Statistica Sinica **8**(1998), 1059-1070

QUASI-SCORE TESTS WITH SURVEY DATA

J. N. K. Rao, A. J. Scott and C. J. Skinner

Carleton University, University of Auckland and University of Southampton

Abstract: Although most survey texts are concerned primarily with problems of estimating finite population parameters, survey data are often used to develop and fit stochastic models describing the underlying structure of the population. In this paper we develop analogues of the score test for use with survey data where the use of multi-stage sampling and variable selection probabilities cause special problems.

Key words and phrases: Estimating equations, jackknife variance estimator, score test, superpopulation parameters, survey data.

1. Introduction

In most conventional statistics courses, a clear distinction is made between sample survey methods on the one hand and the rest of applied statistics on the other. Traditional survey methods are concerned with estimating population means, totals and proportions, along with related quantities like ratios, while the rest of applied statistics concentrates on model-building for explanation, prediction and so on. In reality, many surveys (especially in the health and social sciences) are aimed at exploring relationships and building predictive models, just as in the rest of statistics. Surveys are conducted to find out what effect education has on unemployment or income, what factors affect crib deaths in infants or strokes in older people and so on. For example, the data shown in Table 1 comes from a stratified case control sample drawn from records of people under the age of 35 in northern Malawi as given in Clayton and Hills (1993). Here “cases” are new cases of leprosy and Scar is a binary variable taking the value one if a person has a BCG vaccination scar and zero otherwise. The aim in this study is to gain some insight into whether or not a BCG vaccination affects the chance of contracting leprosy rather than estimating population totals and proportions.

What makes the analysis of survey data different? One obvious problem is that, by their very nature, analytical surveys are observational studies and we are always faced with the difficulty of making causal inferences in situation where we have no control over the assignment of experimental treatments. There are two other major features that distinguish the analysis of survey data. The first is the correlation induced by the hierarchical structure of multi-stage sampling.

Because of the cost, most large scale surveys are carried out in two or more stages. The lack of independence within primary sampling units (census blocks, doctors' practices, schools, households) means that standard errors, confidence intervals and P-values produced by standard computer packages are invalid. This is by no means unique to surveys, however. Many experimenters have to cope with correlation between repeated measurements on the same subject, siblings from the same litter, and so on. As an aside it is interesting to note that a number of techniques developed to handle survey data are starting to find uses in other areas that have to deal with correlated data (see Rao and Scott (1991) for a simple example). Perhaps more important is the use of variable selection probabilities. If some parts of the population are sampled more intensively than others, then the resulting sample can look very different from the population from which it is drawn and about which we want to make inferences. The data in Table 1, from a survey in which cases of leprosy are heavily oversampled, gives an illustration of this phenomenon, although it is difficult to see the impact with a binary response variable. A more graphic illustration is shown in Figure 1 of Scott and Wild (1986) which plots blood alcohol readings against readings from a blood test for a sample of respondents. The sample was a stratified one, with strata defined by values of the response variable and with very different sampling fractions between strata. When the data were weighted to allow for these varying selection probabilities, the fitted straight line gave a perfectly adequate fit.

Table 1. A stratified case-control sample

	Scar = 0		Scar = 1		Total		Popn
Age[a]	Case	Control	Case	Control	Case	Control	Control
7.5	11	10	14	15	25	25	17327
12.5	28	19	22	31	50	50	13172
17.5	16	6	28	38	44	44	10325
22.5	20	13	19	26	39	39	8026
27.5	36	35	11	12	47	47	4981
32.5	47	49	6	4	53	53	6479
Total					258	258	61310

[a] Age is age-group midpoint

In this paper, we attempt to show how standard survey methods can be adapted to produce valid methods for fitting models to survey data, and to test hypotheses about model parameters.

2. Quasi-Score Tests

Suppose that, attached to all units of a finite population of size N, we have measurements $(\mathbf{x}_i, y_i)$ made on a vector of explanatory variables, $\mathbf{x}$, and a response variable, Y. We assume that for a given value of $\mathbf{x}$, Y is generated by some random process with mean

$$E(Y_i) = \mu_i = \mu(\mathbf{x}_i, \boldsymbol{\beta}) \tag{2.1}$$

and we suppose that we have in mind some working model for the variance, say

$$\mathrm{Var}(Y_i) = V_{0i} = V_0(\mu_i) \tag{2.2}$$

for $i = 1, \ldots, N$. The model (2.1) for the mean is assumed to be valid, but the working variance (2.2) may only be a rough approximation at best.

We do not observe values for all the population units but only for those in a sample drawn from the finite population according to some well-defined sampling scheme. We are interested in estimating the parameters $\boldsymbol{\beta}$ and, more particularly, in testing the hypothesis $H_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$ using the sample data, where $\boldsymbol{\beta}$ is partitioned as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ with $\boldsymbol{\beta}_2$ a $q \times 1$ vector, and $\boldsymbol{\beta}_1$ a $r \times 1$ vector.

