LARGE SAMPLE PROPERTIES OF THE LINEARIZATION, JACKKNIFE AND BALANCED REPEATED REPLICATION METHODS

D. Krewski and J.N.K. Rao, Health & Welfare Canada and Carleton University

The asymptotic normality of both linear and nonlinear statistics and the consistency of the variance estimators obtained using the linearization, jackknife and balanced repeated replication (BRR) methods in stratified samples are established. In the nonlinear case, several jackknife statistics are shown to have the same asymptotic distribution as the unjackknifed statistic. The results are obtained as $L \to \infty$ within the context of a sequence of finite populations $\{\Pi_L\}$ with L strata in Π_L and are valid for any stratified multistage design in which the primary sampling units (psu's) are selected *with* replacement and in which independent subsamples are taken for those psu's selected more than once. These results provide a theoretical basis for the commonly made assumption that $(\hat{\theta} - \theta)/\sqrt{v(\hat{\theta})}$ (where $\hat{\theta}$ is an estimator of some parameter θ and $v(\hat{\theta})$ is an estimator of the variance of $\hat{\theta}$) has approximately a standard normal or t distribution.

1. Introduction. Many large scale surveys involve large numbers of strata with relatively few psu's selected within each stratum (Gurney & Jewett, 1975). The asymptotic results for stratified samples presented in this article were developed with such situations in mind.

In section 2, some large sample theory for linear statistics is outlined for the case of stratified simple random sampling. In sections 3-5, these results are extended to the case of nonlinear statistics, where the linearization, jackknife and BRR methods may be employed.

Extensions of the results in sections 2-5 to multiple character stratified multi-stage samples are indicated in section 6. A discussion of the significance of the theoretical results presented here in relation to the empirical results reported previously by Kish & Frankel (1974) is provided in section 7. Detailed proofs of all results may be found in Krewski & Rao (1978a).

2. Stratified simple random sampling. The framework for the large sample theory is provided by a sequence of finite populations $\{\Pi_L\}$ with L strata in Π_L where $L \to \infty$.

Let Y_{Lhi} denote the value of the i^{th} unit in the h^{th} stratum in population Π_L $(i = 1,\ldots,N_{Lh};$ $h = 1,\ldots,L)$ and let $\bar{Y}_{Lh}$ denote the mean of all N_{Lh} units in the h^{th} stratum. The population mean may then be expressed as $\bar{Y}_L = \Sigma_h W_{Lh} \bar{Y}_{Lh}$ where $W_{Lh} = N_{Lh}/N_L$ and $N_L = \Sigma_h N_{Lh}$.

Similarly, let y_{Lhi} denote the value of the i^{th} unit in the sample in the h^{th} stratum in Π_L $(i = 1,\ldots,n_{Lh};$ $h = 1,\ldots,L)$ selected *with* replacement and let $\bar{y}_{Lh}$ denote the mean of the n_{Lh} (≥ 2) units in the sample in the h^{th} stratum. The usual estimator of $\bar{Y}_L$ is given by $\bar{y}_{L,st} = \Sigma_h W_{Lh} \bar{y}_{Lh}$.

Letting σ^2_{Lh} denote the variance within the h^{th} stratum in Π_L, the variance of $\bar{y}_{L,st}$ is given by

$$(2.1) \quad V(\bar{y}_{L,st}) = \Sigma_h W^2_{Lh} \sigma^2_{Lh}/n_{Lh}.$$

An unbiased estimator $v(\bar{y}_{L,st})$ of $V(\bar{y}_{L,st})$ is obtained by substituting

$$s^2_{Lh} = (n_{Lh} - 1)^{-1} \Sigma_{i=1}^{n_{Lh}} (y_{Lhi} - \bar{y}_{Lh})^2$$

for σ^2_{Lh} in (2.1).

Letting $n_L = \Sigma_h n_{Lh}$ denote the total sample size, the regularity conditions involved in the asymptotic theory are now introduced. (In order to avoid unnecessary repetition all limiting processes will be understood to be as $L \to \infty$.)

