MATHEMATICAL MODELING AND COMPUTING, Vol. x, No. x, pp. 1–7 (2017)

Score Tests Statistics for Detecting Extra-Poisson Variation:A Review

Hassan S. U.[1], Mohammed H. N.[2]

[1,2]*Dept. of Statistics*
College of Administration and Economics
University of Al-Qadisiya
Aljamea'a Str., Diwanyiah,IRAQ

(Received 12 August 2022)

When the empirical variance in the data is higher than that expected by the model, count data examined under a Poisson assumption or data in the form of proportions under a binomial assumption sometimes demonstrate overdispersion.Underestimating standard errors of covariate effects is one of the implications of neglecting overdispersion.This paper involves a selective overview of score test statistics that are existing literature and used to detect the extra-Poisson variation.The methodologies of these tests have been covered to aid the statistics practitioners in understanding how to deal with count data and identify whether there is extra-Poisson variation.The baseline formula of all score test statistics had been derived from the negative binomial model. Subsequently, the proposed adjusted test statistics have been presented to approximate the distribution of these tests to standard normal distribution. However, the test result is unreliable unless the correct conditional mean and variance are obtained.

Keywords: *Overdispersion, Poisson, Score test, NB-1, NB-2*

2000 MSC: 62Jxx

UDC: Up to six UDC numbers

1. Introduction

The appeal of Poisson regression models for analyzing count data non-negative integer numbers seems to be widely documented in economics and medical literature. That is due to the Poisson regression of the count model being basic, and the response variable being discrete. The Poisson distribution is well dealing with discrete data and it is assumed that the conditional mean of the response variable is equal to the conditional dispersion (Frome et al.;1973 and Frome;1983). In reality, finding such conditions in real data is not easy, particularly the discrete data that perhaps create a heterogeneity problem. However, when the conditional dispersion in the data exceeds the basic dispersion under a hypothesized model, the phenomenon, which has come to be known as overdispersion, occurs. The conditional mean function's covariates may not adequately reflect overlooked or unobserved heterogeneity, which can lead to overdispersion. As a result, the test statistics obtained from the models will be inadequate and the standard errors, often computed by the maximum likelihood technique, will be biased. That is due the Maximum likelihood estimate (MLE) approach cannot provide parameter estimates when existing formal model cannot adequately specify the probability distribution of a random variable. The Pearson or deviance residuals of the model fitted are primarily too large, resulting in inadequate goodness-of-fit statistics (Manton et al 1981; Breslow 1984; Brillinger 1986).

The unobserved heterogeneity in the data is not the only reason for overdispersion. Due to the appearance of outliers, the impact of other variables results in dependency between two probability events or more, in which the response variable was inflated by excess zeros (Hilbe and Hardin, 2014). Unfortunately, the presence of overdispersion may lead to misleading inferences due to the so-called upward bias and downward bias. The first one can be observed with independent variables since it increases with its significance. Meanwhile, the second bias performs underestimated standard deviation of pa-

rameter estimations (Ismail and Jemain, 2007).
Overdispersion has been studied in the statistical literature since Student (1919) commented, while Fisher (1950) proposed a goodness-of-fit statistic for evaluating the Poisson distribution's, efficacy in the single sample case, that is, when the counts are taken to be an independent variable with a common mean from a Poisson distribution. It is beyond doubt that when the dispersion parameter is not considered in the formulated model, the outcomes will not be reliable. Apart from that, Hilbe and Hardin (2014) classified overdispersion into two types, real and apparent. The source of real dispersion is either zero inflation or correlation, while outliers, missing predictors, and inappropriate link functions cause the apparent overdispersion. Therefore, the concentration of authors in the statistical literature tends to adjust the model to fit any type of overdispersion. It is evident that the decision to adjust the Poisson model for fitting data is based on occurring underdispersion or overdispersion.
As mentioned above, outliers are one reason for apparent overdispersion that motivated the statistics practitioners to use the quasi-Poisson or negative binomial model in their practical research. It is well known that there are two defined types of outliers in the regression model that appears with the response variable, vertical outliers and leverage points, which occur in design matrix X. Both types of outliers are assumed to influence estimating the parameter of Poisson distribution, which perhaps led to misleading inferences. This manuscript discusses the effect of leverage points on the decision-making to apply specific overdispersion or underdispersion models based on the test outcome. For instance, the outcome of the test may refer to the overdispersion case, while the reality, it is the underdispersion case and vis versa, is correct.
In general, the approach of the quasi-likelihood function is that rather than determining a probability distribution for the data, the variance is considered a function of the mean. Hence, it specifies the relationship between the mean and the variance by allowing the variance to be inflated as follows, $Var(Y) = \Phi\dot{\lambda}$ where Φ is the scale parameter or dispersion parameter.
Assume that y is a discrete random variable and $\lambda = \mu(X)$, then the estimated scale parameter can be computed by the formula of generalized χ^2 statistic,

