البواقي ا تطرفة وذج ا دار اللوجستك

الدكتور خليفة ود الزبيدي الدكتور زكي الصراف

كلية الاقتصاد والعلوم الإدارية

جامعة الزرقاء الأهلية

الأردن

الملخص

هنـاك الكـث مـن النمـاذج الـ كـن توفيقهـا للبيانـات الثنائيـة عنـد وجـود مـتغ ات توضـيحية تؤثر متغ الاستجابة. تعد اذج اللوجستك والاحتما واللوغاريتمي التكميلي من أكثر النماذج استخداماً ذا النوع من البيانات.

لقد استخدمنا نموذج اللوجستك في هذا البحث لتوفيق البيانات الثنائية وذلك لكونه أكثر استخداماً وشيوعاً.

وقد استخدمنا عدة طرق تقدير لغرض الحصول على تقديرات معلمات النموذج وهي طريقة الإمكان الأعظم، وتصغير مربع كاي، والمربعات الصغرى الموزونة، وأخيراً طريقة المربعات الصغرى الموزونة المعدلة.

إنَّ الهدف الرئيسي من هذا البحث هو الحصول على البواقي المتطرفة من خلال اعتماد صيغتين: الأولى هي الصيغة الاعتيادية للبواقي، والثانية هي الصيغة المعدلة والمشتقة من قبل (بيرسون). حُصِل على هذه البواقي من خلال توفيق نموذج اللوجستك وتطبيق طرق التقدير الأربع السالفة الذكر.

ولغرض معرفة أي الصيغت تكون أكثر دقة فقد اخت نا دقة القيم ا رجة للبواقي ا تطرفة للصيغت عند مستويات معنوية تلفة، لقد ذلك من خلال الاعتماد على بيانات توليدها باستخدام ا اكاة و جوم عينات تلفة (ثلاثـة حجـوم تلفة) لكل وذج، وعند كل حجم عينة ألف موعة من البيانات، ولكل موعة من موعات البيانات هذه توفيـق وذج اللوجستك وتقدير معلماته باستخدام طرق التقدير الأربع ا ذكورة أعلاه.

تبين نتائج ا اكاة هذا البحث فروقات صغ ة من حيث الدقة فيها إذا استخدمنا صيغة البواقي ا تطرفة مقارنة بصـيغة

يرد هذا البحث باللغة الإنكليزية في الصفحات (٤٧-٧٧).

البواقي ا تطرفة ا عدلة. وكذلك ا ال فإن النتائج أظهرت فارقاً ليس بالكب من حيث دقة مستوى ا عنوية ا سوبة للبواقي ا تطرفة ا عدلة توفـ أ للجهد والوقت.

اشتمل البحث على خمسة أجزاء كانت المقدمة الجزء الأول، في حين تم اشتقاق خواص البواقي في الجزء الثاني. أما الجزء الثالث فتضمن استعراض كيفية الوصول إلى القيم الحرجة التقريبية للبواقي المتطرفة. أما نتائج المحاكاة فقد تضمنها الجزء الرابع وحوى الجزء الخامس والأخير الاستنتاجات والمقترحات.

Extreme Residuals in Logistic Regression Model

Khalifa H. Jaber Zaki J. al-saraf Zarka
Faculty of Economics and Administrative Science
University Zarka
Jordan

Abstract

Many models can be used for *fitting* binary response data when explanatory variables are present.

The most common models are the logistic, probit and complementary log-log model. In this paper we used the logistic model because of its simple interpretation. In order to obtain the parameters estimates of this model we used several methods of estimation. These methods are, the maximum likelihood, minimum chi-square, weighted least squares and modified weighted least squares

The main objective of this paper is to obtain the extreme residuals based on two different ways. The first is the ordinary residuals and the second is the modified Pearson residuals. These two kinds of residuals are obtained through the above mentioned four methods of estimation.

In order to give a clear preference of one of these two methods we examined the accuracy of the approximation to the critical values of these extreme residuals. This was done by using large-scale simulation for three different sample sizes, each sample size was run for one thousand different sets of data. The results show that there is a little gain of accuracy by using the modified residuals compared with the ordinary one. The other important results show that the difference between significance levels when modified residuals are used instead of the ordinary residuals is generally quite small.

This paper includes five sections, the introduction is given in section 1. The approximation to the moments of the Pearson residuals is given in section 2. The approximation to the critical values of the extreme residual statistics is shown in section 3. The simulation results are given in section 4. Finally the conclusion is given in section 5.

1. Introduction

In the statistical analysis of binary response data when explanatory variables are present, models such as the logistic, probit and complementary log –log are commonly used. Interpretation will be in terms of odd ratios and here we shall be restricted. The lot is model in comes with attention to this model.

We suppose that there are k explanatory variabes having g distinct groups of values, which for the ith group are denoted by $x_{i1}, \cdots x_{ik}$. For the ith group we suppose that n_i independent trials are made and let Y_i denote the number of "successes", i=1,…g. if p_i denotes the true probability of success for each trial the ith group, the linear logistic regression model is

$$\log(P_i / Q_i) = X_i^1 \beta, \qquad I=1,\ldots,g \qquad (1.1)$$

where $Q_i = 1 - P_i, Xi = (1, X_{i1},\ldots, X_{ik}) and \beta' = (\beta_0, \beta_1,\ldots\beta_k)$

vector of unknown regression coefficients.

Three methods are commonly used for the estimation of β, namely maximum likelihood (ML), minimum chi- square (MCS) and weighted least aquares (WLS), the last method being sometimes referred to as minimum logit chi- square. For later work it is helpful to consider these methods as members of a class of procedures in which the estimator $\hat{\beta}$ is found by minimizing a function of the from.

$$\phi = \sum_{i=1}^{g} n_1 . \phi_i (p_i , P_i) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1.2)$$

where $p_i = y_i / n_i$ is the sample proportion of successes for the ith group and ϕ_i (p_i, P_i) serves to measure the 'distance' between p_i and P_i. The forms of ϕ_i (p_i, P_i) for the ML, MCS and WLS estimation procedures are

$$\phi_i^{(1)} = -(p_i \log P_i + q_i \log Q_i) \ldots\ldots\ldots\ldots\ldots\ldots (1.3)$$

$$\phi_i^{(2)} = (P_i Q_i)^{-1} (p_i - P_i)^2 \text{.....................(1.4)}$$

$$\phi_i^{(3)} = p_i q_i \{(\log P_i / q_i) - \log(P_i / Q_i)\}^2 \text{...... (1.5)}$$

respectively, where q_i = 1-p_i

The ML and MCS methods both require an iterative solution to estimate β but the WLS procedure yields a non- iterative procedure. Furthermore, in the case of a single explanatory variable there is now considerable evidence with better variance and mean square error properties (Berkson (1955), Al- Sarraf and young (1985)). The WLS method can be applied when p_i =0 or 1 but because the sample logit z_i = log (p_i /q_i) is undefined for these extreme cases, modified logits defined by

$$Z_i^* - \log\left\{(p_i + \frac{1}{2ni})/(q_1 + \frac{1}{2ni})\right\} \text{,i=1,..........g} \qquad (1.6)$$

are sometimes used. We have $E(z_i^*) = x_{i'}\beta + 0(n_i^{-2})$ and an estimate the asymptotic variance of z_i which has very small bias is from gart and zweifel (1967)

$$\frac{(n_i + 1)(n_i + 2)}{n_i^3 (p_i + n_i^{-1})(q_i + n_i^{-1})} = w_i^{*-1} say \text{.....................(1.7)}$$

