# MEASURING THE EFFICIENCY OF IMPLICIT STRATIFICATION IN MULTISTAGE SAMPLE SURVEYS

D.J. Megill, E.E. Gomez, Bureau of the Census; A. Balmaceda, M. Castillo, Ministry of Agriculture, Peru
D. J. Megill, Bureau of the Census[1], SMD, Washington, D.C. 20233

**Introduction**

Systematic selection is commonly used in multistage sampling for large scale household surveys, as it is simple to implement and it makes it possible to increase the precision of survey estimates through implicit stratification by ordering the sampling units in the frame by certain criteria correlated with the key estimates to be obtained from the survey data. In some surveys systematic selection is used at each sampling stage, from primary sampling units (PSU's) ordered geographically or by some socioeconomic characteristics within each stratum, to housing units listed sequentially within ultimate clusters. Of course, care must be taken to avoid any cyclical pattern in the ordering which could result in biased variance estimation. The criteria used for ordering the sampling units at each stage of selection is chosen to provide maximum correlation between adjacent units for the principal variables being measured. For example, in the case of farm household surveys, the PSU's within each explicit stratum are sometimes ordered geographically (e.g., in a serpentine manner) with the expectation that nearby areas are similar in climate, cropping patterns and socioeconomic characteristics of farm households.

Despite the frequent use of systematic sampling, most commonly used computer software packages for variance estimation from stratified multi-stage sample designs do not take into account implicit stratification. The corresponding variance estimators assume that ultimate clusters are randomly selected within a stratum. The sampling errors calculated using this type of variance estimator are probably overestimates, since it is expected that the implicit stratification resulting from systematic sampling from an ordered list would decrease the sampling error of survey estimates compared to those resulting from a completely random selection. Although it is sometimes desirable to use conservative estimates of sampling error, it would also be useful to measure the gains (or losses) in precision from the implicit stratification. By evaluating the corresponding efficiency of the systematic sampling, it is possible to study alternative ways of improving the implicit stratification in future surveys.

The study described in this paper had two main objectives. The first was to develop a procedure for reflecting the implicit stratification in the estimates of standard error for the Peru National Rural Household Survey (NRHS). The variance software package SUPER CARP (Cluster Analysis and Regression Program) was used for calculating variances for the survey estimates in a timely and cost-effective manner. The second objective was to measure the efficiency of the implicit stratification without resorting to complex and time-consuming estimation procedures. As in most developing countries, the resource constraints in government institutions in Peru place a limit on the amount of funds and staff time spent on research of a theoretical nature.

The methodology used for taking into account the implicit stratification in the variance estimation through SUPER CARP involved creating new explicit strata to simulate the effects (i.e., increased precision) of the implicit stratification. The effects of the implicit stratification were then measured by comparing the estimated standard errors based on this new stratification to that based on the original explicit strata. It was found that for a representative group of survey estimates at different levels of disaggregation, there was an overall estimated gain in precision of about 7.2 percent.

Since the main focus of this paper is on the methodology used for variance estimation for the NRHS, only a brief description of the general survey objectives and the sample design is provided below. Persons interested in further details concerning these areas may request survey methodological documents available at the Instituto Nacional de Estadistica (INE) in Lima, Peru.

**General Objectives of the NRHS**

The Peru NRHS, carried out by INE and the Oficina Sectorial de Estadistica (OSE) in the Peru Ministry of Agriculture, was sponsored jointly by the U.S. Agency for International Development (AID) and the Government of Peru. The U.S. Bureau of the Census provided technical assistance in all aspects of the survey and sample design and electronic data processing. A major objective of this survey was to provide a much-needed data base for the agricultural subsector pertaining to farm households, including variables related to crop and animal production and corresponding inputs, as well as socioeconomic characteristics of the farming population. The nonfarm households in rural areas were included in the survey as separate domains in order to compare their socioeconomic status to that of rural farm households. Although the main focus of the survey was on rural households, for comparative purposes the survey also included separate domains for farm households in urban centers where at least 20 percent of the households had farm operations. There was also an interest in obtaining separate estimates for each region of Peru. Therefore, the following 24 domains of study were defined for the survey:

1. Urban Coast - Farm Households
2. Urban Mountains - Farm Households
3. Urban Jungle Highlands - Farm Households
4. Urban Jungle Lowlands - Farm Households
5a. Rural North Coast - Farm Households
5b. Rural North Coast - Nonfarm Households
6a. Rural Central Coast - Farm Households
6b. Rural Central Coast - Nonfarm Households
7a. Rural South Coast - Farm Households
7b. Rural South Coast - Nonfarm Households
8a. Rural North Mountains - Farm Households
8b. Rural North Mountains - Nonfarm Households
9a. Rural Central Mountains - Farm Households
9b. Rural Central Mountains - Nonfarm Households
10a. Rural South Mountains - Farm Households

10b. Rural South Mountains - Nonfarm Households
11a. Rural North Jungle Highlands - Farm Households
11b. Rural North Jungle Highlands - Nonfarm Households
12a. Rural Central Jungle Highlands - Farm Households
12b. Rural Central Jungle Highlands - Nonfarm Households
13a. Rural South Jungle Highlands - Farm Households
13b. Rural South Jungle Highlands - Nonfarm Households
14a. Rural Jungle Lowlands - Farm Households
14b. Rural Jungle Lowlands - Nonfarm Households

**Sample Design for NRHS**

A stratified multi-stage sample design was used for the NRHS. Data and cartographic materials from the 1981 Census of Population and Housing in Peru were used to develop the sampling frame. For that census, a combination of urban blocks and rural enumeration areas (EA's) were defined to cover the entire territory of Peru. The urban sector was defined as cities with a population of at least 2,000. Only the cities with at least 20 percent of the households having farm operations were included in the urban frame for the survey. The ultimate clusters (segments) defined for the survey were rural EA's and urban blocks (or a combination of blocks) with an average of 100 housing units each.

The sampling frame was stratified by 14 regions corresponding to the geographic domains of study. In the urban strata there were three stages of sampling: cities were selected with probability proportional to size (PPS) at the first stage, segments were selected with PPS at the second stage and housing units were selected at the third stage. The measures of size in the urban strata were based on the number of farm households identified in the 1981 census. Systematic selection was used at each sampling stage. Within each urban stratum, the cities were ordered geographically at the first stage, and within sample cities the segments were ordered by the proportion of households with farm operations (as an indicator of the intensity of agricultural activities) at the second stage, to provide an implicit stratification by these criteria. The cities were used basically as counting units in order to select the urban segments more efficiently. The resulting dispersion of the urban sample segments was similar to that which would have been obtained from a two-stage sample design (i.e., there was no clustering above the segment level). For self-representing cities (with the number of housing units larger than the first stage sampling interval), the number of sample segments was allocated proportional to size in order to maintain an approximately self-weighting sample within each stratum.

Each rural stratum was divided into two substrata: (1) small towns with a population of 500 to 2,000; and (2) rural areas with dispersed housing units. The number of sample segments in each stratum was allocated proportionally between the two substrata. In the first substratum, the sampling was carried out in three stages in the same manner as in the urban strata. In the second substratum, only two sampling stages were used, with segments selected systematically with PPS at the first stage and housing units selected systematically at the second stage. The ordering of the segments at the first stage was first by department (state) and then by the proportion of households with farm operations within the segment. The measure of size for each segment in the rural strata was defined as the total number of housing units enumerated in the 1981 Census.

In this manner, 30 sample segments were selected within each of the 14 geographic strata. Within each sample segment, a listing of housing units was carried out about 1 month before the survey. At the last stage of selection, the housing units listed in each sample segment were stratified into farming and nonfarming substrata. Within each sample urban segment 10 housing units with farm operations were selected, while in the sample rural segments a sample of 10 housing units without farm operations was selected in addition to 10 housing units with farm operations. Thus, a total sample size of 7,200 housing units in 420 sample segments was specified in the basic design: 300 housing units per domain of study. Minor adjustments of the design in regions with sociopolitical problems are described in the survey documentation. The final number of completed interviews was 4146 farm households and 1839 nonfarm households. The overall noninterview rate was 14 percent, although 8 percent of this corresponded to entire segments which could not be reached because of sociopolitical problems, so that the effective household noninterview rate was 6 percent.

**Weighting Procedures**

The sample was designed to be approximately self-weighting within each stratum, although within a rural stratum the weights vary by sample segment according to the proportion of households with farm operations in the segment, and in each urban stratum the weights vary by sample segment according to how well the measure of size for the segment approximated the actual number of farm households listed.

The basic sampling weight (or expansion factor) for the sample households interviewed in each sample segment was calculated as the inverse of the final probability of selection (i.e., the product of the probabilities of selection at each sampling stage). These weights were adjusted at the segment level for noninterviews in two stages: one for noninterview households and another for noninterview segments.

