SHOULD ONE OR TWO PSU's PER STRATUM BE SELECTED?

Gary M. Shapiro and Christina L. Olsen
U. S. Bureau of the Census

I. INTRODUCTION

In determining a sample design for a survey, the statistician frequently must decide between one sample primary sampling unit (PSU) per stratum or two sample PSU's per stratum. Generally, the one PSU choice has the advantage of resulting in the lowest actual variance, i.e., the additional stratification possible for an increased number of strata and only one sample PSU per stratum normally reduces the between PSU variance. However, an unbiased variance estimator for such a design does not exist. Some form of the collapsed stratum variance estimator [1] is then generally used, usually resulting in an overestimate of variance if the stratification is somewhat effective. Thus, the two PSU per stratum design might be preferable when unbiased variance estimation is important.

In this paper, the following criterion is examined for deciding between one and two sample PSU's: If we use the survey results for hypothesis testing, which design will lead us to accept a false null hypothesis less frequently, i.e., which method will result in the smaller Type II error? This criterion is explained in detail in the next section. At this point, however, let us summarize the results of the paper: For the set of empirical data used, there was little difference between the Type II error probabilities for one and two PSU's per stratum. However, a one PSU per stratum design was not substantially better under any circumstances, whereas two PSU's was clearly preferable under some circumstances. Thus, by the Type II error criterion, two PSU's per stratum would be preferred over one PSU per stratum.

The authors do not suggest, however, that this criterion always be the sole basis for deciding between one and two sample PSU's per stratum. There are cases where hypothesis testing is irrelevant to the uses of the survey. There are also many cases where little or no perusal of variance estimates are expected to be made by data users: In these cases one sample PSU should always be used. There are many statisticians who always opt for two sample PSU's due to their view (frequently mistaken?) of the importance of unbiased variance estimates. This paper should strengthen their preference for two sample PSU's.

Normally, when comparisons of Type II error are made between methods, the Type I error (the probability of rejecting a true null hypothesis) for the methods is the same. The Type I error for the one PSU and two PSU's per stratum designs is not the same. However, since analysts in practice usually assume they are equal, the authors do not think this is a limitation in their work.

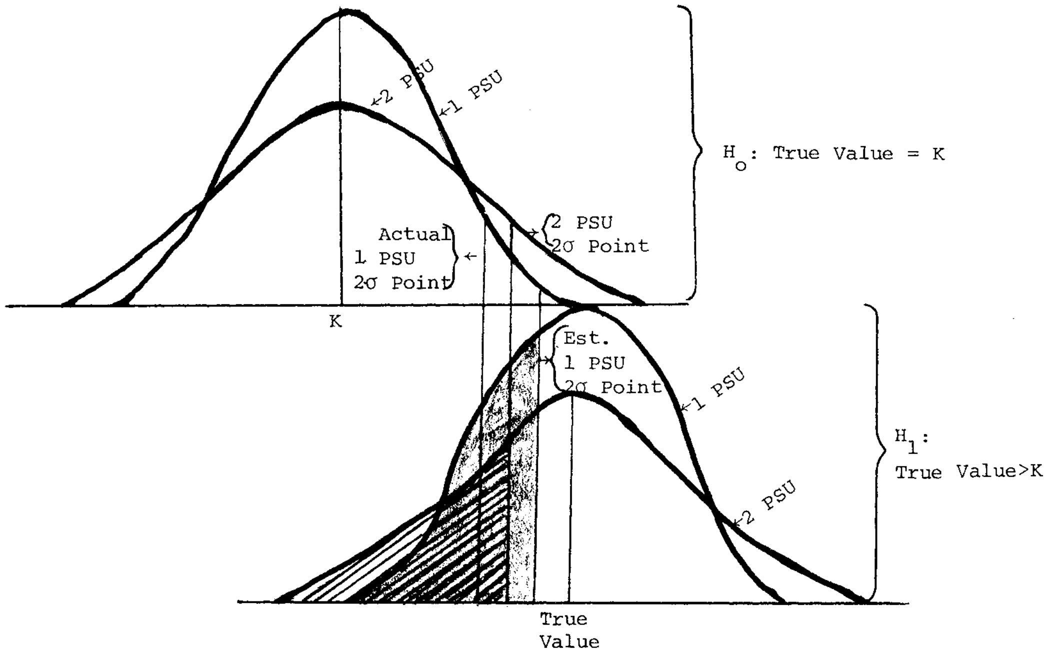
The comparisons made in this paper assume that a collapsed stratum variance estimator with no adjustment for the bias is used in conjunction with the one PSU per stratum design. If a "perfect" adjustment for bias was made, one PSU per stratum would, of course, be preferable.

