7 in 1973. The base of the distribution, from which this median was determined, is 24,099,000 housing units.

1. Table I in conjunction with the formula for the standard error of a percentage shows that the standard error of 50 percent on a base of 24,099,000 is .3 percentage points:

$$.3 = (100)\left(\frac{12,049,500}{24,099,000}\right)\sqrt{\left(\frac{101,000}{12,049,500}\right)^2 - \left(\frac{132,000}{24,099,000}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 49.4 and 50.6.

3. From table A-1, it can be seen by cumulating the frequencies for the first three categories that 10,920,000 housing units, or 45.3 percent, had up to four rooms (actually, the category of four rooms is considered to be from 3.5 to 4.5 rooms) and that an additional 5,410,000 housing units, or 22.4 percent, had five rooms (i.e., 4.5 to 5.5 rooms). By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$4.5 + (5.5-4.5)\left(\frac{49.4-45.3}{22.4}\right) = 4.68$$

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

$$4.5 + (5.5-4.5)\left(\frac{50.6-45.3}{22.4}\right) = 4.74$$

Thus, the 95-percent confidence interval ranges from 4.68 to 4.74 rooms.

**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 ques-

tions, 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 1973 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.*—Comparison 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

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 1973 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 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; i.e., tolerance limits were derived to determine which interviewer passed or failed this reinterview with regard to the above items.

The results of the reinterview study are presented in the following internal memorandum, **Reinterview Results for the Annual Housing Survey-National Sample: 1973.** Some of these results are:

1. "Overall, the results showed a fairly consistent response between the original interview and the reinterview."

2. "For most 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 this study 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.

With respect to errors of coverage and estimation for missing data, 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 (see section on estimation). 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. This assumption, however, has been found to be invalid since there was an estimated 600,000 units with permits issued before 1970 but which were completed after the 1970 census. Therefore, these conventional new construction units had no representation in the AHS sample of conventional new construction. Secondly, unlike the procedure for conventional new construction, there is no sampling procedure connected with the representation of 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 AHS sample misses about one-half of all new mobile homes (i.e., about 700,000 units). The second-stage ratio estimation was employed to correct both these deficiencies.

☆ U. S. GOVERNMENT PRINTING OFFICE : 1975—210-802/1010

# APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN ..... App-28
Selection of sample areas ..... App-28
Selection of sample housing units .. App-28
Splitting of the sample ..... App-28
ESTIMATION ..... App-29
RELIABILITY OF THE ESTIMATES App-30
Sampling errors ..... App-30
Illustration of the use of the standard error tables ..... App-31
Differences ..... App-31
Illustration of the computation of the standard error of a difference . App-31
Nonsampling errors ..... App-32

## SAMPLE DESIGN

The estimates are based on data collected in August through December 1973 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 53,800 sample housing units (both occupied and vacant) were eligible for interview in the 1973 AHS. Of this number, 1,500 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 53,800, there were also 5,500 sample units which were visited but found not to provide information relevant to the 1973 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 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 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.

**Selection of sample housing units.**—The overall sampling rate used in 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 subdivided into segments, i.e., small land areas with well-defined boundaries having an expected size of about four housing units.

The sample of new construction units was selected from building permits issued since the time of the 1970 census. Within each sample PSU, the building permits were chronologically ordered by month issued, and clusters of housing units with an expected size of four were created. These clusters were then sampled for inclusion in AHS at the sampling rate of 2 in 1,366. As a result of the area sampling methods mentioned 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, how-

ever, 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.

## ESTIMATION

The 1973 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 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. This procedure was designed to reduce the contribution to the variance arising from the sampling of PSU's. The first-stage ratio estimation procedure takes into account the differences that existed at the time of the 1970 census in the distribution by tenure and residence of the housing population estimated from the sample NSR PSU's and that of the NSR housing population in each of the four census regions of the country.

