S R D   W O R K I N G   P A P E R

THE EFFECTS OF OMITTING ACREAGE QUESTIONS AND MODIFYING THE OPERATION DESCRIPTION SECTION IN CATTLE SURVEYS

by

Jack Nealon

Statistical Research Division
Economics and Statistics Service
U.S. Department of Agriculture
Washington, D.C.   20250

November, 1980

THE EFFECTS OF OMITTING ACREAGE QUESTIONS AND MODIFYING THE OPERATION DESCRIPTION SECTION IN CATTLE SURVEYS; by Jack Nealon; Statistical Research Divison; Economics and Statistics Service; U.S. Department of Agriculture; Washington, D.C.; November, 1980.

ABSTRACT

This report summarizes the impact of omitting acreage questions and modifying the operation description section in cattle surveys. Results indicated that acreage questions can probably be omitted without significantly affecting the survey estimates and that the test version of the operation description section can probably be used.

Key words: Replicates, Average Significance Level, Univariate and Multivariate Analyses of Variance.

ACKNOWLEDGEMENTS

The author would like to thank the Colorado, Georgia, Indiana, Iowa, Kansas, Minnesota, South Dakota and Wyoming offices for the special effort they made to provide the necessary data for this study.

CONTENTS

SUMMARY

Research was conducted in eight states during the January 1980 Multiple Frame Cattle Survey. The purpose of the research was to determine (1) if acreage questions can be removed from the operational questionnaire and (2) if the operation description section can be modified. Results indicated that the acreage questions can probably be removed and the operation description section modified without significantly affecting the level of the survey estimates.

INTRODUCTION

A test version of the cattle list questionnaire was examined in eight states during the January 1980 Multiple Frame (MF) Cattle Survey. This questionnaire version was identical in form to the version studied during the December 1979 MF Hog Survey [1, pgs. 22-25]. The objectives of the research in the eight states during the MF Cattle Survey were the same as the objectives of the seven-state hog research conducted during the December 1979 MF Hog Survey. The objectives were: (1) determine if the acreage questions can be removed from the operational questionnaire without significantly affecting the level of the estimates or significantly increasing the refusal rate, and (2) determine if a modification in the operation description section can be implemented without significantly influencing the estimates. The reasons for testing a questionnaire with the acreage questions removed and the operation description section modified are given in the report summarizing the findings of the seven-state hog research [1, pgs 1-2].

SURVEY DESIGN

The following eight states were involved in this study: Colorado, Georgia, Indiana, Iowa, Kansas, Minnesota, South Dakota and Wyoming. In each state, the list sample was systematically divided into two groups. One group received the operational questionnaire, and the other group received the alternate or test questionnaire. The list strata included in the analysis for each state and the number of completed reports in each state and the eight states combined are shown in Table 1. The strata include all strata except the no livestock and cattle strata, unbounded extreme operator strata such as 6000 + cattle, the no information stratum in Minnesota and stratum 17 in Kansas. Stratum 17 in Kansas was deleted since no operators in this stratum were assigned the test questionnaire.

Table 1

List Strata and Completed Reports in Each State

State	Strata	Completed Reports
Colorado	2-17	1,288
Georgia	3-17	1,225
Indiana	2, 4-16	1,256
Iowa	3-16	1,048
Kansas	3-16	1,062
Minnesota	5-17	1,367
South Dakota	3-16	1,018
Wyoming	2-16	968
Eight States Combined		9,232

ANALYSIS

The same analytical methods were used in this study as were employed in the hog research [1, pgs. 3-4]. Within each state, observations were assigned randomly to one of ten replicates. Therefore, within a given state there were ten replicates times two questionnaire versions or 20 replicate estimates for the state level analyses of variance. Since there were eight states involved in the research, there were 20 x 8 or 160 replicate estimates available for the eight-state analyses of variance.

It was just mentioned that each sample observation was randomly assigned to one of the ten replicates. To protect against the effect of the random assignment, the random assignment process was done five times. The analyses of variance were then generated for each of the five random assignments to replicates. Significance levels discussed later are the average significance levels from the five random replicate assignments. Average significance level was defined simply as the average or mean of the five significance levels. Average significance levels less than or equal to .10 were considered significant.

