Computational Statistics and Data Analysis 53 (2008) 222–232

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Model diagnostic plots for repeated measures data using the generalized estimating equations approach

Sohee Oh [a], K.C. Carriere [b], Taesung Park [a,*]

[a] Department of Statistics, Seoul National University, Seoul, Republic of Korea
[b] University of Alberta, Edmonton, Canada

ARTICLE INFO

Article history:
Received 21 October 2006
Received in revised form 25 March 2008
Accepted 12 July 2008
Available online 22 July 2008

ABSTRACT

The generalized estimating equations (GEE) approach has been widely used to analyze repeated measures data. However, in the absence of likelihood ratio tests, model diagnostic checking tools are not well established for the GEE approach, whereas they are for other likelihood-based approaches. Diagnostic checking tools are essential for determining a model's goodness of fit, especially for non-normal data. In this paper, we propose simple residual plots to investigate the goodness of fit of the model based on the GEE approach for discrete data. The proposed residual plots are based on the quantile–quantile (Q–Q) plots of a χ^2-distribution, and are particularly useful for comparing several models simultaneously.

1. Introduction

Repeated measures data arise when outcomes are observed repeatedly on each experimental subject at several time points. Liang and Zeger (1986) and Zeger and Liang (1986) presented the generalized estimating equations (GEE) approach to analyzing repeated measures data, which is an extension of the quasi-likelihood and marginal model approach. It requires the researcher to specify only the model for the marginal mean and a working covariance matrix for the vector of repeated measurements from each subject. This approach can be used for both discrete and continuous outcome variables.

The GEE methodology consists of two estimation steps: first, the quasi-likelihood solution for estimators of regression parameters, and second, a robust moment-based method for estimators of correlation parameters, based on Pearson residuals. These steps are repeated until convergence is achieved.

As with any model, the GEE approach also needs diagnostic procedures for checking the model's adequacy, and for detecting outliers and influential observations. However, diagnostic tools such as the likelihood ratio test for generalized linear models are not available for the GEE analysis. Recently, several methods have been developed for assessing the goodness of fit of the models for the GEE approach. Barnhart and Williamson (1998) proposed model-based and robust goodness of fit tests for the GEE method with binary responses by partitioning the covariate space into distinct regions and forming score statistics. Horton et al. (1999) proposed tests for repeat binary responses using predicted deciles of risk. Although easily interpretable, the approach of Horton et al. may not detect important deviations from the fit, and can only test covariates that are in the model. Chang (2000) proposed using the non-parametric Wald–Wolfowitz run test to check the conventional residuals plots for model diagnoses. Lee and Qaqish (2004) presented an approach to repeated binary outcomes based on a modified GEE approach that replaces the working covariance matrix with one based on the multinomial distribution. The test statistic is easy to compute and has a simple reference distribution.

Park and Lee (2004) proposed simple residual plots to investigate the goodness of model fit for normally distributed repeated measures data. The proposed residual plots are based on the quantile–quantile (Q–Q) plots of a χ^2-distribution. In

* Corresponding author. Tel.: +82 2 880 8924; fax: +82 2 883 6144.
E-mail address: tspark@stats.snu.ac.kr (T. Park).

0167-9473/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2008.07.022

particular, the proposed model is useful in comparing several models simultaneously. In this paper, we extend Park and Lee's residual plots to investigate the goodness of fit of the GEE approach. In the GEE approach, Pearson residuals are commonly used to estimate correlation parameters. However, the normality assumption of the Pearson residuals does not necessarily hold for discrete responses. Park et al. (1998) investigated Pearson, Anscombe, and deviance residuals for Poisson and binary responses. They showed that the distribution of Anscombe and deviance residuals is closer to the normal distribution than Pearson residuals. Using these residuals, we propose quantile–quantile (Q–Q) plots for discrete repeated measures data analyses. Q–Q plots use the χ^2-distribution.

This paper is organized as follows. In Section 2, we describe the GEE approach and introduce three types of residuals (Pearson, Anscombe, and deviance). Q–Q plots are then constructed based on the χ^2-distribution. Simulation results are presented in Section 3, followed by two examples illustrating the proposed plots in Section 4. Finally, Section 5 provides a summary and discussion.

2. The GEE method and residual plots

2.1. The GEE approach

Let t be the maximum number of time points at which data are collected. Let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{it_i})^{\mathrm{T}}$ be the $t_i \times 1$ vector of responses and $\boldsymbol{x}_i = (x_{i1}^{\mathrm{T}}, \ldots, x_{it_i}{}^{\mathrm{T}})^{\mathrm{T}}$ be the $t_i \times p$ matrix of covariate values for the ith subject ($i = 1, \ldots, n$). Here, t_i is the number of time points for the ith subject and may be less than t because of missed time points. The marginal density of y_{ij} is assumed to be a member of a generalized exponential family with a scale parameter ϕ, mean μ_{ij}, variance function V_{ij}, and $\eta_i = x_i\boldsymbol{\beta}$, a function of the covariate vector and covariates, where $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{it_i})^{\mathrm{T}}$ with $\eta_{ij} = g(\mu_{ij})$ and $\eta_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\beta}$. Here, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ is the $p \times 1$ vector of unknown parameters to be estimated.

Liang and Zeger (1986) considered the GEE approach to estimating regression parameters by solving the equation:

$$\sum_{i=1}^{n} \mathbf{D}_i^{\mathrm{T}} \mathbf{V}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) = 0,$$

with $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{it_i})^{\mathrm{T}}$, $\mathbf{D}_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}$ and $\mathbf{V}_i$ is a $t_i \times t_i$ covariance matrix of $\boldsymbol{y}_i$. Then they wrote $\mathbf{V}_i = \phi(\mathbf{A}_i)^{1/2}\mathbf{R}_i(\boldsymbol{\alpha})(\mathbf{A}_i)^{1/2}$, where $\mathbf{A}_i = \mathrm{diag}(v_{ij}/\phi)$ and $\mathbf{R}_i(\alpha)$ is a working correlation matrix for each subgroup repeated outcomes.

