A BETTER ALTERNATIVE TO THE COLLAPSED STRATUM VARIANCE ESTIMATE

Gary M. Shapiro and David V. Bateman, U. S. Bureau of the Census

## I. INTRODUCTION

In sample designs with one sample primary sample unit (PSU) per stratum, a collapsed stratum variance estimator is generally employed. Theory and empirical evidence are presented in this paper in support of the premise that a without replacement variance estimator [3] produces an estimate of the variance with both smaller bias and smaller variance than a collapsed stratum variance estimator [4]. Section II explains the premise of the paper more completely, introduces notation, and gives both intuitive and theoretical arguments for why a without replacement estimator should result in a lower bias. Section III presents some empirical data which shows one particular without replacement estimator to have smaller bias and smaller variance. Finally, section IV summarizes the paper, recommends that a without replacement variance estimator be used instead of a collapsed stratum estimator, and draws some inferences from the earlier sections pertaining to the sample design on the question of whether to select one or two PSU's per stratum.

## II. BASIC PREMISE AND THEORY

Surveys are frequently designed and conducted with one PSU selecter per stratum. In such a design, no unbiased estimate of variance is possible. Generally, some form of the collapsed stratum variance estimator is employed. For the estimator, pairs of strata are formed for those strata comprised of more than one PSU. The aim is to pair strata with similar characteristics and approximately equal measures of size. The form of the estimator considered in this paper is for simple unbiased estimates. However, the basic principles also apply when the collapsed strata are used with more sophisticated weighting in conjunction with replication or linearized variance estimation. Also, to keep things simple and eliminate extraneous concerns, it is assumed throughout this paper that a census is conducted within each sample PSU, i.e., there is no within PSU variance, and between PSU variance is equal to the total variance.

Let $y_{ijk}$ = estimate for characteristic of interest for $k^{th}$ PSU in $j^{th}$ stratum within $i^{th}$ pair of strata.

$j$ takes on values of only 1 and 2 and thus $(i,j)$ denotes a unique stratum.

Since a census of sample PSU's is being assumed, $y_{ijk}$ is the estimate obtained from a census of the $k^{th}$ PSU in stratum $(i,j)$.

$L_{ij}$ = the number of PSU's in the $(i,j)^{th}$ stratum.

$N_{ijk}$ = the measure of size for the $k^{th}$ PSU in $(i,j)^{th}$ stratum.

$N_{ij} = \sum_{k}^{L_{ij}} N_{ijk}$ $(i,j)^{th}$ stratum total.

$P_{ijk} = \frac{N_{ijk}}{N_{ij}}$ probability of selecting the $k^{th}$ PSU in a single draw within the $(i,j)^{th}$ stratum.

We are interested in an estimated total,

$$\hat{y} = \sum_i \sum_j \frac{1}{P_{ijk}} y_{ijk} .$$

$P'_{ijk} = N_{ijk}/N_i$ probability of selecting the $k^{th}$ PSU assuming a single draw within the $i^{th}$ stratum.

$\pi_{ijk} = 2P'_{ijk}$ probability of selecting the $k^{th}$ PSU from the $i^{th}$ stratum assuming a Durbin selection of 2 PSU's in the stratum.

$$\pi_{i1k_1,2k_2} = \frac{2P'_{i1k_1}P'_{i2k_2}}{\lambda_i}\left[\frac{1}{1-2P'_{i2k_2}} + \frac{1}{1-2P'_{i1k_1}}\right]$$

joint probability of selecting the $k_1$ and $k_2$ PSU's in the $i^{th}$ stratum assuming a Durbin selection method.

$$\lambda_i = 1 + \sum_j \sum_k \frac{P'_{ijk}}{1-2P'_{ijk}} .$$

The true one PSU per stratum variance of

$$\hat{y} \text{ is } VAR_T = \sum_i \sum_j \sum_k \frac{N_{ijk}}{N_{ij}}\left(\frac{y_{ijk}N_{ij}}{N_{ijk}} - y_{ij}\right)^2 \qquad (1)$$

The usual form of the collapsed stratum variance estimator [4] is:

$$\widehat{VAR}_{CS} = \sum_i \left(\frac{2N_{i2}}{N_i}\,\frac{y_{i1k_1}N_{i1}}{N_{i1k_1}} - \frac{2N_{i1}}{N_i}\,\frac{y_{i2k_2}N_{i2}}{N_{i2k_2}}\right)^2 \qquad (2)$$

where $k_1$ is the sample PSU in $(i,1)$ stratum, and

$k_2$ is the sample PSU in $(i,2)$ stratum.