A modified weighted least squares (MWLS) estimate is then given by the value of B which minimises (Z*B) W* (Z*XB) where z* =(z_1*, …, z_g) and w*= diag ((w_I,…, w_g))

In applications, it is of course important to assess the goodness of fit of the logistic regression model. This is commonly done by computing a over all summary statistic such as the Pearson statistic

$$R = \sum_{i=1}^{g} n_i (p_i - \hat{P}_i)^2 / (\hat{P}_i \hat{Q}_i) \text{...................(1.8)}$$

Where we use $\hat{P}_i = \exp(xi'\beta)/\{1+\exp(x'\beta)\}$ to denote the estimator of p_i under a general estimation procedure within the class defined in (1.2) An alternative summary statistic is the deviance statistic defined by

$$D = 2\sum_{i=1}^{g} n_i \{p_i.\log(p_i/\hat{P}_i) + q_i \log(q_i/\hat{Q}_i)\}$$

although this statistic is only likely to be used if a maximum likelihood fit has been made. The individual group residuals corresponding to these overall statistics are given by

$$R_i = (n_i/\hat{P}_i.\hat{Q}_i)^{\frac{1}{2}}.(p_i - \hat{P}_i), D_i = \pm 2^{\frac{1}{2}} n_i^{\frac{1}{2}} \{p_i \log(p_i/\hat{P}_i) + q_i \log(q_i/\hat{Q}_i)\}^{\frac{1}{2}} \quad (1.10)$$

where ($\pm$)the sign of p_i - P_i, and an assessment of goodness of fit is often based on an inspection of these residuals or of a normal probability plot based on them:

The extreme residuals denoted by

$R_{max}=max_iR_i$,. $R_{min}=min_iR_i$, $R_m=max_i|R_i|$, (1.11)

$D_{max}=max_iD_i$,. $D_{min}=min_iD_i$, $R_m=max_i|D_i|$, *1.12)

are themselves of particular interest and we focus attention on them in this study. A simple and common approach in assessing the extreme residuals appears to be to take them as being approximately distributed as the corresponding extremes in a sample of g independent observations from the N(0,1) distribution. This approach uses only the first order approximations to the mean and variance of the residuals and ignores their correlations, and so can be misleading.

In section 2, second order approximations to the expectations and co-variance matrix of the Pearson residuals are given and used to define modified extreme residual statistics. Approximations to the percentage points of the extreme residual statistics are presented in section 3. Finally, results from a Monte Carlo investigation to assess the adequacy of the approximations are given for the case when there is a single explanatory variable and for various success probability configurations.

2. Approximations to the Moments of the Pearson Residuals

In this section, we first derive approximations correct to $o(\frac{1}{N2})$ for the expectations of the Pearson residuals, where N=$\sum_{i}^{g} n_i$ is the total number of trials. Our approach is very similar to the general approach given by Cox and Snell (1968), but here we assume that g is fixed and the { n_i }are large, whereas in their method neglected terms are 0(g^{-1})

Since $\hat{P}_i = P_i(\hat{\beta})$, we may write

$$R_i = h_i(p_i.\hat{\beta}), \quad \varepsilon_i = h_i(p_i, \beta) \tag{2.1}$$

where

$$h_i(p_i.\hat{\beta}) = \left(\frac{n_i}{P_i Q_i}\right)^{\frac{1}{2}} \left\{ p_i - \frac{\ell^{\chi_i^{/}} \beta}{1 + \ell^{\chi_i^{/}} \beta} \right\} \tag{2.2}$$

A Taylor seies expansion about B gives to second order

$$R_i = \varepsilon_i + \left\{ (P_i - \frac{1}{2})\varepsilon_i - (n_i P_i Q_i)^{\frac{1}{2}} \right\} \sum_r x_{ir} (\hat{\beta}_r - \beta_r) \tag{2.3}$$

so

$$E(R_i) \approx (n_i.P_i.Q_i)^{\frac{1}{2}} \sum_r x_{ir} \left\{ (P_i - \frac{1}{2})(P_i Q_i)^{-1} a_{ir} - b_r) \right\} \tag{2.4}$$

where b_r denotes the bias of $\hat{B}r$ and

$$a_{ir} = E\{(p_i - P_i)(\hat{\beta}_r - \beta_r)\} \tag{2.5}$$

The biases have been found by Sarraf and Young (1985) Correct to $0(N^{-1})$ for the ML, MCS and WLS estimation procedures and are given by.

$$b_r^{(1)} = -\frac{1}{2}\sum_S \sum_t \sum_u I^{rs} I^{tu} K_{stu} \qquad (2.6)$$

$$b_r^{(2)} = b_r^{(3)} = \frac{1}{2}\sum_S I^{rs} \sum_i x_{is}(Q_i - P_i) + +2b_r^{(1)} \qquad (2.7)$$

respectively, where

$$I_{rs} = \sum_i n_i P_i Q_i x_{ir} x_{is}, \qquad k_{rst} = \sum_i n_i P_i Q_i (Q_i - P_i) x_{ir} x_{is} x_{it} \quad (2.8)$$

and I^{rs} denotes the elements in the (r+1)st row and (s+1) column in the inverse of $I=((I_{rs}))$. All summations over r, s,…. Run from 0, 1,……..

To find a_{ir} correct to $o(N^{-1})$, we may use the first order approxcimation

$$\hat{\beta}_r - \beta_r = -\sum_S \lambda^{rs} U_s \qquad (2.9)$$

where

$$U_r = \frac{\partial \phi}{\partial \beta_r}, \qquad \lambda_{rs} = E\left(\frac{\partial^2 \phi}{\partial \beta_r \partial \beta_s}\right) \qquad (2.1)$$

and λ_{rs} denoted the element in the (r+1)st row and (s+1)st column of $\lambda = ((\lambda_{rs}))$

