European Journal of Scientific Research
ISSN 1450-216X Vol.60 No.4 (2011), pp. 584-592

http://www.eurojournals.com/ejsr.htm

The Effect of Overdispersion on Regression Based Decision with Application to Churn Analysis on Indonesian Mobile Phone Industry

Asep Saefuddin
Department of Statistics, Faculty of Mathematics and Natural Sciences
Bogor Agricultural University, Indonesia
E-mail: asaefuddin@gmail.com

Nur Andi Setiabudi
Innovation Center for Regional Resource Development and Community Empowerment
(i-Crescent), Indonesia
E-mail: nurandi.mail@gmail.com

Noer Azam Achsani
Department of Economics and Graduate School of Management and Business
Bogor Agricultural University, Indonesia
E-mail: achsani@yahoo.com; achsani@mb.ipb.ac.id

Abstract

Extra binomial variation or commonly known as overdispersion in logistic regression will provide incorrect conclusions. Overdispersion may be caused by the occurrence of variation in the response probabilities or correlation within the response variable. On the other hand, independent assumption of response variable is required in the logistic regression. In the case of correlated outcomes, although maximum likelihood gives unbiased estimates, their standard errors are underestimate.

This study was aimed at showing the effect of overdispersion on the hypothesis test of logistic regression. The example was taken from telecommunication industry to analyze churn of the subscribers. A simple method proposed by William was used to correct the effect of overdispersion by taking inflation factor into consideration. The result showed that the William method adjusted the standard error of estimates and provided more precise conclusion which was important in marketing strategies.

Keywords: Binomial logistic regression, overdispersion, William method, churn analysis

1. Introduction

In logistic regressions, response variable is assumed to be independent and probability of success has no extra variation. If these assumptions are violated, such as response variable is positively correlated, variation of response variable will be greater than binomial variation. This condition is referred to overdispersion. The standard generalized linear models under a binomial assumption often exhibit overdispersion (Hinde and Demétrio 1998; Dean 1992). McCullagh and Nelder (1989) stated that non-independent data is common phenomena, while the independence is a special case. Hence, clarifying the effect of overdispersion is necessary to obtain right conclusion.

If the response variable contains overdispersion, maximum likelihood produces unbiased estimates with narrow standard errors. As a result, effect of factors tends to be significant to the response variable. This situation is certainly unexpected and has to be avoided in decision making process in every industry. In the telecom industry, for example in churn modeling, the problem of overdispersion should get serious attention so that conclusions drawn provide correct insight on how the company's strategy should be implemented.

The approach in handing overdispersion was introduced firstly by William (1982). William equates the value of Pearson's chi-square statistic of the model to its approximate expected value in order to obtain the optimal weighting value for parameter estimation. In addition to William approch, there are many methods have been applied to handle overdispersion, for example, beta-binomial regression (Hajarisman and Saefuddin 2008; Kurnia *et al* 2002; Hinde and Demétrio 1998), mixed effect model (Handayani and Kurnia 2006) and logistics normal model (Hinde and Demétrio 1998). Overdispersion can also accommodated by using Quasi-likelihood method (Baggerly *et al* 2004) and double-exponential method approach (*see* Lambert and Roeder 1995).

In term of the application, many literatures have showed the overdispersion phenomenon in many areas, such as education (Hajarisman and Saefuddin 2008; Kurnia *et al* 2002), actuarial sciences (Cheong 2008; Ismail and Jemain 2005, 2007), bioinformatics (Lu *et al* 2005; Baggerly *et al* 2004) and biology (Etterson 2009; Rushton 2004; Lammertyn 2000; Hinde and Demétrio 1998).

In this paper we analyze the overdispersion phenomenon in Indonesian, especially in mobile phone industry -- a very dynamic market since Indonesia is a fast growing nation with more than 240 million inhabitants. Understanding the costumer behavior is in this sense a key element in order to win the competition. Logistic regression model is employed to analyze the loyalty of customer (commonly named as 'subscriber'). Such approach is well-known as 'churn analysis' which is to predict 'the churn of subscriber'. A subscriber is considered as churn when he (or she) stops his (or her) subscription from the company. Appropriate approach to handle the subscribers is very important as it is related to spending capital. Overdispersion in the mobile phone industries is common due to non independence of subscribers. The subscribers commonly tend to follow their group or community in using the services, i.e. because of the surrounding or collaborative effect (Ariely *et al* 2004) or social network effect.

