THE USE OF CLASSICAL QUADRATIC TEST STATISTICS FOR TESTING HYPOTHESIS WITH COMPLEX SURVEY DATA: AN APPLICATION TO TESTING INFORMATIVENESS OF SAMPLING WEIGHTS IN MULTIPLE LINEAR REGRESSION ANALYSIS

Barry I. Graubard and Edward L. Korn, National Cancer Institute

1. INTRODUCTION

Complex surveys usually involve multistage cluster sampling where the selection of primary sampling units (PSU's) is within well-defined strata at the first stage of sampling. The standard error of an estimator can be computed from the variability of the estimator using data from different subsets of the sampled PSU's. For parameters involving nonlinear combinations of population means, methods including Taylor series linearization, the jackknife, balanced half sample replication, and the bootstrap have been developed to estimate standard errors (Efron 1982; Wolter 1985; Rao and Wu 1985) . The beauty of these replication methods is that they accommodate quite complicated sampling and estimation at the second and further stages of sampling.

In this paper, we are interested in testing hypotheses about an infinite population parameter $\theta = g(\mu)$ where μ is a mean vector. Following Krewski and Rao (1981), we do not specify the infinite population. Instead, it is assumed that there is sequence of increasing finite populations whose means converge to μ. Another aspect of the inferential approach of this paper is that we utilize replicated estimates of standard errors instead of model-based estimates for hypothesis testing. It can be difficult to model the complex variance structure from data that comes from a multistage stratified cluster sample, and even if modeled, the model-based variances can be sensitive to deviations from model assumptions (Skinner 1989). Replicated standard errors have the advantage of requiring fewer model assumptions. However, they can be more unstable than model-based estimates when there are few PSU's. This can be a serious problem with inference for a multidimensional parameter θ when the number of sampled PSU's is small (Skinner 1989). For example, consider testing a p-dimensional vector of regression coefficients equaling zero in a sampling design with a total of k sampled PSU's from L strata. Although we can estimate the covariance ($\hat{V}$) of the estimated regression coefficients $\overline{\theta}$ using a replication method, the degrees of freedom associated with $\hat{V}$ is only $k - L - p + 1$. Thus, the Wald statistic $X_W = \overline{\theta}'\hat{V}^{-1}\overline{\theta}$ can have very low power (Korn and Graubard 1990).

Alternative test procedures the Wald test, along with their asymptotic properties, are presented in the next section of this paper. These procedures are based upon the work of Rao and Scott (1981) and Fay (1985). An application to testing whether sampling weights are informative (i.e., reduce bias) in a multiple linear regression analysis is given in the third section. In the last section we discuss related work.

2. TEST STATISTICS AND THEIR ASYMPTOTIC PROPERTIES

2.1 Framework for Hypothesis Testing and Construction of Test Statistics

Let θ be a p-dimensional vector of parameters that is equal to zero under the null hypothesis, i.e., H_o: $\theta = 0$. We assume that θ can be expressed as a (nonlinear) function of a q-dimensional mean vector μ, i.e., $\theta = g(\mu)$. Additionally, we assume the existence of a quadratic test statistic that would be utilized if we had a simple random sample (srs) of size n from the population, namely, $nT(\overline{y}_{srs}) = n\,\overline{\theta}'_{srs}\,M(\overline{y}_{srs})\,\overline{\theta}_{srs}$, where $\overline{y}_{srs}$ is the sample mean, $\overline{\theta}_{srs} = g(\overline{y}_{srs})$ and $M(\overline{y}_{srs})$ is a pxp matrix. For example, nT could be a Wald statistic for testing a vector of regression coefficients equaling zero. In this case, μ would equal the vector of expected values of the dependent variable, the independent variables, and their cross products. For a complicated sample design, the weighted mean $\overline{y}$ incorporating the sampling weights will be a (design) unbiased estimator of the finite population mean $\overline{Y}$. This suggests calculating $nT(\overline{y}) = n\,\overline{\theta}'\,M(\overline{y})\,\overline{\theta}$, where $\overline{\theta} = g(\overline{y})$. Under the null hypothesis we would expect $nT(\overline{y})$ to be small but with a complicated distribution. Under this framework, the asymptotic null distribution of $nT(\overline{y})$ is given by:

Proposition 1. Under the null hypothesis, as $k \to \infty$, and under conditions C1-C8 given in Graubard (1991)

(i) $k^{1/2}(\bar{y} - \mu) \xrightarrow{D} N(0, \Gamma)$

(ii) $M(\bar{y}) \xrightarrow{P} M(\mu)$

(iii) $nT(\bar{y}) \xrightarrow{D} \sum_{i=1}^{p} \lambda_i \chi^2_{(i)}$

where Γ is a covariance matrix, the $\chi^2_{(i)}$ are i.i.d. chi-squared random variables with one degree of freedom, and λ_i are the eigenvalues of $c\Sigma M(\mu)$ where $n/k \rightarrow c$ and $\Sigma = G(\mu)\Gamma G'(\mu)$ with $G(\mu)$ the pxq matrix $[\partial g(x)/\partial x]\big|_{x=\mu}$.

Proof of Proposition 1. Parts (i) and (ii) follow from results in Krewski and Rao (1981) and a further assumption (C5) about the rate of convergence of $\bar{Y}$ to μ. (iii) follows from an extension of well-known results concerning quadratic forms (Johnson and Kotz 1970, pp. 150-151); see Graubard (1991) for details.

The λ_i' s are referred to as generalized design effects (Rao and Scott 1981).

2.2 Rao and Scott's Test Procedures

The approach Rao and Scott (1981) take to utilizing $nT(\bar{y})$ for testing is based on the following result:

Proposition 2 (Rao-Scott). Let λ_i be the eigenvalues of $c\Sigma M(\mu)$ as in Proposition 1, and let $\hat{\lambda}_i$ be the eigenvalues of $(n/k)\hat{\Sigma}M(\bar{y})$, where $\hat{\Sigma}$ is a consistent estimate of Σ. Under the null hypothesis, as $k \rightarrow \infty$, and under conditions C1-C8 given in Graubard (1991)

(i) $nT(\bar{y})/\hat{\bar{\lambda}} \xrightarrow{D} \sum_{i=1}^{p} (\lambda_i/\bar{\lambda}) \chi^2_{(i)}$

(ii) $nT(\bar{y})/[\hat{\bar{\lambda}}(1+\hat{a}^2)] \xrightarrow{D}$

$$\sum_{i=1}^{p} \{ \lambda_i/[\bar{\lambda}(1+a^2)] \} \chi^2_{(i)}$$

where $\hat{\bar{\lambda}} = \frac{1}{p}\sum_{i=1}^{p} \hat{\lambda}_i$ and $\hat{a}^2 = \frac{1}{p}\sum_{i=1}^{p} (\hat{\lambda}_i - \hat{\bar{\lambda}})^2/(\hat{\bar{\lambda}}^2)$,

and $\bar{\lambda}$ and a^2 are analogously defined using the λ_i 's.

Proof of Proposition 2. Parts (i) and (ii) follow from (iii) of Prop. 1 and the consistency of $\hat{\Sigma}$.

When the $\lambda_i \equiv \lambda$ and in particular when the $\lambda_i \equiv 1$, the right hand sides of (i) and (ii) in Proposition 2 reduce to chi-squared distributions with p degrees of freedom.

Rao and Scott (1981) give two procedures, denoted here by RS1 and RS2, which utilize $nT(\bar{y})$ for testing the null hypothesis. The rejection regions for level α tests are

RS1 : $nT(\bar{y})/\hat{\bar{\lambda}} > \chi^2_{p,\ 1-\alpha}$

and

RS2 : $nT(\bar{y})/[\hat{\bar{\lambda}}(1+\hat{a}^2)] > \chi^2_{p/(1+\hat{a}^2),\ 1-\alpha}$

where $\chi^2_{w,1-\alpha}$ is the $1-\alpha^{th}$ upper tail of a (central) chi-squared distribution with w degrees of freedom. RS1 uses an average design effect correction to approximate the asymptotic distribution of $nT(\bar{y})$ whereas RS2 uses a Satterthwaite correction to approximate it.