Effect on Refusal Rate:

The first issue to be addressed is whether the refusal rate was significantly different between the questionnaire versions. If an operator refused to provide any data or did not supply enough data so that the report was usable, the operator was considered a refusal. It was assumed that the test on refusal rate measured the reaction of the operators to the removal of the "acres in operation" question asked prior to the cattle and calves inventory questions. Changes in the operation description section should not have influenced the refusal rate, because this section was completed after inventory data was obtained. The effect on refusal rate is of great concern because a significantly higher refusal rate for the test version would justify continued usage of the operational version.

The refusal rate for a state was defined as $\Sigma\, r_h \frac{N_h}{N}$ where r_h is the number of mail, telephone and personal interview refusals divided by the total number of interviews and N_h/N is the stratum weight. Inaccessibles were not included in the calculation of the refusal rate since these operators neither responded nor refused, but were simply not found. Known zero operations were not included in the calculation since they were precoded and therefore not given an opportunity to use one of the questionnaire versions.

A univariate analysis of variance was run using each of the five replicate assignments within each state and across all eight states to determine if the refusal rate was significantly different between questionnaire versions. The refusal rate is presented for each questionnaire version by state and for the eight states combined in Table 2. Also shown in this table is the average significance level from the five random assignments of observations to the replicates for each state and the eight states combined.

Table 2

Refusal Rate for Each Questionnaire Version and the Average Significance Level for Each State and the Eight States Combined 1/

State	Refusal Rate (%)		Average Significance Level
	Operational	Test	
Colorado	11.65	11.34	.865
Georgia	2.42	2.39	.988
Indiana	3.93	4.60	.556
Iowa	17.27	13.07	.076*
Kansas	10.05	12.23	.324
Minnesota	9.87	9.48	.799
S. Dakota	22.62	24.24	.529
Wyoming	7.92	8.10	.922
Eight States Combined	10.83	10.51	.960

1/ Average significance level $\leq$ .100 was considered significant. The symbol, *, was used to denote significance.

At the eight state level, the refusal rate was not significantly different between questionnaire versions. At the individual state level, there was no significant difference in seven of the eight states. A significant difference only occurred in Iowa, but the refusal rate was greater for the operational version. Therefore, it appears that the removal of the "acres in operation" question did not adversely influence the refusal rate.

Effect on Proportion of Zero Cattle Operations and on Survey Items:

Since the estimates for the proportion of zero cattle operations and for the survey items such as cattle inventory could be affected not only by the removal of the acreage question prior to inventory questions but also by the changes in the operation description section, statistical tests made on these variables were considered as tests between the entire operational and test questionnaires.

Before analyzing survey items that come directly from the questionnaire, the variable, proportion of operations with zero cattle, was examined. If the questionnaire versions tend to yield a significantly different number of zero cattle operations, then it is possible that the omission of the acreage question and/or the changes in the operation description section affected the number of times operators reported that they had no cattle.

A univariate analysis of variance was run using each of the five replicate assignments within each state and across the eight states. Known zero operators were excluded from the analysis since these operators were never given one of the questionnaires. In Table 3 the proportion of operations with zero cattle is given by questionnaire version and the average significance level is stated for each state and the eight states combined. The percentages shown in Table 3 are the expanded or weighted percentages of zero cattle operations from the list strata analyzed.

At the eight state level, the proportion of zero cattle operations was not significantly different between questionnaire versions. Of the eight states Indiana was the only state where there was a significant difference. The proportion of zero cattle operations for the test version was higher than the operational version in Indiana.

Table 3

Proportion of Operations with Zero Cattle for Each Questionnaire and the Average Significance Level for Each State and the Eight States Combined 1/

State	Proportion of Zero Cattle Operations		Average Significance Level
	Operational	Test	
Colorado	41.27	41.98	.792
Georgia	37.00	31.95	.133
Indiana	47.22	52.31	.074*
Iowa	19.63	19.48	.928
Kansas	20.36	22.03	.562
Minnesota	45.61	44.85	.781
S. Dakota	23.46	21.16	.449
Wyoming	19.19	17.37	.578
Eight States Combined	33.71	34.06	.785

1/ Average significance level $\leq$ .100 was considered significant. The symbol, *, was used to denote significance.

The next step in the analysis was to examine the effect of the questionnaire versions on selected survey items. The four survey items tested were:

1) Cattle inventory

2) Beef and milk cows

3) Cattle on feed

4) Calves born since January 1, 1979

A univariate analysis of variance was run on each of the four survey items for each state and the eight states combined for each of the five replicate assignments. In addition, a multivariate analysis of variance using all four survey items was run for each state and the eight states combined. The Wilk's criterion for judging significance was used for the multivariate tests.