This paper will be organized as follows. After the introduction in Section 1, we describe the data and methodology in Section 2. In this section, standard binary logistic model, overdispersion and how to handle overdispersion on binary logistic regression are reviewed. Section 3 provides the empirical results and discussion whereas Section 4 gives summary and concluding remarks.

2. Data and Methodology

2.1. Data

This study used call detail records (CDR) data taken from a well-known Indonesian mobile telecommunication provider to predict probability of a post-paid subscriber being churn at a certain time. This study involved 60 thousand phone numbers (MSISDNs) which were taken randomly from the database of the company. Each MSISDN consists of two explanatory variables namely 'invoice' and 'tenure'. The invoice is monthly bill charged to subscriber related to the use of telecommunication service, while the tenure is a subscription period. Both of variables were stated in the categorical scale, in this case invoice consists of four categories and tenure consists of six categories. Thus, there would be $(4 \times 6) = 24$ binomial observations (or $k = 24$).

The response variable of interest was 'churn' status in the next three months. The value of churn would be one (y=1) if the corresponding MSISDN has not being active or the subscriber was churn. Meanwhile, if the MSISDN was still active, churn status would be zero (y=0). Among 60 thousand numbers, about 96.1% MSISDN still active until the next three months, while the 3.9% has churned.

In addition to estimating the standard logistic regression model, this study also examined the goodness of the model through deviance and Pearson's chi-square statistic. If occurrence of overdispersion is found, William methods would be used to solve the problem. Furthermore, the results obtained by the William method were compared to the standard logistic regression.

2.2. Linear Logistic Regression

Suppose there are k binomial observations written in the form of '*event* (y_i) / *trial* (n_i)' where y_i is number of occurrences of 'success' (event) and n_i is number of replications (trial) on i^{-th} observation (i = 1, 2 , ..., k). Consider that π_i is probability of 'success' and $E(Y_i) = \pi_i n_i$ is expected value for each random variable, Y_i. Logistic regression model which correspond every π_i with p explanatory variables, $X_1, X_2, ..., X_p$ is:

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi})}{1 + \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi})} \tag{1}$$

In the linear form, the equation is often expressed as logit function of 'success', written as follow:

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_p X_{pi} \tag{2}$$

where β_0, β_1, β_2, …, β_p are parameters of logistic regression model. The model is known as Generalized Linear Model (GLM) with logit link function (McCullagh and Nelder 1989).

Parameters are estimated using maximum likelihood method, where its likelihood function is expressed as follows:

$$L(\beta) = \prod_{i=1}^{k} \binom{n_i}{y_i} \pi_i^{y_i} (1-\pi_i)^{n_i - y_i} \tag{3}$$

To obtain parameter estimates, $\hat{\beta}$, $L(\beta)$ or $\log L(\beta)$ must be iterated to reach the maximum value.

2.3. Deviance and Pearson's Chi-Square

Deviance statistic measures lack of fit of the model. In a linear logistic regression model, if fitted value of number of 'success' denoted by $\hat{y}_i$, where $\hat{y}_i = \hat{\pi}_i n_i$, deviance statistics ($D$) is

$$D = 2\sum_{i=1}^{k} \left\{ y_i \log\left(\frac{y_i}{\hat{y}_i}\right) + (n_i - y_i) \log\left(\frac{n_i - y_i}{n_i - \hat{y}_i}\right) \right\} \tag{4}$$

D statistic follows chi-square (χ^2) distribution with $k - p$ degree of freedom, where k is number of binomial observations and p is numbers of parameters in the model. If fitted model is satisfactory, D statistic will be close to its number of degree of freedom. Or in other word, ratio of D statistic to its degree of freedom will be close to unity (Collett 2003).

Other approach to evaluate lack of fit of model is Pearson's chi-square statistic, X^2. It is obtained by the following formula:

$$X^2 = \sum_{i=1}^{k} \frac{(y_i - \hat{\pi}_i n_i)^2}{\hat{\pi}_i n_i (1-\hat{\pi}_i)} \tag{5}$$

X^2 statistic follows chi-square (χ^2) distribution with $k - p$ degree of freedom. If fitted model is satisfactory, X^2 statistic will be close to its number of degree of freedom, or ratio X^2 statistic to its number of degree of freedom will be close to one (Collett 2003).

