COMPOSITE ESTIMATORS IN THE JUNE HOG SERIES

Charles R. Perry, Paul W. Cook, Martin L. Holko, Steve D. Wiyatt and Thomas R. Birkett
National Agricultural Statistics Service, USDA
Charles R. Perry, NASS/USDA, South Agri. Bld., Washington, D.C. 20250

KEYWORDS: hog series, composite estimation

ABSTRACT

The goal of this study was to improve the precision of the statistical estimates used by the Agricultural Statistics Board of the National Agricultural Statistics Service, a committee of senior statisticians from headquarters and major state offices, in setting official hog and pig inventory statistics. Six composite estimators were evaluated using historical data from the June survey. Each composite was a weighted average of the National Agricultural Statistics Service farm, tract, weighted, and multiple frame estimators.

Three methods of analyses were used in the evaluation: multivariate analyses of composite biases with respect to national board estimates; nonparametric analyses of composite biases, average absolute differences, standard deviations, and root mean square errors with respect to national board estimates; and model interpretations of historical board actions in terms of the six composites.

All three methods of analyses indicated that the smoothed inverse variance composite performs as well as or better than the other composites in all but one major state. This outcome was not surprising because the smoothed inverse variance weighting formula more nearly conforms to the weighting formula for the theoretically optimal composite than any of the other six composites.

This summary paper is based on Technical Report No. 88-09, 1988, NASS/USDA, which is available on request from the authors.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to Jacquline Penn for her assistance throughout this project. Jacquline's knowledge of the historical hog series and her untiring effort to provide reliable data has prevented many errors. Her enthusiasm has made this project more enjoyable. We also wish to thank Ben Klugh, George Hanuschak, and Bill Pratt for their assistance with this research and documentation.

SUMMARY

The purpose of this research was to improve the precision of the statistical estimates used by the Agricultural Statistics Board (board) of the National Agricultural Statistics Service (NASS) in setting official hog and pig inventory statistics. Six composite estimators were evaluated using historical data from the June survey for eight years from eight major hog producing states. Each composite was a weighted average of the NASS farm, tract, weighted, and multiple frame estimators.

Three methods of analyses were used in evaluating the six composites: multivariate analyses of composite biases with respect to national board estimates; nonparametric analyses of composite biases, average absolute differences, standard deviations, and root mean square errors with respect to national board estimates; and model interpretations of historical board actions in terms of the six composites.

Each of these three methods of analyses indicated that the smoothed inverse variance composite performs as well as or better than the other six composites. Iowa was the only exception to this conclusion. In Iowa, the midrange outperformed the other composites. Multivariate analyses of the biases relative to board values indicated that in most states those composites depending most heavily on the multiple frame estimator (multiple frame, inverse variance, and smoothed inverse variance composites) were least biased.

Nonparametric analyses strongly indicated that the smoothed inverse variance composite was the "best" composite when all four evaluation criteria were considered. Model interpretation analyses indicated that, with the exception of Iowa, past board actions were most closely mimicked by either the inverse variance or smoothed inverse variance composite.

On the basis of this research, the following recommendations were made: 1. that the smoothed inverse variance composite estimates be the primary indications used by the individual states and the national board in setting June hog and pig inventory statistics, 2. that the revised summary system provide estimates of the covariance matrix associated with the farm, tract, weighted, and multiple frame indication for each hog series item, and 3. that further research be conducted to determine how best to relate the smoothed inverse variance composite in June to the hog series estimates produced in the three other quarters.

INTRODUCTION

The objective of this study was to evaluate six composite estimators in the context of the hog series using historical farm, tract, weighted, and multiple frame summary statistics from the June survey. The original analysis plan included summary statistics from the December Enumerative Survey (DES), but these data were excluded from consideration when the agency eliminated the DES series from the NASS program. The study was motivated by three primary considerations:

1. The need to bring practice closer to theory, since there always exists a composite estimator which has variance less than or equal to any of its component estimators,
2. The desire to have a method of combining a set of estimates based on statistical theory, and
3. The desire to have a repeatable method of combining survey data for board use that is not influenced by changes in the membership of the board.

Six composite estimators were evaluated in the context of the June survey for eight hog series items. There were three major parts to the evaluation:

1. Multivariate analyses of variance for differences in biases relative to national board estimates,

2. Nonparametric analyses of biases, average absolute differences, standard deviations, and root mean square errors relative to national board estimates, and
3. Model interpretation of historical board actions -- first, the composite weights that best predict historical board statistics were calculated, then the nearest of the six composites was determined.