The first-stage ratio 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 three categories of new construction units (i.e., two categories for conventional new construction units and one for new construction mobile homes). These independent estimates were considered to be the best estimates available for current estimates 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.

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 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 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 four types of vacant housing units and for 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-stage 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 "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. 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, the average result would be included in the interval from one standard error below and above the sample estimate for about 68 out of 100 possible samples. For about 90 out of 100 possible samples, the interval from 1.6 standard errors below and above the sample estimate would include the average result; and for about 95 out of 100 possible samples, the interval from two standard errors below and above the sample estimate would include the average result.

All statements of comparison appearing in this report are significant at the 1.6 standard error level or higher, and most are significant at a level of more than 2.0 standard errors. This means that for most differences cited in the text, the estimated difference is greater than twice the standard error of the difference. Statements of comparison qualified in some way (e.g., by the use of the phrase, "some evidence") have a level of significance between 1.6 and 2.0 standard errors.

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 in this report except those pertaining to selected items in the tables. The standard errors shown in table II should be used for those selected 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: 1973 (Excludes estimates of housing units pertaining to Garbage Collection Service, Exterminator Service, Common Stairways, Light Fixtures in Public Halls, Roof, Basement, Water Supply, Sewage Disposal, Flush Toilet, Heating Equipment, Selected Neighborhood Characteristics, Households With Negro Head, or Households With Head of Spanish Origin)**

**(68 chances out of 100)**

| Size of estimate (000) | Standard error (000) | Size of estimate (000) | Standard error (000) |
|---|---|---|---|
| 25 | 5 | 2,500 | 50 |
| 50 | 7 | 5,000 | 70 |
| 100 | 10 | 10,000 | 96 |
| 250 | 16 | 25,000 | 134 |
| 500 | 23 | 50,000 | 138 |
| 1,000 | 32 | | |

TABLE II. **Standard Errors of Estimated Number of Housing Units Pertaining to Garbage Collection Service, Exterminator Service, Common Stairways, Light Fixtures in Public Halls, Roof, Basement, Water Supply, Sewage Disposal, Flush Toilet, Heating Equipment, Selected Neighborhood Characteristics, Households With Negro Head, or Households With Head of Spanish Origin: 1973**

**(68 chances out of 100)**

| Size of estimate (000) | Standard error (000) | Size of estimate (000) | Standard error (000) |
|---|---|---|---|
| 25 | 6 | 1,000 | 40 |
| 50 | 9 | 2,500 | 63 |
| 100 | 13 | 5,000 | 87 |
| 250 | 20 | 10,000 | 119 |
| 500 | 28 | 25,000 | 165 |

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. An approximation to the standard error of a percentage may be obtained by using the following formula:

Let $x$ = the numerator

$y$ = the denominator

$\sigma_x$ = the standard error of the numerator

$\sigma_y$ = the standard error of the denominator

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

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

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula underestimates 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 changing the sign in the formula from minus (-) to plus (+).

*Illustration of the use of the standard error tables.*—Table A-3 of this report shows that in the United States there were 1,073,000 owner-occupied housing units (occupied three months or longer) in 1973 with water supplied from a public system or a private company or an individual well that had a breakdown in the water supply. Interpolation in table II above shows that the standard error of an estimate of this size is approximately 41,000 (table II should be used since this estimate pertains to water supply). Consequently, the 68-percent confidence interval, as shown by these data, is from 1,032,000 to 1,114,000 housing units. Therefore, a conclusion that the average estimate of 1973 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 1,007,000 to 1,139,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 991,000 to 1,155,000 housing units with 95 percent confidence.