2.4. Over Dispersion

A logistic model is satisfactory if the ratio of D and X^2 statistic to number of degree of freedom is approximately one. When this ratio is significantly exceed one, assumption of binomial variation is invalid and overdispersion is occurred. Otherwise, a very rare phenomenon called underdispersion, presents when the ratio of D and X^2 statistic to number of degree is smaller than one (Collett 2003). For more advanced method, Lambert and Roeder (1995) introduced a convexity or C plot and relative variance curve and relative variance test to examine overdispersion on Generalized Linear Model (GLM). Overdispersion diagnostics could be done also by a score test (Dean 1992).

Theoretically, overdispersion does not affect the parameter estimates of logistic models, but the standard errors of estimates are attenuated (Kurnia *et al*, 2002). Then, confidence interval of parameter will be narrow, causing the null hypothesis (H_0 : β_i = 0) tend to be rejected. In other word, the explanatory variables affect significantly to the response.

Modeling of overdispersion is often expressed in equation of variance of response variable, Y_i, as follow:

$$\text{var}(Y_i) = \pi_i n_i (1 - \pi_i)\{1 + (n_i - 1)\phi\} \tag{6}$$

Where $\{1 + (n_i - 1)\phi\}$ is overdispersion's scale and ϕ denote inflation factor. When overdispersion is not occurred or very small, ϕ will equal to or approximately zero, so Y_i exactly follows binomial distribution, $B(n_i,p_i)$, and $\text{var}(Y_i) = \pi_i n_i (1 - \pi_i)$ (Collett 2003). However, when overdispersion is existed, ϕ exceeds zero and leads var(Y_i) to be greater than $\pi_i n_i (1 - \pi_i)$. Therefore the actual variance of response variable is greater than the variance calculated from binomial distribution.

2.5. William Method

Parameter estimate of ϕ, denoted by $\hat{\phi}$, is obtained by equating X^2 statistic of the model to its approximate expected value, written as :

$$X^2 = \sum_{i=1}^{k} \frac{w_i (y_i - \hat{\pi}_i n_i)^2}{\hat{\pi}_i n_i (1 - \hat{\pi}_i)} \quad \text{and} \quad E\left(X^2\right) \simeq \sum_{i=1}^{k} w_i (1 - w_i v_i d_i)\{1 + \phi(n_i - 1)\} \tag{7}$$

where $v_i = \pi_i n_i (1 - \pi_i)$, w_i is the weight and v_i is diagonal element of the variance-covariance matrix of the linear predictor, $\hat{\eta}_i = \sum \hat{\beta}_j x_{ji}$. The value of X^2 statistic depends on $\hat{\phi}$, so iteration process is needed to find optimum value. This procedure was firstly introduced by William (1982), and then called William method. The algorithm of William method is described as follow.

1. Assumed $\phi = 0$, calculate parameter estimate of logistic regression model, $\hat{\beta}$, using maximum likelihood method. Calculate the X^2 statistics of fitted model.
2. Compared X^2 statistics to $\chi^2_{(k-p)}$ distribution. If X^2 statistic is too large, conclude that $\phi > 0$ and calculated initial estimates of ϕ using following formula :

$$\hat{\phi}_0 = \frac{X^2 - (k - p)}{\sum_{i=1}^{k}\{(n_i - 1)(1 - v_i d_i)\}} \tag{8}$$

3. Using initial weight $w_{i0} = \left[1 + (n_i - 1)\hat{\phi}_0\right]^{-1}$, recalculate the value of $\hat{\beta}$ dan X^2 statistic.
4. If X^2 statistic closes to its number of degree of freedom, $k - p$, estimated value of ϕ is sufficient. If not, re-estimate ϕ using following expression:

$$\hat{\phi}_1 = \frac{X^2 - \sum_{i=1}^{k}\left[w_i(1 - w_i v_i d_i)\right]}{\sum_{i=1}^{k}\left\{w_i(n_i - 1)(1 - w_i v_i d_i)\right\}} \tag{9}$$

If X^2 statistic remain large, return to step 3 until optimum value of estimated ϕ obtained.