No comparison is made here for the "without replacement variance estimator," which can be used with the one PSU per stratum design and is indicated in [3] to be preferable to the collapsed stratum estimator. This would clearly improve the properties for one PSU and will be the subject for a future paper planned for the 1980 American Statistical Association Convention. Also, no comparison is made for a regression variance estimator which is indicated in [2] to be preferable to the collapsed stratum estimator.

II. CRITERION STATEMENT

The exact meaning of the Type II error criterion can be best understood by studying the figure. Assume that the purpose of the survey to be designed is to determine whether a population mean (labeled "true value" in the figure) is equal to some fixed constant K (our null hypothesis) or whether the population mean is greater than K (the alternative hypothesis). We assume that both one sample PSU and two sample PSU designs result in sample estimates that are normally distributed and unbiased. On the bottom portion of the figure are two normal curves representing the actual distribution of sample estimates for one PSU per stratum and for two PSU's per stratum. The one PSU method naturally results in the more concentrated curve and in lower variance, assuming there are additional gains achieved by doubling the number of strata.

On the top portion of the figure, two normal curves represent what the distributions of sample estimates for the two designs would be under the null hypothesis that the "true value" is K. The vertical line labeled "Actual 1 PSU" represents the 2σ significance level point for the one PSU per stratum method; this line separates the acceptance region for the null hypothesis from the rejection region. If the "true value" were K, only 5 percent of all possible samples of one PSU per stratum would fall to the right of this line. The vertical line labeled "Estimated 1 PSU" represents the biased estimate of the 2σ point for the one PSU per stratum method. The statistician acts as if, for the case where the "true value" is K, 5 percent of all possible samples of one PSU per stratum would fall to the right of this line. In reality, of course, it would be less than 5 percent. Finally, the vertical line labeled "2 PSU" represents the 2σ point (actual and estimated) for the two PSU per stratum method.



Now look at the lower portion of the figure. The area under the "1 PSU method" curve to the left of the "Estimated 1 PSU" line (shaded in the figure) is where the curve representing the alternative hypothesis overlaps the acceptance region of the null hypothesis and represents the probability for the one PSU method of incorrectly deciding that the true value is K (Type II error) when, in fact, the true value occurs as shown in the figure. The area under the "2 PSU method" curve to the left of the "2 PSU" line (containing slanting lines in the figure) represents the same probability for the two PSU method. The one PSU method is preferable if the shaded area is smaller than the hatched area, and the two PSU method is preferable if the hatched area is smaller than the shaded area.

III. RESULTS

Variances calculated for one and two PSU's per stratum designs from one set of data were used to derive Type II error probabilities. The variances were compared to known theoretical relationships in order to support their use. The results of these calculations are presented in the following section and the conclusions in the last section.

A. Derivation of Table

All the data relate to the Current Population Survey (CPS), 357 Area Design. This particular design was used from March 1963 to January 1967 and consisted of 357 strata comprising 701 counties and independent cities with coverage in each of the 50 States and the District of Columbia. One hundred twelve strata consisted of 1 PSU and were self-representing, while the other 245 strata contained more than 1 PSU and were non-self-representing.