For Ml- estimation, we have

$$U_r = -\sum_i n_i x_{ir}(p_i - P_i), \quad \lambda_{rs} = I_{rs}, E\{(p_i - P_i)U_s\} = x_{is} P_i Q_i$$

so correct to $0(N^{-1})$we have

$$a_{ir}^{(1)} = \sum_s I^{rs} x_{is} P_i Q_i$$

for MCS estimation, we have

$$U_r = -2\sum_i n_i x_{ir}(p_i - P_i) - (P_i - \frac{1}{2})(p_i - P_i)^2 (P_i Q_i)^{-1},$$

$$\lambda_{rs} = 2I_{rs} + 0(1), E\{(p_i - P_i)U_s\} = -2x_{is}P_iQ_i + 0(n_i^{-1}$$

hence correct to $0(N^{-1})$ we have

$$a_{ir}^{(2)} = \sum_s I^{rs} x_{is} P_i Q_i$$

For WLS estimation, we have

$$U_r = -2\sum_i n_i x_{ir} p_i q_i \{\log(p_i / q_i) - \log(P_i / Q_i)\}$$

$$\lambda_{rs} = 2I_{rs} + 0(1), \quad E\{(p_i - P_i)U_s\} = 2x_{is}P_iQ_i + 0(n_i^{-1})$$

giving

$$a_{ir}^{(3)} = \sum_s I^{rs} x_{is} P_i Q_i \text{ ....(2.13)}$$

Since $a_{ir}^{(1)} = a_{ir}^{(2)} = a_{ir}^{(3)}$, we have the general formula correct to $0(N^{-\frac{1}{2}})$ covering the threee methods of estimation

$$E(R_i) \approx (n_i P_i Q_i)^{\frac{1}{2}} \left\{ (P_i - \frac{1}{2}) \sum_r \sum_s x_{ir} x_{is} I^{rs} - \sum_r x_{ir} b_r \right\} \quad (2.14)$$

where the biases b_r one time by (2.6) and (2.7)

The second order expression for the covariance matrix of the Pearson residuals is given by

$$\text{cov}(R) = I_g - nV^{\frac{1}{2}} P^{*\prime} I^{-1} P^* V^{-\frac{1}{2}} n \quad (2.15)$$

where

$R' = (R_1,..R_g), n = diag(n_1,..,n_g), V = diag(n_1 P_1 Q_1,..,n_g, P_g, Q_g)$ I_g is the identity matrix of order g and

$$P^* = \begin{bmatrix} \partial P_1 / \partial \beta_0 & \cdots & \partial P_g / \partial \beta_0 \\ \cdot & & \\ \cdot & & \\ \partial P_1 / \partial \beta_k & & \partial P_g / \partial \beta_k \end{bmatrix} \quad (2.16)$$

The covariance result holds for ML. MCS and WLS estimation and for arbitrary model specification for the $\{p_i\}$. for the logistic regression model we have $\partial P_i / \partial \beta_0 = x_{ij} P_i Q_i$ so $P^* = X' n^{-1} V$ and hence

$$\mathrm{cov}(R) = I_g - V^{\frac{1}{2}} X (X'VX)^{-1} X'V^{\frac{1}{2}} \quad (2.17)$$

If c_{ij} denotes the (i, j) the element in

$$C = V^{\frac{1}{2}} X (X^1 VX)^{-1} X^1 V^{\frac{1}{2}}$$

then $c_{ij} = (n_i P_i Q_i)^{\frac{1}{2}} (n_j P_j Q_j)^{\frac{1}{2}} \sum_r \sum_s x_{ir} x_{js} I^{rs}$ (2.18)

Hence

$$\mathrm{var}(R_i) = 1 - c_{ii} = 1 - n_i P_i Q_i \sum_r \sum_s x_{ir} x_{is} I^{rs} \quad (2.19)$$

and

$$corr(R_i, R_j) = -c_{ij} / \{(1 - c_{ii})(1 - c_{jj})\}^{\frac{1}{2}} \quad (2.20)$$

Although the variances of the $\{R_i\}$ depend on (P_i) which are unknown, their average variance is independent of the $\{ P_i \}$. Thus using (2.17), the sum of the variances is given by

$$\sum_i \mathrm{var}(\dot{R}_i) = tr(I_g) = tr\left\{ (V^{\frac{1}{2}} X (X'VX)^{-1} X'V^{\frac{1}{2}} \right\}$$

$$= tr(I_g) - tr(I_{k+1)} = g - k - 1 \quad (2.21)$$

In order to obtain information about the magnitude of the approximate expectations and variances given by (2.14) and (2.19), their values have been computed for the case of a single explanatory variable with x_i =i-1 , i=1,2….,g for g=5,10,n =25. 50, 100 and six (Bo, B_i) cofigurations give a wide range of values for the success probabilities { p_i }. The configurations are shown in table 1.

TABLE 1

Parameter values (B_0,B_1) and success probabilities

	G=5	{ p_i }
(i)	B_o =-2.0, B_1 =0.4	0.119, 0.168, 0.232, 0.310, 0.401
(ii)	B_o =-1.0, B_1 =0.5	0.269, 0.378, 0.500, 0.623, 0.731
(iii)	B_o =0.5, B_1 =0.5	0.623, 0.731, 0.818, 0.881, 0.924

g=10

(iv)	B_o =-2.0, B_1 =0.2	0.119, 0.142, 0.168, 0.198, 0.231 0.269, 0.310, 0.354, 0.401, 0.450
(v)	B_o =-0.4, B_1 =0.2	0.401, 0.450, 0.500, 0.550, 0.591 0.646, 0.690, 0.731, 0.769, 0.802
(vi)	B_o =-0.5, B_1 =0.2	0.623, 0.668, 0.711, 0.750, 0.785 0.818, 0.846, 0.870, 0.891, 0.908

Values of E (R_i) based on the ML, MCS and WLS methods of estimation are shown in table 2 for the six configurations given in table

1 and for sample sizes n_i =n= 25, 50, 100, i=1,….,g. Values of the approximate variances of the { R_i } given by (2.19) are also shown, these variances being independent of the common sample size n.

TABLE 2

Approximate expectations and variances of residuals $E_i^{(1)} = 10 \times E(R_i)$ based on MLS fit, $Ei^{(2)} = 10 \times E(Ri) based$ on MCS or WLS fit.

		n=25		n=50		n=100		
Configuration	i	$E_i^{(1)}$	$E_i^{(2)}$	$E_i^{(1)}$	$E_i^{(2)}$	$E_i^{(1)}$	$E_i^{(1)}$	Var (R_i)
	1	-0.34	-1.62	-0.24	-1.15	-0.17	-0.81	0.504
	2	0.14	-1.00	0.10	-0.71	0.07	-0.50	0.671
(i)	3	0.28	-0.63	0.20	-0.45	0.14	-0.32	0.780
	4	0.09	-0.49	0.07	-0.35	0.05	-0.25	0.707
	5	-0.21	-0.39	-0.15	-0.28	-0.11	-0.20	0.338
	1	-0.07	-0.56	-0.05	-0.40	-0.07	-0.56	0.438
	2	0.12	-0.15	0.08	-0.11	0.12	-0.15	0.674
(ii)	3	0.00	0.00	0.00	0.00	0.00	0.00	0.775
	4	-0.12	0.15	-0.08	0.11	-0.12	0.15	0.674
	5	0.07	0.56	0.05	0.40	0.07	0.56	0.438
	1	0.28	0.39	0.20	0.27	0.14	0.19	0.309
	2	-0.16	0.63	-0.11	0.44	-0.08	0.31	0.710
(iii)	3	-0.37	0.92	-0.26	0.65	-0.18	0.46	0.767
	4	-0.13	0.15	-0.09	1.03	-0.06	0.73	0.663
	5	0.45	0.22	0.32	0.15	0.23	0.11	0.551
	1	-0.22	-1.84	-0.16	-1.30	-0.11	-0.92	0.733
	2	-0.08	-1.63	-0.06	-1.15	-0.04	-0.81	0.774
	3	0.03	-1.42	-0.02	-1.01	0.02	-0.71	0.817
	4	0.10	-1.23	0.07	-0.87	0.05	-0.61	0.855
(iv)	5	0.13	-1.05	0.09	-0.74	0.07	-0.52	0.883
	6	0.12	-0.87	0.08	-0.62	0.06	-0.44	0.892
	7	0.07	-0.71	0.05	-0.50	0.03	-0.36	0.892
	8	0.00	-0.54	0.00	-0.38	0.00	-0.27	0.824
	9	-0.07	-0.35	-0.05	-0.25	-0.04	-0.18	0.737
	10	-0.11	-0.13	-0.08	-0.09	-0.06	-0.06	0.612
	1	0.01	-0.32	-0.01	-0.22	-0.01	-0.16	0.648

