2022

Analysis of Count Data in One-Way Layout with Missing Response

Poonam Shrestha Malakar
University of Windsor

Recommended Citation

Malakar, Poonam Shrestha, "Analysis of Count Data in One-Way Layout with Missing Response" (2022). *Electronic Theses and Dissertations*. 8986.
https://scholar.uwindsor.ca/etd/8986

ANALYSIS OF COUNT DATA IN ONE-WAY LAYOUT WITH MISSING RESPONSE

by

Poonam Shrestha Malakar

A Dissertation
Submitted to the Faculty of Graduate Studies
through the Department of Mathematics and Statistics
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy at the
University of Windsor

Windsor, Ontario, Canada

2023

ANALYSIS OF COUNT DATA IN ONE-WAY LAYOUT WITH MISSING RESPONSE

by

Poonam Shrestha Malakar

APPROVED BY:

S.B. Provost, External Examiner
Western University

G. Bhandari
Odette School of Business

M. Hlynka
Department of Mathematics and Statistics

A. Hussein
Department of Mathematics and Statistics

S. Paul, Advisor
Department of Mathematics and Statistics

December 9, 2022

Author's Declaration of Originality

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

Abstract

Discrete data in the form of counts arise in many health science disciplines such as biology and epidemiology. Poisson models are widely used in the regression analysis of count data. The Poisson distribution has a property that the mean and the variance are equal. However in practice many count data sets often display extra-variation or over-dispersion relative to a Poisson model. Thus the Poisson distribution is not an ideal choice for analysing count data in many applications. One very convenient and common model to accommodate this extra dispersion is the two parameter negative binomial distribution. Count data in the form of one-way layout arise in many practical situations. These data often exhibit extra variation that cannot be explained by a simple model, such as the binomial or the Poisson. These data may further be complicated when some of the observations are missing as in the continuous and some other discrete data situations. In this dissertation we study the performance $C(\alpha)$ statistics recommended by Barnwal and Paul (1988) for testing the equality of means of several groups of count data in presence of a common dispersion parameter. We also study the performance of the three $C(\alpha)$ statistics developed by Saha (2008) in terms of level and power. We develop estimation procedures for the parameters involved in the one way layout of count data under different missing data scenarios and study the effect of missingness on the $C(\alpha)$ statistics through simulations.

Dedication

Dedicated to my family.

Acknowledgements

All glorification and gratitude to Almighty God who has given me the ability to accomplish this dissertation.

I would like to express my deepest gratitude to my supervisor, Dr. S. Paul for his guidance, help, continuous encouragement and financial support. This dissertation would not have been completed without his continuous academic advice.

I would like to thank the advisory committee members, Dr. M. Hlynka, Dr. A. Hussein and Dr. G. Bhandari for reading my dissertation, and offering constructive suggestions. My great gratitude also goes to the external examiner of my dissertation Dr. Serge Provost, Department of Statistical and Actuarial Sciences, Western University, London, Ontario, for his critical review, valuable comments and advice.

My sincere thanks also goes to my friends for providing me continuous moral support and motivation.

Last but not least, I would like to thank my husband and my family members for their support, patience, love and prayers. Their love, expectation and trust always motivated me to work harder and helped me achieve my goal.

Table of Contents

List of Tables

List of Abbreviations

CC Complete case analysis

EM Expected maximization algorithm

MAR Missing at random

MCAR Missing Completely at random

MNAR Missing not at random

QL Quasi Likelihood

EQL Extended Quasi Likelihood

DEQL Double Extended Quasi Likelihood

Chapter 1

Introduction

Discrete data in the form of counts arise in many health science disciplines such as biology and epidemiology. For examples of discrete count data see Anscombe (1949); Bliss and Fisher (1953); McCaughran and Arnold (1976); Margolin, Kaplan, and Zeiger (1981); Bohning, Dietz, Schlattmann, Mendonca, and Kirchner (1999); Paul and Deng (2000), and Deng and Paul (2005).

Poisson models are widely used in the regression analysis of count data. The Poisson distribution has a property that the mean and the variance are equal. However in practice many count data sets often display extra-variation or over/under dispersion relative to a Poisson model. Thus the Poisson distribution is not an ideal choice for analysing count data in many applications. One very convenient and common model to accommodate this extra dispersion is the two parameter negative binomial distribution.

For applications of the negative binomial distribution, see, for example Margolin et al. (1989); Engel (1984); Breslow (1984); Lawless (1987); Collings and Margolin (1985). Different authors have used different parameterizations for the negative binomial distribution. For example, see, Paul and Plackett (1978); Barnwal and Paul (1988); Piegorsch (1990); Paul and Banerjee (1998); Paul and Deng

(2000), and Deng and Paul (2005).

Count data may further be complicated by the existence of missing values. Extensive work has been done on analysis of continuous response data under normality assumption. See, for example, Rubin (1976), Anderson and Taylor (1976), Geweke (1986), Little and Rubin (1987), Raftery, Madigan and Hoeting (1997), Chen, Hubbard and Rubin (2001), Kelly (2007), and Zhang, and Huang (2008).

Some work on missing values has also been done on logistic regression analysis of binary data. See, for example, Ibrahim (1990); Lipsitz and Ibrahim (1996), Ibrahim and Lipsitz (1996); Ibrahim, Chen and Lipsitz (1999); Ibrahim, Chen and Lipsitz (2001); Sinha and Maiti (2008); Maiti and Pradhan (2009).

Some work on missing values with count data has been done. See, for example, Mian and Paul (2016), Luo and Paul (2018).

Rubin (1976) and Little and Rubin (1987) discuss various missingness mechanisms. If the missingness does not depend on observed data, then the missing data are called missing completely at random (MCAR). If the missing data mechanism depends only on observed data, then the data are missing at random (MAR). The MAR is also known as ignorable missing That is, in this case, the missing data mechanism is ignored. If the missing data mechanism depends on both observed and unobserved data, that is, failure to observe a value depends on the value that would have been observed, then the data are said to be missing not at random (MNAR) in which case the missingness is nonignorable . For more detailed discussion on missing data mechanism, see, Ibrahim et al. (2005, p333).

Count data in the form of one-way layout arise in practice. See, for example, Beal (1939), Blish and Fisher (1953), McCaughran and Arnold (1976), and Hutto, Pletschet, and Hendricks (1986). These data often exhibit extra variation that cannot be explained by a simple model, such as the binomial or the Poisson.

Barnwal and Paul (1988) developed two C(α) tests to test the equality of the means of several groups of count data with negative binomial variation. Saha (2008) developed similar C(α) tests when the data are over/under dispersed but the data distribution is not known.

In this dissertation we develop inference procedures for one-way layout of count data with over/under dispersion in presence of missing responses. We develop estimation procedures for the parameters involved in one way layout of count data under different missing data scenarios and study the effect of missingness on the $C(\alpha)$ statistics when data can be assumed to have come from a specific over/under dispersed model, such as, the negative binomial distribution or when data are assumed to have come from an unspecified over/under dispersed model based on the knowledge of only the first two moments of the counts obtained using the double extended quasi-likelihood.

In chapter 2 we discuss some preliminaries and review some literature related to the count data model with extra-variation, missing values, maximum likelihood estimation by using weighted expected maximization algorithm and $C(\alpha)$tests.

In chapter 3 we study the two $C(\alpha)$ statistics recommended by Barnwal and Paul (1988). A simulation study is conducted to study the performance of these statistics in terms of level and power.

In chapter 4, we study the performance of the $C(\alpha)$ statistics based on the quasi-likelihood, extended quasi-likelihood and double extended quasi-likelihood in terms of size and power.

In chapter 5, a study of the effect of missingness on the $C(\alpha)$ statistic based on maximum likelihood and double extended quasi-likelihood is presented.

Finally, conclusions of the thesis with the summary of findings and a plan for future study are presented in chapter 6.

Chapter 2

Preliminaries and Literature Review

2.1 Count data model with extra-variation

2.1.1 Poisson Model

Let Y be the count data which follows the Poisson distribution. The probability mass function for Poisson distribution is given by

$$f(y;\mu) = \frac{e^{-\mu}\mu^{y}}{y!}, \tag{2.1.1}$$

where μ is the mean parameter. The mean and variance of the Poisson distribution are equal to μ.

2.1.2 Negative Binomial model

Let Y be a negative binomial random variable with mean parameter m and dispersion parameter c. Then, using the terminology of Paul and Plackett (1978),

Y has the probability mass function

$$f(y; m, c) = \tfrac{\Gamma(y+c^{-1})}{y!\Gamma(c^{-1})} \left(\tfrac{cm}{1+cm}\right)^{y} \left(\tfrac{1}{1+cm}\right)^{c^{-1}} \tag{2.1.2}$$

for $y = 0, 1, ..., m > 0$. Now, for a typical Y, $Var(Y) = m(1+cm)$ and $c > -1/m$. Obviously, when $c = 0$, variance of the $NB(m, c)$ distribution becomes that of the Poisson(m) distribution. Moreover, it can be shown that the limiting distribution of the $NB(m, c)$ distribution, as $c \to 0$, is the Poisson(m).