Table A-3 also shows that of the 1,073,000 owner-occupied housing units (occupied three months or longer) with water from a public system or a private company or an individual well that had a breakdown in the water supply, 115,000 or 10.7 percent were located in central cities of SMSA's. Table II above shows that the standard error for 115,000 is approximately 14,000. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph following table II above), the standard error of the 10.7 percent is approximately 1.2 percentage points:

$$1.2 = (100)\left(\frac{115{,}000}{1{,}073{,}000}\right)\sqrt{\left(\frac{14{,}000}{115{,}000}\right)^2 - \left(\frac{41{,}000}{1{,}073{,}000}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 9.5 to 11.9 percent; the 90-percent confidence interval is from 8.8 to 12.6 percent; and the 95-percent confidence interval is from 8.3 to 13.1 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 standard error.

*Illustration of the computation of the standard error of a difference.*—Table A-3 of this report shows that in 1973, there were 237,000 renter-occupied housing units (occupied three months or longer) in central cities of the SMSA's in the United States with water supplied from a public system or a private company or an individual well that had a breakdown in the water supply. Thus, the apparent difference between the number of owner-occupied housing units and the number of renter-occupied housing units (occupied three months or longer) in central cities of the SMSA's in the United States with water supplied from a public system or a private company or an individual well that had a breakdown in the water supply is 122,000. The standard error of 115,000 is 14,000 as is shown above. Table II shows that the standard error on an estimate of 237,000 is approximately 19,000. Therefore, the standard error of the estimated difference of 122,000 is about

$$24{,}000 = \sqrt{(14{,}000)^2 + (19{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 122,000 difference is from 98,000 to 146,000 housing units. Therefore, a conclusion that the average

estimate of the 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 74,000 to 170,000 housing units, and thus we can conclude with 95 percent confidence that the number of renter-occupied housing units (occupied three months or longer) in central cities of the SMSA's in the United States with water supplied from a public system or a private company or an individual well that had a breakdown in the water supply is greater than the number of owner-occupied housing units of this type in 1973.

**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 1973 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 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 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; i.e., tolerance limits were derived to determine which interviewer passed or failed this reinterview with regard to the above items.

The results of the reinterview study are presented in the following internal memorandum: **Reinterview Results for the Annual Housing Survey-National Sample: 1973.** Some of these results are:

1. "Overall, the results showed a fairly consistent response between the original interview and the reinterview."
2. "For most 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 this study 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.

With respect to errors of coverage and estimation for missing data, 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 (see section on estimation). 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. This assumption, however, has been found to be invalid since there was an estimated 600,000 units with permits issued before 1970 but which were completed after the 1970 census. Therefore, these conventional new construction units had no representation in the AHS sample of conventional new construction. Secondly, unlike the procedure for conventional new construction, there is no sampling procedure connected with the representation of 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 AHS sample misses about one-half of all new mobile homes (i.e., about 700,000 units). The second-stage ratio estimation was employed to correct both these deficiencies.

# APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN . . . . . App-30
Selection of sample areas. . . . . App-30
Selection of sample housing units. App-30
Splitting of the sample . . . . . App-30
ESTIMATION. . . . . App-31
RELIABILITY OF THE ESTIMATES. . . . . App-32
Sampling errors . . . . . App-32
Illustrations 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-34
Illustration of the computation of the standard error of a median . . . . . App-34
Nonsampling errors. . . . . App-35

## SAMPLE DESIGN

The estimates are based on data collected in August through December 1973 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 53,800 sample housing units (both occupied and vacant) were eligible for interview in the 1973 AHS. Of this number, 1,500 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 53,800, there were also 5,500 sample units which were visited but found not to provide information relevant to the 1973 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 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 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.

**Selection of sample housing units.**—The overall sampling rate used in 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 subdivided into segments, i.e., small land areas with well-defined boundaries having an expected size of about four housing units.