	2	0.05	-0.11	0.03	-0.08	0.02	-0.05	0.744
	3	0.04	0.07	0.03	0.05	0.02	0.03	0.819
	4	0.00	0.21	0.00	0.15	0.00	0.11	0.869

(v)	5	-0.04	0.35	-0.03	0.25	-0.02	0.17	0.891
	6	-0.07	0.48	-0.05	0.34	-0.04	0.24	0.886
	7	-0.08	0.63	-0.06	0.45	-0.04	0.31	0.859
	8	-0.05	0.79	-0.03	0.56	-0.02	0.40	0.816
	9	-0.02	0.98	0.02	0.69	0.01	0.49	0.762
	10	0.13	0.12	0.09	0.83	0.07	0.59	0.706
	1	0.17	0.34	0.12	0.24	0.09	0.17	0.596
	2	0.08	0.58	0.05	0.41	0.04	0.29	0.735
	3	-0.03	0.79	-0.02	0.56	-0.01	0.39	0.827
	4	-0.11	0.98	-0.08	0.69	-0.05	0.49	0.877
(vi)	5	-0.16	1.17	-0.11	0.83	-0.08	0.58	0.892
	6	-0.16	1.36	-0.11	0.96	-0.08	0.68	0.881
	7	-0.11	1.57	-0.08	1.11	-0.06	0.79	0.854
	8	-0.02	1.79	-0.02	1.26	-0.01	0.89	0.818
	9	0.11	2.01	0.08	1.42	0.06	1.01	0.779
	10	0.27	2.23	0.19	1.58	0.14	1.12	0.743

The results show that the absolute values of the approximate expectation of the residuals based on a ML fit are generally much smaller than those based on a MCS or WLS fit. Also, the approximate variances of the residuals are often appreciably less than one, particularly for the extreme values of the index i of the residuals.

The approximations to the expectation and variances of the residuals allow modified residuals to be used whose mean and variance are likely to be closer to zero and one, respectively, than those of the corresponding moments of the unmodified residuals $\{R_i\}$.

Modified residuals allowing for variance adjustment are defined by

$$R_i^* = R_i /(1-\hat{c}_{ii})^{\frac{1}{2}}, \qquad i=1,\ldots\ldots,g \qquad (2.22)$$

where

$$\hat{c}_{ii} = n_i \hat{P}_i \hat{Q}_i \sum_r \sum_s x_{ir} x_{is} \hat{I}^{rs} \qquad (2.23)$$

denotes the estimates of c_{ii} using $\{p_i\}$

With variance and expectation adjustment, the modified residuals are defined by

$$R_i^{**} = (R_i - E_i)/(1-\hat{c}_{ii})^{\frac{1}{2}} \quad i=1,\ldots\ldots g \qquad (2.24)$$

where $\hat{E}_i$ denotes the estimate of $E(R_i)$

Using the modified residual, extreme residual statistics will be denoted by

$$R^*_{\max} = \max_i R_i^*, \quad R^*_{\min} = \min_i R_i^*, \quad R^*_m = \max_i R_i^* \qquad (2.25)$$

$$R^{**}_{\max} = \max_i R_i^{**}, \quad R^{**}_{\min} = \min_i R_i^{**}, \quad R^{**}_m = \max_i R_i^{**} \qquad (2.26)$$

3. Approximations to the Critical Values of the Extreme Residual Statistics

In order to use the extreme residual statistics in formal goodness of fit tests for the logistic regression model, we need approximations to the percentage points of their distributions when the model is correct. Simple approximations are available based on the use of the Banferroni inequality we illustrate in the approach for the $R_{\max}$ statistic.

We have

$$\sum_i P(R_i \geq r) - \sum_i \sum_{<j} P(R_i \geq r, R_j \geq r) \leq P(R_{\max} \geq r) \leq \sum_i P(R_i \geq r) \qquad (3.1)$$

Since nearly all pairs of residuals are negatively correlated, we are led to the conjecture that :

$$\sum_{icj} p(R \geq r, R_j \geq r) < P \sum_{i<j} P(R_i \geq r) p(R_i \geq r) \qquad (3.2)$$

This leads to the inequalitets

$$\sum_i P(R_i \geq r) - \frac{1}{2}\left\{\sum_i P(R_i \geq r)\right\}^2 \leq P(R_{\max} \geq r) \leq (\sum_i P(R_i \geq r) \qquad (3.3)$$

For large r, we use the approximation.

$$P(R_{\max} \geq r) \approx \sum_i P(R_i \geq r) \qquad (3.4)$$

The error in the approximation being less than $\frac{1}{2}(1-g^{-1})\left\{\sum_i P(R_i \geq r)\right\}^2$ if the conjective given by (3.2) is true. Since var (R_i) given by (2.19) depends on the $\{P_i\}$which are unknown, we may either use the estimates of var(R_i) based on the fitted model or ignore the variations in the variances work with their average value (g-k-1)/g. Using the latter approach we take.

$$\left(\frac{g}{g-k-1}\right)^{\frac{1}{2}} R_i \text{ approx } N(0,1) \qquad (3.5)$$

and use of (3.4) gives

$$P(R_{\max} \geq r) \approx 1-\phi\left\{\left(\frac{g}{g-k-1}\right)^{\frac{1}{2}} r\right\} \qquad (3.6)$$

where $\phi(.)$ denote the c.d.f. of the N(0,1) distribution. If we let $r_{max}(1-\alpha)$ denote the upper 100α percentage point of the ditribution of R_{max}, we have the approximation

$$r_{man}^{(1-\alpha)} \approx \left(\frac{g-k-1}{g}\right)^{\frac{1}{2}} U_1 - \frac{\alpha}{g} \qquad (3.7)$$

where $U_{1-}\alpha$ is the 100(1-α) percentoale of the N(0,1) distribution. If $r_{min}(\alpha)$ at $r_m(1-\alpha)$ denote the lower and upper 100α persentage pointage of the distributions of R_{min} and R_m respectively, similar measurement lead to the approximations

$$r_{\min}(\alpha) \approx -\left(\frac{g-k-1}{g}\right)^{\frac{1}{2}} U_{1-\frac{\alpha}{g}} \qquad (3.8)$$

$$r_m(1-\alpha) \approx -\left(\frac{g-k-1}{g}\right)^{\frac{1}{2}} U_{1-\frac{\alpha}{2g}} \qquad (3.9)$$

since to the same order of approximation var(D_i)=var(R_i), approximation to the critical values $d_{max(1-\alpha)}$, $d_{min(\alpha)}$ and $d_{m(1-\alpha)}$ for the D_{max}, D_{min} and D_m statistics are also given by (3.7), (3.8) and (3.9), respectively.