The collapsed stratum variance estimator acts as if each pair of collapsed strata had actually been one stratum in the first place, and two sample PSU's had been selected with replacement with probability proportionate of $N_{ijk}$.[1] Had such a sample design actually been implemented, (2) would be an unbiased estimate of variance.[2] For one sample PSU per stratum sample designs, however, when $N_{i1} = N_{i2}$, i.e., the two paired strata have equal measures of size, (2) is always an overestimate of variance.

When $N_{i1} \neq N_{i2}$, (2) doesn't have to be an overestimate of variance, but if the stratification is somewhat effective and/or if $N_{i1} \doteq N_{i2}$ (2) will generally be an overestimate (see [1]).

The basic point of this paper is that a variance estimator with a smaller variance and a smaller bias can be used instead of the collapsed stratum estimator. Again, assuming that the stratification used is somewhat effective, or that $N_{i1} \doteq N_{i2}$, a variance estimator that assumes that two sample PSU's have been selected without replacement will have a smaller bias. The rationale for this is sample: Collapsed stratum assumes selection of two sample PSU's with no restrictions whatsoever; a without replacement estimator assumes a slight restriction, namely that no PSU be selected twice; the true variance in this case is obtained when even more stringent restrictions hold, namely, that only those combinations of two sample PSU's are allowed in which one PSU comes from one stratum of the collapsed pair and the second PSU comes from the other stratum. Thus, without replacement falls in between the extremes and thus should have expected value between the two. A without replacement variance estimator can be used that utilizes work done by Durbin or Hartley-Rao in estimating the joint probabilities of selection. The discussion in this paper will be with regards to Durbin probabilities since we feel that it is likely to be subject to low variance compared with alternative estimators.

Some theory supporting the intuitive discussion above is now presented. The expected value of the collapsed stratum variance estimator is:

$$VAR_{CS} = \sum_{i} \sum_{k_1}^{L_{i1}} \sum_{k_2}^{L_{i2}} P_{i1k_1} P_{i2k_2} \left( \frac{2N_{i2}}{N_i} \frac{y_{i1k_1} N_{i1}}{N_{i1k_1}} - \frac{2N_{i1}}{N_i} \frac{y_{i2k_2} N_{i2}}{N_{i2k_2}} \right)^2 \quad (3)$$

The expected value of the Yates-Grundy without replacement variance estimator using Durbin probabilities[3] is:

$$VAR_{DUR} = \sum_{i} \sum_{k_1}^{L_{i1}} \sum_{k_2}^{L_{i2}} P_{i1k_1} P_{i2k_2} \left( \frac{\pi_{i1k_1}\pi_{i2k_2} - \pi_{i1k_1,2k_2}}{\pi_{i1k_1,2k_2}} \right) \left( \frac{y_{i1k_1}}{\pi_{i1k_1}} - \frac{y_{i2k_2}}{\pi_{i2k_2}} \right)^4 \quad (4)$$

Both the expected value and the sample estimate of the Durbin estimator is less than that of collapsed stratum for a particular stratum, when $\frac{\pi_{i1k_1}\pi_{i2k_2}}{\pi_{i1k_1,2k_2}} < 2$, providing that $N_{i1} = N_{i2}$.

Although this inequality need not hold, it usually does. For example, we found that it failed to hold for only 4 of 65 possible sample PSU combinations in the Longitudinal Manpower Survey, a recurring survey conducted by the Census Bureau for the Employment and Training Administration.

In the more general case when $N_{i1} \neq N_{i2}$, a complex inequality results for the expected value although the same simple inequality still applies for the sample estimates. The complex inequality and its derivation are given in appendix A.

