The Annals of Statistics
1991, Vol. 19, No. 4, 2275–2279

A NOTE ON THE LARGE SAMPLE PROPERTIES OF LINEARIZATION, JACKKNIFE AND BALANCED REPEATED REPLICATION METHODS FOR STRATIFIED SAMPLES

By Edward L. Korn and Barry I. Graubard

National Cancer Institute

Krewski and Rao consider inference for a (nonlinear) function of a vector of finite population means $\theta = g(\bar{Y})$. For a sequence of finite populations with increasing number of strata, they demonstrate that $\hat{\theta} = g(\bar{y})$ is asymptotically normal, where $\bar{y}$ is the usual unbiased stratified estimator of $\bar{Y}$. Additionally, they demonstrate that $(\hat{\theta} - \theta)/v^{1/2}(\hat{\theta})$ is asymptotically a standard normal distribution, where $v(\hat{\theta})$ is a variance estimator obtained using linearization, jackknife or balanced repeated replication (BRR) methods. In this note we extend their results to when the partial first derivatives $(g_1(\mu), g_2(\mu), \dots, g_p(\mu)) \equiv 0$, where μ is the limit of $\bar{Y}$ with increasing number of strata. We explore the asymptotic distribution of $(\hat{\theta} - \theta)/v^{1/2}(\hat{\theta})$ and show (1) that it is no longer normal and (2) that it depends upon which variance estimator is used. We describe an application of these results to hypothesis testing using complex survey data.

Large sample surveys frequently have complicated multistage designs. Fortunately for the purposes of inference, the first stage of these designs can usually be approximated by a probability proportional to size with replacement sample of a small number of primary sampling units (PSU's) from within strata. With these designs, Taylor series linearization, jackknife and balanced repeated replication (BRR) methods can be used to estimate the variance of complicated statistics; see Wolter (1985) for a complete review. Krewski and Rao (1981) put these estimation methods on a firm asymptotic footing by considering a sequence of finite populations with L strata, $L \to \infty$. In particular, they considered inference for a (nonlinear) function of a p-vector of population means $\theta = g(\bar{Y})$. They showed for $\bar{y}$ being the usual unbiased stratified estimator of $\bar{Y}$, that $\hat{\theta} = g(\bar{y})$ is asymptotically normal with variance that can be estimated by any of the above mentioned methods. An implicit assumption in their work is that the partial first derivatives $g_k(\mu)$ are not identically zero, where μ is the limit of $\bar{Y}$. (Subscripts of L are suppressed throughout this note, and all limits should be interpreted as $L \to \infty$.) In this note we utilize the second-order asymptotics of Rao and Wu (1985) to explore the asymptotic distribution of $\hat{\theta}$ and the variance estimators when the partial first derivatives are identically zero.

We first briefly describe the variance estimators that Krewski and Rao (1981) consider. For the finite population with L strata, n_h PSU's are sampled

Received December 1989; revised July 1990.

AMS 1980 *subject classifications*. Primary 62D05; secondary 62F12.

Key words and phrases. Stratified sampling, linearization method, jackknife method, balanced repeated replication, hypothesis testing.

from strata h, $h = 1, 2, \ldots, L$. The total number of sampled PSU's is $n = \sum n_h$. The linearization variance estimator is defined by

$$v_L(\hat{\theta}) = \sum \sum g_i(\bar{y}) g_j(\bar{y}) \hat{D}_{ij}(\bar{y}),$$

where $\hat{D}_{ij}(\bar{y})$ is the (ij)th element of the usual unbiased stratified estimator $\hat{D}(\bar{y})$ of the covariance matrix of $\bar{y}$, $D(\bar{y})$. Regularity condition C4 of Krewski and Rao (1981) is that $n\ D(\bar{y})$ converges to Γ, say, while Theorem 3.1 of Krewski and Rao (1981) which utilizes this condition and three additional regularity conditions (their C1–C3) insures

$$n^{1/2}(\bar{y} - \bar{Y}) \to_{\mathscr{D}} N(0, \Gamma). \tag{1}$$

Krewski and Rao (1981) define six related jackknife estimators of the variance. However, Rao and Wu (1985) show that they are asymptotically equivalent to order $O_p(n^{-3})$. Therefore, we consider only one of their jackknife estimators here, namely,