All the composite estimators investigated were weighted averages of the NASS farm, tract, weighted, and multiple frame estimators used in the June survey. Each composite can be written symbolically as:

$$Y_c = W_1Y_1 + W_2Y_2 + W_3Y_3 + W_4Y_4$$

where the weights W_1, W_2, W_3, and W_4 sum to one and Y_1, Y_2, Y_3, and Y_4 represent the June farm, tract, weighted, and multiple frame indications, respectively, for a specified hog series item.

The six composites evaluated were chosen on the basis of practical and theoretical considerations. The composites are listed below along with a short descriptive name to facilitate future reference. The multiple frame is included in this listing because it is treated as a composite in some analyses.

1. Equal: Each indication is weighted equally.
2. Inv.var: Each indication is weighted proportional to the inverse of its estimated variance.
3. Inv.cv: Each indication is weighted proportional to the inverse of its estimated coefficient of variation.
4. Mid.range: The largest and smallest indication are weighted by one half and all other indications by zero.
5. S.inv.var: Each indication is weighted proportional to an exponential smoothed historical average of the inv.var weights defined for composite 2 above.
6. S.inv.cv: Each indication is weighted proportional to an exponential smoothed historical average of the inv.cv weights defined for composite 3 above.
7. Multi.frame: The multiple frame indication is weighted by one.

DESCRIPTION OF DATA SETS

Two data sets were used in this study: one data set to calculate the six composites and one to evaluate the composites. Changes in the June Survey program weight groups and variation in the states included in Multiple Frame States left a usable data set for eight states -- Illinois, Indiana, Iowa, Kansas, Minnesota,Missouri, Nebraska, and Ohio for the eight years 1979 through 1986. The estimators and items contained in the data sets are summarized below.

The indications data set contained estimates for:

1. Farm Direct Expansion
2. Tract Direct Expansion
3. Weighted Direct Expansion
4. Multiple Frame Direct Expansion

The evaluation data set contained estimates for:

1. First state recommendation
2. First national board estimate
3. Second state recommendation
4. Second national board estimate
5. Third state recommendation
6. Third national board estimate

Each data set contained the items:

1. Total hogs
2. Market hogs
   a. Less than 60 pounds
   b. 60 through 119 pounds
   c. 120 through 179 pounds
   d. 180 pounds and greater
3. Hogs for Breeding
4. Previous quarter pig crop (births)

ANALYSES

The next three sub-sections present results that characterize the behavior of the six composite estimators relative to historical board estimates. Each section presents a different approach to the basic question: How do the differences between the six composite estimates and historical board values behave?

Multivariate Analysis for Biases

Multivariate analyses of variance techniques were used to determine if, relative to board values, some composites were less biased than others. This determination involved two sets of analyses. One set of analyses (state) was used to determine if differences between the composites and national board values differed significantly by either composite, board, or state. Another set of analyses (regional) was used to determine if differences between the composites and national board values differed significantly by either composite, board, or method used to compute the regional composite estimate.

The two methods of computing regional composite estimates were: 1. compute the composite estimates for the individual states and sum the results over states to get a regional composite estimate or 2. compute the composite estimate directly from the regional indications (when the weights are constant both methods yield the same values). The three board estimates were: initial, first revision, second revision. The analyses were multivariate because differences for each of the eight hog item categories (total, breeding, market, under 60, 60-119, 120-179, 180 and up, and pig crop) were analyzed simultaneously.

The regional analyses showed that for some hog item categories both the composite and the method used to calculate the regional composite estimates significantly effected the differences between the composite estimates and the board estimates. However, these differences were not significantly effected by the board value used in their computation. For the first method of computing a regional composite (sum the state composites), there was a significant composite effect in five of the eight hog item categories -- total, breeding, market, under60, and 120-179. For the second method of computing a regional composite (composite the regional indications), there was a significant composite effect in four of the eight hog item categories -- total, breeding, market, and under60. In general, the midrange composite and the multiple frame showed the least bias.

The state analyses showed that both the composites and the states significantly effected the differences and that these effects changed from state to state. The multiple frame is statistically different from other composites for the total, breeding, market, under60, 60-119, and the 120-179 categories for all states except Minnesota and Missouri. No composite was statistically different for any state for the 180up and the births categories. Biases were examined by determining the least biased composite for each state and hog item category. Tabulating the number of times each composite was the least biased by state or by category gave the following frequencies: five equal weight, nine inverse variance, four inverse coefficient of variation, 12 midrange, six smoothed inverse variance, four smoothed inverse coefficient of variation, and 24 multiple frame. Six of the 12 times the midrange was least biased occurred in Iowa and 21 of the 24 times the multiple frame was least biased occurred in four states: Illinois, Indiana, Minnesota, and Missouri. There were no other patterns discerned in these analyses.