Once ϕ has been estimated by $\hat{\phi}$, $w_i = 1/[1 + (n_i - 1)\phi]$ could be used as weights in fitting new model (Collett 2003; William 1982).

3. Results and Discussion

3.1. Result of Standard Logistic Regression

Setiabudi and Saefuddin (2009), conducted a logistic regression model to estimate the probability of each subscriber being churn based on invoice and tenure following linear form of

$$\text{logit } (\pi_{jk}) = \beta_0 + \beta_1 * invoice + \beta_2 * tenure$$

Invoice and tenure were recorded in categorical scales; hence these two variables were transformed into dummy variables. The parameterization was done using reference method (*see* SAS Institute, 2009). In this case, the last category of each explanatory variable was reference for other categories. Thus, the 4th category of invoice i.e. subscribers who spent more than IDR150,000 a months in using telecommunication services, written as $invoice_4$, would be reference for $invoice_1$, $invoice_2$, and $invoice_3$; while $tenure_6$ i.e. subscribers who have been subscribed for 60 months or more, would be a reference for the $tenure_1$ until $tenure_5$. Table 1 shows completely how explanatory variables were categorized.

Using standard logistic regression, parameter estimates of model were listed in Table 2, and evaluations to goodness of this model were in Table 3.

Table 1: Explanatory variables and their dummy variables

Explanatory Variable	Dummy	Description
Invoice	1	0 – IDR 25,000 / month
	2	IDR 25,000 – IDR 50,000 / month
	3	IDR 50,000 – IDR 150,000 / month
	4	> IDR 150,000 / month *
Tenure	1	0 – 3 months
	2	3 – 6 months
	3	6 – 12 months
	4	12 – 24 months
	5	24 – 60 months
	6	> 60 months *

IDR = Indonesian currency *) Reference

Table 2: Parameter estimates of model using standard logistic regression

Explanatory Variable	$\hat{\beta}$	$SE(\hat{\beta})$	*df*	*p-value*
Intercept	–5.708	0.107	1	<.0001
Invoice			3	
$Invoice_1$	0.816	0.097	1	<.0001
$Invoice_2$	1.299	0.062	1	<.0001

Table 2: Parameter estimates of model using standard logistic regression - continued

$Invoice_3$	0.399	0.063	1	<.0001
Tenure			5	
$Tenure_1$	3.710	0.109	1	<.0001
$Tenure_2$	3.038	0.113	1	<.0001
$Tenure_3$	2.147	0.121	1	<.0001
$Tenure_4$	2.481	0.113	1	<.0001
$Tenure_5$	0.724	0.119	1	<.0001

The value of deviance statistic was 182.849 with 15 degree of freedom and the value of Pearson’s chi-square was 186.919 with 15 degree of freedom, indicated the churn data contained overdispersion. This phenomenon was reasonable since the behavior of a person to subscribe to a mobile phone service provider depend on, or affected by, other persons or community in the same groups or segments. Therefore the response variable was not independent or there was correlation between subscribers within binomial observation.

Table 3: Goodness of fit of standard logistic regression model based on deviance and Pearson’s chi-square statistic

Statistic	***Value***	***df***	***Value/df***	***p-value***
Deviance	182.849	15	12.190	<.0001
Pearson’s chi square	186.919	15	12.461	<.0001

3.2. Result of Weighted Logistic Regression Model Using William Method

Since overdispersion occurred, William method will be used to estimate weighted logistic regression. Through iteration process, estimated parameter of inflation factor, $\hat{\phi}$, was found to be 0.008657, thus overdispersion scale or weights was $w_i = 1/[1+0.008657(n_i-1)]$, where n_i denotes number of subscribers in the *i*-th group. Using weights of w_i we obtained new logistic model, namely weighted logistic regression model using William method, whose parameter estimates presented on Table 4. Furthermore, the goodness of this model according to deviance and Person’s chi-square statistic was given by Table 5.