Liang and Zeger (1986) proposed a robust estimator of variance for $\hat{\boldsymbol{\beta}}$, which was shown to be consistent even when $\mathbf{R}_i(\alpha)$ is misspecified. The covariance matrix $\mathbf{V}_\beta$ is estimated by

$$\hat{\mathbf{V}}_{\boldsymbol{\beta}} = n \left(\sum_{i=1}^{n} \mathbf{D}_i^{\mathrm{T}} \mathbf{V}_i^{-1} \mathbf{D}_i \right)^{-1} \left\{ \sum_{i=1}^{n} \mathbf{D}_i^{\mathrm{T}} \mathbf{V}_i^{-1} \mathrm{cov}(\boldsymbol{y}_i) \mathbf{V}_i^{-1} \mathbf{D}_i \right\} \left(\sum_{i=1}^{n} \mathbf{D}_i^{\mathrm{T}} \mathbf{V}_i^{-1} \mathbf{D}_i \right)^{-1}. \tag{1}$$

For a given estimate for $\boldsymbol{\beta}$, the parameters $\boldsymbol{\alpha}$ and ϕ can be estimated using the current residuals.

First, the Pearson residual is defined by

$$e_{ij}^{P} = \frac{y_{ij} - \hat{\mu}_{ij}}{\{[\hat{V}_i]_{jj}\}^{1/2}}, \tag{2}$$

where $[\hat{V}_i]_{jj}$ is the jth diagonal element of $\hat{V}_i$. This is a simple residual scaled by the estimated standard deviation of y_{ij}.

Two other residuals, Anscombe and deviance, are modified versions of Pearson residuals.

Anscombe (1953) proposed a residual using the function $G(y)$ in place of y, where $G(\cdot)$ is chosen to make the distribution of $G(y)$ as normal as possible. For univariate generalized linear models, the function $G(\cdot)$ is given by

$$G(\cdot) = \int \frac{1}{V^{1/3}(\mu)} \mathrm{d}\mu. \tag{3}$$

For Poisson outcomes after stabilizing the square root of the variance of $G(\cdot)$, the Anscombe residual, to be denoted by e^A, is given by

$$e_{ij}^{A} = \frac{\frac{3}{2}(y_{ij}^{2/3} - \hat{\mu}_{ij}^{2/3})}{\hat{\mu}_{ij}^{1/6}}. \tag{4}$$

The Anscombe residual for binary outcomes is given by

$$e_{ij}^{A} = \frac{A(y_{ij}) - A(\hat{y}_{ij})}{\sqrt{\mathrm{Var}(A(y_{ij}))}}, \tag{5}$$

where $A(u) = \int_0^u t^{-1/3}(1-t)^{-1/3}\mathrm{d}t$, for $0 \le u \le 1$. $A(u)$ can be computed using the incomplete beta function $I_u(\frac{2}{3}, \frac{2}{3})$, which is symmetric about $u = 0.5$; multiplication by $B(\frac{2}{3}, \frac{2}{3}) = 2.0533$ gives the value of $A(u)$.

In univariate generalized linear models, the deviance is often used as a measure of discrepancy. The ith observation contributes a quantity e_i to the deviance D so that $\sum e_i = D$. The deviance residual is defined to be

$$e_{ij}^D = \text{sign}(y_{ij} - \hat{\mu}_{ij})\sqrt{d_i}, \tag{6}$$

which increases (or decreases) by the sign of $(y - \mu)$. For the Poisson outcomes, the deviance residual is defined by

$$e_{ij}^D = \text{sign}(y_{ij} - \hat{\mu}_{ij})\{2(y_{ij}\log(y_{ij}/\hat{\mu}_{ij}) - (y_{ij} - \hat{\mu}_{ij}))\}^{1/2}.$$

The deviance residual for binary outcomes is given by

$$e_{ij}^D = \begin{cases} -\sqrt{2|\log(1-\hat{\mu}_{ij})|} & \text{if } y_{ij} = 0, \\ \sqrt{2|\log(\hat{\mu}_{ij})|} & \text{if } y_{ij} = 1. \end{cases}$$

2.2. Residual plots

Park et al. (1998) investigated Pearson, Anscombe, and deviance residuals for Poisson and binary responses with arbitrary covariance structures. They showed that there are no substantive distinctions among the three residuals and concluded that there is no compelling reason to consider using residuals other than Pearson residuals. However, they noted that using Anscombe and deviance residuals produces distributions that are closer to normal. The distribution of Pearson residuals for non-normal distributions is often markedly skewed, and may fail to have properties similar to those of normal residuals. Although the Anscombe and deviance residuals appear to have very different functional forms for non-normal distributions, the values they take for given y and μ are often remarkably similar. In this paper, we further investigate these residuals for the purpose of comparing models.

In univariate regression models, if a model fits the data well, the residuals are expected to follow the normal distribution. If the assumed mean and covariance structures are correct, then the residual vector $\boldsymbol{e}$ is also expected to be normally distributed with a zero mean vector. Using this idea, Park and Lee (2004) proposed simple residual plots to check the goodness of fit. We extend Park and Lee's residual plots to investigate the goodness of fit of the GEE approach.