$$\hat{\Phi} = \frac{\chi^2}{n-p} = \frac{1}{n-p}\sum_{i=1}^{n}\frac{(Y_i - \hat{\lambda}_i)^2}{\hat{\lambda}_i} \tag{1}$$

where $\lambda = \mu(X_i)$. Inflating the estimated variance by $\hat{\Phi}$ has motivated the authors to propose several variance forms by combining the variance function and scale parameter for modeling overdispersion. Therefore, the statistic $\hat{\Phi}$ is approximately unbiased estimate to calculate the ratio of the Pearson χ^2 to its degree of freedom for detecting overdispersion (Rodriguez, 2015). When the Poisson model's claim of equal mean and variance is true, the scale parameter, $\hat{\Phi} \approx 1$. The overdispersion problem has garnered the authors' attention. Some options are proposed to estimate the statistic $\hat{\varphi}$ in the literature, such as score statistic, likelihood ratio, Pearson statistic and Wald statistic. This paper is designed in the following manner. Subsequent to the introduction, the score test from the negative binomial is given in Section 2. Meanwhile, Section 3 gives the score tests, standardized and adjusted its statistics. Finally, Section 4 provided the discussion.

2. Score Test from Negative Binomial Model

Let $Y_i \sim Poisson(\nu_i\lambda_i)$ be a mixed distribution, where $Y_i$is a response variable, ν_i be iid random variables having only positive values, such that $E(\nu_i) = 1$ and $Var(\nu_i) = \tau$, in this case Collings and Margolin (1985) formulate $Var(Y_i/x_i) = \lambda + \tau\lambda^2$ which equivalents to the traditional variance of negative binomial 2. Later, Cameron and Trivedi (1986) rewrite this formula of inflated variance as,$Var(Y) = \lambda + \tau\lambda^q$. So, one can use hypotheses test rely on the value of τ, the null hypothesis Poisson model where $H_0 : \tau\alpha = 0$ versus $H_1 : \tau > 0$ the extra-Poisson model for alternative hypothesis. When ν_i's are distributed Gamma the distribution of $Y_i's$ should be negative binomial. If $q = 1$, it results in

$Var(Y) = \lambda(1+\tau)$ implying that the variance is linearly proportional to mean which matches with Negative Binomial-1 (NB-1) variance and equivalent to $\Phi\dot{\lambda}$ where $\Phi = (1+\tau)$.
Let $\ell(\tau)$ is the log-likelihood of $y_i \sim NB(\lambda_i, \tau\alpha)$ as a function of τ.

$$\ell(\tau) = \sum_{i=0}^{n}[y_i \log \tau\lambda_i - (y_i + \tau^{-1}) \log(1+\tau\lambda_i) + \sum_{t=0}^{y_i-1} \log(t+\tau^{-1}) - \log y_i!] \tag{2}$$

The third term of (2) can be rewrite as

$$\begin{aligned} \sum_{t=0}^{y_i-1} \log(t+\tau^{-1}) &= \sum_{t=0}^{y_i-1} \log \tau^{-1}(1+\tau t) \\ &= \sum_{t=0}^{y_i-1} \log \tau^{-1} + \sum_{t=0}^{y_i-1} \log \tau^{-1}(1+\tau t) \\ &= -y_i \log \tau + \sum_{t=0}^{y_i-1} (1+\tau t) \end{aligned} \tag{3}$$