Table 4: Parameter estimates of weighted logistic regression model using William method

Explanatory Variable	$\hat{\beta}$	$SE(\hat{\beta})$	***df***	***p-value***
Intercept	–5.296	0.516	1	<.0001
Invoice			3	
$Invoice_1$	0.881	0.253	1	0.0005
$Invoice_2$	1.178	0.235	1	<.0001
$Invoice_3$	0.392	0.257	1	0.1275
Tenure			5	
$Tenure_1$	3.229	0.498	1	<.0001
$Tenure_2$	2.616	0.507	1	<.0001
$Tenure_3$	1.787	0.529	1	0.0007
$Tenure_4$	2.178	0.517	1	<.0001
$Tenure_5$	0.612	0.608	1	0.3138

Table 5: Goodness of weighted logistic regression model with William method based on deviance and Pearson’s chi-square statistic

Statistic	***Value***	***df***	***Value/df***	***p-value***
Deviance	15.443	15	1.030	0.4200
Pearson’s chi square	15.000	15	1.000	0.4514

The William method corrected the goodness of fit of model indicated by the value of deviance and Pearson's chi square statistic. The value of deviance (D=15.443) and Pearson's chi-square (X^2=15.443) which very close to their degree of freedom (df=15).

In the weighted logistic regression model with William method, $invoice_3$ and $tenure_5$ did not significantly differ from zero (Table 4). Whereas, parameter estimates of the standard logistic regression model were all significant at 95% confidence interval (Table 1). Since a standard logistic regression model did not accommodate overdispersion, the value of SE would be underestimated, thus decision making about relationship of explanatory variables and corresponding response variable would be invalid. In other hand, William method provided a larger *SE* of estimates, as a result of adjustment to extra binomial variation within data (Kurnia *et al*, 2002).

3.3. Best Model Selection

When parameter estimate of $invoice_3$ did not significantly differ, the model can be simplified by joining $invoice_3$ and its reference, $invoice_4$, thus all subscribers who had invoice more than IDR50,000 were classified as one category. The same way applied for $tenure_5$ and $tenure_6$. Thus, the invoice would consist of three categories while the tenure consists of five categories where the last category of each variable was reference of other categories.

Using the modified categories, standard logistic regression was implemented to estimate a new model with D=164.571 on df=8 and X^2=163.129 on 8. Thus, overdispersion was still existed and the model might be not appropriate. Hence, the William method applied, and then obtained $\hat{\phi} = 0.009347$ which equivalent with the weights of $w_i = 1/[1+0.009347*(n_i - 1)]$. Since D=8.649 on df=8 and X^2=8.000 on 8, the model whose parameter estimates presented on Table 6 was satisfactory.

Table 6: The modified model of weighted logistic regression using William method

Explanatory Variable	$\hat{\beta}$	$SE(\hat{\beta})$	***df***	***p-value***
Intercept	-4.5780	0.4666	1	<.0001
Invoice			2	
$Invoice_1$	0.5578	0.2450	1	0.0228
$Invoice_2$	0.8919	0.2248	1	<.0001
Tenure			4	
$Tenure_1$	2.7949	0.4602	1	<.0001
$Tenure_2$	2.1623	0.4732	1	<.0001
$Tenure_3$	1.2832	0.5055	1	0.0111
$Tenure_4$	1.8528	0.4825	1	0.0001

The model on Table 4 called original model, while the modified model is on Table 6. The later model produced satisfactory result and looked simpler than the original one. Practically the modified model might recommend more efficient guidance on how the marketing strategies have to be performed to minimize the occurrence of subscribers churn. In addition, the AIC (*Akaike Information Criterion*) statistic was employed to obtain the best between the original and modified model following Agresti (2007) and McCullagh and Nelder (1989). Based on the AIC statistics, the modified model was better than the original one, i.e. the AIC statistics of modified model was 855.54 while the AIC statistics of the original one was 1183.97.

4. Concluding Remarks

Overdispersion problem cause the value of standard error of parameter estimates to be underestimated which then yields significant effect of explanatory variables. Nonetheless, overdispersion does not produce biased estimates. The William method adjusts the standard error of parameter estimate to the occurrence of overdispersion. This approach is appropriate in the binary logistic regression.

In the case of churn analysis in the Indonesian telecommunication industry, model with overdispersion will recommend a misleading marketing strategy. It was shown that if overdispersion ignored, statistical recommendations slightly distorted. In the example, standard logistic regression found that all the category of invoice and tenure looks significant. After correction to overdispersion, there were several categories of invoice which were actually in one category. Likewise, there were several categories of tenure which were actually in one category. Relation to marketing strategy, other than this distortion will cause wastage due to excessive cost allocation and inappropriate strategy. Correcting the overdispersion will provide simple and appropriate marketing campaign efficiently and effectively.