The mean value for each of the four survey items for each state and the eight states combined is shown in Table 4 for the operational and test questionnaire versions. In Table 5 the average significance level for the univariate tests on each of the four survey items and the multivariate test is given for each state and the eight states combined.

Table 4

Mean Value of Each Survey Item for Each Questionnaire for Each State and the Eight States Combined

State	Cattle Inventory		Beef and Milk Cows		Cattle on Feed		Calves Born Since January 1, 1979	
	Oper.	Test	Oper.	Test	Oper.	Test	Oper.	Test
Colorado	103.78	101.77	40.39	42.22	15.13	12.09	39.03	40.24
Georgia	43.10	45.33	23.31	23.87	0.67	1.35	19.36	20.25
Indiana	22.29	24.61	9.32	9.03	3.58	3.54	8.14	8.18
Iowa	88.55	87.44	27.21	24.72	23.48	29.28	26.55	22.67
Kansas	97.51	88.76	31.88	28.16	15.22	10.93	28.79	26.08
Minnesota	35.95	39.25	13.92	14.98	6.77	6.12	12.98	13.86
S. Dakota	116.75	126.36	53.36	57.03	5.94	8.59	51.30	53.54
Wyoming	214.86	210.26	102.90	105.10	4.57	1.53	101.99	103.00
Eight States Combined	67.28	67.83	25.58	25.25	10.48	10.90	23.95	23.37

Table 5

Average Significance Level for Each Survey Item and the Four Survey Items Combined for Each State and the Eight States Combined 1/

State	Cattle Inventory	Beef and Milk Cows	Cattle on Feed	Calves Born Since Jan. 1, 1979	Four Survey Items Combined
Colorado	.771	.604	.358	.740	.781
Georgia	.487	.760	.209	.486	.582
Indiana	.383	.730	.887	.947	.679
Iowa	.766	.199	.070*	.055*	.119
Kansas	.296	.135	.316	.241	.357
Minnesota	.275	.355	.644	.412	.574
S. Dakota	.242	.309	.185	.476	.627
Wyoming	.699	.801	.128	.881	.536
Eight States Combined	.937	.812	.712	.876	.930

1/ Average significance level $\leq$ .100 was considered significant. The symbol, *, was used to denote significance.

For the eight states combined, the univariate and multivariate tests were not significant. At the individual state level, the multivariate test was not significant in any state and the univariate tests were not significant in seven of the eight states. In Iowa, significant differences occurred for two of the survey items, namely, cattle on feed and calves born. Cattle on feed was greater for the test version and calves born greater for the operational version.

Since it is possible that a few very large observations can greatly influence the test results, it was decided to reanalyze the data set when the influence of very large observations is reduced. An estimator, $\bar{y}_t$, which replaces all sample values greater than a cutoff point, t, by the value of t itself was used. The cutoff point, t, is determined by selecting the point t where the relative efficiency of $\bar{y}_t$ to $\bar{y}$, the sample mean, is maximum. The derivation of the formulae for this estimator and the results of the outlier analysis are given in the APPENDIX.

The mean of each survey item for each questionnaire version and the average significance level from the univariate and multivariate tests are shown in Tables 6 and 7, respectively, for each state and the eight states combined when the very large observations have been trimmed. The results from the statistical tests do not differ from the results previously stated on the original data set.

Table 6

Mean Value of Each Survey Item for Each Questionnaire for Each State and the Eight States Combined When Large Observations Were Trimmed

State	Cattle Inventory		Beef and Milk Cows		Cattle on Feed		Calves Born Since January 1, 1979	
	Oper.	Test	Oper.	Test	Oper.	Test	Oper.	Test
Colorado	103.35	100.87	40.39	41.62	15.01	12.09	38.91	39.94
Georgia	42.86	45.14	23.09	23.87	0.64	1.34	19.27	20.17
Indiana	22.29	23.79	9.32	8.95	3.58	3.33	8.14	8.11
Iowa	87.71	87.41	26.70	24.67	23.13	29.24	25.84	22.67
Kansas	97.51	87.85	31.70	27.92	15.00	10.93	28.66	25.87
Minnesota	35.66	38.72	13.91	14.83	6.41	6.12	12.94	13.76
S. Dakota	116.75	125.43	53.36	56.49	5.94	8.24	51.13	53.35
Wyoming	214.86	208.93	102.87	104.27	3.94	1.53	101.86	102.26
Eight States Combined	67.28	67.68	25.56	25.22	10.40	10.90	23.95	23.36

