

DIAGNOSTICS FOR MULTIPLE RESPONSE DATA

THOMAS SUESSE — IVY LIU

ABSTRACT. Surveys often allow arbitrary number of responses. Each of the categorical responses is referred as an item. Marginal modeling of items simultaneously requires to incorporate the dependence between items. We investigate deletion diagnostics as Cook distance and DBETA for these marginal models based on homogenous linear predictor (HLP) model fitting and compare results with the generalized estimation equations (GEE) approach.

1. Introduction

Surveys often contain qualitative variables for which respondents may select any number of the outcome categories. This type of response is called multiple responses. Each outcome category refers to an item, where the items are dependent. Agresti and Liu [2] introduced marginal models based on marginal counts of each item. The models describe the association between the items and some explanatory variables taking into account the dependence. We want to consider two model approaches: generalized estimation equations (GEE) [9], a generalization of quasi-likelihood, and maximum likelihood (ML) estimation for homogeneous linear predictor (HLP) models [8]. This article focuses on the computation of deletion diagnostics such as DBETA and the Cook distance considering HLP models and compares these with GEE diagnostics. Deletion diagnostics determine the influence of a deleted set of observations on the parameters estimate. For GEE, Preisser and Qaqish [11] considered Cook distance and derived generalized one-step approximation formulas. We consider Cook distance for HLP models for analyzing multiple response data. We provide a variety of partly equivalent deletion methods for full solutions and one-step approximations of the Cook distance. The methods are illustrated using an example. Below, an example for multiple responses is introduced, then section 2 discusses marginal models and model fitting for GEE and HLP models. In the

2000 Mathematics Subject Classification: 62J20, 62J02, 62J12, 62J99, 62-07.

Keywords: multiple responses, diagnostics, Cook distance, marginal model fitting, GEE, HLP models.

next section we introduce deletion diagnostics and discuss their application on both fitting approaches. The article finishes with applying of the diagnostics on the introduced example and its discussion.

Many surveys allow each respondent to respond to more than one category or “item”. The respondent can pick any number out of c categories. This multiple outcome variable is referred to as pick any/ c variables and the corresponding data as pick any/ c data, where “/” stands for “out of” [4]. Each of the c category responses is called an item [1].

Let c items be given and let the multiple response be $\mathbf{l}_i = (l_{i1}, \dots, l_{ic})$, $l_{ij} \in \{0, 1\}$ for subject $i = 1, \dots, n$. There are 2^c such responses. Let \mathbf{X} be a column vector of predictor variables with K possible combinations of settings with $\sum_{k=1}^K n_k = n$, where n_k denotes the number of observations for setting k . We assume independent multinomial distributions with 2^c possible outcomes for each of the K settings. The multinomial cell counts for setting $i = 1, \dots, K$ and multiple response $l = 1, \dots, 2^c$ are denoted by v_{il} , equivalently, let τ_{il} denote the multinomial probabilities and m_{il} be the mean cell counts. The marginal counts for item $j = 1, \dots, c$ and setting $i = 1, \dots, K$ is denoted by y_{ij} and equivalently, the marginal probabilities and means by π_{ij} and μ_{ij} . Let \mathbf{v} denote the column vector including all v_{il} , equivalently, \mathbf{y} , $\boldsymbol{\tau}$, $\boldsymbol{\pi}$, \mathbf{m} and $\boldsymbol{\mu}$. The multinomial quantities can be transformed to marginal quantities by:

$$\pi_{ij} = \sum_{l=(l_1, \dots, l_c): l_j=1} \tau_{il}, \quad y_{ij} = \sum_{l=(l_1, \dots, l_c): l_j=1} v_{il} \quad \text{and} \quad \mu_{ij} = \sum_{l=(l_1, \dots, l_c): l_j=1} m_{il}$$

equivalently being denoted by $\mathbf{y} = \mathbf{B}\mathbf{v}$, $\boldsymbol{\pi} = \mathbf{B}\boldsymbol{\tau}$, $\boldsymbol{\mu} = \mathbf{B}\mathbf{m}$ with $K \cdot c \times K \cdot 2^c$ matrix \mathbf{B} containing only 0s and 1s. Another notation being used for GEE are the original observations $\mathbf{y}_i = \mathbf{l}_i$ for subject $i = 1, \dots, n$ appropriately define $\boldsymbol{\mu}_i = \boldsymbol{\pi}_i$.