$$v_J(\hat{\theta}) = \sum_{h=1}^{L} n_h^{-1}(n_h - 1) \sum_{i=1}^{n_h} \{g(\bar{y}^{hi}) - g(\bar{y})\}^2,$$

where $\bar{y}^{hi}$ is the usual estimator of $\bar{Y}$ computed from the sample after omitting the data from the ith sampled PSU of the hth stratum: see Krewski and Rao (1981) for details. For $n_h = 2$ for all h, the BRR estimators of variance considered by Krewski and Rao (1981) are as follows:

$$v_B^{(1)}(\hat{\theta}) = \sum \{g(\bar{y}^{(j)}) - g(\bar{y})\}^2 \Big/ S,$$

$$v_B^{(2)}(\hat{\theta}) = \sum \{g(\bar{y}^{(j)}) - g(\bar{y}_c^{(j)})\}^2 \Big/ (4S),$$

$$v_B^{(3)}(\hat{\theta}) = \sum \left[\{g(\bar{y}^{(j)}) - g(\bar{y})\}^2 + \{g(\bar{y}_c^{(j)}) - g(\bar{y})\}^2\right] \Big/ (2S),$$

where the sums are over the S half-samples and $\bar{y}^{(j)}$ and $\bar{y}_c^{(j)}$ are the estimators of $\bar{Y}$ based on the jth half-sample and the complement of the jth half-sample; see Krewski and Rao (1981) for details. For convenience, we restate the relevant asymptotic theorem of Krewski and Rao (1981).

THEOREM [Krewski and Rao (1981)]. *Under the regularity conditions* C1–C6 *of Krewski and Rao* (1981), (i) $n^{1/2}(\hat{\theta} - \theta) \to_{\mathscr{D}} N(0, \sigma^2)$; (ii) $nv(\hat{\theta}) \to \sigma^2$ *in probability and* [*if* $g_k(\mu) \not\equiv 0$] (iii) $T = (\hat{\theta} - \theta)/v^{1/2}(\hat{\theta}) \to_{\mathscr{D}} N(0, 1)$, *where* $\sigma^2 = \sum\sum g_i(\mu) g_j(\mu) \gamma_{ij}$, $\Gamma = ((\gamma_{ij}))$ *and* $v(\hat{\theta})$ *is the linearization, jackknife or any one of the BRR variance estimators.*

To demonstrate the differences in the asymptotic behavior of $\hat{\theta}$ when $g_k(\mu) \equiv 0$, we first consider a simple univariate example. Suppose we are interested in testing the null hypothesis that $\mu = \mu_0$. A test statistic could be based on $g(\bar{y})$, where $g(t) = (t - \mu_0)^2$. We consider the distribution of $g(\bar{y})$

under a sequence of finite populations satisfying the null hypothesis in the limit. Although the pointwise limit $\bar{Y}$ is assumed to exist by regularity condition C5 of Krewski and Rao (1981), the asymptotic behavior of $\hat{\theta} - \theta \equiv (\bar{y} - \mu_0)^2 - (\bar{Y} - \mu_0)^2$ depends upon the speed at which $\bar{Y} \to \mu_0$. In particular, if $n^{1/2}(\bar{Y} - \mu_0) \to C$, then it is easy to show that $n(\hat{\theta} - \theta) \to_{\mathscr{D}} 2C\gamma^{1/2}Z + \gamma Z^2$, where Z has a standard normal distribution and γ is the asymptotic variance of $n^{1/2}(\bar{y} - \bar{Y})$. Using the linearization variance estimator described above, $v_L(\hat{\theta}) = 4(\bar{y} - \mu)^2\hat{D}(\bar{y})$. Under the regularity conditions C1–C4 of Krewski and Rao (1981), $n\ \hat{D}(\bar{y}) \to \gamma$ in probability, so that it is easy to show that $n^2v_L(\hat{\theta}) \to_{\mathscr{D}} 4\gamma(\gamma^{1/2}Z + C)^2$, where Z is the same standard normal random variable utilized for the asymptotic distribution of $\hat{\theta}$. Thus, for example, when $C = 0$, $(\hat{\theta} - \theta)/v_L^{1/2}(\hat{\theta}) \to_{\mathscr{D}} |Z|/2$.