For the modified residulas R_i^* which use the variance estimates and R_i^{**} which use both expectation and variance estimates, we take both sets of residuals to be approximately distributed as N(0,1). If the logistic regression model is correct, we are led to the approximations.

$$r^*_{\max}(1-\alpha) = r^{**}_{\max}(1-\alpha) \approx U_1 - \frac{\alpha}{g} \qquad (3.10)$$

$$r^*_{\min}(\alpha) = r^{**}_{\min x}(\alpha) \approx U_1 - \frac{\alpha}{g} \qquad (3.11)$$

$$r^*_m(1-\alpha) = r^{**}_m(1-\alpha) \approx U_1 - \frac{\alpha}{2g} \qquad (3.12)$$

4. Mote Carlo Results

in order to examine the accuracy of the approximations to the critical values of the extreme residul statistics, a Monte Carlo investigation was made for the case of a single explanatory variable using the parameter configurations given in table 1. equal sample sizes n_i=n=25, 50, 100, i=1,……g were used. The model fits by ML, MCS, WLS and MWLS estimation were made using the statistical package GLIM and a run-size of 2000 was used in each case. The empirical distributions of the extreme residual were used to obtain the upper and lower critical values for significance levels α =0.10, 0.05, 0.025 and 0.01.

The actual significance levels associated with the approxcimating critical values given by (3.7) to (3.12) were also determined and contrasted with the nomial values of α .

The broad findings reached from the investigation are:

- I) the differences between the values of the extreme statistics based on the modified residuals R_i^* and R_i^{**} are in general very small and so little is gained by making an additional adjustment for estimated expectations, once the ordinary residuals have been adjusted for estimated variance.
- II) The differences between the significance levels when variance adjusted residuals are used instead of the ordinary Pearson residuals are generally quite small. The extreme statistics based on values of the residuals variance adjustment lead to some improvement for the smaller values of α .
- III) No one estimation procedure systematically provides better control over the significance level of the test.

Tables 3 to 14 show the actual significance levels as estimated by simulation associated with the approximating values for the extreme residual statistics based on the ordinaary residuals and the variance asjusted residuals. It is encouraging to see there is generally good agreement with the nomial significance levels, particularly for the modulus statistics R_m and R_m^*.

Table 3

Estimated singificance levels for approximate critical valuse for (a) R_{max} and (b) R_{max}^* based on ML fit

	∝=0.10	∝=0.05	∝=0.025	∝=0.01

Configuration	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
n=25	0.123	0.107	0.064	0.057	0.038	0.029	0.017	0.011
i) n=50	0.116	0.101	0.070	0.055	0.039	0.029	0.018	0.013
n=100	0.109	0.090	0.057	0.047	0.032	0.025	0.017	0.012
n=25	0.109	0.101	0.056	0.050	0.031	0.022	0.015	0.009
ii) n= 50	0.089	0.088	0.052	0.042	0.026	0.022	0.012	0.011
N=100	0.092	0.088	0.045	0.044	0.023	0.021	0.011	0.009
N=25	0.067	0.071	0.031	0.030	0.014	0.010	0.004	0.004
iii) n=50	0.084	0.081	0.043	0.040	0.019	0.016	0.007	0.005
N=100	0.090	0.083	0.050	0.038	0.023	0.018	0.009	0.007
n=25	0.142	0.137	0.082	0.080	0.046	0.044	0.021	0.019
iv) n=50	0.114	0.112	0.060	0.055	0.026	0.023	0.012	0.011
n=100	0.122	0.111	0.061	0.062	0.033	0.033	0.017	0.018
n=25	0.083	0.079	0.049	0.042	0.018	0.016	0.007	0.007
v) n=50	0.088	0.092	0.042	0.040	0.018	0.018	0.007	0.009
n=100	0.089	0.086	0.042	0.042	0.021	0.021	0.008	0.008
n=25	0.044	0.013	0.000	0.019	0.008	0.008	0.001	0.002
vi) n=50	0.070	0.070	0.029	0.030	0.012	0.010	0.002	0.002
n=100	0.077	0.074	0.037	0.033	0.013	0.014	0.005	0.005

Table 4

Estimated significance levels for approximate critical values for (a) R_{min} and (b) R_{min} based on ML fit

	∝=0.10		∝=0.05		∝=0.025		∝=0.01	
Configuration	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
n=25	0.085	0.082	0.036	0.038	0.015	0.020	0.005	0.004
i) n=50	0.087	0.079	0.042	0.038	0.019	0.018	0.008	0.004
n=100	0.080	0.084	0.045	0.037	0.022	0.021	0.012	0.007
n=25	0.097	0.095	0.053	0.046	0.026	0.024	0.012	0.007
ii) n= 50	0.097	0.091	0.050	0.049	0.028	0.025	0.013	0.010
N=100	0.010	0.088	0.047	0.044	0.025	0.023	0.011	0.008
N=25	0.141	0.124	0.080	0.061	0.043	0.003	0.025	0.012
Iii) n=50	0.128	0.110	0.074	0.061	0.041	0.029	0.019	0.012
N=100	0.100	0.089	0.055	0.043	0.026	0.021	0.012	0.008
n=25	0.067	0.065	0.029	0.029	0.012	0.012	0.005	0.003
iv) n=50	0.078	0.075	0.028	0.026	0.013	0.011	0.005	0.003
n=100	0.081	0.078	0.034	0.033	0.013	0.015	0.005	0.004
n=25	0.100	0.095	0.047	0.047	0.026	0.025	0.016	0.014
v) n=50	0.114	0.112	0.058	0.058	0.031	0.029	0.011	0.010
n=100	0.109	0.103	0.058	0.050	0.032	0.027	0.013	0.012
n=25	0.160	0.150	0.095	0.086	0.057	0.052	0.026	0.026
Vi) n=50	0.130	0.128	0.072	0.067	0.039	0.035	0.018	0.014
n=100	0.132	0.125	0.078	0.073	0.046	0.041	0.024	0.019

Table 5

Estimated significance levels for approximate critical values for (a) R_m and (b) R^*_m Statistics, based on ML fit