2.2 Missing data

Missing data or missing values occur when no information is available on the responses or some of the covariates or both responses and covariates for some subject of interest in the study. Missing observations are very common obstacles faced by researchers in real-world contexts which makes the data analysis more complicated. There can be several reasons why some observations in the data set may be missing. Non-response occurs when the respondent does not respond to certain questions due to stress, lack of knowledge or some questions may be sensitive. Most of the standard statistical methods are based on complete information for all the variables under study. The use of observed information only in the analysis may produce biased and inefficient parameter estimates and the results can be misleading. Even a small number of missing observations can have dramatic effect on the statistical analysis. Missing data may degrade the performance of confidence intervals, reduce statistical power and bias parameter estimates (Nakai and Ke, 2011).

2.2.1 Missing data mechanism

It is essential to know how the observations in a data set are missing. Missing data mechanism is the way how observations are missing in a data set. Rubin (1976) and Little and Rubin (1987) came up with the classification system that is in practice today: Missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR).

Missing Completely at random(MCAR)

Data are MCAR if the probability of missing data on a variable X is unrelated to other observed as well as unobserved values of the variable itself or any other variable in the data set. Under MCAR, the probability of missingness is same for all the observations. For example, consider a child in an educational study that moves to another district midway through the study. The missing values are MCAR if the reason for the move is unrelated to other variables in the data set.

Missing at random(MAR)

Data are MAR, if missingness is related to only observed variables in the analysis, but not to the underlying values of the incomplete variables. Here probability of missingness depends only on available observations, not on unobserved observations. For example, missing data on income depends on a house value but is not related to income given house value. The MAR is also known as ignorable missing.

Missing not at random (MNAR)

Data are MNAR, if the missingness depends on observed as well as unobserved observations. Here the probability of missingness depends on both observed and unobserved observations. For example, dropouts in the medical studies. A person in a study may not like the previous results and may be worried about the future results of the study and drops out. The MNAR is also known as non-ignorable missing.

2.2.2 Procedures for handling missing data

2.2.2.1 Complete case analysis

This method of analysis also known as listwise deletion, considers only those subjects which have all the informations available. Here, any subjects having missing observations are deleted before the analysis. The main advantage of this method is, one can use any standard statistical software for the analysis. It is easy to use since no special computational methods are required. The drawback of the method is when a large fraction of data is missing and considering only complete cases in the analysis will result in reduction of sample size, also loss of some important features of the data hence the study may not be reliable. The method works well when data are missing completely at random (MCAR) , which is rare in reality.

2.2.2.2 Methods based on Imputation

Imputation method is one of the widely used methods in practice, where each missing observation is replaced by some guess or estimate based on available or observed data. Once the imputation is complete the analysis is straightforward using conventional software. Imputation can be single value imputation or sometimes multiple imputation. The basic advantage of the method is that no observation is removed, all the information is used in the analysis. Based on the process how missing values are replaced this is further classified as:

(1) Last value carried forward imputation:

This is one of the most widely used technique in longitudinal analysis. Under this method, each missing observation is substituted by the last observed value for the same subject. The method is simple but it uses strong assumption that the missing value does not change which is quite unlikely in many situations. One of few settings where the method may be appropriate is in some studies where the missingness is due to recovery or cure (Nakai and Ke, 2011) .

(2) Imputation by related observation:

Here the missing observations are replaced by some related observations. For example, missing observation about income can be substituted by income of another person doing similar job.

(3) Marginal mean imputation:

The missing values are filled in by the average of observed values of the variable. It is easy to substitute the missing values however the distribution of variable is distorted and this method assumes data missing completely at random, hence the method is not recommended these days.

(4) Conditional mean imputaiton:

This method of imputation was discussed by Buck (1960) and Little and Rubin (1987). At first, based on the complete data, mean and covariance matrix is estimated. Using these estimates, least square regression of missing values on observed values is computed for each missing data pattern. The missing values are filled in by the conditional mean in the second step. This method yields reasonable estimates of mean if the normality assumptions are plausible.

(5) Hot deck imputation:

This technique is common in survey practice. Here the missing values are substituted from similar responding units in the sample. For example, the missing information about the total number of individuals in a household is filled in by the total number of persons in a similar household in that area.

(6) Cold deck imputation:

Under this method, information from external source is used to replace the missing observation, such as value from a previous survey.

(7) Substitution method:

This technique is used at the fieldwork stage of survey. In case of nonresponse, information is collected from the other available units which were not previously a part of the sample. For example if the initially selected respondents are not available then the information is gathered from a substitute who was not a part of the sample before.

(8) Regression imputation:

Under this method, the estimated values are obtained from the regression of missing observations on the observed values. First regression equation is computed based on the completed observation which is then used to predict the missing observations.

(9) Multiple Imputation:

Multiple Imputation (MI) is one of the most popularly used technique in handling the missing values. Here, each missing value is replaced by two or more plausible estimates representing a distribution of possibilities (Allison, 2001). MI works on the principle that larger random samples yield more certainty about estimates and the estimate of the missing observation is more robust when many plausible values are sampled (Little et al. 2014). The advantage of method is that analysis can be performed using any statistical package once the imputed data set is generated. The limitation of the method is that data should be missing at random.

2.2.2.3 Methods based on Likelihood

Another widely used method in handling missing data is the Expectation Maximization (EM) algorithm (Dempster et al., 1997). In usual situation where data are assumed to have come from a distribution with parameter $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ can be vector valued, a likelihood L or log-likelihood l is constructed and maximized to obtain the maximum likelihood estimates (MLE) of the parameter $\boldsymbol{\theta}$. However the situation becomes complicated when some of the observations are missing and EM algorithm is useful in such situations. EM algorithm is an iterative algorithm that finds the estimates of parameters which maximize the log likelihood in presence of missing observations. Each iteration of EM comprises two steps, expectation (E-step) and maximization (M-step). A cautionary note is that the current EM theory only guarantees convergence to a local maximum of the likelihood, which is not necessarily the MLE.

2.3 Estimation procedures for the parameters

There are a few methods of estimation available for estimating the parameters. namely maximum likelihood, Bayesian estimation techniques.

In the model based procedure, a parametric model can be specified for the variable with missing observations. In likelihood based estimation, the likelihood function often is factored based on the observed or missing observations. In this type of situation, maximum likelihood estimation technique is easily applicable to estimate the parameters. Maximum likelihood estimates can be used to estimate the variance components from the second derivative of the log likelihood function. The complete data log-likelihood function can be maximized using available algorithms such as Newton Raphson (NR), Nelder Maid (NM) and other similar algorithms.

For data with missing values, maximum likelihood estimates of the parameters can be obtained using EM algorithm by Dempster, Laird and Rubin (1977), and the Weighted EM algorithm by Ibrahim (1990).

Multiple Imputation is another likelihood based approach. In this approach, multiple complete data sets are created by filling in the missing observations. Parameter estimates are obtained for each complete (imputed) data sets. The average estimates of parameters are then obtained for the multiple data sets. For detailed discussion see Little and Rubin (2002). In many practical situations, likelihood based estimation may not be possible to find due to incorrect distributional assumptions. Weighted Estimating Equations can be used to estimate the parameters in such situations. More details about weighted estimating equations in the presence of missing observations are given in Lipsitz, Ibrahim, and Zhao (1999). Bayesian approach is another technique for handling data with missing observations. In this approach prior distributions are specified for all the parameters in the model. Distributional assumptions for the variables having missing observations are also necessary in this approach. For detailed discussion see Ibrahim, Lipsitz, and Chen (2002). Application of any one of these techniques depends on the situation needed to be addressed. There is no unique superiority of these techniques. For more detailed discussion, see, Ibrahim, Chen, Lipsitz, and Herring (2005). We have applied the maximum likelihood estimation technique (using weighted EM algorithm) to estimate the parameters of over/under dispersed count data model.