References

[1] Agresti, A. (2007) *An Introduction to Categorical Data Analysis,* Second edition. John Wiley & Sons, New Jersey, US.

[2] Ariely, D., J.G. Lynch and M. Aparicio (2004) Learning by collaborative and individual-based recommendation agents. *Journal of Consumer Psychology* 14(1): 81–94.

[3] Baggerly, K.A., L. Deng, J.S. Morris and C.M. Aldaz (2004) Overdispersed logistic regression for SAGE: modelling multiple groups and covariates. *BMC Bioinformatics* 5:144. http://www.biomedcentral.com/1471-2105/5/144

[4] Cheong, P.W., A.A. Jemain and N. Ismail (2008) Practice and pricing in non-life insurance: the malaysian experience. *Journal of Quality Measurement and Analysis* 4(1): 11-24.

[5] Collett, D. (2003) *Modelling Binary Data*, Second edition. Chapman & Hall/CRC, London, UK.

[6] Dean, C.B. (1992) Testing for overdispersion in poisson and binomial regression models. *Journal of the American Statistical Association* 87(418) : 451-457.

[7] Etterson, M.A., G.J. Niemi and N.P. Danz (2009) Estimating the effects of detection heterogeneity and overdispersion on trends estimated from avian point counts. *Ecological Applications* 19(8): 2049–2066.

[8] Hajarisman, N. and A. Saefuddin (2008) The beta-binomial multivariate model for correlated categorical data. *Jurnal Statistika* 8(1): 61–68.

[9] Handayani, D. and A. Kurnia (2006) *Mixed effect model approach for logistic regression model with overdispersion.* Proceeding of ICoMS-1, Bandung 19-21 June 2006.

[10] Hinde, J. and C.G.B. Demétrio (1998) Overdispersion: models and estimation. *Computational Statistics and Data Analysis* 27: 151-170.

[11] Ismail, N. and A.A. Jemain (2005) Generalized poisson regression: an alternative for risk classification. *Jurnal Teknologi* 43(C): 39–54.

[12] Ismail, N. and A.A. Jemain (2007) *Handling overdispersion with negative binomial and generalized poisson regression models.* Proceeding of Casualty Actuarial Society Forum, Winter 2007 Edition : 103 – 158.

[13] Kurnia, A., A. Saefuddin and E. Sutisna (2002) *Overdispersi dalam regresi logistik.* Proceeding of Seminar Nasional Statistika, Bogor, 28 September 2002 : 11-17.

[14] Lambert, D. and K. Roeder (1995) Overdispersion diagnostics for generalized linear models. *Journal of the American Statistical Association* 90(432): 1225-1236.

[15] Lammertyn, J., M. Aerts, B.E. Verlinden, W. Schotsmans and B.M. Nicolaï (2000) Logistic regression analysis of factors influencing core breakdown in 'Conference' pears. *Postharvest Biology and Technology* 20: 25–37.

[16] Lu, J., J.K. Tomfohr and T.B. Kepler (2005) Identifying differential expression in multiple SAGE libraries: an overdispersed log-linear model approach. *BMC Bioinformatics* 6:165. http://www.biomedcentral.com/1471-2105/6/165

[17] McCullagh, P., and J.A. Nelder (1989) *Generalized Linear Models*, Second ed. Chapman and Hall, London, UK.
[18] Rushton, S.P., S.J. Ormerod and G. Kerby (2004) New paradigms for modelling species distributions? *Journal of Applied Ecology* 41: 193–200.
[19] SAS Institute Inc. (2009) *SAS/STAT® 9.2 User’s Guide*, Second Edition, SAS Institute Inc., Cary, NC, US.
[20] Setiabudi, N.A. and A. Saefuddin (2009) *Customer segmentation and churn analysis in the mobile telecommunications industry*, Book of Abstract ICCS-X Conference, Cairo, December 20–23, 108.
[21] Williams, D. A. (1982) Extra-binomial variation in logistic linear models. *Applied Statistics* 31: 144–148.