In a study [14], 262 farmers were questioned about their veterinary information sources. They could choose the following categories: (A) professional consultant, (B) veterinarian, (C) state or local extension service, (D) magazines, and (E) feed companies and reps. Agresti and Liu [2] used “education” and “size” of a farm as explanatory variables. The farmers were questioned whether they had any kind of college education (“some college”) or not (“no college”). The size of farms depended on the number of pigs they marketed annually: less than 1,000, 1,000 to 2,000, 2,000 to 5,000, more than 5,000. Table 1 shows the marginal counts for education and size of a farm, whereas table 2 shows the multinomial counts.

DIAGNOSTICS FOR MULTIPLE RESPONSE DATA

TABLE 1. Marginal Table of Farmers' Veterinary Information Sources by Education and Number of Pigs [14].

<i>Number of Positive Responses</i>							
Education	Number of Pigs	Information Source					Subjects
		A	B	C	D	E	
No College	<1000	2	13	18	22	17	42
	1000-2000	2	15	10	11	15	27
	2000-5000	7	10	10	14	11	22
	>5000	13	10	7	14	7	27
Some College	<1,000	3	16	21	33	22	53
	1000-2000	2	10	15	22	10	42
	2000-5000	1	7	7	7	6	20
	>5000	14	9	7	8	5	29
Total		44	90	95	131	93	262

TABLE 2. Joint Table of Farmers' Veterinary Information Sources by Education and Number of Pigs [14].

Number of Joint Counts									
	No College				Some College				
Number= (Binary Sequence)	Number of Pigs								Total
	<1000	1000-2000	2000-5000	>5000	<1000	1000-2000	2000-5000	>5000	
1=(00000)	0	0	0	0	0	0	0	0	0
2=(00001)	3	4	1	2	11	6	3	2	32
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
32=(11111)	1	2	3	2	3	0	0	1	12
Subjects	42	27	22	27	53	42	20	30	262

2. Marginal models

We want to model the marginal probabilities with 1:1 link function g , linear predictor $\boldsymbol{\eta}$, parameter estimate $\boldsymbol{\beta}$ and design matrix \mathbf{X} :

$$\boldsymbol{\eta} = g(\boldsymbol{\pi}) = g(\mathbf{B}\boldsymbol{\tau}) = \mathbf{X}\boldsymbol{\beta} \quad \text{with} \quad p = |\boldsymbol{\beta}|. \quad (1)$$

2.1. GEE

Maximum likelihood (ML) estimates for generalized linear models (GLM) [10] can be obtained by fitting each of the c models separately. However, models depend on items and items are not independent. More efficient parameter estimation can be obtained by the generalized estimation equation (GEE) method [9],

where marginal models are fitted simultaneously and a chosen correlation structure is incorporated, which is an extension of the quasi-likelihood method [16]. Let $\text{Var}(\mathbf{Y}) = \mathbf{f} \cdot \phi^{-1}$ with variance function $\mathbf{f} = \mathbf{f}(\boldsymbol{\mu})$ and scale or dispersion parameter ϕ . GEE estimates are obtained by computing the root of the GEE (or quasi-score) equations:

$$\sum_{i=1}^n (\partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta})^T (\mathbf{A}_i \mathbf{R}_i(\alpha) \mathbf{A}_i)^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = 0$$

where $\partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}$ is a $c \times p$ matrix, $\mathbf{A}_i = \sqrt{\mathbf{f}_i}$ is a $c \times c$ diagonal matrix, $\mathbf{R}_i(\alpha)$ is the $c \times c$ correlation matrix for observation (cluster) i ($i = 1, \dots, n$) which is a function of parameter(s) α . If design matrix \mathbf{X} has full column rank, $\boldsymbol{\beta}$ can be estimated by iterated weighted least squares [11]:

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \left(\mathbf{X}^T \mathbf{W}^{(k)} \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{W}^{(k)} \mathbf{Z}^{(k)}, \quad k = 0, 1, 2, \dots \quad (2)$$

with $\mathbf{Z} = \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{D}(\mathbf{y} - \boldsymbol{\mu})$, block diagonal matrices $\mathbf{W} = \text{diag}(\mathbf{W}_1, \dots, \mathbf{W}_n)$ and $\mathbf{D} = \text{diag}(\mathbf{D}_1, \dots, \mathbf{D}_n)$ with $\mathbf{W}_i = \mathbf{D}_i^{-1} \mathbf{A}_i^{-1} \mathbf{R}_i^{-1}(\alpha) \mathbf{A}_i^{-1} \mathbf{D}_i^{-1}$, and $\mathbf{D}_i = \partial \boldsymbol{\eta}_i / \partial \boldsymbol{\mu}_i$ for cluster i assuming the dispersion parameter ϕ and the correlation matrix $\mathbf{R}(\alpha) = \text{diag}(\mathbf{R}_1(\alpha), \dots, \mathbf{R}_K(\alpha))$ are known and given. If unknown, they must be estimated consistently for every iterate by e.g. using the Pearson residuals as suggested in [9]. Under mild regularity conditions *Liang* and *Zeger* [9] show, as $n \rightarrow \infty$, $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is multivariate normal with mean 0 and some robust covariance matrix (see [9]).

2.2. HLP models

Liang (2005) generalized the class of generalized log-linear models to homogeneous linear predictor (HLP) models. For previous work see *Liang* and *Agresti* [6], *Liang* [7], [8] and *Bergsma* [3]. HLP models have the form

$$\mathbf{L}(\mathbf{m}) = \mathbf{X} \boldsymbol{\beta} \quad (3)$$

where \mathbf{L} is a smooth function from the multinomial counts satisfying certain conditions, see *Liang* (2005) for details. The marginal model (1) depends on $\boldsymbol{\pi}$ and therefore on $\boldsymbol{\tau}$. According to *Liang* (2005), models being expressed in terms of $\boldsymbol{\tau}$ are automatically HLP models. We have $\mathbf{L}(\mathbf{m}) = g(\boldsymbol{\pi}) = g(\mathbf{A} \boldsymbol{\tau})$ with $\mathbf{L} : \mathbb{R}^{K^{2^c}} \rightarrow \mathbb{R}^{K^c}$. Let \mathbf{U} be the orthogonal complement of \mathbf{X} (\mathbf{X} full column rank), then define $\mathbf{h}(\mathbf{m}) := \mathbf{U}^T \mathbf{L}(\mathbf{m})$ and $\mathbf{H} := \frac{\partial \mathbf{h}(\mathbf{m})}{\partial \mathbf{m}} = \frac{\partial \mathbf{L}^T}{\partial \mathbf{m}} \mathbf{U}$. From (3), it follows $\mathbf{h}(\mathbf{m}) = \mathbf{0}$. The following iteration scheme with reparametrization $\boldsymbol{\xi} = \log \mathbf{m}$ was recommended by *Liang* (2005) based on constrained (by the model) maximization of the likelihood kernel $l(\boldsymbol{\xi}; \mathbf{v}) = \mathbf{v}^T \boldsymbol{\xi}$:

$$\hat{\boldsymbol{\theta}}^{new} = \hat{\boldsymbol{\theta}} - \mathbf{G}(\hat{\boldsymbol{\theta}})^{-1}g(\hat{\boldsymbol{\theta}}) \quad (4)$$

with

$$g(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{v} - \mathbf{e}^{\boldsymbol{\xi}} + \mathbf{H}(\boldsymbol{\xi})\boldsymbol{\lambda} \\ \mathbf{h}(\boldsymbol{\xi}) \end{bmatrix} \quad \text{and} \quad \mathbf{G}(\boldsymbol{\theta}) = \begin{pmatrix} -\mathbf{D}(\mathbf{e}^{\boldsymbol{\xi}}) & \mathbf{H}(\boldsymbol{\xi}) \\ \mathbf{H}(\boldsymbol{\xi})^T & \mathbf{0} \end{pmatrix}$$