Dr. Benjamin J. Tepping, at the time with the Bureau of the Census, calculated between PSU variances for four different situations for the 245 non-self-representing strata using 1960 Decennial Census data. He calculated the actual variance under a one PSU per stratum design, the expectation of the collapsed stratum variance estimator under a one PSU per stratum design (i.e., the 1 PSU per stratum biased variance), the variance when two PSU's are selected with replacement, and the variance when two PSU's are selected without replacement, with probability proportionate to size. The actual one-PSU-per-stratum variance was based on one PSU chosen from each of the 245 noncertainty strata, and the other three variances assumed 2 PSU's chosen from each of 120 strata where, except for five cases of three strata, each of the 120 strata was formed by collapsing two of the 245 strata. Variances were calculated for 24 demographic characteristics which ranged from total employed to retail sales, to total housing units, to marriages. Ratios of the collapsed stratum variance and both two-PSU-per-stratum variances to the actual one-PSU-per-stratum variance were calculated and then averaged over the 24 characteristics. Since the calculations were based on census data, the three average relationships represent between PSU variance only. The averages were as follows:

2 PSU Method, Without Replacement: 1 PSU Method Actual = 1.23

2 PSU Method, With Replacement: 1 PSU Method Actual = 1.39

1 PSU Method, Biased: 1 PSU Method Actual = 1.72

Table 1 was derived using these three averages. It gives probabilities that one would fail to reject the hypothesis that the true value is K when the true value is not K. Probabilities are given for the four types of situations considered by Tepping. Since Tepping's variances did not include the within PSU component, we examined different relative amounts of within PSU variances, always assuming that the within PSU variance was the same for the four situations.

The 1.39 and 1.72 ratios are consistent with the following formulae, which are easily derived:

Actual variance for 1 sample PSU per stratum

$$= \sum_i (\sigma_{1i}^2 + \sigma_{2i}^2)$$

Expected value of collapsed stratum variance estimator with no adjustment for bias (1 sample PSU per stratum)

$$= \sum_i \left[\sigma_{1i}^2 + \sigma_{2i}^2 + (Y_{1i} - Y_{2i})^2\right]$$

Expected value of 2 PSU's per stratum, with replacement, variance estimator

$$= \sum_i \left[\sigma_{1i}^2 + \sigma_{2i}^2 + \tfrac{1}{2}(Y_{1i} - Y_{2i})^2\right]$$

The notation is as follows:

We have pairs of strata (1i,2i). With selection of two PSU's per stratum, each pair of strata is collapsed before two sample PSU's are selected. With selection of one PSU per stratum, the collapsing for variance estimation is according to the same pairs. It is assumed that the measures of size are the same for the two strata in each pair, and that selection of sample PSU's is with probability proportionate to the measure of size. If the measures of size are not equal, these results do not hold exactly.

Y_{ji} = The actual total for the characteristic of interest for the j^{th} stratum in the i^{th} pair of strata.

σ_{ji}^2 = Actual variance for the total Y_{ji} for the j^{th} stratum in the i^{th} pair of strata. Includes both the within PSU variance and the between PSU variance.

These formulae yield the following ratios:

2 PSU Method, With Replacement:

$$\frac{\text{2 PSU Method, With Replacement}}{\text{1 PSU Method Actual}} = 1 + \frac{\frac{1}{2}\sum_i (Y_{1i} - Y_{2i})^2}{\sum_i (\sigma_{1i}^2 + \sigma_{2i}^2)}$$

1 PSU Method, Biased:

$$\frac{\text{1 PSU Method, Biased}}{\text{1 PSU Method Actual}} = 1 + \frac{\sum_i (Y_{1i} - Y_{2i})^2}{\sum_i (\sigma_{1i}^2 + \sigma_{2i}^2)}$$

From our empirical data, the second term of the first ratio is .39 and the second term of the second ratio is .72. The ratio of these numbers, $\frac{.39}{.72} = .54$, is quite close to the theoretically expected relationship of .50. No similar relationship for 2 PSU's without replacement can apparently be derived.

B. Conclusions

As expected, the Type II error probabilities are always lowest for the ideal procedure, one PSU per stratum with a hypothetical variance estimator with expected value equal to the actual variance. Also, as expected, probabilities are always lower for two PSU's without replacement than for two PSU's with replacement. The following are observations on comparing probabilities for the one sample PSU design with a biased variance estimator to probabilities for a two sample PSU design.

1. Two PSU's with replacement is significantly better than one PSU when the distance between the true value and K is roughly in the 1.0σ to 3.0σ range and when the within PSU variance is a low percentage of the total variance.

2. Two PSU's without replacement is significantly better than one PSU when the distance between the true value and K is roughly in the 1.0σ to 4.5σ range, even when the within PSU variance is a fairly high percentage of the total variance.