Suppose that, if we had values for the whole finite population, we could obtain a consistent estimator of $\boldsymbol{\beta}$ by solving the estimating equations

$$\mathbf{S}(\boldsymbol{\beta}) = \sum_1^N \mathbf{u}_i(\boldsymbol{\beta}) = \mathbf{0}, \tag{2.3}$$

where $\mathbf{u}_i(\boldsymbol{\beta})$ has kth component $u_{ik} = (\partial\mu_i/\partial\beta_k)(y_i - \mu_i)/V_{0i}$. Thus we are working in the general estimating equations framework considered by Godambe and Thomson (1986) and Godambe (1991), although there is no requirement that the estimating equations be optimal or that the units be sampled independently from the superpopulation. Note that the resulting estimator is the quasi-likelihood estimator if the finite population is regarded as a random sample from the superpopulation, but the estimator is consistent under much more general conditions. Essentially, all we need to assume is that the finite population can be regarded as a self-weighting sample from the superpopulation. Note that the equations (2.3) are similar to the generalized estimating equations of Liang and Zeger (1986).

In reality, we do not know the values for the whole finite population but only for those units in a sample drawn from the population. We suppose only that the sample design provides consistent, asymptotically normal estimators of population totals, and associated standard errors. Then, since $\mathbf{S}(\boldsymbol{\beta})$ is a vector of population totals for fixed $\boldsymbol{\beta}$, we can produce an estimator of $\mathbf{S}(\boldsymbol{\beta})$ as

$$\hat{\mathbf{S}}(\boldsymbol{\beta}) = \sum_{i \in s} w_{is}\mathbf{u}_i(\boldsymbol{\beta}), \tag{2.4}$$

where the survey weights, w_{is}, may depend on the sample s (e.g., post-stratified weights). Our sample estimator, $\hat{\boldsymbol{\beta}}$, is obtained by solving $\hat{\mathbf{S}}(\hat{\boldsymbol{\beta}}) = 0$. This approach was suggested by Fuller (1975) for linear regression with two-stage sampling, and by Binder (1983) for generalized linear models and any survey design.

Under suitable conditions (see Binder (1983) for details), $\hat{\boldsymbol{\beta}}$ is asymptotically normal with mean $\boldsymbol{\beta}$, and we can estimate Cov $(\hat{\boldsymbol{\beta}})$ consistently by

$$\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}) = [\mathbf{J}(\hat{\boldsymbol{\beta}})]^{-1}\hat{\mathbf{V}}_S(\hat{\boldsymbol{\beta}})[\mathbf{J}(\hat{\boldsymbol{\beta}})]^{-1}, \tag{2.5}$$

where

$$\mathbf{J}(\boldsymbol{\beta}) = -\frac{\partial \hat{\mathbf{S}}}{\partial \boldsymbol{\beta}'} = -\sum_{i\in s} w_{is}\frac{\partial \mathbf{u}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \tag{2.6}$$

and $\hat{\mathbf{V}}_S(\hat{\boldsymbol{\beta}})$ is the estimated covariance matrix of $\hat{\mathbf{S}}(\boldsymbol{\beta})$ under the specified survey design evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$. Note that $\hat{\mathbf{V}}_S(\hat{\boldsymbol{\beta}})$ is obtained from the standard survey variance estimator for a total since $\hat{\mathbf{S}}(\boldsymbol{\beta})$, given by (2.4), is the estimator of the total $\mathbf{S}(\boldsymbol{\beta})$ given by (2.3).

Now consider the problem of testing the null hypothesis that $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$. One approach is to base the test on the corresponding Wald statistic

$$X_W^2 = (\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{20})'\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_2)^{-1}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{20}) \tag{2.7}$$

which is asymptotically a χ_q^2 variable under H_0. This has the usual problems associated with the Wald test. For example, it is not invariant to nonlinear transformations of the parameters $\boldsymbol{\beta}$ and often has poor small sample behaviour. In addition, with survey data the effective degrees of freedom for estimating Cov $(\hat{\boldsymbol{\beta}})$ is often rather small, resulting in instability of $\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_2)^{-1}$ when the dimension of $\boldsymbol{\beta}_2$ is large (see Rao and Thomas (1987)). Ideally we would prefer to use a likelihood ratio test, which is invariant and usually has better small sample properties, but we have no likelihood from which to construct such a test here. However, the score test shares many of the desirable properties of the likelihood ratio test, and it is relatively straightforward to construct a simple analogue of the score test in our framework. Our development of the test and its properties parallels the development of Boos (1992) for the case of random sampling from an infinite population.

Let $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_1', \boldsymbol{\beta}_{20}')'$ be the solution of $\hat{\mathbf{S}}_1(\tilde{\boldsymbol{\beta}}) = 0$ where $\hat{\mathbf{S}} = (\hat{\mathbf{S}}_1', \hat{\mathbf{S}}_2')'$ is partitioned in the same way as $\boldsymbol{\beta}$. The analogue of the score test, which we shall call the quasi-score test, is based on the statistic

$$X_S^2 = \tilde{\mathbf{S}}_2'\tilde{\mathbf{V}}_{2S}^{-1}\tilde{\mathbf{S}}_2, \tag{2.8}$$

where $\tilde{\mathbf{S}}_2 = \hat{\mathbf{S}}_2(\tilde{\boldsymbol{\beta}})$ and $\tilde{\mathbf{V}}_{2S}$ is a consistent estimator of Cov $(\tilde{\mathbf{S}}_2)$. The asymptotic distribution of $\tilde{\mathbf{S}}_2$ under H_0 can be obtained as in Boos (1992), who treated the case of random sampling from an infinite population, by expanding $\hat{\mathbf{S}}_1(\tilde{\boldsymbol{\beta}})$ and $\hat{\mathbf{S}}_2(\tilde{\boldsymbol{\beta}})$ as a function of $\tilde{\boldsymbol{\beta}}$ about the true value, $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_1^{*\prime}, \boldsymbol{\beta}_{20}^{\prime})^{\prime}$. We give a brief sketch of the development here but omit technical details for simplicity.

