COMPOSITE ESTIMATION OF TOTALS FOR LIVESTOCK SURVEYS. By Lynn Kuo, Statistical Survey Institution; Statistical Research Division, Statistical Reporting Service; U.S. Department of Agriculture, Washington, D.C. 20250; Staff Report No. SF&SRB 92.

Abstract

Four different preliminary estimators are employed by the Statistical Reporting Service (SRS) of the U.S. Department of Agriculture to obtain the final estimate for livestock inventories of major States. A composite estimation model is proposed here to solve the dilemma of how to combine these four estimators. The composite estimator is derived by minimizing a quadratic function subject to linear constraints. The variance and mean squared error of the composite estimator are evaluated by the jackknife method. The author analyzed estimator bias by assuming the tract estimator to be unbiased when nonsampling errors are considered. Numerical results based on the data from the 1984 June Enumerative Survey conducted by SRS support the use of composite estimation.

Key Words

Stratified sampling, multiple frame methodology, domain estimation, common total, convex programming, jackknife method.

Acknowledgment

A major portion of this work was done while the author was a research fellow at the Statistical Survey Institute sponsored by the American Statistical Association and the Statistical Reporting Service of the U.S. Department of Agriculture. The author wishes to thank both sponsors. She especially thanks Fred Leone for encouragement. The author wishes to thank Carney, Davies, Ford, Kleweno, Nealon, Steiner, and Vogel for their encouragement and discussions. She also wishes to thank Elaine Devine for typing. Special thanks go to Richard Coulter for discussion and providing the author much detailed information about the existing procedures; Ron Fecso and Robert Tortora for discussions and enthusiastic encouragement. The author owes Michael P. Cohen at the Bureau of Labor Statistics a special debt of gratitude for many discussions on composite estimation, sample reuse methods, other topics in applied statistics, and for reviewing an earlier draft of this paper.

Table of Contents

Composite Estimation of Totals
for Livestock Surveys

Lynn Kuo

1. Introduction

The June Enumerative Survey (JES) conducted by SRS is a multi-purpose probability survey where basic information concerning crop acreages, livestock inventory, and other agricultural characteristics are collected. The sampling units for the annual survey are selected from two sampling frames that have been constructed and are maintained by SRS.

The Area Sampling Frame is stratified by land use. This frame represents 100 percent of the geographical area of interest. Selection of the sampling units (segments) is from within each land use stratum. These units will vary in size but are targeted to be 1 square mile for concentrated cropland areas. Information is collected from the operators of the land within these segments by personal interview around the June 1 reference date.

A List Sampling Frame has been constructed by SRS to contain known farm operators. This frame is stratified by type and size of farm. Information is collected from the selected list units by mail, telephone, or personal visit around the June 1 reference date.

Different estimators are often produced for the same characteristics. For example, four estimators, tract, farm, weighted, and multiple frame screening estimator, are produced for livestock items for each of the 10 major States. These 10 States usually account for more than 80 percent of the U.S. hogs and cattle inventory. Three of the estimators are derived from the same primary sampling units. Due to different methods of associating the farm products with the segments (primary sampling units) from the area frame, three different estimators are produced. The tract estimator counts only the farm inventory within the segment. The farm estimator would include the farm inventory beyond the segment, so long as those farm products belong to the same operator residing in the segment. The weighted estimator uses the ratio of tract to farm acreages operated to prorate the farm inventory for each survey item to a tract level. A fourth estimator called the multiple frame screening estimator is predominantly computed from list sample data. To compensate for the incompleteness of the list frame, an area frame estimate of operators sampled but not found on the list is computed and is added to the list estimate.

One of the problems faced by the statisticians at SRS is finding a method to combine the four estimators into one. A composite estimation model is proposed here. This composite estimator is motivated by minimizing the mean squared errors of a family of weighted averages of the four preliminary estimators.

A brief discussion of the present procedure used by SRS to derive the final estimate can be found in the section headed by "forming the estimates" in Hog and Pig Reports (or Cattle Reports): A Handbook on Surveying and Estimating Procedures (Crop Reporting Board 1979 and 1981). To summarize: each State office obtains the summary for all the different estimators and makes its recommendations and comments to the Washington, D.C. office. In Washington, the Crop Reporting Board (CRB) is responsible for the final estimates. The CRB consists of the Chairperson and Secretary of the CRB, Director of the Estimates Division, Branch Chief, Section Head, and several other commodity and sampling specialists. The CRB meets to review the current estimates, previously published estimates, and other check data at the State and national levels. The review process is assisted by graphs as in Figure 0 which plot the different estimates over time. The check data include slaughtering information from commercial packing plants, import and export information, and U.S. Census of Agriculture information, available every 5 years.

Basically, the CRB combines four preliminary estimates into one final estimate published by SRS according to two processes. One is a judgmental process exercised by both the State offices and the CRB to obtain a final number. This process has also been described as a subjective weighting scheme by SRS (see p. 40 of Bynum et al 1985). This number is further examined against check data by balance sheet methods. Revisions might be employed in light of the check data.

The judgmental process puts the CRB in a potentially vulnerable position to defend the repeatability, accuracy, and ability to assess the variance of their final estimators. Composite estimation is proposed to replace the judgmental process.

In addition to the balance sheet methods, statistical methodology using past data and the Census of Agriculture information is also needed for the revision process. Further improvement on the composite estimator can be obtained. However, it is beyond the scope of this paper.

Other approaches such as empirical Bayes and linear Bayes were also explored by the author to solve this problem. Composite estimation has been pursued. The strictly frequentist and nonparametric features of composite estimation are also shared by classical survey sampling. These two features give composite estimation the greatest potential for implementation by SRS.

As can be seen from the numerical results in Section 6, not only variances but also nonnegligible biases affect the accuracy of the preliminary estimators. Consequently, analysis of biases has to be incorporated. The author assumes that the tract estimator is unbiased, and all other estimators are biased when nonsampling errors are considered. This assumption is also supported by Nealon (1984), where discussion on the biases of the weighted and multiple frame screening estimator can be found. The tract estimator by design is least susceptible to nonsampling errors. An unbiased estimator of the bias squared term developed in Section 4 is used for the biased preliminary estimators.

Some of the major recommendations given by a 1980 statistical review panel of non-USDA statisticians are as follows (see p. 2 of Bynum et al 1985). 1. The CRB should have standard errors, biases, and historical errors available to them. 2. State statisticians should provide recommendations expressed as point estimates and their ranges. 3. The bias component of error on probability based estimators should be quantified. 4. SRS should publish at least the probability based estimates. 5. CRB should set national estimates that lie within bounds of some form of confidence limit or some weighted combination of estimates adjusted for bias. The analysis developed in Section 4 provides a solution to quantify biases. The composite estimation developed in this paper provides an adjustment for component weights depending on biases.

Mosteller (1948) discusses the desirability of pooling the data. He describes several ways of pooling data from two samples to estimate the mean of one of the populations. He illustrates it by using data from the normal distribution, but his ideas are applicable in a broader context. A stout believer in unbiasedness would only use the tract estimator which is least susceptive to nonsampling errors. However, most statisticians are willing to accept some bias to reduce the mean squared error. This is done by pooling all the available data.

Theoretical work on composite estimation for independent observations from the normal distribution is given by Graybill and Deal (1959). To combine two independent unbiased preliminary estimators for the common mean, they show the composite estimator has uniformly smaller variance than any of the preliminary estimators so long as each sample size is greater than 10. Further improvement and other related references are given by Brown and Cohen (1974). Although the situation at SRS is much more complicated, these theoretical works shed light on the advantage of intelligently combining estimators.

It would be desirable to have theoretical results for composite estimators without distributional assumptions. Many of the estimates SRS produces are influenced by large farm operators in the sample. It would be difficult to justify a particular distribution assumption, especially for repeated use.

Composite estimation has been used by numerous statisticians in applications. Schaible (1978 and 1979) uses it to estimate small area statistics for the Health Interview Survey. Brock, French, and Peyton (1980) provide an empirical evaluation of mean squared errors of composite estimators, and suggestions for component estimators for small area estimation. Cohen and Sommers (1984) provide empirical evaluation of composite estimation of cost weights for the Consumer Price Index. There is also extensive literature on composite estimation for the Current Population Survey for panel studies and rotation designs. See Wolter (1979) for the theory, applications, and other references.

Composite estimation has been anticipated by the statisticians at SRS. Houseman (1971) proposes composite estimators which combine estimators from a probability survey and indicators from a nonprobability survey. He indicates that weights from the probability survey should be a function of the variances and covariances of the estimators. Weights for the nonprobability survey are assigned according to past performance and other information.

Bosecker and Ford (1976) at SRS develop a composite estimator by generalizing Hartley's multiple frame estimator to stratified samples. The total for each stratum in the overlap domain is estimated by a composite estimator. They show by empirical results from two States that the sampling errors of this estimator are 14 percent lower than that of the present multiple frame screening estimator. This proposed estimator has not been adopted by SRS in its operational program.

In Framework for the Future, a report of the Long-Range Planning Group of SRS, Allen et al (1983) make the following suggestions regarding CRB standards. An objective procedure for weighting the different estimators should be developed. These weights could be determined by reviewing previous estimators prior to the availability of the current data. Nealon (1984) reviews the strengths and weaknesses of the four estimators and compares them to the official statistics published by the CRB to gain insight on the objective weighting scheme. It is not clear to the author how SRS intends to pursue these suggestions.

A possible simplifying assumption is that the variance and covariance of the preliminary estimators are quite stable over the years. Therefore, weights determined previously could be applied to current survey estimators. The composite weights proposed by the author are derived from the current survey for the following reasons. First, the assumption of stabilized variances has not been validated. Second, if weighting is derived from the covariance matrix, it would be more efficient and accurate to derive it from the present data.

Nealon's report has assisted the author to formulate the present study. Moreover, Nealon's observations should be useful for future survey research. To incorporate them, more complicated analysis is required such as developing an empirical Bayes or Bayes method. These methods could be of future interest to SRS.

In the SRS National Conference Proceedings (1984), Ford summarizes group discussion regarding composite estimation. Most SRS statisticians agree that it may be a good time to try composite estimation. However, the main difficulty remains in deciding the weights. Suggestions made for weighting include use of standard errors of the preliminary estimators for the probability surveys. Equal weighting or weighting depending on the historical relationship of preliminary estimates to CRB estimates is suggested for survey estimators in general. The author is skeptical of the latter suggestion. Two dangers are also pointed out by the group. First, the use of composite estimators might preclude the statisticians looking at its components and their properties. Second, the use of composite estimators might deter SRS from deleting some of the components which are not very useful. Perhaps both cautions are well-founded. The evaluation of the variance and mean squared error of the composite estimator is proposed here. It is sufficient to use only the composite estimator if its mean squared error is smaller than those of its components. Numerical results in Section 6 reveal that the methodology proposed here also has potential for providing justification for deleting some of the less useful estimators. This point will be expanded later.

In a recent publication entitled Crop Reporting Board Standards, Bynum et al (1985) voices the need for defendable statistical methodology to replace subjective judgments exercised by the CRB. They specify that the optimum weighting scheme depending on current or historical sampling errors should be produced for the following reports: acres planted, acres harvested, yield, production, stocks, hogs and pig inventory, and cattle inventory. This paper provides a method for generating the optimum weights.

The four preliminary estimators presently in use at SRS are described in Section 2. A review of composite estimation and its specialization to SRS applications are given in Section 3. Estimation of the second moment term needed in composite estimation is discussed in Section 4. Variance and mean squared error evaluations of the composite estimators are discussed in Section 5. Numerical results for total hogs and pigs inventory from the 1984 June Enumerative Survey are given in Section 6. Finally, the conclusion and recommendations are given in Section 7.

2. Description of Presently Used Estimators

As mentioned earlier, both area and list frames are used by SRS to select samples for probability surveys.

The area frame for each State used by SRS is stratified by land use, for example, more than 75 percent cultivated, 50-74 percent cultivated, 15-49 percent cultivated, agriculture mixed with urban, non-agricultural land, etc. Each stratum is further subdivided into more homogeneous geographic substrata called paper strata. Segments (parcels of land) treated as the primary sampling units are selected as a simple stratified sample from each paper stratum. A detailed description on how the segments are constructed from aerial photographs with identifiable boundaries, how segment sizes and the number of segments are determined, and how the segments are selected via count units can be found in Houseman (1975) and Geuder (1984). The first segment selected in each paper stratum is designated as replicate 1, the second segment as replicate 2, etc. Approximately 20 percent of the segments are replaced annually on a rotational basis.

The list frame consisting of names of farmers is stratified by the size of farms contained in the control information. For example, for hogs and pigs inventory, typical strata are no hogs, 1-99 hogs, 100-199 hogs, 200-399 hogs, 400-999 hogs, 1000-2499 hogs, and more than 2500 hogs. Systematic sampling from each stratum is usually used to select the list sample. See Section 5 of the June Supervising and Editing Manual (1984).

For each area sample, there are three different methods of evaluating the farm inventory. A tract is a piece of land within the boundary of the segment under one management. A tract may be the entire farm if all of it is in the segment, or a portion of the farm, if the farm's boundary extends outside of the segment. The area tract estimator is an expansion of inventory on all the tracts of the selected segments. The area farm estimator is an expansion of inventory on the farms where the operator resides in the segment. The area weighted estimator is computed using farm inventory weighted by the ratio of tract acreage to farm acreage, for all tracts regardless of the residency of the operator. There are no such complications for the list sample. The list sample uses the inventory of the entire farm.

Three different domains are needed to explain the four estimators presently in use. Domain D1, the nonoverlap domain, refers to the farms not in the list frame. (This domain is automatically in the area frame, since the area frame is complete). Domain D2 refers to the farms in both frames not classified as "extreme operators." Domain D3 refers to the extreme operators in both frames. (Extreme operators are farmers with very large livestock inventories. The exact definition for the list sample in the Domain D3 will be given later.)

A version of a multiple frame estimator for estimating the population total can be written as

$$\hat{Y} = \hat{Y}_{D1 \cup D2, A1} + p\,\hat{Y}_{D3, A1} + (1-p)\,\hat{Y}_{D3, L}$$

where $\hat{Y}_{D1 \cup D2, A1}$ is the tract estimator for the D1 u D2 domain, $\hat{Y}_{D3, A1}$ and $\hat{Y}_{D3, L}$ denote the two estimators for D3 expanded from the tract and the list sample respectively. The quantity p is determined by minimizing the variance of $\hat{Y}$. The estimator $\hat{Y}_{D3, A1}$ usually has a large variance. Consequently, p is set to zero in the SRS current procedure. The operational tract, farm, and weighted estimators denoted by Y_1, Y_2, and Y_3, can be expressed as follows:

$$\hat{Y}_i = \hat{Y}_{D1 \cup D2, Ai} + \hat{Y}_{D3, L}, \text{ where } i = 1, 2 \text{ or } 3. \tag{2.1}$$

The estimator $\hat{Y}_{D1 \cup D2, Ai}$ is computed by

$$\hat{Y}_{D1 \cup D2, Ai} = \sum_{h \in H} e_h \sum_{k=1}^{n_h} y_{i,hk} \tag{2.2}$$

where H = the collection of paper strata,

e_h = the inverse of the probability of selection of each segment in the h^{th} paper stratum,

n_h = the number of segments sampled in the h^{th} paper stratum,

$$y_{1,hk} = \sum_{l=1}^{g_{hk}} t_{hkl} \, \delta_{hkl}$$

$$y_{2,hk} = \sum_{l=1}^{g_{hk}} f_{hkl} \, d_{hkl} \, \delta_{hkl}$$

$$y_{3,hk} = \sum_{l=1}^{g_{hk}} f_{hkl} \frac{a_{hkl}}{b_{hkl}} \delta_{hkl} \quad \text{with.}$$

t_{hkl} = the value of the characteristic for the l^{th} tract in the k^{th} segment of the h^{th} stratum,

f_{hkl} = the value of the characteristic for the l^{th} farm overlap with the k^{th} segment of the h stratum,

a_{hkl} = acreage of the hkl^{th} tract,

b_{hkl} = acreage of the hkl^{th} farm,

g_{hk} = total number of tracts in the hk^{th} segment,

$$d_{hkl} = \begin{cases} 1 & \text{if the operator of } hkl^{th} \text{ farm resides in the } hk^{th} \text{ segment} \\ 0 & \text{otherwise,} \end{cases}$$

$$\delta_{hkl} = \begin{cases} 1 & \text{if } hkl^{th} \text{ farm is in } D_1 \cup D_2 \\ 0 & \text{otherwise,} \end{cases}$$

The estimator $\hat{Y}_{D3,\,L}$ is computed from the list samples in the extreme operator (EO) strata :

$$\hat{Y}_{D3,\,L} = \sum_{l \in EO} \frac{N_l}{n_l} \sum_{k=1}^{n_l} y_{lk}$$

where y_{lk} = the value of the k^{th} farm in the l^{th} stratum,
N_l = the population size of the l^{th} stratum,
n_l = the sample size of the l^{th} stratum,
EO = collection of list strata with extreme operators.

Remark 2.1: The definition of EO strata from the list population depends on the State. For example, the EO strata for Indiana hogs consist of three strata defined by the size of the farms: 1000-1999 hogs, 2000-4999 hogs, and more than 5000 hogs. The largest stratum is sampled with probability one. The other EO strata are sampled at varying rates approximately one-quarter and one-half.

If there are nonresponses from the list sample, then the estimator $Y_{D3,\,L}$ is computed by

$$\hat{Y}_{D3,\,L} = \sum_{l \in EO} \frac{N_l}{r_l} \sum_{k'=1}^{r_l} y_{lk'}, \tag{2.3}$$

where $y_{lk'}$ = the value of the $k'^{\,th}$ respondent farm in the l^{th} stratum,
r_l = the number of farms responding in the l^{th} stratum.

Remark 2.2: The information on area samples is collected by the enumerators via personal visits. If the person cannot be contacted, the enumerator fills in his or her best assessments which are treated as sampled values. Consequently, no further treatment for nonresponse in the area is used to obtain summary statistics.