Nonparametric Analysis for Four Criteria

Nonparametric ranks analyses were used to evaluate the composites at the state and regional level for four evaluation criteria: bias, absolute difference, standard deviation, and root mean square error. In all cases, the second revised national board estimates were assumed to be the actual population values. The three sets of state analyses were: across states for each hog item category, across hog item categories for each state, and across states and hog item categories.

The regional analyses were across hog item categories for each of the two methods of computing a regional composite. In all analyses the methodology was essentially the same: Average the ranks over one or more classification variables and then test for significant differences among the averages. The methodology of transforming the original data to ranks and then analyzing the ranks by standard multivariate procedures to produce nonparametric tests are described in the works of Conover and Iman.

The regional analyses showed that in terms of bias, absolute difference, and root mean square error and for equal treatment of all four evaluation criteria the mid.range was the best performing composite -- has the smallest average rank. In terms of standard deviation, the s.inv.var was the best performing composite. The results were essentially the same for both methods of computing a regional composite. For some of the evaluation criteria there was a significant difference between the mid.range and the s.inv.var.

The state analyses across hog item categories showed no consistent pattern across states. However, the inv.var or the s.inv.var composite was the minimal mean rank composite in five, four, four, five, and five states, of the eight states, for the five evaluation criteria: absolute bias, absolute difference, root mean square error, standard deviation, and equal treatment of all four evaluation criteria, respectively.

The state analyses across states showed a clear pattern. For most of the eight hog item categories, the minimal mean rank composite was either the inv.var or the s.inv.var for each of the four evaluation criteria. The inv.var or s.inv.var was the minimal mean rank composite for all evaluation criteria for the three major hog item categories: total, market, and breeding hogs. Table 1 below gives the average rank over states for equal treatment of the four evaluation criteria for each hog item category.

The state analyses across states and hog item categories identified the s.inv.var composite as minimal mean rank composite for all evaluation criteria. Table 2 below gives the results for equal treatment of all four evaluation criteria.

The state analyses strongly suggested that the smoothed inverse variance composite was the composite of choice relative to board values. This choice was especially true when the four evaluation criteria were given equal weight. However, there were differences between the results from the regional and the across state analyses. These differences were primarily due to the influence of Iowa in the regional analyses. In the state analyses each of the eight states was treated equally. In the regional analyses, each state was treated proportional to its size. These two different methods of weighting the state data and the fact that not all states behaved the same in the state analyses tended to explain the differences between the two sets of analyses.

Composite No.	Name	total	mrkt	brd	undr-60	60-119	120-179	180 up	pig crop
1	equal	6.0	5.5	6.2	5.0	6.1	6.2	5.0	4.8
2	inv.var	3.1	3.2	2.9	4.0	2.8	2.0*	4.5	3.8
3	inv.cv	3.7	3.4	3.8	3.4	3.9	4.0	4.3	3.3
4	mid.rang	5.9	5.4	5.9	5.1	6.0	5.8	4.2	5.0
5	s.inv.vr	2.3*	2.7*	2.1*	2.7*	2.4*	2.6	3.1*	3.4*
6	s.inv.cv	3.9	3.8	4.0	3.4	4.0	4.5	3.3	3.6
7	multi.fm	3.2	4.0	3.1	4.4	2.8	2.9	3.5	4.1

Table 1: The mean rank over states for equal treatment of absolute bias, absolute difference, root mean square error, and standard deviation. Asterisks denote the minimal mean ranked composite.

Composite		Average Rank	Significant differences
1	equal	5.6	52736
2	inv.var	3.3	1465
3	inv.cv	3.7	145
4	mid.range	5.4	52367
5	s.inv.var	2.7*	231467
6	s.inv.cv	3.8	1452
7	multi.frame	3.5	14

Table 2: The mean rank and significant differences over all states and categories for equal treatment of absolute bias, absolute difference, root mean square error, and standard deviation. The relative magnitude of the significant differences is given by the left to right ordering of the composites in the last column.