Using the same notation of Park and Lee (2004), let $\boldsymbol{e}_i$ be the the $t_i \times 1$ vector of the ith subject residual vector. Let $\boldsymbol{W}_i$ be the variance matrix of $\boldsymbol{e}_i$, which is a block diagonal submatrix of $\boldsymbol{W}$. When $\boldsymbol{W}_i$ is known and the mean model is correct, $q_i = \boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{W}_i^{-1}\boldsymbol{e}_i$, $i = 1, \ldots, n$, are approximately distributed as the χ^2-distribution with t_i degrees of freedom. That is,

$$q_i = \boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{W}_i^{-1}\boldsymbol{e}_i \sim \chi^2(t_i). \tag{7}$$

When the number of responses from the same subject is equal to t for all i, we can construct a Q–Q plot easily with observed q_is using the χ^2-distribution. Let $q_{(1)} \le \cdots \le q_{(n)}$ be ordered values of q_is. Then, $q_{(i)}$ is in fact the empirical $100 \times i/n$ percentile, and from the χ^2-distribution with t degrees of freedom, we can obtain the corresponding quantiles $\psi_{(1)} \le \cdots \le \psi_{(n)}$. Then, the Q–Q plot is the graph of $(q_{(i)}, \psi_{(i)})$ from which we can investigate the model fit and identify outliers.

When the number of responses from the same subject t_i differ from subject to subject due to unbalanced or incomplete observations, the degrees of freedom q_i differ, and it is not possible to construct a Q–Q plot based on the χ^2-distribution. In that case, we suggest using the Q–Q plot with the following simple fourth root transformation proposed by Hawkins and Wixley (1986) to achieve approximate normality, so that

$$q_i^N = \frac{q_i^{1/4} - (t_i - .5)^{1/4}}{(8\sqrt{t_i})^{-1/2}}, \quad i = 1, \ldots, n, \tag{8}$$

is approximately distributed as the standard normal distribution $N(0, 1)$. Here, the mean and variance are approximated using the Taylor expansion of the exact expression of the logs of the first and second moments. Park and Lee (2004) showed that this transformation works well for normally distributed repeated measures. We call q_i^N the *normalized residual* for the ith subject. Similarly, we can construct a Q–Q plot with observed q_i^Ns using the standard normal distribution.

3. Simulation studies

3.1. Poisson responses

To investigate the performance of the residual Q–Q plots in discrete responses, we performed simulation studies. First, we considered a two-sample configuration with correlated Poisson responses with $t = 4$.

For Poisson responses, the true marginal model is given by

$$\log(\mu_{ij}) = \beta_0 + \beta_1 x_i + \beta_2 j + \beta_3 x_i j,$$

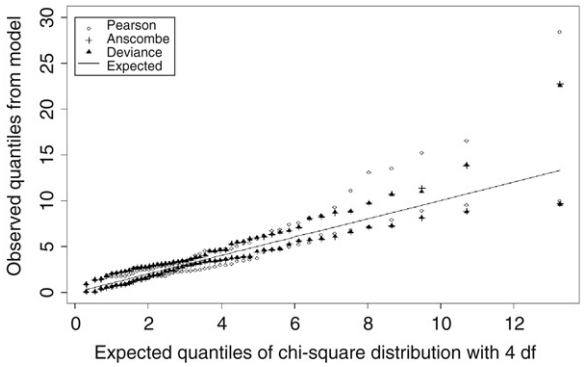


Fig. 1. Poisson. Min–max band plots for Pearson, Anscombe, and deviance residuals based on simulated 100 Q–Q plots. The plots are based on Model 4 with a total sample size of 50.

where x_i is a group variable having value 0(1) in group 1(2) and j represents the time covariate. In each group, we used the same sample size of 25. We considered three structures of **R**: (1) exchangeable correlation with $\rho = 0.25, 0.5, 0.75$, (2) AR-1 with $\rho = 0.25, 0.5, 0.75$, and (3) unstructured correlations. However, since the correlation structures did not have much effect on the performance of residual plots, we present a simple case of the exchangeable structure for $\rho = 0.5$.

We considered the following four models in our simulation:

Model 1: $\log(\mu_{ij}) = \beta_0$,
Model 2: $\log(\mu_{ij}) = \beta_0 + \beta_1 x_i$,
Model 3: $\log(\mu_{ij}) = \beta_0 + \beta_1 x_i + \beta_2 j$, and
Model 4: $\log(\mu_{ij}) = \beta_0 + \beta_1 x_i + \beta_2 j + \beta_3 x_i j$.

The true parameter values were chosen as $(\beta_0, \beta_1, \beta_2, \beta_3) = (0.1, -0.5, 0.6, -0.4)$. For simplicity, $\phi = 1$ is assumed.

In the GEE framework, there is currently no formal diagnostic tool available to compare the adequacy of models such as the likelihood ratio tests (LRT) for generalized linear models. For the simulated dataset, we considered the proposed residual plots. Using this approach, a model that satisfies the expected quantiles can be judged as the best model. This can be used for non-nested model comparisons as well. In our simulations, under Model 4, a total of 100 datasets were generated for correlated Poisson responses.

Figs. 1 and 2 show the Q–Q plots. Fig. 1 shows the ranges of qs for Pearson, Anscombe, and deviance residuals based on simulated 100 Q–Q plots. The X-axis is the expected quantiles from the χ^2-distribution with four degrees of freedom, and the Y-axis is the observed quantiles of $q_1, \ldots, q_{50}$ obtained from each model. For each x value, there are two y values: the maximum and minimum of 100 simulated quantiles. In this figure, the Pearson residual bands are highly skewed, demonstrating that at this sample size Pearson residuals do not achieve asymptotic normality. On the other hand, the Anscombe and deviance residuals seem to follow the χ^2-distribution quite well.

To examine how well the plots can distinguish models, we drew the Q–Q plots for the mean of 100 simulated quantiles (Fig. 2). The first graph illustrates plots for the Pearson residuals, the second Anscombe residuals, and the third deviance residuals. Each plot contains the Q–Q plots from the four models: Models 1, 2, 3, and 4. For the Pearson residuals, all models show some systematic departures from the $Y = X$ line. Even the true Model 4 has a few observations far from the $Y = X$ line, indicating that the qs from the Pearson residuals do not seem to follow the χ^2-distribution for this sample size. However, the Q–Q plots still show that Model 4 is the best model. On the other hand, the Anscombe and deviance residuals show that Models 3 and 4 are quite close to the $Y = X$ line, while Models 1 and 2 are far from the $Y = X$ line. For Model 4, the observed Anscombe and deviance quantiles are consistent with the expected quantiles of the χ^2-distribution, suggesting that they follow the asymptotic chi-square distribution for the true model.