C1. $\Sigma_{h=1}^{L} W_{Lh} E|y_{Lhi} - \bar{Y}_{Lh}|^{2+\delta} = O(1)$ for some $\delta > 0$.

C2. $\frac{n_L}{N_L} \max_{1\leq h\leq L} \frac{N_{Lh}}{n_{Lh}} = O(1)$.

C2'. $\max_{1\leq h\leq L} W_{Lh} = O(L^{-1})$.

C3. $n_L \Sigma_{h=1}^{L} W^2_{Lh} \sigma^2_{Lh}/n_{Lh} \to \sigma^2 > 0$.

A sufficient condition for C1 to hold is that the sequence of finite populations $\{\Pi_L\}$ be uniformly bounded, i.e., $\max_{1\leq i\leq N_{Lh};\ 1\leq h\leq L} |Y_{Lhi}| = O(1)$. C2 is satisfied under *proportional allocation* $(n_{Lh}/n_L = W_{Lh})$ and reduces to C2' under *bounded allocation* $(\max_{1\leq h\leq L} n_{Lh} = O(1))$.

Conditions under which the asymptotic normality of $\bar{y}_{L,st}$ and the consistency of $v(\bar{y}_{L,st})$ may be established are given in Theorems 2.1 and 2.2 respectively.

THEOREM 2.1 Under C1-C3, $n_L^{\frac{1}{2}} (\bar{y}_{L,st} - \bar{Y}_L) \to_d N(0,\sigma^2)$.

THEOREM 2.2 Under C1 and C2, $n_L\{v(\bar{y}_{L,st}) - V(\bar{y}_{L,st})\} \to_p 0$.

3. The linearization method. Some parameters of interest may be expressed as a nonlinear function of the population mean, say $\Theta_L = g(\bar{Y}_L)$. A natural estimator of Θ_L in such cases is $\hat{\Theta}_L = g(\bar{y}_{L,st})$. Conditions under which the asymptotic normality of $\hat{\Theta}_L$ and the consistency of the corresponding "linearization" variance estimator may be established are given in Theorems 3.1 and 3.2 respectively. Two additional regularity conditions are required.

C4. $\bar{Y}_L \to \mu$.

C5. g has a continuous first derivative g' in a neighbourhood of μ.

THEOREM 3.1 Under C1 - C5, $n_L^{\frac{1}{2}} (\hat{\Theta}_L - \Theta_L) \to_d N(0, |g'(\mu)|^2\sigma^2)$.

THEOREM 3.2 Under C1 - C5, $n_L v_L \to_p |g'(\mu)|^2\sigma^2$ where $v_L = |g'(\bar{y}_{L,st})|^2 v(\bar{y}_{L,st})$ is the "linearization" variance estimator.

It follows immediately from Theorems 3.1 and 3.2 that

$$(3.1)\quad (\hat{\Theta}_L - \Theta_L)/v_L^{\frac{1}{2}} \to_d N(0, 1),$$

providing a theoretical basis for the construction of approximate confidence intervals and tests of hypothesis involving Θ_L when the number of strata is large.

4. The jackknife method. While $\bar{y}_{L,st}$ is an unbiased estimator of $\bar{Y}_L$, $\hat{\Theta}_L = g(\bar{y}_{L,st})$ is not in general an unbiased estimator of $\Theta_L = g(\bar{Y}_L)$ due to the nonlinearity of the transformation. The jackknife statistic may then be considered as an estimator of Θ_L on the basis of bias reduction (Jones, 1974). Perhaps more importantly, the jackknife method also provides a simple variance estimator which does not require the evaluation of partial derivatives as in the case of the linearization variance estimator.

Miller (1974) has given an excellent review of the jackknife method in the case of simple random sampling. The extension to stratified sampling, however, is not immediate and several different versions have been proposed.