Modeling interpretation

Models were used to interpret historical board actions in terms of the six composites. There were two steps to the procedure which were performed independently for each state and the eight state region. First, determine the best fitting composite model for each board series and each composite series -- this associates a unique set of weights with each board series and each composite series. Second, determine for each of the six board series the composite series whose weights were nearest in terms of four-dimensional Euclidean distance. The model interpretation related each board series with the composite series that had the closest weighting structure. Summary results for the eight states are given in Table 3. Results for the eight state region are given in Table 4. The numbering convention used for the board estimates and State recommendations (which have been referred to simply as board series) are given in the second note to Table 3.

The results given in Tables 3 and 4 indicate that, with one exception, the board's treatment of states has been essentially the same. With the exception of Iowa, all states were best modeled by either the inverse variance composite or by the smoothed inverse variance composite. The choice between these two composites matters little since they are almost the same in construction and behavior. The similarity between the eight state regional results and the Iowa results; and, the dissimilarity between the eight state regional results and the seven other individual states suggested that Iowa dominates the eight state regional results.

COMPOSITE	BOARD 1	2	3	4	5	6	SUM
equal							
inv.var	5	2	2	2	2	2	15
inv.cv							
mid.range	1	1	1	1			4
s.inv.var	2	5	5	5	5	5	27
s.inv.cv					1	1	2

Table 3: The total number of states for which each composite was nearest the indicated board estimate. Iowa, the only state not best modeled by the s.inv.var of the s.inv.cv, corresponds to the 1's in the body of the table. The first state recommendation, the first national board, the second state recommendation, the second national board, the third state recommendation, and the third national board are denoted by board 1,2,3,4,5,and 6 respectively.

COMPOSITE	BOARD 1	2	3	4	5	6	SUM
equal						*	1
inv.var							
inv.cv							
mid.range		*	*	*			3
s.inv.var							
s.inv.cv	*				*		2

Table 4: The eight state regional composite that is nearest the regional board (sum of board estimates or state recommendations).

CONCLUSIONS

All three methods of analyses indicated that the smoothed inverse variance composite performs as well as or better than the other six composites evaluated. Iowa was the exception to this conclusion. In Iowa the midrange outperformed the other composites. Multivariate analyses of the biases relative to board values indicated that in most states those composites depending most heavily on the multiple frame estimator (multi.frame, inv.var, and s.inv.var composites) were least biased.

Nonparametric analyses strongly suggested that the s.inv.var composite was the "best" composite when all four evaluation criteria were considered. Model interpretation analyses indicated that, with the exception of Iowa, past board actions were most closely mimicked by either the inverse variance or smoothed inverse variance composite.

In many of the analyses reported in the study there were differences between those conducted at the regional level and those conducted across states. These differences were generally explained when the effect of Iowa on the analyses was examined more closely.

In summary, the analyses indicated that the smoothed inverse variance composite performs as well as or better than the other composites examined in all but one major state. This outcome was not surprising because the smoothed inverse variance composite weighting formula more nearly conforms to the weighting formula for the theoretically optimal composite than any of the other six composites.

GRAPHS OF SMOOTHED INVERSE VARIANCE COMPOSITE ESTIMATES AND NATIONAL BOARD ESTIMATES FOR TOTAL NUMBER OF HOGS

Figures 1-9 of the Appendix graphically display the total number of hog for the smoothed inverse variance composite and the third national board estimates of total hogs for the eight states and the eight state region. The 90% confidence intervals superimposed on the graphs were derived from multiple frame statistics. The intervals should provide a rough approximation to the appropriate confidence intervals, if the results from analyses of one small data set for one of the major hog producing states holds up for other states. The analyses for the one state showed that for the years 1982 through 1986 the multiple frame and the smoothed inverse variance composite had essentially the same length confidence intervals. The appropriate confidence intervals were not computed because covariance matrices necessary for their computation were not available from historical NASS data -- the current summary system does not produce covariances estimates.

RECOMMENDATIONS

1. That the smoothed inverse variance composite estimates be the primary indication used by the individual states and the national board in setting June hog and pig inventory statistics.
2. That the smoothed inverse variance composite estimates be computed at the state and national levels for all hog items; and that the same composite estimates be computed for all states.
3. That estimators for the covariance matrix associated with the farm, tract, weighted, and multiple frame hog item be included in the next revision to the summary system, if the cost is not prohibitive. These covariances will permit statistical estimates of the variance associated with the smoothed inverse variance composite.

4. That further research be conducted to determine how best to relate the smoothed inverse variance composite in June to the hog series estimates produced in the three other quarters. This research is necessary since the indications are only available in June to compute the composite.