Expectation Maximization (EM) algorithm by Demster, Laird and Rubin (1997) has been used to find the maximum likelihood estimates of the regression parameters of the model for the data having incomplete or missing observations. Ibrahim (1990) used the EM algorithm by the method of weights for incomplete data in generalized linear models. Following Ibrahim (1990), a number of articles have been published for the application of the EM algorithm by method of weights. For more details, see, Lipsitz and Ibrahim (1996), Ibrahim and Lipsitz (1996), Ibrahim, Chen, and Lipsitz (1999, 2001), Ibrahim, Chen, Lipsitz, and Herring (2005), Sinha and Maiti (2008), and Maiti and Pradhan (2009). The implementation of EM algorithm is straight forward and is computationally more feasible. In this approach, the log-likehood function of the parameters can be separated for the regression parameters, parameters of the covariate distribution from the parameters of the missingness mechanism. This feature of the log likelihood facilitates the separate maximization and helps to separate the nuisance parameters from the parameters of interest. These characteristics of the EM algorithm motivate us to use this algorithm to find the maximum likelihood estimates of the over/under dispersed count data model with missing observations. More details of the EM algorithm by the method of weights are explained in the chapters that follow.

2.4 C(α) statistics

The C(α) test is based on partial derivatives of the log-likelihood function with respect to the nuisance parameters and the parameters of interest evaluated at the null hypothesis. Let L($\boldsymbol{\theta}, \boldsymbol{\phi}$;y) be the likelihood function and l be the log-likelihood function for the data, where $\boldsymbol{\phi} = (\phi_1, \phi_2, \cdots, \phi_s)$ is the nuisance parameters. Define partial derivatives of the log-likelihood which are evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0 = (\theta_{01}, \theta_{02}, \cdots, \theta_{0k})'$ as

$$\boldsymbol{\psi} = \left.\frac{\partial l}{\partial \boldsymbol{\theta}}\right|_{\theta=\theta_0} = \left.\left[\frac{\partial l}{\partial \theta_1}, \frac{\partial l}{\partial \theta_2}, \cdots, \frac{\partial l}{\partial \theta_k}\right]'\right|_{\theta=\theta_0}$$

and

$$\boldsymbol{\gamma} = \left.\frac{\partial l}{\partial \boldsymbol{\phi}}\right|_{\theta=\theta_0} = \left.\left[\frac{\partial l}{\partial \phi_1}, \frac{\partial l}{\partial \phi_2}, \cdots, \frac{\partial l}{\partial \phi_s}\right]'\right|_{\theta=\theta_0}.$$

Under the null hypothesis and mild regularity conditions, $(\frac{\partial l}{\partial \boldsymbol{\theta}}, \frac{\partial l}{\partial \boldsymbol{\phi}})$ follow a multivariate normal distribution with the mean vector $\mathbf{0}$ and variance-covariance matrix, $I^{-1}(\boldsymbol{\theta}, \boldsymbol{\phi})$ (Cramer,1946), where

$$I(\boldsymbol{\theta}, \boldsymbol{\phi}) = \begin{bmatrix} I_{\boldsymbol{\theta\theta}} & I_{\boldsymbol{\theta\phi}} \\ I'_{\boldsymbol{\theta\phi}} & I_{\boldsymbol{\phi\phi}} \end{bmatrix}$$

is the Fisher information matrix with elements

$$I_{\boldsymbol{\theta\theta}} = E\left(-\left.\frac{\partial^2 l}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right|_{\theta=\theta_0}\right), I_{\boldsymbol{\phi\phi}} = E\left(-\left.\frac{\partial^2 l}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'}\right|_{\theta=\theta_0}\right), \text{and} \quad I_{\boldsymbol{\theta\phi}} = E\left(-\left.\frac{\partial^2 l}{\partial \boldsymbol{\theta} \partial \boldsymbol{\phi}}\right|_{\theta=\theta_0}\right)$$

which are $(k \times k), (s \times s),$ and $(k \times s)$ matrices, respectively.

The $C(\alpha)$ test is based on the adjusted score $S = \frac{\partial l}{\partial \boldsymbol{\theta}} - B\frac{\partial l}{\partial \boldsymbol{\phi}}$, where B is the matrix of partial regression coefficients that is obtained by regressing $\frac{\partial l}{\partial \boldsymbol{\theta}}$ on $\frac{\partial l}{\partial \boldsymbol{\phi}}$. According to Bartlett (1953), B and the variance-covariance matrix of S can be expressed as $B = I_{\boldsymbol{\theta\phi}} I^{-1}_{\boldsymbol{\phi\phi}}$, and $I_{\boldsymbol{\theta\theta}.\boldsymbol{\phi}} = I_{\boldsymbol{\theta\theta}} - I_{\boldsymbol{\theta\phi}} I^{-1}_{\boldsymbol{\phi\phi}} I_{\boldsymbol{\theta\phi}}$. Thus the distribution of adjusted score $S \sim MN(\mathbf{0}, I_{\boldsymbol{\theta\theta}.\boldsymbol{\phi}})$ and hence the distribution of $S' I^{-1}_{\boldsymbol{\theta\theta}.\boldsymbol{\phi}} S \sim \chi^2_{(k)}$ (Neyman, 1959).

The statistic involves nuisance parameters $\boldsymbol{\phi} = (\phi_1, \phi_2, \cdots, \phi_s)$ which need to be replaced by some suitable estimates for testing the null hypothesis. Following Moran (1970) and replacing the nuisance parameters by some $\sqrt{n}$ consistent estimators $\tilde{\boldsymbol{\phi}} = (\tilde{\phi}_1, \tilde{\phi}_2, \cdots, \tilde{\phi}_s)'$ evaluated from the data, the test statistic becomes

$$\chi^2_{C(\alpha)} = \tilde{S}'\tilde{I}^{-1}_{\boldsymbol{\theta\theta}.\boldsymbol{\phi}}\tilde{S},$$

which is asymptotically distributed as chi-square with k degrees of freedom (Neyman, 1959).

Note, if the nuisance parameter $\boldsymbol{\phi}$ is replaced by its maximum likelihood estimates $\hat{\boldsymbol{\phi}}$, then the adjusted score function S reduces to $\boldsymbol{\psi}$. The $C(\alpha)$ statistics then becomes

$$S_1 = \hat{\boldsymbol{\psi}}'\hat{I}^{-1}_{\boldsymbol{\theta\theta}.\boldsymbol{\phi}}\hat{\boldsymbol{\psi}},$$

which is a score test (Rao, 1948).

The score test or $C(\alpha)$ class of test has many advantages: it often maintains at least approximately, a preassigned level of significance, it requires estimates of parameters only under the null hypothesis, and it often produces a statistic that is simple to calculate (Deng and Paul, 2000).

Chapter 3

Analysis of One-way layout of count data:Complete Data with Parametric Model

3.1 Introduction

Discrete data in the form of counts arise in many health science disciplines such as biology and epidemiology. For examples of discrete count data see Deng and Paul (2000, 2005); Anscombe (1949); Bliss and Fisher (1953); McCaughran and Arnold (1976); Margolin, Kaplan, and Zeiger (1981); Bohning, Dietz, Schlattmann, Mendonca, and Kirchner (1999).

Poisson models are widely used in the regression analysis of count data. The Poisson distribution has a property that the mean and the variance are equal. However, in practice many count data often display extra-variation or over-dispersion relative to a Poisson model. Thus Poisson distribution is not an ideal choice for analysing count data in many applications. One very convenient and common model to accommodate this extra dispersion is the two parameter negative binomial distribution.

For applications of the negative binomial distribution, see, for example Engel (1984); Breslow (1984); Margolin et al. (1989); Lawless (1987); Collings and Margolin (1985). Different authors have used different parameterizations for the negative binomial distribution. For example, see, Paul and Plackett (1978); Barnwal and Paul (1988); Paul and Banerjee(1998); Piegorsch (1990); Paul and Deng (2000), and Deng and Paul (2005).

Count data in the form of one-way layout arise in practice. See, for example, Beal (1939), Blish and Fisher (1953), McCaughran and Arnold (1976), and Hutto, Pletschet, and Hendricks (1986). These data often exhibit extra variation that cannot be explained by a simple model, such as the binomial or the Poisson.

Barnwal and Paul (1988) developed two C(α) tests to test the equality of the means of several groups of count data with negative binomial variation. The performance of these statistics were compared with the likelihood ratio statistic and other two statistic based on transformed data (Anscombe, 1948) and C(α) statistics were recommended.

Here, we study the performance of these two C(α) statistics recommended by Barnwal and Paul (1988) in terms of size and power.

3.1.1 Negative Binomial model

Let Y be a negative binomial random variable with mean parameter m and dispersion parameter c. Then, using the terminology of Paul and Plackett (1978), Y has the probability mass function

$$f(y; m, c) = \tfrac{\Gamma(y+c^{-1})}{y!\Gamma(c^{-1})} \left(\tfrac{cm}{1+cm}\right)^y \left(\tfrac{1}{1+cm}\right)^{c^{-1}} \tag{3.1.1}$$

for $y = 0, 1, ..., m > 0$. Now, for a typical Y, $Var(Y) = m(1+cm)$ and $c > -1/m$. Obviously, when $c = 0$, variance of the $NB(m, c)$ distribution becomes that of the Poisson(m) distribution. Moreover, it can be shown that the limiting distribution of the $NB(m, c)$ distribution, as $c \to 0$, is the Poisson(m).