3.2. Binary responses

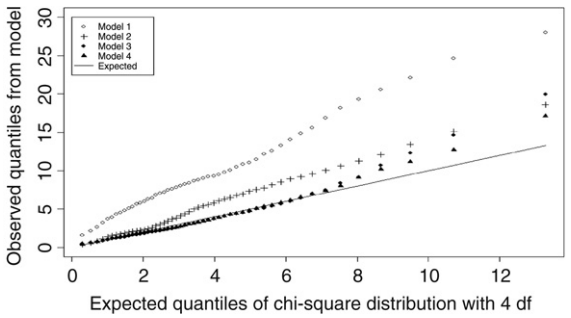
We also considered similar models for the correlated binary responses. The true marginal model is given by

$$\text{logit}(\mu_{ij}) = \beta_0 + \beta_1 x_i + \beta_2 j + \beta_3 x_i j,$$

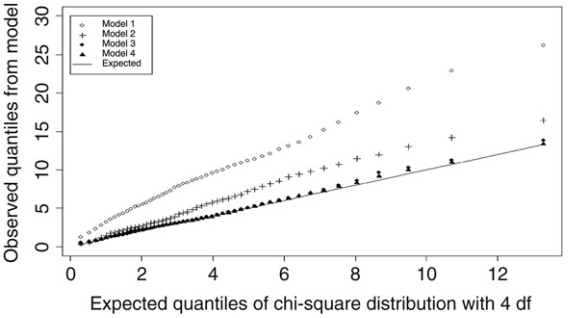
where x_i is a group variable having two values and j represents the time covariate. Like Poisson responses, we considered four models.

Model 1: $\text{logit}(\mu_{ij}) = \beta_0$,
Model 2: $\text{logit}(\mu_{ij}) = \beta_0 + \beta_1 x_i$,
Model 3: $\text{logit}(\mu_{ij}) = \beta_0 + \beta_1 x_i + \beta_2 j$, and
Model 4: $\text{logit}(\mu_{ij}) = \beta_0 + \beta_1 x_i + \beta_2 j + \beta_3 x_i j$,

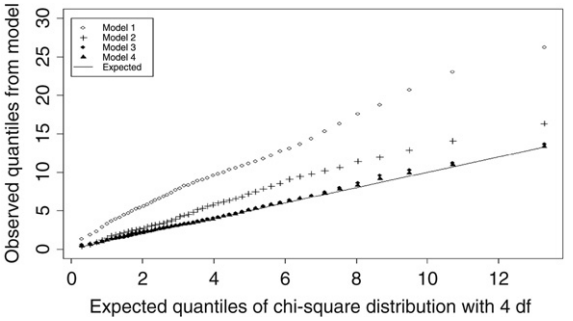
where $\text{logit}(\mu_{ij}) = \log(\mu_{ij}/(1 - \mu_{ij}))$.



(a) Pearson residuals.



(b) Anscombe residuals.



(c) Deviance residuals.

Fig. 2. Poisson. Q–Q plots for the mean of 100 simulated quantiles. Model 4 is the true model.

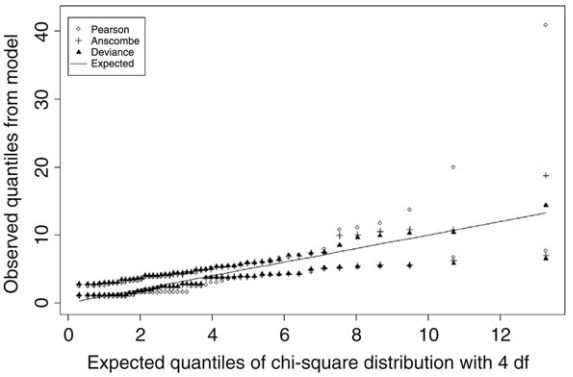


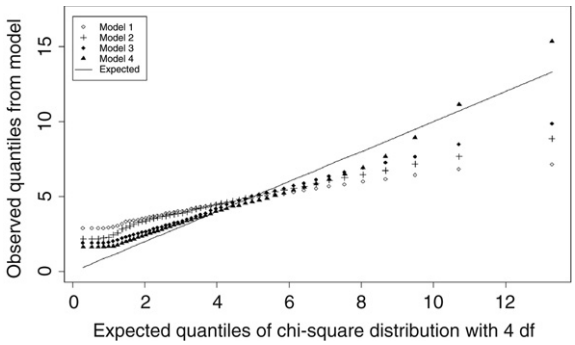
Fig. 3. Binary. Min–max band plots for Pearson, Anscombe, and deviance residuals based on simulated 100 Q–Q plots. The plots are based on Model 4 with a total sample size of 50.

The true parameter values were chosen as $(\beta_0, \beta_1, \beta_2, \beta_3) = (-2, 1, 1, -1)$. We also assumed scale parameter $\phi = 1$ and tried several different correlation structures. Since the correlation structures did not have much effect on the performance of residual plots, we present a simple case of the exchangeable structure for $\rho = 0.5$. Under Model 4, a total of 100 datasets were generated for correlated binary measures.

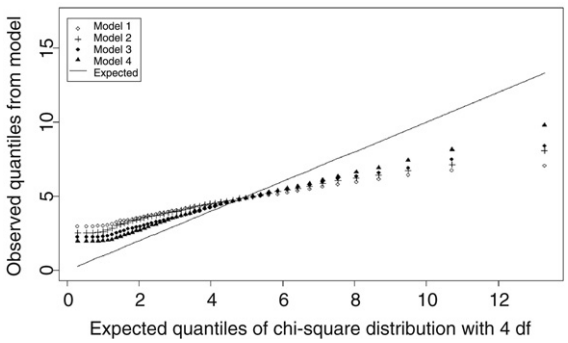
Fig. 3 shows the ranges of qs for Pearson, Anscombe, and deviance residuals based on 100 simulated datasets. Pearson residuals for binary responses seem to be more skewed and have much wider ranges than Anscombe and deviance residuals.