The above three estimators are area-oriented. The fourth estimator is list-oriented. A version of it can be written with

$$\hat{Y} = \hat{Y}_{D1,\,A3} + q\,\hat{Y}_{D2,\,A3} + (1-q)\,\hat{Y}_{D2,\,L} + p\,\hat{Y}_{D3,\,A3} + (1-p)\,\hat{Y}_{D3,\,L}, \tag{2.4}$$

where $\hat{Y}_{Di,\,A3}$ denotes the weighted area estimator for domain Di, and $\hat{Y}_{Di,\,L}$ denotes the list estimator for domain Di. The constants p and q are set to zero in the present procedures. Therefore, the fourth estimator, called the multiple frame screening estimator, is given by

$$\hat{Y}_4 = \hat{Y}_{D1,\,A3} + \hat{Y}_{D2,\,L} + \hat{Y}_{D3,\,L} \tag{2.5}$$

The component $\hat{Y}_{D1,\ A3}$ is defined as

$$\hat{Y}_{Di,\ A3} = \sum_{h \varepsilon H} e_h \sum_{k=1}^{n_h} \sum_{l=1}^{g_{hk}} f_{hkl} \frac{a_{hkl}}{b_{hkl}} \delta'_{hkl}, \tag{2.6}$$

where $\delta'_{hkl} = \begin{cases} 1 \text{ if } hkl^{th} \text{ farm } \varepsilon \text{ D1} \\ 0 \text{ otherwise} \end{cases}$

and all the other terms are defined as before.

The component $\hat{Y}_{D2,L}$ is defined as $\hat{Y}_{D3,\ L}$ in equation (2.3) except the summation is over $l \varepsilon EO^c$.

The set EO^c denotes the collection of the list strata which are not the EO strata.

Remark 2.3: The indicator functions δ_{hkl} and δ'_{hkl} in equations (2.2) and (2.6) are used to define the required domain estimators.

Remark 2.4: Discussion of multiple frame methodology can be found in Hartley (1962 and 1974), and in Section 5A.15 of Cochran (1977).

Remark 2.5: The estimators $\hat{Y}_i$, i = 1, 2, or 3, are basically derived from the area frame. However, the list estimator replaces the area estimator for the farmers classified as extreme operators. This perhaps could be interpreted as a robust procedure taken by SRS to reduce the influence of the big farms in the area sample. Further study of robust estimation in surveys is needed.

Remark 2.6: The author questions the desirability of setting p and q to zero, especially in equation (2.4). A major portion of the area information is thrown out.

The variances of the four preliminary estimators used by SRS are as follows:

For i = 1, 2 or 3,

$$\hat{v}_{ii} = \hat{v}(\hat{Y}_i) = \hat{v}(\hat{Y}_{D1\ u\ D2,\ Ai}) + \hat{v}(\hat{Y}_{D3,\ L})$$

$$= \sum_{h \varepsilon H} \frac{e_h (e_h - 1)}{(1 - \frac{1}{r_l})} \sum_{k=1}^{n_h} (y_{i,\ hk} - \bar{y}_{i,\ h.})^2$$

$$+ \sum_{l \varepsilon EO} \frac{N_l (N_l - r_l)}{r_l (r_l - 1)} \sum_{k'=1}^{r_l} (y_{lk'} - \bar{y}_{l.})^2, \tag{2.7}$$

where $\bar{y}_{i,h.} = \sum_{k=1}^{n_h} y_{i,hk} / n_h$ and $\bar{y}_{l.} = \sum_{k'=1}^{r_l} y_{lk'} / r_l.$

$$\hat{v}_{44} = \hat{v}(\hat{Y}_4) = \hat{v}(\hat{Y}_{D1,\,A3}) + \hat{v}(\hat{Y}_{D2\,\cup\,D3,\,L})$$

$$= \sum_{h=1}^{H} \frac{e_h\,(e_h - 1)}{1 - \frac{1}{n_h}} \sum_{k=1}^{n_h} (y'_{3,\,hk} - \bar{y}'_{3,\,h.})^2$$

$$+ \sum_{l \in L} \frac{N_l\,(N_l - r_l)}{r_l\,(r_l - 1)} \sum_{k'=1}^{r_l} (y_{lk'} - \bar{y}_{l.})^2 \qquad (2.8)$$

where $y'_{3,\,hk} = \sum_{l=1}^{g_{hk}} f_{hkl} \frac{a_{hkl}}{b_{hkl}} \delta'_{hkl}$ as in (2.1)

$\bar{y}'_{3,\,h.}$ is defined by $\sum_{k=1}^{n_h} y'_{3,\,hk} / n_h$

and L is the collection of all list strata including EO **strata.**

For the composite estimator developed later, we need the estimators of Cov (Y_1, Y_i), i = 2, 3, or 4, denoted by v_{1i} given as follows:

For i = 2 or 3,

$$\hat{v}_{1i} = \sum_{h \varepsilon H} \frac{e_h \quad (e_h - 1)}{1 - \frac{1}{n_h}} \sum_{k=1}^{n_h} (y_{1,\, hk} - \bar{y}_{1,\, h.}) \quad (y_{i,\, hk} - \bar{y}_{i,\, h.})$$

$$+ \sum_{1 \varepsilon EO} \frac{N_1 \quad (N_1 - r_1)}{r_1 \quad (r_1 - 1)} \sum_{k'=1}^{r_1} (y_{1k'} - \bar{y}_{1.})^2$$

$$\hat{v}_{14} = \sum_{h \varepsilon H} \frac{e_h \quad (e_h - 1)}{1 - \frac{1}{n_h}} \sum_{k=1}^{n_h} (y_{1,\, hk} - \bar{y}_{1,\, h.}) \quad (y'_{3,\, hk} - \bar{y}'_{3,\, h.})$$

$$+ \sum_{1 \varepsilon EO} \frac{N_1 \quad (N_1 - r_1)}{r_1 \quad (r_1 - 1)} \sum_{k'=1}^{r_1} (y_{1k'} - \bar{y}_{1.})^2$$

3. Composite Estimation

In this section, composite estimation is explained and is specialized to the SRS situation. A heuristic argument for composite estimator for the simplest case is given below.

Let us assume there are two independent and unbiased estimators $\hat{Y}_1$ and $\hat{Y}_2$ for the same parameter Y with known variances σ_1^2 and σ_2^2 respectively. Let us propose

$$\hat{Y}_C = c\hat{Y}_1 + (1 - c)\hat{Y}_2,$$

where c is a constant with values between 0 and 1. Then

$$E\hat{Y}_C = Y$$

$$V(\hat{Y}_C) = c^2\sigma_1^2 + (1-c)^2\sigma_2^2. \qquad (3.1)$$

To minimize $V(\hat{Y}_C)$, we should choose c to be

$$c_o = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_1^2}$$

The minimum variance can be obtained from (3.1):

$$V(\hat{Y}_{c_o}) = \frac{\sigma_1^2 \quad \sigma_2^2}{\sigma_1^2 + \sigma_1^2} \qquad (3.2)$$

Note that (3.2) is always smaller than σ_1^2 and σ_2^2. For $\sigma 1 < \sigma 2$, so long as we choose c between $(\sigma_2^2 - \sigma_1^2) / (\sigma_2^2 + \sigma_1^2)$ and 1, we obtain an estimator with smaller variance than $\sigma_1 2$. See Schaible (1987 and 1979) and Royall (1979, pages 85-86) for more discussion on composite estimation.

In general, the variances of $\hat{Y}_1$ and $\hat{Y}_2$ are unknown. However, they can be estimated from the data. The estimated variances are denoted by $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$. Therefore, the composite estimator is given by

$$\hat{Y}_{\hat{c}} = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}\hat{Y}_1 + \frac{\hat{\sigma}_1^2}{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}\hat{Y}_2.$$

Since the weight for the composite estimator is now a function of the data, equation (3.1) can no longer be used to evaluate the variance of the composite estimator. Nevertheless, the variance of the composite estimator can be estimated by sample reuse methods such as jackknife, bootstrap, random group, and balanced repeated replication.

To generalize the above idea to the situation at SRS, let us propose a family of linear combinations of the four preliminary estimators:

$$\hat{Y}_{\underset{\sim}{W}} = \sum_{i=1}^{4} w_i \hat{Y}_i \tag{3.3}$$

where $o \le w_i \le 1$ for all i and $\Sigma w_i = 1$.

We search for the one which minimizes the mean squared errors (MSE's) of the estimators in the linear family. Note that

$$f(\underset{\sim}{W}) = \text{MSE of } \hat{Y}_{\underset{\sim}{W}} = E(\hat{Y}_{\underset{\sim}{W}} - Y)^2$$

$$= \sum_{i=1}^{4} w_i^2 \; E(\hat{Y}_i - Y)^2 + \sum\sum_{i \neq j} w_i w_j \; E(\hat{Y}_i - Y)(\hat{Y}_j - Y) \tag{3.4}$$

where Y denotes the population total.

Since all the second moment terms are unknown, they have to be estimated from the data. The estimation of the second moment terms will be treated in the next section. Let $\hat{m}_i^2$ and $\hat{m}_{ij}$ denote the estimated terms $E(\hat{Y}_i-Y)^2$ and $E(\hat{Y}_i-Y)(\hat{Y}_j-Y)$ respectively. The composite estimator, denoted by $\hat{Y}_{\underset{\sim}{W}_o}$ is derived from minimizing

$$\hat{f}(\underset{\sim}{W}) = \sum_i w_i^2 \; \hat{m}_i^2 + \sum\sum_{i \neq j} w_i w_j \; \hat{m}_{ij}. \tag{3.5}$$

subject to linear constraints $o \le w_i \le 1$, for i=1 to 4, and $\Sigma w_i = 1$.

A further refinement, motivated by the limited translation idea in Efron and Morris (1971, 1972) and Fay and Herriot (1979), is used to derive the final composite estimator. It depends on a "safety factor" K, a positive number specified in advance.

$$\hat{Y}_f = \begin{cases} \hat{Y}_{\underset{\sim}{W}_o} & \text{if } |\hat{Y}_{W_o} - \hat{Y}_1| \le K.SD(\hat{Y}_1), \\ \hat{Y}_1 - K.SD(\hat{Y}_1) & \text{if } \hat{Y}_1 - \hat{Y}_{W_o} > K.SD(\hat{Y}_1), \\ \hat{Y}_1 + K.SD(\hat{Y}_1) & \text{if } \hat{Y}_1 - \hat{Y}_{W_o} < -K.SD(\hat{Y}_1), \end{cases} \tag{3.6}$$

where the estimated standard error $SD(\hat{Y}_1)$ is given by the square root of v_{ii} as in equation (2.7).

This refinement, which limits the amount the composite estimator can deviate from the unbiased estimator, is employed to guard against instability. One can still achieve substantial gain from the composite estimation.

Remark 3.1: A program using Lagrange multipliers and the PROC MATRIX procedure in SAS has been written by the author to solve equation (3.5), a convex programming problem with constraints. See Appendix II for a detailed explanation.

Remark 3.2: A flow chart of the entire SAS program is in Appendix I, and the entire program is given in Appendix III.

Remark 3.3: Another version of a composite estimator can be developed by minimizing the mean squared errors of the estimators:

$$\hat{Y} = \sum_{i=1}^{3} w_{1i} \hat{Y}_{D1,\ Ai} + \sum_{i=1}^{3} w_{2i} \hat{Y}_{D2,\ Ai} + w_{24} \hat{Y}_{D2,\ L} + \sum_{i=1}^{3} w_{3i} \hat{Y}_{D3,\ Ai}$$

$$+ w_{34} \hat{Y}_{D3,\ L}, \qquad (3.7)$$

subject to $\sum_{i=1}^{3} w_{1i} = 1$, $\sum_{i=1}^{4} w_{hi} = 1$ for fixed $h=2$ or 3,

and $o \leq w_{hi} \leq 1$ for all w.

The author has pursued the earlier formulation (3.3) and (3.5) for the following reasons. (1) The solution to equation (3.5) is less sensitive to misclassification errors due to domain determination. (2) It mimics the process used by the CRB. The relative importance of the four preliminary estimators (optimal weights in composite estimation) is of particular interest to the CRB and other statisticians at SRS. (3) The solution is simpler to equation (3.3) than to equation (3.7).

4. Estimation of Second Moments

Development of the estimation of the second moment terms incorporates bias analysis and is discussed in this section.

As is seen from equation (3.5), there are four MSE's and six mixed central moments to be estimated. To estimate these terms, it is assumed:

$$E\,\hat{Y}_1 = Y$$

$$E\,\hat{Y}_2 = Y + b_2\,(Y),$$

$$E\,\hat{Y}_3 = Y + b_3\,(Y),$$

$$E\,\hat{Y}_4 = Y + b_4\,(Y),$$

where $b_i\,(Y)$ denotes the bias of the i^{th} estimator.

All the following identities are used for the estimation procedure.

$$m_1^2 = E\,(\hat{Y}_1 - Y)^2 = V\,(\hat{Y}_1), \qquad (4.1)$$

$$m_i^2 = E\,(\hat{Y}_i - Y)^2 = E\,(\hat{Y}_i - \hat{Y}_1)^2 + 2\,C_{ov}\,(\hat{Y}_1, \hat{Y}_i) - V\,(\hat{Y}_1),$$
for $i=2$, 3, or 4, (4.2)

$$m_{1j} = C_{ov}\,(\hat{Y}_1, \hat{Y}_j), \text{ for } j\neq 1, \qquad (4.3)$$

$$m_{ij} = E\,(\hat{Y}_i - \hat{Y}_1)\,(\hat{Y}_j - \hat{Y}_1) + C_{ov}\,(\hat{Y}_1, \hat{Y}_i) + C_{ov}\,(\hat{Y}_1, \hat{Y}_j) - V\,(\hat{Y}_1)$$
for i and $j\neq 1$. (4.4)

It is straightforward to verify these identities. For example, for $i\neq 1\neq j$

$$m_{ij} = E\,(\hat{Y}_i - Y)\,(\hat{Y}_j - Y)$$

$$= E\,(\hat{Y}_i - \hat{Y}_1 + \hat{Y}_1 - Y)\,(\hat{Y}_j - \hat{Y}_1 + \hat{Y}_1 - Y)$$

$$= E\,(\hat{Y}_i - \hat{Y}_1)\,(\hat{Y}_j - \hat{Y}_1) + E\,(\hat{Y}_i - Y + Y - \hat{Y}_1)\,(\hat{Y}_1 - Y)$$

$$+ E\,(\hat{Y}_j - Y + Y - \hat{Y}_1)\,(\hat{Y}_1 - Y) + V_1\,\hat{Y}_1)$$

$$= E\,(\hat{Y}_i - \hat{Y}_1)\,(\hat{Y}_j - \hat{Y}_1) + C_{ov}\,(\hat{Y}_i, \hat{Y}_1) + C_{ov}\,(\hat{Y}_j, \hat{Y}_1) - V(\hat{Y}_1).$$

Using identities (4.1) - (4.4), unbiased estimates of the mixed central moment terms, and refinements over the unbiased estimates of the MSE terms can be obtained as follows.

$$\hat{m}_1^2 = v_{11}, \tag{4.5}$$

$$\hat{m}_i^2 = \max \left\{ (\hat{Y}_i - \hat{Y}_1)^2 + 2\, v_{1i} - v_{11}, \quad v_{ii} \right\} \text{ for } i = 2, 3 \text{ or } 4, \tag{4.6}$$

$$\hat{m}_{1j} = v_{1j}, \text{ for } j \neq 1 \tag{4.7}$$

$$\hat{m}_{ij} = (\hat{Y}_i - \hat{Y}_1)(\hat{Y}_j - \hat{Y}_1) + v_{1i} + v_{1j} - v_{11}, \text{ for } i, j \neq 1, \tag{4.8}$$

where v_{ij}'s are given in Section 2.

The maximum function in $\hat{m}_i^2$ is employed to ensure that the estimators for the bias squared terms are nonnegative.

Remark 4.1: Equation (4.2) without the covariance terms has been used by Brock, French, and Peyton (1980) to estimate the MSE's for independent estimators. Equation (4.2) has been used by Cohen and Sommers (1984) to estimate the MSE's of regional mean expenditure and composite estimators.

Remark 4.2: Equation (4.6) enables us to obtain an unbiased estimate of the bias squared term b_i^2 (Y), for $i \neq 1, \ldots, 4$. A refinement over this unbiased estimate is given by $\hat{b}_i^2 = \max \left\{ (\hat{Y}_i - \hat{Y}_1)^2 + 2\, v_{1i} - v_{11} - v_{ii}, 0 \right\}$.

5. Variance and Mean Squared Error Evaluation of The Composite Estimator

The heuristic argument for using composite estimation has been given. The variance and mean squared error estimates for the composite estimator are needed to justify the gain in using composite estimation. The jackknife method is used to estimate the variance and mean squared error. This method is adopted because of its simplicity of explanation and ease of programming. See Efron (1982), Wolter (1985) for excellent expositions on sample reuse methods.

Assume the data are divided into g independent groups. Let $Y\,(i)$ be an estimator derived from the data with i^{th} group deleted. The i^{th} pseudo-value of Y is defined to be $\hat{Y}^*\,(i) = g\hat{Y} - (g-1)\,Y\,(i)$, where $\hat{Y}$ is the estimator based on the full sample.