Expanding $\hat{\mathbf{S}}_1(\tilde{\boldsymbol{\beta}})$ and $\hat{\mathbf{S}}_2(\tilde{\boldsymbol{\beta}})$ gives

$$0 = \hat{\mathbf{S}}_1(\tilde{\boldsymbol{\beta}}) \approx \hat{\mathbf{S}}_1(\boldsymbol{\beta}^*) - \mathbf{J}_{11}^*(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^*) \tag{2.9}$$

and

$$\hat{\mathbf{S}}_2(\tilde{\boldsymbol{\beta}}) \approx \hat{\mathbf{S}}_2(\boldsymbol{\beta}^*) - \mathbf{J}_{21}^*(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^*), \tag{2.10}$$

where

$$\mathbf{J}^* = \mathbf{J}(\boldsymbol{\beta}^*) = \begin{bmatrix} \mathbf{J}_{11}^* & \mathbf{J}_{12}^* \\ \mathbf{J}_{21}^* & \mathbf{J}_{22}^* \end{bmatrix}.$$

Then, replacing $\mathbf{J}^*$ by its expected value, $\mathbf{I}^*$ say, and substituting for $(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^*)$ from (2.9) into (2.10) yields

$$\tilde{\mathbf{S}}_2 = \hat{\mathbf{S}}_2(\tilde{\boldsymbol{\beta}}) \approx \hat{\mathbf{S}}_2(\boldsymbol{\beta}^*) - \mathbf{I}_{21}^*\mathbf{I}_{11}^{*-1}\hat{\mathbf{S}}_1(\boldsymbol{\beta}^*) = \sum_{i \in s} w_{is}\mathbf{z}_i, \tag{2.11}$$

where $\mathbf{z}_i = \mathbf{u}_{2i}(\boldsymbol{\beta}^*) - \mathbf{A}\mathbf{u}_{1i}(\boldsymbol{\beta}^*)$ with $\mathbf{A} = \mathbf{I}_{21}^*\mathbf{I}_{11}^{*-1}$ and $\mathbf{u}_i = (\mathbf{u}_{1i}^{\prime}, \mathbf{u}_{2i}^{\prime})^{\prime}$. It then follows from our assumptions about the survey estimator of a total that $\tilde{\mathbf{S}}_2$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix $\mathrm{Cov}\,(\tilde{\mathbf{S}}_2)$ under H_0. Thus, X_S^2 is asymptotically a χ_q^2 variable under H_0.

The quasi-score test based on X_S^2 shares most of the advantages of its infinite population counterpart. With an appropriate choice for the variance estimator $\tilde{\mathbf{V}}_{2S}$ (we discuss this point further in Section 3), the test is invariant to nonlinear transformation of the parameters $\boldsymbol{\beta}$ (Boos (1992)). Moreover, we need only ever fit the simple null model, which is a considerable advantage if the full model contains a large number of terms as will be the case, for example, with a factorial structure of explanatory variables containing a large number of interactions. The few studies that have been carried out so far indicate that the small sample behaviour tends to be better than that of the corresponding Wald test, but much more work needs to be done.

If the effective degrees of freedom is small, $\tilde{\mathbf{V}}_{2S}^{-1}$ can become unstable when the dimension of $\boldsymbol{\beta}_2$ (or $\tilde{\mathbf{S}}_2$) is large. Thus both the Wald test (2.7) and the quasi-score test (2.8) suffer from instability. We discuss some alternative tests in Section 4 in an attempt to overcome the problem.

3. Estimation of $\mathrm{Cov}(\tilde{\mathbf{S}}_2)$

Calculation of the quasi-score statistic X_S^2 requires an estimator of $\mathrm{Cov}(\tilde{\mathbf{S}}_2)$. A resampling method, such as the jackknife or balanced repeated replication

(BRR), is particularly attractive in the case of stratified multistage sampling because post-stratification and unit nonresponse adjustment are automatically taken into account. For example, a jackknife estimator of $\mathrm{Cov}(\tilde{\mathbf{S}}_2)$ under stratified multistage sampling with n_g sampled clusters from the gth stratum is given by

$$\hat{\mathbf{V}}_J(\tilde{\mathbf{S}}_2) = \sum_{g=1}^{L} \frac{n_g - 1}{n_g} \sum_{j=1}^{n_g} (\tilde{\mathbf{S}}_{2(gj)} - \tilde{\mathbf{S}}_2)(\tilde{\mathbf{S}}_{2(gj)} - \tilde{\mathbf{S}}_2)'. \tag{3.1}$$

Here $\tilde{\mathbf{S}}_{2(gj)}$ is obtained in the same manner as $\tilde{\mathbf{S}}_2$ when the data from the (gj)-th sample cluster is deleted, but using jackknife weights (see the Appendix) and recalculating $\tilde{\boldsymbol{\beta}}$, say $\tilde{\boldsymbol{\beta}}_{(gj)}$. Computation of $\tilde{\boldsymbol{\beta}}_{(gj)} = (\tilde{\boldsymbol{\beta}}'_{1(gj)}, \boldsymbol{\beta}'_{20})'$ can be simplified by performing only a single Newton-Raphson iteration for the solution of $\hat{\mathbf{S}}_{1(gj)}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{20}) = \mathbf{0}$, using $\tilde{\boldsymbol{\beta}}$ as the starting value, where $\hat{\mathbf{S}}_{1(gj)}$ $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{20})$ uses the jackknife weights instead of the original weights. We refer the reader to Shao and Tu (1995), Chapter 6 and Rao (1996) for details on the jackknife method and other resampling methods under stratified multistage sampling. A proof of the asymptotic consistency of the jackknife variance estimator (3.1) as $\sum n_h \to \infty$ is sketched in the Appendix.