Let $\hat{\Theta}_L^{hi} = g(\bar{y}_{L,st}^{hi})$ and $\hat{\Theta}_L^{h} = n_{Lh}^{-1} \Sigma_{i=1}^{n_{Lh}} \hat{\Theta}_L^{hi}$, where $\bar{y}_{L,st}^{hi}$ denotes the estimator of $\bar{Y}_L$ computed from the sample after deleting the i^{th} observation in the h^{th} stratum. Then Jones (1974) shows that the jackknife estimator

$$(4.1)\quad J_1(\hat{\Theta}_L) = (1 + n_L - L)\,\hat{\Theta}_L - \Sigma_{h=1}^{L}(n_{Lh} - 1)\,\hat{\Theta}_L^{h}$$

reduces the bias to second order moments in large samples. Jones' jackknife variance estimator is given by

$$(4.2)\quad v_{LJ}^{(1)} = \Sigma_{h=1}^{L} \frac{(n_{Lh}-1)}{n_{Lh}} \Sigma_{i=1}^{n_{Lh}}(\hat{\Theta}_L^{hi} - \hat{\Theta}_L^{h})^2.$$

McCarthy (1966) has also proposed a jackknife estimator when $n_{Lh} = 2 (h = 1,\ldots,L)$. With *pseudovalues* defined by

$J_{Lhi} = n_{Lh}\,\hat{\Theta}_L - (n_{Lh} - 1)\,\hat{\Theta}_L^{hi}$, either

$$(4.3)\quad J_2(\hat{\Theta}_L) = \frac{1}{n_L} \Sigma_{h=1}^{L} \Sigma_{i=1}^{n_{Lh}} J_{Lhi} \quad \text{or}$$

$$(4.4)\quad J_3(\hat{\Theta}_L) = \frac{1}{L} \Sigma_{h=1}^{L} \frac{1}{n_{Lh}} \Sigma_{i=1}^{n_{Lh}} J_{Lhi}$$

represent natural extensions of McCarthy's jackknife estimator to the case of arbitrary $\{n_{Lh}\}$. The corresponding jackknife variance estimators are

$$(4.5)\quad v_{LJ}^{(2)} = \Sigma_{h=1}^{L} \frac{1}{n_{Lh}} \Sigma_{i=1}^{n_{Lh}} \frac{(J_{Lhi} - J_2(\Theta_L))^2}{(n_{Lh} - 1)} \quad \text{and}$$

$$(4.6)\quad v_{LJ}^{(3)} = \Sigma_{h=1}^{L} \frac{1}{n_{Lh}} \Sigma_{i=1}^{n_{Lh}} \frac{(J_{Lhi} - J_3(\hat{\Theta}_L))^2}{(n_{Lh} - 1)}$$

Kish & Frankel (1974) have also proposed several jackknife variance estimators when $n_{Lh} = 2$ $(h = 1,\ldots,L)$. Their JRR-D estimator in the case of arbitrary $\{n_{Lh}\}$ is equivalent to $v_{LJ}^{(1)}$ while their JRR-S estimator is given by

$$(4.7)\quad v_{LJ}^{(4)} = \Sigma_{h=1}^{L} \frac{(n_{Lh} - 1)}{h_{Lh}} \Sigma_{i=1}^{n_{Lh}} (\hat{\Theta}_L^{hi} - \hat{\Theta}_L)^2.$$

REMARK 4.1. when $n_{Lh} = 2$, Kish & Frankel actually compute $\hat{\Theta}_L^{hi}$ by deleting the i^{th} unit in the h^{th} stratum *and including the other unit twice*. Since $\hat{\Theta}_L$ is assumed here to be a function of $\bar{y}_{L,st}$, the values of $\hat{\Theta}_L^{hi}$ based on this method of calculation will be identical to thosed based on the previously described method.

Conditions under which the asymptotic normality of $J_i(\hat{\Theta}_L)$ $(i = 1,2,3)$ and the consistency of $v_{LJ}^{(i)}$ $(i = 1,2,3,4)$ may be established are given in Theorems 4.1 and 4.2 respectively. As in the case of simple random sampling, a slightly stronger condition on g is required for the asymptotic normality of $J_i(\hat{\Theta}_L)$ than for the asymptotic normality of $\hat{\Theta}_L$.

C5'. g has a bounded second derivative in a neighbourhood of μ.