To summarize the above theory, we have established the following for a particular stratum for the situation when

$$N_{i1} = N_{i2} \text{ and } \frac{\pi_{i1k_1}\pi_{i2k_2}}{\pi_{i1k_1,2k_2}} < 2 :$$

1) $VAR_{CS}$ is an overestimate of the variance.

2) $E(VAR_{DUR}) \leq E(VAR_{CS})$ and the sample estimate for the Durbin estimator is less or equal to the sample collapsed stratum estimate.

We have not, however, been able to show that the expected Durbin estimator is greater than the true variance. Thus, it is possible that the Durbin estimator could result in an underestimate of variance and thus could conceivably have a larger bias than collapsed stratum. It is intuitively clear, however, that if there is any gain in efficiency from selecting one sample PSU per stratum instead of two sample PSU's per collapsed strata, then the Durbin estimator should result in an overestimate of variance. In this circumstance, Durbin is definitely subject to a smaller bias than collapsed stratum.

## III. EMPIRICAL RESULTS

Empirical results are available at this point only for rural South Dakota. It was convenient to make comparisons for one State that was restratified for the 1976 expansion of the Current Population Survey (CPS). Thus, the results presented here are for that part of South Dakota that was non-self-representing in the CPS. We plan to do more extensive empirical investigations in the future. The list of PSU's for the 12 strata and the six collapsed strata is given in table 1 of appendix B. The 1970 census population estimate in these 12 strata was 416,633 and the 1960 census estimate of unemployment was 10,207. There were restrictions on the restratification for the State based on the stratification used for CPS prior to expansion. The main criteria for restratification, however, was a regression estimate of unemployment rate. The most important variables in the regression for South Dakota were percent nonwhite, proportion of service workers, and proportion in service industry. (See [2] for more details on the restratification.) The collapsed pairs of strata were formed by pairing strata according to their mean regression estimates of unemployment rate, with no regard for the population of the strata. 1970 Census data was used to form the regression estimates. Variances were estimated for the 1960 census estimate of total unemployment. Thus, the same characteristic was used as the characteristic of interest and the basic stratification variable, but with a 10-year time difference to keep the correlation from being unreasonably high.

Table A compares the true variance, the expected value of the Durbin variance estimator (Formula (4)), and the expected value of the collapsed stratum variance estimator (Formula (3)) for these six pairs of strata.

TABLE A. EXPECTED VALUES OF 2 VARIANCE ESTIMATORS COMPARED TO TRUE VARIANCE

| True Variance (1) | Expected Value of Durbin (2) | Expected Value of Collapsed Stratum (3) | $\|(1) - (2)\|$ | $\|(1) - (3)\|$ |
|---|---|---|---|---|
| 453,416 | 513,618 | 722,863 | 60,202 | 269,447 |

As expected, both the collapsed stratum and Durbin estimators are biased upwards, but the bias of the Durbin estimator is relatively small.

Two hundred sets of 12 sample PSU's one from each stratum, were selected with probability proportionate to 1970 census populations. Each of the 200 selections was independent, so that the same selections could be repeated more than once. For each set, the Durbin variance estimate, the collapsed stratum variance estimate (2) and the actual (deviation)$^2$ of the sample estimate from truth (the true error for the particular sample) were computed. The formula used for the Durbin estimate was:

$$\widehat{VAR}_D = \sum_{i=1}^{6}\left(\frac{\pi_{i1k_1}\pi_{i2k_2} - \pi_{i1k_1,2k_2}}{\pi_{i1k_1,2k_2}}\right)\left(\frac{y_{i1k_1}}{\pi_{i1k_1}} - \frac{y_{i2k_2}}{\pi_{i2k_2}}\right)^2 \qquad (5)$$

The formula used for the actual deviation for a particular set of PSU's was

$$DEV_K^2 = \left(\sum_{i=1}^{6}\sum_{j}\frac{y_{ijk}N_{ij}}{N_{ijk}} - y\right)^2 \qquad (6)$$

Table B compares the results of these quantities for the 200 samples combined and for four subsets of 50 samples each. The first 17 of the 200 samples are shown individually in table 2 of appendix B.