3.1.2 The negative binomial likelihood

Let $Y_{ij},\ i = 1, ..., K,\ j = 1, ..., n_i$ be the counts for the j^{th} individual of the i^{th} treatment group. We assume that $Y_{ij} \sim NB(m_i, c_i)$, with mean m_i and dispersion parameter c_i, which has probability mass function

$$Pr(Y_{ij} = y_{ij}|m_i, c_i) = \tfrac{\Gamma(y_{ij}+c_i{}^{-1})}{y_{ij}!\Gamma(c_i{}^{-1})} \left(\tfrac{c_i m_i}{1+c_i m_i}\right)^{y_{ij}} \left(\tfrac{1}{1+c_i m_i}\right)^{c_i{}^{-1}} \tag{3.1.2}$$

for $y_{ij} = 0, 1, ...,$ and $m_i > 0$. The mean and variance of Y_{ij} are

$$E(Y_{ij}) = m_i \quad \text{and } \mathrm{var}(Y_{ij}) = m_i(1 + c_i m_i), \tag{3.1.3}$$

where $c_i > -1/m_i$. The log-likelihood, apart from some constant terms, can be written as

$$l = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[y_{ij}\ln\ (m_i) - \Big(y_{ij} + \tfrac{1}{c_i}\Big)\ln(1 + c_i m_i) + \sum_{l=1}^{y_{ij}}\ln\{1 + c(l-1)\}\Big]. \tag{3.1.4}$$

3.1.3 Estimation of the Parameters

We are interested in testing $H_0 : m_1 = m_2 = \cdots m_K = m$ against H_1 : not all $m_i^{'}$s are equal for all $c > -1/m$. The log likelihood under the hypothesis H_0 is

$$l_0 = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[y_{ij}\ln m - (y_{ij} + c^{-1})\ln(1 + cm) + \sum_{l=1}^{y_{ij}}\ln\{1 + c(l-1)\}\Big]. \tag{3.1.5}$$

Under H_0, the maximum likelihood estimator of m is $\hat{m} = \bar{y} = \sum_{i=1}^{K} \frac{\bar{y_{i.}}}{n}$, where $n = \sum_{i=1}^{K} n_i$. The maximum likelihood estimator c_0 of c under H_0 is obtained as a solution to

$$n\log(1 + c_0\bar{y}) - \sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{l=1}^{y_{ij}} \left\{\frac{c_0}{1 + c_0(l-1)}\right\} = 0. \tag{3.1.6}$$

3.1.4 Testing of Hypothesis

In this section we develop procedures for testing the composite hypotheses $H_0 : m_1 = \cdots = m_K$ against H_1: at least two m_i's are not the same, with the assumption $c_1 = \cdots = c_K = c$, where c is unknown and $c > -1/m_i$. For the convenience of the derivation of the $C(\alpha)$ statistics we reparameterize m_i under H_1 by $m_i = m + \delta_i,\ i = 1, ..., K-1$, with $\delta_K = 0$. Then testing $H_0 : m_1 = ... = m_K = m$ reduces to testing $H_0 : \delta_1 = ... = \delta_{K-1}$, where m and c are treated as nuisance parameters. This technique was employed by many authors. For example, Tarone (1985) used this technique to obtain $C(\alpha)$ statistic for testing the equality of several odds ratios. Barnwal and Paul (1988) used this same technique to derive these statistics for testing equality of means in the presence of common negative binomial over-dispersion. The log-likelihood in terms of reparameterization of $m_i = m + \delta_i$ and $c_i = c$, apart from some constant terms, can be written as

$$\begin{aligned} l = & \sum_{i=1}^{K}\sum_{j=1}^{n_i} \Big[y_{ij}\ln(m + \delta_i) - \Big(y_{ij} + \tfrac{1}{c}\Big)\ln(1 + cm + c\delta_i) \\ & + \sum_{l=1}^{y_{ij}}\log\{1 + c(l-1)\}\Big]. \end{aligned} \tag{3.1.7}$$

Now, define $\boldsymbol{\delta} = (\delta_1, ..., \delta_{K-1})$ and $\boldsymbol{\nu} = (\nu_1, \nu_2)^{'} = (m,\ c)^{'}$. Then,

Following the theory in §2.4, we obtain

$\phi_i = \frac{\partial l}{\partial \delta_i}\Big|_{\delta=0} = \frac{n_i(\bar{y_{i.}} - m)}{m(1+cm)}, \quad i = 1, ..., K-1,$

$$\eta_1 = \frac{\partial l}{\partial \nu_1}\Big|_{\delta=0} = \sum_{i=1}^{K} \frac{n_i(\bar{y_{i.}} - m)}{m(1+cm)}$$

and

$$\eta_2 = \frac{\partial l}{\partial \nu_2}\Big|_{\delta=0} = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{1}{c^2}\ln(1+cm) - \sum_{r=0}^{y_{ij}-1}\frac{1}{c(1+cr)}\Big],$$

where $\bar{y_{i.}} = (1/n_i)\sum_{j=1}^{n_i} y_{ij}$ is the sample mean of the i^{th} treatment. The $C(\alpha)$ test is based on $\lambda_i(\hat{\nu}) = \phi_i(\hat{\nu}) - \beta_{1i}\eta_1(\hat{\nu}) - \beta_{2i}\eta_2(\hat{\nu})$, where β_{1i} and β_{2i} are, respectively, the partial regression coefficient of δ_i on η_1 and δ_i on η_2, and where $\hat{\nu}$ is some $\sqrt{n}$ (where $n = \sum_{i=1}^{K} n_i$) consistent estimator of ν under the null hypothesis. The regression coefficients $\beta = (\beta_1, \beta_2)$ with $\beta_1 = (\beta_{11}, ..., \beta_{1K-1})$ and $\beta_2 = (\beta_{21}, ..., \beta_{2K-1})$ are obtained as $\Gamma\gamma^{-1}$, and the variance-covariance of $\Lambda(\nu) = [\lambda_1(\nu), ..., \lambda_{K-1}(\nu)]^{'}$ is $\Delta - \Gamma\gamma^{-1}\Gamma^{'}$, where the (s,t)th components of Δ, Γ, and γ are

$$\Delta_{st} = E\Big(-\frac{\partial^2 l}{\partial\delta_s\partial\delta_t}\Big|_{\delta=0}\Big) = \begin{cases} n_s/m(1+cm), & s = t = 1, \ldots, K-1, \\ 0 & \text{otherwise,} \end{cases}$$

$$\Gamma_{st} = E\Big(-\frac{\partial^2 l}{\partial\delta_s\partial\nu_t}\Big|_{\delta=0}\Big) = \begin{cases} n_s/m(1+cm) & , s = 1, ..., K-1, t = 1, \\ 0, & s = 1, ..., K-1, t = 2, \end{cases}$$

and

$$\gamma_{st} = E\Big(-\frac{\partial^2 l}{\partial\nu_s\partial\nu_t}\Big|_{\delta=0}\Big) = \begin{cases} n/m(1+cm), & s=t=1, \\ b, & s=t=2, \\ 0 & \text{otherwise}, \end{cases}$$

respectively. Note that the above terms involve the nuisance parameter $\nu = (m, c)$. Thus, using $\hat{\nu}$ in Λ, Δ, Γ, and γ, the $C(\alpha)$ statistic is obtained as $\Lambda'(\Delta - \Gamma\gamma^{-1}\Gamma')^{-1}\Lambda$, which is approximately distributed as a chi-square distribution with K-1 degrees of freedom. After replacing $\hat{\nu} = (\hat{m}, \hat{c})$ by the maximum likelihood estimtes of ν, the $C(\alpha)$ statistic, after some algebra, becomes

$$S_c(ml) = \sum_{i=1}^{K} \frac{n_i(\bar{y_{i.}} - \hat{m})^2}{\hat{m}(1+\hat{m}\hat{c})},$$

The derivation is given in appendix A.

This $C(\alpha)$ statistic based on the maximum likelihood has been derived by Barnwal and Paul(1988).

Using the method of moment estimates $\tilde{m}$ and $\tilde{c}$ of m and c, the $C(\alpha)$ statistic is

$$S_c(mm) = \sum_{i=1}^{K} \frac{n_i(\bar{y_{i.}} - \tilde{m})^2}{\tilde{m}(1+\tilde{c}\tilde{m})},$$

where $\tilde{m} = \bar{y} = \sum_{i=1}^{K} \frac{\bar{y_{i.}}}{n}$, $\tilde{c} = (s^2 - \bar{y})/\bar{y}^2$, and $s^2 = \sum_{i=1}^{K}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2/(n-1)$.