To further explore the asymptotic distributions of $\hat{\theta}$ and the variance estimators, we will utilize the following additional regularity conditions:

(C5′) $n^{1/2}(\bar{Y} - \mu) \to 0$.

(C6′) The partial second derivatives $g_{ij}(\cdot)$ and third partial derivatives of g are continuous in a neighborhood of μ, $i, j = 1, 2, \dots, p$, the vector of partial first derivatives $(g_1(\mu), g_2(\mu), \dots, g_p(\mu))$ is identically zero and the matrix of partial second derivatives, $G \equiv ((g_{ij}(\mu)))$, is not identically zero. Recall that p is the dimension of $\bar{Y}$ and μ.

Theorem. *Under the regularity conditions* C1–C4 *of Krewski and Rao* (1981) *and* C5′–C6′,

(i) $n(\hat{\theta} - \theta) \to_{\mathscr{D}} \sum_{i=1}^{p}\lambda_i X_i$,
(ii) $n^2v(\hat{\theta}) \to_{\mathscr{D}} 4\sum_{i=1}^{p}\lambda_i^2 X_i$,
(iii) $T = (\hat{\theta} - \theta)/v^{1/2}(\hat{\theta}) \to_{\mathscr{D}} \sum_{i=1}^{p}\lambda_i X_i/(4\sum_{i=1}^{p}\lambda_i^2 X_i)^{1/2}$,

where X_i are independent chi-square random variables with 1 *degree of freedom, the λ_i are the eigenvalues of $\frac{1}{2}\Gamma G$ and $v(\hat{\theta})$ is either $v_L(\hat{\theta})$, $v_J(\hat{\theta})$ or $v_B^{(2)}(\hat{\theta})$.*

Proof. Under the regularity conditions C1–C3, Krewski and Rao (1981) show in their Theorem 3.2 that $n\{\hat{D}(\bar{y}) - D(\bar{y})\} \to 0$ in probability. This result with (1), C5′ and C6′ yields by standard Taylor series arguments the following:

$$n(\hat{\theta} - \theta) = \tfrac{1}{2}\{n^{1/2}(\bar{y} - \bar{Y})\}'G\{n^{1/2}(\bar{y} - \bar{Y})\} + o_p(1)$$

and

$$n^2v_L(\hat{\theta}) = \{n^{1/2}(\bar{y} - \bar{Y})\}'G\Gamma G\{n^{1/2}(\bar{y} - \bar{Y})\} + o_p(1)$$

Using (1) and standard results for quadratic forms [Johnson and Kotz (1970), pages 150–151], the conclusions of the theorem follow for $v_L(\hat{\theta})$, since the eigenvalues of $\Gamma G \Gamma G$ are four times the squares of the eigenvalues of $\frac{1}{2}\Gamma G$. Since $g_j(\bar{Y}) = o(n^{-1/2})$ by conditions C5′ and C6′, equation (31b) of Rao and

Wu (1985) implies $v_J(\hat{\theta}) = v_L(\hat{\theta}) + O_p(n^{-2.5})$, while equations (37) and (39) of Rao and Wu (1985) imply $v_B^{(2)}(\hat{\theta}) = v_L(\hat{\theta}) + O_p(n^{-2.5})$. The conclusions of the theorem therefore follow for these two variance estimators also. □

Using C5′ and the results of Rao and Wu (1985), it is easy to show that $n^2 v_B^{(1)}(\hat{\theta})$ and $n^2 v_B^{(3)}(\hat{\theta})$ are not asymptotically equivalent to the other variance estimators. For example, when $g(t) = (t - \mu_0)^2$, the asymptotic mean of these two statistics is $7\gamma^2$ compared to $4\gamma^2$ for $n^2 v_B^{(2)}(\hat{\theta})$, a large difference.