The jackknife estimator of the variance of Y is given by

$$v_J\,(\hat{Y}_Q) = \frac{1}{g(g-1)} \sum_{i=1}^{g} (Y^*_{(i)} - \bar{Y}^*)^2 \tag{5.1}$$

where $\bar{Y}^* = \sum_{i=1}^{g} Y^*_{(i)}/g.$

If $\ddot{Y}$ is an estimator other than $\hat{Y}_1$, then the mean squared error of $\hat{Y}$ can also be estimated by the jackknife method.

$$mse_J\,(\dot{Y}) = (\dot{Y} - \hat{Y}_1)^2 + \frac{2}{g(g-1)} \sum_{i=1}^{g} (Y^*_{(i)} - \bar{Y}^*)\,(Y^*_{1(i)} - \bar{Y}^*_1) \tag{5.2}$$

$$- \frac{1}{g(g-1)} \sum_{i=1}^{g} (Y^*_{(i)} - \bar{Y}^*_1)^2, \text{ or}$$

$$mse_{AJ}\,(\dot{Y}) = \frac{1}{g} \sum_{i=1}^{g} (Y_{(i)} - \hat{Y}_{1(i)})^2 + \frac{2}{g(g-1)} \sum_{i=1}^{g} (Y^*_{(i)} - \bar{Y}^*)\,(Y^*_{1(i)} - \bar{Y}^*)$$

$$- \frac{1}{g(g-1)} \sum_{i=1}^{g} (Y^*_{1(i)} - \bar{Y}^*_1)^2 \tag{5.3}$$

6. Numerical Results

The data are from the 1984 June Enumerative Survey conducted by SRS. Six States are selected from the 10 major hog States which account for about 79 percent of the U.S. hogs and pigs inventory (Crop Reporting Board 1984). Described below are summary statistics prepared for each of the six States (Tables 1-6).

Seven estimates, denoted by $\hat{Y}_i$, i = 1 .., 7, are given for the total hogs and pigs inventory. The estimates $\hat{Y}_i$, i=1, .., 4 are the tract, farm, weighted, and multiple frame screening estimates defined before. The estimate $\hat{Y}_5$ is the composite estimate defined by (3.6) for any $K > 1$. The estimate $\hat{Y}_6$ is derived similarly to $\hat{Y}_5$ except by setting $W_2 = W_3 = 0$. In other words, $\hat{Y}_6$ is the composite estimate by combining just the tract and the multiple frame screening estimators. The estimate $\hat{Y}_7$ denotes the official CRB statistics published in the Livestock Series: Hogs and Pigs (Crop Reporting Board 1984). The optimal weights (denoted by W) for the components of $\hat{Y}_5$ derived from equation (3.5) are given in the table. The optimal weights for $\hat{Y}_6$ are denoted by $\hat{W}$.

All the standard errors and root mean squared errors of the four preliminary estimators are also estimated from equations (2.7), (2.8), (4.5), and (4.6) by taking square roots of $\hat{v}_{ii}$ and $\hat{m}_i^2$. The reader should note the root mean squared errors published by the Statistical Reporting Service in their Crop Reporting Board official reports are computed differently. These estimates are given in the tables (denoted by SD_i and $\sqrt{MSQ_i}$). Two estimates of bias of $\hat{Y}_i$, i=2, 3, 4 are given. One is obtained from Remark 4.2, i.e, $\hat{b}_i$. A second estimate is an unbiased estimate of the bias, i.e., $\hat{\hat{b}}_i = \hat{Y}_i - \hat{Y}_1$.

For the variance evaluations of the composite estimator, the author has only completed Indiana and Minnesota.

Due to different frame constructions by SRS of the replication codes of the area and list samples, formulations of the groups of data for the jackknife method are slightly different between the area and the list sample. The replication codes in the area sample which usually run from 1 to 10 or 1 to 5 for each land use stratum are used. A land use stratum defined at the beginning of Section 2 is a collection of paper strata. The replication codes for each list stratum were generated by the author using random numbers. Several (3 or 4) replications are constructed for each list stratum.

The i^{th} $(1 \leq i \leq d)$ where $d \leq g$ jackknife estimate is computed by deleting the i^{th} replicate of each land use stratum (i.e. deleting each segment from each paper stratum in the same land use stratum). The expansion factor e_h is adjusted by multiplying the number of replicates/(number of replicates - 1) in each land use stratum. The number d is the total number of replicates for all land use strata.

The i^{th} $(d + 1 \leq i \leq g)$ jackknife estimate is computed by deleting each replicate from each list stratum sequentially. The adjustment on the expansion factor is automatically obtained by using equation (2.3) where r_e and $y_{ek'}$ are obtained after deleting the i^{th} subgroup from the data No data are deleted from the self-representing stratum (the largest EO stratum).

The numbers d and g for Indiana are 31 and 55, i.e. the area sample is divided into 31 groups, the nonself-representing list sample is divided into 24 approximately independent groups (6 strata with 4 groups each). The numbers d and g for Minnesota are 30 and 60. The area sample for Minnesota is divided into 30 independent groups. The list sample is divided into 30 approximately independent groups (30 groups deriving from 10 strata with 3 groups in each stratum).

The jackknife method can be used for any estimators from probability surveys. Therefore, for each of the estimators Y_i, $i=1, \ldots, 6$, we compute its variance and mean squared error estimates by using equations (5.1), (5.2) or (5.3) with $\hat{Y}$ replaced by $\hat{Y}_i$. Empirical results reveal that there are no big differences between (5.2) and (5.3). Therefore, equation (5.2) is used for estimating mean squared error. Empirical evaluations for the two States also reveal that the variance estimates for the composite estimators are more sensitive to outliers from the pseudo-values. Consequently, the Winsorized variance estimates and mean squared error estimates (Winsorized methods are applied to the covariance and variance terms in equation (5.2)) are used here in tables 1 and 2. They are denoted by $SDJKR_i$ and $MSQJKR_i$ $i=1, \ldots, 6$. When $i=1, \ldots, 4$, the quantities $SDJKR_i$ and $\sqrt{MSQJKR_i}$ can be compared to SD_i and $\sqrt{MSQ_i}$, which are computed from the full sample using the stratified design, to determine the goodness of the variance estimates by the jackknife method. The discrepancy can be explained by the variances of the jackknife variance estimator. The large operators in the area sample, not classified in the D3 domain, are a major cause of this discrepancy.

Discussion of the Winsorized method can be found in Huber (1981, p 151) and Elashoff and Elashoff (1978). Ten percent from each end of the pseudo-values of Indiana's data and 15 percent from that of Minnesota are Winsorized to obtain the variance estimates. The same percentages are also used for the covariance estimate used in equation (5.2), where the terms $(Y^*(i) - \bar{Y}^*_w)(Y^*_1(i) - \bar{Y}^*_{1,w})$ are Winsorized with $\bar{Y}^*_w$ and $\bar{Y}^*_{1,w}$ denoting the Winsorized mean of $\bar{Y}^*$ and $\bar{Y}^*_1$.

Due to the time constraints of the fellowship, the author did not explore other jackknife and sample reuse methods to obtain more satisfactory variance and mean squared error estimate of the composite estimator. However, the numerical results in each of the Tables present enough evidence to show that the composite estimator performs very well. Examining the mean squared errors and mixed moments of the preliminary estimators, it can be seen that the composite estimate is very effective in selecting the desirable components, i.e. the components with small mean squared errors or with negative correlations.

Five summary points, $\hat{Y}_i$, $Y_i \pm SD_i$, and $\hat{Y}_i \pm 2\,SD_i$ for the preliminary estimates are plotted in Figures 1 to 6 for each State. The two estimates $\hat{Y}_5$ (composite) and $\hat{Y}_7$ (CRB) are also plotted in the last column. These schematic plots spell out the necessity of analyzing the biases in some of the preliminary estimators. Among all the approximately unbiased estimates, the composite estimate is always quite close to the preliminary one with smallest variance. Hence, composite estimators $\hat{Y}_5$ should perform better than all the other estimates.

The weighted estimators (equation (2.6)) are used for the nonoverlap domain in the present multiple frame screening estimator. However, it would be simpler to consider the tract estimator for the nonoverlap domain. The term $f_{hkl} \cdot a_{hkl}/b_{hkl}$ is replaced by t_{hkl} in equation (2.6). The resulting multiple frame estimator denoted by $\hat{Y}'_4$ is defined by $\hat{Y}_{D1,A1} + \hat{Y}_{D2,L} + \hat{Y}_{D3,L}$. Consequently, composite estimators can also be considered by combining $\hat{Y}_1$, $\hat{Y}_2$, $\hat{Y}_3$, and $\hat{Y}_4$, or just $\hat{Y}_1$ and $\hat{Y}'_4$. The numerical results for this case are tabulated in the second half of each table. The only changes from the first half would be all the related parameter estimates from $\hat{Y}'_4$. These changes are given in the tables. The terms which have not been changed are left blank in the tables. The terms with bar "__" have not been evaluated.

The motivation for considering $\hat{Y}'_4$ instead of $\hat{Y}_4$ is to reduce biases and respondents' burden. It is seen from the tables that the bias of $\hat{Y}'_4$ is less than that of $\hat{Y}_4$ in four out of the six States. The composite estimator combining $\hat{Y}_1$ and $\hat{Y}_4$ requires only information on the tract inventories and information from the list sample.

The results from Iowa are particularly interesting. They reveal that the optimal composite estimator is formed by combining just the tract and multiple frame screening estimators. The use of $\hat{Y}'_4$ is considerably better than $\hat{Y}_4$ in terms of mean squared errors. If there were enough evidence showing the same behavior over the years, the alternative of deleting the farm and weighted estimators could be considered. Both cost saving and improved accuracy of the estimate (by reducing respondent burden) can be expected for this alternative procedure.

It is also interesting to note that the estimator $w_1 \hat{Y}_1 + w_4 \hat{Y}_4$ (without the outlier adjustment as discussed in Remark 2.5) is equivalent to $\hat{Y}_{D1, A1} + w_1 \hat{Y}_{D2 \cup D3, A1} + w_4 \hat{Y}_{D2 \cup D3, L}$. This latter estimator is exactly the full multiple frame estimator proposed by Hartley (1962). Hartley proposes to choose optimal w's from the design information. The optimal w's in this paper are derived from current data.

The farm estimator has traditionally been considered as an inferior estimator by the statisticians at SRS. This judgment has also been confirmed in the analysis. The farm estimator is only used in two of six States with less than 25 percent weight in composite estimation. The present multiple frame screening estimator (which has been considered from its variance evaluation to be a superior estimator by SRS) is shown to have nonnegligible bias. The causes for this bias has been discussed by Nealon (1984) such as misclassification errors due to domain determination, imputation techniques for nonresponses in the list sample, etc. The bias component of the multiple frame screening estimator has reduced its weight in composite estimation.

7. Conclusion and Suggestions for Further Research

This paper discusses a composite estimation methodology which combines the different preliminary estimators used by SRS into one by minimizing the mean squared errors of the combined estimators. Some numerical results from the 1984 June Enumerative Survey are presented. These results support the use of composite estimation. It is recommended that SRS incorporate composite estimation methodology into their Crop Reporting Board procedures.

Although limited numerical results are presented here, SRS can easily apply the technique and the computer program to other commodities, to other States (including the States with only two or three preliminary estimators), and to data from past years to gain insight on the performance of the preliminary estimators. Numerical results may also suggest whether or not to delete some of the less useful estimators.

Research on the following topics is recommended:

(1) composite estimation for second stage sampling used in the December Enumerative Survey,
(2) research on variance evaluation (The recent work on sample reuse methods of Bickel and Freedman (1984), and Rao and Wu (1984) could be explored.),
(3) research on sample reuse methods for multiple frame sampling,
(4) robust estimation for finite population sampling (See Varceman and Meeden (1983) for some research results on trimmed and Winsorized estimators for finite populations.), and
(5) empirical Bayes methodology (This methodology incorporates past data to improve upon the composite estimators. See Robbins (1983) for discussions and ideas.).

8. References

Allen, R., Clampet, J., Dunkerley, C., Tortora, R., and Vogel, F. (1983), "Framework for the Future," Statistical Reporting Service, U.S. Department of Agriculture.

Avriel, M. (1976), "Nonlinear Programming, Analysis and Methods," Prentice Hall.

Bickel, P.J. and Freedman, P.A. (1984). "Asymptotic Normality and the Bootstrap in Stratified Sampling," Annals of Statistics, 12, 470-482.

Bosecker, R. and Ford, B. (1976), "Multiple Frame Estimation with Stratified Overlap Domain," Statistical Reporting Service, U.S. Department of Agriculture.

Brock, D.B., French, D.K., and Peyton, B.W. (1980), "Small Area Estimation: Empirical Evaluation of Several Estimators for Primary Sampling Units," Proceedings, Section on Survey Research Methods, American Statistical Association, 766-771.

Brown, L.D., and Cohen, A. (1974), "Point and Confidence Estimation of a Common Mean and Recovery of Interblock Information," Annals of Statistics, 2, 963-976.

Bynum, H., Dowdy, W., Hanuschak, G., Hudson, C., Murphy, R., Steinberg, J., and Vogel, F. (1985), "Crop Reporting Board Standards," Statistical Reporting Service, U.S. Department of Agriculture.

Cochran, W. G. (1977), Sampling Techniques, third edition, New York, Wiley.

Cohen, M.P., and Sommers, J.P. (1984), "Evaluation of Methods of Composite Estimation of Cost Weights for the CPI," Proceedings, Section on Business and Economic Statistics, American Statistical Association, 466-471.

Crop Reporting Board (1979), Hog and Pig Reports: A Handbook on Surveying and Estimating Procedures, Economics, Statistics, and Cooperative Service, U.S. Department of Agriculture.

Crop Reporting Board (1981), Cattle Reports: A Handbook on Surveying and Estimating Procedures, Statistical Reporting Service, U.S. Department of Agriculture.

Crop Reporting Board (1984), Livestock Series, Hogs and Pigs, Statistical Reporting Service, U.S. Department of Agriculture.

Efron, B., and Morris, C. (1971), "Limiting the Risk of Bayes and Empirical Bayes Estimators--Part I: The Bayes Case," Journal of the American Statistical Association, 66, 807-815.

Efron, B., and Morris, C. (1972), "Limiting the Risk of Bayes and Empirical Bayes Estimators--Part II: The Empirical Bayes Case," Journal of the American Statistical Association, 67, 117-130.

Efron, B. (1982), The Jackknife, the Bootstrap and Other Resampling Plans, Philadelphia, Society for Industrial and Applied Mathematics.

Elashoff, J. and Elashoff, R., (1978) "Errors: Effects of Errors in Statistical Assumptions," International Encyclopedia of Statistics, eds, Kruskal, W.H. and Tanur, J.M., New York, Macmillan.

Fay, R. E. III, and Herriot, R. A. (1979), "Estimates of Income in Small Places: An Application of James-Stein Procedures to Census Data," Journal of the American Statistical Association, 74, 269-277.

Geuder, J. (1984), "Paper Stratification in SRS Area Sampling Frames," Statistical Reporting Service, U.S. Department of Agriculture.

Graybill, F. A. and Deal, R. B. (1959), "Combining Unbiased Estimators," Biometrics 15, 543-550.

Hartley, H. O. (1962), "Multiple Frame Surveys," Proceedings, Section on Social Statistics, American Statistical Association, 203-206.

__________(1974), "Multiple Frame Methodology and Selected Application," Sankhya, 99-118.

Houseman, E. (1971), "Composite Estimation," unpublished memo. Statistical Reporting Service, U.S. Department of Agriculture.

__________(1975), "Area Frame Sampling in Agriculture," Statistical Reporting Service, U.S. Department of Agriculture.

Huber, P.J. (1981), Robust Statistics, New York, Wiley.

Mosteller, F. (1948), "On Pooling Data," Journal of the American Statistical Association, 43, 231-242.

Nealon, J. (1984), "Review of the Multiple and Area Frame Estimators," Statistical Reporting Service, U.S. Department of Agriculture.

Rao, J.N.K. and Wu, C.F.J. (1984), "Bootstrap Inference for Sample Surveys," Proceedings, Section on Survey Research Methods, American Statistical Association, 106-112.

Robbins, H. (1983), "Some Thoughts on Empirical Bayes Estimation," Annals of Statistics, 11, 713-723.

Royall, R. M. (1979), "Prediction Models in Small Area Estimation," in Synthetic Estimators for Small Areas: Statistical Workshop Papers and Discussion, NIDA Research Monograph 24, Washington: U.S. Government Printing Office, 63-87.

Schaible, Wesley J. (1978), "Choosing Weights for Composite Estimators for Small Area Statistics," Proceedings, Section on Survey Research Methods, American Statistical Association, 741-746.

Schaible, Wesley J. (1979), "A Composite Estimator for Small Areas (with discussion)," in Synthetic Estimators for Small Areas: Statistical Workshop Papers and Discussion, NIDA Research Monograph 24, Washington: U.S. Government Printing Office, 36-62.

Statistical Reporting Service (1984), SRS National Conference Proceedings, U.S. Department of Agriculture.

Statistical Reporting Service (1984), June Supervising and Editing Manual, U.S. Department of Agriculture.

Vardeman, S. and Meeden, G. (1983), "Admissible Estimators of the Population Total Using Trimming and Winsorization," Statistics and Probability Letters, 1, 317-321.

Wolter, K. (1979), "Composite Estimation in Finite Populations," Journal of the American Statistical Association, 74, 604-613.

Wolter, K. (1985), Variance Estimation, New York, Springer-Verrlag.

9. Table Captions

Each of six major hog producing States have numerical results from the 1984 JES summarized for comparative purposes (table 1-6). Each table has two sections. Results using the weighted nonoverlap domain estimator (2.6) in the multiple frame screening estimator are presented in the top half of each table. An alternate method which uses the tract nonoverlap domain estimator is presented in the lower half of each table. Seven estimates ($\hat{Y}_i$) are described. The notation is given below for the estimators and summary statistics.

Notation	Description
$\hat{Y}_i$	i=1 tract estimate (2.2) i=2 farm estimate (2.2) i=3 weighted estimate (2.2) i=4 multiple frame screening estimate (2.6) i=5 full composite estimate (3.6) i=6 composite estimate using i=1, 4 i=7 official Crop Reporting Board estimate
$\hat{W}_i$	optimal weights for the components of Y_5
$\check{W}_i$	optimal weights for the components of Y_6
SD_i	standard error estimate of $\hat{Y}_i$, i=1, ..., 4
$SDJKR_i$	jackknife standard error estimate of Y_i, i=1, ..., 6
MSQ_i	root mean squared error estimate of Y_i, i=1, ..., 4
$MSQJKR_i$	root mean squared error jackknife estimate of Y_i, i=1, ..., 6
$\grave{b}_i$	estimate of bias (Remark 4.2)
$\hat{\hat{b}}_i$	unbiased estimate of bias ($\dot{Y}_i$ - $\dot{Y}_1$)
$\hat{M}_{ij}$	mean squared error matrix of the four estimators

Variance estimates are computed for composite estimators (i=5, 6) in only two States--Indiana and Minnesota (see tables 1 and 2).