	∝=0.10		∝=0.05		∝=0.025		∝=0.01	
Configuration	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
n=25	0.091	0.077	0.050	0.041	0.029	0.018	0.012	0.006
i) n=50	0.101	0.079	0.055	0.042	0.032	0.019	0.014	0.009
n=100	0.089	0.069	0.049	0.037	0.032	0.019	0.016	0.010
n=25	0.096	0.081	0.052	0.040	0.029	0.020	0.013	0.008
ii) n= 50	0.089	0.075	0.049	0.040	0.029	0.022	0.028	0.010
N=100	0.079	0.076	0.043	0.039	0.024	0.020	0.014	0.007
N=25	0.101	0.074	0.054	0.043	0.034	0.023	0.016	0.013
iii) n=50	0.108	0.086	0.058	0.039	0.030	0.017	0.012	0.009
N=100	0.096	0.069	0.045	0.033	0.026	0.016	0.008	0.006
n=25	0.106	0.104	0.058	0.056	0.031	0.026	0.014	0.012
iv) n=50	0.085	0.078	0.038	0.034	0.019	0.017	0.009	0.006
n=100	0.093	0.092	0.045	0.047	0.026	0.025	0.013	0.011
n=25	0.083	0.084	0.042	0.041	0.026	0.024	0.012	0.008
v) n=50	0.094	0.090	0.047	0.043	0.022	0.022	0.007	0.007
n=100	0.093	0.086	0.049	0.046	0.027	0.023	0.012	0.009
n=25	0.112	0.101	0.064	0.060	0.034	0.030	0.018	0.018
vi) n=50	0.096	0.094	0.049	0.044	0.025	0.022	0.010	0.007
n=100	0.111	0.102	0.058	0.054	0.036	0.030	0.014	0.013

Table 6

Estimated significance levels for approximate critical values for (a) R_m and (b) R_{max} Statistics, based on MCS fit

	∝=0.10		∝=0.05		∝=0.025		∝=0.01	
Configuration	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
n=25	0.056	0.046	0.027	0.020	0.014	0.009	0.004	0.003
i) n=50	0.074	0.058	0.033	0.027	0.017	0.011	0.009	0.005
n=100	0.075	0.062	0.038	0.027	0.019	0.014	0.008	0.006
n=25	0.107	0.100	0.057	0.051	0.027	0.022	0.014	0.011
ii) n= 50	0.095	0.087	0.052	0.049	0.025	0.022	0.011	0.010
n=100	0.093	0.088	0.046	0.044	0.023	0.019	0.012	0.010
n=25	0.138	0.136	0.065	0.069	0.028	0.032	0.013	0.017
iii) n=50	0.140	0.132	0.071	0.074	0.040	0.053	0.014	0.011
N=100	0.124	0.113	0.073	0.066	0.041	0.031	0.017	0.013
n=25	0.079	0.080	0.036	0.038	0.017	0.017	0.009	0.009
iv) n=50	0.073	0.064	0.030	0.026	0.014	0.013	0.005	0.005
n=100	0.087	0.084	0.044	0.042	0.022	0.022	0.012	0.010
n=25	0.096	0.091	0.051	0.045	0.027	0.020	0.009	0.009
v) n=50	0.106	0.102	0.050	0.053	0.023	0.024	0.008	0.009
n=100	0.103	0.101	0.051	0.050	0.023	0.023	0.010	0.010
n=25	0.078	0.074	0.030	0.030	0.012	0.017	0.004	0.004
vi) n=50	0.103	0.109	0.051	0.049	0.021	0.024	0.005	0.005
n=100	0.109	0.106	0.052	0.047	0.025	0.024	0.009	0.008

Table 7

Estimated significance levels for approximate critical values for (a) R_{min} and (b) R^*_{min} Statistics, based on MCS fit

	∝=0.10		∝=0.05		∝=0.025		∝=0.01	
Configuration	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
n=25	0.124	0.129	0.071	0.069	0.034	0.040	0.012	0.019
i) n=50	0.127	0.122	0.065	0.062	0.035	0.036	0.018	0.016
n=100	0.109	0.109	0.063	0.061	0.034	0.035	0.017	0.015
n=25	0.100	0.097	0.055	0.051	0.028	0.023	0.010	0.011
ii) n= 50	0.099	0.087	0.051	0.049	0.031	0.026	0.014	0.013
N=100	0.095	0.090	0.047	0.045	0.025	0.021	0.013	0.009
N=25	0.048	0.041	0.025	0.017	0.014	0.008	0.007	0.002
iii) n=50	0.063	0.048	0.031	0.020	0.015	0.009	0.005	0.002
N=100	0.056	0.046	0.027	0.018	0.011	0.007	0.005	0.003
n=25	0.100	0.098	0.047	0.045	0.020	0.020	0.007	0.007
iv) n=50	0.108	0.101	0.049	0.045	0.017	0.018	0.008	0.006
n=100	0.086	0.099	0.045	0.042	0.022	0.021	0.006	0.005
n=25	0.075	0.074	0.039	0.037	0.020	0.020	0.009	0.006
v) n=50	0.094	0.095	0.049	0.049	0.023	0.020	0.008	0.004
n=100	0.100	0.091	0.046	0.044	0.025	0.023	0.012	0.010
n=25	0.069	0.064	0.035	0.030	0.017	0.016	0.008	0.006
vi) n=50	0.075	0.068	0.034	0.030	0.018	0.011	0.004	0.003
n=100	0.094	0.092	0.051	0.048	0.029	0.025	0.010	0.009

Table 8

Estimated significance levels for approximate critical values for (a) R_m and (b) R_m Statistics, based on MLS fit

	∝=0.10		∝=0.05		∝=0.025		∝=0.01	
Configuration	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
N=25	0.091	0.076	0.045	0.044	0.020	0.023	0.007	0.009
i) n=50	0.091	0.076	0.049	0.043	0.031	0.024	0.013	0.008
n=100	0.090	0.075	0.053	0.042	0.029	0.022	0.014	0.008
n=25	0.101	0.089	0.051	0.041	0.029	0.022	0.013	0.011
ii) n= 50	0.092	0.082	0.053	0.042	0.028	0.027	0.017	0.013
N=100	0.082	0.076	0.043	0.037	0.026	0.021	0.015	0.007
N=25	0.082	0.074	0.040	0.034	0.023	0.016	0.007	0.006
iii) n=50	0.094	0.082	0.052	0.037	0.023	0.016	0.009	0.005
N=100	0.095	0.074	0.051	0.033	0.024	0.017	0.011	0.009
n=25	0.080	0.080	0.037	0.036	0.018	0.018	0.010	0.007
iv) n=50	0.077	0.068	0.031	0.029	0.016	0.013	0.007	0.005
n=100	0.034	0.081	0.044	0.043	0.022	0.021	0.009	0.008
n=25	0.085	0.077	0.045	0.039	0.021	0.018	0.008	0.004
v) n=50	0.092	0.093	0.045	0.042	0.018	0.016	0.006	0.007
n=100	0.090	0.087	0.045	0.043	0.026	0.022	0.011	0.010
n=25	0.064	0.058	0.028	0.031	0.015	0.013	0.005	0.004
vi) n=50	0.082	0.075	0.036	0.035	0.015	0.010	0.003	0.003
n=100	0.097	0.092	0.053	0.48	0.025	0.020	0.008	0.005

Table 9

Estimated significance levels for approximate critical values for (a) R_{max} and (b) R^*_{max} Statistics, based on WLS fit