In addition to C1 - C5, condition C6 will be required in order to establish the consistency of $v_{LJ}^{(2)}$ as will C7 in the case of $v_{LJ}^{(3)}$.

C6. $n_L^{-2}(\Sigma_{h=1}^{L} n_{Lh}^2(n_{Lh}-1))\ \ \Sigma_{h=1}^{L}(n_{Lh} - 1) = O(1)$.

C7. $L^{-2}(\Sigma_{h=1}^{L} (n_{Lh} - 1))\ \ \Sigma_{h=1}^{L}(n_{Lh} - 1)^{-1} = O(1)$.

(Note that both these conditions are satisfied in the important case of bounded allocation.)

THEOREM 4.1. Under C1 - C4 and C5', $n_L^{\frac{1}{2}}(J_i(\hat{\Theta}_L) - \Theta_L) \to_d N(0, |g'(\mu)|^2\sigma^2)$ for $i = 1,2,3$.

THEOREM 4.2. Under C1 - C5, $n_L\, v_{LJ}^{(i)} \to_p |g'(\mu)|^2\sigma^2$ for $i = 1,2,3,4$ provided in addition that C6 and C7 hold in the case of $v_{LJ}^{(2)}$ and $v_{LJ}^{(3)}$ respectively.

As in (3.1), it follows from Theorems 4.1 and 4.2 that

$$(4.8)\quad (J_i(\hat{\Theta}_L) - \Theta_L)/(v_{LJ}^{(j)})^{\frac{1}{2}} \to_d N(0, 1)$$

for $i = 1,2,3$ and $j = 1,2,3,4$. By Theorems 3.1 and 4.2, any of the jackknife variance estimators may be used in conjunction with the unjackknifed estimator $\hat{\Theta}_L$ in that

$$(4.9)\quad (\hat{\Theta}_L - \Theta_L)/(v_{LJ}^{(j)})^{\frac{1}{2}} \to_d N(0,1).$$

5. <u>Balanced repeated replication (BRR)</u>. When $n_{Lh} = 2$ $(h = 1,\ldots, L)$ McCarthy (1966, 1969b) has developed a replication method of variance estimation in which the replicates are half-samples formed by deleting one unit in each stratum. The set of R_L half-samples used here will have *full orthogonal balance* (Frankel, 1971). Plackett & Burman (1946) provide a method of constructing sets of half-samples with full orthogonal balance where $L + 1 \le R_L \le L + 4$.

Letting $\bar{y}_{L,st}^{(r)}$ denote the estimator of $\bar{Y}_L$ based on the r^{th} half-sample, the orthogonality conditions imply that

$$(5.1)\quad R_L^{-1} \Sigma_{r=1}^{R_L} \bar{y}_{L,st}^{(r)} = \bar{y}_{L,st} \quad \text{and}$$

$$(5.2)\quad R_L^{-1} \Sigma_{r=1}^{R_L} (\bar{y}_{L,st}^{(r)} - \bar{y}_{L,st})^2 = v(\bar{y}_{L,st}).$$

To apply the method in the nonlinear case, let $\hat{\Theta}_L^{(r)} = g(\bar{y}_{L,st}^{(r)})$. Then McCarthy has proposed

$$(5.3)\quad H(\hat{\Theta}_L) = R_L^{-1} \Sigma_{r=1}^{R_L} \hat{\Theta}_L^{(r)}$$

as an estimator of Θ_L and

$$(5.4)\quad v_{LH}^{(1)} = R_L^{-1} \Sigma_{r=1}^{R_L} (\hat{\Theta}_L^{(r)} - \hat{\Theta}_L)^2 \quad \text{and}$$

$$(5.5)\quad v_{LH}^{(2)} = R_L^{-1} \Sigma_{r=1}^{R_L} (\hat{\Theta}_L^{(r)} - H(\hat{\Theta}_L))^2$$

as estimators of the variance of either $H(\hat{\Theta}_L)$ or $\hat{\Theta}_L$. Conditions under which the asymptotic normality of $H(\hat{\Theta}_L)$ and the consistency of the BRR variance estimators $v_{LH}^{(i)}$ $(i = 1,2)$ may be established are given in Theorems 5.1 and 5.2 respectively. (Note that since all $n_{Lh} = 2$, C2 reduces to C2'.)