Substitute this outcome instead of $\sum_{t=0}^{y_i-1} \log(t+\tau^{-1})$ in equation (2)

$$\ell(\tau) = \sum_{i=0}^{n}\Big[y_i \log \tau\lambda_i - (y_i + \tau^{-1}) \log(1+\tau\lambda_i) - y_i \log \tau + \sum_{t=0}^{y_i-1} (1+\tau t) - \log y_i!\Big]$$

$$\ell(\tau) = \sum_{i=0}^{n}\Big[y_i \log \lambda_i - (y_i + \tau^{-1}) \log(1+\tau\lambda_i) + \sum_{t=0}^{y_i-1} (1+\tau t) - \log y_i!\Big]$$

$$\ell(\tau) = \sum_{i=0}^{n}\Big[y_i \log \lambda_i - (y_i \log(1+\tau\lambda_i) - \tau^{-1}\Big\{\tau\lambda_i - \frac{(\tau\lambda_i)^2}{2} + \cdots\Big\} + \sum_{t=0}^{y_i-1} (1+\tau t) - \log y_i!\Big]$$

$$\ell(\tau) = \sum_{i=0}^{n}\Big[y_i \log \lambda_i - (y_i \log(1+\tau\lambda_i) - \lambda_i - \frac{\tau\lambda_i^2}{2} + \cdots + \sum_{t=0}^{y_i-1} (1+\tau t) - \log y_i!\Big] \tag{4}$$

From the (4) one can conclude that $\ell(0) = y_i \log \lambda_i - \lambda_i - \log y_i!$
According to the traditional Neyman-Pearson concept, the strongest test rejects H_0 for large values of $\ell(\tau) - \ell(0)$ if we take into account that λ_i's is known.

$$\ell(\tau) - \ell(0) = \sum_{i=1}^{n}\sum_{t=0}^{y_i-1}\Big[\log(1+\tau t) - y_i \log(1+\tau t)\Big]$$

Even though the λ_i's are available, a consistently most powerfull test of H_0 vs H_1 doesn't somehow exist because the test is not independent of τ. A locally most strong test for H_0 vs H_1 can indeed be achieved as a result. Based on the approach of Ferguson (2014) the local strongest test rejects H_0 for large values of the outcome $\frac{\partial \ell}{\partial \tau}$ at $\alpha = 0$

$$\frac{\partial \ell}{\partial \tau} = \sum_{i=1}^{n}[-\frac{y_i\lambda_i}{(1+\tau\lambda_i)} - \Big\{\frac{\lambda_i^2}{2} + \frac{\lambda_i^3}{3} - \cdots\Big\} + \sum_{t=0}^{y_i-1} \frac{t}{(1+\tau t)}$$

$$\frac{\partial \ell}{\partial \tau}\,|_{\tau=0} = \sum_{i=1}^{n}\Big[-y_i\lambda_i + \frac{\lambda_i^2}{2} + \sum_{t=0}^{y_i-1} t\Big]$$

$$= \sum_{i=1}^{n} \Big[\frac{\lambda_i^2 - 2y_i\lambda_i}{2} - \frac{y_i(y_i-1)}{2} \Big]$$

$$= \sum_{i=1}^{n} \Big[\frac{\lambda_i^2 - 2y_i\lambda_i + y_i^2}{2} - \frac{y_i}{2} \Big] = \frac{1}{2} \sum_{i=1}^{n} \Big[(y_i - \lambda_i) - y_i \Big] \tag{5}$$

It is considered to be the fundamental overdispersion test which is the so-called Partial Score Statistic and is denoted as (S),

$$S = \frac{1}{2} \sum_{i=1}^{n} \Big[(y_i - \lambda_i) - y_i \Big] \tag{6}$$

3. Score Tests Statistics

Numerous approaches have been presented in the literature to deal with extra-Poisson variation or overdispersion and its impact on the Poisson-based tests (see: Lawless 1987a,b ;Alam, 2015and Cox 1983). such as Pearson chi-square, deviance, and score test statistic. In this section we present some of familiar score tests statistics which were widely used by practitioners of statistics.

3.1. Standardized Score Test Statistic

Dean and Lawless (1989) pointed out that the distribution of S can be roughly calculated using the classic maximum likelihood large-sample theory. Thus, they demonstrated that when $(n \longrightarrow \infty)$ the $\frac{S}{\sqrt{\frac{1}{2}\sum_{i=1}^{n}\hat{\lambda}_i^2}}$ is an asymptotically comparable standardized statistic when the distribution converges to a standard normal. Hence, the suggested form is based on the accuracy estimate of mean and variance.