TABLE B. SOME SAMPLE ESTIMATES OF THE DURBIN VARIANCE ESTIMATE, THE COLLAPSED STRATUM VARIANCE ESTIMATE, AND THE ACTUAL DEVIATION SQUARED

| Means | Durbin Variance Estimate (1) | Collapsed Stratum Variance Estimate (2) | Actual (Deviation)$^2$ (3) | $\frac{1}{G}\sum_{g}^{G}\|(1_g) - (3)_g\|$[1] (4) | $\frac{1}{G}\sum_{g}^{G}\|(2_g) - (3_g)\|$[2] (5) |
|---|---|---|---|---|---|
| 1st 50 | 566,675 | 801,067 | 413,489 | 471,707 | 663,417 |
| 2nd 50 | 485,260 | 691,737 | 703,515 | 633,677 | 703,502 |
| 3rd 50 | 494,408 | 706,444 | 441,684 | 434,163 | 542,283 |
| 4th 50 | 530,387 | 757,799 | 373,376 | 508,031 | 686,132 |
| All 200 samples | 519,183 | 739,262 | 483,016 | 511,895 | 648,834 |

[1]The absolute values of the Durbin estimate minus the actual deviation squared averaged over the sets of sample PSU's.

[2]The absolute values of the collapsed stratum estimate minus the actual deviation squared averaged over the sets of sample PSU's.

The average Durbin estimate agrees quite closely to its expected value; the collapsed stratum estimate also agrees quite closely to its expected value, and the average actual (deviation)$^2$ is quite close to the true variance. Most importantly, comparing columns (4) and (5) shows that the Durbin estimate tends to be somewhat closer to the actual (deviation)$^2$ than collapsed stratum does. A word of explanation on the meaning and importance of these columns is needed. The (deviation)$^2$ for a particular sample is the "true error" for that sample. In the best of all worlds, one would like to know the "true error" for each sample estimate. If the variance estimator resulted exactly in the (deviation)$^2$ for each sample, one would know this "true error." Thus, one measure of a variance estimator's worth is how close it comes to the (deviation)$^2$. This is what columns (4) and (5) show for the Durbin and collapsed stratum estimators respectively; e.g., the first entry in column (4) was calculated by taking the absolute difference between the Durbin estimate and the (deviation)$^2$ for each of the first 50 samples, and then taking the average of the 50 absolute differences. Quite impressively, of the 200 samples, in only 57 (28.5 percent) was the collapsed stratum estimate closer to the actual (deviation)$^2$ than the Durbin estimate.

Table C shows the estimated variances on the Durbin estimate and collapsed stratum estimate, and the absolute differences between Durbin and the $(\text{deviation})^2$ (column 4 of table B) and between collapsed stratum and the $(\text{deviation})^2$ (column 5 of table B). The formula used for the first variance is as follows:

$$\text{VAR(Durbin Estimate)} = \sum_{g}^{200}\left(\text{VAR}_D(g) - \frac{\sum_{g}^{200}\text{VAR}_D(g)}{200}\right)^2 \Bigg/ 199.$$

Similar formulae were used for the other variances.

TABLE C. COMPARISON OF THE VARIANCE OF VARIANCE ESTIMATES

(ALL NUMBERS X $10^9$)

| | |
|---|---|
| Variance (Durbin Estimate) | 106 |
| Variance (Collapsed Stratum Estimate) | 195 |
| Variance $\lvert(1_g) - (3_g)\rvert$ (From table B) | 287 |
| Variance $\lvert(2_g) - (3_g)\rvert$ (From table B) | 292 |

The most important comparison in this table is between the Durbin and the collapsed stratum estimators. The much lower variance for Durbin indicates that not only is Durbin subject to a smaller bias than collapsed stratum, but it is also more stable.

In summary, the empirical evidence points unequivocally to Durbin being substantially preferable to collapsed stratum in every possible respect. Its expected value is better, it is generally closer to the actual $(\text{deviation})$,[2] and it is subject to a smaller variance. For all of these, the differences between Durbin and collapsed stratum are relatively large.

