VARIANCES FOR SUBDOMAINS OF MULTISTAGE CLUSTER SAMPLES:
THE DELTA AND BOOTSTRAP METHODS

Alan B. Forsythe, BMDP/UCLA
Richard Hough, UCLA
Audrey Burnam, UCLA

It is often not practical to form a sampling frame by listing all the people in the population. When taking a sample from a city, for example, it is much more economical to draw a sample of city blocks, list and sample the houses within the chosen blocks, and then to select a person at random from each household. One advantage is that we need only enumerate the houses within the selected blocks rather than all blocks. Further, only the selected houses have to have their members listed for possible inclusion in the sample. This form of sampling is called multistage cluster sampling and is described in many textbooks (Kish, 1965; Cochran, 1977). The blocks are the clusters and are also called the primary sampling units.

It is often interesting to form separate estimates for subgroups of subjects in the population. The subgroup of males from 30 to 50 years old is a subdomain of all people. If the characteristic being estimated is the presence or absence of a given event, then the rate for middle-aged males might be the statistic of interest. Of course, it is also very valuable to know the variability of the rate and not just the rate itself.

In this paper we consider the question of estimating the variance of a rate estimated for a subdomain in multistage cluster samples. The Taylor Series (delta) method gives an asymptotic estimate of variance and is inexpensive and easy to implement (Cochran, 1977). A relatively new technique for variance estimation has been defined by Bradley Efron (1979; 1981a,b) and is called the Bootstrap method. Although it is not as economical to implement, it has a strong intuitive appeal and conceptual simplicity.

In broad terms one can imagine what could be done to estimate the variance of any statistic. If we know the probability distribution function, all we need to do is sample from that distribution over and over again. Each sample yields a value for our statistic and the variance is the variance of those values. What if we do not know the distribution function? One possible solution is to assume a distribution and use that as the basis of our calculations. The normal distribution is often assumed. Efron's approach is to take the sample data values as an empirical estimate of the distribution and then sample from that distribution. This amounts to drawing a sample of the same size as the original sample -- with replacement. Each resample is used to estimate the statistic of interest, and the collection yields the estimated variance. There are enhancements to the method that include having each point represent a smear of points and also Bayesian mixing of prior information or belief about the distribution with the observed sample distribution.

In this paper we restrict our attention to the simple bootstrap and the delta method. We will not discuss other resampling schemes such as the jackknife, interpenetrating samples, balanced repeated replicates, or random subsamples (Tukey, 1958; McCarthy, 1966; Deming, 1956, 1960; Frankel, 1971; Kish and Frankel, 1974; Hartigan, 1969; Forsythe and Hartigan, 1970; Miller, 1974; and Quenouille, 1949).

The motivating example for this research is the estimation of depression rate in middle-aged males. We let X be the number of middle-aged males (about 20% of the population in this example) and Y be those middle-aged males who are depressed. The depression rate is $P = Y/X$. A depression rate of 5% and 20% will be used in this empirical comparison of variances of estimated subdomain rates. One can anticipate problems if the sample value of X could be close to zero. This might happen if X has a large coefficient of variation or a moderate coefficient of variation with a skewed distribution.

Methods

This paper reports a Monte Carlo (simulation) study of sampling from a hypothetical population of 2065 persons. There are 4.13 persons per household and there is an average of 5 houses per block (cluster). The distributions of persons per household and houses per cluster are given in Tables 1 and 2. There are 413 middle-aged males of the 2065 persons (20%) and we shall consider estimation of the depression rate when the population is generated at 5.085% and 9.677%. We shall examine two sample sizes:

1) Sample 30 clusters, 4 houses per cluster, 1 person per house. Of these 120 people, 24 are expected to be middle-aged males.

2) Sample 10 clusters, 3 houses per cluster, 1 person per house. This sample of 30 people would have an average of 6 males. The samples are self-weighting with probability proportional to size.

Each trial of this computer experiment consisted of fabricating the population,

Table 1. Distribution of Number of Persons per Household.			Table 2. Distribution of Number of Houses per Cluster.	

Number of People in House	Proportion of Population
1	.05
2	.10
3	.20
4	.35
5	.15
6	.05
7	.03
8	.03
9	.02
10	.02

Number of Houses per Cluster	Proportion of Houses
1	.00
2	.00
3	.00
4	.50
5	.30
6	.06
7	.05
8	.04
9	.03
10	.02

drawing a random sample from the population, calculating the ratio estimate of the depression rate (P), the variance of P as estimated by the delta method (D), and the bootstrap estimate of variance (B) based on 128 resamplings. These trials were repeated many times, enabling the calculation of the variance of P, the average estimate of variance from the delta method, D, and the average estimate of variance from the bootstrap method, B. This program was run on a Digital Equipment Corporation VAX/780 using an IMSL random number generator.

Results

The results are given in Table 3. There are four experimental conditions reflecting two sample sizes and two population depression rates. The variances, as calculated by the delta and bootstrap methods, are expressed as a percentage of the observed variance across trials. A value of 100% would represent an unbiased estimate of variance, while values less than 100% indicate that the estimate is biased to give variances that are lower than the true variance. Thus, the delta method estimated variance that is only 74% as large as it should be in the 10,000 trials of the first experimental condition. Both methods give a false sense of greater precision of estimation than warranted. The delta method is significantly more biased than the bootstrap

in each of four experiments. All four matched t tests yielded p values less than 0.0001. The magnitude of the bias does not vary strongly with the population rate, but the bias does decrease with greater sample size.