a) When $E(Y_i) = \lambda_i$ and $Var(Y_i) = \lambda_i + \tau\lambda_i^2$ and τ is small, the score test statistic can be written as,

$$S_1 = \frac{\sum_{i=1}^{n}\{(Y_i - \hat{\lambda}_i) - Y_i\}}{\sqrt{2\sum_{i=1}^{n}\hat{\lambda}_i^2}} \tag{7}$$

b) When $E(Y_i) \simeq \lambda_i$ and $Var(Y_i) = \lambda_i + \tau\lambda_i^2$ it can be replaced $\hat{\lambda}_i$ instead of Y_i and then rewrite S_1 as

$$S_1^* = \frac{\sum_{i=1}^{n}\{(Y_i - \hat{\lambda}_i) - \hat{\lambda}_i\}}{\sqrt{2\sum_{i=1}^{n}\hat{\lambda}_i^2}} \tag{8}$$

c) When $E(Y_i) = \lambda_i$ and $Var(Y_i) = \lambda_i(1+\tau)$ and τ is large enough, the score test statistic can be written as,

$$S_2^* = (2n)^{-\frac{1}{2}} \sum_{i=1}^{n} \Big\{ \frac{(Y_i - \hat{\lambda}_i)^2 - Y_i}{\hat{\lambda}_i} \Big\} \tag{9}$$

All tests have a conventional normal asymptotic distribution. The statistic makes it evident that higher fluctuation between observed and expected values would raise the score statistic's value, implying overdispersion brought on by data heterogeneity and other variables. If score statistic S_1 or S_1^* is bigger than the 95th percentile of the N(0,1) distribution, we may reject the hypothesis of equidispersion at a 0.05 level of significance using a one-sided test (Payne et al., 2018). Collings and Margolin (1985) concentrated on preparing powerful tests against alternative hypotheses based on a specific mixed Poisson model. They derived their score tests from special cases of the negative binomial model and derivatives of the asymptotic distribution theory of these tests. To be more specific, three score tests were suggested to rely on the $\bar{Y}$ either constant, single covariate with regression line computed without intercept, and finally takes fixed values according to the one-way layout. Dean 1989 considered these tests are just extensions to Fisher's (1950) and Maron's (1971) tests when the observations are identically independently distributed processes, following the Poisson distribution. Apart from that, they provide another approach to score tests, which is denoted as S_2 and its limiting distribution is obtained by both Fisher (1950) and Collings-Margolin (1985) dispersion statistics,

$$S_2 = \sum_{i=1}^{n} \frac{(Y_i - \hat{\lambda}_i)^2}{\bar{Y}} \tag{10}$$

where $\bar{Y} = (n)^{-1} \sum_{i=1}^{n} Y_i$ Dean and Lawless (1989) found special cases of the limiting distributions from S_2 equivalent to Fisher (1950) and Collings and Margolin (1985) score tests.
Suppose that n is to be fixed and $\lambda_i^{\prime}s \longrightarrow \infty$ such that λ_i/λ_+converges to a positive constant. $S_3 = n\lambda_+^{-1}(Y - \hat{\lambda})^{\prime}(Y - \hat{\lambda})$, as another form of S_2, where Y and $\hat{\lambda}$ are two vectors and $\lambda_+ = \sum_{i=1}^{n} Y_i$. Since $\lambda_+^{-1} \sum_{i=1}^{n} Y_i \xrightarrow{\text{P}} 1$ when $\tau = 0$, both S_2 and S_3 statistics have the same limiting distribution, and then $\sqrt{\lambda_+}(Y - \hat{\lambda}) \xrightarrow{\text{D}} N(0, \Delta)$ where $\Delta = \lambda_+^{-1}\sqrt{W}(I - H)\sqrt{W}$ is the asymptotic covariance matrix. $W = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and $H = \sqrt{W}X(X^{\prime}WX)^{-1}X^{\prime}\sqrt{W}$ is the leverage matrix of Poisson regression. Other than that, the estimation of regression coefficients in the presence of leverage points lead to $E\{\sum_{i=1}^{n}(Y_i - \hat{\lambda}_i)^2\} < \sum_{i=1}^{n} \lambda_i$. We can write $E\{\sum_{i=1}^{n}(Y_i - \hat{\lambda}_i)^2\} = (1 - h_{ii})\lambda_i$, where h_{ii} is the i^{th} diagonal elements of H. Thus, the null distribution of S_1 and S_2 are from $\sqrt{\lambda_+}(Y - \hat{\lambda}) \xrightarrow{\text{d}} N(0, \Delta)$ that of a linear combination of independent $\chi^2_{(1)}$ random variables (Rao,1973). Particularly, $S_2 \xrightarrow{\text{converge}} \sum_{i=1}^{n} \pi_i \gamma_i$, where $\pi_i{}^{\prime}s \sim \chi^2_{(1)}$ independent random variables, and $\gamma_i{}^{\prime}s$ are the n eigenvalues of Δ.