and $\boldsymbol{\theta} = (\boldsymbol{\xi}^T, \boldsymbol{\lambda}^T)^T$, where $\mathbf{D}(\mathbf{x})$ denotes a diagonal matrix with vector \mathbf{x} on the diagonal and $\boldsymbol{\lambda}$ denoting the Lagrange multiplier. Suppose a final solution $\hat{\mathbf{m}}$ exists, the parameter estimate is computed by $\hat{\boldsymbol{\beta}} = \mathbf{R}_{\mathbf{X}}\mathbf{L}(\hat{\mathbf{m}})$ with $\mathbf{R}_{\mathbf{X}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. The asymptotic covariance for our model is (HLP model of zero order):

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = \left(\mathbf{X}^T \frac{\partial \mathbf{L}(\hat{\mathbf{m}})}{\partial \mathbf{m}^T} \mathbf{D}(\hat{\mathbf{m}}) \frac{\partial \mathbf{L}(\hat{\mathbf{m}})^T}{\partial \mathbf{m}} \mathbf{X} \right)^{-1}.$$

The inverse of the matrix \mathbf{G} can be simplified [8] by applying the inverse formula of a partitioned matrix (e.g., in Searle, p. 261 [15]) and is:

$$\mathbf{G}^{-1} = \begin{pmatrix} -\mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{H}(\mathbf{H}^T\mathbf{D}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{D}^{-1} & \mathbf{D}^{-1}\mathbf{H}(\mathbf{H}^T\mathbf{D}^{-1}\mathbf{H})^{-1} \\ (\mathbf{H}^T\mathbf{D}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{D}^{-1} & (\mathbf{H}^T\mathbf{D}^{-1}\mathbf{H})^{-1} \end{pmatrix}. \quad (5)$$

Note that ML estimates are not properly defined for zero cell counts as in our example and estimates are then called “extended ML estimates”.

3. Deletion diagnostics

Let $\hat{\boldsymbol{\beta}}$ be parameter estimate of all observations and let $\hat{\boldsymbol{\beta}}_{[d]}$ be the parameter estimate having a set d of observations deleted. For given d the diagnostics DBETA and Cook distance (CD) [5] are defined as follows:

$$\begin{aligned} \text{DBETA}_{[d]} &= \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{[d]}, \\ \text{CD}_{[d]} &= (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{[d]})^T \text{Cov}(\hat{\boldsymbol{\beta}})^{-1} (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{[d]}) / p. \end{aligned} \quad (6)$$

3.1. GEE-diagnostics

The full solution for $\hat{\beta}_{[d]}$ given index set d of observations to be deleted is obtained by applying (2) with starting value $\hat{\beta}$ and replacing \mathbf{X} , \mathbf{W} and \mathbf{Z} by $\mathbf{X}_{[d]}$, $\mathbf{V}_{[d]}^{-1}$ and $\mathbf{Z}_{[d]}$ respectively with $\mathbf{V} = \mathbf{W}^{-1}$, \mathbf{W} and \mathbf{Z} partitioned as

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{[d]} & \mathbf{W}_{[d]d} \\ \mathbf{W}_{d[d]} & \mathbf{W}_d \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} \mathbf{Z}_{[d]} \\ \mathbf{Z}_d \end{pmatrix}$$

such that without loss of generality the deleted observations are at the “end” of \mathbf{W} and \mathbf{Z} . Matrices and vectors are now partitioned similarly. It follows $\mathbf{V}_{[d]}^{-1} = \mathbf{W}_{[d]} - \mathbf{W}_{[d]d} \mathbf{W}_d^{-1} \mathbf{W}_{d[d]}$. Computing full solutions for DBETA and Cook distance for all possible index sets d can be very time consuming. Preiser and Qaqish [11] derived one step approximations for these measures generalizing one step approximations for GLM [17] and logistic regression models [13]. One step approximation for $\hat{\beta}_{[d]}$ are obtained by applying only one iteration. For ordinary linear models the introduction of dummy variables is equivalent to deletion [12]. We will apply this now for GEE. The solution $\tilde{\beta}_{[d]}$ is equal to $\hat{\beta}_{[d]}$ with