Numerical results using the tract nonoverlap domain estimator is presented in the bottom half of each table. Entries in the table reflect changes when the multiple frame screening estimator ($\hat{Y}'_4$) uses the tract nonoverlap domain estimator. Jackknife variance estimates are not calculated.

Table 1. Summary statistics for Indiana by estimator $\hat{Y}_i$, 1984 June Enumerative Survey 1/

i	$\hat{Y}_i$ (1,000)	$\hat{W}_i$	$\hat{\hat{W}}_i$	SD_i	$SDJKR_i$	$\sqrt{MSQ_i}$	$\sqrt{MSQJKR_i}$	$\hat{b}_i$	$\hat{\hat{b}}_i$
Wtd. NOL Domain Estimator:									
1	3,367	0	0.8293	412,634	349,212	412,634	349,212	0	0
2	3,797	0		470,456	506,901	587,733	568,014	352,280	429,977
3	3,616	1		249,265	204,888	249,265	204,888	0	249,070
4	4,331	0	0.1707	178,276	183,289	884,276	908,925	866,119	963,743
5	3,616				253,425		279,298		
6	3,532				476,814		484,162		
7	4,300								
Tract NOL Domain Estimator:									
1		0	0.7519						
2		0							
3		1							
4	4,141	0	0.2481	185,946	___	686,851	___	661,202	773,647
5	3,616				___		___		
6	3,559				___		___		
7									

1/ Reference page 24 for a description of table captions.

Table 2. Summary statistics for Minnesota by estimator $\hat{Y}_i$, 1984 June Enumerative Survey 1/

i	$\hat{Y}_i$ (1,000)	$\dot{W}_i$	$\ddot{W}_i$	SD_i	$SDJKR_i$	$\sqrt{MSQ_i}$	$\sqrt{MSQJKR_i}$	$\dot{b}_i$	$\ddot{b}_i$
Wtd. NOL Domain Estimator:									
1	4,899	0	0.6516	698,755	550,851	698,755	550,851	0	0
2	5,226	0.2326		715,786	765,496	753,163	765,496	234,318	327,512
3	4,645	0.5630		394,873	483,465	394,873	483,465	0	-253,687
4	3,753	0.2044	0.3484	236,931	278,648	941,385	1,054,340	911,081	1,145,488
5	4,598				665,203		709,733		
6	4,500				685,620		777,168		
7	3,870								
Tract NOL Domain Estimator:									
1		0	0.6543						
2		0.2190							
3		0.5905							
4	3,768	0.1905	0.3457	250,228	______	939,799	______	905,874	1,131,409
5					______		______		
6					______		______		
7									

1/ Reference page 24 for a description of table captions.

Table 3. Summary Statistics for Iowa by estimator $\hat{Y}_i$, 1984 June Enumerative Survey 1/

i	$\hat{Y}_i$ (1,000)	$\hat{W}_i$	$\hat{\hat{W}}_i$	SD_i	$\sqrt{MSQ_i}$	$\hat{b}_i$	$\hat{\hat{b}}_i$
Wtd. NOL Domain Estimator:							
1	12,674	0.6132	0.6132	1,022,345	1,022,345	0	0
2	13,709	0		1,320,368	1,464,836	634,329	1,035,081
3	14,703	0		1,076,946	2,114,770	1,820,011	2,028,578
4	14,149	0.3868	0.3868	662,858	1,240,187	1,048,181	1,475,235
5	13,245						
6	13,245						
7	13,800						
Tract NOL Domain Estimator:							
1		0.3774	0.3774				
2		0					
3		0					
4	13,786	0.6226	0.6226	723,831	861,289	466,784	1,112,458
5	13,367						
6	13,367						
7							

1/ Reference page 24 for a description of table captions.

Table 4. Summary Statistics for Kansas by estimator $\hat{Y}_i$, 1984 June Enumerative Survey 1/

i	$\hat{Y}_i$ (1,000)	$\hat{W}_i$	$\check{W}_i$	SD_i	$\sqrt{MSQ_i}$	b_i	$\dot{b}_i$
Wtd. NOL Domain Estimator:							
1	1,418	0.1382	0.2421	192,280	192,280	0	0
2	1,669	0		262,756	281,738	101,664	250,868
3	1,656	0.1699		172,686	228,935	150,302	238,514
4	1,485	0.6919	0.7579	125,479	125,479	0	67,031
5	1,505						
6	1,469						
7	1,485						
Tract NOL Domain Estimator:							
1		0	0.1604				
2		0					
3		0.3181					
4	1,397	0.6819	0.8396	133,601	133,601	0	20,858
5	1,479						
6	1,400						
7							

1/ Reference page 24 for a description of table captions.

Table 5. Summary Statistics for Missouri by estimator $\hat{Y}_i$, 1984 June Enumerative Survey 1/

i	$\hat{Y}_i$ (1,000)	$\hat{W}_i$	$\tilde{W}_i$	SD_i	$\sqrt{MSQ_i}$	$\hat{b}_i$	$\hat{\hat{b}}_i$
Wtd. NOL Domain Estimator:							
1	3,960	0	0	796,198	796,198	0	0
2	3,915	0		782,721	782,721	0	44,750
3	3,859	0.4627		380,727	380,727	0	100,198
4	3,418	0.5373	1	305,689	305,689	0	542,118
5	3,622						
6	3,418						
7	3,400						
Tract NOL Domain Estimator:							
1		0	0.0264				
2		0					
3		0.6385					
4	3,375	0.3615	0.9736	554,084	554,084	0	584,618
5	3,684						
6	3,390						
7							

1/ Reference page 24 for a description of table captions.

Table 6. Summary Statistics for Ohio by estimator $\hat{Y}_i$, 1984 June Enumerative Survey 1/

i	$\hat{Y}_i$ (1,000)	$\hat{W}_i$	$\tilde{W}_i$	SD_i	$\sqrt{MSQ_i}$	$\hat{b}_i$	$\hat{\hat{b}}_i$
Wtd. NOL Domain Estimators:							
1	1,326	0.7623	0.9037	205,630	205,630	0	0
2	1,547	0.2377		267,451	267,451	0	221,728
3	1,710	0		175,623	392,840	351,397	384,2751
4	1,848	0	0.0963	150,652	512,190	489,533	522,051
5	1,378						
6	1,376						
7	1,800						
Tract NOL Domain Estimator:							
1		0.7623	0.9022				
2		0.2377					
3		0					
4	1,756	0	0.0978	173,494	437,467	401,593	430,517
5	1,378						
6	1,368						
7							

1/ Reference page 24 for a description of table captions.

10. Figure Captions

Figure 0:

This chart illustrates some information available to the Crop Reporting Board at the State, regional and national levels for different estimators. The estimators are plotted over time for certain commodities of interest. Varied check data are also available to the Board during their review. The plot is for illustration purposes only and does not represent actual relationships among estimators. The estimators shown are $\hat{Y}_i$, where i=1,..,4, and are the tract, farm, weighted and multiple frame screening estimates respectively. The official Board level is noted as $\hat{Y}_7$.

Figures 1-6:

These are schematic plots by State. Five summary points, $\hat{Y}_i$, $\hat{Y}_i + SD_i$, and $\hat{Y}_i +2\ SD_i$ are plotted vertically for the tract ($\hat{Y}_1$), farm ($\hat{Y}_2$), weighted ($\hat{Y}_3$) and multiple frame screening ($\hat{Y}_4$) estimates. The full composite estimate ($\hat{Y}_5$) and the Crop Reporting Board estimate ($\hat{Y}_7$) are also given.

Figure 0: Total number of hogs and pigs
v.s. years of a State

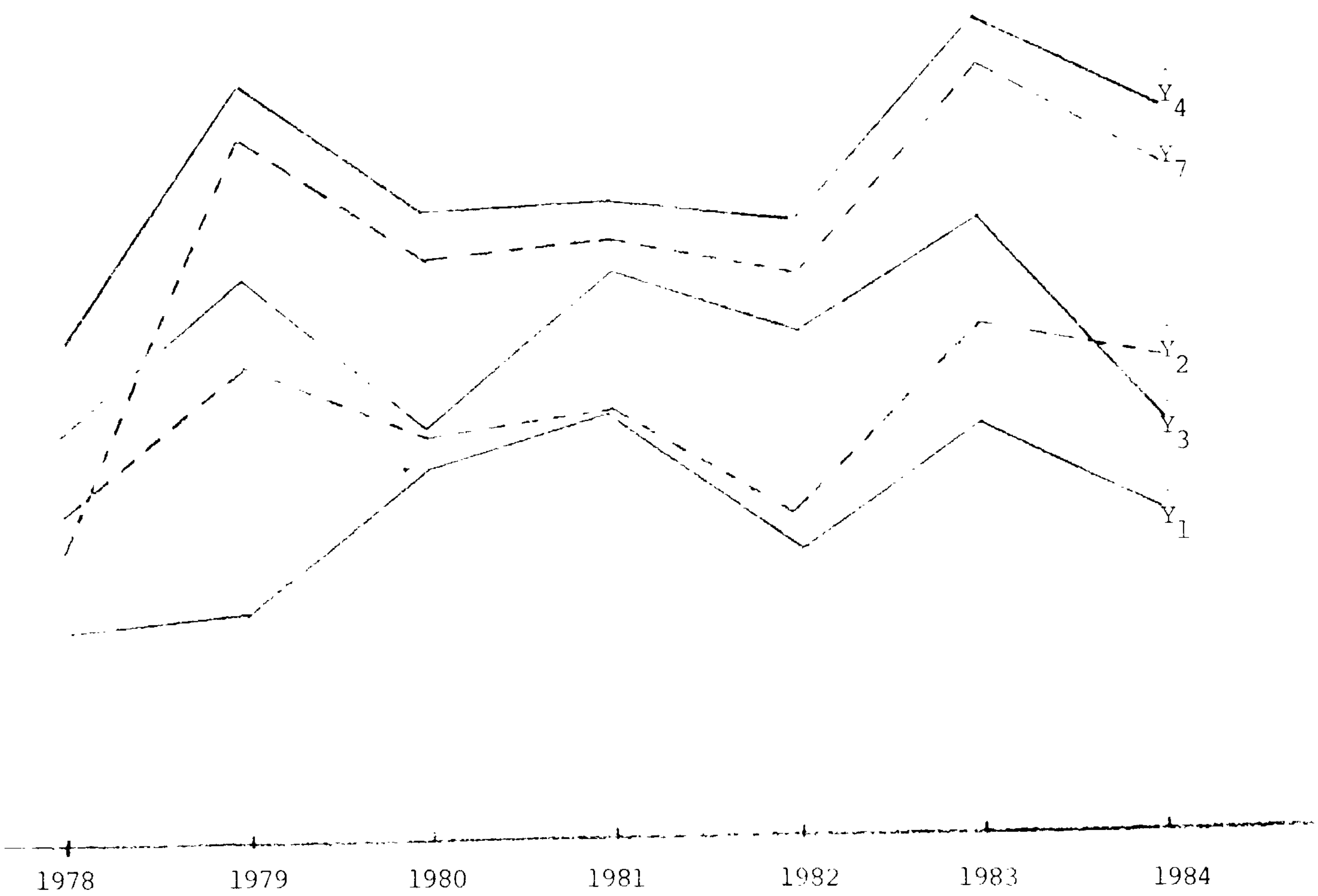
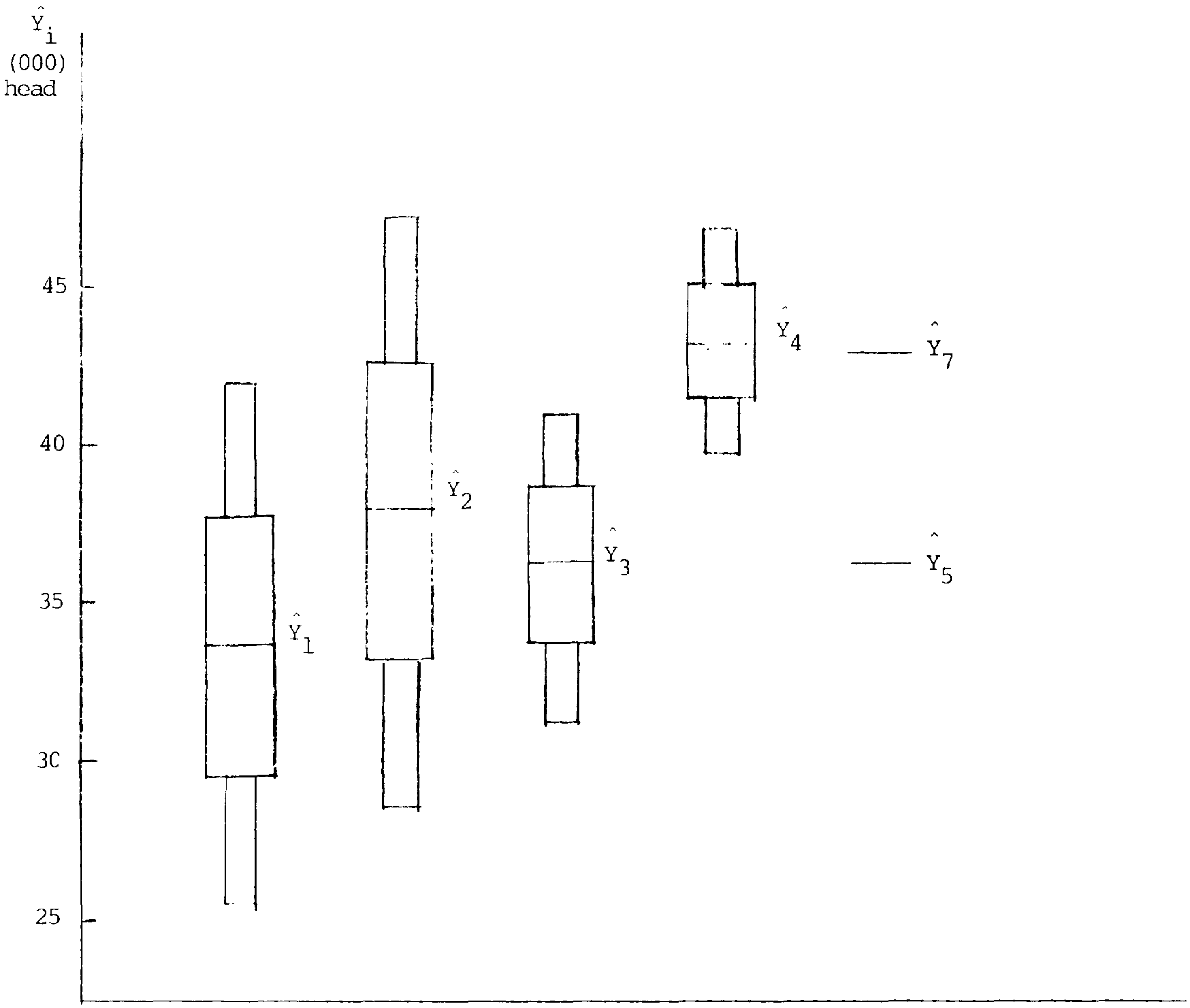
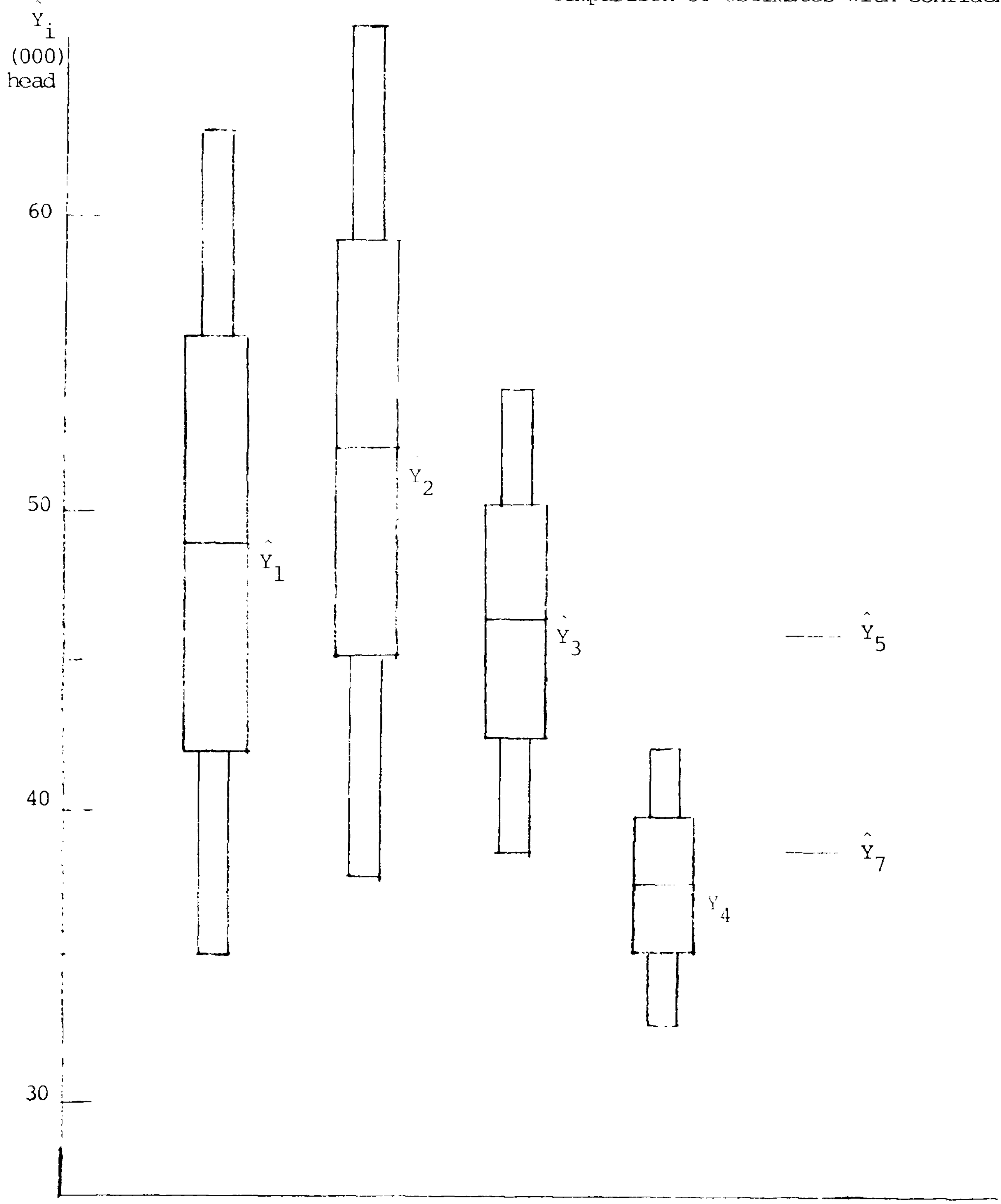