THEOREM 5.1. Under C1 - C4 and C5',

$\sqrt{2L}\ (H(\hat{\Theta}_L) - \Theta_L) \to_d N(0, |g'(\mu)|^2\sigma^2)$.

THEOREM 5.2 Under C1 - C5,

$$(2L)v_{LH}^{(i)} \to_p |g'(\mu)|^2\sigma^2 \text{ for } i = 1,2.$$

As in (4.9), it follows from Theorems 3.1 and 5.2 that

$$(5.6)\quad (\hat{\Theta}_L - \Theta_L)/(v_{LH}^{(i)})^{\frac{1}{2}} \to_d N(0,1) \text{ for } i = 1,2.$$

Although $\hat{\Theta}_L$ may be replaced by $H(\hat{\Theta}_L)$ in (5.6) by virtue of Theorem 5.1, the latter estimator may be of limited interest in practice because of its greater bias.

6. Multiple character stratified multi-stage sampling. The results of section 2-5 may be extended to the case where the parameter of interest Θ_L is a nonlinear function of the population means for several characters. As illustrated in the following example, the asymptotic results presented here are thus applicable in the case of nonlinear statistics such as ratios and regression and correlation coefficients.

EXAMPLE 6.1 (Correlation coefficients). Consider $\underset{\sim}{Y}_{Lhi} = (Y_{Lhi1},\ldots,Y_{Lhi5})$ where $Y_{Lhi3} = Y_{Lhi1}^2$, $Y_{Lhi4} = Y_{Lhi2}^2$ and $Y_{Lhi5} = Y_{Lhi1}\, Y_{Lhi2}$. The correlation coefficient for variables one and two may then be expressed as

$$(6.1)\qquad \rho_L = \frac{\overline{Y}_{L5} - \overline{Y}_{L1}\overline{Y}_{L2}}{((\overline{Y}_{L3} - \overline{Y}_{L1}^2)\ (\overline{Y}_{L4} - \overline{Y}_{L2}^2))^{\frac{1}{2}}} = g(\overline{Y}_{L1},\ldots,\overline{Y}_{L5})$$

and estimated by $\hat{\rho}_L = g(\overline{y}_{L1,st},\ldots,\overline{y}_{L5,st})$.

The preceding results also hold in the case of any stratified multi-stage design in which the psu's may be selected with arbitrary probabilities (*with* replacement), provided that unbiased estimates of the totals for selected psu's are available and that independent subsamples are taken within those psu's selected more than once.

7. Discussion. Under the regularity conditions applicable in (3.1), (4.9) and (5.6),

$$(7.1)\qquad (\hat{\Theta}_L - \Theta_L)/(v_L^*)^{\frac{1}{2}} \to_d N(0,1),$$

where v_L^* denotes any variance estimator based on the linearization, jackknife or BRR methods. This result provides theoretical support for the empirical work undertaken by Kish & Frankel (1974). Using data from the Current Population Survey and sample designs involving the selection of two psu's per stratum, Kish & Frankel studied the degree to which the studentized statistic in (7.1) follows a t distribution with L degrees of freedom. For a variety of statistics (with the possible exception of the multiple correlation coefficient), this approximation was found to be adequate for designs involving as few as six or twelve strata. The BRR method was found to perform consistently better than the jackknife method which in turn performed better than the linearization method, although differences were small for relatively simple nonlinear statistics such as ratios.

As indicated in (4.8), $\hat{\Theta}_L$ may be replaced by $J_i(\hat{\Theta}_L)$ $(i = 1,2,3)$ in (7.1). However, since a number of empirical studies (see Kish, Namboodiri & Pillai (1962), McCarthy (1969a), Frankel (1971) and Bean (1975) for example) have indicated that most estimators are approximately unbiased in large scale surveys, the use of the jackknife as a means of bias reduction may be of limited importance in practice. (One exception to this consensus may be the case of partial and multiple correlation coefficients, where Frankel (1971) found the relative bias to be as high as twenty to twenty-five percent.)

