ANALYSIS OF DOMAIN MEANS IN COMPLEX SURVEYS

D.R. Bellhouse, University of Western Ontario and J.N.K. Rao, Carleton University
D.R. Bellhouse, Department of Statistical and Actuarial Sciences, London, ON N6A 5B7

Key Words: analysis of variance, complex surveys, generalized linear models.

1. INTRODUCTION AND NOTATION

Under simple random sampling all the standard techniques of analysis apply. When a complex design - for example, stratified multistage cluster sampling - is used, it can have a significant impact on the analysis of the data. Confidence intervals may not have the stated coverage properties and calculated p-values associated with testing a hypothesis may be inaccurate. Here we examine the analysis of domain means under complex sampling designs. Under the assumption of a normal distribution of the domain sample estimates, results are given corresponding to one and two-way analyses of variance (§2). The analysis of domain means is extended to include generalized linear models (§3) and an example using a Poisson model is examined (§§4 and 5).

Consider a population of size N divided into I domains each of size N_i, $i = 1, \ldots, I$. Let $\bar{Y}$ be the population mean, $\bar{Y}_i$ the mean of the i-th domain and $W_i = N_i / N$ the i-th domain weight. We denote the $I \times 1$ vector of domain means by $\bar{\mathbf{Y}} = (\bar{Y}_1, \ldots, \bar{Y}_I)'$ and its estimate under any sampling design by $\bar{\mathbf{y}} = (\bar{y}_1, \ldots, \bar{y}_I)'$. The $I \times I$ covariance matrix of $\bar{\mathbf{y}}$ under any sampling design is denoted by $\mathbf{V} = [V_{ij}]$. The matrix $\mathbf{V_d} = diag(V_{11}, \ldots, V_{II})$ contains only the variances of $\bar{\mathbf{y}}$ on the diagonal. Assuming simple random sampling and ignoring the finite population correction factor, $\mathbf{V}$ reduces to $\mathbf{S} = diag(S_1^2 / n_1, \ldots, S_I^2 / n_I)$, where S_i^2 is the finite population variance in the ith domain and n_i is the sample size observed in that domain. The design effect in the i-th domain is the ratio of the standard error of the estimate under the complex design to that of simple random sampling, or $deft_i = \sqrt{V_{ii} / (S_i^2 / n_i)}$. The estimate of $\mathbf{V}$ is denoted by $\hat{\mathbf{V}}$. A consistent estimate $\hat{\mathbf{S}}$ of $\mathbf{S}$ may be obtained under a complex design if the sample weights are available. The estimate of $deft_i$ is found on substituting the variance estimate v_{ii} for V_{ii} and a consistent estimate of S_i^2 / n_i obtained under the complex design. In §2 we develop methods of analysis under the assumption that $\hat{\mathbf{V}}_\mathbf{d}$ or the $deft_i$ are available. The results of §§3 - 5 are obtained under the assumption that the full estimated covariance matrix $\hat{\mathbf{V}}$ is available.

2. NORMAL MODELS - PARALLELS TO THE ANALYSIS OF VARIANCE

We assume that a finite population Central Limit Theorem holds so that $\bar{\mathbf{y}} \rightarrow_d MVN(\bar{\mathbf{Y}}, \mathbf{V})$. Consider the null hypothesis

$$H_0: \ \mathbf{C}\bar{\mathbf{Y}} = \mathbf{0} \tag{1}$$

where $\mathbf{C}$ is a $k \times I$ matrix of contrast coefficients such that $\mathbf{C1} = \mathbf{0}$. The column vectors $\mathbf{1}$ and $\mathbf{0}$ contain 1's and 0's respectively. If $\hat{\mathbf{V}}$ is available then the Wald statistic

$$X_W^2 = (\mathbf{C}\bar{\mathbf{y}})'(\mathbf{C}\hat{\mathbf{V}}\mathbf{C})^{-1}(\mathbf{C}\bar{\mathbf{y}}) \tag{2}$$

may be used to test H_0. Under H_0, $X_W^2 \rightarrow_d \chi_k^2$.