The results found here may not, of course, hold for all situations. In general, we think that Durbin should be better than collapsed stratum with respect to its expected value. It may not, however, be closer to the actual $(\text{deviation})^2$ and may not be subject to a smaller variance in many situations. In this empirical example, collapsing of strata was done without regard to $N_{i1}$ being close to $N_{i2}$. This could have led to collapsed stratum performing unusually poorly here, although only one of the pairings turned out highly unequal ($N_{11}$= 31,269 and $N_{12}$= 48,336), and this pairing could not have been improved upon greatly (the second highest stratum population to have matched with the 48,336 stratum had only 37,718 population.

## IV. CONCLUSION

Traditionally, a collapsed stratum variance estimator has been used in conjunction with a one PSU per stratum sample design. The purpose of this paper is to propose an alternative to this type of estimator. Conceptually, and in the case of this empirical study, it appears that a better variance estimate can be obtained by using a variance estimator associated with a design for two PSU's per stratum drawn without replacement. Due to the complex formulae involved, we have not been able to develop a theoretical model from a mean square error viewpoint that will present the conditions under which the Durbin estimator is preferred over the collapsed stratum estimator. Certainly this would have been a more definitive approach to the problem, and we would encourage further research along these lines. At the present time, we recommend that a variance estimator for two PSU's per stratum selected without replacement be used for a one PSU per stratum sample design.

A major decision that has to be made when one is designing a sample survey is whether or not to stratify PSU's beyond a two PSU per stratum selection method. Historically, a major drawback in stratifying to a point where only one PSU is selected from a stratum is the inability to get an unbiased variance estimate. The relatively small bias and variance of the Durbin estimator as indicated by this empirical study would suggest that this may no longer be such an important argument.


## V. ACKNOWLEDGMENTS

The idea for this paper developed from comments/questions made by Earl Bryant, National Center for Health Statistics, to the first author. The empirical data resulted from programming done by Susan J. Dellinger. Significant improvements were made to the paper as the results of comments by Henry F. Woltman, Robert Fay, Kirk Wolter, and Nash Monsour of the Census Bureau. The typing was done by Edith L. Oechsler.


## FOOTNOTES

[1]A collapsed stratum variance estimator with a finite correction factor which would not assume with replacement selection could be used; the properties of such an estimator have not been studied. Also, this statement is not completely accurate in a technical sense when $N_{i1} \neq N_{i2}$.

[2]This statement, like the preceding statement, is not completely accurate in a technical sense when $N_{i1} \neq N_{i2}$.

[3]For the remainder of this paper, this variance estimator will be referred to simply as the Durbin estimator.

## REFERENCES


[1] Capps, Grant, "Some Interesting Results in Unequal Probability Sampling from a Finite Population." Proceedings of the Social Statistics Section; American Statistical Association, 1977, pp. 301-307.

[2] Dippo, Cathryn, "Expansion of CPS to Provide Reliable State Estimates of Unemployment." Proceedings of the Social Statistics Section; American Statistical Association, 1975, pp. 387-391.

[3] Durbin, J. "Design of Multi-Stage Surveys for the Estimation of Sampling Errors." Applied Statistics, 1967, Vol. XVI, pp. 152-164.

[4] Hansen, Morris H., Hurwitz, William N., and Madow, William G., Sample Survey Methods and Theory. (New York: Wiley, 1953) Vol. I, Chap. 9, pp. 399-401

[5] Yates, F. and Grundy, P. M., "Selection Without Replacement from Within Strata with Probability Proportional to Size." Journal of the Royal Statistical Society, Series B (Methodological), 1953, Vol. XV, pp. 253-261.

APPENDIX A.