3.2. Adjusted Standardized Score Test Statistic

As a result of slowness for S_1, S_1^*, S_2^* distribution to be approximate as a standard normal distribution when $n \longrightarrow \infty$, Dean and Lawless (1989) and Dean and Lundy (2016) adjusted these statistics with the same asymptotic properties to find the significant levels (p-value) and critical values.
Let us begin to adjusted S_1 statistics to reflect a more suitable distributional assumption as,

$$S_\alpha = \frac{\sum_{i=1}^{n}(Y_i - \hat{\lambda}_i)^2 - Y_i + h_{ii}\hat{\lambda}_i}{\sqrt{2\sum_{i=1}^{n} \hat{\lambda}_i^2}} \tag{11}$$

Using the limiting distribution of S_2 when $\lambda_i{}^{\prime}s \longrightarrow \infty$, the previous authors inserted $\hat{\lambda}_i$ into Δ matrix and thus into eigenvalues,γ_i to approximate $Pr\{\ S_2 > s\}$. Then, compute a probability of $\chi^2_{(1)}$ independent random variables. For this purpose, the authors used two moments $c\chi^2_{(d)}$ approximation for this procedure. The mean and variance of limiting distribution for S_2 are,

$$(\bar{S_2}) = n.tr(\Delta) = n\lambda_+^{-1}(1 - h_{ii})\lambda_i$$

$$Var(S_2) = 2n^2 tr(\Delta^{\prime}\Delta) = 2n^2\lambda_+^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j (1 - h_{ii})^2$$

Selecting approximation $S_2\ c\chi^2_{(d)}$ such that the asymptotic mean and variance are correct allows estimating c and d as follows,

$$\hat{c} = n\lambda_+^{-1} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \hat{\lambda}_i \hat{\lambda}_j (1 - h_{ii})^2}{\sum_{i=1}^{n} \hat{\lambda}_i (1 - h_{ii})}$$

$$\hat{d} = \frac{\left\{\sum_{i=1}^{n} \hat{\lambda}_i (1 - h_{ii})\right\}^2}{\sum_{i=1}^{n} \sum_{j=1}^{n} \hat{\lambda}_i \hat{\lambda}_j (1 - h_{ii})^2}$$

When $\hat{d} \geqslant 10$, it is preferable to transfer $\chi^2_{(.)}$ variables using Wilson-Hilferty transformation. However, this approximation gives,

$$S_b = (4.5\hat{d})^{1/2}\left\{ \sqrt[3]{\frac{S_2}{\hat{d}\hat{c}}} + \frac{2}{9\hat{d}-1} \right\} \tag{12}$$

This formula involves standard normal distribution when $\hat{c} \doteq 1$ and $\hat{d} \doteq n-p$. S_b is adequate statistics when $\hat{d}$ is large enough, otherwise $S_2 \sim c\chi^2_{(d)}$ is more appropriate than it.
Dean and Lundy (2016) modified the S_1^* and S_2^* for testing the null hypothesis $\tau = 0$ two times, the first one for the model with $Var(Y_i) = \lambda_i + \tau\lambda_i^2$ and other for the model with $Var(Y_i) = \lambda_i(1+\tau)$ respectively, as

$$S_1^{**} = \frac{\sum_{i=1}^{n} \left\{ (Y_i - \hat{\lambda}_i)^2 - (1-h_{ii})\hat{\lambda}_i \right\}}{\sqrt{2\sum_{i=1}^{n} \hat{\lambda}_i^2}} \tag{13}$$

$$S_2^{**} = \frac{1}{(2n)^{\frac{1}{2}}} \sum_{i=1}^{n} \left\{ \frac{(Y_i - \hat{\lambda}_i)^2 - (1-h_{ii})\hat{\lambda}_i}{\hat{\lambda}_i} \right\} \tag{14}$$

However, the distribution of $S_1^{(**)}$ and $S_2^{(**)}$ asymptotically $n \longrightarrow \infty$ distributed as $N(0,1)$ and converge to normality faster than S_1^* and S_2^*. It is notable that when $\lambda_i \longrightarrow \infty$ asymptotic for fixed positive constant n the $S_2^{(**)}$ can be written as follows,

$$\widetilde{S_2^*} = \sum_{i=1}^{n} \left[\frac{(Y_i - \hat{\lambda}_i)^2}{\hat{\lambda}_i} \right] \tag{15}$$

Which is equivalent to Pearson statistic in equation (1) with $\chi^2(n-p)$ as a limiting distribution.