Figure 1: Schematic plot for Indiana 1/
Comparison of estimates with confidence limits



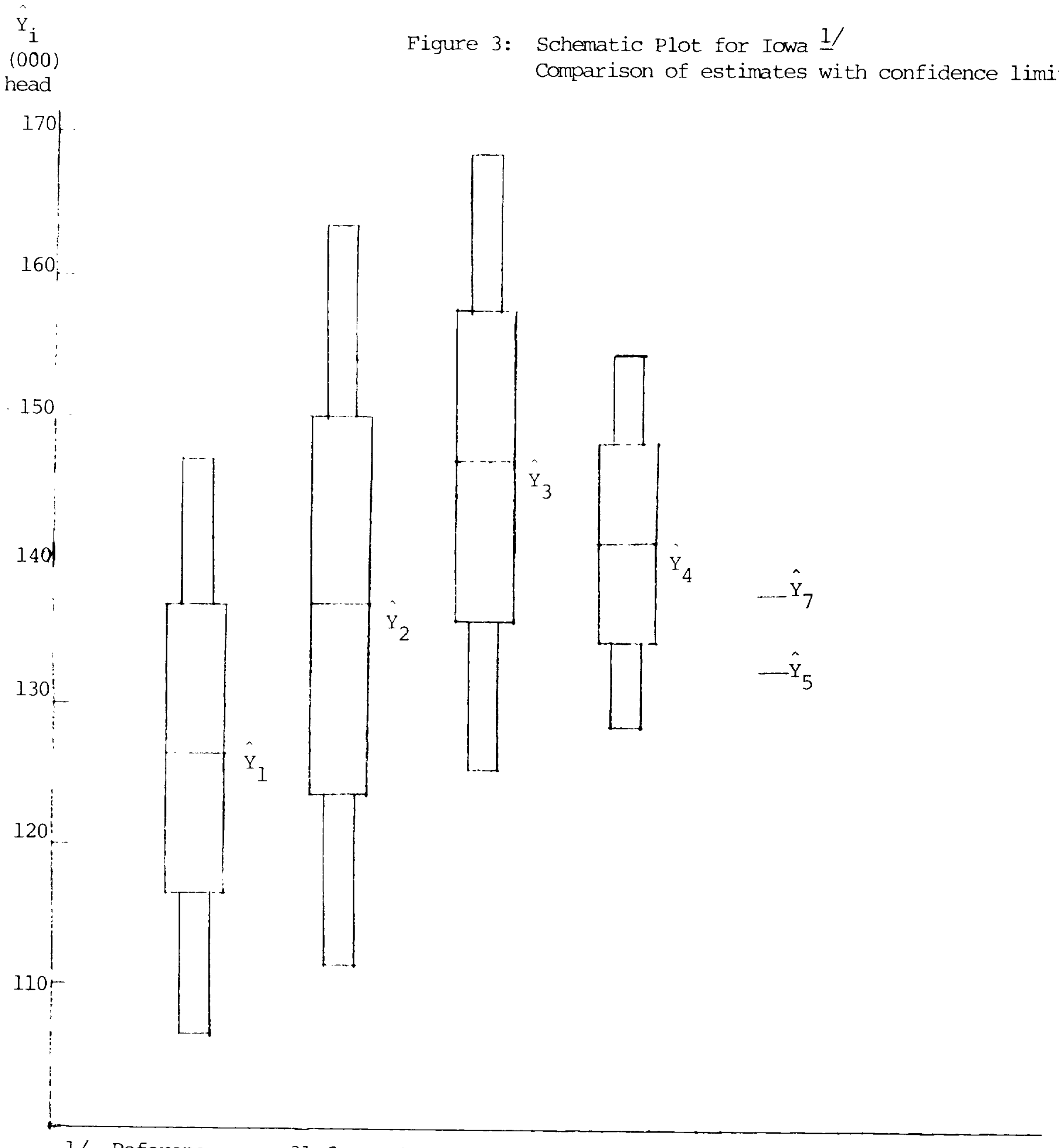
1/ Reference page 31 for a description of plot.

Figure 2: Schematic Plot for Minnesota 1/
Comparison of estimates with confidence limits



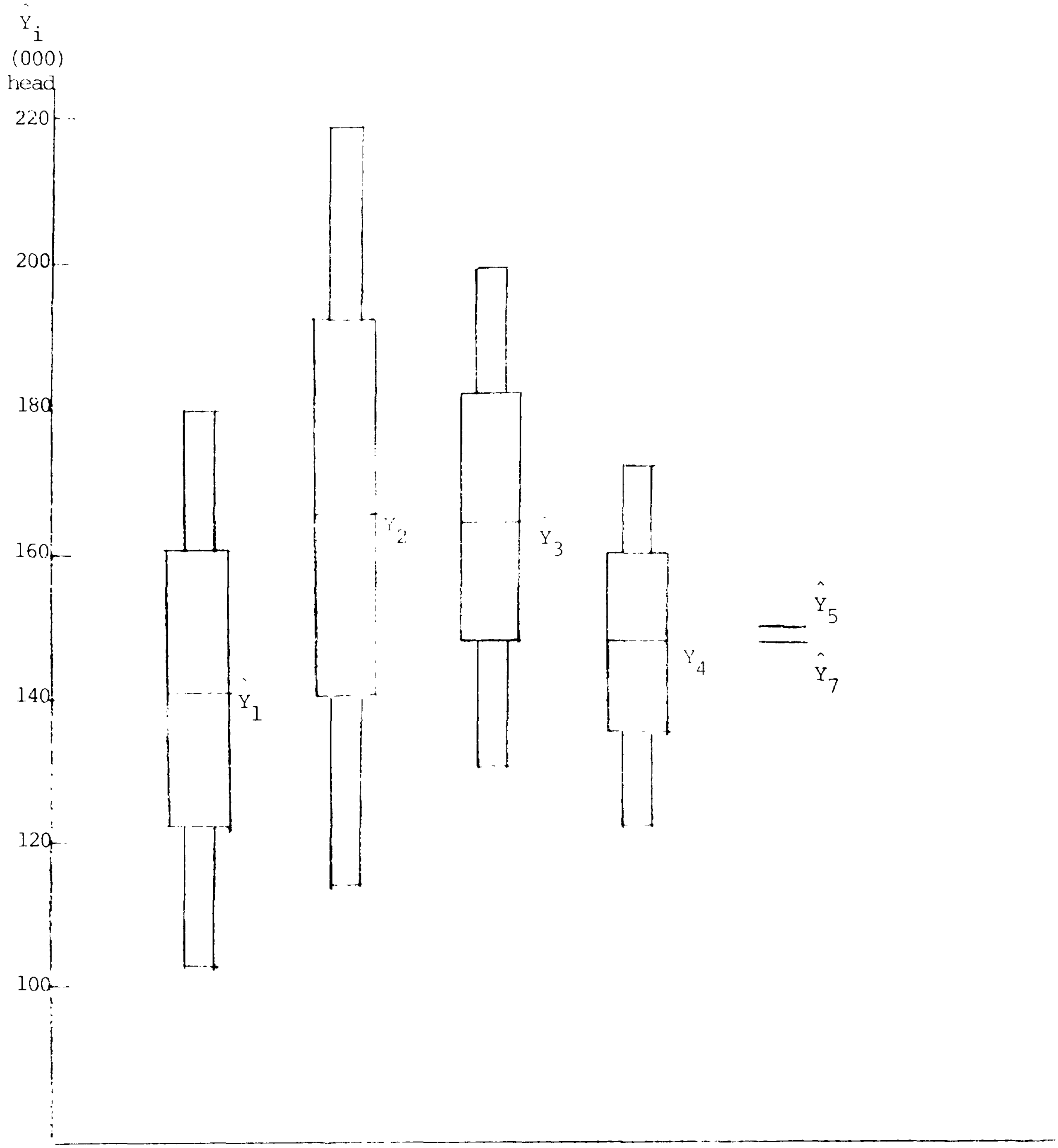
1/ Reference page 31 for a description of plot.

Figure 3: Schematic Plot for Iowa 1/
Comparison of estimates with confidence limits



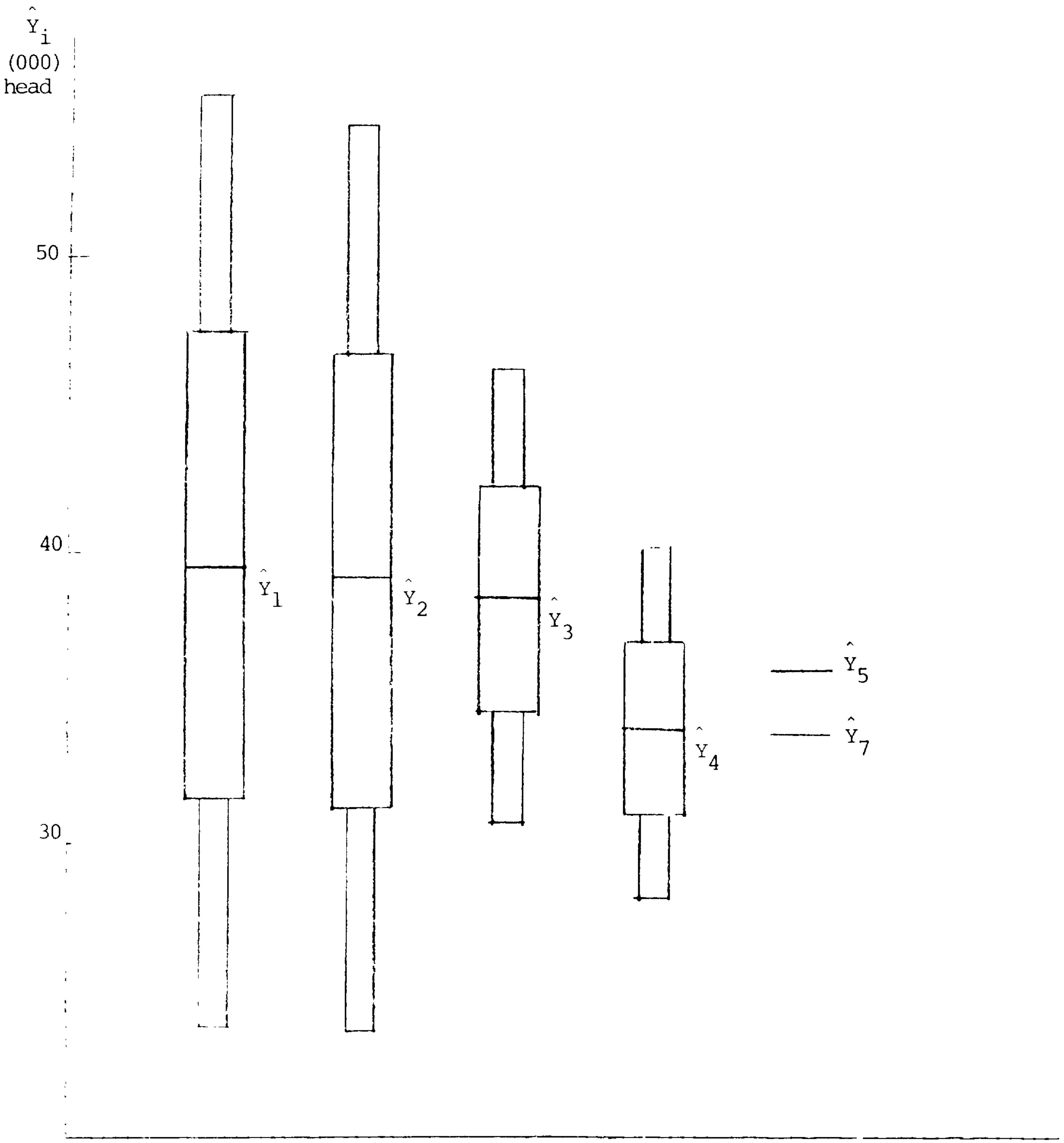
1/ Reference page 31 for a description of plot.

Figure 4: Schematic Plot for Kansas [1/]
Comparison of estimates with confidence limits



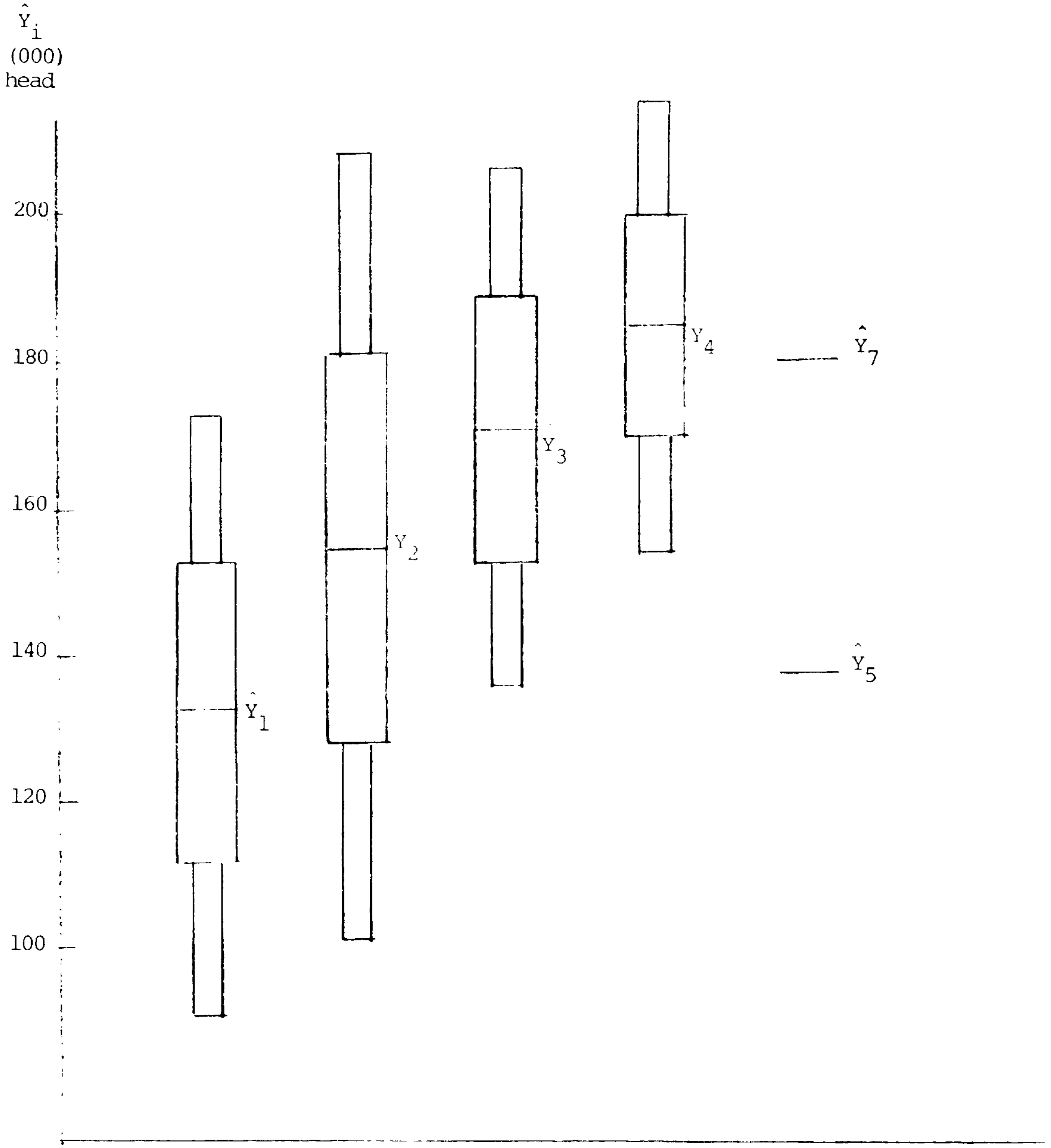
1/ Reference page 31 for a description of plot.