2.3 Fay's Jackknife Procedure and Modifications

Following the approach of Fay (1985), consider the test statistic

$$X = \frac{nT(\bar{y}) - \hat{C}}{\hat{V}_1^{1/2}}$$

where $\hat{C}$ is an estimate of the expected value of $nT(\bar{y})$ under the null hypothesis, and $\hat{V}_1$ is an estimate of the variance of $nT(\bar{y})$. After a square root transformation, we have

$$X_{J1} = \frac{[\, nT(\bar{y})\,]^{1/2} - \hat{C}^{1/2}}{\left\{ \hat{V}_1/[4\, nT(\bar{y})] \right\}^{1/2}}$$

using the delta method. (Note that $\hat{V}_1$ here is half the corresponding quantity defined in Fay (1985).) The estimators $\hat{C}$ and $\hat{V}_1$ depend upon the replication method (i.e., balanced half-sample repeated replication (BRR) or jackknife) employed. We describe an approach which uses BRR for sample designs consisting of two PSU's per stratum. (The use of the jackknife method of replication for computing $\hat{C}$ and $\hat{V}_1$ is described elsewhere (Graubard 1991)). Let $\bar{z}^{(i,1)}$ and $\bar{z}^{(i,2)}$ be the half-sample and complement half-sample replicate estimates of $\bar{Y}$, respectively (i.e., $\bar{z}^{(i,1)} + \bar{z}^{(i,2)} = 2\bar{y}$). We can write estimators for the mean and variance of $nT(\bar{y})$ as follows:

$$\hat{C} = n/(2d^2r)\sum_{i=1}^{r}\sum_{j=1}^{2}\left[\, T(\bar{y} + q^{(i,j)}) - T(\bar{y}) \,\right]$$

and

$$\hat{V}_1 = n^2/(4d^2r)\sum_{i=1}^{r}\sum_{j=1}^{2}\left[\, T(\bar{y} + q^{(i,j)}) - T(\bar{y}) \,\right]^2 ,$$

where $q^{(i,j)} = d(\bar{z}^{(i,j)} - \bar{y})$, r is the number of replicates and d is a small constant (following Fay (1985) it is chosen to be .05). When the estimator $\hat{C}$ is negative it is set to zero (Fay 1985) for computing the test statistic X_{J1} (also for X_{J2} defined latter). The statistic X_{J1} corresponds to the jackknife chi-squared test statistic of Fay (1985) for testing independence with categorical data using the Pearson chi-squared statistic and related test statistics.

In the next proposition, the asymptotic properties of $\hat{C}$ and $\hat{V}_1$ are given as $d \to 0$ (see Graubard, 1991, for further details).

Proposition 3 (Fay). Under the null hypothesis as $k \to \infty$ and under conditions C1-C9 given in Graubard (1991)

(i) $nT(\bar{y}) \xrightarrow{D} \sum_{i=1}^{p} \lambda_i \chi^2_{(i)}$

(ii) $\hat{C} \xrightarrow{P} \sum_{i=1}^{p} \lambda_i$

(iii) $\hat{V}_1 \xrightarrow{D} 2\sum_{i=1}^{p} \lambda_i^2 \chi^2_{(i)}$

(iv) $X_{J1} \xrightarrow{D}$

$$\frac{\left(\sum_{i=1}^{p} \lambda_i \chi^2_{(i)}\right)^{1/2} - \left(\sum_{i=1}^{p} \lambda_i\right)^{1/2}}{\left[\left(\sum_{i=1}^{p} \lambda_i^2 \chi^2_{(i)}\right) \Big/ \left(2\sum_{i=1}^{p} \lambda_i \chi^2_{(i)}\right)\right]^{1/2}}$$

where the convergence in (i) and (iii) is joint convergence. The λ_i, $i = 1, \ldots, p$ are the eigenvalues of $c\Sigma M(\mu)$ where Σ is as given in Proposition 1.

Proof of Proposition 3. Part (i) is Proposition 1(iii). (ii) and (iii) follow from asymptotic arguments given in Graubard (1991). (iv) follows from (i)-(iii).