**Variance Estimation**

Given the complexity of developing customized computer programs to calculate variances and the large amount of time this would require, it was decided to use an existing variance software package, SUPER CARP. The variance estimators included in SUPER CARP take into account a stratified multistage sample design such as that used for the NRHS. This software package provides for the calculation of variances for estimates of totals, means, proportions and ratios, as well as regression coefficients.

SUPER CARP uses an ultimate cluster type of variance estimator based on the squared difference between weighted segment totals. The variance estimator for ratios is based on a Taylor series expansion. The variance formulas are presented in the SUPER CARP manual in the form of matrices. The following formulas, presented in a simpler form, are used by SUPER CARP to calculate the variance of totals and ratios (without a finite population correction factor).

(1) Variance of a Total Estimate ($\hat{X}$)

$$\mathrm{Var}(\hat{X}) = \sum \frac{m_h}{m_h-1} \sum \left(\hat{X}_{hi} - \frac{\hat{X}_h}{m_h}\right)^2,$$

where:

$m_h$ = number of sample segments selected in stratum h

$\hat{X}_{hi} = \sum_{j=1}^{n_{hi}} W_{hij} X_{hij}$ = weighted total of variable X for the i-th sample segment in stratum h

$n_{hi}$ = number of sample households in the i-th sample segment in stratum h

$W_{hij}$ = final weight for the j-th sample household in the i-th sample segment in stratum h

$X_{hij}$ = value of variable X for the j-th sample household in the i-th sample segment in stratum h

$\hat{X}_h = \sum^{m_h} \hat{X}_{hi}$ = weighted total of X for substratum h

$\hat{X} = \sum_h X_h$

(2) Variance of a Ratio Estimate ($\hat{Y}/\hat{X}$)

$$\mathrm{Var}\left(\frac{\hat{Y}}{\hat{X}}\right) = \sum_h \frac{m_h}{m_h-1} \left(\frac{1}{\hat{X}^2}\right) \sum_{i=1}^{m_h} \left[ \left(\hat{Y}_{hi} - \frac{\hat{Y}_h}{m_h}\right)^2 + \left(\frac{\hat{Y}}{\hat{Y}}\right)^2 \left(\hat{X}_{hi} - \frac{\hat{X}}{m_h}\right)^2 - 2\left(\frac{\hat{Y}}{\hat{X}}\right)\left(\hat{Y}_{hi} - \frac{\hat{Y}_h}{m_h}\right)\left(\hat{X}_{hi} - \frac{\hat{X}_h}{m_h}\right)\right],$$

where $\hat{Y}$, $\hat{Y}_h$, $\hat{Y}_{hi}$ and $Y_{hij}$ are defined for variable Y in the same manner as $\hat{X}$, $\hat{X}_h$, $\hat{X}_{hi}$ and $X_{hij}$, respectively.

Note: Means and proportions would be special types of ratios, in which the variable $X_{hij}$ (in the case of means) or both $X_{hij}$ and $Y_{hij}$ (in the case of proportions) are defined as variables equal to 1 or 0.

Even though SUPER CARP has the option of a finite population correction (FPC) factor, it is based on the assumption that the ultimate clusters are selected with equal probability within a stratum. However, the sample segments for the NRHS were selected with probability proportional to size. For this reason, the FPC option was not used for this survey. In any case, an FPC factor would only have a small effect on the variance estimates, since the overall sampling rate is relatively low. Also, for the analysis described later on measuring the efficiency of the implicit stratification, the FPC factors would cancel out in the estimates of percentage change in standard errors. A representative group of different types of survey estimates at various levels of disaggregation was selected from the survey tables in order to calculate the corresponding variances using SUPER CARP.

**Methodology Used to Reflect Implicit Stratification in Variance Estimates**

It can be seen from the formulas for the SUPER CARP variance estimators specified in the previous section that they assume the random selection of ultimate clusters within each stratum. The sample segments for the NRHS were actually selected systematically PPS, so these variance estimators would not reflect the gain (or loss) in precision from the implicit stratification in the ordered sampling frame. In order to improve the variance estimation so that it would take into account such implicit stratification, the 30 sample segments within each stratum were subdivided into smaller strata of two or three sequentially selected segments each. These new explicit strata were created to simulate the effects of the implicit stratification. This procedure was carried out based on the assumption that implicit stratification would be equivalent to having had the original ordered list of segments in the frame for each stratum subdivided into 15 equal-sized strata with two sample segments selected from each.