Figure 5: Schematic Plot for Missouri 1/
Comparison of estimates with confidence limits



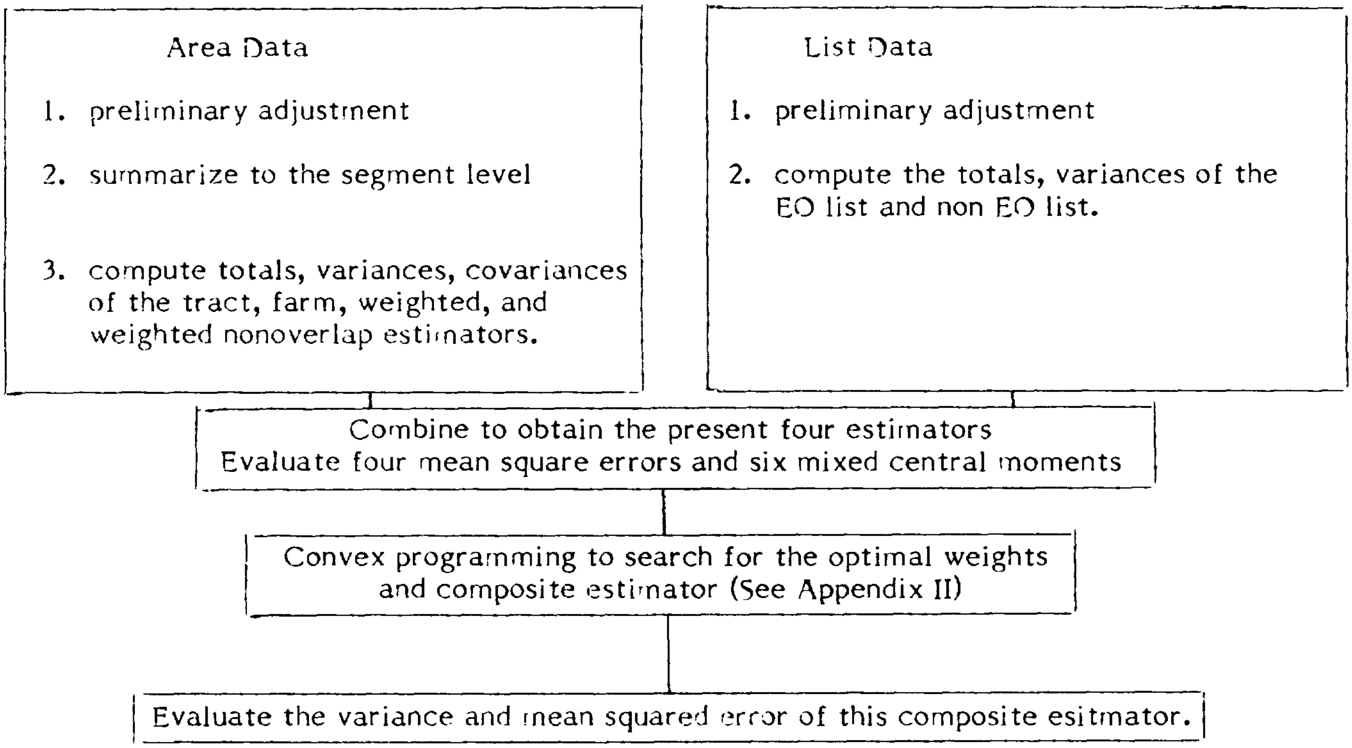
1/ Reference page 31 for a description of plot.

Figure 6: Schematic Plot for Ohio 1/
Comparison of estimates with confidence limits



1/ Reference page 31 for a description of plot.

Appendix I: Flow Chart of the SAS Program



Appendix II. Convex Programming to Search for Composite Weights

The application of the PROC MATRIX procedure used to solve the convex programming problem described in Section 3 is given here.

From equation (3.5) in the text we need to minimize

$$f(w) = \sum_{i=1} w_i^2 \hat{m}_i^2 + \sum_{i \neq j} w_i w_j \hat{m}_{ij} \tag{II.1}$$

subject to $0 \leq w_i \leq 1$ for all i, and $\sum_{i=1}^{4} w_i = 1$.

Let the inequality constraint functions be denoted by $g_i(w) = w_i$ for all i. The Lagrange multiplier technique is applied. A necessary and sufficient condition for the minimum to exist is as follows (see page 152 of Avriel (1976)):

There exists u, w_i and positive constraints λ_i, i=1,...,4,

such that

$$\lambda_i g_i(\underset{\sim}{w}) = 0 \text{ for all i,} \tag{II.2}$$

$$\nabla \hat{f}(\underset{\sim}{w}) - \Sigma_i \lambda_i \nabla g_i(\underset{\sim}{w}) - u \nabla (\Sigma w_i - 1) = 0, \text{ and the w's}$$

satisfy the constraints of (II.1).

Equation (II.2) can be rewritten as

$$\begin{aligned} &\lambda_i w_i = 0 \quad (i=1 \text{ to } 4), \\ &\hat{m}_i^2 w_i + \sum_{j \neq i} (\hat{m}_{ij} w_i - \lambda_i/2) - u/2 = 0, (i=1 \text{ to } 4), \\ &\sum_{i=1}^{4} w_i = 1. \end{aligned} \tag{III.3}$$

There are nine equations with nine variables (four w's, four λ's, and one u) to be solved simultaneously. It would be easier to solve the first four equations, i.e., $\lambda i=0$ or $w_i=0$, then substitute those values in the last five equations and solve them. To solve the first four equations, there are 16 cases to be considered. The cases for all $w_i=0$ is ruled out. Therefore, there are 15 cases left. Each of the 15 cases can be written as linear equations with 5 variables. The function SOLVE in the PROC MATRIX procedure is used to reach a solution. The minimum of equation (II.1) corresponds to the solution with all non-negative variables. Basically, the program searches for the minimal value of (II.1) among all 15 cases of the possible combinations of the preliminary estimators: combination of all the four (one case), combinations of three at a time (four cases), combinations of two at a time (six cases), and just the preliminary estimators (four cases).

Appendix III contains the complete program for all the 15 cases. A few cases are given here to explain the PROC MATRIX program in Appendix III.

Case 1: $\lambda_i=0$ for $i=1...,4$, then equation (III.3) reduces to

$$
\begin{array}{lllllll}
w_1 \hat{m}_1^2 & + w_2 \hat{m}_{12} & + w_3 \hat{m}_{13} & + w_4 \hat{m}_{14} & - u/2 & & = 0 \\
w_1 \hat{m}_{12} & + w_2 \hat{m}_2^2 & + w_3 \hat{m}_{13} & + w_4 \hat{m}_{14} & - u/2 & & = 0 \\
w_1 \hat{m}_{13} & + w_2 \hat{m}_{23} & + w_3 \hat{m}_3^2 & + w_4 \hat{m}_{34} & - u/2 & & = 0 \\
w_1 \hat{m}_{14} & + w_2 \hat{m}_{24} & + w_3 \hat{m}_{34} & + w_4 \hat{m}_4^2 & - u/2 & & = 0 \\
w_1 & + w_2 & + w_3 & + w_4 & & - 1 & = 0
\end{array}
$$

The variables w_i's and u are solved by the SOLVE function. If all the $w_i > 0$, then the solution to (II.1) is given by this solution. This solution corresponds to the occasion that the minimum of f (w) is obtained in the interior of the feasible region which is a tetrahedron.

Case 2: $w_1=0, \lambda_2=\lambda_3=\lambda_4=0$, then equation (III.3) reduces to

$$
\begin{array}{llllllll}
w_2 \hat{m}_{12} & + w_3 \hat{m}_{13} & + w_4 \hat{m}_{14} & - \lambda_1/2 & - u/2 & & = 0 \\
w_2 \hat{m}_2^2 & + w_3 \hat{m}_{23} & + w_4 \hat{m}_{24} & & - u/2 & & = 0 \\
w_2 \hat{m}_{23} & + w_3 \hat{m}_3^2 & + w_4 \hat{m}_{34} & & - u/2 & & = 0 \\
w_2 \hat{m}_{24} & + w_3 \hat{m}_{34} & + w_4 \hat{m}_4^2 & & - u/2 & & = 0 \\
w_2 & + w_3 & + w_4 & & & - 1 & = 0
\end{array}
$$

Having solved this linear equation, if w_2, w_3, w_4, and λ_1, are all positive, then the minimum of f(w) within the feasible region is given by this solution. This minimum is obtained at the boundary of the tetrahedron, i.e. one of the four faces. In terms of composite estimation, this minimum is obtained by just combining Y_2, Y_3, and Y_4.

Cases 3, 4 and 5 can be derived similarly.

Case 6: $w_3=w_4=\lambda_1=\lambda_2 = 0$

$$
\begin{array}{lllll}
w_1 \hat{m}_1^2 & + w_2 \hat{m}_{12} & & - u/2 & = 0 \\
w_1 \hat{m}_{12} & + w_2 \hat{m}_2 & - \lambda_3/2 & - u/2 & = 0 \\
w_1 \hat{m}_{13} & + w_2 \hat{m}_{23} & - \lambda_4/2 & - u/2 & = 0 \\
w_1 \hat{m}_{14} & + w_2 \hat{m}_{24} & & - u/2 & = 0 \\
w_1 & + w_2 & & & = 1
\end{array}
$$

If the solutions w_1, w_2, λ_3, and λ_4 are all positive, then it is the minimum of the equation (II.1). It corresponds to the occasion that the minimum of f(w) within the tetrahedron is attained at one of the six edges. It is best to use just the combination of Y_1 and Y_2.

Cases 7-11 are derived similarly.