Note that $\hat{V}_1$ is not a consistent estimator of the asymptotic variance of $nT(\bar{y})$ but is asymptotically unbiased. As a result of Proposition 3, when all the $\lambda_i \equiv \lambda$,

$$\left[\, (X_{J1}/\sqrt{2}) + \sqrt{p} \,\right]^2 \xrightarrow{D} \chi^2_p .$$

Following Fay (1985), we use this result, even though the λ_i are not all equal, to obtain the test procedure FJ1 which rejects the null hypothesis when

$$\text{FJ1: } X_{J1} > \sqrt{2}\left[\, (\chi^2_{p,\,1-\alpha})^{1/2} - p^{1/2} \,\right].$$

Note that for X_{J1} to be useful, (1) $\hat{C}$ should remain close to its expected value under the null hypothesis even under alternative hypotheses, and (2) $\hat{V}_1/\, nT(\bar{y})$ should not be very variable. These considerations lead to other possible choices for $\hat{C}$ and $\hat{V}_1$ which will now be discussed.

Another estimator for the variance of T which is commonly used to estimate the variance of nonlinear statistics (Wolter 1985, Ch. 3.4) is

$$\hat{V}_2 = (1/8r)\sum_{i=1}^{r}\left[\, nT(\bar{z}^{(i,1)}) - nT(\bar{z}^{(i,2)}) \,\right]^2 .$$

Substituting $\hat{V}_2$ for $\hat{V}_1$ in X_{J1}, we obtain the test statistic

$$X_{J2} = \frac{[\, nT(\bar{y}) \,]^{1/2} - \hat{C}^{1/2}}{\left\{\hat{V}_2/\, [4\, nT(\bar{y})]\right\}^{1/2}} .$$

Under the null hypothesis and under the same set of conditions as stated in Proposition 3, $\hat{V}_2$ has the same limiting distribution with the same chi-squared random variables as on the right hand side of Proposition 3(iii) (Korn and Graubard in press). Therefore, X_{J2} has the same limiting distribution as X_{J1}. We will refer to the test procedure FJ2 which rejects the null hypothesis when

$$\text{FJ2: } X_{J2} > \sqrt{2}\left[\, (\chi^2_{p,\,1-\alpha})^{1/2} - p^{1/2} \,\right].$$

It should be noted that there exist consistent estimators of the variance of $nT(\bar{y})$, e.g., the usual BRR variance that utilizes only the replicates and not the complement replicates. Such consistent estimators are not as correlated

with $nT(\bar{y})$ as are $\hat{V}_1$ and $\hat{V}_2$, making them less useful.

We can substitute $p\hat{\bar{\lambda}}$, a consistent estimate of $\sum \lambda_i$, for $\hat{C}$ in X_{J2} to obtain the test statistic

$$X_{J3} = \frac{[\, nT(\bar{y})\,]^{1/2} - (\, p\hat{\bar{\lambda}}\,)^{1/2}}{\left\{\hat{V}_2/\,[4\, nT(\bar{y})]\right\}^{1/2}} .$$

Thus, X_{J3} has the same limiting distribution as X_{J1} and X_{J2}. We will refer to the test procedure FJ3 which rejects when

$$\text{FJ3:}\quad X_{J3} > \sqrt{2}\left[\,(\,\chi^2_{p,\,1-\alpha}\,)^{1/2} - p^{1/2}\,\right].$$

2.4 The Wald Procedure

One of the earliest methods used for testing H_o with complex survey data is the Wald procedure (WP) (Koch, Freeman, and Freeman 1975). It utilizes the Wald test statistic

$$X_W = k\,\bar{\theta}'\,\hat{\Sigma}^{-1}\,\bar{\theta},$$

where $k^{-1}\hat{\Sigma}$ is a consistent estimator of the covariance matrix of $\bar{\theta}$, e.g., one obtained using a replicated variance estimator. WP rejects the null hypothesis when

$$\text{WP}: \frac{(k-L-p+1)}{(k-L)p} X_W > F_{p,\,(k-L-p+1),\,1-\alpha},$$

where $F_{u,\,v,\,1-\alpha}$ is the $1-\alpha$ upper tail of a F-distribution with u and v degrees of freedom; see Thomas and Rao (1987) and Korn and Graubard (1990). It should be noted that the RS procedures reduce to WP for one dimension and the FJ procedures do not.