Computing the farm, tract, weighted, and multiple frame indications in all states will support the computation of a true national balance sheet estimate with known statistical properties. With estimates for the covariances between the various indications, the national balance sheet estimate and the state composites can then be used to implement a statistically based state reallocation process.

BIBLIOGRAPHY

1. Conover, W.J. and Iman, R.L., "On Some Alternative Procedures Using Ranks for the Analysis of Experimental Designs", Communications in Statistics, A5, 14, 1976.
2. Conover, W.J. and Iman, R.L., "Rank Transformations as a Bridge between Parametric and Nonparametric Statistics", The American Statistician, 35, 1981.
3. Graybill, F.A. and Deal, R.B., "Combining Unbiased Estimators", Biometrics, Dec. 1959.
4. Houseman, E.E., "Composite Estimation", Unpublished Note, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C. 20250, 1969.
5. Kuo, Lynn, "Composite Estimation of Totals for Livestock Surveys", National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C. 20250, Staff Report No. SF&SRB-92, 1986.
6. Perry, C.R., Cook, P.W., Holko, M.L., Wiyatt, S.D., and Birkett, T.R. "Composite Estimators in the June Hog Series", National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C. 20250, Staff Report No. SRB 88-09, 1988.
7. Raj, Des, Sampling Theory, McGraw-Hill, 1968.
8. Wigton, W.H. and Van der Vaart, R.H., "A Note on Combining Estimates", Unpublished Note, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C. 20250, 1973.
9. Wolter, K.M., "Composite Estimation in Finite Populations", JASA, Sept. 1979.

APPENDIX: TOTAL HOG GRAPHS

Figures 1-9 graphically display the total number of hog for the smoothed inverse variance composite and the third national board estimates of total hogs for the eight states and the eight state region. The national board estimate is denoted by B; the smoothed inverse variance composite is denoted by C; and the confidence intervals are approximately 90%.

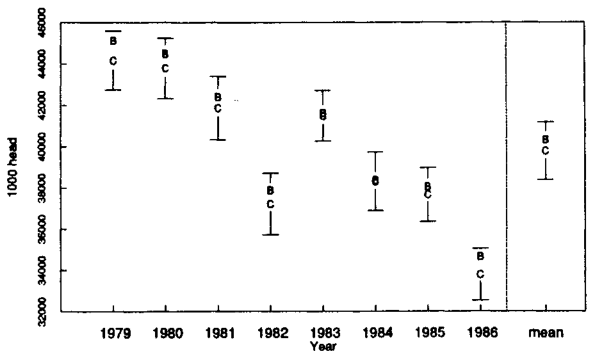

Figure 1: Eight Region Total Hogs.

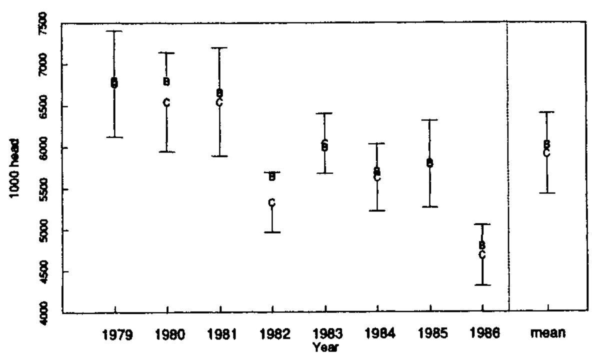

Figure 2: Illinois Total Hogs.

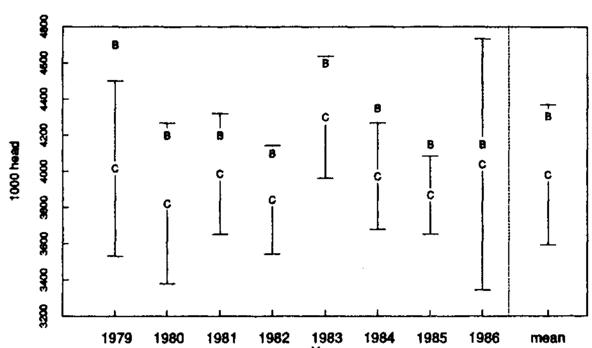

Figure 3: Indiana Total Hogs.