To show under what conditions the expected Durbin variance is less than the expected collapsed stratum estimator:

$$\mathrm{Var}_{CS} = \sum_i \sum_{k_1}^{L_{i1}} \sum_{k_2}^{L_{i2}} P_{i1k_1} P_{i2k_2} \left( \frac{2N_{i2}}{N_i} \frac{N_{i1}}{N_{i1k_1}} y_{i1k_1} - \frac{2N_{i1}}{N_i} \frac{N_{i2}}{N_{i2k_2}} y_{i2k_2} \right)^2$$

$$= \sum \sum \sum \frac{4N_{i1k_1}}{N_{i1}} \frac{N_{i2k_2}}{N_{i2}} \frac{N_{i1}^2 N_{i2}^2}{N_i^2} \left( \frac{y_{i1k_1}}{N_{i1k_1}} - \frac{y_{i2k_2}}{N_{i2k_2}} \right)^2$$

$$= \sum_i \sum_{k_1} \sum_{k_2} \left[ \frac{4(N_{i1})\,(N_{i2})\,(N_{i1k_1})(N_{i2k_2})}{N_i^2} \right] \left[ \frac{y_{i1k_1}}{N_{i1k_1}} - \frac{y_{i2k_2}}{N_{i2k_2}} \right]^2$$

In terms of Durbin notation

$$= \sum_i \sum_{k_1} \sum_{k_2} \left[ (N_{i1})(N_{i2})\; \pi_{i1k_1} \pi_{i2k_2} \right] \left[ \frac{y_{i1k_1}}{N_{i1k_1}} - \frac{y_{i2k_2}}{N_{i2k_2}} \right]^2$$

---

$$\mathrm{Var}_{DUR} = \sum_i \sum_{k_1}^{L_{i1}} \sum_{k_2}^{L_{i2}} \frac{1}{4} \frac{N_{i1k_1}}{N_{i1}} \frac{N_{i2k_2}}{N_{i2}} \left[ \frac{\pi_{i1k_1}\pi_{i2k_2} - \pi_{i1k_1,2k_2}}{\pi_{i1k_1,2k_2}} \right] \left[ \frac{N_i}{N_{i1k_1}} y_{i1k_1} - \frac{N_i}{N_{i2k_2}} y_{i2k_2} \right]^2$$

$$= \sum \sum \sum \frac{N_i^2}{N_{i1}N_{i2}} \frac{N_{i1k_1}N_{i2k_2}}{4} \left[ \frac{\pi_{i1k_1}\pi_{i2k_2} - \pi_{i1k_1,2k_2}}{\pi_{i1k_1,2k_2}} \right] \left[ \frac{y_{i1k_1}}{N_{i1k_1}} - \frac{y_{i2k_2}}{N_{i2k_2}} \right]^2$$

For a particular stratum the expected Durbin variance will be less than the expected collapsed variance if

$$\frac{N_i^2 N_{i1k_1} N_{i2k_2}}{N_{i1}N_{i2}\; 4} \left[ \frac{\pi_{i1k_1}\pi_{i2k_2}}{\pi_{i1k_1,2k_2}} - 1 \right] < 4N_{i1}N_{i2} \frac{N_{i1k_1}}{N_i} \frac{N_{i2k_2}}{N_i} .$$

If $N_{i1} = N_{i2}$ for every i, then we get

$$\left[ \frac{\pi_{i1k_1}\pi_{i2k_2}}{\pi_{i1k_1,2k_2}} - 1 \right] < 1 \quad \text{which implies} \quad \frac{\pi_{i1k_1}\pi_{i2k_2}}{\pi_{i1k_1,2k_2}} < 2.$$

In the more general case where $N_{i1} \neq N_{i2}$,

$$\frac{\pi_{i1k_1}\pi_{i2k_2}}{\pi_{i1k_1,2k_2}} < \frac{16 N_{i1}^2 N_{i2}^2}{N_i^4} + 1 .$$

APPENDIX B.