Cases 12-15 correspond to the occasion that the minimum of f(w) is attained at the vertices of the tetrahedron. It is best to use just one of the preliminary estimators.

```
//[illegible]   JOB ([illegible],[illegible] ,SR,B6),'LYNN',
// USER=[illegible],PASSWORD=[illegible] ,
// MSGLEVEL=(2,0),CLASS=U
/*ROUTE PRINT RMT478
//STEP1 EXEC SAS,TIME=(30,30)
//IN DD DSN=SR780.LK.INDIHOGS.JUNE84.DATA,UNIT=SYSDA,
// DISP=OLD
//OUT DD DSN=SR780.LK.PSEUDO55.INDI.DATA,UNIT=SYSSR,
// DISP=(NEW,CATLG),SPACE=(TRK,(100,10),RLSE)
//SYSIN DD *
DATA REDUCED;
SET IN.INDIHOGS;
KEEP ID5 ID6 ID9 P7 P12 P200 P300 P403 P815 P818 P840 P900;

DATA AREA;
SET REDUCED;
IF ID5>=100;

DATA ASSIGN;
SET REDUCED;
IF ID5<=99;
RETAIN SEED1 234678903;
PRAN=RANUNI(SEED1);

PROC SORT DATA=ASSIGN;
BY ID5 PRAN;

DATA LISTREP;
SET;
PREP=MOD(_N_-1,4);

RUN;
%MACRO CREATE;

%DO I=1 %TO 10;
DATA DATA&I;
SET REDUCED;
IF 1101<=ID5<=1119 AND P7=&I THEN DELETE;
IF 1101<=ID5<=1119 THEN P12=P12*10/9;
RUN;
%END;

%DO I= 11 %TO 15;
DATA DATA&I;
SET REDUCED;
IF 1201<=ID5<=1211 AND P7=%EVAL(&I-10) THEN DELETE;
IF 1201<=ID5<=1211 THEN P12=P12*5/4;
RUN;
%END;

%DO I=16 %TO 20;
DATA DATA&I;
SET REDUCED;
IF 2001<=ID5<=2006 AND P7=%EVAL(&I-15) THEN DELETE;
```

```
IF 2001<=ID5<=2006 THEN P12=P12*5/4;
RUN;
%END;

%DO I=21 %TO 25;
DATA DATA&I;
SET REDUCED;
IF 3101<=ID5<=3105 AND P7=%EVAL(&I-20) THEN DELETE;
IF 3101<=ID5<=3105 THEN P12=P12*5/4;
RUN;
%END;

%DO I=26 %TO 30;
DATA DATA&I;
SET REDUCED;
IF 4001<=ID5<=4003 AND P7=%EVAL(&I-25) THEN DELETE;
IF 4001<=ID5<=4003 THEN P12=P12*5/4;
RUN;
%END;

%DO I=31 %TO 34; DATA DATA&I;
SET LISTREP AREA;
IF ID5=85 AND PREP=%EVAL(&I-31) THEN DELETE;
RUN;
%END;

%DO I=35 %TO 38;
DATA DATA&I;
SET LISTREP AREA;
IF ID5=86 AND PREP=%EVAL(&I-35) THEN DELETE;
RUN;
%END;

%DO I=39 %TO 42;
DATA DATA&I;
SET LISTREP AREA;
IF ID5=87 AND PREP=%EVAL(&I-39) THEN DELETE;
RUN;
%END;

%DO I=43 %TO 46;
DATA DATA&I;
SET LISTREP AREA;
IF ID5=88 AND PREP=%EVAL(&I-43) THEN DELETE;
RUN;
%END;

%DO I=47 %TO 50;
DATA DATA&I;
SET LISTREP AREA;
IF ID5=93 AND PREP=%EVAL(&I-47) THEN DELETE;
RUN;
%END;
```

```
%DO I=51 %TO 54;
DATA DATA&I;
SET LISTREP AREA;
IF ID5=94 AND PREP=%EVAL(&I-51) THEN DELETE;
RUN;
%END;

%DO I=55 %TO 55;
DATA DATA&I;
SET REDUCED;
IF ID5=3201 OR ID5=3301 OR ID5=5001 THEN DELETE;
RUN;
%END;

%DO I=56 %TO 56;
DATA DATA&I;
SET REDUCED;
RUN;
%END;
%MEND CREATE;
%CREATE

DATA NSIZE;
INPUT ID5 NPOP;
CARDS;
85 14599
86 2924
87 2307
88 1579
93 444
94 136
98 26
;

RUN;
%MACRO PLAY;

%DO I=1 %TO 56;
DATA DA&I;
SET DATA&I;
IF ID5>100;
IF P403=3 THEN  P301=0;
  ELSE P301=P300;
IF P403=3 THEN P201=0;
  ELSE P201=P200;
IF P818=0 THEN P500=P301*P840/P900;
   ELSE P500=0;
IF P403=1 THEN P600=P500;
   ELSE P600=0;
P302=P301*P815;

PROC SUMMARY;
CLASS ID6;
VAR P201 P302 P500 P600;
```

```
ID ID5 P12;
OUTPUT OUT=SEG SUM=TR FARM WT WT1;

DATA SEGCLN;
SET SEG;
IF _TYPE_=1;
TRE=TR*P12;
FARME=FARM*P12;
WTE=WT*P12;
WT1E=WT1*P12;

PROC SORT OUT=SORSEG;
BY ID5;

PROC SUMMARY DATA=SORSEG;
VAR TRE FARME WTE WT1E P12;
BY ID5;
OUTPUT OUT=TOTAL N=NN MEAN(P12)=AP12 SUM=STR SFR SWT SWTNOL SP12;

PROC SUMMARY DATA=TOTAL;
VAR STR SFR SWT SWTNOL;
OUTPUT OUT=ARTL SUM=ATR AFR AWT AWTNOL;

PROC CORR DATA=SORSEG NOPRINT COV OUTP=A;
VAR TRE FARME WTE WT1E;
BY ID5;

DATA NEW;
SET A;
IF _TYPE_  EQ 'MEAN' THEN DELETE;
IF _TYPE_  EQ 'STD' THEN DELETE;
IF _TYPE_  EQ 'N' THEN DELETE;
IF _TYPE_  EQ 'CORR' THEN DELETE;
DROP _TYPE_;

DATA NEWA;
MERGE NEW TOTAL;
BY ID5;
VTR=TRE*NN*(1-1/AP12);
VFR=FARME*NN*(1-1/AP12);
VWT=WTE*NN*(1-1/AP12);
VWT1=WT1E*NN*(1-1/AP12);

DATA STEP;
SET NEWA;
ROW=MOD(_N_-1,4);
PROC SORT  OUT=SORSTP;
BY ROW;

PROC SUMMARY DATA=SORSTP;
VAR VTR VFR VWT VWT1;
BY ROW;
OUTPUT OUT=ACOV SUM=COVTR COVFR COVWT COVWT1;
```

```
DATA LIST;
SET DATA&I;
IF ID9 LE 5 AND ID5 LT 99;

PROC SORT DATA=LIST;
BY ID6;

DATA LISTCL;
SET LIST;
BY ID6;
RETAIN TOT 0;
P305=P300*P403;
TOT=SUM(TOT,P305);
IF LAST.ID6 THEN DO;
P305=TOT;
TOT=0;
OUTPUT;
END;

PROC SUMMARY DATA=LISTCL;
CLASS ID5;
VAR P305;
OUTPUT OUT=STAT N=COUNT MEAN=MLST VAR=VLST;

DATA STATCL;
SET STAT;
IF _TYPE_=0 THEN DELETE;

PROC SORT DATA=STATCL OUT=STASOR;
BY ID5;

DATA NEWLI;
MERGE NSIZE STASOR;
BY ID5;
NMLST=MLST*NPOP;
NVLST=VLST*NPOP**2*(1-COUNT/NPOP)/COUNT;

DATA EO;
SET NEWLI;
IF 93<=ID5<=99;
PROC SUMMARY;
VAR NMLST NVLST;
OUTPUT OUT=EOT SUM=EOTL EOV;

DATA NONEO;
SET NEWLI;
IF 81<=ID5<=92;
PROC SUMMARY;
VAR NMLST NVLST;
OUTPUT OUT=LIOUT SUM=LITL LIV;

DATA COMP;
MERGE ARTL EOT LIOUT;
HTR=ATR+EOTL;
```

```
HFR=AFR+EOTL;
HWT=AWT+EOTL;
HMF=AWTMOL+LITL+EOTL;

PROC MATRIX ERRMAX=300;
FETCH X DATA=ACOV;
FETCH Y DATA=COMP;
M1=X(1,4)+Y(1,8);
MP2=(Y(1,12)-Y(1,11))**2-M1 +(X(1,5)+Y(1,8))*2;
V2=X(2,5)+Y(1,8);
M2=MP2<>V2;
MP3=(Y(1,13)-Y(1,11))**2-M1+(X(1,6)+Y(1,8))*2;
V3=X(3,6)+Y(1,8);
M3=MP3<>V3;
MP4=(Y(1,14)-Y(1,11))**2-M1+(X(1,7)+Y(1,8))*2;
V4=X(4,7)+Y(1,8)+Y(1,10);
M4=MP4<>V4;
M12=X(1,5)+Y(1,8);
M13=X(1,6)+Y(1,8);
M14=X(1,7)+Y(1,8);
M23=(Y(1,12)-Y(1,11))*(Y(1,13)-Y(1,11))-M1+X(1,5)+X(1,6)+2*Y(1,8);
M24=(Y(1,12)-Y(1,11))*(Y(1,14)-Y(1,11))-M1+X(1,5)+X(1,7)+2*Y(1,8);
M34=(Y(1,13)-Y(1,11))*(Y(1,14)-Y(1,11))-M1+X(1,6)+X(1,7)+2*Y(1,8);
MA11=M1||M12;
MA12=M13||M14;
MA1=MA11||MA12;
MA21=M12||M2;
MA22=M23||M24;
MA2=MA21||MA22;
MA31=M13||M23;
MA32=M3||M34;
MA3=MA31||MA32;
MA41=M14||M24;
MA42=M34||M4;
MA4=MA41||MA42;
MA5=1 1 1 1 0;
MC1=MA1//MA2;
MC2=MA3//MA4;
MC=MC1//MC2;
MCC=0.5/0.5/0.5/0.5;
MV=MC||MCC;
MVV=MV//MA5;
B=0/0/0/0/1;
W=SOLVE(MVV,B);
W(5,1)=0;
WW=%STR(W%');
YY=Y(1,11)//Y(1,12);
YZ=Y(1,13)//Y(1,14);
EST=YY//YZ;
EST1=EST//W(5,1);
FEST=WW*EST1;

MB1=M1||M14;
MB2=M14||M4;
```

```
MB=MB1//MB2;
MBC=0.5/0.5;
MBV=MB||MBC;
MB3=1 1 0;
MBVV=MBV//MB3;
BB=0/0/1;
WB=SOLVE(MBVV,BB);
ESTB=Y(1,11)*WB(1,1)+Y(1,14)*WB(2,1);

SD1=SQRT(M1);
SD2=SQRT(V2);
SD3=SQRT(V3);
SD4=SQRT(V4);
RMSQ2=SQRT(M2);
RMSQ3=SQRT(M3);
RMSQ4=SQRT(M4);
MOM2=(SD1||SD2)||(SD3||SD4)||(RMSQ2||RMSQ3)||RMSQ4;
OUTPUT MOM2 OUT=T1(RENAME=(COL1=SD1 COL2=SD2 COL3=SD3 COL4=SD4 COL5=RMSQ2
COL6=RMSQ3 COL7=RMSQ4));

MVER1=(M1//M12)//(M13//M14);
MVER2=(M12//M2)//(M23//M24);
MVER3=(M13//M23)//(M3//M34);
MVER4=(M14//M24)//(M34//M4);
U=0.5/0.5/0.5/0.5;
MU=-U;
MVER12=MVER1||MVER2;
MVER34=MVER3||MVER4;
MVER=MVER12||MVER34;
MVERU=MVER||MU;
MCON4=MVERU//MA5;
WCN4=SOLVE(MCON4,B);
ESTCN4=%STR(WCN4%')*EST1;
WCON4=(WCN4(1,1)//WCN4(2,1))//(WCN4(3,1)//WCN4(4,1));
MINCN4=%STR(WCON4%')*MVER*WCON4;

MVERM1=MVER2||MVER34;
LAMDA1=-0.5/0/0/0;
MM1U=(MVERM1||LAMDA1)||MU;
LAST3=1 1 1 0 0;
MCNM1=MM1U//LAST3;
WCNM1=SOLVE(MCNM1,B);
YM1=(Y(1,12)//YZ)//(0//0);
ESTM1=%STR(WCNM1%')*YM1;
WM1=(0//WCNM1(1,1))//(WCNM1(2,1)//WCNM1(3,1));
CN3M1=%STR(WM1%')*MVER*WM1;

MVERM2=MVER1||MVER34;
LAMDA2=0/-0.5/0/0;
MM2U=(MVERM2||LAMDA2)||MU;
MCNM2=MM2U//LAST3;
WCNM2=SOLVE(MCNM2,B);
YM2=(Y(1,11)//YZ)//(0//0);
ESTM2=%STR(WCNM2%')*YM2;
```

```
WM2=(WCNM2(1,1)//0)//(WCNM2(2,1)//WCNM2(3,1));
CN3M2=%STR(WM2%')*MVER*WM2;

MVERM3=MVER12||MVER4;
LAMDA3=0/0/-0.5/0;
MM3U=(MVERM3||LAMDA3)||MU;
MCNM3=MM3U//LAST3;
WCNM3=SOLVE(MCNM3,B);
YM3=(YY//Y(1,14))//(0//0);
ESTM3=%STR(WCNM3%')*YM3;
WM3=(WCNM3(1,1)//WCNM3(2,1))//(0//WCNM3(3,1));
CN3M3=%STR(WM3%')*MVER*WM3;

MVERM4=MVER12||MVER3;
LAMDA4=0/0/0/-0.5;
MM4U=(MVERM4||LAMDA4)||MU;
MCNM4=MM4U//LAST3;
WCNM4=SOLVE(MCNM4,B);
YM4=(YY//Y(1,13))//(0//0);
ESTM4=%STR(WCNM4%')*YM4;
WM4=(WCNM4(1,1)//WCNM4(2,1))//(WCNM4(3,1)//0);
CN3M4=%STR(WM4%')*MVER*WM4;

L2D3=0/0/-0.5/0;
L2D4=0/0/0/-0.5;
M34U=(MVER12||L2D3)||(L2D4||MU);
LAST=1 1 0 0 0;
MCNM34=M34U//LAST;
WCNM34=SOLVE(MCNM34,B);
TRO=0/0/0;
YM34=YY//TRO;
ESTM34=%STR(WCNM34%')*YM34;
WM34=(WCNM34(1,1)//WCNM34(2,1))//(0//0);
CNM34=%STR(WM34%')*MVER*WM34;

L2D2=0/-0.5/0/0;
M24U=((MVER1||MVER3)||(L2D2||L2D4))||MU;
MCNM24=M24U//LAST;
WCNM24=SOLVE(MCNM24,B);
YM24=(Y(1,11)//Y(1,13))//TRO;
ESTM24=%STR(WCNM24%')*YM24;
WM24=(WCNM24(1,1)//0)//(WCNM24(2,1)//0);
CNM24=%STR(WM24%')*MVER*WM24;

L2D1=-0.5/0/0/0;
M12U=((MVER3||MVER4)||(L2D1||L2D2))||MU;
MCNM12=M12U//LAST;
WCNM12=SOLVE(MCNM12,B);
YM12=YZ//TRO;
ESTM12=%STR(WCNM12%')*YM12;
WM12=(0//0)//(WCNM12(1,1)//WCNM12(2,1));
CNM12=%STR(WM12%')*MVER*WM12;

M13U=((MVER2||MVER4)||(L2D1||L2D3))||MU;
```

```
MCNM13=M13U//LAST;
WCNM13=SOLVE(MCNM13,B);
YM13=(Y(1,12)//Y(1,14))//TRO;
ESTM13=%STR(WCNM13%')*YM13;
WM13=(0//WCNM13(1,1))//(0//WCNM13(2,1));
CNM13=%STR(WM13%')*MVER*WM13;

M14U=((MVER2||MVER3)||(L2D1||L2D4))||MU;
MCNM14=M14U//LAST;
WCNM14=SOLVE(MCNM14,B);
YM14=(Y(1,12)//Y(1,13))//TRO;
ESTM14=%STR(WCNM14%')*YM14;
WM14=(0//WCNM14(1,1))//(WCNM14(2,1)//0);
CNM14=%STR(WM14%')*MVER*WM14;

M23U=((MVER1||MVER4)||(L2D2||L2D3))||MU;
MCNM23=M23U//LAST;
WCNM23=SOLVE(MCNM23,B);
YM23=(Y(1,11)//Y(1,14))//TRO;
ESTM23=%STR(WCNM23%')*YM23;
WM23=(WCNM23(1,1)//0)//(0//WCNM23(2,1));
CNM23=%STR(WM23%')*MVER*WM23;

OUTPUT ESTM23 OUT=OUTM23;
LASTT=1 0 0 0 0;
M234U=((MVER1||L2D2)||(L2D3||L2D4))||MU;
MCNM234=M234U//LASTT;
WCNM234=SOLVE(MCNM234,B);
ESTM234=Y(1,11);
CNM234=M1;

M134U=((MVER2||L2D1)||(L2D3||L2D4))||MU;
MCNM134=M134U//LASTT;
WCNM134=SOLVE(MCNM134,B);
ESTM134=Y(1,12);
CNM134=M2;

M124U=((MVER3||L2D1)||(L2D2||L2D4))||MU;
MCNM124=M124U//LASTT;
WCNM124=SOLVE(MCNM124,B);
ESTM124=Y(1,13);
CNM124=M3;

M123U=((MVER4||L2D1)||(L2D2||L2D3))||MU;
MCNM123=M123U//LASTT;
WCNM123=SOLVE(MCNM123,B);
ESTM123=Y(1,14);
CNM123=M4;

IF WCON4>=0 THEN OUTPUT ESTCN4 OUT=T;
IF WCON4>=0 THEN PRINT WCON4;
WCNM1S=WCNM1(1 2 3 4,1);
IF WCNM1S>=0 THEN OUTPUT ESTM1 OUT=T;
IF WCNM1S>=0 THEN PRINT WCNM1S;
```

```
WCNM2S=WCNM2(1 2 3 4,1);
IF WCNM2S>=0 THEN OUTPUT ESTM2 OUT=T;
IF WCNM2S>=0 THEN PRINT WCNM2S;
WCNM3S=WCNM3(1 2 3 4,1);
IF WCNM3S>=0 THEN OUTPUT ESTM3 OUT=T;
IF WCNM3S>=0 THEN PRINT WCNM3S;
WCNM4S=WCNM4(1 2 3 4,1);
IF WCNM4S>=0 THEN OUTPUT ESTM4 OUT=T;
IF WCNM4S>=0 THEN PRINT WCNM4S;
WCNM34S=WCNM34(1 2 3 4,1);
IF WCNM34S>=0 THEN OUTPUT ESTM34 OUT=T;
IF WCNM34S>=0 THEN PRINT WCNM34S;
WCNM24S=WCNM24(1 2 3 4,1);
IF WCNM24S>=0 THEN OUTPUT ESTM24 OUT=T;
IF WCNM24S>=0 THEN PRINT WCNM24S;
WCNM12S=WCNM12(1 2 3 4,1);
IF WCNM12S>=0 THEN OUTPUT ESTM12 OUT=T;
IF WCNM12S>=0 THEN PRINT WCNM12S;
WCNM13S=WCNM13(1 2 3 4,1);
IF WCNM13S>=0 THEN OUTPUT ESTM13 OUT=T;
IF WCNM13S>=0 THEN PRINT WCNM13S;
WCNM14S=WCNM14(1 2 3 4,1);
IF WCNM14S>=0 THEN OUTPUT ESTM14 OUT=T;
IF WCNM14S>=0 THEN PRINT WCNM14S;
WCNM23S=WCNM23(1 2 3 4,1);
IF WCNM23S>=0 THEN OUTPUT ESTM23 OUT=T;
IF WCNM23S>=0 THEN PRINT WCNM23S;
WCNM234S=WCNM234(1 2 3 4,1);
IF WCNM234S>=0 THEN OUTPUT ESTM234 OUT=T;
IF WCNM234S>=0 THEN PRINT WCNM234S;
WCNM134S=WCNM134(1 2 3 4,1);
IF WCNM134S>=0 THEN OUTPUT ESTM134 OUT=T;
IF WCNM134S>=0 THEN PRINT WCNM134S;
WCNM124S=WCNM124(1 2 3 4,1);
IF WCNM124S>=0 THEN OUTPUT ESTM124 OUT=T;
IF WCNM124S>=0 THEN PRINT WCNM124S;
WCNM123S=WCNM123(1 2 3 4,1);
IF WCNM123S>=0 THEN OUTPUT ESTM123 OUT=T;
IF WCNM123S>=0 THEN PRINT WCNM123S;


DATA NEWT1;
SET T1;
DROP ROW;

DATA SCOMP;
SET OUTM23;
RENAME COL1=COMP14;
DROP ROW;

DATA KFACTR&I;
MERGE T NEWT1 COMP SCOMP;
DROP ROW _TYPE_ _FREQ_;
K=2;
```

```
DIFF=HTR-COL1;
KSD1=K*SD1;
IF DIFF>=KSD1 THEN LTEST=HTR-KSD1;
   ELSE IF -DIFF>=KSD1 THEN LTEST=HTR+KSD1;
   ELSE LTEST=COL1;
PROC PRINT;
RUN;
%END;
%MEND PLAY;

%PLAY
RUN;
DATA KFULL;
SET KFACTR56;
KEEP COL1 HTR HFR HWT HMF COMP14 LTEST K SD1 SD2 SD3 SD4 RMSQ2 RMSQ3 RMSQ4;
RENAME COL1=FCOL1 HTR=FHTR HFR=FHFR HWT=FHWT HMF=FHMF
COMP14=FCOMP14  LTEST=FLTEST SD1=FSD1 SD2 =FSD2 SD3=FSD3 SD4=FSD4