3. One PSU is not significantly better than two PSU's with or without replacement for any distance.

4. Overall, particularly when comparing two PSU's with replacement to one PSU, one is struck with how little the Type II errors really differ.

In summary, the probabilities of incorrectly accepting the hypothesis that the true value is some constant K do not differ dramatically between two PSU's per stratum and one PSU per stratum (using the collapsed stratum estimator). However, two PSU's per stratum proves to be unequivocably superior to one PSU per stratum. Thus, if one wishes to use as a criterion of choice the minimization of the probability of accepting such a false hypothesis, two PSU's per stratum should be selected in preference to one PSU per stratum (when used in conjunction with a collapsed stratum variance estimator unadjusted for bias). (Also, of course, two PSU's without replacement is preferable to with replacement.)

Note, however, that these conclusions have been reached based on one survey situation, and even though some of the results are supported by the theoretical relationship between expected values, there may well be other actual situations under which the conclusions are not valid.

IV. ACKNOWLEDGMENTS

About 10 years ago Joseph Waksberg originated the idea to compare one PSU selection with two PSU selection in this manner. The results presented in this paper are a refinement to earlier work done by Gary M. Shapiro under the direction of Joseph Waksberg. The authors of this paper, however, accept full responsibility for any errors herein.

Improvements were made to this paper as the result of comments by Larry Ernst and Cary Isaki of the Census Bureau. The typing was done by Edith L. Oechsler.

REFERENCES

[1] 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.

[2] Hartley, H. O., Rao, J.N.K., and Kiefer, Grace, "Variance Estimation with One Unit per Stratum." *Journal of the American Statistical Association,* 64 (Sept. 1969), pp. 841-851.

[3] Shapiro, Gary M. and Bateman, David V. "A Better Alternative to the Collapsed Stratum Variance Estimate." Proceedings of the Social Statistics Section, *American Statistical Association,* 1978, pp. 451-456.

TABLE 1

PROBABILITIES OF FAILURE TO REJECT NULL HYPOTHESIS THAT TRUE POPULATION VALUE IS K FOR DIFFERENT RELATIONS BETWEEN THE TRUE VALUE AND K (PROBABILITIES IN %)

Distance in Actual σ's for 1 PSU Method Between True Value & K	1 PSU Method Actual	2 PSU METHOD, WITHOUT REPLACEMENT: Within PSU Variance as Percent of Total Variance[1]		2 PSU METHOD, WITH REPLACEMENT: Within PSU Variance as Percent of Total Variance[1]		1 PSU METHOD, BIASED: Within PSU Variance as Percent of Total Variance[1]	
		50	70	50	70	50	70
6.00	0.00	0.01	0.01	0.02	0.01	0.01	0.01
5.00	0.13	0.31	0.23	0.50	0.32	0.38	0.26
4.50	0.62	1.19	0.93	1.72	1.20	1.51	1.09
4.00	2.28	3.69	3.08	4.85	3.72	4.77	3.64
3.75	4.01	6.04	5.19	7.63	6.08	7.82	6.12
3.50	6.68	9.43	8.30	11.47	9.48	12.15	9.77
3.25	10.56	14.06	12.64	16.53	14.12	17.94	14.81
3.00	15.87	20.02	18.36	22.83	20.09	25.22	21.34
2.75	22.66	27.28	25.47	30.31	27.36	33.81	29.30
2.50	30.85	35.66	33.80	38.71	35.74	43.34	38.42
2.00	50.00	54.22	52.62	56.77	54.29	63.02	58.14
1.50	69.15	71.89	70.86	73.49	71.93	79.74	75.97
1.00	84.13	85.38	84.92	86.11	85.40	90.86	88.60
0.50	93.32	93.66	93.53	93.85	93.66	96.66	95.59
0.01	97.67	97.67	97.67	97.68	97.67	98.99	98.59

[1]Refers to actual total variance for a one PSU per stratum design.