The systematic PPS sampling actually has the effect of dividing a geographic stratum into $m_h$ implicit substrata with boundaries defined by multiples of the first stage sampling interval mapped onto the cumulated measures of size (although a PSU may overlap a boundary, in which case it has a probability associated with more than one substratum); one PSU is selected from each of these substrata. The methodology for variance estimation used in this study involves grouping such substrata in consecutive pairs. With two sample PSU's per stratum, the ultimate cluster variance estimator used in SUPER CARP is equivalent to the estimator presented as equation 7.7.3 in "Introduction to Variance Estimation" (Kirk M. Wolter, 1985), page 287. An alternative variance estimator under systematic sampling shown in the same text, equation 7.7.4, involves the sum of squared differences between all sequentially selected pairs of sample PSU's within a geographic stratum. That equation may be expressed in the terms defined previously for the variance of a total estimate, as follows:

$$\mathrm{Var}(\hat{X})' = \sum_h \frac{m_h}{2(m_h-1)} \sum_{i=1}^{m_h-1} \left[\hat{X}_{hi} - \hat{X}_{h(i+1)}\right]^2,$$

where: $\hat{X}_{h(i+1)}$ = weighted total of variable X for the sample segment following the i-th one.

Unlike the SUPER CARP variance estimator used in this study (based on two sample PSU's per stratum), the estimator Var $(\hat{X})'$ utilizes overlapping differences. As Wolter points out, this estimator aims to increase the number of "degrees of freedom." In "Sample Survey Methods and Theory, Volume I" (Hansen, Hurwitz, and Madow, 1953), Chapter 11, Section 8, it is also stated that the precision of the variance estimates (under systematic sampling) would be increased if the grouping is extended to all possible pairs of adjacent substrata.

As indicated previously, it was necessary to use the software package SUPER CARP to calculate the variances for the Peru survey because of resource constraints. However, in comparing Var $(X)'$ to the corresponding ultimate cluster variance estimator Var $(X)$ with two sample segments per stratum, it can be seen that the results should be similar. Both are based on a similar assumption about the grouping of implicit substrata.

In the urban strata, each self-representing (SR) city with two or more sample segments was treated as a separate stratum, and SR cities with one sample segment were collapsed. For the non-self-representing segments in the urban strata and those in each of the rural substrata (small towns and rural areas with dispersed housing units), the new strata were defined separately within each department. A total of 188 new strata with two or three sample segments each were defined in this way. The stratum numbers in the SUPER CARP data file were recoded accordingly.

**Methodology Used for Measuring the Efficiency of Implicit Stratification**

Since the variance estimation procedure based on the restratification described in the previous section is expected to reflect the gain (or loss) in precision due to the implicit stratification of segments, it would be desirable to quantify the difference between the resulting estimates of standard errors and those which would have resulted from a corresponding random PPS selection of sample segments within each of the original 14 strata. Given the resource and time constraints related to carrying out this research in Peru, it was necessary to find a simple and cost effective procedure for measuring the efficiency of the implicit stratification. A very simple approach to this problem was to recalculate the variances for the same set of survey estimates after changing the stratification back to the original 14 geographic strata with 30 sample segments each. This was quite easy to carry out with SUPER CARP, since it was only necessary to recode the stratum numbers on the SUPER CARP data file and rerun the programs using the same parameter cards. The sampling errors thus obtained were then compared to those resulting from the 188 new strata to measure the efficiency of the implicit stratification. Since the two alternative ultimate cluster variance estimates are not unbiased, the difference between them should only be considered an approximation to the true effect of the implicit stratification. By examining the results for different types of estimates and various levels of disaggregation, it was possible to evaluate the relative efficiency of the implicit stratification within each of the 14 geographic strata for the different estimates. The results of this study are presented below.

**Results of Implicit Stratification Study**

The difference between the sampling error resulting from the original explicit stratification by the 14 regions and that resulting from the 188 new strata to simulate the implicit stratification was calculated for a representative group of estimates from the survey tables. The types of survey estimates included in this study only involved the 4096 farm household records for the NRHS (i.e., unweighted sample size for the subpopulation).