In many situations (a secondary analysis of the data, for example), $\hat{\mathbf{V}}$ may not be available. A modified Wald statistic, denoted by X_M^2 may be obtained on replacing $\hat{\mathbf{V}}$ in (1) by another matrix, say $\hat{\mathbf{A}} = diag(1/\hat{a}_1, \ldots, 1/\hat{a}_I)$, where $\hat{a}_i$ is a consistent estimate under the sampling design for a constant a_i for $i = 1, \ldots, I$. Then

$$X_M^2 = (\mathbf{C}\bar{\mathbf{y}})'(\mathbf{C}\hat{\mathbf{A}}\mathbf{C})^{-1}(\mathbf{C}\bar{\mathbf{y}}) \rightarrow_d \sum_{j=1}^{k} \delta_j Z_j^2 \tag{3}$$

where the Z_j are independent standard normal random variables and δ_j (j = 1, ..., k) are the eigenvalues of $(\mathbf{CAC'})^{-1}(\mathbf{CVC'})$. The Rao-Scott correction to this test is

$$X_M^2 / \bar{\delta} \tag{4}$$

where $\bar{\delta}$ is the mean of the δ's. A conservative correction to X_M^2, using Scott and Styan (1985), is given by

$$X_M^2 / \bar{\lambda}, \tag{5}$$

where $\bar{\lambda}$ is the mean of $\lambda_1, \ldots, \lambda_I$ which are the eigenvalues of $\mathbf{A}^{-1}\mathbf{V}$. The correction $\bar{\lambda}$ may be estimated by

$$\sum_{i=1}^{I} \hat{a}_i v_{ii} / I. \tag{6}$$

Consider two special cases:

I. Only the domain standard errors are available. In this case it is reasonable to set $\hat{\mathbf{A}} = \hat{\mathbf{V}}_d$ in X_M^2 in (3). The Rao-Scott correction, appearing in (4), is not estimable since estimates of the covariances are not available. Consequently, (4) is not calculable. The test statistic in (5) is calculable and reduces to X_M^2.

II. The domain design effects are available and a microdata file is available containing the sample weights from which a consistent estimate of S may be obtained. In this case set $\hat{\mathbf{A}} = \hat{\mathbf{S}}$ in (3). The test statistic in (5) is calculable since $\bar{\lambda}$ in (6) is given by the average of the I domain design effects.

In a one-way analysis we are interested in testing the hypothesis

$$H_0\colon\ \bar{Y}_1 = \bar{Y}_2 = \cdots = \bar{Y}_I. \tag{7}$$

Then **C** in (1) may be expressed as the $(I\text{-}1)\times I$ matrix

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & -1 \\ 0 & 1 & 0 & \cdots & 0 & -1 \\ 0 & 0 & 1 & \cdots & 0 & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix}. \tag{8}$$

The modified Wald statistic reduces to

$$X_M^2 = \sum_{i=1}^{I} \hat{a}_i (\bar{y}_i - \bar{y}')^2 \tag{9}$$

$$-\frac{1}{\sum_{i=1}^{I} \hat{a}_i} \sum_{i=1}^{I} \sum_{j=1}^{I} \hat{a}_i \hat{a}_j (\bar{y}_i - \bar{y}')(\bar{y}_i - \bar{y}');$$

the correction $\bar{\delta}$ in (4) is

$$[\sum_{i=1}^{I} a_i V_{ii} - \sum_{i=1}^{I} \sum_{j=1}^{I} a_i a_j V_{ij} / \sum_{i=1}^{I} a_i] / (I-1) \tag{10}$$

and the correction $\bar{\lambda}$ in (5) is

$$\sum_{i=1}^{I} a_i V_{ii} / I. \tag{11}$$

Special case I for the general hypothesis in equation (1) is obtained on setting $\hat{a}_i = 1/v_{ii}$ in (9). The correction in (11) reduces to $\bar{\lambda} = 1$. Special case II is obtained on setting $\hat{a}_i = n_i / \hat{S}_i$ in (9). The correction in (11) is the average of the I domain design effects as in the general case.