Table 7

Average Significance Level for Each Survey Item and the Four Survey Items Combined for Each State and the Eight States Combined When Large Observations Were Trimmed 1/

State	Cattle Inventory	Beef and Milk Cows	Cattle on Feed	Calves Born Since Jan. 1, 1979	Four Survey Items Combined
Colorado	.714	.725	.379	.782	.830
Georgia	.465	.646	.186	.471	.619
Indiana	.524	.660	.699	.968	.628
Iowa	.875	.258	.055*	.090*	.136
Kansas	.241	.127	.337	.227	.421
Minnesota	.274	.400	.797	.429	.643
S. Dakota	.280	.364	.233	.472	.701
Wyoming	.617	.851	.128	.896	.566
Eight States Combined	.933	.824	.810	.875	.948

1/ Average significance level $\leq$ .100 was considered significant. The symbol, *, was used to denote significance.

Effect on Partnerships:

A different approach was used on the test questionnaire in the operation description section to detect partnership arrangements. This approach is shown in the APPENDIX of the seven-state hog study [1, pgs. 22-25]. The variable selected to compare the detection of partnerships between questionnaire versions was partnership arrangements reported by operations sampled as individual operations. A univariate analysis of variance was run on partnerships reported by sampled individuals as a percentage of all positive cattle operations for each of the five replicate assignments. Analysis was limited to positive cattle operations because a change in the operation type does not affect the estimate if an operator has no cattle.

The proportion of the positive reports for each questionnaire version where partnerships were reported by sampled invididuals and the average significane level for each state and the eight states combined are given in Table 8. The estimates from the operational and test versions were not significantly different for any of the individual states and at the eight state level.

Table 8

The Proportion of Sampled Individual Operations that Reported Partnership Arrangements for Each Questionnaire and the Average Significance Level for Each State and the Eight States Combined (Positive Reports Only) 1/

State	Proportion of Individual Operations Reporting Partnership Arrangements		Average Significance Level
	Operational	Test	
Colorado	3.76	3.83	.945
Georgia	4.56	5.85	.490
Indiana	8.26	11.16	.207
Iowa	6.33	6.72	.880
Kansas	5.87	5.32	.752
Minnesota	4.50	6.12	.432
S. Dakota	2.35	1.95	.666
Wyoming	4.29	2.28	.190
Eight States Combined	5.46	6.13	.568

1/ Average significance level ≤ .100 was considered significant. The symbol, *, was used to denote significance.

CONCLUSIONS AND RECOMMENDATIONS

Results indicated that the removal of the "acres in operation" question did not significantly increase the refusal rate. The estimate of the proportion of zero cattle operations and the estimates of the four selected survey items were rarely significantly different between the operational and test questionnaires. Finally, the proportion of individual operations reporting partnership arrangements was not significantly different between the operation description sections. Therefore, the acreage questions can probably be removed from the operational questionnaire and the test version of the operation description section implemented without significantly affecting the survey estimates. This statement implies only that the estimates should not change significantly and not that the estimates are or are not accurate.

REFERENCES

1/ Nealon, Jack, "The Effects of Omitting Acreage Questions and Modifying the Operation Description Section in Hog Surveys," Economics, Statistics and Cooperatives Service, USDA, 1980, Washington, D.C.

APPENDIX

INTRODUCTION

When testing an alternative questionnaire version for surveys such as the Hog or Cattle Multiple Frame (MF) Surveys, the test results may be highly influenced by the presence of a few very large observations. It is assumed that these very large observations are actually true observations, but it is not known if the questionnaire version caused these observations to be very large or if the observations would have been very large regardless of which questionnaire version they were randomly assigned. Because the second viewpoint is highly possible, the approach the Survey Research Section has taken has been to conduct the analysis first using all the data and then analyze the data with the very large or outlier observations removed. For recent studies, an observation was considered an outlier if its Z score was greater than a specified positive integer. For the seven state study conducted during the September 1978 MF Hog Survey, an observation was considered an outlier if $Z > 5$ [A1]. For the seven state study during the December 1979 MF Hog Survey, a $Z > 8$ defined the observation as an outlier [A2]. Clearly, the definition of an outlier has been subjective. The purpose of this paper is to describe an objective outlier definition for a single-stage stratified sample design such as the hog or cattle single-stage stratified sample designs. This method will then be applied to data collected in eight states for the January 1980 MF Cattle Survey.