	∝=0.10		∝=0.05		∝=0.025		∝=0.01	
Configuration	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
n=25	0.052	0.039	0.023	0.017	0.013	0.008	0.004	0.004
i) n=50	0.069	0.052	0.033	0.026	0.017	0.009	0.009	0.004
n=100	0.072	0.058	0.038	0.026	0.020	0.012	0.008	0.006
n=25	0.111	0.104	0.061	0.057	0.035	0.032	0.018	0.015
ii) n= 50	0.095	0.088	0.053	0.048	0.028	0.024	0.012	0.011
N=100	0.099	0.088	0.047	0.045	0.024	0.021	0.012	0.010
N=25	0.199	0.188	0.121	0.112	0.073	0.081	0.039	0.047
iii) n=50	0.157	0.146	0.085	0.090	0.056	0.049	0.025	0.023
N=100	0.130	0.199	0.080	0.069	0.048	0.037	0.019	0.017
n=25	0.076	0.076	0.037	0.038	0.019	0.020	0.009	0.006
iv) n=50	0.070	0.064	0.030	0.028	0.016	0.013	0.006	0.005
n=100	0.087	0.082	0.045	0.042	0.022	0.023	0.011	0.012
n=25	0.108	0.104	0.057	0.057	0.034	0.030	0.014	0.013
v) n=50	0.111	0.109	0.056	0.058	0.027	0.026	0.009	0.011
n=100	0.104	0.103	0.052	0.051	0.027	0.024	0.012	0.011
n=25	0.116	0.116	0.059	0.053	0.023	0.024	0.010	0.011
vi) n=50	0.112	0.121	0.062	0.062	0.030	0.032	0.013	0.010
n=100	0.112	0.111	0.055	0.050	0.028	0.027	0.011	0.008

Table 10

Estimated significance levels for approximate critical values for (a) R_{min} and (b) R^*_{min} Statistics, based on WLS fit

	∝=0.10		∝=0.5		∝=0.025		∝=0.1	
Configuration	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
N=25	0.152	0.157	0.009	0.097	0.058	0.063	0.031	0.036
i) n=50	0.137	0.133	0.076	0.069	0.043	0.046	0.025	0.021
n=100	0.113	0.113	0.064	0.064	0.039	0.037	0.018	0.017
n=25	0.104	0.104	0.063	0.054	0.037	0.029	0.014	0.014
ii) n= 50	0.094	0.090	0.053	0.052	0.032	0.027	0.017	0.014
N=100	0.096	0.089	0.048	0.045	0.025	0.022	0.013	0.010
N=25	0.039	0.028	0.018	0.013	0.011	0.006	0.005	0.002
iii) n=50	0.056	0.039	0.027	0.016	0.013	0.006	0.003	0.002
N=100	0.053	0.042	0.026	0.017	0.010	0.007	0.005	0.002
n=25	0.125	0.125	0.065	0.064	0.035	0.035	0.013	0.014
iv) n=50	0.117	0.112	0.061	0.053	0.023	0.022	0.009	0.009
n=100	0.102	0.104	0.048	0.047	0.023	0.022	0.006	0.006
n=25	0.075	0.076	0.039	0.040	0.020	0.023	0.011	0.009
v) n=50	0.093	0.098	0.050	0.049	0.024	0.022	0.008	0.004
n=100	0.097	0.093	0.047	0.044	0.025	0.025	0.013	0.011
n=25	0.061	0.056	0.031	0.030	0.015	0.015	0.008	0.006
vi) n=50	0.071	0.064	0.035	0.029	0.019	0.012	0.004	0.003
n=100	0.094	0.091	0.050	0.049	0.029	0.025	0.010	0.010

Table 11

Estimated significance levels for approximate critical values for (a) R_m and (b) R^*_m Statistics, based on WLS fit

	∝=0.10		∝=0.05		∝=0.025		∝=0.01	
Configuration	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
n=25	0.115	0.104	0.068	0.067	0.040	0.044	0.023	0.021
i) n=50	0.102	0.083	0.058	0.052	0.038	0.029	0.018	0.013
n=100	0.092	0.078	0.055	0.044	0.031	0.022	0.016	0.010
n=25	0.113	0.010	0.066	0.058	0.035	0.032	0.016	0.019
ii) n= 50	0.095	0.085	0.056	0.046	0.033	0.029	0.019	0.014
N=100	0.085	0.078	0.045	0.039	0.028	0.022	0.015	0.007
N=25	0.134	0.124	0.081	0.083	0.052	0.054	0.023	0.032
iii) n=50	0.106	0.097	0.067	0.052	0.034	0.028	0.016	0.012
N=100	0.101	0.077	0.057	0.039	0.028	0.020	0.013	0.010
n=25	0.100	0.099	0.053	0.056	0.026	0.028	0.013	0.012
iv) n=50	0.089	0.076	0.038	0.034	0.019	0.016	0.008	0.008
n=100	0.090	0.086	0.045	0.044	0.024	0.023	0.010	0.009
n=25	0.092	0.093	0.052	0.051	0.030	0.028	0.011	0.009
v) n=50	0.099	0.099	0.050	0.046	0.022	0.019	0.007	0.007
n=100	0.093	0.088	0.049	0.046	0.028	0.025	0.012	0.011
n=25	0.088	0.079	0.037	0.036	0.017	0.020	0.007	0.006
vi) n=50	0.094	0.088	0.048	0.044	0.023	0.020	0.004	0.006
n=100	0.101	0.096	0.056	0.050	0.026	0.023	0.008	0.008

Table 12

Estimated significance levels for approximate critical values for (a) R_{max} and (b) R^*_{max} Statistics, based on MWLS fit

	∝=0.10		∝=0.05		∝=0.025		∝=0.01	
Configuration	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
n=25	0.046	0.036	0.022	0.017	0.013	0.007	0.004	0.003
i) n=50	0.059	0.043	0.026	0.021	0.016	0.008	0.006	0.003
n=100	0.060	0.046	0.030	0.022	0.016	0.011	0.008	0.005
n=25	0.111	0.108	0.061	0.058	0.034	0.032	0.018	0.014
ii) n= 50	0.099	0.090	0.053	0.050	0.030	0.024	0.012	0.011
N=100	0.095	0.090	0.094	0.047	0.025	0.022	0.012	0.011
N=25	0.183	0.166	0.089	0.085	0.044	0.040	0.019	0.016
iii) n=50	0.177	0.131	0.090	0.088	0.054	0.046	0.024	0.018
N=100	0.151	0.136	0.089	0.082	0.054	0.043	0.022	0.019
n=25	0.075	0.066	0.033	0.036	0.017	0.018	0.010	0.090
iv) n=50	0.053	0.052	0.025	0.025	0.012	0.011	0.005	0.005
n=100	0.070	0.069	0.035	0.036	0.020	0.021	0.010	0.010
n=25	0.110	0.108	0.058	0.056	0.032	0.028	0.012	0.010
v) n=50	0.117	0.116	0.062	0.060	0.031	0.028	0.009	0.011
n=100	0.112	0.109	0.056	0.055	0.030	0.025	0.013	0.011
n=25	0.110	0.098	0.045	0.040	0.020	0.020	0.009	0.008
vi) n=50	0.134	0.134	0.070	0.067	0.030	0.032	0.011	0.010
n=100	0.134	0.131	0.064	0.057	0.035	0.030	0.011	0.011

Table 13

Estimated significance levels for approximate critical values for (a) R_{min} and (b) R^*_{min} Statistics, based on MWLS fit