3. AN APPLICATION TO TESTING THE INFORMATIVENESS OF SAMPLING WEIGHTS FOR MULTIPLE REGRESSION

In this application, we utilize a multiple regression analysis of systolic blood pressure on 25 dietary and blood chemistry variables in a sample of 2377 white males from the second National Health and Nutrition Examination Survey (NHANES II) (McDowell, Engel, Massey, and Mauer 1981) controlling for four background variables (age, age^2, body mass index, and total dietary calories). NHANES II is a national household survey which used a stratified multistage probability sample. We consider the problem of testing whether in a regression analysis the weighted least squares estimate ($\hat{\beta}_w$) and the ordinary least squares estimate ($\hat{\beta}$) are estimating the same population quantity. Loosely speaking, this is a test of whether the weights "matter" in terms of bias reduction. A Wald statistic can be constructed, $k(\hat{\beta}_w - \hat{\beta})'\hat{\Sigma}^{-1}(\hat{\beta}_w - \hat{\beta})$, where now $k^{-1}\hat{\Sigma}$ is a BRR estimate of the covariance of $\hat{\beta}_w - \hat{\beta}$. Fuller (1984) suggests a similar Wald statistic based on a linear transformation of $\hat{\beta}_w - \hat{\beta}$, and using a replicated Taylor series linearization estimate of the covariance of this linear transformation. These tests may lack power when there are limited degrees of freedom for the covariance estimation. We consider all 30 regression coefficients (including the intercept) so that $\hat{\beta}$ and $\hat{\beta}_w$ are both vectors of length 30. The Wald statistic is $X_W = 813.0$ with 30 and 2 degrees of freedom, yielding a p-value $=0.43$.

An alternative approach to testing the effect of the sample weights is suggested by DuMouchel and Duncan (1983) which has the form $n(\hat{\beta}_w - \hat{\beta})'\hat{V}^{-1}(\hat{\beta}_w - \hat{\beta})$, where $\hat{V}$ is a model-based estimate of the covariance of $\hat{\beta}_w - \hat{\beta}$ under simple random sampling. This testing procedure did not allow for cluster sampling. We show here how the DuMouchel and Duncan (1983) test statistic can be replicated so that its use can be extended to multistage sampling. In our framework, the population parameter $\theta = g(\mu)$ is the difference between the vector of population regression coefficients and the population regression coefficients estimated by the unweighted regression (OLS). We are interested in testing the null hypothesis H_0: $\theta = 0$. First, we define the vector of finite population means

$$\bar{Y} = \frac{1}{N}\left(\sum w_j^{-1}u_j^2,\ \sum u_j x_{ji_1},\ \sum w_j^{-1}u_j x_{ji_1},\ \sum w_j u_j x_{ji_1},\ \sum x_{ji_1}x_{ji_2},\ \sum w_j^{-1}x_{ji_1}x_{ji_2},\ \text{and}\ \sum w_j x_{ji_1}x_{ji_2};\ i_1, i_2 = 1, \ldots 30\right)',$$

where u_j , $(x_{j1}, \ldots, x_{j30})$ and w_j are the values of the dependent, independent and sample weight for the jth individual in the target finite population with $x_{j1} \equiv 1$. Next, we define $g(\bar{Y}) = (X'X)^{-1}X'U - (X'W^{-1}X)^{-1}X'W^{-1}U$

where X and U are as in the first example, and W is the diagonal matrix whose jth element on the diagonal (w_j) is the sample weight for the jth individual in the finite population. The quadratic test statistic to be replicated is $nT(\bar{y}) = ng(\bar{y})'M(\bar{y})g(\bar{y})$ where $\bar{y}$ is the weighted estimate of $\bar{Y}$ and $M(\bar{y})$ is obtained by substituting $\bar{y}$ for $\bar{Y}$ in

$$M(\bar{Y})^{-1} = [\,(X'X)^{-1}(X'WX)(X'X)^{-1} - (X'W^{-1}X)^{-1}\,]\,\sigma^2,$$

with

$$\sigma^2 = U'W^{-1}U - U'W^{-1}X(X'W^{-1}X)^{-1}X'W^{-1}U - g(\bar{Y})'\,[\,(X'X)^{-1}(X'WX)(X'X)^{-1} - (X'W^{-1}X)^{-1}]g(\bar{Y}).$$