THE ESTIMATOR

An estimator, $\bar{y}_t$, which is formed by replacing all sample values larger than a cutoff point, t, by the value of t itself, was developed by Searls [A3] for simple random sampling. The cutoff point, t, is determined by selecting the point t where the relative efficiency of $\bar{y}_t$ to $\bar{y}$ is maximum. The formulae for computing $\bar{y}_t$ and $\bar{y}$ are $\bar{y}_t = \frac{\sum_{i=1}^{r} y_i + (n - r)t}{n}$, where $r \leq n$ and $y_i \leq t$, and $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$.

The relative efficiency of $\bar{y}_t$ to $\bar{y}$ is given by R.E. $= \frac{V(\bar{y})}{MSE(\bar{y}_t)} \times 100$, where

$V(\bar{y}_t) = \frac{p}{n} [\sigma_t^2 + q (t - \mu_t)^2]$ and $MSE(\bar{y}_t) = V(\bar{y}_t) + (p\mu_t + qt - \mu)^2$.

The quantities μ_t and σ_t^2 are the mean and variance of the truncated distribution, respectively, p is the proportion of population values $\leq t$, $q = 1 - p$ and μ is the population mean of the untruncated distribution. It was shown by Searls that a region of t always exists where the relative efficiency of $\bar{y}_t$ to $\bar{y}$ is greater than 100.

Two other estimators were found in the literature which adjust for sample observations > t. The first estimator, $\bar{y}_{t,W}$, is obtained by assigning a reduced weight to observations > t. The second estimator, $\bar{y}_{t_d}$, is simply the mean of all observations below the cutoff. Ernst [A4] proved that among these three estimators, that is, $\bar{y}_t$, $\bar{y}_{t,W}$ and $\bar{y}_{t_d}$, that $\bar{y}_t$ is the best in the sense that given any t and W there exists t_1 and t_2 such that $MSE(\bar{y}_{t_1}) \leq MSE(\bar{y}_{t,W})$ and $MSE(\bar{y}_{t_2}) \leq MSE(\bar{y}_{t_d})$. Therefore, the estimator, $\bar{y}_t$, as developed by Searls, was selected as the estimator to be investigated.

The formulae given by Searls were extended to a single-stage stratified sample design since this is the type of design ESS encounters in their split-list tests. Since hypothesis testing is done at the state level and at the combined or aggregate state level for split-list testing, an optimum outlier definition will be generated for each of these two inference levels. The classical formulae for the mean and variance in a single-stage stratified random sample design and the mean and mean square error for the same design when values > t are replaced by the value of t itself will now be detailed at the state level and then the aggregate level.

The estimate of the mean at the state level is

$$\bar{y}_s = \sum_{h=1}^{L_s} W_{sh}\, \bar{y}_{sh}\ , \text{ where } s = 1, 2, \ldots, M,$$

M = the number of states,

L_s = the number of strata in state s,

$$\bar{y}_{sh} = \sum_{i=1}^{n_{sh}} \frac{y_{shi}}{n_{sh}} \quad \text{and}$$

$$W_{sh} = \frac{N_{sh}}{N_s}\ .$$

The estimate of the variance of the mean at the state level is

$$V(\bar{y}_s) = \sum_{h=1}^{L_s} W_{sh}^{\,2}\, V(\bar{y}_{sh}), \text{ where } V(\bar{y}_{sh}) = \frac{S_{sh}^{\,2}}{n_{sh}} (1 - f_{sh}),$$

$$S_{sh}^{\,2} = \sum_{i=1}^{n_{sh}} \frac{(y_{shi} - \bar{y}_{sh})^2}{n_{sh} - 1} \quad \text{and}$$

$$f_{sh} = \frac{n_{sh}}{N_{sh}}\ .$$

The state estimate of the mean formed by replacing all sample values larger than a cutoff point, t_{sh}, by the value of t_{sh} itself is

$$\bar{y}'_s = \sum_{h=1}^{L_s} W_{sh}\, \bar{y}'_{sh}\ , \quad \text{where} \quad s = 1, 2, \ldots, M\ ,$$

$$\bar{y}'_{sh} = \frac{\sum_{i=1}^{r_{sh}} y_{shi} + (n_{sh} - r_{sh})\, t_{sh}}{n_{sh}}\ ,$$

r_{sh} is the number of sample values less than or equal to t_{sh} and

$y_{shi} \le t_{sh}$.