A test statistic which has a form similar to the treatment sum of squares in the one-way analysis of variance is obtained on setting $\hat{a}_i = \hat{W}_i$ where $\hat{W}_i$ is the estimate domain weight. Then (9) reduces to

$$X_M^2 = \sum_{i=1}^{I} \hat{W}_i (\bar{y}_i - \hat{\bar{Y}})^2$$

where

$$\hat{\bar{Y}} = \sum_{i=1}^{I} \hat{W}_i \bar{y}_i \quad \text{and} \quad \sum_{i=1}^{I} \hat{W}_i = 1.$$

The Rao-Scott correction in (10) to the test statistic in (9) is estimated by

$$[\sum_{i=1}^{I} \hat{W}_i v_{ii} - \mathrm{var}(\hat{\bar{Y}})] / (I-1),$$

where $\mathrm{var}(\hat{\bar{Y}})$ is the estimated variance of the population mean. The correction in (11) is estimated by

$$\sum_{i=1}^{I} \hat{W}_i v_{ii} / I.$$

In a two-way analysis we have I domains of one type and J domains of another for a cross classification of IJ domains. The notation extends in a natural way by using double subscripts; for example $\overline{Y}_{ij}$ is the population cell mean for the cross classification of domain i of Type I and domain j of Type II. The test for equality of the marginal means follows in the same fashion as the one-way analysis. One additional hypothesis of interest in the two-way analysis is the test for interaction. This may be written as

$$H_0\colon\ \overline{Y}_{ij} - \overline{Y}_{i'j} - \overline{Y}_{ij'} + \overline{Y}_{i'j'} = 0$$

for all $i \neq i'$ and $j \neq j'$. The matrix $\mathbf{C}$ in (1) is the Kroenecker product of two matrices of the form given in (8). Similar to the one-way analysis we can set $\hat{\mathbf{A}} = \hat{\mathbf{S}}$ or $\hat{\mathbf{A}} = \hat{\mathbf{V}}_\mathbf{d}$ in (3) which would require knowledge of cell design effects or the standard errors of the cell mean estimates respectively. A third approach is to set $\hat{\mathbf{A}} = \mathbf{I}$ the identity matrix. This yields

$$X_M^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{y}_{ij} - \bar{y}_{+j} - \bar{y}_{i+} + \bar{y}_{++})^2$$

where the "+" notation means that the appropriate subscript has been summed over and the result divided by the number in the sum. The correction to the test statistic given by (6) is calculable and is estimated using the IJ standard errors of the cell means.

3. GENERALIZED LINEAR MODELS

The use of generalized linear models under complex sampling designs has been considered by Molina and Skinner (1992). Here we provide corrections to the Wald statistic and apply these results (§4) to data under a Poisson model.

Consider a model for the domain means in which $E(\bar{y}_i) = \mu_i(\boldsymbol{\beta})$ and the model covariance matrix is $\mathbf{V}_\mathrm{m}(\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is a $p \times 1$ vector of paramaters. For simplicity we denote $\mu_i(\boldsymbol{\beta})$ by μ_i and $\mathbf{V}_\mathrm{m}(\boldsymbol{\beta})$ by $\mathbf{V}_\mathrm{m}$. The vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_I)'$. A hypothesis for goodness-of-fit of the model is

$$H_0\colon\ E(\bar{y}_i) = \mu_i(\boldsymbol{\beta}). \qquad (12)$$

An estimate of $\boldsymbol{\beta}$ is obtained by solving the quasi-likelihood equations

$$\mathbf{S}(\hat{\boldsymbol{\beta}}) = \mathbf{D}(\hat{\boldsymbol{\beta}})'\mathbf{V}_\mathbf{m}(\hat{\boldsymbol{\beta}})^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu}(\hat{\boldsymbol{\beta}})) = 0 \qquad (13)$$

for $\hat{\boldsymbol{\beta}}$, where $\mathbf{D}(\hat{\boldsymbol{\beta}}) = (\mathbf{D}_1(\hat{\boldsymbol{\beta}}), \ldots, \mathbf{D}_I(\hat{\boldsymbol{\beta}}))'$ and

$$\mathbf{D}_i(\hat{\boldsymbol{\beta}}) = \partial\mu_i(\hat{\boldsymbol{\beta}})/\partial\hat{\boldsymbol{\beta}}$$

is a $p \times 1$ vector. For simplicity we denote $\mathbf{D}(\hat{\boldsymbol{\beta}})$ by $\mathbf{D}$, $\mu_i(\hat{\boldsymbol{\beta}})$ by $\hat{\mu}_i$, $(\hat{\mu}_1, \ldots, \hat{\mu}_I)'$ by $\hat{\boldsymbol{\mu}}$ and $\mathbf{V}_\mathbf{m}(\hat{\boldsymbol{\beta}})$ by $\hat{\mathbf{V}}_\mathbf{m}$.