$$\tilde{\beta} = \begin{pmatrix} \tilde{\beta}_{[d]} \\ \tilde{\beta}_d \end{pmatrix}. \quad (7)$$

where $\tilde{\beta}$ is the solution for all observations but with design matrix $\tilde{\mathbf{X}}$:

$$\tilde{\mathbf{X}} = \begin{pmatrix} \mathbf{X}_{[d]} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_d \end{pmatrix}, \quad (8)$$

where \mathbf{I}_d is the identity matrix of length $|d|$ and creates the dummy variables $\tilde{\beta}_d$ for the deleted observations. However, this is only true, if the correlations of both methods are the same for each step (e.g. known). It is sufficient to show that applying one step yields the same new iterates $\tilde{\beta}_{[d]}$ and $\hat{\beta}_{[d]}$ with the same starting values.

Proof. (Applying formula for inverse of partitioned matrix, e.g. in Searle, p. 261 [15]):

$$\begin{aligned} (\tilde{\mathbf{X}}^T \mathbf{W} \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \mathbf{W} \mathbf{Z} &= \begin{pmatrix} \mathbf{X}_{[d]}^T \mathbf{W}_{[d]} \mathbf{X}_{[d]} & \mathbf{X}_{[d]}^T \mathbf{W}_{[d]d} \\ \mathbf{W}_{d[d]} \mathbf{X}_{[d]} & \mathbf{W}_d \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}_{[d]}^T \mathbf{W}_{[d]} \mathbf{Z}_{[d]} + \mathbf{X}_{[d]}^T \mathbf{W}_{[d]d} \mathbf{Z}_d \\ \mathbf{W}_{d[d]} \mathbf{Z}_{[d]} + \mathbf{W}_d \mathbf{Z}_d \end{pmatrix} \\ &= \begin{pmatrix} (\mathbf{X}_{[d]}^T \mathbf{V}_{[d]}^{-1} \mathbf{X}_{[d]})^{-1} & (\mathbf{X}_{[d]}^T \mathbf{V}_{[d]}^{-1} \mathbf{X}_{[d]})^{-1} \mathbf{X}_{[d]}^T \mathbf{W}_{[d]d} \mathbf{W}_d^{-1} \\ \mathbf{W}_d^{-1} \mathbf{W}_{d[d]} \mathbf{X}_{[d]} (\mathbf{X}_{[d]}^T \mathbf{V}_{[d]}^{-1} \mathbf{X}_{[d]})^{-1} & \mathbf{W}_d^{-1} \mathbf{W}_{d[d]} \mathbf{X}_{[d]} (\mathbf{X}_{[d]}^T \mathbf{V}_{[d]}^{-1} \mathbf{X}_{[d]})^{-1} \mathbf{X}_{[d]}^T \mathbf{W}_{[d]d} \mathbf{W}_d^{-1} \end{pmatrix} \\ &\quad \begin{pmatrix} \mathbf{X}_{[d]}^T \mathbf{W}_{[d]} \mathbf{Z}_{[d]} + \mathbf{X}_{[d]}^T \mathbf{W}_{[d]d} \mathbf{Z}_d \\ \mathbf{W}_{d[d]} \mathbf{Z}_{[d]} + \mathbf{W}_d \mathbf{Z}_d \end{pmatrix} \end{aligned}$$

$$= \begin{pmatrix} (\mathbf{X}_{[d]}^T \mathbf{V}_{[d]}^{-1} \mathbf{X}_{[d]})^{-1} \mathbf{X}_{[d]}^T \mathbf{V}_{[d]}^{-1} \mathbf{Z}_{[d]} \\ -\mathbf{W}_d^{-1} \mathbf{W}_{d[d]} \left(\mathbf{X}_{[d]} (\mathbf{X}_{[d]}^T \mathbf{V}_{[d]}^{-1} \mathbf{X}_{[d]})^{-1} \mathbf{X}_{[d]}^T \mathbf{V}_{[d]}^{-1} \mathbf{Z}_{[d]} - \mathbf{Z}_{[d]} \right) + \mathbf{Z}_d \end{pmatrix}.$$