3.1.5 Simulation

A simulation study is conducted to examine the comparative behaviour of the test statistics $S_c(ml)$ and $S_c(mm)$ in terms of size and power. The test of equality of means of two and three negative binomial distributions with common dispersion parameter is considered. Empirical significance levels and power of the tests were all based on 10,000 samples from the negative binomial distribution for different values of m and c. For all combinations of $m = 7, 15, 20, 30, 40$ and $c = 0.05, 0.25, 0.50$ simulations are run based on 10,000 samples of sizes

$n_1 = n_2 = 10, 20$, and 50 for two groups. The values of parameters are same for $n_1 = n_2 = n_3 = 10, 20$, and 50 for three groups. Table 3.1 displays empirical levels based on a nominal significance level of $\alpha = 5\%$ for $S_c(ml)$ and $S_c(mm)$ for data generated from the NB distributions for two groups. Table 3.2 displays empirical power based on a nominal significance level of $\alpha = 5\%$ for $S_c(ml)$ and $S_c(mm)$ for data generated from the NB distributions. For two groups, $c = 0.25$, $m_1 = m$, $m_2 = m + \phi$, and $\delta = \phi/m$, where $m = 10, 20, 50$ and $\delta = 0.0, 0.2, 0.4, 0.6, 0.8$. The sample sizes considered were $n_1 = n_2 = 10, 20,$ and 50. Table 3.3 displays empirical levels based on a nominal significance level of $\alpha = 5\%$ for $S_c(ml)$ and $S_c(mm)$. For each replication, three samples are generated from the NB(m,c). Table 3.4 displays empirical power based on a nominal significance level of $\alpha = 5\%$ for $S_c(ml)$ and $S_c(mm)$. In this case also for each replication three samples are generated from the NB(m,c). For three groups, c $= 0.25$, $m_1 = m$, $m_2 = m + \phi_1$, $m_3 = m + \phi_2$, $\delta_1 = \phi_1/m$,$\delta_2 = \phi_2/m$, where $m = 10, 20, 50$ and $\boldsymbol{\delta} = (0, 0), (0, 0.2), (0.2, 0.4), (0.4, 0.6), (0.4, 0.8)$. The sample sizes considered were $n_1 = n_2 = n_3 = 10, 20,$ and 50.

Table 3.1: $10^3\times$ empirical levels: α=0.05; based on 10,000 replications

		m=7			m=15			m=20			m=30			m=40		
		c			c			c			c			c		
$n_1 = n_2$	Statistic	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50
10	$S_c(ml)$	43.5	49.3	46.1	48.1	49.5	42 .2	51.5	49.3	50.3	48.7	49	49.8	54.8	49.1	51.6
	$S_c(mm)$	43.2	46.5	43.0	44.6	43.8	40.5	43.8	44.8	45.1	43.7	44.4	44.8	48.2	42.9	45.2
20	$S_c(ml)$	48.2	51.7	44.2	56.5	52.9	48.5	55	50.2	48.9	54.5	52.6	49	52.9	51.5	59.9
	$S_c(mm)$	47.4	51	44.4	53.3	49.5	46.5	52.8	48.5	47.2	51.6	50.8	46.7	50.2	48.9	45.3
50	$S_c(ml)$	54.1	63.9	56.7	50.4	51.2	47.9	53.4	52.3	51	53.7	50.6	50.1	52.6	53.1	53.7
	$S_c(mm)$	52.3	51	49.9	49.5	49.7	47.3	51.7	50	50.2	52.4	48.3	46.8	51.3	52	53.3

Table 3.2: $10^3\times$ empirical power corresponding to nominal significance level α=0.05;based on 10,000 replications for 2 groups; $m_1 = m,\ m_2 = m + \phi,\ \delta = \phi/m$; c=0.25

		$n_1 = n_2 = 10$ δ					$n_1 = n_2 = 20$ δ					$n_1 = n_2 = 50$ δ				
m	Statistic	0.0	0.2	0.4	0.6	0.8	0.0	0.2	0.4	0.6	0.8	0.0	0.2	0.4	0.6	0.8
10	$S_c(ml)$	48.9	98	226.9	403.5	580.2	49.6	158.8	425.7	703	886.6	52.7	341.3	823.8	982	998.9
	$S_c(mm)$	43.5	88.4	207.3	376.5	547	46.9	154.5	409	688.2	875	51.3	340.1	820	980.9	998.6
20	$S_c(ml)$	49.3	114.6	266.7	452.6	636.8	50.2	179.9	484	766.4	918.1	52.3	379.7	863.8	991.6	999.1
	$S_c(mm)$	44.8	103.2	239.5	420.5	598.6	48.5	171.9	465.5	751.8	910.8	50	376.1	860.8	991	999.1
50	$S_c(ml)$	50.2	114.1	276.6	482.4	671.4	50.5	195.8	517.4	797.2	937.3	53.3	420.7	893.3	994.3	999.9
	$S_c(mm)$	44.9	102.7	250.7	444	628.8	48.1	184.7	498.4	782.5	927	52.1	417	889.7	993.6	999.9

Table 3.3: $10^3\times$ empirical levels: α=0.05; based on 10,000 replications

		m=7			m=15			m=20			m=30			m=40		
		c			c			c			c			c		
$n_1 = n_2 = n_3$	Statistic	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50
10	$S_c(ml)$	45.4	47	44.6	53.2	44.2	46.6	48	46.7	47	52.8	48.3	43	48.3	43.9	43.9
	$S_c(mm)$	44.7	42.5	42.3	48	40.5	43.4	43.2	41.8	40.8	45.8	43.8	37.7	43	40.9	40.2
20	$S_c(ml)$	49.6	43.6	45.7	49.9	52.4	49.2	52.1	46.6	51.2	49.3	48.7	49.4	51.5	48.9	46.8
	$S_c(mm)$	48.1	42.8	43.9	46.8	50.2	49.3	49.5	45.4	48.2	46.4	47	49.3	48.8	47.3	44
50	$S_c(ml)$	51.5	52.6	51.5	51	52.3	51.6	51.2	47.6	49.7	49.9	45.9	50.9	46.5	50	49.8
	$S_c(mm)$	50.4	51.5	49	50.7	50.6	49.2	50	47.4	48.7	48.5	43.3	49.9	46.2	49.4	47.5

Table 3.4: $10^3\times$ empirical power corresponding to nominal significance level α=0.05;based on 10,000 replications for 3 groups; $m_1 = m,\ m_2 = m + \phi_1,\ m_3 = m + \phi_2,\ \delta_1 = \phi_1/m,\ \delta_2 = \phi_2/m$; c=0.25

		$n_1 = n_2 = n_3 = 10$ δ					$n_1 = n_2 = n_3 = 20$ δ					$n_1 = n_2 = n_3 = 50$ δ				
m	Statistic	(0,0)	(0,0.2)	(0.2,0.4)	(0.4, 0.6)	(0.4,0.8)	(0,0)	(0,0.2)	(0.2,0.4)	(0.4, 0.6)	(0.4,0.8)	(0,0)	(0,0.2)	(0.2,0.4)	(0.4, 0.6)	(0.4,0.8)
10	$S_c(ml)$	43.1	95.6	163.9	296.3	449.7	49.5	155.3	328.7	613.9	809.6	51.5	352.1	724.6	963.7	996.1
	$S_c(mm)$	39.8	86.8	150.5	274.2	418.1	47.9	148.5	319.2	595.3	795.9	50.3	346.9	717	961.6	996
20	$S_c(ml)$	46.7	96.4	188.8	342.4	508.9	46.6	168.9	365.6	669	859.4	47.6	393	784.5	981.4	998.6
	$S_c(mm)$	41.8	86.5	171.9	314.1	469.3	45.4	162.9	351.2	656.4	844	47.4	389.7	779	980.1	998.6
50	$S_c(ml)$	46.1	106.4	201	357.9	542.6	48.5	192	412.6	712.2	881	48.9	433.7	827	989.9	999.5
	$S_c(mm)$	42.3	96	180.5	327.8	506.8	46.5	183.1	398	697.1	866.9	47.9	426.9	821.4	989.7	999.4

3.1.6 Discussion and Conclusion

The results in table 3.1 shows that both statistics $Sc(ml)$ and $Sc(mm)$ maintain the significance level well. From table 3.3 we observe that $Sc(ml)$ performs better compared to $Sc(mm)$ when the sample size is small. However, for larger sample size ($n = 50$) $Sc(mm)$ also performs well. The results in table 3.2 and table 3.4 shows the performance of these statistics in terms of power. From table 3.2 we observe that the level of significance of $Sc(ml)$ is better than that for $Sc(mm)$ in almost all the cases. The power of $Sc(ml)$ is always higher than those for $Sc(mm)$. Similar results are observed in table 3.4. In both tables, a significant increase in power is achieved as we increase the sample size. Based on our simulation study $Sc(m)$ is recommended for the analysis of one way layout of count data with negative binomial distribution. This conclusion is in agreement with that Barnwal and Paul (1988).