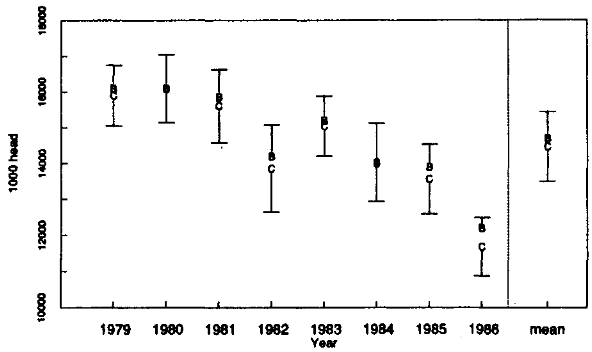

Figure 4: Iowa Total Hogs.

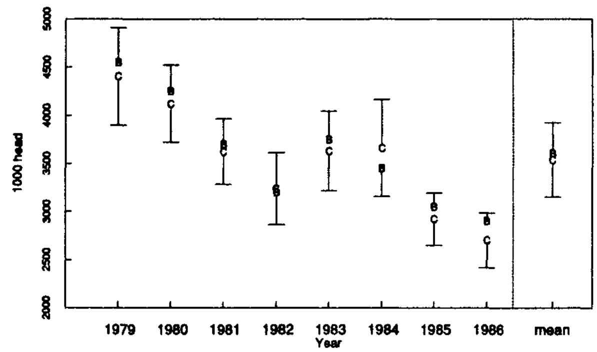

Figure 7: Missouri Total Hogs.

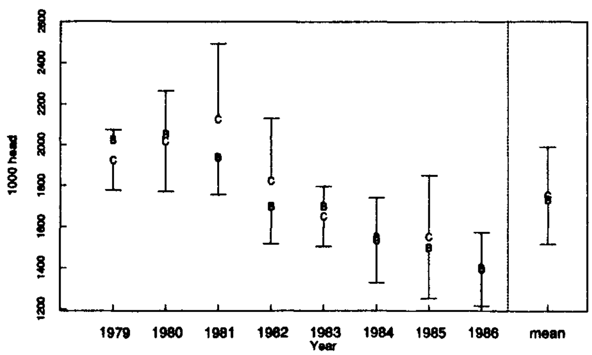

Figure 5: Kansas Total Hogs.

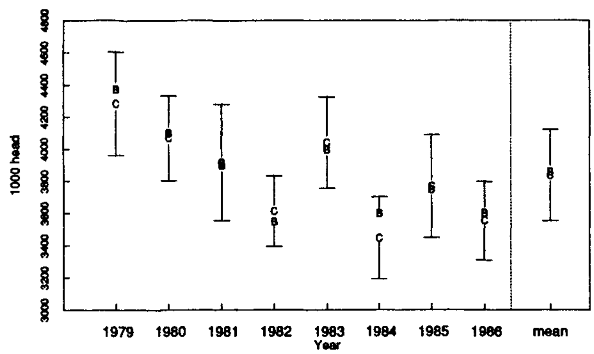

Figure 8: Nebraska Total Hogs.

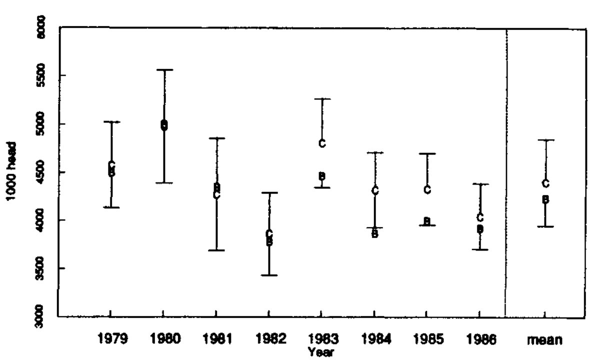

Figure 6: Minnesota Total Hogs.

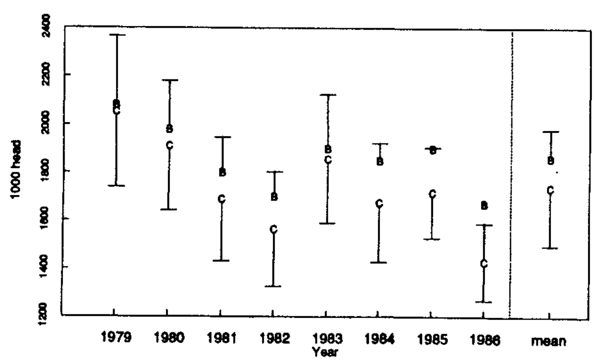

Figure 9: Ohio Total Hogs.