The jackknife quasi-score test resulting from $\hat{\mathbf{V}}_J(\tilde{\mathbf{S}}_2)$ is invariant to a one-to-one reparametrization of $\boldsymbol{\beta}$ with non-singular Jacobian, unlike the Wald test X_W^2. Alternatively, we can use a Taylor linearization variance estimator $\hat{\mathbf{V}}_L(\tilde{\mathbf{S}}_2)$. This amounts to applying the survey variance estimator for a total to the representation in (2.11), replacing $\mathbf{z}_i$ by

$$\tilde{\mathbf{z}}_i = \mathbf{u}_{2i}(\tilde{\boldsymbol{\beta}}) - \tilde{\mathbf{A}}\mathbf{u}_{1i}(\tilde{\boldsymbol{\beta}}), \tag{3.2}$$

where $\tilde{\mathbf{A}}$ is an estimator of $\mathbf{A}^* = \mathbf{I}_{21}^* \mathbf{I}_{11}^{*-1}$. Note that the survey variance estimator used should account for post-stratification and unit nonresponse.

There are several possible choices for $\tilde{\mathbf{A}}$. It might seem natural to use $\mathbf{J}(\tilde{\boldsymbol{\beta}})$ in place of $\mathbf{I}^*$, where $\mathbf{J}(\boldsymbol{\beta})$ is defined by (2.6) and, in fact, this form of the quasi-score statistic (2.8) (with $q = 1$) is used by Binder and Patak (1994) to construct confidence intervals for β_2, although their derivation is quite different from that given here. However, this choice does not have the desired invariance property in general. We can get an invariant test by taking the expectation of $\mathbf{J}(\boldsymbol{\beta})$ under the mean specification defined by (2.1), giving

$$\mathbf{I}(\tilde{\boldsymbol{\beta}}) = \sum_{i \in s} w_{is} \mathbf{D}_i(\tilde{\boldsymbol{\beta}}) \mathbf{D}_i(\tilde{\boldsymbol{\beta}})' / V_{0i}(\tilde{\mu}_i), \tag{3.3}$$

where $\mathbf{D}_i(\boldsymbol{\beta}) = \partial \mu_i(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}$, $\tilde{\mu}_i = \mu(\mathbf{x}_i, \tilde{\boldsymbol{\beta}})$. We suspect that $\mathbf{I}(\tilde{\boldsymbol{\beta}})$ is also more stable than $\mathbf{J}(\tilde{\boldsymbol{\beta}})$ and it is the one we recommend as the choice for $\tilde{\mathbf{A}}$, although again, much more work is needed here. Note that $\mathbf{I}(\tilde{\boldsymbol{\beta}})$ and $\mathbf{J}(\tilde{\boldsymbol{\beta}})$ are identical for models with canonical link functions (e.g., logistic regression). We could replace $\tilde{\boldsymbol{\beta}}$ by $\hat{\boldsymbol{\beta}}$ in either of the above choices, but this would require fitting the full model and thus would negate one of the principal attractions of the quasi-score test.

4. Alternative Tests

We now consider some alternative tests in an attempt to overcome the degrees of freedom problem in the context of stratified multistage sampling. The degrees of freedom for estimating $\mathrm{Cov}(\hat{\boldsymbol{\beta}}_2)$ or $\mathrm{Cov}(\tilde{\mathbf{S}}_2)$ is usually taken as $f = n - L$, where $n = \sum n_h$ is the total number of sample clusters and L is the number of design strata. In the case of subgroups (or domains), the degrees of freedom can be much less if the subgroup population is not uniformly distributed across strata. For a subgroup, the degrees of freedom is taken as $f = n' - L'$, where L' is the number of strata that contain at least one sample member from the subgroup and n' is the total number of sample clusters that contain at least one sample member from the subgroup. We refer the reader to Rust and Rao (1996) for further discussion on degrees of freedom for variance estimation.

If the degrees of freedom, f, is not large, an F-version of the Wald test is often used. This test treats

$$F_W = [(f - q + 1)/(fq)]X_W^2 \tag{4.1}$$

as an F-variable with q and $f - q + 1$ degrees of freedom under H_0 (see e.g., Korn and Graubard (1990)). Empirical results suggest that F_W might perform better than X_W^2 in controlling the size of the test when f is not large (see e.g., Thomas and Rao (1987)). An F-version of the score test treats

$$F_S = [(f - q + 1)/(fq)X_S^2 \tag{4.2}$$

as an F-variable with q and $f - q + 1$ degree of freedom.