With this notation $(n-2p)T(\bar{y})$ is precisely p times the DuMouchel-Duncan test statistic. (For our example, $(n-2p)T(\bar{y})=25.28$. Ignoring the clustered sample design and comparing the DuMouchel-Duncan test statistic to a $F_{30,\ 2317}$, the p-value is 0.71.) Using the BRR estimate of the covariance of $\hat{\beta}_w - \hat{\beta}$, we compute the 30 eigenvalues of $[(n-2p)/k]\hat{\Sigma}M(\bar{y})$. These eigenvalues range from .0021 to 7.85 with a mean of $\bar{\lambda} = 1.05$, and a coefficient of variation $\hat{a} = 1.46$. Using the Rao-Scott procedure (RS2), we compare 7.65 to a chi-squared distribution with 9.6 degrees of freedom. The p-value is 0.63. Using the Fay procedure (FJ1), we replicate with d=.05 and find $\hat{C} = 0.820$ and $\hat{V}_1 = 0.116$. The statistic $X_{J1} = .096$, and $[\ (X_{J1}/\sqrt{2}) + \sqrt{30}\ ]^2 = 30.75$, which can be compared to a chi-squared distribution with 30 degrees of freedom. The p-value is 0.43. (For test procedures FJ2 and FJ3, $\hat{V}_2 = 139.69$, X_{J2}=0.059 and X_{J3}= − 0.497, and the p-values are 0.44 and 0.66, respectively.) These p-values suggest that there is not strong evidence that the OLS analysis is different from the design-based analysis (i.e., the population parameter θ is not different from zero). This inference is useful because there can be large losses in efficiency if a weighted analysis of survey data is performed when the weights are unimportant (Korn and Graubard, 1991).

4. DISCUSSION

Six procedures for simultaneously testing the null hypothesis that a parameter vector θ is zero with complex survey data have been discussed in this paper: the Wald procedure (WP) and five procedures which utilize $nT(\bar{y})$ and replication methods to approximate the distribution of $nT(\bar{y})$. Another procedure based on the Bonferroni adjusted t-statistic consists of taking the maximum of the absolute value of p univariate replicated test statistics and comparing it to the α/p cut-off from a t-distribution with degrees of freedom equal to the number of sampled PSU's (k) minus the number of strata (Korn and Graubard 1990). Each replicated t-test is formed by dividing a different component of $\bar{\theta}$, i.e., $\bar{\theta}_i$, by its design-based standard error estimate. The Bonferrroni procedure can have better power than the WP when only a few of the components of θ are different from zero and the dimension of θ is large relative to k. Unlike the procedures discussed in this paper, one potential drawback of the Bonferroni procedure is that it is not invariant to linear transformations of the data. For example, suppose we are testing 25 parameters for which the first five correspond to dummy variables of a single categorical variable with six categories. The results of the Bonferroni procedure depend on how the dummy variables are created. A reasonable modified Bonferroni procedure would be to use any of the procedures discussed here to test the first five variables at the $\alpha/21$ level, and then use replicated t-tests at the $\alpha/21$ level for the last 20 variables.

Other possible procedures intended specifically for multiple linear regression applications have been described by Wu, Holt and Holmes (1988). They propose modified F-tests in which the modifications take account of intra-cluster correlation of the observations from two-stage cluster samples. Their test procedures are asymptotically equivalent to the RS1 procedure when there is a common intra-cluster correlation coefficient among the residuals. Their procedures do not apply to more complex sampling schemes in which there are differential sampling weights or more than two stages. Earlier work on the effect of two-stage cluster sampling on OLS regression analysis has been done by Scott and Holt (1982).