The state estimate of the variance of $\bar{y}'_s$ is

$$V(\bar{y}'_s) = \sum_{h=1}^{L_s} W_{sh}^2 \, V(\bar{y}'_{sh}),$$

where $V(\bar{y}'_{sh}) = \frac{p_{sh}}{n_{sh}} (1 - f_{sh}) \{S'^2_{sh} + q_{sh} [t_{sh} - \bar{y}'_{sh}]^2\}$,

$\bar{y}'_{sh}$ and S'^2_{sh} are the mean and variance of the truncated sample distribution,

p_{sh} is the proportion of sample values $\leq t_{sh}$ and

$q_{sh} = 1 - p_{sh}$.

The state estimate of the mean square error of $\bar{y}'_s$ is

$$MSE(\bar{y}'_s) = V(\bar{y}'_s) + \left\{ \sum_{h=1}^{L_s} W_{sh} \, b_{sh} \right\}^2 = V(\bar{y}'_s) + b_s^2$$

where $b_{sh} = p_{sh} \bar{y}'_{sh} + q_{sh} t_{sh} - \bar{y}_{sh}$ = estimate of the bias in the estimator.

Similarly, the estimate of the mean and variance of the mean when the state data is aggregated are

$$\bar{y} = \sum_{s=1}^{M} W_s \bar{y}_s \quad \text{and} \quad V(\bar{y}) = \sum_{s=1}^{M} W_s^2 \, V(\bar{y}_s)$$

where $W_s = \frac{N_s}{N}$.

The estimate of the mean and mean square error of the mean at the aggregate level formed by replacing all sample values $> t_{sh}$ by t_{sh} itself are

$$\bar{y}' = \sum_{s=1}^{M} W_s \bar{y}'_s \quad \text{and} \quad MSE(\bar{y}') = \sum_{s=1}^{M} W_s^2 \, V(\bar{y}'_s) + \{ \sum_{s=1}^{M} W_s \, b_s \}^2 .$$

Finally, the relative efficiency at the state level for cutoff points, t_{sh};

$h = 1, 2, \ldots, L_s$, is $RE_s = \dfrac{V(\bar{y}_s)}{MSE(\bar{y}_s')} * 100$

and at the aggregate level is $RE = \dfrac{V(\bar{y})}{MSE\ (\bar{y}')} * 100$

for cutoff points t_{sh}; $h = 1, 2, \ldots, L_s$ and $s = 1, 2, \ldots, M$.

The values for t_{sh} which maximize RE_s for each state provide an optimum outlier definition at the state level. The values for t_{sh} which maximize RE provide an optimum outlier definition at the aggregate level.

When computing the $MSE(\bar{y}_s')$ or $MSE(\bar{y}')$ for the purpose of arriving at the maximum RE_s or RE, a very large number of calculations would be required if the number of cutoff points, t_{sh}, examined and the number of strata are large. For example, if n cutoff points are examined in each stratum and there are m strata, then the number of calculations of the mean square error would be n^m. Therefore, to reduce computer expenses, the constraint that

$$t_{sh}/\bar{y}_{sh} = \text{constant for } h = 1, 2, \ldots, L_s$$

can be imposed on each stratum within a given state when computing $MSE(\bar{y}_s')$ or on each stratum when computing $MSE(\bar{y}')$. With this constraint only n calculations rather than n^m are necessary within a state. This was the approach taken to arrive at the outlier definitions for the data from the January 1980 MF Cattle Survey.

OUTLIER ANALYSIS

The same data set was used for the outlier analysis that is shown in Table 1. Outliers were examined for the four survey items (cattle inventory, beef and milk cows, calf crop and cattle on feed) at the state and aggregate levels. The relative efficiency was computed separately at the state and aggregate levels for various values of $t_{sh}/\bar{y}_{sh}$ in order to determine for each level the region $t_{sh}/\bar{y}_{sh}$ where the relative efficiency was greater than 100. Once the regions were located at the state and aggregate levels, the optimum value of $t_{sh}/\bar{y}_{sh}$ within one decimal point was determined where the relative efficiency was maximum. Recall that the constraint is being made that $t_{sh}/\bar{y}_{sh}$ is constant from stratum to stratum.