	∝=0.10		∝=0.05		∝=0.025		∝=0.01	
Configuration	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
n=25	0.150	0.151	0.092	0.081	0.049	0.047	0.019	0.022
i) n=50	0.155	0.149	0.088	0.075	0.045	0.046	0.026	0.022
n=100	0.135	0.133	0.075	0.071	0.044	0.042	0.020	0.018
n=25	0.105	0.109	0.059	0.053	0.035	0.026	0.012	0.013
ii) n= 50	0.102	0.089	0.055	0.053	0.032	0.030	0.017	0.014
N=100	0.095	0.090	0.050	0.045	0.025	0.024	0.014	0.010
N=25	0.038	0.033	0.021	0.016	0.012	0.008	0.007	0.003
iii) n=50	0.047	0.037	0.022	0.015	0.010	0.004	0.003	0.002
N=100	0.042	0.030	0.018	0.013	0.009	0.007	0.004	0.002
n=25	0.124	0.123	0.061	0.058	0.030	0.026	0.011	0.012
iv) n=50	0.133	0.128	0.067	0.060	0.028	0.023	0.010	0.009
n=100	0.119	0.114	0.057	0.055	0.027	0.024	0.009	0.008
n=25	0.069	0.070	0.036	0.039	0.020	0.022	0.009	0.007
v) n=50	0.090	0.091	0.047	0.046	0.023	0.019	0.007	0.003
n=100	0.087	0.085	0.044	0.043	0.023	0.024	0.012	0.010
n=25	0.055	0.054	0.031	0.027	0.015	0.015	0.006	0.008
vi) n=50	0.057	0.051	0.025	0.025	0.013	0.010	0.003	0.002
n=100	0.080	0.074	0.043	0.040	0.023	0.019	0.008	0.006

Table 14

Estimated significance levels for approximate critical values for (a) R_m and (b) R^*_m Statistics, based on MWLS fit

	∝=0.10		∝=0.05		∝=0.025		∝=0.01	
Configuration	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
n=25	0.107	0.087	0.058	0.051	0.029	0.030	0.011	0.014
i) n=50	0.107	0.084	0.059	0.051	0.039	0.028	0.016	0.012
n=100	0.095	0.081	0.057	0.048	0.033	0.025	0.016	0.011
n=25	0.109	0.099	0.065	0.053	0.036	0.030	0.016	0.015
ii) n= 50	0.097	0.088	0.058	0.050	0.032	0.030	0.019	0.014
N=100	0.088	0.080	0.046	0.043	0.029	0.022	0.015	0.009
N=25	0.103	0.090	0.054	0.044	0.028	0.024	0.012	0.009
iii) n=50	0.105	0.093	0.062	0.049	0.034	0.023	0.012	0.007
N=100	0.102	0.087	0.062	0.045	0.034	0.022	0.011	0.011
n=25	0.092	0.091	0.046	0.043	0.023	0.022	0.015	0.090
iv) n=50	0.090	0.082	0.039	0.044	0.018	0.016	0.008	0.007
n=100	0.091	0.088	0.047	0.033	0.023	0.023	0.010	0.008
n=25	0.090	0.091	0.050	0.049	0.026	0.025	0.009	0.008
v) n=50	0.102	0.101	0.051	0.045	0.022	0.017	0.007	0.007
n=100	0.093	0.091	0.050	0.046	0.028	0.025	0.012	0.011
n=25	0.073	0.065	0.034	0.033	0.017	0.018	0.007	0.006
vi) n=50	0.092	0.088	0.042	0.042	0.019	0.017	0.003	0.005
n=100	0.103	0.094	0.047	0.047	0.024	0.023	0.008	0.008

5.Conclusions

The logistic, Probit and complementary log-log models are commonly used in the analysis of binary response data when explanatory variables are present. Since logistic model is of simple interpretation therefore the analysis is focused on ~
model. Three methods of estimation for B have been used, they are the maximum, likelihood, Minimum chi-square, and weighted least squares. To examine the accuracy of the approximation of the extreme residuals statistics a Monte Carlo investigation was made for the case of a single explanatory variable. The important results reached from the investigation that the differences between the values of the extreme statistics based on the modified residuals and the ordinary residuals very small so little is gained by making an additional adjustment for estimated expectation. The other result is the difference between significance levels when variance adjusted residuals are used instead of the ordinary Pearson residuals are generally quite small. For the extreme statistics based on the absolute values of the residuals, variance adjustment leads to some improvement for the smaller values of the nominal level of significance
The final important finding from this paper is that no one of the three methods of estimation procedure provides better control over the significance level.

Therefore as the previous tables show that there is a good agreement in general between the ordinary residuals and the variance adjusted residuals. So we suggest that using the ordinary residuals will be sufficient enough with out complicate the computation of involving the use of adjusted residuals no matter what the method of estimation is.

References

المراجع باللغة العربية:

١-الزوري، مثنى علي عبد الله، ""خواص تقديرات معلمات نموذج وحدة الاحتمال مع تطبيق عملي""، كلية الاقصاد- جامعة بغداد- ١٩٩٥م.
٢-الطائي، رواء ابراهيم مهدي، ""خواص تقديرات نموذج اللوغاريتم المزدوج التكميلي"" كلية الإدارة والاقتصاد — جامعة بغداد- ١٩٩٥م.
٣-متى، نغم نافع، ""خواص قوة الاختبار وحدود الثقة لمعاملات اللوجستك الخطي مع دراسة مقارنة"".

المراجع باللغة الإنكليزية:

4. Aitchison J. and Silvey R. (1957) “The generalization of probit to the case of multiple response”. Biometrika; 44, 13 1-140.
5. Aldrich J. H and Nelson F. D. (1985) “Linear probability logit and probit models” Sage publication Inc. USA.
6. Al-Sarraf and Young, D. H. (1985) Properties of estimators of parameters in logistic regression models. Technical Report TR 105185, Departement of Mathematics and statistics, Brunel University.
7. Berkson, J (1955) Maximum likelihood and minimum X^2 estimates of the logistic function. J-Amer. Statist Assoc. 50 130-162.
8. Cordeiro, G. M., McCullagh P (191) Bais correction in generalized linear models J. R. statistical Soc. B, 629-643.
9. Cox, D. R. and SneIl, E. J. (1968). A general definition of residuals J. R. Statist. Soc.B, 30, 248-265.
10. Cox, D. R. SnelI,j E. J. (1989) Analysis of Binary Data 2nd Edition. London chapman, Hall.
11. Farewell V. T. (1982) “Anote on regression analysis of ordinal with variability of classification, Biometrika vol 69, 533.
12. Firth, D. (1993) Bias reduction of maximum likelihood estimates, Biometrika, 80, 27-3 8.

13. Gait, J. J. and Zweifel, J. R. (1976) On the bias of various estimators of the logistic function and its variance with application to quantal bioassay, Biometrika, *54,* 181-187.
14. Maurice M. T. (1971) Multivariate analysis techniques for educational and psycological research. John whiley and sons Inc.
15. McCullagh P. (1980) Regression models for ordinal data J. R. Statist. Soc. B.
16. McCullagh P. and Nelder J.A. (1983). Generalized linear models London chapman and hail Ltd.

Received, 20 July /2000.