Rao and Scott (1984) proposed corrections to naive tests that ignore the survey data features. A naive test, using normalized survey weights $\tilde{w}_{is} = w_{is}/\bar{w}_s$, is given by $X_{NS}^2 = \tilde{\mathbf{S}}_{2N}'\tilde{\mathbf{V}}_{2NS}^{-1}\tilde{\mathbf{S}}_{2N}$, where $\bar{w}_s$ is the mean of the weights w_{is}, $i \in s$, $\tilde{\mathbf{V}}_{2NS} = \tilde{\mathbf{I}}_{22} - \tilde{\mathbf{I}}_{21}\tilde{\mathbf{I}}_{11}^{-1}\tilde{\mathbf{I}}_{12}$, $\tilde{\mathbf{S}}_{2N}$ is obtained by changing w_{is} to $\tilde{w}_{is}$ in $\tilde{\mathbf{S}}_2$, and $\tilde{\mathbf{I}}$ is given by $\mathbf{I}(\tilde{\boldsymbol{\beta}})$ with w_{is} changed to $\tilde{w}_{is}$. In terms of the original weights w_{is}, we can express X_{NS}^2 as $X_{NS}^2 = \bar{w}_s^{-1}\tilde{\mathbf{S}}_2'\mathbf{V}_{2NS}^{-1}\tilde{\mathbf{S}}_2$, where $\mathbf{V}_{2NS} = \mathbf{I}_{22} - \mathbf{I}_{21}\mathbf{I}_{11}^{-1}\mathbf{I}_{12}$ and $\mathbf{I} = \mathbf{I}(\tilde{\boldsymbol{\beta}})$ is given by (3.3). If $w_{is} = w$, i.e., equal weights, then X_{NS}^2 reduces to the classical score statistic.

Following Rao and Scott (1984), a first-order correction to X_{NS}^2 is given by

$$X_S^2(1) = X_{NS}^2/\delta., \tag{4.3}$$

where $\delta. = \sum \delta_i/q$ and $\delta_1, \ldots, \delta_q$ are the nonzero eigenvalues of $\bar{w}_s^{-1}[\mathbf{V}_{2NS}^{-1}\tilde{\mathbf{V}}_{2S}]$. Under H_0, $X_S^2(1)$ is treated as a χ_q^2 variable. Note that $X_S^2(1)$ avoids the inversion of $\tilde{\mathbf{V}}_{2S}$, unlike X_S^2; $\tilde{\mathbf{V}}_{2NS}^{-1}$ remains stable even if f is small. It follows from (4.3) that normalization of weights is not necessary in using $X_S^2(1)$ because it reduces to

$$X_S^2(1) = [\tilde{\mathbf{S}}_2' \mathbf{V}_{2NS}^{-1} \tilde{\mathbf{S}}_2]/\tilde{\delta}., \tag{4.4}$$

where $\tilde{\delta}. = \sum \tilde{\delta}_i/q$ and $\tilde{\delta}_1, \ldots, \tilde{\delta}_q$ are the nonzero eigenvalues of $\mathbf{V}_{2NS}^{-1}\tilde{\mathbf{V}}_{2S}$. A more accurate, second-order correction to X_{NS}^2 is given by

$$X_S^2(2) = X_S^2(1)/(1 + \tilde{a}^2), \tag{4.5}$$

where $\tilde{a}^2 = \sum(\tilde{\delta}_i - \tilde{\delta}.)^2/[(q-1)\tilde{\delta}.^2]$ is the square of the coefficient of variation of the $\tilde{\delta}_i$'s. Under H_0, $X_S^2(2)$ is treated as a χ^2 variable with degrees of freedom $q/(1+\tilde{a}^2)$. The first-order correction $X_S^2(1)$ is adequate if the coefficient of variation of the $\tilde{\delta}_i$'s is small. Rotnitzky and Jewell (1990) proposed corrections to score tests, similar to (4.4) and (4.5), in the context of longitudinal data on a simple random sample of individuals.

Korn and Graubard (1990) proposed Bonferroni t statistics in the context of a Wald test. The Bonferroni procedure for nominal level α rejects H_0 when

$$\max_{1 \le i \le q} (|\hat{\beta}_{r+i} - \beta_{0,r+i}|/s(\hat{\beta}_i)) \ge t_f(\alpha/2q), \tag{4.6}$$

where $s(\hat{\beta}_i)$ is the standard error of $\hat{\beta}_i$, $t_f(\alpha/2q)$ is the upper $\alpha/(2q)$-point of a t variable with f degrees of freedom, $\hat{\boldsymbol{\beta}}_2 = (\hat{\beta}_{r+1}, \ldots, \hat{\beta}_{r+q})'$ and $\boldsymbol{\beta}_{20} = (\beta_{0,r+1}, \ldots, \beta_{0,r+q})'$. One could use either the Taylor linearization variance estimator (2.5) or a resampling variance estimator, such as the jackknife, to calculate $s(\hat{\beta}_i)$. The asymptotic size of the Bonferroni test (4.6) is less than or equal to α. Korn and Graubard (1990) demonstrated the benefits of the Bonferroni procedure (4.6) in terms of both size and power of the test. Thomas, Singh and Roberts (1996) found similar benefits in the context of tests of independence in a two-way table under cluster sampling. However, (4.6) is not invariant to linear transformations.