To examine how well the plots can distinguish models, we drew the Q–Q plots for the mean of 100 simulated quantiles (Fig. 4). The first graph illustrates plots for Pearson residuals, the second Anscombe residuals, and the third deviance residuals. Each graph contains the Q–Q plots from the four models: Models 1, 2, 3, and 4. The Q–Q plot of Pearson residuals shows that the true Model 4 is closest to the $Y = X$ line with clear differences among four models. Although the Q–Q plots of Anscombe and deviance residuals show some systematic departures from the $Y = X$ line, they consistently show that Model 4 is the best among the four models.

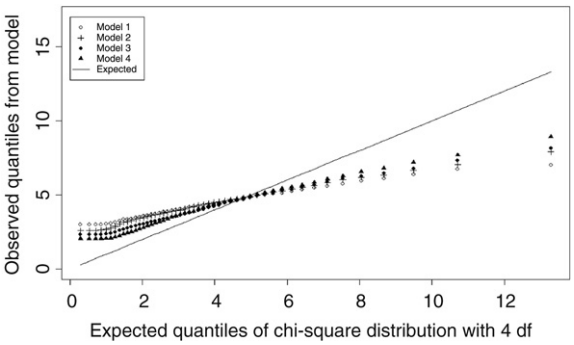
Fig. 5 is the Q–Q plot of Model 4 for the three residuals from one randomly selected dataset showing that the qs from the Pearson residuals tend to follow the χ^2-distribution more closely than the other two residuals. The mean plots in Fig. 4 also



(a) Pearson residuals.



(b) Anscombe residuals.



(c) Deviance residuals.

Fig. 4. Binary. Q–Q plots for the mean of 100 simulated quantiles. Model 4 is the true model.

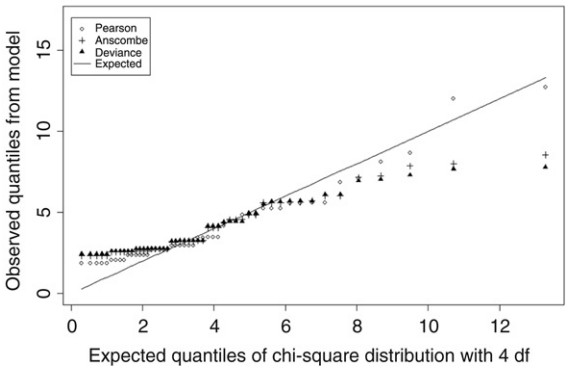


Fig. 5. The Q–Q plot using a randomly selected dataset.

show this tendency. However, Fig. 3 shows that the qs from the Pearson residuals have larger variabilities than the other two residuals. Thus, we conclude as follows. In general, the qs from Pearson residuals tend to follow the χ^2-distribution more closely than other two residuals. However, they are more sensitive to the outliers than the other two residuals.

4. Examples

4.1. Poisson responses

Leppik et al. (1985) conducted a clinical trial involving 59 epileptic subjects. In this study, individuals suffering from simple or complex partial seizures were randomized to receive either the antiepileptic drug progabide (31 subjects) or a placebo (28 subjects). At each of four successive postrandomization clinical visits, the number of seizures occurring during the previous two weeks was reported. The medical question of interest is whether or not progabide reduces the frequency of epileptic seizures.

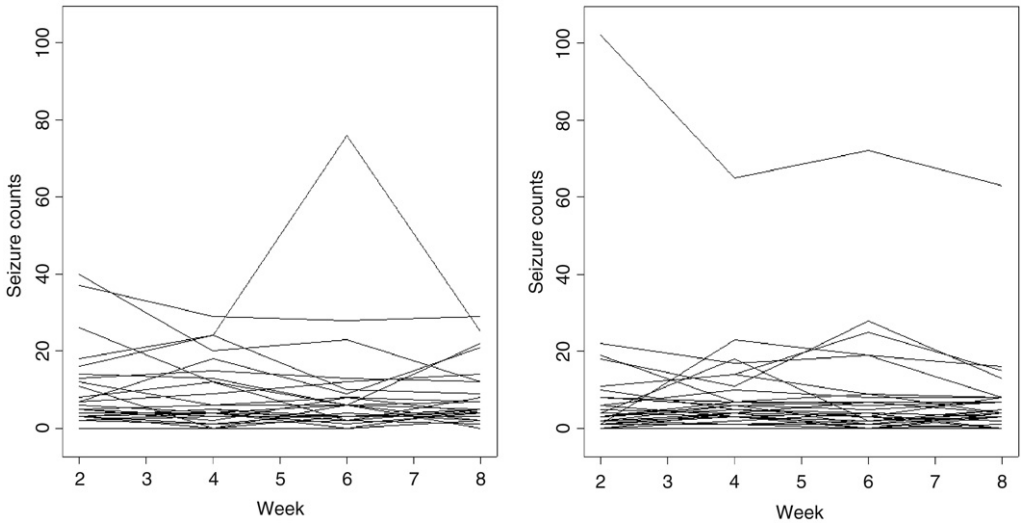


Fig. 6. Individual plots for seizure data for placebo (left) and progabide (right) groups.

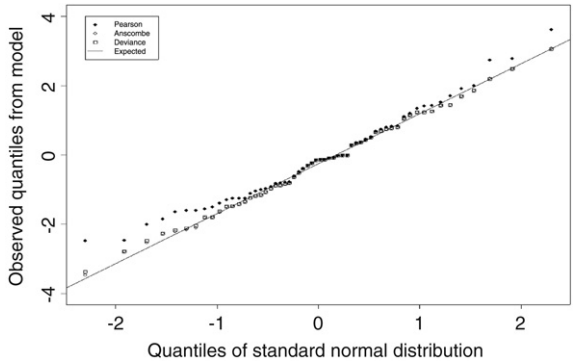


Fig. 7. Normal Q–Q plot for epileptic seizure data at time = 8 for model 6.