As expected, there was an overall increase in the precision of survey estimates resulting from the implicit stratification of the segments by geography and percentage of households with farm operations in the segment. The overall unweighted average percentage decrease in sampling error (i.e., gain in precision) due to implicit stratification across all the estimates tabulated was 6.2 percent, with a standard deviation of 14.8 percent, indicating a large variability. Some of the differences were negative, indicating a loss of precision, although since most of these cases are related to estimates for subpopulations with a small sample size and having less reliable standard error estimates, these negative differences are probably mostly due to random variation.

In order to take into account the relative sample size involved in each subpopulation estimate when averaging the differences across estimates, the weighted average percentage differences were calculated by multiplying each difference by the corresponding number of observations. These tabulations were facilitated by the use of a microcomputer spreadsheet program, which provided the flexibility to summarize these results in Tables 1 through 5. In order to indicate the overall efficiency of the implicit stratification for each type of estimate, the weighted average percentage difference across subpopulations was obtained.

The overall weighted average percentage gain in precision due to implicit stratification across all the estimates tabulated was 7.2 percent. Comparing this figure to the unweighted estimate of 6.2 percent, it can be seen that the negative differences were concentrated in estimates for subpopulations with small sample sizes and therefore small weights. In the case of estimates at the national level, there was always a gain in precision, although the magnitude varied for different estimates. The weighted average difference for each type of estimate was also positive. There were also a few small negative differences at the regional level.

In order to compare the percentage difference in S.E. (between the S.E. based on 14 strata and that based on 188 strata) for the various estimates, Tables 1 through 5 present a summary by type of disaggregation. A positive difference indicates a gain in precision from the implicit stratification. These tables also show the overall average percentage difference in S.E. across all the estimates for each category in order to determine the overall efficiency of the implicit stratification for a multi-purpose type of survey. It can be seen from each table that the implicit stratification resulted in an overall gain in precision for each estimate. The overall weighted average percentage difference in S.E. across all the regional estimates in Table 1 is 11.2 percent, which appears to be a reasonable gain in precision. One interesting observation from this table is that for a particular region there is considerable variation in the percentage difference in S.E. for the different estimates, indicating that even within a geographic stratum the implicit stratification is more efficient for some variables than for others. Across all the estimates, the implicit stratification is most efficient in the Urban Jungle Lowlands and the Rural Central Jungle Highlands.

Across all the estimates only one region, the Rural Central Coast, showed a minor loss in precision (-2 .8 percent) from the implicit stratification. Although such a small negative difference may have resulted from sampling variability, the distribution of EA's in this region was closely examined to determine whether there was any particular feature of the new stratification which may have resulted in a higher sampling error. There are four departments in the Rural Central Coast (Ancash, Ica, Lima, and Callao), and each new stratum was defined within a department, except for the case of Callao with only one sample rural segment, which was assigned to form a new stratum with a sample segment from the adjacent department. It would be interesting in a more in-depth analysis to actually examine the sample segment totals for certain estimates to determine which new strata in that region made the largest contribution to the variance.

In the case of the estimates by region and farm size, Table 5 summarizes the weighted average percentage difference in S.E. by region, in order to examine the overall efficiency of the implicit stratification within each geographic stratum. Perhaps as a result of the higher degree of disaggregation for these estimates, their overall weighted average percentage difference in S.E. (4.9 percent) was less than half that for the regional estimates (11.2 percent), although it should be pointed out that the groups of estimates being compared are different.

It should be noted that the results presented in these tables are based on preliminary tabulations from the NRHS data. Given the low number of observations for some cross-tabulation cells by region/farm size and region/income group, the final survey results were published at a more aggregated level.

**General Conclusions**

The main conclusion from the study was that this methodological approach represents a simple and cost-effective manner of accounting for the effects of implicit stratification in the calculation of variances for estimates from multistage sample designs, and of measuring the efficiency of the design. In the case of the results from the Peru NRHS, it was shown that by ordering the PSU's geographically and by percentage of households with farm operations for the systematic selection, the corresponding implicit stratification resulted in an overall gain in precision in the survey estimates over what would have resulted from a simple random sample of PSU's.

**Considerations for Further Analysis**

As stated previously, the scope of this study was limited by resource constraints in carrying out the analysis in Peru. However, the study also pointed out interesting areas for further analysis on the efficiency of implicit stratification. Given that systematic sampling is commonly used in large scale sample surveys, it is important to research the most efficient criteria on which to base the ordering of PSU's for a particular type of survey. When data from a similar survey carried out previously is available, it may be possible to simulate the effects of different types of implicit stratification by creating new strata with two or three sample PSU's each based on different criteria to determine which simulated alternative provides the lowest S.E's for various estimates. Such results would also be useful for determining the most efficient stratification variables for sample designs with two sample PSU's per explicit stratum.