Shown in Table A-1 for each of the four variables for each state are the optimum value of $t_{sh}/\bar{y}_{sh}$, the number of outliers or very large observations, the bias as a percentage of the mean and the relative efficiency at the state level. Notice from Table A-1 that the value $t_{sh}/\bar{y}_{sh}$ often varied considerably from state to state and between cattle on feed and the other survey items. The bias as a percentage of the mean was less than 1 percent in most instances. Finally, the relative efficiency was always greater than 100.

The results at the aggregate (eight-state) level are given in Table A-2. Again the value of $t_{sh}/\bar{y}_{sh}$ varied considerably for some of the survey items. The bias as a percentage of the mean was negligible and the relative efficiency was always greater than 100.

Table A-1: The Value of $t_{sh}/\bar{y}_{sh}$, the Number of Outliers, the Bias as a Percentage of the Estimate and the Relative Efficiency for Each State by Survey Item

Survey Item	State	$\frac{t_{sh}}{\bar{y}_{sh}}$	Number of Outliers	$(\frac{b_s}{\bar{y}_s}) * 100$	Relative Efficiency
Cattle Inventory	Colorado	11.6	4	0.64	104.15
	Georgia	7.5	5	0.36	101.74
	Indiana	20.3	4	1.77	114.35
	Iowa	3.8	4	0.27	101.55
	Kansas	7.7	2	0.47	103.71
	Minnesota	10.4	3	1.08	108.15
	S. Dakota	7.0	3	0.38	102.75
	Wyoming	7.1	5	0.32	102.41
Beef and Milk Cows	Colorado	12.7	3	0.72	106.75
	Georgia	9.3	3	0.48	103.03
	Indiana	20.0	3	0.48	101.79
	Iowa	7.1	3	0.25	100.76
	Kansas	5.9	10	0.51	101.06
	Minnesota	10.6	5	0.54	102.26
	S. Dakota	6.3	5	0.49	104.27
	Wyoming	8.5	3	0.42	104.13
Calf Crop	Colorado	9.8	5	0.53	103.20
	Georgia	8.7	5	0.42	102.38
	Indiana	21.7	2	0.49	101.94
	Iowa	7.5	4	0.50	103.16
	Kansas	6.4	7	0.40	101.51
	Minnesota	8.9	9	0.48	100.92
	S. Dakota	6.7	4	0.35	102.39
	Wyoming	8.6	3	0.43	103.95
Cattle on Feed	Colorado	128.0	1	0.46	100.32
	Georgia	95.0	3	1.87	100.85
	Indiana	66.0	2	3.03	111.15
	Iowa	22.5	3	0.66	101.83
	Kansas	79.0	1	0.88	102.53
	Minnesota	76.0	1	2.80	116.86
	S. Dakota	56.0	1	2.31	108.09
	Wyoming	114.0	2	9.99	132.49

Table A-2: The Value of $t_{sh}/\bar{y}_{sh}$, the Number of Outliers, the Bias as a Percentage of the Estimate and the Relative Efficiency for Each Survey Item for the Eight States Combined

Survey Item	$\frac{t_{sh}}{\bar{y}_{sh}}$	Number of Outliers	$(\frac{b_s}{\bar{y}_s}) * 100$	Relative Efficiency
Cattle Inventory	20.2	4	0.12	101.08
Beef and Milk Cows	18.3	4	0.07	100.21
Calf Crop	30.8	3	0.03	100.10
Cattle on Feed	113.5	7	0.39	101.56

REFERENCES

[A1] Hall, K. B., "The Effect of Telephone Enumerators and Evaluating a New Operation Description," ESCS, USDA, September 1979, Washington, D.C.

[A2] Nealon, J., "The Effects of Omitting Acreage Questions and Modifying the Operation Description Section in Hog Surveys," ESCS, USDA, March 1980, Washington, D.C.

[A3] Searls, D. T., "An Estimator for a Population Mean Which Reduces the Effect of Large True Observations," JASA, December 1966.

[A4] Ernst, L. R., "Some Results on the "Large" Observation Problem," Bureau of the Census, U.S. Department of Commerce, March 1978.