Fig. 6 shows the individual plots. There are some outlying observations. Since in the progabide group one subject had high responses that might falsely increase the treatment effects, we excluded this subject. Let y_{ij} denote the log-transformed response (seizure) for the ith subject ($i = 1, \ldots, 58$) at time j ($j = 2, 4, 6, 8$). The covariates in the models are baseline seizure rate (*Base*, computed as the logarithm of $\frac{1}{4}$ the eight-week prerandomization seizure count), logarithm of age in years (*Age*), the binary indicator for the progabide group (*Trt*), and postrandomization clinic visit (*Visit*).

We have considered the following six models:

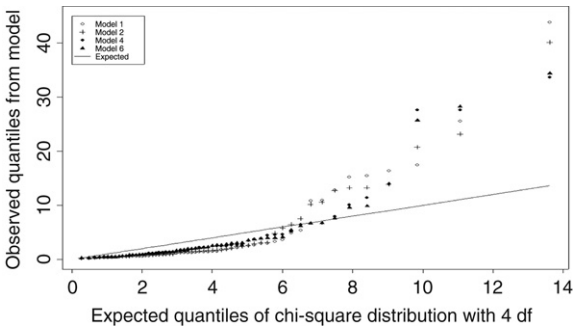
Model 1: $\log(\mu_{ij}) = \beta_0$,
Model 2: $\log(\mu_{ij}) = \beta_0 + \beta_1 \cdot Trt$,
Model 3: $\log(\mu_{ij}) = \beta_0 + \beta_1 \cdot Trt + \beta_2 \cdot Age$,
Model 4: $\log(\mu_{ij}) = \beta_0 + \beta_1 \cdot Trt + \beta_2 \cdot Age + \beta_3 \cdot Base$,
Model 5: $\log(\mu_{ij}) = \beta_0 + \beta_1 \cdot Trt + \beta_2 \cdot Age + \beta_3 \cdot Base + \beta_4 \cdot Base \cdot Trt$, and
Model 6: $\log(\mu_{ij}) = \beta_0 + \beta_1 \cdot Trt + \beta_2 \cdot Age + \beta_3 \cdot Base + \beta_4 \cdot Base \cdot Trt + \beta_5 \cdot Visit$.

Thall and Vail (1990) considered Model 6 the best. Models 1–5 were defined by removing the covariates sequentially from Model 6 in order to compare the goodness of fit. The model fitting results are summarized in Table 1. To investigate how close the distribution of residuals is to the normal distribution, we drew the normal Q–Q plots for each time point. For example, Fig. 7 shows the normal Q–Q plots of Model 6 for the three residuals at time 8. Anscombe and deviance quantiles are consistent with the normal quantile. However, Pearson quantiles show some systematic departures from normal quantiles.

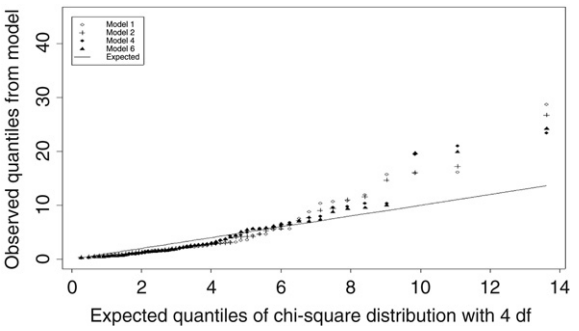
Fig. 8 shows the Q–Q plots based on the χ^2-distribution. The X-axis is the expected quantiles from the χ^2-distribution with four degrees of freedom, and the Y-axis is the observed quantiles of $q_1, \ldots, q_{58}$ obtained from the underlying models. We do not show the plots of Models 3 and 5 because the plot of Model 3 is similar to that of Model 2, and Model 5 is similar to Model 6 except for the three largest observations. Fig. 8(a) shows the Q–Q plot for Pearson residuals for Models 1, 2, 4, and 6. All models depart far from the $Y = X$ line.

Table 1
Epileptic seizure data model fits: Estimates, standard errors, and *P*-values

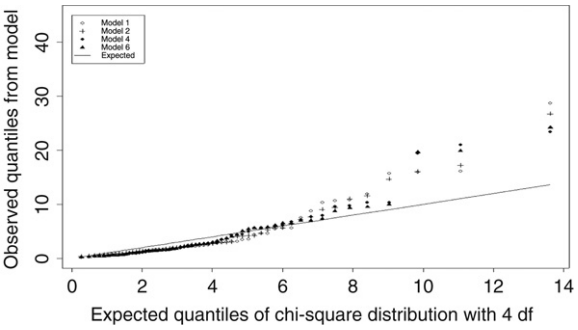
Parameter	Model 1			Model 2			Model 3			Model 4			Model 5			Model 6		
	Esti.	S.E.	*p*-value	Esti.	S.E.	*p*-value	Esti.	S.E.	*p*-value	Esti.	S.E.	*p*-value	Esti.	S.E.	*p*-value	Esti.	S.E.	*p*-value
Intercept	1.9629	0.1309	<0.0001	2.1529	0.1892	<0.0001	1.1808	1.6221	0.4666	−2.2338	0.8730	0.0105	−2.3369	0.8681	0.0071	−2.2707	0.8686	0.0089
Treatment				−0.4074	0.2456	0.0971	−0.3951	0.2405	0.1004	−0.2311	0.1598	0.1481	−0.5028	0.4082	0.2181	−0.5075	0.4084	0.2140
Age							0.2900	0.4988	0.5610	0.7086	0.2469	0.0041	0.7629	0.2525	0.0025	0.7748	0.2535	0.0022
Baseline										0.9828	0.0833	<0.0001	0.9482	0.0965	<0.0001	0.9472	0.0972	<0.0001
Baseline*Treatment													0.1294	0.1913	0.4986	0.1321	0.1905	0.4878
Visit																−0.0211	0.0188	0.2617



(a) Pearson residuals.