It is also desirable to conduct out a more in-depth analysis of the efficiency of the sample design for a particular survey already carried out, such as the Peru survey, by examining which new strata have the highest contribution to the variance. By researching the sources of large variance contributions, it may be possible to improve the sample design for future surveys.

**REFERENCES**


Hansen, Morris H., Hurwitz, William N. and Madow, William G., Sample Survey Methods and Theory, Volume I "Methods and Applications". A Wiley Publication in Mathematical Statistics. 1953.

Wolter, Kirk M. , Introduction to Variance Estimation, Springer Series in Statistics. 1985.


[1] This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the author(s) and do not necessarily reflect those of the Census Bureau.

SUMMARY TABLES FOR WEIGHTED PERCENTAGE DIFFERENCES IN STANDARD ERRORS (FROM 14 STRATA AND 188 STRATA) BY TYPE OF DISAGGREGATION

TABLE 1. ESTIMATES FOR PERU

| PERCENTAGE DIFFERENCE IN S.E. | | | | | | | |
|---|---|---|---|---|---|---|---|
| ESTIMATE | | | | | | | |
| NO.5 | NO.7 | NO.11 | NO.15 | NO.18 | NO.21 | NO.24 | AVG. |
| % | % | % | % | % | % | % | % |
| 3.9 | .3 | 5.8 | 10.0 | 2.2 | 23.8 | 2.0 | 6.9 |

TABLE 2. ESTIMATES BY REGION

| REGION | PERCENTAGE DIFFERENCE IN S.E. | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | ESTIMATE | | | | | | | |
| | NO.1 | NO.4 | NO.8 | NO.12 | NO.13 | NO.17 | NO.23 | AVG. |
| | % | % | % | % | % | % | % | % |
| URBAN COAST | -10.9 | 20.9 | 3.1 | 3.1 | 13.1 | .6 | 27.9 | 8.3 |
| URBAN MOUNTAINS | 14.8 | -2.4 | 14.4 | 21.1 | -1.0 | -1.5 | 5.6 | 7.3 |
| URBAN JUNGLE HIGHLANDS | 15.0 | -.7 | 17.9 | 0.0 | 40.4 | 22.5 | 23.3 | 16.9 |
| URBAN JUNGLE LOWLANDS | 24.0 | 28.0 | 37.0 | 0.0 | 34.3 | -2.5 | 42.9 | 23.4 |
| RURAL NORTH COAST | 21.0 | -13.1 | -6.3 | -7.1 | 12.3 | -1.8 | 3.8 | 1.3 |
| RURAL CENTRAL COAST | 1.3 | -2.0 | -2.4 | -4.5 | -15.9 | -1.1 | 5.0 | -2.8 |
| RURAL SOUTH COAST | 50.2 | 20.6 | -17.9 | 0.0 | .7 | 18.5 | 14.4 | 12.4 |
| RURAL N. MOUNTAINS | 1.5 | 26.6 | 13.5 | -12.2 | -2.6 | .6 | 19.4 | 6.7 |
| RURAL C. MOUNTAINS | 8.1 | 31.7 | 10.9 | 6.5 | 9.1 | -21.6 | 4.0 | 7.0 |
| RURAL S. MOUNTAINS | 25.5 | -17.7 | 26.7 | 13.7 | .5 | 62.6 | 24.3 | 19.4 |
| RURAL N. J. HIGHLANDS | -5.7 | -12.8 | 7.3 | 41.4 | -4.1 | 22.7 | 0.0 | 7.0 |
| RURAL C. J. HIGHLANDS | 3.0 | 0.0 | -3.6 | -2.6 | -4.7 | 16.5 | 31.6 | 5.7 |
| RURAL S. J. HIGHLANDS | 13.7 | 5.7 | 18.1 | 23.4 | 19.1 | 81.3 | 17.8 | 25.6 |
| RURAL JUNGLE LOWLANDS | -6.4 | 15.8 | -2.3 | 19.4 | 10.0 | 30.3 | 9.2 | 10.9 |
| WEIGHTED AVG. DIFF. | 11.4 | 7.1 | 8.1 | 9.9 | 7.7 | 17.8 | 16.0 | 11.2 |