A Bonferroni procedure in the context of the quasi-score test (2.8) can also be constructed. This procedure for nominal level α rejects H_0 when

$$\max_{1 \le i \le q} [|\tilde{S}_{r+i}|/s(\tilde{S}_{r+i})] \ge t_f(\alpha/2q), \tag{4.7}$$

where $s(\tilde{S}_{r+i})$ is the standard error of $\tilde{S}_{r+i}$ and $\tilde{\mathbf{S}}_2 = (\tilde{S}_{r+1}, \ldots, \tilde{S}_{r+q})'$. One could use either the Taylor linearization variance estimator $\hat{\mathbf{V}}_L(\tilde{\mathbf{S}}_2)$ or a resampling variance estimator such as $\hat{\mathbf{V}}_J(\tilde{S}_2)$ to calculate s.e.$(\tilde{S}_i)$. An advantage of (4.7) over (4.6) is that only the simpler null model need be fitted to calculate $\tilde{\boldsymbol{\beta}}_1$. The Bonferroni t-test (4.7) may be preferred to the quasi-score test (2.8) when f is not large relative to q. However it is not invariant even to linear transformations, unlike the quasi-score test which is invariant under re-parametrization.

5. Simple Logistic Regression

We illustrate the preceding theory in the simple special case in which the response variable, Y, is binary and we are fitting a simple linear logistic regression model. Thus $E(Y_i) = \mu_i = \exp(\beta_1 + \beta_2 x_i)/[1 + \exp(\beta_1 + \beta_2 x_i)]$. As a working model for the variance, we take the standard binomial form with $V_{0i} = \mu_i(1-\mu_i)$ so that $\mathbf{u}_i = \mathbf{x}_i(y_i - \mu(\mathbf{x}_i, \boldsymbol{\beta}))$ where $\mathbf{x}_i = (1, x_i)'$ and $\boldsymbol{\beta} = (\beta_1, \beta_2)'$. Suppose that we want to test the null hypothesis that $\beta_2 = 0$. Then $\hat{S}_1(\boldsymbol{\beta}) = \sum_{i\in s} w_{is}(y_i - \mu(\mathbf{x}_i, \boldsymbol{\beta}))$ and $\hat{S}_2(\boldsymbol{\beta}) = \sum_{i\in s} w_{is} x_i (y_i - \mu(\mathbf{x}_i, \boldsymbol{\beta}))$. Setting $\hat{S}_1(\tilde{\beta}) = 0$ gives $\tilde{\beta}_1 = \log[\tilde{p}/(1-\tilde{p})]$ with $\tilde{p} = \sum_s w_{is} y_i / \sum_s w_{is}$. Thus, the quasi-score test is particularly simple in this case since we can write down $\tilde{S}_2$ explicitly, viz., $\tilde{S}_2 = \sum_s w_{is} x_i (y_i - \tilde{p})$. Note that since this is a canonical model, $\mathbf{I}(\boldsymbol{\beta})$ and $\mathbf{J}(\boldsymbol{\beta})$ are identical:

$$\mathbf{I}(\boldsymbol{\beta}) = \mathbf{J}(\boldsymbol{\beta}) = \sum_{i\in s} w_{is}\mu_i(1-\mu_i)\begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix}.$$

Substituting $\tilde{p}$ for $\tilde{\mu}_i = \mu_i(\tilde{\boldsymbol{\beta}})$ gives $\tilde{A} = \tilde{I}_{21}\tilde{I}_{11}^{-1} = \sum_{i\in s} w_{is} x_i / \sum_{i\in s} w_{is} = \widehat{\overline{X}}$ say, and $\tilde{z}_i = u_{2i}(\tilde{\boldsymbol{\beta}}) - \tilde{A} u_{1i}(\tilde{\boldsymbol{\beta}}) = (y_i - \tilde{p})(x_i - \widehat{\overline{X}})$. The linearization estimator, $\hat{V}_L(\tilde{S}_2)$, is then the standard variance estimator for the total of the "synthetic" variable $\tilde{z}_i$ under the specified design, denoted by $\hat{V}(\sum_s w_{is}\tilde{z}_i)$.

The quasi-score test statistic X_S^2 based on the linearization variance estimator reduces to $X_S^2 = [\sum_{i\in s} w_{is} x_i (y_i - \tilde{p})]^2 / \hat{V}(\sum_{i\in s} w_{is}\tilde{z}_i)$, which is asymptotically a χ_1^2 variable under the null hypothesis. If the jackknife method is used under stratified multistage sampling, then we replace $\hat{V}(\sum_s w_{is}\tilde{z}_i)$ by $\hat{V}_J(\tilde{S}_2) = \sum_g [(n_g - 1)/n_g] \sum_j (\tilde{S}_{2(gj)} - \tilde{S}_2)^2$.

6. Example

Consider the data shown in Table 1. Here the response variable, is binary and takes the value 1 if the person has leprosy (case) and 0 otherwise (control); the explanatory variables are Age and Scar, where Scar takes value 1 if a person has a BCG vaccination scar and 0 otherwise. The sample design is stratified random sampling with seven strata. All cases are sampled and the sampling fractions for the six age strata can be found from Table 1. For reasons outside the scope of this illustration, we consider a logistic regression model with $\log[\mu_i/(1-\mu_i)] = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$, where $x = (\text{Age}+7.5)^{-2}$ and $x_2 = \text{Scar}$. We take the Bernoulli variance, $V_{0i} = \mu_i(1-\mu_i)$, as our working model variance. Interest centers on whether the BCG vaccination has any impact on the incidence of leprosy, i.e., in testing the hypothesis that $\beta_2 = 0$. Thus we set $\boldsymbol{\beta}_1 = (\beta_0, \beta_1)'$ and $\boldsymbol{\beta}_2 = \beta_2$. We obtain $\tilde{\boldsymbol{\beta}}_1$ by solving