(b) Anscombe residuals.



(c) Deviance residuals.

Fig. 8. Q–Q plots for the epileptic seizure data.

Fig. 8(b) and (c) are the Q–Q plots for the Anscombe and deviance residuals, respectively. These Q–Q plots are much closer to the $Y = X$ line. In particular, Model 6 shows that the observed quantiles agree with the expected quantiles except for the three largest observations. Model 4 shows some slight departures from the $Y = X$ line, while Models 1 and 2 show systematic departures.

4.2. Binary responses

Koch et al. (1977) presented a longitudinal study comparing a new drug with a standard drug in the treatment of 340 subjects suffering mental depression. Subjects were diagnosed as "mild" or "severe". In each group, subjects were randomly assigned to one of the two drugs. Following 1, 2, and 4 weeks of treatment, each subject was classified as normal or abnormal in terms of suffering from mental depression at each follow-up. In this study, we used the following covariates: diagnosis, treatment, and time ($j = 1, 2, 4$). Koch et al. (1977) used time $(0, 1, 2)$, the logs to base 2 of week numbers $(1, 2, 4)$ for time as covariates. This dataset was also analyzed by Barnhart and Williamson (1998) and Agresti (2002). In our analysis, we have considered the following four models:

Model 1: $\text{logit}(\mu_{ij}) = \beta_0$,
Model 2: $\text{logit}(\mu_{ij}) = \beta_0 + \beta_1 \cdot \mathit{Trt}$,
Model 3: $\text{logit}(\mu_{ij}) = \beta_0 + \beta_1 \cdot \mathit{Trt} + \beta_2 \cdot \mathit{Diagnosis}$, and
Model 4: $\text{logit}(\mu_{ij}) = \beta_0 + \beta_1 \cdot \mathit{Trt} + \beta_2 \cdot \mathit{Diagnosis} + \beta_3 \cdot \mathit{Time}$.

Models from 1 to 4 were defined by removing the covariates sequentially from Model 4 in order to compare the goodness of fit. We used an unstructured correlation structure. The model-fitting results are summarized in Table 2. To investigate how close the distribution of residuals is to the χ^2-distribution, we drew χ^2 Q–Q plots. Fig. 9 shows the Q–Q plots of Pearson, Anscombe, and deviance residuals in Model 4. In this data analysis, Anscombe and deviance quantiles locate far below the $Y = X$ line.

Fig. 10 shows the Q–Q plots based on the χ^2-distribution. The X-axis is the expected quantiles from the χ^2-distribution with three degrees of freedom, and the Y-axis is the observed quantiles of $q_1, \ldots, q_{340}$ obtained from the underlying models. Each Q–Q plot contains four models. Fig. 10(a) shows the Q–Q plot for Pearson residuals. Model 4 is closest to the $Y = X$ line. Fig. 10(b) and (c) are the Q–Q plots for Anscombe and deviance residuals, respectively. Although these Q–Q plots are much more apart from the $Y = X$ line than those of the Pearson residuals, they consistently show that the observed quantiles of Model 4 are much closer to the $Y = X$ line than for other models.

Table 2
Depression data model fits: Estimates, standard errors, and P-values

Parameter	Model 1			Model 2			Model 3			Model 4		
	Esti.	S.E.	p-value	Esti.	S.E.	p-value	Esti.	S.E.	p-value	Esti.	S.E.	p-value
Intercept	0.1116	0.0647	0.0846	−0.2196	0.0899	0.0145	0.3864	0.1118	0.0005	−0.4989	0.1566	0.0014
Treatment				0.7405	0.1269	<0.0001	0.8001	0.1225	<0.0001	0.9200	0.1384	<0.0001
Diagnosis							−1.1254	0.1269	<0.0001	−1.2780	0.1436	<0.0001
Time										0.9002	0.0931	<0.0001

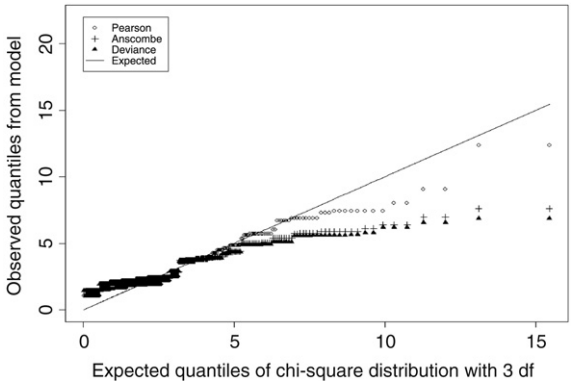
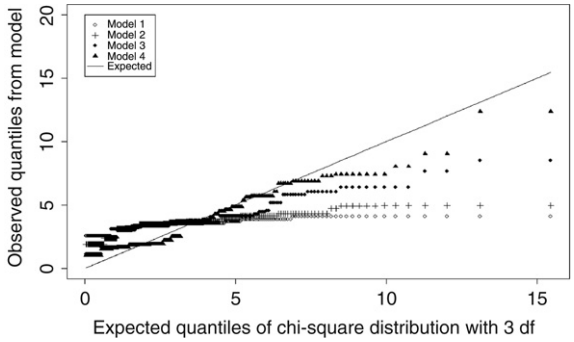
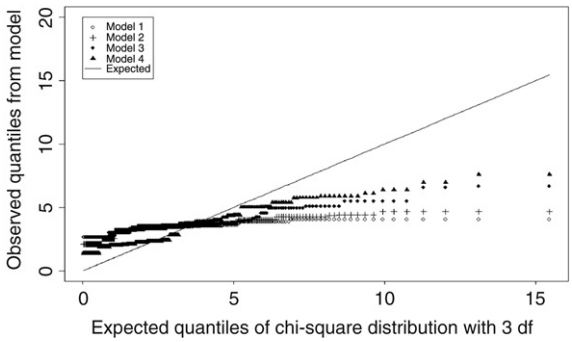