Chapter 4

Analysis of One-way Layout of Count Data: Complete data with Semi-parametric Models

4.1 Introduction

Discrete data in the form of counts arise in many health science disciplines such as biology and epidemiology. For examples of discrete count data see Deng and Paul (2000, 2005); Anscombe (1949); Bliss and Fisher (1953); McCaughran and Arnold (1976); Margolin, Kaplan, and Zeiger (1981); Bohning, Dietz, Schlattmann, Mendonca,L., and Kirchner (1999).

Poisson models are widely used in the regression analysis of count data. The Poisson distribution has a property that the mean and the variance are equal. However, in practice many count data often display extra-variation or over/under dispersion relative to a Poisson model. Thus Poisson distribution is not an ideal choice for analysing count data in many applications. One very convenient and common model to accommodate this extra dispersion is the two parameter negative binomial distribution.

For applications of the negative binomial distribution, see, for example Engel (1984); Breslow (1984); Lawless (1987); Margolin et al. (1989); Collings and Margolin (1985). Different authors have used different parameterizations for the negative binomial distribution. For example, see, Paul and Plackett (1978); Barnwal and Paul (1988); Paul and Banerjee(1998); Piegorsch (1990); Paul and Deng (2000), and Deng and Paul (2005).

However, in many practical data analysis situations, the full distributional assumption becomes too restrictive and one can perform robust analysis using some semi-parametric models which require specification of only the first two moments of the counts. To avoid the full distributional assumption Wedderburn (1974) introduced the quasi-likelihood based on the assumption of only the first two moments of the response variable.The quasi-likelihood methodology is useful for estimating only mean or the regression parameters. Nelder and Pregibon (1987) proposed the extended quasi-likelihood which can be used to jointly estimate the mean and the dispersion parameter. Lee and Nelder (2001) introduced the double extended quasi-likelihood for the joint estimation of the mean and the dispersion parameters.

Saha (2008) developed test statistics for the homogeneity of the means of several treatment groups of count data in presence of over/under dispersion when there is no likelihood available. The $C(\alpha)$ statistics were derived based on the semi-parametric models quasi-likelihood, extended quasi-likelihood, and double extended quasi-likelihood and compared to the $C(\alpha)$ statistic based on the negative binomial model in terms of size and power. The $C(\alpha)$ statistic based on double extended quasi-likelihood is recommended.

Here we study the performance the statistics based on the semi-parametric models in terms of size and power.

4.2 The Likelihood

4.2.1 The quasi-likelihood

The quasi-likelihood methodology of Wedderburn(1974) is based on the knowledge of the form of first two moments of the random variable Y_{ij}, which coincides with those based on the negative binomial model. The quasi-log-likelihood (see Breslow, 1990) for the counts Y_{ij} $(i = 1, 2, ..., K,\ j = 1, 2, ..., n_i)$ is given by

$$Q = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\left[\Big(y_{ij}+\frac{1}{c_i}\Big)\ln\Big(\tfrac{(1+c_iy_{ij})}{1+c_im_i}\Big) - y_{ij}\ \ln\ \Big(\tfrac{y_{ij}}{m_i}\Big)\right], \tag{4.2.1}$$

where the mean-variance relationship is

$$E(Y_{ij}) = m_i \quad \text{and } \mathrm{var}(Y_{ij}) = m_i(1 + c_im_i).$$

4.2.2 The extended quasi-likelihood

The quasi-likelihood is useful for estimating only the mean or the regression parameters. By introducing a normalizing factor to the quasi-likelihood, Nelder and Pregibon (1987) and Godambe and Thompson (1989) proposed the extended quasi-likelihood (EQL), which can be used for the simultaneous estimation of the parameters m_i and c_i. The modified extended quasi-log-likelihood (for details see

Clark and Perry, 1989), apart from a constant, becomes

$$\begin{aligned} Q^{+*} = & \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{1}{2}\ln\ \{1+c_iy_{ij}+c_i/6\} - \tfrac{1}{2}\ln\{(y_{ij}+1/6)(1+c_iy_{ij})^2(1+c_i/6)\} \\ & +\Big(y_{ij}+\tfrac{1}{c_i}\Big)\ln\Big(\tfrac{1+c_iy_{ij}}{1+c_im_i}\Big) - y_{ij}\ \ln\Big(\tfrac{y_{ij}}{m_i}\Big)\Big]. \end{aligned} \tag{4.2.2}$$

4.2.2.1 The double extended quasi-likelihood

In a generalized linear model setup Lee and Nelder (2001) developed hierarchical likelihood procedure for the joint estimation of the mean and the variance components. For the situation in which a full distributional assumption is not available, Lee and Nelder(2001) introduced the double extended quasi-likelihood (DEQL) for estimation of the mean and the dispersion parameters of the response variable. For joint estimation of the mean and the dispersion parameters the DEQL has been derived by Paul and Saha (2007). Omitting details of the derivation, the profile DEQL with the modified Stirling approximation is

$$\begin{aligned} p_v{}^*(DEQ) = & \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[y_{ij}\ln(m_i) + \Big(y_{ij}+\tfrac{1}{c_i}\Big)\ln\Big(\tfrac{1+c_iy_{ij}}{1+c_im_i}\Big) - \tfrac{1}{2}\ln(1+c_iy_{ij}) \\ & -\Big(y_{ij}+\tfrac{1}{2}\Big)\ln(y_{ij}) + \tfrac{c_i}{12(1+c_iy_{ij})} - \tfrac{c_i}{12} - \tfrac{1}{12y_{ij}} - \tfrac{1}{2}\ln(2\pi)\Big]. \end{aligned} \tag{4.2.3}$$

4.3 Estimation

We are interested in testing $H_0 : m_1 = m_2 = \cdots m_K = m$ against H_1 : not all $m_i^{'}$s are equal for all $c > -1/m$.

4.3.1 Quasi-likelihood estimates

The quasi-log-likelihood under H_0 is

$$Q_0 = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\left[\Big(y_{ij}+\frac{1}{c}\Big)\ln\Big(\tfrac{(1+cy_{ij})}{1+cm}\Big) - y_{ij}\ \ln\ \Big(\tfrac{y_{ij}}{m}\Big)\right]. \tag{4.3.1}$$

The estimating equation for m is

$$\tfrac{\partial Q_0}{\partial m} = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\left[\frac{y_{ij}}{m} - \frac{(1+cy_{ij})}{(1+cm)}\right] = 0. \tag{4.3.2}$$

No such estimating function exists for c. However, following Breslow (1984, 1990) and Saha (2008) an unbiased estimating function for c can be obtained as

$$V(m,c) = \ \sum_{i=1}^{K}\sum_{j=1}^{n_i}\frac{(y_{ij}-m)^2}{m(1+cm)} - (n-K). \tag{4.3.3}$$

4.3.2 Extended Quasi-likelihood estimates

The modified extended quasi-log-likelihood under H_0 is

$$\begin{aligned} Q_0{}^{+*} = \ & \sum_{i=1}^{K}\sum_{j=1}^{n_i}\left[\frac{1}{2}\ln\ \{1+cy_{ij}+c/6\} - \tfrac{1}{2}\ln\{(y_{ij}+1/6)(1+cy_{ij})^2(1+c/6)\}\right. \\ & \left. +\Big(y_{ij}+\tfrac{1}{c}\Big)\ln\Big(\tfrac{1+cy_{ij}}{1+cm}\Big) - y_{ij}\ \ln\Big(\tfrac{y_{ij}}{m_i}\Big)\right]. \end{aligned} \tag{4.3.4}$$

The estimating equations for m and c are

$$\tfrac{\partial Q_0{}^{+*}}{\partial m} = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\left[\frac{y_{ij}}{m} - \frac{(1+cy_{ij})}{(1+cm)}\right] = 0 \tag{4.3.5}$$

and

$$\begin{aligned}\frac{\partial {Q_0}^{+*}}{\partial c} = & \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{y_{ij}-m}{c(1+cm)} + c^{-2}\ln\Big(\tfrac{1+cm}{1+cy_{ij}}\Big) + \tfrac{1+6y_{ij}}{2(c+6+6cy_{ij})} \\ & -\tfrac{y_{ij}}{1+cy_{ij}} - \tfrac{1}{2(c+6)}\Big] = 0.\end{aligned} \tag{4.3.6}$$

The maximum extended quasi-likelihood estimate of m obtained from the first equation above is $\hat{m} = \bar{y}$. The maximum extended quasi-likelihood estimate $\hat{c}_{eql}$ of c is obtained by iteratively solving the second equation after replacing m by $\bar{y}$.