Conclusions and Suggestions for Future Work

In the conditions of this simulation study the delta method is found to be substantially and significantly more biased than the bootstrap in the estimation of the variance of a subdomain percentage. The variance as estimated by the delta method ranged from 66% to 74% of the observed variance. The bootstrap also gives estimated variances that are downward biased, but to a lesser extent (86% to 94%). See Table 4.

The conditions of these experiments are very stringent since the subdomain sample sizes are at most 24, on the average. Since the bias appears to decrease with increasing sample size, one might feel much more comfortable if the subdomain had a denominator in the hundreds. There is, perhaps, a tendency to subdivide a sample into many subgroups (such as age, gender, socio-economic status, and geographic location) without looking carefully at effects on variance. Unfortunately, the very variance we might want to consider might be giving us a

Table 3. Results of Simulation Study.

Number of Clusters in Sample	Number of Houses per Cluster in Sample	Depression Rate in Population %	Depression Rate in Sample R %	Var(R)	$\frac{\bar{D}}{Var(R)}$ %	$\frac{\bar{B}}{Var(R)}$ %	Number of Times
30	4	5.085	5.016	0.264	74	94	10,000
30	4	20.097	20.531	0.904	71	94	200
10	3	5.085	5.135	1.087	66	84	10,000
10	3	20.097	20.162	3.459	68	86	10,000

Table 4.
ESTIMATED VARIANCE AS A PERCENT OF TRUE VARIANCE

B BOOTSTRAP
D DELTA
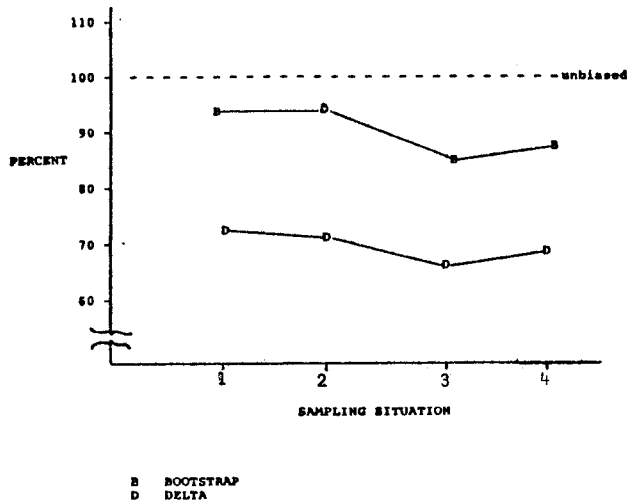
false sense of security, since it is biased downward. Such problems may also be present in biomedical research where, for example, a very large sample of births can be divided into categories for indirect standardization. The mechanical tools enable one to simultaneously consider birth weight in 50 gram intervals, gestational age, ethnic group, and gender. Soon, some cells might contain so few observations that they cast doubt upon the estimated variances.

Further work in this area might look at other resampling schemes such as the jackknife. Perhaps one could jackknife the variance estimate itself to reduce its bias. It would also be interesting to know the consequences of subdomain rates that are much smaller than the 5.1% considered in this paper.

References

Cochran, W.G. (1977). Sampling Techniques. John Wiley and Sons, New York.

Deming, W.E. (1956). On simplifications of sampling design through replication with equal probabilities and without stages. J. Am. Stat. Assoc. 51, 24-53.

Deming, W.E. (1960). Sample Design in Business Research. John Wiley and Sons, New York.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. Ann. Stat. 7, 1-26.

Efron, B. (1981a). Censored data and the bootstrap. J. Am. Statist. Assoc. 76, 312-319.

Efron, B. (1981b). Nonparametric estimates of standard error: the jackknife, the bootstrap, and other resampling methods. Biometrika.

Forsythe, A.B. and J.A. Hartigan (1969). Efficiency of confidence intervals generated by repeated subsample calculations. Biometrika 57(3), 629.

Frankel, M.R. (1971). Inference from survey samples. Institute for Social Research, Ann Arbor, Mich.

Hartigan, J.A. (1969). Using subsample values as typical values. J. Am. Stat. Assoc. 64, 1303-17.

Kish, L. (1965). Survey Sampling. John Wiley and Sons, New York.

Kish, L., and M.R. Frankel (1974). Inference from complex samples. J. Roy. Stat. Soc. B36, 1-37.

McCarthy, P.J. (1966). Replication: An approach to the analysis of data from complex surveys. National Center for Health Statistics, Washington, D.C., Series, 2, 14.

Miller, R.G., Jr. (1974). The jackknife - a review. Biometrika 61, 1-17.

Quenouille, M. (1949). Approximate tests of correlation in time series. J. Roy. Statist. Soc. Ser. B 11, 18-84.

Tukey, J.W. (1958). Bias and confidence in not-quite large samples. Ann. Math. Stat. 29, 614.