$$\hat{\mathbf{S}}_1(\tilde{\boldsymbol{\beta}}) = \sum_{i \in s} w_{is} \begin{bmatrix} y_i - \tilde{\mu}_i \\ x_{1i}(y_i - \tilde{\mu}_i) \end{bmatrix} = \mathbf{0},$$

where $\log[\tilde{\mu}_i/(1-\tilde{\mu}_i)] = \tilde{\beta}_0 + \tilde{\beta}_1 x_i$. This gives $\tilde{\beta}_0 = -4.6$ and $\tilde{\beta}_1 = -427.0$. Then $\tilde{S}_2 = \sum_{i \in s} w_{is} x_{2i}(y_i - \tilde{\mu}_i) = -32.61$ and the linearization variance estimate is $\hat{V}_L(\tilde{S}_2) = 99.10$, leading to a value of 10.73 for the quasi-score statistic X_S^2. Thus there does seem to be a strong association (P-value $= Pr(\chi_1^2 < 10.73) < 0.001$) although, as with any observational study, great care needs to be taken with any interpretation of this result.

7. Concluding Remarks

We have developed quasi-score tests on regression parameters, assuming only a general mean specification. These tests take proper account of the survey design features such as clustering and unequal selection probabilities, and hence can be used with complex survey data. An advantage of the proposed quasi-score tests is that we need only fit the simple null model, which is a considerable advantage if the full model contains a large number of parameters. Also, the tests are invariant under re-parametrization.

Alternative tests to handle small degrees of freedom associated with the variance estimators are also considered. We hope to conduct a detailed simulation study on the finite sample performance of the proposed tests in terms of size and power.

Acknowledgements

We are thankful to a referee for helpful comments. This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

Appendix: Asymptotic Consistency of $\hat{\mathbf{V}}_J(\tilde{\mathbf{S}}_2)$

Shao (1992) established the asymptotic consistency of the jackknife variance estimator of $\hat{\boldsymbol{\beta}}$ in the case of independent responses, y_i. A rigorous proof of the consistency of $\hat{\mathbf{V}}_J(\tilde{\mathbf{S}}_2)$ for stratified multistage sampling would involve an extension of the method used in Shao (1992), using the asymptotic set-up for stratified multistage sampling as outlined, for example, in Shao and Tu (1995), Chapter 6. Here we provide a sketch of the proof of consistency of $\hat{\mathbf{V}}_J(\tilde{\mathbf{S}}_2)$ when only the basic survey weights are used in the estimating equations $\hat{\mathbf{S}}(\boldsymbol{\beta})$ given by

(2.4). Extension to post-stratified weights would involve the asymptotic theory for post-stratified estimators of totals (Yung (1996)).

The estimator $\hat{\mathbf{S}}(\boldsymbol{\beta})$ for stratified multi-stage sampling may be written as

$$\hat{\mathbf{S}}(\boldsymbol{\beta}) = \sum_{hik\in s} w_{hik}\mathbf{u}_{hik}(\boldsymbol{\beta}),$$

where w_{hik} is the basic survey weight attached to the kth element of the ith cluster in the hth stratum ($i = 1,\ldots,n_h$; $h = 1,\ldots,L$). Using the variance estimator for a total (Rao (1996), Section 3.1), it follows from (2.11) and (3.2) that a Taylor linearization variance estimator of $\tilde{\mathbf{S}}_2 = \hat{\mathbf{S}}_2(\tilde{\boldsymbol{\beta}})$ is given by

$$\hat{\mathbf{V}}_L(\tilde{\mathbf{S}}_2) = \sum_{h=1}^{L}\frac{1}{n_h(n_h-1)}\sum_{i=1}^{n_h}(\tilde{\mathbf{z}}_{hi}-\tilde{\mathbf{z}}_{h\cdot})(\tilde{\mathbf{z}}_{hi}-\tilde{\mathbf{z}}_{h\cdot})', \tag{A.1}$$

where $\tilde{\mathbf{z}}_{hi} = \sum_k (n_h w_{hik})\tilde{\mathbf{z}}_{hik}$ with $\tilde{\mathbf{z}}_{hik} = \mathbf{u}_{2hik}(\tilde{\boldsymbol{\beta}}) - \tilde{\mathbf{A}}\mathbf{u}_{1hik}(\tilde{\boldsymbol{\beta}})$ and $\tilde{\mathbf{z}}_{h\cdot} = n_h^{-1}\sum_i \tilde{\mathbf{z}}_{hi}$. Note that $\tilde{\mathbf{z}}_{hik}$ is simply given by (3.2) with the subscript i replaced by hik. The asymptotic consistency of $\hat{\mathbf{V}}_L(\tilde{\mathbf{S}}_2)$ follows along the lines of Binder (1983). The jackknife weights when the (gj)-th sample cluster is deleted are given by $w_{hik(gj)} = w_{hik}b_{gj}$, where b_{gj} is 0 if $(hi) = (gj)$; is $n_g/(n_g-1)$ if $h = g$ and $i \neq j$; is 1 if $h \neq g$. Replacing w_{hij} by $w_{hik(gj)}$ we get $\hat{\mathbf{S}}_{(gj)}(\boldsymbol{\beta})$, $\tilde{\boldsymbol{\beta}}_{(gj)}$ and $\tilde{\mathbf{S}}_{2(gj)} = \hat{\mathbf{S}}_{2(gj)}(\tilde{\boldsymbol{\beta}}_{(gj)})$.