4. Discussion

All the previous overdispersion test postulates are derived under the assumption that the regression specification is accurate. According to that, two alternative hypotheses should be considered, $\lambda_i = \lambda_i(x_i, \beta)$ is not accurate specified or $Var(y) = \Phi.\lambda$. Otherwise, the apparent dispersion is present, for instance, missing covariates (for more details, see: Hilbe and Hardin, 2014; Dean and Lundy, 2016). However, unless the λ_i had been pretty adequately described, we would not interpret a considerably large Pearson statistic as suggesting overdispersion in a generalized linear model. On the contrary, the apparent overdispersion could be present, and one can look at the reasons that are result-in, as it may be methodical lack-to-fit, missing variables, it might not be suitable to use the mean functional form and others.

Reference

[1] "Student." An explanation of deviations from PoissonвЂ™s law in practice. Biometrika **12**, (1919).

[2] Alam N. Detecting Overdispersion in Count Data: Comparison of Tests. (Doctoral dissertation, East West University), (2015).

[3] Breslow N. Extra-Poisson Variation in Log-Linear Models., J R Stat Soc Ser C Appl Stat **33**,(1984)

[4] Brillinger D. R. The Natural Variability of Vital Rates and As- sociated Statistics (with discussion). Biometrics **13**1, (1986).

[5] Collings B.J., Margolin BH Testing goodness-of-fit for the Poisson assumption when observations are not identically distributed. J. Am. Stat. Assoc **80(930)**,(1985).

[6] Cox D. R. Some Remarks on Overdispersion. Biometrika, **70**11 (1983).

[7] Dean C.,Lawless J.F. Tests for detecting overdispersion in Poisson regression models. J. Am. Stat. Assoc, **84** (1989). Dean, C.B. (1992) Testing for overdispersion in Poisson and binomial regression models. J. Am. Stat. Assoc. 87, 451вЂ“457.

[8] Dean C. Testing for overdispersion in Poisson and binomial regression models. J. Am. Stat. Assoc, **87**(1992). Dean, C.B. and Lundy, E.R (2016), Overdispersion, Encyclopedia of Clinical Trials, Wiley, N. Balakrishnan, P. Brandimarte, B. Everitt, G. Molenberghs, W. Piegorsch and F. Ruggeri (eds.), Wiley StatsRef: Statistics Reference Online. 19, DOI: 10.1002/9781118445112.stat06788.pub2.

[9] Dean C., Lundy E. R. Overdispersion, Encyclopedia of Clinical Trials. Wiley, N. Balakrishnan, P. Brandimarte, B. Everitt, G. Molenberghs, W. Piegorsch and F. Ruggeri (eds.), Wiley StatsRef: Statistics Reference Online. 19, DOI: 10.1002/9781118445112.stat06788.pub2. , **19** (2016).

[10] Ferguson T. S. Mathematical statistics: A decision theoretic approac. Academic press.вЂЦ; (2014).

[11] Fisher R. A. The Significance of Deviations From Expectation in a Poisson Series. Biometrics., **6** (1950).

[12] Frome E. L. The Analysis of Rates using Poisson Regression Models. Biometrics, **39**, (1983).

[13] Harris T., Hilbe J., Hardin J. Modeling count data with generalized distributions. Stata J., **14**, (2014).

[14] Ismail N., Jemain A. Handling Overdispersion with Negative Binomial and Generalized Poisson Regression Models. Casualty actuarial society forum, (2007).

[15] Lawless J. F. Regression methods for poisson process data. J. Am. Stat. Assoc, **82**, (1987a).

[16] Lawless J. F. Negative Binomial Regression Models. Can. J. Stat., **15**, (1987b).

[17] Manton K.,Woodbury M.,Stallard E. Variance components approach to categorical data models with heterogenous cell populations: analysis of spatial gradients in lung cancer mortality rates in North Carolina counties. Biometrics, (1998).

[18] Moran P. A. Maximum Likelihood Estimation in Non-Standard Conditions Proceedings of the Cambridge Philosophical Society, **70**, (1971).

[19] Rao C. R. Linear Statistical Inference and Its Application. Wiley, **2**, (1973).

[20] Rodriguez G. Models for over-dispersed count data Generalized Linear Models.Princeton University, Web. 14 Oct. 2015. Available at: http://data.princeton.edu/wws509/stata/overdispersion.html, (2015).

1