4.3.3 The Double Extended Quasi-likelihood estimates

The profile DEQL under H_0 is

$$\begin{aligned}{p_0}^*(DEQ) = & \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[y_{ij}\ln(m) + \Big(y_{ij}+\tfrac{1}{c_i}\Big)\ln\Big(\tfrac{1+c_i y_{ij}}{1+cm}\Big) - \tfrac{1}{2}\ln(1+cy_{ij}) \\ & -\Big(y_{ij}+\tfrac{1}{2}\Big)\ln(y_{ij}) + \tfrac{c}{12(1+cy_{ij})} - \tfrac{c}{12} - \tfrac{1}{12y_{ij}} - \tfrac{1}{2}\ln(2\pi)\Big].\end{aligned} \tag{4.3.7}$$

The estimating equations for m and c are

$$\frac{\partial {p_0}^*(DEQ)}{\partial m} = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{y_{ij}}{m} - \frac{(1+cy_{ij})}{(1+cm)}\Big] = 0 \tag{4.3.8}$$

and

$$\begin{aligned}\frac{\partial {p_0}^*(DEQ)}{\partial c} = & \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{y_{ij}-m}{c(1+cm)} + \frac{1}{c^2}\ln\big(\tfrac{1+cm}{1+cy_{ij}}\big) \\ & -\tfrac{y_{ij}}{2(1+cy_{ij})} - \tfrac{cy_{ij}(2+cy_{ij})}{12(1+cy_{ij})^2}\Big] = 0.\end{aligned} \tag{4.3.9}$$

The maximum double extended quasi-likelihood estimate of m obtained from the first equation above is $\hat{m} = \bar{y}$. The maximum double extended quasi-likelihood estimate $\hat{c}_{deql}$ of c is obtained by iteratively solving the second equation after replacing m by $\bar{y}$.

4.4 Testing Hypothesis

In this section we develop procedures for testing the composite hypotheses $H_0 : m_1 = \cdots = m_K$ against H_1: at least two m_i's are not the same, with the assumption $c_1 = \cdots = c_K = c$, where c is unknown and $c > -1/m_i$. For the convenience of the derivation of the $C(\alpha)$ statistics we reparameterize m_i under H_1 by $m_i = m + \delta_i,\ i = 1, ..., K-1$, with $\delta_K = 0$. Then testing $H_0 : m_1 = ... = m_K = m$ reduces to testing $H_0 : \delta_1 = ... = \delta_{K-1}$, where m and c are treated as nuisance parameters. This technique was employed by many authors. For example, Tarone (1985) used this technique to obtain $C(\alpha)$ statistic for testing the equality of several odds ratios. Barnwal and Paul (1988) used this same technique to derive these statistic for testing equality of means in the presence of a common negative binomial over-dispersion.

4.4.1 The C(α) statistic based on quasi-likelihood

The quasi-log-likelihood in terms of reparameterization of $m_i = m + \delta_i$ and $c_i = c$, apart form some constant terms, can be written as

$$Q = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\left[(y_{ij} + c^{-1})\ln\left(\tfrac{1+cy_{ij}}{1+c(m+\delta_i)}\right) - y_{ij}\ln\left(\tfrac{y_{ij}}{m+\delta_i}\right)\right]. \qquad (4.4.1)$$

Given c the unbiased estimating functions for the parameters $\delta_1, \delta_2, \cdots, \delta_{K-1}, m$ are , $U_i(\delta_i, m, c) = \frac{\partial Q}{\partial \delta_i},\ \ i = 1, 2, \cdots K-1$ and $V_1(\delta_i, m, c) = \frac{\partial Q}{\partial m}$. No such estimating function exists for c. However, following Breslow (1984, 1990) and Moore and Tsiatis (1991), given $\delta_1, \delta_2, \cdots, \delta_{K-1}, m$, an unbiased estimating function for

c can be obtained as

$$V_2(\delta_i, m, c) = \sum_{i=1}^{K}\sum_{j=1}^{n_i} \frac{(y_{ij} - m - \delta_i)^2}{(m+\delta_i)(1+cm+c\delta_i)} - (n-K), \tag{4.4.2}$$

where $n = \sum_{i=1}^{K} n_i$. We obtain the $C(\alpha)$ statistic based on the Quasi-likelihood, treating $U_i,\ i = 1, 2, \cdots, K-1, V_1$, and V_2 as the likelihood score analogs. Now, define $\boldsymbol{\delta} = (\delta_1, ..., \delta_{K-1})$ and $\boldsymbol{\nu} = (\nu_1, \nu_2)' = (m,\ c)'$. Following the theory in §2.4 we obtain

$U_i = \frac{\partial Q}{\partial \delta_i}\Big|_{\delta=0} = \frac{n_i(\bar{y_{i.}} - m)}{m(1+cm)}, \quad i = 1, ..., K-1,$

$V_1(\delta_i, m, c) = \frac{\partial Q}{\partial \nu_1}\Big|_{\delta=0} = \sum_{i=1}^{K} \dfrac{n_i(\bar{y_{i.}} - m)}{m(1+cm)},$

where $\bar{y_{i.}} = (1/n_i)\sum_{j=1}^{n_i} y_{ij}$ is the sample mean of the i^{th} treatment. The $C(\alpha)$ test is based on $\lambda_i(\hat{\nu}) = U_i(\hat{\nu}) - \beta_{1i}V_1(\hat{\nu}) - \beta_{2i}V_2(\hat{\nu})$, where β_{1i} and β_{2i} are, respectively, the partial regression coefficient of δ_i on V_1 and δ_i on V_2, where $\hat{\nu}$ is some $\sqrt{n}$ (where $n = \sum_{i=1}^{K} n_i$) consistent estimator of ν under the null hypothesis. The regression coefficients $\beta = (\beta_1, \beta_2)$ with $\beta_1 = (\beta_{11}, ..., \beta_{1K-1})$ and $\beta_2 = (\beta_{21}, ..., \beta_{2K-1})$ are obtained as $\Gamma\gamma^{-1}$, and the variance-covariance of $\Lambda(\nu) = [\lambda_1(\nu), ..., \lambda_{K-1}(\nu)]'$ is $\Delta - \Gamma\gamma^{-1}\Gamma'$, where the (s,t)th components of Δ, Γ, and γ are

$$\Delta_{st} = E\Big(-\frac{\partial^2 Q}{\partial\delta_s\partial\delta_t}\Big|_{\delta=0}\Big) = \begin{cases} n_s/m(1+cm), & s = t = 1, \ldots, K-1, \\ 0 & \text{otherwise,} \end{cases}$$

$$\Gamma_{st} = E\Big(-\frac{\partial^2 Q}{\partial\delta_s\partial\nu_t}\Big|_{\delta=0}\Big) = \begin{cases} n_s/m(1+cm) & , s = 1, ..., K-1, t = 1, \\ 0, & s = 1, ..., K-1, t = 2, \end{cases}$$

and

$$\gamma_{st} = E\Big(-\frac{\partial^2 Q}{\partial \nu_s \partial \nu_t}\Big|_{\delta=0} \Big) = \begin{cases} n/m(1+cm), & s=t=1, \\ b, & s=t=2, \\ 0 & \text{otherwise}, \end{cases}$$

respectively. Note that the above terms involve the nuisance parameter $\nu = (m, c)$. Thus, using $\hat{\nu}$ in Λ, Δ, Γ, and γ, the $C(\alpha)$ statistic is obtained as $\Lambda^{'}(\Delta - \Gamma\gamma^{-1}\Gamma^{'})^{-1}\Lambda$, which is approximately distributed as a chi-square distribution with K-1 degrees of freedom. After replacing $\hat{\nu} = (\hat{m}, \hat{c}_{mm})$, the $C(\alpha)$ statistic, after some algebra, becomes

$$C(ql) = \sum_{i=1}^{K} \frac{n_i(\bar{y_{i.}} - \hat{m})^2}{\hat{m}(1+\hat{m}c_{\hat{mm}})},$$

where $\hat{m} = \bar{y} = \sum_{i=1}^{K} \frac{\bar{y_{i.}}}{n}$, $\hat{c}_{mm} = (s^2 - \bar{y})/\bar{y}^2$, and $s^2 = \sum_{i=1}^{K}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2/(n-1)$. The derivation is given in appendix B. This $C(\alpha)$ statistic can also be obtained by method of moments (see Barnwal and Paul, 1988).