The result (i) is well known for chi-square tests involving contingency table data from complex surveys [Fay (1989)]. For example, $\hat{\theta}$ may be the usual Pearson chi-square test statistic for testing independence in a two-dimensional table. In this application, $\bar{y}$ is the sample cell proportions and μ is the limiting cell proportions associated with the sequence of finite populations. Under the null hypothesis, $g(\mu) = 0$, but we see from (iii) that the sum of the eigenvalues will need to be estimated for centering T for hypothesis testing [Rao and Scott (1987), Fay (1985)]. For this application, Fay (1985) derives (ii) using $v_J(\hat{\theta})$, but for the BRR, he uses a different estimator with the same asymptotic distribution. The theorem suggests that he could have used $v_B^{(2)}(\hat{\theta})$. Note that the expected value of $v(\hat{\theta})$ for large samples is $4\sum_{i=1}^{p}\lambda_i^2/n$, which is twice the large sample variance of $\hat{\theta} - \theta$, a point also noted by Simonoff (1986). Besides chi-square statistics for contingency table applications, the theorem will have relevance whenever a quadratic test statistic is used for hypothesis testing with complex survey data. For example, consider testing whether several regression coefficients are simultaneously zero. In this application, $\hat{\theta}$ could be the classical F statistic with $\bar{y}$ being the vector of sample means, squares and cross products of the variables. The theorem suggests ways to adjust the reference distribution to create a valid test statistic under the null hypothesis. Note that a Wald statistic that estimates the covariance matrix of the regression coefficients using a replication method will also produce a valid test statistic that incorporates the survey design. However, such a Wald statistic may have poor properties when the number of regression coefficients is approaching the number of PSU's available for the covariance estimation [Korn and Graubard (1990)]. As a final example, consider testing whether the sampling weights matter in a regression analysis. DuMouchel and Duncan (1983) and Fuller (1984) suggest computing the difference in the weighted and unweighted regression coefficients. For computing a test statistic, DuMouchel and Duncan (1983) use a model based estimate of the covariance matrix of this difference, while Fuller (1984), incorporating the survey design, uses a Taylor series linearization method to estimate a linear transformation of this difference. The DuMouchel and Duncan (1983) procedure does not take into account the possible clustering of the sample, while the Fuller (1984) procedure may not work well with limited number of PSU's for the covariance estimation. The theorem suggests ways one could use the DuMouchel and Duncan (1983) test statistic but modify its (null) reference distribution to account for clustering in the survey design.

REFERENCES

DuMouchel, W. H. and Duncan, G. J. (1983). Using sample survey weights in multiple regression analyses of stratified samples. *J. Amer. Statist. Assoc.* **78** 535–543.

Fay, R. E. (1985). A jackknifed chi-squared test for complex samples. *J. Amer. Statist. Assoc.* **80** 148–157.

Fay, R. E. (1989). Rao and Scott (type) tests. In *Encyclopedia of Statistical Sciences, Supplement Volume* 126–128. Wiley, New York.

Fuller, W. A. (1984). Least squares and related analyses for complex survey designs. *Survey Methodology* **10** 97–118.

Johnson, N. L. and Kotz, S. (1970). *Continuous Univariate Distributions* **2**. Wiley, New York.

Korn, E. L. and Graubard, B. I. (1990). Simultaneous testing of regression coefficients with complex survey data: Use of Bonferroni *t*-statistics. *Amer. Statist.* **44** 270–276.

Krewski, D. and Rao, J. N. K. (1981). Inference from stratified samples: Properties of the linearization, jackknife, and balanced repeated replication methods. *Ann. Statist.* **9** 1010–1019.

Rao, J. N. K. and Scott, A. J. (1987). On simple adjustments to chi-square tests with sample survey data. *Ann. Statist.* **15** 385–397.

Rao, J. N. K and Wu, C. F. J. (1985). Inference from stratified samples: Second-order analysis of three methods for nonlinear statistics. *J. Amer. Statist. Assoc.* **80** 620–630.

Simonoff, J. S. (1986). Jackknifing and bootstrapping goodness-of-fit statistics in sparse multinomials. *J. Amer. Statist. Assoc.* **81** 1005–1011.

Wolter, K. M. (1985). *Introduction to Variance Estimation*. Springer, New York.

Biometric Research Branch
National Cancer Institute
Bethesda, Maryland 20892

Biometry Branch
National Cancer Institute
Bethesda, Maryland 20892