TABLE 1. PSU's, STRATA, AND COLLAPSED PAIRS OF STRATA IN SOUTH DAKOTA USED TO PRODUCE THE EMPIRICAL DATA

| Collapsed Pair No. | Stratum No. | PSU Description (Counties) |
|---|---|---|
| 1 | 1 | Brule |
| 1 | 1 | Haakon |
| 1 | 1 | Hand |
| 1 | 1 | Jackson |
| 1 | 1 | Marshall |
| 1 | 1 | Perkins |
| 1 | 1 | Potter |
| 1 | 464 | Aurora-Douglas |
| 1 | 464 | Clark-Hamlin |
| 1 | 464 | Custer |
| 1 | 464 | Faulk |
| 1 | 464 | Harding |
| 1 | 464 | Jerauld-Sanborn |
| 1 | 464 | Lyman |
| 1 | 464 | McPherson |
| 1 | 464 | Sully |
| 2 | 2 | Hughes |
| 2 | 2 | Spink |
| 2 | 2 | Edmunds |
| 2 | 2 | Gregory |
| 2 | 3 | Union |
| 2 | 3 | Charles Mix |
| 2 | 3 | Kingsbury-Miner |
| 3 | 4 | Tripp |
| 3 | 4 | Beadle |
| 3 | 4 | Jones |
| 3 | 5 | Codington-Deuel |
| 3 | 5 | Lincoln |
| 4 | 6 | Clay |
| 4 | 6 | Lake-McCook |
| 4 | 6 | Day |
| 4 | 7 | Washabaugh |
| 4 | 7 | Campbell-Walworth |
| 4 | 7 | Hutchinson-Turner |
| 5 | 8 | Meade |
| 5 | 8 | Davison-Hanson |
| 5 | 8 | Butte |
| 5 | 9 | Grant |
| 5 | 9 | Roberts |
| 5 | 9 | Stanley |
| 6 | 10 | Lawrence |
| 6 | 10 | Buffalo-Hyde |
| 6 | 10 | Fall River |
| 6 | 10 | Mellette |
| 6 | 11 | Shannon |
| 6 | 11 | Todd |
| 6 | 11 | Bennett |
| 6 | 11 | Corson |
| 6 | 11 | Dewey |
| 6 | 11 | Ziebach |

APPENDIX B.

TABLE 2. VARIANCES FOR 17 OF THE 200 SAMPLES OF PSU's

| | SAMPLE | DURBIN VAR. EST. (1) | COLLAPSED STR. VAR. EST. (2) | DEVIATION (3) | ABS((1)-(3)) (4) | ABS((2)-(3)) (5) |
|---|---|---|---|---|---|---|
| 1. | 42 27 40 38 18 7<br>10 44 36 23 4 55 | 408022. | 809121. | 2130. | 405892. | 806991. |
| 2. | 45 32 41 25 8 22<br>21 44 36 20 4 52 | 277500. | 399665. | 381794. | 104294. | 17871. |
| 3. | 48 32 40 47 18 22<br>21 44 36 51 4 13 | 313918. | 516340. | 21471. | 292447. | 494869. |
| 4. | 48 29 17 47 18 22<br>21 10 36 6 16 55 | 609285. | 685045. | 7228. | 602058. | 677817. |
| 5. | 50 27 17 47 46 7<br>39 44 15 20 16 52 | 849878. | 1278279. | 454301. | 395578. | 823978. |
| 6. | 5 27 41 47 18 7<br>21 19 15 6 4 52 | 199902. | 338949. | 94499. | 105403. | 244449. |
| 7. | 49 28 41 47 8 7<br>39 19 15 6 4 12 | 339480. | 370799. | 431606. | 92126. | 60807. |
| 8. | 43 29 40 47 8 7<br>21 44 15 6 11 37 | 385845. | 464839. | 262478. | 123367. | 202361. |
| 9. | 50 29 17 38 18 22<br>21 19 36 20 4 52 | 378923. | 583097. | 237300. | 141623. | 345798. |
| 10. | 5 34 17 38 18 22<br>39 19 15 20 54 13 | 768940. | 836965. | 272556. | 496384. | 564409. |
| 11. | 49 29 24 38 8 7<br>21 44 36 20 54 52 | 167220. | 246954. | 2044615. | 1877395. | 1797661. |
| 12. | 50 27 41 25 8 7<br>21 44 15 20 4 53 | 205643. | 332736. | 734308. | 528665. | 401571. |