Under simple random sampling the domain means are independent. Then a test statistic for goodness-of-fit (the hypothesis in (12)) is given by

$$X_G^2 = (\bar{\mathbf{y}} - \hat{\boldsymbol{\mu}})'\hat{\mathbf{V}}_\mathrm{m}^{-1}(\bar{\mathbf{y}} - \hat{\boldsymbol{\mu}}) \rightarrow_d \chi^2_{I-p}. \qquad (14)$$

We examine the distribution of X_G^2 in (14) under a complex design in which the covariance matrix of $\bar{\mathbf{y}}$ is $\mathbf{V}$.

On performing a Taylor expansion of $\mathbf{S}$ around $\hat{\boldsymbol{\beta}}$ we obtain from (13)

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \approx [\mathbf{D}'\mathbf{V}_\mathrm{m}^{-1}\mathbf{D}]^{-1}\mathbf{S}(\hat{\boldsymbol{\beta}}). \qquad (15)$$

A Taylor series expansion of $\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}$ around $\hat{\boldsymbol{\beta}}$ yields

$$\begin{aligned}\hat{\boldsymbol{\mu}} - \boldsymbol{\mu} &\approx \mathbf{D}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ &= \mathbf{D}[\mathbf{D}'\mathbf{V}_\mathrm{m}^{-1}\mathbf{D}]^{-1}\mathbf{D}\mathbf{V}_\mathrm{m}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu})\end{aligned}$$

from (13) and (15). Since $\hat{\boldsymbol{\beta}} \rightarrow_p \boldsymbol{\beta}$

$$\bar{\mathbf{y}} - \hat{\boldsymbol{\mu}} \approx \mathbf{B}(\bar{\mathbf{y}} - \boldsymbol{\mu})$$

where

$$\mathbf{B} = \mathbf{I} - \mathbf{D}\boldsymbol{\Delta}^{-1}\mathbf{D}'\mathbf{V}_\mathrm{m}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu}) \qquad (16)$$

and

$$\boldsymbol{\Delta} = \mathbf{D}'\mathbf{V}_\mathrm{m}^{-1}\mathbf{D}.$$

From (16) and in view of $\hat{\mathbf{V}}_\mathrm{m} \rightarrow_p \mathbf{V}_\mathrm{m}$ we find that

$$X_G^2 \approx (\bar{\mathbf{y}}-\boldsymbol{\mu})'\mathbf{B}'\mathbf{V}_{\mathbf{m}}^{-1}\mathbf{B}(\bar{\mathbf{y}}-\boldsymbol{\mu})\,.$$

Then

$$X_G^2 \to_d \sum_{i=1}^{I-p} \phi_i Z_i^2$$

where the Z_i are independent standard normal random variables and the ϕ_i are the non-zero eigenvalues of $(\mathbf{BVB}')\mathbf{V}_{\mathbf{m}}^{-1}$. The first order Rao-Scott correction to this test is

$$X_G^2 / \bar{\phi}\ , \tag{17}$$

where

$$(I-p)\bar{\phi} = tr[(\mathbf{B}'\mathbf{VB})\mathbf{V}_{\mathbf{m}}^{-1}]\,. \tag{18}$$

After some algebra it may be shown that

$$\bar{\phi} \le [I/(I-p)]\bar{\psi}$$

where the ψ's are the eigenvalues of $\mathbf{V}_{\mathbf{m}}^{-1}\mathbf{V}$. Then $\bar{\psi} = tr(\mathbf{V}_{\mathbf{m}}^{-1}\mathbf{V})/I$ provides a conservative correction to the test. Note that $\bar{\psi}$ depends only on the design effects if $\mathbf{V}_{\mathbf{m}}$ is a diagonal matrix.

4. A POISSON MODEL

Assume in (12) that

$$\ln(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}\,, \tag{19}$$

where $\mathbf{X}$ is an $I \times p$ matrix of known constants. The model variance

$$\begin{aligned}\mathbf{V}_{\mathbf{m}} &= diag(\mu_1/n_1,\ldots,\mu_I/n_I)\\ &= \mathbf{D}_{\mu}\mathbf{D}_{\mathbf{w}}^{-1}/n\end{aligned} \tag{20}$$

where

$$\mathbf{D}_{\mu} = diag(\mu_1,\ldots,\mu_I)\,,$$

$$\mathbf{D}_{\mathbf{w}} = diag(n_1/n,\ldots n_I/n)$$

and n is the total sample size. The quasi-likelihood equations in (13) reduce to

$$\mathbf{X}'\mathbf{D}_{\mathbf{w}}(\bar{\mathbf{y}}-\boldsymbol{\mu}) = \mathbf{0}$$