While Kish & Frankel found the BRR variance estimators best in their empirical study according to the distributional properties discussed above, they also found the BRR variance estimators to be less stable than the jackknife variance estimator, although the differences encountered were small. (In order to resolve this conflict, Kish & Frankel recommend the use of the linearization variance estimator for relatively simple statistics such as ratios and BRR for more complex statistics such as correlation and regression coefficients.) Analytical results on the bias and stability of these alternative variance estimators in the case of ratio estimation in small samples are forthcoming (Krewski & Rao, 1978b).

REFERENCES

(1) BEAN, J.A. (1975). Distribution and properties of variance estimators for complex multistage probability samples. *Vital and Health Statistics*, Series 2, No. 65. U.S. Government Printing Office, Washington, D.C.

(2) FRANKEL, M.R. (1971). *Inference from Survey Samples*. Institute for Social Research, University of Michigan, Ann Arbor.

(3) GURNEY, M. and JEWETT, R.S. (1975). Constructing orthogonal replications for variance estimation. *J. Amer. Statist. Assoc.* 78 819-821.

(4) JONES, H.L. (1974). Jackknife estimation of functions of stratum means. *Biometrika* 61 343-348.

(5) KISH, L. and FRANKEL, M.R. (1974). Inference from complex samples (with discussion). *J. R. Statist. Soc. Ser. B* 36 1-37.

(6) KISH, L., NAMBOODIRI, N.K. and PILLAI, R.K. (1962). The ratio bias in surveys. *J. Amer. Statist. Assoc.* 57 863-876.

(7) KREWSKI, D. and RAO, J.N.K. (1978a). Inference from stratified samples I: large sample properties of the linearization, jackknife and balanced repeated replication methods. Technical Report No. 155, Carleton University Mathematical Series.

(8) KREWSKI, D. and RAO, J.N.K. (1978b). Inference from stratified samples II: small sample properties of the linearization, jackknife and balanced repeated replication methods. In preparation.

(9) McCARTHY, P.J. (1966). Replication: an approach to the analysis of data from complex surveys. *Vital and Health Statistics*, Series 2, No. 14. U.S. Government Printing Office, Washington, D.C.

(10) McCARTHY, P.J. (1969a). Pseudoreplication: further evaluation and application of the balanced half-sample technique. *Vital and Health Statistics*, Series 2, No. 31. U.S. Government Printing Office, Washington, D.C.

(11) McCARTHY, P.J. (1969b). Pseudoreplication: half-samples. *Rev. Int. Statist. Inst.* 37 239-264.

(12) MILLER, R.G. (1974). The jackknife - a review. *Biometrika* 61 1-15.

(13) PLACKETT, R.L. and BURMAN, J.P. (1946). The design of optimum multifactorial experiments. *Biometrika* 33 305-325.

DISCUSSION

Philip J. McCarthy, Cornell University

These two papers are quite different in character, but both make useful contributions to sample survey methodology. The Krewski-Rao paper is concerned with the asymptotic behavior of the linearization, jackknife and balanced repeated replication methods of estimating the variances of complex stratified sample estimators as the number of strata approaches infinity. Asymptotic normality of the estimators and consistency of the variance estimators is established. In the past, survey statisticians have had no choice but to assume that these results hold, but it is comforting to have a theoretical basis for such an assumption.

The Shapiro-Bateman paper deals with the problem of variance estimation when only one primary sampling unit is selected from each stratum. It suggests that the use of a without replacement variance estimate, derived from the Durbin model for selecting two PSU's without replacement from a stratum, provides a better approximation to the true variance than does the ordinary collapsed stratum variance estimate. Some empirical and theoretical evidence is presented to demonstrate that this approach usually provides a variance estimate with smaller bias and variance. This study should be supplemented with further empirical studies, and should possibly include consideration of other models for selecting two PSU's without replacement.