TABLE 1. PROBABILITIES OF FAILURE TO REJECT NULL HYPOTHESIS THAT TRUE POPULATION VALUE IS K FOR DIFFERENT RELATIONS BETWEEN THE TRUE VALUE AND K (PROBABILITIES IN %)

Distance in Actual σ's for 1 PSU Method Between True Value & K	1 PSU Method Actual	2 PSU METHOD, WITHOUT REPLACEMENT: Within PSU Variance as Percent of Total Variance[1]								2 PSU METHOD, WITH REPLACEMENT: Within PSU Variance as Percent of Total Variance[1]								1 PSU METHOD, BIASED: Within PSU Variance as Percent of Total Variance[1]							
		40	50	55	60	65	70	75	80	40	50	55	60	65	70	75	80	40	50	55	60	65	70	75	80
6.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01
5.00	0.13	0.36	0.31	0.29	0.27	0.25	0.23	0.21	0.19	0.62	0.50	0.45	0.40	0.36	0.32	0.28	0.24	0.46	0.38	0.35	0.32	0.29	0.26	0.23	0.21
4.50	0.62	1.33	1.19	1.12	1.05	0.99	0.93	0.88	0.82	2.01	1.72	1.58	1.44	1.32	1.20	1.09	0.98	1.76	1.51	1.39	1.29	1.18	1.09	1.00	0.91
4.00	2.28	4.01	3.69	3.53	3.38	3.23	3.08	2.94	2.80	5.47	4.85	4.56	4.27	3.99	3.72	3.45	3.20	5.41	4.77	4.47	4.18	3.90	3.64	3.38	3.14
3.75	4.01	6.49	6.04	5.82	5.61	5.40	5.19	4.98	4.78	8.44	7.63	7.23	6.84	6.46	6.08	5.71	5.35	8.74	7.82	7.37	6.94	6.52	6.12	5.73	5.36
3.50	6.68	10.01	9.43	9.15	8.86	8.58	8.30	8.02	7.75	12.49	11.47	10.97	10.47	9.97	9.48	9.00	8.52	13.42	12.15	11.53	10.93	10.34	9.77	9.22	8.68
3.25	10.56	14.76	14.06	13.70	13.35	13.00	12.64	12.29	11.95	17.73	16.53	15.92	15.32	14.72	14.12	13.52	12.92	19.58	17.94	17.14	16.35	15.57	14.81	14.06	13.33
3.00	15.87	20.83	20.02	19.60	19.19	18.78	18.36	17.95	17.53	24.18	22.83	22.15	21.47	20.78	20.09	19.39	18.69	27.20	25.22	24.24	23.26	22.30	21.34	20.40	19.47
2.75	22.66	28.17	27.28	26.83	26.38	25.93	25.47	25.01	24.55	31.72	30.31	29.58	28.85	28.11	27.36	26.60	25.83	36.07	33.81	32.68	31.55	30.43	29.30	28.18	27.07
2.50	30.85	36.56	35.66	35.20	34.74	34.27	33.80	33.32	32.84	40.11	38.71	37.99	37.25	36.50	35.74	34.96	34.17	45.75	43.34	42.13	40.90	39.66	38.42	37.17	35.91
2.00	50.00	54.98	54.22	53.83	53.43	53.03	52.62	52.20	51.77	57.91	56.77	56.17	55.56	54.93	54.29	53.62	52.94	65.30	63.02	61.84	60.63	59.40	58.14	56.85	55.53
1.50	69.15	72.37	71.89	71.64	71.38	71.12	70.86	70.59	70.31	74.21	73.49	73.12	72.74	72.34	71.93	71.50	71.07	81.42	79.74	78.85	77.93	76.97	75.97	74.94	73.86
1.00	84.13	85.60	85.38	85.27	85.15	85.04	84.92	84.79	84.67	86.43	86.11	85.94	85.77	85.59	85.40	85.21	85.01	91.82	90.86	90.34	89.79	89.21	88.60	87.95	87.27
0.50	93.32	93.72	93.66	93.62	93.59	93.56	93.53	93.50	93.46	93.94	93.85	93.81	93.76	93.71	93.66	93.61	93.55	97.08	96.66	96.42	96.16	95.89	95.59	95.28	94.94
0.01	97.67	97.67	97.67	97.67	97.67	97.67	97.67	97.67	97.67	97.68	97.68	97.67	97.67	97.67	97.67	97.67	97.67	99.14	98.99	98.90	98.81	98.71	98.59	98.47	98.34

[1] Refers to actual total variance for a one PSU per stratum design.