where $\mathbf{D}$ in (13) is $\mathbf{D}_{\mu}\mathbf{X}$. The matrix $\mathbf{B}$ in (16) may be expressed aa

$$\mathbf{B} = \mathbf{I} - \mathbf{D}_{\mathbf{w}}^{-1}(\mathbf{D}_{\mathbf{w}}\mathbf{D}_{\mu})\mathbf{X}(\mathbf{X}'\mathbf{D}_{\mathbf{w}}\mathbf{D}_{\mu}\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}_{\mathbf{w}}\,.$$

The uncorrected test statistic for goodness-of-fit, given by (14) reduces to

$$X_G^2 = \sum_{i=1}^{I}(n_i\bar{y}_i - n_i\hat{\mu}_i)^2/(n_i\hat{\mu}_i)\,. \tag{21}$$

for the Poisson case. The correction to this test statistic, given by (18), may be expressed as

$$\begin{aligned}(I-p)\bar{\phi} &= tr[n\mathbf{V}\mathbf{D}_{\mathbf{w}}\mathbf{D}_{\mu}^{-1}]\\ &\quad - tr[n\mathbf{V}\mathbf{D}_{\mathbf{w}}\mathbf{X}(\mathbf{X}'\mathbf{D}_{\mathbf{w}}\mathbf{D}_{\mu}\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}_{\mathbf{w}}].\end{aligned} \tag{22}$$

On replacing $\mathbf{V}$ by $\hat{\mathbf{V}}$ and $\mathbf{D}_{\mu}$ by $\mathbf{D}_{\bar{\mathbf{y}}} = diag(\bar{y}_1,\ldots,\bar{y}_I)$ in (22) an estimate of the correction to the test is obtained. The conservative correction to this test is the average of the $deft_i^2$.

These results are similar to those in Roberts *et al.* (1987) except that their binomial covariance matrix $\mathbf{D}_{\mathbf{w}}^{-1}\mathbf{D}_{\mathbf{f}}/n$, where $\mathbf{D}_{\mathbf{f}} = diag(f_1(1-f_1),\ldots, f_I(1-f_I))$, is replaced by $\mathbf{D}_{\mathbf{w}}^{-1}\mathbf{D}_{\mu}/n$.

Nested hypotheses may be treated in the following way. Suppose that the matrix $\mathbf{X}$ is partitioned into $(\mathbf{X}_1,\mathbf{X}_2)$, where $\mathbf{X}_1$ is $I \times q$ and $\mathbf{X}_2$ is $I \times r$ ($p = q + r$). Then (19) may be written as

$$\ln(\boldsymbol{\mu}) = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2\,.$$

We are interested in testing the hypthesis

$$H_0\colon\ \boldsymbol{\beta}_2 = \mathbf{0}\,. \tag{23}$$

The uncorrected test statistic is given by

$$X_G^2(2\mid 1) = \sum_{i=1}^{I}(n_i\hat{\mu}_i - n_i\tilde{\mu}_i)^2/n_i\tilde{\mu}_i\,, \tag{24}$$

where $\hat{\mu}_i$, $i = 1, \ldots, I$, are the solutions to the quasi-likelihood equations in (13) under the full model given by (19) and the $\tilde{\mu}_i$ are the solutions to (19) for the reduced model under the null hypothesis in (23). The corrected test statistics is given by

$$X_G^2(2\mid 1)/\bar{\gamma}$$

where $\bar{\gamma}$ is estimated by

$$tr[n(\tilde{\mathbf{X}}_2{}'\mathbf{D}_w\mathbf{D}_{\bar{y}}\tilde{\mathbf{X}}_2)^{-1}(\tilde{\mathbf{X}}_2{}'\mathbf{D}_w\hat{\mathbf{V}}\mathbf{D}_w\tilde{\mathbf{X}}_2)/r\,. \quad (25)$$

The matrix

$$\tilde{\mathbf{X}}_2 = \mathbf{X}_2 - \mathbf{X}_1(\mathbf{X}_1{}'\mathbf{D}_w\mathbf{D}_{\bar{y}}\mathbf{X}_1)^{-1}(\mathbf{X}_1{}'\mathbf{D}_w\mathbf{D}_{\bar{y}}\mathbf{X}_2)\,.$$

5. NUMERICAL EXAMPLE FROM THE WORLD FERTILITY SURVEY

Little ((1978) has fit generalized linear models under a Poisson assumption to Fijian data from the World Fertility Survey. His analysis ignores any complex sampling design that may have been used. We use part of his data to illustrate the adjustments to the test statistic that can be made to account for the complex sampling design.