□

3.2. HLP diagnostics

In general, the marginal counts \mathbf{y} have lower dimension than the joint counts \mathbf{v} . There are several ways of deleting. Let us assume that (*) deleting a set d of marginal observations is equivalent to deleting a unique set d' of joint observations. The most natural way is deleting/manipulating the joint counts. We can show that the following methods are equivalent under (*):

- (1) Deleting the set d' of joint counts and hence deleting the set d of marginal counts.
- (2) Deleting only the set d of marginal counts by changing function L , s.t. L maps all joint counts to the marginal counts with the deleted set d excluded (simply omitting these components of L being in set d).
- (3) Applying design matrix $\tilde{\mathbf{X}}$ (8) in iteration scheme (4) instead of \mathbf{X} .

Assumption (*) is not always fulfilled, however, methods (2.) and (3.) are always equivalent (resulting in identical steps). When one deletes marginal counts with assumption (*) not holding, e.g., deleting a single item, methods (1.), (2.) and (3.) still yield equal full solutions, only 1-step approximations differ slightly (see Figure 2), where method (1.) refers now to manipulating the joint counts such that the marginal counts are automatically deleted.

For deletion of joint counts without assumption (*), s.t. no marginal observations are deleted, these simplifications cannot be made. The proof of equivalence is omitted and left to the reader.

A one-step solution for $\beta_{[d]}$ is:

$$\hat{\beta}_{[d]} = \mathbf{R}_{\mathbf{X}_{[d]}} \mathbf{L} \left(\exp \left[(-\mathbf{D}^{-1} + \mathbf{D}^{-1} \mathbf{H}_{[d]} \mathbf{A} \mathbf{H}_{[d]}^T \mathbf{D}^{-1}) (\mathbf{v} - \hat{\mathbf{m}}) + \mathbf{D}^{-1} \mathbf{H}_{[d]} \mathbf{A} \mathbf{h}_{[d]} \right] \right) \quad (9)$$

with $\mathbf{A} (\mathbf{H}_{[d]}^T \mathbf{D}^{-1} \mathbf{H}_{[d]})^{-1}$. One could also expand \mathbf{L} in a first order Taylor series, but the formula would not simplify.

4. Example and discussion

One model denoted by “LIN S” (linear in farm size) for the farmer’s data [2] is

$$\log \left(\frac{\pi_{ij}}{1 - \pi_{ij}} \right) = \alpha_j + \beta_j^S x_i \quad (10)$$

with equally spaced scores $x_i = 1, 2, 3, 4$ depending on farm size ($< 1,000, \dots, > 5,000$). We computed the Cook distance for the farmers' model (10) with GEE and HLP fitting algorithms deleting farmsize and item \times education (see Figure 1), where full solutions and one step approximations (only 1 iteration, for details see [13],[17]) are shown. The results show that GEE and ML diagnostics are

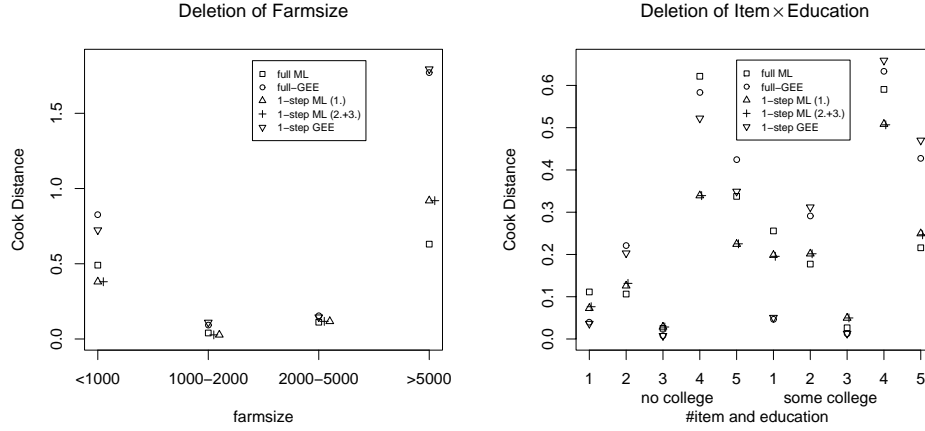


FIGURE 1. Cook distance for model (10) and deletion of farmsize left and deletion of item \times Education right.