The sample of new construction units was selected from building permits issued since the time of the 1970 census. Within each sample PSU, the building permits were chronologically ordered by month issued, and clusters of housing units with an expected size of four were created. These clusters were then sampled for inclusion in AHS at the sampling rate of 2 in 1,366. As a result of the area sampling methods mentioned 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

# APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN ..... App-28
Selection of sample areas ..... App-28
Selection of sample housing units .. App-28
Splitting of the sample ..... App-28
ESTIMATION ..... App-29
RELIABILITY OF THE ESTIMATES App-30
Sampling errors ..... App-30
Illustration of the use of the standard error tables ..... App-31
Differences ..... App-31
Illustration of the computation of the standard error of a difference . App-31
Medians ..... App-31
Illustration of the computation of the standard error of a median ... App-32
Nonsampling errors ..... App-32

## SAMPLE DESIGN

The estimates are based on data collected in August through December 1973 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 53,800 sample housing units (both occupied and vacant) were eligible for interview in the 1973 AHS. Of this number, 1,500 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 53,800, there were also 5,500 sample units which were visited but found not to provide information relevant to the 1973 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 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 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.

**Selection of sample housing units.**—The overall sampling rate used in 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 subdivided into segments, i.e., small land areas with well-defined boundaries having an expected size of about four housing units.

The sample of new construction units was selected from building permits issued since the time of the 1970 census. Within each sample PSU, the building permits were chronologically ordered by month issued, and clusters of housing units with an expected size of four were created. These clusters were then sampled for inclusion in AHS at the sampling rate of 2 in 1,366. As a result of the area sampling methods mentioned 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.

## ESTIMATION

The 1973 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 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:

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

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

The first-stage ratio 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 three categories of new construction units (i.e., two categories for conventional new construction units and one for new construction mobile homes). These independent estimates were considered to be the best estimates available for current estimates 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.

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 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 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 four types of vacant housing units and for 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 "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. 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, the average result would be included in the interval from one standard error below and above the sample estimate for about 68 out of 100 possible samples. For about 90 out of 100 possible samples, the interval from 1.6 standard errors below and above the sample estimate would include the average result; and for about 95 out of 100 possible samples, the interval from two standard errors below and above the sample estimate would include the average result.

All statements of comparison appearing in this report are significant at the 1.6 standard error level or higher, and most are significant at a level of more than 2.0 standard errors. This means that for most differences cited in the text, the estimated difference is greater than twice the standard error of the difference. Statements of comparison qualified in some way (e.g., by the use of the phrase, "some evidence") have a level of significance between 1.6 and 2.0 standard errors.

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 in this report except those pertaining to selected items in the tables. The standard errors shown in table II should be used for those selected 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: 1973 (Excludes estimates of housing units pertaining to Mortgage on Property, Parking Facilities, Garbage and Trash Collection Service, Public or Subsidized Housing, Basement, Public Sewer, Households With Negro Head, or Households With Head of Spanish Origin)**

(68 chances out of 100)

| Size of estimate (000) | Standard error (000) | Size of estimate (000) | Standard error (000) |
|---|---|---|---|
| 25 | 5 | 2,500 | 50 |
| 50 | 7 | 5,000 | 70 |
| 100 | 10 | 10,000 | 96 |
| 250 | 16 | 25,000 | 134 |
| 500 | 23 | 50,000 | 138 |
| 1,000 | 32 | | |

TABLE II. **Standard Errors of Estimated Number of Housing Units Pertaining to Mortgage on Property, Parking Facilities, Garbage and Trash Collection Service, Public or Subsidized Housing, Basement, Public Sewer, Households With Negro Head, or Households With Head of Spanish Origin: 1973**

(68 chances out of 100)

| Size of estimate (000) | Standard error (000) | Size of estimate (000) | Standard error (000) |
|---|---|---|---|
| 25 | 6 | 1,000 | 40 |
| 50 | 9 | 2,500 | 63 |
| 100 | 13 | 5,000 | 87 |
| 250 | 20 | 10,000 | 119 |
| 500 | 28 | 25,000 | 165 |

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. An approximation to the standard error of a percentage may be obtained by using the following formula:

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

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

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

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula underestimates 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 changing the sign in the formula from minus (–) to plus (+).