The data in Table 1 are adapted from Table 3 of Little (1978). The data in the table in Roman script show the mean number of children ever born for women of Indian race cross classified by education and years since the woman's first marriage. The data in italics in the table are the cell sample sizes.

Table 1

		Education	
	None	Lower Primary	Upper Primary or Higher
Years Since First Marriage			
<5	0.97	0.96	0.90
	62	*102*	*154*
5 - 9	2.44	2.71	2.42
	70	*117*	*102*
10 - 14	4.14	4.14	3.85
	88	*132*	*59*
15 - 19	5.06	5.59	4.42
	114	*86*	*31*
20 - 24	6.46	6.34	5.48
	117	*68*	*25*
25+	7.48	7.81	5.80
	195	*59*	*10*

No variance estimates are given in Little's (1978) report. For the purposes of illustration we construct $\hat{\mathbf{V}}$ in the following way from $\mathbf{V}_m$, given by (20) and the cell means in Table 1. Suppose $deft_i^2 = b$, a constant for all i and the design effect for the difference of two means $deft_{ij} = c$ for all i and j. Then

$$V(\bar{y}_i) = b\frac{\mu_i}{n_i} \quad (26)$$

and

$$V(\bar{y}_i - \bar{y}_j) = c\left(\frac{\mu_i}{n_i} + \frac{\mu_j}{n_j}\right) \quad (27)$$

From (26) and (27) we obtain

$$Cov(\bar{y}_i, \bar{y}_j) = d\left(\frac{\mu_i}{n_i} + \frac{\mu_j}{n_j}\right), \quad (28)$$

where $d = (b-c)/2$. The diagonal elements of $\hat{\mathbf{V}}$ are obtained from (26) and the off-diagonal elements from (28) with μ_i and μ_j replaced by $\bar{y}_i$ and $\bar{y}_j$ respectively. For the purposes of illustration we chose $b = 2.0$ and $c = 1.2$.

Consider the loglinear model

$$\ln(\mu_{ij}) = \beta_0 + \beta_{1i} + \beta_{2j}, \quad (29)$$

where $i = 1, \ldots, 6$ and j = 1, 2, 3 and the restrictions $\beta_{16} = 0$ and $\beta_{23} = 0$ to ensure full rank.

On using (20), the uncorrected test statistic for goodness-of-fit of the model (29) is

$$X_G^2 = 5.9.$$

The correction to this test statistic, calculated from (22), is

$$\hat{\bar{\phi}} = 1.2,$$

which yields a corrected value of the test statistic of 4.915. This value is compared to the 5% point of χ_{10}^2 (18.3) is not significant so that it may be concluded that the Poisson model is a reasonable assumption.

The hypothesis that there is no effect due to education on the average number of children born givne model (29) may be written as

$$H_0: \beta_{21} = \beta_{22} = 0. \qquad (30)$$

The uncorrected test statistic for the hypothesis in (30), calculated from (24), reduces to

$$X_G^2(2 \mid 1) = 12.942.$$

The correction to this test statistic, calculated from (25), is

$$\hat{\bar{\gamma}} = 1.340,$$

which yields a corrected value of the test statistic of 9.66. When this value is compared to a χ_2^2 the test is significant at the 1% level.

REFERENCES

Little, R.J.A. (1978). Generalized linear models for cross-classified data from the WFS. *World Fertility Survey Technical Bulletins No. 5/Tech. 834.*

Molina, C., E.A. and Skinner, C.J. (1992). Pseudo-likelihood and quasi-likelihood estimation for complex sampling schemes. *Computational Statistics & Data Analysis* **13**: 395 - 405.

Roberts, G., Rao, J.N.K. and Kumar, S. (1987). Logistic regression analysis of sample survey data. *Biometrika* **74**: 1 - 12.

Scott, A.J. and Styan, G.P.H. (1985). On a separation theorem for generalized eigenvalues and a problem in the analysis of sample surveys. *Linear Algebra and Its Applications 70*: 209 - 224.