approximately the same and methods 1, 2 and 3 are also the same for full solutions and very close for one step approximations even if condition (*) does not hold. GEE and HLP (ML) deletion diagnostics both have their limitations. GEE is not based on maximum likelihood and should only be applied if HLP diagnostics are not applicable due to either too many zero cell counts or the huge number of multinomial parameter. We investigated marginal models for multiple response data, however, the introduced deletion methods do not depend on marginal models only, but are generally applicable for GEE and HLP models. Furthermore, the deletion methods do not depend on GEE and HLP models only, but also on modeling based on the introduced iteration schemes - (2) and (4). Looking at the results of the Farmer's example, one can see less variation of the Cook distance for HLP, which can be explained by the more unstable behaviour of GEE due to the simultaneous estimation of the correlation and its parameter. Thus, high Cook distance values obtained by GEE are suspicious of being only a result of GEE fitting and not by influential observations.

REFERENCES

- [1] AGRESTI, A.—LIU, I.-M.: *Modeling responses to a categorical variable allowing arbitrarily many category choices*. Technical Report 575, University of Florida, Department of Statistics, 1998.
- [2] AGRESTI, A.—LIU, I.: *Strategies for modeling a categorical variable allowing multiple category choices*, *Sociol. Methods and Research* **29** (2001), 403–434.
- [3] BERGSMAN, W. P.: *Marginal Models for Categorical Data*. Tilburg University Press, Tilburg, 1997.
- [4] COOMBS, C. H.: *A Theory of Data*. Wiley, New York, 1964.
- [5] COOK, R. D.: *Detection of influential observation in linear-regression*, *Technometrics* **19** (1977), 15–18.
- [6] LANG, J. B.—AGRESTI, A.: *Simultaneously modeling joint and marginal distributions of multivariate categorical responses*, *J. Amer. Statist. Assoc.* **89** (1994), 625–632.
- [7] LANG, J. B.: *Multinomial-poisson homogeneous models for contingency tables*, *Ann. Statist.* **32** (2004), 340–383.
- [8] LANG, J. B.: *Homogeneous linear predictor models for contingency tables*, *J. Amer. Statist. Assoc.* **100** (2005), 121–134.
- [9] LIANG, K. Y.—ZEGER, S. L.: *Longitudinal data-analysis using generalized linear-models*, *Biometrika* **73** (1986), 13–22.
- [10] MCCULLAGH, P.—NELDER, J. A.: *Generalized Linear Models*, (2nd ed.), Chapman and Hall, New York, 1989.
- [11] PREISSER, J. S.—QAQISH, B. F.: *Deletion diagnostics for generalised estimating equations*, *Biometrika* **83** (1996), 551–562.
- [12] PEIXOTO, J. L.—LAMOTTE, L. R.: *Simultaneous identification of outliers and predictors using variable selection techniques*, *J. Statist. Plann. Inference* **23** (1989), 327–343.
- [13] PREGIBON, D.: *Logistic-regression diagnostics*, *Ann. Statist.* **9** (1981), 705–724.
- [14] RICHERT, B. T.—TOKACH, M. D.—GOODBAND, R. D.—NELSEN, J. L.: *Integrated swine systems: “the animal component”—phase 1; the Kansas State University survey, swine day 1993*. Technical report, Kansas State University, 1993.
- [15] SEARLE, S. R.: *Matrix Algebra Useful for Statistics*. Wiley, New York, 1982.
- [16] WEDDERBURN, R. W. M.: *Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method*, *Biometrika* **61** (1974), 439–447.
- [17] WILLIAMS, D. A.: *Generalized linear-model diagnostics using the deviance and single case deletions*, *J. Roy. Statist. Soc. Ser. C* **36** (1987), 181–191.

Received September 28, 2006

*Department of Mathematics,
Statistics and Computer Science
Victoria University
P. O. Box 600
Wellington 6140
NEW ZEALAND
E-mail: thomas.suesse@mcs.vuw.ac.nz*