*Illustration of the use of the standard error tables.*—Table A-1 of this report shows that in the United States there were 4,392,000 owner-occupied housing units occupied by recent movers in 1973. Interpolation in table I above shows that the standard error of an estimate of this size is approximately 65,000. Consequently, the 68-percent confidence interval, as shown by these data, is from 4,327,000 to 4,457,000 housing units. Therefore, a conclusion that the average estimate of 1973 owner-occupied housing units occupied by recent movers lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate of 1973 owner-occupied housing units occupied by recent movers, derived from all possible samples, lies within the interval from 4,288,000 to 4,496,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 4,262,000 to 4,522,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 4,392,000 owner-occupied housing units occupied by recent movers, 394,000 or 9.0 percent had six persons or more. Table I above shows that the standard error for 394,000 is approximately 20,000. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph following table II above), the standard error of the 9.0 percent is approximately .4 percentage points:

$$.4 = (100)\left(\frac{394{,}000}{4{,}392{,}000}\right)\sqrt{\left(\frac{20{,}000}{394{,}000}\right)^2 - \left(\frac{65{,}000}{4{,}392{,}000}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 8.6 to 9.4 percent; the 90-percent confidence interval is from 8.4 to 9.6 percent; and the 95-percent confidence interval is from 8.2 to 9.8 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 501,000 owner-occupied housing units occupied by recent movers with five persons in 1973. 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 107,000. The standard error of 394,000 is 20,000 as is shown above. Table I shows that the standard error on an estimate of 501,000 is approximately 23,000. Therefore, the standard error of the estimated difference of 107,000 is about

$$30{,}000 = \sqrt{(20{,}000)^2 + (23{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 107,000 difference is from 77,000 to 137,000 housing units. Therefore, a conclusion that the average estimate of the 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 47,000 to 167,000 housing units, and thus we can conclude with 95 percent confidence that the number of 1973 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.

*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 tables I and II 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 3.1 in 1973. The base of the distribution, from which this median was determined, is 4,392,000 housing units.

1. Table I in conjunction with the formula for the standard error of a percentage shows that the standard error of 50 percent on a base of 4,392,000 is .7 percentage points:

$$.7 = (100)\left(\frac{2,196,000}{4,392,000}\right)\sqrt{\left(\frac{46,000}{2,196,000}\right)^2 - \left(\frac{65,000}{4,392,000}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 48.6 to 51.4.
3. From table A-1, it can be seen by cumulating the frequencies for the first two categories that 1,597,000 owner-occupied housing units occupied by recent movers, or 36.4 percent, had up to two persons (actually, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 977,000 housing units, or 22.2 percent, had three persons (i.e., 2.5 to 3.5 persons). By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$2.5 + (3.5-2.5)\left(\frac{48.6-36.4}{22.2}\right) = 3.05$$

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

$$2.5 + (3.5-2.5)\left(\frac{51.4-36.4}{22.2}\right) = 3.18$$

Thus, the 95-percent confidence interval ranges from 3.05 to 3.18 persons.

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

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

However, for the 1973 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 accuracy of the AHS estimates.

As part of the reinterview, a check was made at each of these households to determine if the following were 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 interviewer passed or failed this reinterview with regard to the above items.

The results of the reinterview study are presented in the following internal memorandum: **Reinterview Results for the Annual Housing Survey-National Sample: 1973.** Some of these results are:

1. "Overall, the results showed a fairly consistent response between the original interview and the reinterview."
2. "For most 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 this study 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.

With respect to errors of coverage and estimation for missing data, 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 (see section on estimation). 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. This assumption, however, has been found to be invalid since there was an estimated 600,000 units with permits issued before 1970 but which were completed after the 1970 census. Therefore, these conventional new construction units had no representation in the AHS sample of conventional new construction. Secondly, unlike the procedure for conventional new construction, there is no sampling procedure connected with the representation of 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 AHS sample misses about one-half of all new mobile homes (i.e., about 700,000 units). The second-stage ratio estimation procedure was employed to correct both these deficiencies.