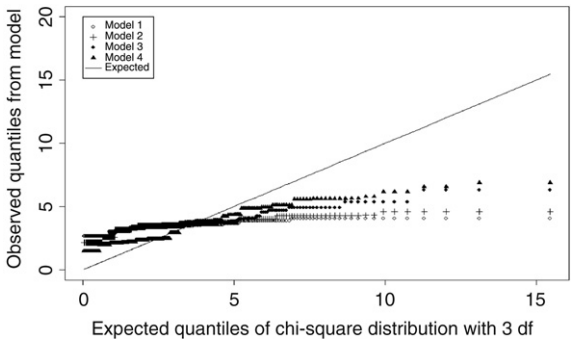
Fig. 9. Q–Q plots for the depression data for model 4.



(a) Pearson residuals.



(b) Anscombe residuals.



(c) Deviance residuals.

Fig. 10. Q–Q plots for the depression data.

In summary, the proposed Q–Q plots are effective in comparing several models simultaneously. Pearson residuals seem to provide a closer approximation of the χ^2-distribution than the other two residuals. However, all three residuals provided very consistent results. Interesting, Anscombe and deviance residuals are quite similar. Finally, though not presented here, we also examined the Q–Q plots for different correlation structures with the same mean model, and found the Q–Q plots are

less sensitive to correlation structures. This is mainly because the estimated regression parameters are consistent regardless of correlation structures.

5. Discussion

In this paper, we focused on diagnostic plots for the discrete repeated measures data based on the GEE methodology. Like the multivariate regression models used to analyze repeated measures data, the GEE approach suffers from outliers and influential observations. Thus, it is important to have useful tools for checking the adequacy of models and detecting outliers and influential observations. The proposed residual plots use the Q–Q plots of the χ^2-distribution and the standard normal distribution. Residuals from the same subject form multivariate data, and our approach uses a univariate summary measure of these residuals.

We compared the Q–Q plots of three types of residuals through simulation studies. For Poisson responses, we showed that Anscombe and deviance residuals tend to follow a χ^2-distribution when the model fits the data well. For binary responses, Pearson residuals provide a closer χ^2-distribution when the model fits data well. Application to a dataset shows that the proposed plots are quite effective in detecting outliers as well as in determining the adequacy of model fit. In particular, the proposed method visually distinguishes good fitting models from poor ones. That is, the proposed Q–Q plots are especially useful in comparing several different models simultaneously, nested or non-nested.

The proposed Q–Q plots can easily be applied where either the sample size or the number of repeated observations is large. It is because the repeated observations from the same subject are summarized by one point in the Q–Q plot, regardless of the dimension of the time points of measurements. As a result, the proposed plot can effectively compare several models for their goodness of fit.

From the proposed Q–Q plot, we can easily tell which model is better than another in the same Q–Q plot. However, a more formal test procedure is required to determine whether a specific model is suitable, and whether a specific observation is an outlier.

This paper focused on mean model diagnostics. That is, we have used the same covariance structure for repeated discrete responses to compare different mean models. Although not reported here, we have investigated the Q–Q plots for a poor choice of $\mathbf{V}$ for the same mean model. It turns out that the Q–Q plots are less sensitive to the covariance structures employed, mainly because the estimated mean parameters are unbiased as long as a consistent estimator of $\mathbf{V}$ is used.

Acknowledgments

We thank the associate editor for many helpful comments and suggestions that improved the manuscript greatly. This research was supported by the Korean Research Foundation (KRF-2004-015-C0086).

References

Agresti, A., 2002. Categorical Data Analysis, 2nd ed. Wiley.
Anscombe, F.J., 1953. Contribution to the discussion of H. Hotelling's paper. J. Roy. Statist. Soc. B 15, 229–230.
Barnhart, H.X., Williamson, J.M., 1998. Goodness-of-fit tests for GEE modeling with binary responses. Biometrics 54, 720–729.
Chang, Y.C., 2000. Residuals analysis of the generalized linear models for longitudina data. Stat. Med. 19, 1277–1296.
Hawkins, D.M., Wixley, R.A.J., 1986. A note of the transformation of Chi-squared variables to normality. Amer. Statist. 40, 296–298.
Horton, N.J., Bebchuk, J., Jones, C.L., Lipsitz, S.R., Catalano, P.J., Zahner, G.E.P., Fitzmaurice, G.M., 1999. Goodness-of-fit for GEE: An example with mental health service utilization. Stat. Med. 18, 213–222.
Koch, G.G., Landis, J.R., Freeman, J.L., Freeman, D.H., Lehnen, R.G., 1977. A general methodology for the analysis of experiments with repeated measurement of categorical data. Biometrics 33, 133–158.
Lee, J.H., Qaqish, B.F., 2004. Modified GEE and Goodness of the Marginal Fit (GOMF) test with correlated binary responses for contingency tables. Biom. J. 46, 675–686.
Leppik, I.E., et al., 1985. A double-blind crossover evaluation of progabide in partial seizures. Neurology 35, 285.
Liang, K.Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13–22.
Park, T., Davis, C.S., Ning, Li., 1998. Alternative GEE estimation procedures for discrete longitudinal data. Comput. Statist. Data Anal. 28, 243–256.
Park, T., Lee, S.Y., 2004. Model diagnostic plots for repeated measures data. Biom. J. 46, 441–452.
Thall, P.F., Vail, S.C., 1990. Some covariance models for longitudinal count data overdispersion. Biometrics 46, 657–671.
Zeger, S.L., Liang, K.Y., 1986. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42, 121–130.