Now expand $\tilde{\mathbf{S}}_{2(gj)}$ around $\tilde{\boldsymbol{\beta}}$ to get

$$\tilde{\mathbf{S}}_{2(gj)} \approx \hat{\mathbf{S}}_{2(gj)}(\tilde{\boldsymbol{\beta}}) - \mathbf{I}_{21}(\tilde{\boldsymbol{\beta}})(\tilde{\boldsymbol{\beta}}_{1(gj)} - \tilde{\boldsymbol{\beta}}_1). \tag{A.2}$$

Similarly,

$$0 = \hat{\mathbf{S}}_{1(gj)}(\tilde{\boldsymbol{\beta}}_{(gj)}) \approx \hat{\mathbf{S}}_{1(gj)}(\tilde{\boldsymbol{\beta}}) - \mathbf{I}_{11}(\tilde{\boldsymbol{\beta}})(\tilde{\boldsymbol{\beta}}_{1(gj)} - \tilde{\boldsymbol{\beta}}_1). \tag{A.3}$$

Also,

$$\hat{\mathbf{S}}_{(gj)}(\tilde{\boldsymbol{\beta}}) = \hat{\mathbf{S}}(\tilde{\boldsymbol{\beta}}) + \frac{1}{n_g-1}(\tilde{\mathbf{u}}_{g\cdot} - \tilde{\mathbf{u}}_{gj}), \tag{A.4}$$

where $\tilde{\mathbf{u}}_{g\cdot} = n_g^{-1}\sum_j \tilde{\mathbf{u}}_{gj}$ and $\tilde{\mathbf{u}}_{gj} = \sum_k (n_g w_{gjk})\tilde{\mathbf{u}}_{gjk}$. Now substituting for $\tilde{\boldsymbol{\beta}}_{1(gj)} - \tilde{\boldsymbol{\beta}}_1$ from (A.3) into (A.2) and using (A.4), we get

$$\tilde{\mathbf{S}}_{2(gj)} - \tilde{\mathbf{S}}_2 \approx -\frac{1}{n_g-1}(\tilde{\mathbf{z}}_{gj} - \tilde{\mathbf{z}}_{g\cdot}), \tag{A.5}$$

noting that $\hat{\mathbf{S}}_1(\tilde{\boldsymbol{\beta}}) = \mathbf{0}$.

Finally, substituting (A.5) into (3.1) we get $\hat{\mathbf{V}}_J(\tilde{\mathbf{S}}_2) \approx \hat{\mathbf{V}}_L(\tilde{\mathbf{S}}_2)$, and hence $\hat{\mathbf{V}}_J(\tilde{\mathbf{S}}_2)$ is asymptotically consistent.

References

Binder, D. A. (1983). On the variance of asymptotically normal estimators from complex surveys. *Internat. Statist. Rev.* **51**, 279-292.

Binder, D. A. and Patak, Z. (1994). Use of estimating functions for estimation from complex surveys. *J. Amer. Statist. Assoc.* **89**, 1035-1043.

Boos, D. D. (1992). On generalized score tests. *Amer. Statist.* **46**, 327-333.

Clayton, D. and Hills, M. (1993). *Statistical Methods in Epidemiology.* Oxford University Press, Oxford.

Fuller, W. A. (1975). Regression analysis for sample survey. *Sankya Ser. C* **37**, 117-132.

Godambe, V. P. (1991). Orthogonality of estimating functions and nuisance parameters. *Biometrika* **78**, 143-151.

Godambe, V. P. and Thompson, M. E. (1986). Parameters of superpopulation and survey population: their relationship and estimation. *Internat. Statist. Rev.* **54**, 127-138.

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.

Liang, K. Y. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.

Rao, J. N. K. (1996). Developments in sample survey theory: an appraisal. *Canad. J. Statist.* **25**, 1-21.

Rao, J. N. K. and Scott, A. J. (1992). A simple method for the analysis of clustered binary data. *Biometrics* **48**, 577-585.

Rotnitzky, A. and Jewell, N. P. (1990). Hypothesis testing of regression in semiparametric generalized linear models for cluster correlated data. *Biometrika* **77**, 485-497.

Rust, K. F. and Rao, J. N. K. (1996). Variance estimation for complex surveys using replication techniques. *Statist. Meth. Medical Research* **5**, 283-310.

Scott, A. J. and Wild, C. J. (1986). Fitting logistic models under case-control or choice-based sampling. *J. Roy. Statist. Soc. Ser. B* **48**, 170-182.

Shao, J. (1992). Jackknifing in generalized linear models. *Amer. Inst. Statist. Math.* **44**, 673-686.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap.* Springer, New York.

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. *J. Amer. Statist. Assoc.* **82**, 630-636.

Thomas, D. R., Singh, A. C. and Roberts, G. R. (1996). Tests of independence in two-way tables under cluster sampling: an evaluation. *Internat. Statist. Rev.* **64**, 295-311.

Yung, W. (1996). Contributions to poststratification in stratified multi-stage samples. Ph.D. Thesis, Carleton University, Ottawa, Canada.

Department of Mathematics and Statistics, Carleton University, Ottawa, K1S 5B6, Canada.

E-mail: jrao@math.carleton.ca

Department of Statistics, University of Auckland, Auckland, New Zealand.

E-mail: scott@stat.auckland.ac.nz

Department of Social Statistics, University of Southampton, Southampton, S09 5NH, United Kingdom.

E-mail: cjs@alcd.soton.ac.uk

(Received June 1997; accepted April 1998)