TABLE 3. ESTIMATES BY FARM SIZE

| FARM SIZE | PERCENTAGE DIFFERENCE IN S.E. | | | |
|---|---|---|---|---|
| | ESTIMATE | | | |
| | NO.2 | NO.9 | NO.20 | AVG. |
| | % | % | % | % |
| 0.1-0.99 HA. | -1.6 | 6.1 | -2.1 | .8 |
| 1-1.99 HA. | 10.4 | 2.7 | 1.6 | 4.9 |
| 2-4.99 HA. | 0.0 | -4.6 | 1.8 | -.9 |
| 5-9.99 HA. | -2.2 | 4.8 | 0.0 | .9 |
| 10-19.99 HA. | -3.0 | -1.5 | 0.0 | -1.5 |
| 20-99.99 HA. | -7.0 | 0.0 | .6 | -2.1 |
| 100 HAS. OR MORE | 0.0 | -25.5 | 0.0 | -8.5 |
| WITHOUT LAND | 7.2 | 49.1 | | 28.2 |
| WEIGHTED AVG. DIFF. | 4.7 | 1.3 | .4 | 2.1 |

TABLE 4. ESTIMATES BY INCOME GROUP

| INCOME GROUP | PERCENTAGE DIFFERENCE IN S.E. | | | | |
|---|---|---|---|---|---|
| | ESTIMATE | | | | |
| | NO.10 | NO.14 | NO.16 | NO.25 | AVG. |
| | % | % | % | % | % |
| NEGATIVE ANNUAL INCOME | 12.7 | 2.7 | -1.0 | -.5 | 3.5 |
| 0-3000 INTIS | 3.5 | .9 | 11.7 | 6.8 | 5.7 |
| 3001-6000 INTIS | 14.0 | 9.2 | 3.2 | 11.5 | 9.5 |
| 6001-15000 INTIS | 14.4 | .9 | 25.4 | 10.3 | 12.8 |
| 15001-30000 INTIS | .6 | 1.0 | 1.0 | -5.0 | -.6 |
| 30001-70000 INTIS | -2.0 | 71.6 | 9.9 | 3.9 | 20.9 |
| 70000 INTIS OR MORE | 0.0 | 0.0 | 0.0 | 0.0 | .0 |
| WEIGHTED AVG. DIFF. | 7.0 | 3.1 | 10.2 | 6.6 | 6.7 |

TABLE 5. ESTIMATES BY REGION AND FARM SIZE (WEIGHTED AVERAGE PERCENTAGE DIFFERENCES BY REGION)

| REGION | WEIGHTED AVERAGE PERCENTAGE DIFFERENCE IN S.E. | | | |
|---|---|---|---|---|
| | ESTIMATE | | | |
| | NO.3 | NO.6 | NO.19 | AVG. |
| | % | % | % | % |
| URBAN COAST | 7.7 | 3.2 | 1.9 | 4.3 |
| URBAN MOUNTAINS | -2.6 | 4.4 | 18.3 | 6.7 |
| URBAN JUNGLE HIGHLANDS | 5.8 | 16.3 | 5.5 | 9.2 |
| URBAN JUNGLE LOWLANDS | 15.4 | 2.9 | 17.4 | 11.9 |
| RURAL NORTH COAST | 5.4 | 7.0 | 6.8 | 6.4 |
| RURAL CENTRAL COAST | 3.2 | -7.4 | 6.3 | .7 |
| RURAL SOUTH COAST | -4.0 | -5.3 | 8.1 | -.4 |
| RURAL N. MOUNTAINS | 4.6 | 7.5 | 10.1 | 7.4 |
| RURAL C. MOUNTAINS | .1 | 2.5 | 5.9 | 2.8 |
| RURAL S. MOUNTAINS | 3.1 | 11.0 | .4 | 4.8 |
| RURAL N. J. HIGHLANDS | -1.6 | .5 | 12.8 | 3.9 |
| RURAL C. J. HIGHLANDS | 1.1 | 2.7 | 1.7 | 1.8 |
| RURAL S. J. HIGHLANDS | -1.7 | 15.5 | 5.1 | 6.3 |
| RURAL JUNGLE LOWLANDS | 3.9 | 3.0 | 4.6 | 3.8 |
| WEIGHTED AVG. DIFF. | 2.7 | 4.6 | 7.4 | 4.9 |