In Graubard (1991), simulations are conducted for simultaneously testing vectors of 2, 14 and 25 means and multiple regression coefficients. The results are presented for sample designs of 16, 32 and 64 strata and under null

and alternative hypotheses. It is based on these simulation results that we make the following recommendations concerning the choice of testing procedure: 1) choose the WP when the number of degrees of freedom for covariance estimation is large relative to the number of dimensions of θ and otherwise, 2) use FJ2 for testing means and RS2 for testing regression coefficients. Additionally, in this latter case consider the Bonferroni procedure when it is thought that only a few of the components of θ are different from zero. Recommendation (2) should be viewed as tentative; further research will hopefully define a more generally applicable test statistic.

REFERENCES

Cochran, W. G. (1977), Sampling Techniques (3rd ed.), New York: John Wiley.

DuMouchel, W. H., and Duncan, G. J. (1983), "Using Sample Survey Weights in Multiple Regression Analysis of Stratified Samples," Journal of the American Statistical Association, 78, 535-543.

Efron, B. (1982), The Jackknife, the Bootstrap and Other Resampling Plans, Philadelphia: Society for Industrial and Applied Mathematics.

Fay, R. E. (1985), "A Jackknifed Chi-squared Test for Complex Samples," Journal of the American Statistical Association, 80, 148-157.

Fuller, W. A. (1984), "Least Squares and Related Analyses for Complex Survey Designs," Survey Methodology, 10, 97-118.

Graubard, B. I. (1991), "Statistical Methods for the Analysis of Complex Survey Data with Biomedical Application," unpublished, University of Maryland, Dept. of Mathematics, College Park, Maryland.

Johnson, N. L. and Kotz S. (1970), Distributions in Statistics: Continuous Univariate Distributions — 2, New York: John Wiley.

Koch, G. G., Freeman, D. H., and Freeman, J. L. (1975), "Strategies in the Multivariate Analysis of Data From Complex Surveys," International Statistical Review, 43, 59-78.

Korn, E. L., and Graubard, B. I. (1990), "Simultaneous Testing of Regression Coefficients with Complex Survey Data: Use of Bonferroni t Statistics," The American Statistician, 44, 270-276.

Korn, E. L., and Graubard, B. I. (in press), "A Note on the Large Sample Properties of Linearization, Jackknife and Balanced Repeated Replication Methods for Stratified Samples," Annals of Statistics.

Korn, E. L., and Graubard, B. I. (1991), "Epidemiologic Studies Utilizing Surveys: Accounting for the Sampling Design," American Journal of Public Health, 81, 1163-1173.

Krewski, D., and J. N. K. Rao (1981), "Inference from Stratified Samples: Properties of the Linearization, Jackknife and Repeated Replication Methods," Annals of Statistics, 9, 1010-1019.

McDowell, A., Engel, A., Massey, J. T., and Mauer, K. (1981), Plan and Operation of the Second National Health and Nutrition Examination Survey, 1976-1980, Vital and Health Statistics, Ser. 1, No. 15, Washington, DC: U. S. Government Printing Office. [Public Health Service Publication 81-1317.]

Rao, J. N. K., and Scott, A. J. (1981), "The Analysis of Categorical Data from Complex Sample Surveys: Chi-Squared Tests for Goodness of Fit and Independence in Two-Way Tables, " Journal of the American Statistical Association, 76, 221-230.

Rao, J. N. K., and Wu, C. F. J. (1985), "Inference from Stratified Samples: Second-Order Analysis of Three Methods for Nonlinear Statistics, " Journal of the American Statistical Association, 80, 620-630.

Scott, A. J. and Holt, D. (1982), "The Effect of Two Stage Sampling on Ordinary Least Squares Methods," Journal of the American Statistical Association, 77, 848-854.

Skinner, C. J. (1989), "Introduction to Part A," in Analysis of Complex Surveys, eds. C. J. Skinner, D. Holt, and T. M. F. Smith, West Sussex: John Wiley, pp. 23-58.

Thomas, D. R., and Rao, J. N. K. (1987), "Small-sample Comparisons of Level and Power for Simple Goodness-of-Fit Statistics Under Cluster Sampling," Journal of the American Statistical Association, 82, 630-636.

Wolter, K. M. (1985), Introduction to Variance Estimation, New York: Springer-Verlag.

Wu, C. F. J., Holt, D., and Holmes, D. J. (1988), "The Effect of Two-Stage Sampling on the F Statistic," Journal of the American Statistical Association, 83, 150-159.