RMSQ2=FRMSQ2 RMSQ3=FRMSQ3 RMSQ4=FRMSQ4;
PROC PRINT;


DATA ALL1;
SET KFACTR1 KFACTR2 KFACTR3 KFACTR4 KFACTR5
    KFACTR6 KFACTR7 KFACTR8 KFACTR9 KFACTR10
    KFACTR11 KFACTR12 KFACTR13 KFACTR14 KFACTR15
    KFACTR16 KFACTR17 KFACTR18 KFACTR19 KFACTR20
    KFACTR21 KFACTR22 KFACTR23 KFACTR24 KFACTR25
    KFACTR26 KFACTR27 KFACTR28 KFACTR29 KFACTR30
    KFACTR31 KFACTR32 KFACTR33 KFACTR34 KFACTR35
    KFACTR36 KFACTR37 KFACTR38 KFACTR39 KFACTR40
    KFACTR41 KFACTR42 KFACTR43 KFACTR44 KFACTR45;

DATA ALL2;
SET KFACTR46 KFACTR47 KFACTR48 KFACTR49 KFACTR50
    KFACTR51 KFACTR52 KFACTR53 KFACTR54 KFACTR55;

DATA ALL;
SET ALL1 ALL2;

DATA PSEUDO;
MERGE ALL KFULL;
BY K;
G=55;
PCOL1=G*FCOL1-(G-1)*COL1;
PHTR=G*FHTR-(G-1)*HTR;
PHFR=G*FHFR-(G-1)*HFR;
PHWT=G*FHWT-(G-1)*HWT;
PHMF=G*FHMF-(G-1)*HMF;
PCOMP14=G*FCOMP14-(G-1)*COMP14;
PLTEST=G*FLTEST-(G-1)*LTEST;
B2SQ=(HFR-HTR)**2;
B3SQ=(HWT-HTR)**2;
B4SQ=(HMF-HTR)**2;
B5SQ=(LTEST-HTR)**2;
```

```
B6SQ=(COMP14-HTR)**2;
PROC PRINT;

PROC CORR DATA=PSEUDO COV OUTP=POUT;
VAR PHTR PHFR PHWT PHMF PLTEST PCOMP14;
PROC PRINT DATA=POUT;

PROC CORR DATA=PSEUDO COV OUTP=ALLOUT;
VAR HTR HFR HWT HMF LTEST COMP14;
PROC PRINT DATA=ALLOUT;

PROC MEANS DATA=PSEUDO;
VAR B2SQ B3SQ B4SQ B5SQ B6SQ;
OUTPUT OUT=BIASQ MEAN=MN2 MN3 MN4 MN5 MN6;

DATA PCOV;
SET POUT;
IF _TYPE_='COV' AND _NAME_='PHTR';

DATA MSQ;
MERGE BIASQ PCOV KFULL;
G=55;
EPMSQ2=MN2+(PHFR/G)*2-PHTR/G;
EPMSQ3=MN3+(PHWT/G)*2-PHTR/G;
EPMSQ4=MN4+(PHMF/G)*2-PHTR/G;
EPMSQ5=MN5+(PLTEST/G)*2-PHTR/G;
EPMSQ6=MN6+(PCOMP14/G)*2-PHTR/G;
MSQ2F=(FHFR-FHTR)**2+(PHFR/G)*2-PHTR/G;
MSQ3F=(FHWT-FHTR)**2+(PHWT/G)*2-PHTR/G;
MSQ4F=(FHMF-FHTR)**2+(PHMF/G)*2-PHTR/G;
MSQ5F=(FLTEST-FHTR)**2+(PLTEST/G)*2-PHTR/G;
MSQ6F=(FCOMP14-FHTR)**2+(PCOMP14/G)*2-PHTR/G;
REPMSQ2=SQRT(EPMSQ2);
REPMSQ3=SQRT(EPMSQ3);
REPMSQ4=SQRT(EPMSQ4);
REPMSQ5=SQRT(EPMSQ5);
REPMSQ6=SQRT(EPMSQ6);
RMSQ2F=SQRT(MSQ2F);
RMSQ3F=SQRT(MSQ3F);
RMSQ4F=SQRT(MSQ4F);
RMSQ5F=SQRT(MSQ5F);
RMSQ6F=SQRT(MSQ6F);
PROC PRINT;

DATA SELMN;
SET POUT;
IF _TYPE_='MEAN';
RENAME PHTR=JKHTR PHFR=JKHFR PHWT=JKHWT PHMF=JKHMF PLTEST=JKLTEST
PCOMP14=JKCOMP14;

DATA SELSD;
SET POUT;
IF _TYPE_='STD';
G=55;
```

```
VJK1=PHTR**2/G;
VJK2=PHFR**2/G;
VJK3=PHWT**2/G;
VJK4=PHMF**2/G;
VJK5=PLTEST**2/G;
VJK6=PCOMP14**2/G;
PROC PRINT;

DATA;
MERGE KFULL SELMN SELSD;
G=55;
VP1=VJK1+(FHTR-JKHTR)**2/(G-1);
VP2=VJK2+(FHFR-JKHFR)**2/(G-1);
VP3=VJK3+(FHWT-JKHWT)**2/(G-1);
VP4=VJK4+(FHMF-JKHMF)**2/(G-1);
VP5=VJK5+(FLTEST-JKLTEST)**2/(G-1);
VP6=VJK6+(FCOMP14-JKCOMP14)**2/(G-1);
SDJKF1=SQRT(VP1);
SDJKF2=SQRT(VP2);
SDJKF3=SQRT(VP3);
SDJKF4=SQRT(VP4);
SDJKF5=SQRT(VP5);
SDJKF6=SQRT(VP6);
SDJK1=SQRT(VJK1);
SDJK2=SQRT(VJK2);
SDJK3=SQRT(VJK3);
SDJK4=SQRT(VJK4);
SDJK5=SQRT(VJK5);
SDJK6=SQRT(VJK6);
PROC PRINT;

DATA OUT.PSEUDO;
SET PSEUDO;
/*
```

```
//xxxxxx JOB (xxxxxxxxxxx      R,B6),'LYNN',
// USER=xxxxxxx,PASSWORD=xxxxx,
// MSGLEVEL=(2,0),CLASS=K
/*ROUTE PRINT RMT478
//STEP1 EXEC SAS,TIME=(2,30)
//IN DD DSN=SR780.LK.PSEUDO55.INDI.DATA,UNIT=SYSDA,
// DISP=OLD
//SYSIN DD *
DATA TRACK;
SET IN.PSEUDO;
KEEP PHTR K;

PROC SORT OUT=SORTR;
BY PHTR;

DATA WTTR;
SET SORTR;
ALPHA=0.1;
CUTOFF=FLOOR(55*ALPHA);
UPCUT=55-(CUTOFF-1);
IF _N_<=CUTOFF OR _N_>=UPCUT THEN W=0;
IF _N_=(CUTOFF+1) OR _N_=(UPCUT-1) THEN W=1+CUTOFF;
IF (CUTOFF+2)<=_N_<=(UPCUT-2) THEN W=1;

PROC UNIVARIATE;
VAR PHTR;
WEIGHT W;
ID K;
OUTPUT OUT=TROUT MEAN=TRMEAN ;

DATA;
MERGE TRACK TROUT;
BY K;
TRSQ=(PHTR-TRMEAN)**2;

PROC SORT OUT=SORTRSQ;
BY TRSQ;

DATA WTTRSQ;
SET SORTRSQ;
ALPHA=0.1;
CUTOFF=FLOOR(55*ALPHA);
UPCUT=55-(CUTOFF-1);
IF _N_<=CUTOFF OR _N_>=UPCUT THEN W=0;
IF _N_=(CUTOFF+1) OR _N_=(UPCUT-1) THEN W=1+CUTOFF;
IF (CUTOFF+2)<=_N_<=(UPCUT-2) THEN W=1;

PROC UNIVARIATE;
VAR TRSQ;
WEIGHT W;
ID K;
OUTPUT OUT=TRSQOUT MEAN=TRVAR ;

DATA FARM;
```

```
SET IN.PSEUDO;
KEEP PHFR K;

PROC SORT OUT=SORFR;
BY PHFR;

DATA WTFR;
SET SORFR;
ALPHA=0.1;
CUTOFF=FLOOR(55*ALPHA);
UPCUT=55-(CUTOFF-1);
IF _N_<=CUTOFF OR _N_>=UPCUT THEN W=0;
IF _N_=(CUTOFF+1) OR _N_=(UPCUT-1) THEN W=1+CUTOFF;
IF (CUTOFF+2)<=_N_<=(UPCUT-2) THEN W=1;

PROC UNIVARIATE;
VAR PHFR;
WEIGHT W;
ID K;
OUTPUT OUT=FROUT MEAN=FRMEAN ;

DATA;
MERGE FARM FROUT;
BY K;
FRSQ=(PHFR-FRMEAN)**2;

PROC SORT OUT=SORFRSQ;
BY FRSQ;

DATA WTFRSQ;
SET SORFRSQ;
ALPHA=0.1;
CUTOFF=FLOOR(55*ALPHA);
UPCUT=55-(CUTOFF-1);
IF _N_<=CUTOFF OR _N_>=UPCUT THEN W=0;
IF _N_=(CUTOFF+1) OR _N_=(UPCUT-1) THEN W=1+CUTOFF;
IF (CUTOFF+2)<=_N_<=(UPCUT-2) THEN W=1;

PROC UNIVARIATE;
VAR FRSQ;
WEIGHT W;
ID K;
OUTPUT OUT=FRSQOUT MEAN=FRVAR ;

DATA WEIGHT;
SET IN.PSEUDO;
KEEP PHWT K;

PROC SORT OUT=SORWT;
BY PHWT;

DATA WTWT;
SET SORWT;
ALPHA=0.1;
```

```
CUTOFF=FLOOR(55*ALPHA);
UPCUT=55-(CUTOFF-1);
IF _N_<=CUTOFF OR _N_>=UPCUT THEN W=0;
IF _N_=(CUTOFF+1) OR _N_=(UPCUT-1) THEN W=1+CUTOFF;
IF (CUTOFF+2)<=_N_<=(UPCUT-2) THEN W=1;

PROC UNIVARIATE;
VAR PHWT;
WEIGHT W;
ID K;
OUTPUT OUT=WTOUT MEAN=WTMEAN;

DATA;
MERGE WEIGHT WTOUT;
BY K;
WTSQ=(PHWT-WTMEAN)**2;

PROC SORT OUT=SORWTSQ;
BY WTSQ;

DATA WTWTSQ;
SET SORWTSQ;
ALPHA=0.1;
CUTOFF=FLOOR(55*ALPHA);
UPCUT=55-(CUTOFF-1);
IF _N_<=CUTOFF OR _N_>=UPCUT THEN W=0;
IF _N_=(CUTOFF+1) OR _N_=(UPCUT-1) THEN W=1+CUTOFF;
IF (CUTOFF+2)<=_N_<=(UPCUT-2) THEN W=1;

PROC UNIVARIATE;
VAR WTSQ;
WEIGHT W;
ID K;
OUTPUT OUT=WTSQOUT MEAN=WTVAR ;

DATA MF;
SET IN.PSEUDO;
KEEP PHMF K;

PROC SORT OUT=SORMF;
BY PHMF;

DATA WTMF;
SET SORMF;
ALPHA=0.1;
CUTOFF=FLOOR(55*ALPHA);
UPCUT=55-(CUTOFF-1);
IF _N_<=CUTOFF OR _N_>=UPCUT THEN W=0;
IF _N_=(CUTOFF+1) OR _N_=(UPCUT-1) THEN W=1+CUTOFF;
IF (CUTOFF+2)<=_N_<=(UPCUT-2) THEN W=1;

PROC UNIVARIATE;
VAR PHMF;
WEIGHT W;
```

```
ID K;
OUTPUT OUT=MFOUT MEAN=MFMEAN;

DATA;
MERGE MF MFOUT;
BY K;
MFSQ=(PHMF-MFMEAN)**2;

PROC SORT OUT=SORMFSQ;
BY MFSQ;

DATA WTMFSQ;
SET SORMFSQ;
ALPHA=0.1;
CUTOFF=FLOOR(55*ALPHA);
UPCUT=55-(CUTOFF-1);
IF _N_<=CUTOFF OR _N_>=UPCUT THEN W=0;
IF _N_=(CUTOFF+1) OR _N_=(UPCUT-1) THEN W=1+CUTOFF;
IF (CUTOFF+2)<=_N_<=(UPCUT-2) THEN W=1;

PROC UNIVARIATE;
VAR MFSQ;
WEIGHT W;
ID K;
OUTPUT OUT=MFSQOUT MEAN=MFVAR ;

DATA EST;
SET IN.PSEUDO;
KEEP PLTEST K;

PROC SORT OUT=SOREST;
BY PLTEST;

DATA WTEST;
SET SOREST;
ALPHA=0.1;
CUTOFF=FLOOR(55*ALPHA);
UPCUT=55-(CUTOFF-1);
IF _N_<=CUTOFF OR _N_>=UPCUT THEN W=0;
IF _N_=(CUTOFF+1) OR _N_=(UPCUT-1) THEN W=1+CUTOFF;
IF (CUTOFF+2)<=_N_<=(UPCUT-2) THEN W=1;

PROC UNIVARIATE;
VAR PLTEST;
WEIGHT W;
ID K;
OUTPUT OUT=ESTOUT MEAN=ESTMEAN;

DATA;
MERGE EST ESTOUT;
BY K;
ESTSQ=(PLTEST-ESTMEAN)**2;

PROC SORT OUT=SORESTSQ;
```

```
BY ESTSQ;

DATA WTESTSQ;
SET SORESTSQ;
ALPHA=0.1;
CUTOFF=FLOOR(55*ALPHA);
UPCUT=55-(CUTOFF-1);
IF _N_<=CUTOFF OR _N_>=UPCUT THEN W=0;
IF _N_=(CUTOFF+1) OR _N_=(UPCUT-1) THEN W=1+CUTOFF;
IF (CUTOFF+2)<=_N_<=(UPCUT-2) THEN W=1;

PROC UNIVARIATE;
VAR ESTSQ;
WEIGHT W;
ID K;
OUTPUT OUT=ESTSQOUT MEAN=ESTVAR ;

DATA COMP14;
SET IN.PSEUDO;
KEEP PCOMP14 K;

PROC SORT OUT=SOR14;
BY PCOMP14;

DATA WT14;
SET SOR14;
ALPHA=0.1;
CUTOFF=FLOOR(55*ALPHA);
UPCUT=55-(CUTOFF-1);
IF _N_<=CUTOFF OR _N_>=UPCUT THEN W=0;
IF _N_=(CUTOFF+1) OR _N_=(UPCUT-1) THEN W=1+CUTOFF;
IF (CUTOFF+2)<=_N_<=(UPCUT-2) THEN W=1;

PROC UNIVARIATE;
VAR PCOMP14;
WEIGHT W;
ID K;
OUTPUT OUT=SROUT MEAN=SRMEAN ;

DATA;
MERGE COMP14 SROUT;
BY K;
SESTSQ=(PCOMP14-SRMEAN)**2;

PROC SORT OUT=SORSRSQ;
BY SESTSQ;

DATA WTSESTSQ;
SET SORSRSQ;
ALPHA=0.1;
CUTOFF=FLOOR(55*ALPHA);
UPCUT=55-(CUTOFF-1);
IF _N_<=CUTOFF OR _N_>=UPCUT THEN W=0;
IF _N_=(CUTOFF+1) OR _N_=(UPCUT-1) THEN W=1+CUTOFF;
```

```
IF (CUTOFF+2)<=_N_<=(UPCUT-2) THEN W=1;

PROC UNIVARIATE;
VAR SESTSQ;
WEIGHT W;
ID K;
OUTPUT OUT=SRSQOUT MEAN=SESTVAR ;

DATA TRFR;
MERGE IN.PSEUDO TROUT FROUT;
BY K;
KEEP PHTR PHFR TRMEAN FRMEAN PDTRFR;
PDTRFR=(PHTR-TRMEAN)*(PHFR-FRMEAN);

PROC SORT OUT=SORTRFR;
BY PDTRFR;

DATA WTTRFR;
SET SORTRFR;
ALPHA=0.1;
CUTOFF=FLOOR(55*ALPHA);
UPCUT=55-(CUTOFF-1);
IF _N_<=CUTOFF OR _N_>=UPCUT THEN W=0;
IF _N_=(CUTOFF+1) OR _N_=(UPCUT-1) THEN W=1+CUTOFF;
IF (CUTOFF+2)<=_N_<=(UPCUT-2) THEN W=1;

PROC UNIVARIATE;
VAR PDTRFR;
WEIGHT W;
OUTPUT OUT=TRFROUT MEAN=COVTRFR ;

DATA TRWT;
MERGE IN.PSEUDO TROUT WTOUT;
BY K;
KEEP PHTR PHWT TRMEAN WTMEAN PDTRWT;
PDTRWT=(PHTR-TRMEAN)*(PHWT-WTMEAN);

PROC SORT OUT=SORTRWT;
BY PDTRWT;

DATA WTTRWT;
SET SORTRWT;
ALPHA=0.1;
CUTOFF=FLOOR(55*ALPHA);
UPCUT=55-(CUTOFF-1);
IF _N_<=CUTOFF OR _N_>=UPCUT THEN W=0;
IF _N_=(CUTOFF+1) OR _N_=(UPCUT-1) THEN W=1+CUTOFF;
IF (CUTOFF+2)<=_N_<=(UPCUT-2) THEN W=1;

PROC UNIVARIATE;
VAR PDTRWT;
WEIGHT W;
OUTPUT OUT=TRWTOUT MEAN=COVTRWT ;
```

```
DATA TRMF;
MERGE IN.PSEUDO TROUT MFOUT;
BY K;
KEEP PHTR PHMF TRMEAN MFMEAN PDTRMF;
PDTRMF=(PHTR-TRMEAN)*(PHMF-MFMEAN);

PROC SORT OUT=SORTRMF;
BY PDTRMF;

DATA WTTRMF;
SET SORTRMF;
ALPHA=0.1;
CUTOFF=FLOOR(55*ALPHA);
UPCUT=55-(CUTOFF-1);
IF _N_<=CUTOFF OR _N_>=UPCUT THEN W=0;
IF _N_=(CUTOFF+1) OR _N_=(UPCUT-1) THEN W=1+CUTOFF;
IF (CUTOFF+2)<=_N_<=(UPCUT-2) THEN W=1;

PROC UNIVARIATE;
VAR PDTRMF;
WEIGHT W;
OUTPUT OUT=TRMFOUT MEAN=COVTRMF ;

DATA TREST;
MERGE IN.PSEUDO TROUT ESTOUT;
BY K;
KEEP PHTR PLTEST TRMEAN ESTMEAN PDTREST;
PDTREST=(PHTR-TRMEAN)*(PLTEST-ESTMEAN);

PROC SORT OUT=SORTREST;
BY PDTREST;

DATA WTTREST;
SET SORTREST;
ALPHA=0.1;
CUTOFF=FLOOR(55*ALPHA);
UPCUT=55-(CUTOFF-1);
IF _N_<=CUTOFF OR _N_>=UPCUT THEN W=0;
IF _N_=(CUTOFF+1) OR _N_=(UPCUT-1) THEN W=1+CUTOFF;
IF (CUTOFF+2)<=_N_<=(UPCUT-2) THEN W=1;

PROC UNIVARIATE;
VAR PDTREST;
WEIGHT W;
OUTPUT OUT=TRESTOUT MEAN=COVTREST ;

DATA TRP14;
MERGE IN.PSEUDO TROUT SROUT;
BY K;
KEEP PHTR PLTEST TRMEAN SRMEAN PDTRSR;
PDTRSR=(PHTR-TRMEAN)*(PCOMP14-SRMEAN);

PROC SORT OUT=SORTRSR;
BY PDTRSR;
```

```
DATA WTTRSR;
SET SORTRSR;
ALPHA=0.1;
CUTOFF=FLOOR(55*ALPHA);
UPCUT=55-(CUTOFF-1);
IF _N_<=CUTOFF OR _N_>=UPCUT THEN W=0;
IF _N_=(CUTOFF+1) OR _N_=(UPCUT-1) THEN W=1+CUTOFF;
IF (CUTOFF+2)<=_N_<=(UPCUT-2) THEN W=1;

PROC UNIVARIATE;
VAR PDTRSR;
WEIGHT W;
OUTPUT OUT=TRSROUT MEAN=COVTRSR ;

DATA FULL;
SET IN.PSEUDO;
IF _N_=1;
KEEP FHTR FHFR FHWT FHMF FLTEST FCOMP14;

DATA ALL;
MERGE  TROUT TRSQOUT FROUT FRSQOUT  WTOUT WTSQOUT MFOUT MFSQOUT
 ESTOUT ESTSQOUT   SROUT SRSQOUT TRFROUT TRWTOUT
       TRMFOUT TRESTOUT TRSROUT FULL;
N=55;
ALPHA=0.1;
G=FLOOR(55*ALPHA);
C=N/(N-2*G)/(N-2*G-1);
MSQ2=(FRMEAN-TRMEAN)**2+2*COVTRFR*C-TRVAR*C;
MSQ3=(WTMEAN-TRMEAN)**2+2*COVTRWT*C-TRVAR*C;
MSQ4=(MFMEAN-TRMEAN)**2+2*COVTRMF*C-TRVAR*C;
MSQ5=(ESTMEAN-TRMEAN)**2+2*COVTREST*C-TRVAR*C;
MSQ6=(SRMEAN-TRMEAN)**2+2*COVTRSR*C-TRVAR*C;
MSQF2=(FHFR-FHTR)**2+2*COVTRFR*C-TRVAR*C;
MSQF3=(FHWT-FHTR)**2+2*COVTRWT*C-TRVAR*C;
MSQF4=(FHMF-FHTR)**2+2*COVTRMF*C-TRVAR*C;
MSQF5=(FLTEST-FHTR)**2+2*COVTREST*C-TRVAR*C;
MSQF6=(FCOMP14-FHTR)**2+2*COVTRSR*C-TRVAR*C;
VJK1=TRVAR*C;
VJK2=FRVAR*C;
VJK3=WTVAR*C;
VJK4=MFVAR*C;
VJK5=ESTVAR*C;
VJK6=SESTVAR*C;
VP1=VJK1+(FHTR-TRMEAN)**2*C;
VP2=VJK2+(FHFR-FRMEAN)**2*C;
VP3=VJK3+(FHWT-WTMEAN)**2*C;
VP4=VJK4+(FHMF-MFMEAN)**2*C;
VP5=VJK5+(FLTEST-ESTMEAN)**2*C;
VP6=VJK6+(FCOMP14-SRMEAN)**2*C;
SDJK1=SQRT(VJK1);
SDJK2=SQRT(VJK2);
SDJK3=SQRT(VJK3);
SDJK4=SQRT(VJK4);
SDJK5=SQRT(VJK5);
```

```
SDJK6=SQRT(VJK6);
SDJKF1=SQRT(VP1);
SDJKF2=SQRT(VP2);
SDJKF3=SQRT(VP3);
SDJKF4=SQRT(VP4);
SDJKF5=SQRT(VP5);
SDJKF6=SQRT(VP6);
RMSQ2=SQRT(MSQ2);
RMSQ3=SQRT(MSQ3);
RMSQ4=SQRT(MSQ4);
RMSQ5=SQRT(MSQ5);
RMSQ6=SQRT(MSQ6);
RMSQF2=SQRT(MSQF2);
RMSQF3=SQRT(MSQF3);
RMSQF4=SQRT(MSQF4);
RMSQF5=SQRT(MSQF5);
RMSQF6=SQRT(MSQF6);
PROC PRINT;
```