4.4.2 The C(α) statistic based on extended quasi-likelihood

Using the parameters $\delta_1, ..., \delta_{K-1}, m$, and c, the modified extended quasi-likelihood, apart from a constant term, is obtained as

$$\begin{aligned} Q^{+*} = \; & \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{1}{2}\ln\{1 + ay_{ij} + a/6\} - \tfrac{1}{2}\ln\{(y_{ij}+1/6)(1+ay_{ij})^2(1+a/6)\} \\ & + \Big(y_{ij} + \tfrac{1}{a}\Big)\ln\Big(\tfrac{1+ay_{ij}}{1+a(m+\delta_i)}\Big) - y_{ij}\ln\Big(\tfrac{y_{ij}}{m+\delta_i}\Big)\Big]. \end{aligned} \tag{4.4.3}$$

Now, similar to the procedure in Section 4.4.1 and using Q^{+*} as the log-likelihood of $\delta_1, ..., \delta_{K-1}, m$, and c, it can be shown that the $C(\alpha)$ statistic, $S_c(eql)$, based on the modified extended quasi-likelihood is

$$S_c(eql) = \sum_{i=1}^{K} \frac{n_i(\bar{y_{i.}} - \hat{m})^2}{\hat{m}(1 + \hat{m}\hat{c_{eql}})},$$

where $\hat{m} = \bar{y} = \sum_{i=1}^{K} \frac{\bar{y_{i.}}}{n}$ and $\hat{c}_{eql}$ is the maximum extended quasi-likelihood estimate of c, under H_0, obtained by solving
$\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left[\frac{y_{ij} - \hat{m}}{c(1 + c\hat{m})} + c^{-2}\ln\left(\frac{1+c\hat{m}}{1+cy_{ij}}\right) + \frac{1+6y_{ij}}{2(c+6+6cy_{ij})} - \frac{y_{ij}}{1+cy_{ij}}\right] = \frac{n}{2(c+6)}.$
The derivation is given in appendix C. This $C(\alpha)$ statistic has been derived by Saha (2008).

4.4.3 The C(α) statistic based on double extended quasi-likelihood

The double extended quasi-likelihood excluding a constant term, using the reparameteriztion of m_i under H_1, can be written as

$$\begin{aligned} p_v{}^*(DEQ) = & \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[y_{ij}\ln(m + \delta_i) + \Big(y_{ij} + \tfrac{1}{c}\Big)\ln\Big(\tfrac{1+cy_{ij}}{1+c(m+\delta_i)}\Big) - \tfrac{1}{2}\ln(1 + cy_{ij}) \\ & + \tfrac{c}{12(1+cy_{ij})} - \tfrac{c}{12} - \tfrac{1}{12y_{ij}}\Big]. \end{aligned} \tag{4.4.4}$$

Now, similar to the procedure in Section 4.4.1 and using $p_v{}^*(DEQ)$ as the log-likelihood of $\delta_1, ..., \delta_{K-1}, m,$ and c, it can be shown that the $C(\alpha)$ statistic, $S_c(deql)$, based on the double extended quasi-likelihood is

$$S_c(deql) = \sum_{i=1}^{K} \frac{n_i(\bar{y_{i.}} - \hat{m})^2}{\hat{m}(1 + \hat{m}\hat{c_{deql}})} \ ,$$

where $\hat{m} = \bar{y} = \sum_{i=1}^{K} \frac{\bar{y_{i.}}}{n}$ and $\hat{c}_{deql}$ is the maximum double extended quasi-likelihood estimate of c, under H_0, obtained by solving

$$\sum_{i=1}^{K} \sum_{j=1}^{n_i} \Big[\frac{y_{ij} - \hat{m}}{c(1 + c\hat{m})} + \frac{1}{c^2} \ln\big(\tfrac{1+c\hat{m}}{1+cy_{ij}}\big) - \tfrac{y_{ij}}{2(1+cy_{ij})} - \tfrac{cy_{ij}(2+cy_{ij})}{12(1+cy_{ij})^2} \Big] = 0. \qquad (4.4.5)$$

The derivation is given in appendix D. This $C(\alpha)$ statistic has been derived by Saha (2008).

4.5 Simulations

A simulation study is conducted to examine the comparative behaviour of the test statistics $S_{c(ql)}$, $S_{c(eql)}$, and $S_{c(deql)}$ in terms of size and power. The test of equality of means of two and three negative binomial distributions with common dispersion parameter is considered. Empirical significance levels and power of the tests were all based on 10,000 samples from the negative binomial distribution for different values of m and c. For all combinations of $m = 7, 15, 20, 30, 40$ and $c = 0.05, 0.25, 0.50$ simulations are run based on 10,000 samples of sizes $n_1 = n_2 = 10, 20,$ and 50 for two groups. The values of parameters are same for $n_1 = n_2 = n_3 = 10, 20,$ and 50 for three groups. Table 4.1 displays empirical levels based on a nominal significance level of $\alpha = 5\%$ for $S_c(ql)$, $S_c(eql)$, and $S_c(deql)$ for data generated from the NB distributions for two groups. Table 4.2 displays empirical power based on a nominal significance level of $\alpha = 5\%$ for $S_c(ql)$, $S_c(eql)$, and $S_c(deql)$ for data generated from the NB distributions. For two groups, $c = 0.25$, $m_1 = m$, $m_2 = m + \phi$, and $\delta = \phi/m$, where $m = 10, 20, 50$ and $\delta = 0.0, 0.2, 0.4, 0.6, 0.8$ .The sample sizes considered were $n_1 = n_2 = 10, 20,$ and 50. Table 4.3 displays empirical levels based on a nominal significance level of $\alpha = 5\%$ for for $S_c(ql)$, $S_c(eql)$, and $S_c(deql)$. For each replication three samples are generated from the NB(m,c). Table 4.4 displays empirical power based on a nominal significance level of $\alpha = 5\%$ for $S_c(ql)$, $S_c(eql)$, and $S_c(deql)$. In this case also for each replication three samples are generated from the NB(m,c). For three groups, $c = 0.25$, $m_1 = m$, $m_2 = m + \phi_1$, $m_3 = m + \phi_2$, $\delta_1 = \phi_1/m$, $\delta_2 = \phi_2/m$, where $m = 10, 20, 50$ and $\boldsymbol{\delta} = (0,0), (0,0.2), (0.2,0.4), (0.4,0.6), (0.4,0.8)$. The sample sizes considered were $n_1 = n_2 = n_3 = 10, 20$, and 50.

Table 4.1: $10^3\times$ empirical levels: α=0.05; based on 10,000 replications

		m=7			m=15			m=20			m=30			m=40		
		c			c			c			c			c		
$n_1 = n_2$	Statistic	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50
10	$S_c(ql)$	43.2	46.5	43.0	44.6	43.8	40.5	43.8	44.8	45.1	43.7	44.4	44.8	48.2	42.9	45.2
	$S_c(eql)$	43.4	49.2	45.5	48.1	49.4	41.9	51.5	49.2	49.6	48.7	48.5	49.4	54.8	49	51.3
	$S_c(deql)$	43.6	49.3	46.6	48.1	49.5	42.5	51.5	49.5	50.5	48.7	49.1	50.1	54.8	49.2	51.7
20	$S_c(ql)$	47.4	51	44.4	53.3	49.5	46.5	52.8	48.5	47.2	51.6	50.8	46.7	50.2	48.9	45.3
	$S_c(eql)$	48.2	51.4	43.7	56.5	52.7	48.1	55.1	50.1	48.5	54.5	52.5	48.8	52.8	51.3	47.6
	$S_c(deql)$	48.2	51.8	44.8	50.6	53	48.5	55.1	50.2	49.3	54.5	52.6	49.4	52.8	51.6	48.7
50	$S_c(ql)$	52.3	51	49.9	49.5	49.7	47.3	51.7	50	50.2	52.4	48.3	46.8	51.3	52	53.3
	$S_c(eql)$	52.9	51.8	48.9	50.4	51.2	47.2	53.4	52.2	50.5	53.7	50.4	49.7	52.6	53	53.4
	$S_c(deql)$	52.9	52.1	49.7	50.4	51.2	48	53.6	52.4	51.1	53.7	50.8	50.3	52.6	53.2	53.8

Table 4.2: $10^3\times$ empirical power corresponding to nominal significance level α=0.05; based on 10,000 replications for 2 groups; $m_1 = m,\ m_2 = m + \phi,\ \delta = \phi/m$; c=0.25

		$n_1 = n_2 = 10$					$n_1 = n_2 = 20$					$n_1 = n_2 = 50$				
		δ					δ					δ				
m	Statistic	0.0	0.2	0.4	0.6	0.8	0.0	0.2	0.4	0.6	0.8	0.0	0.2	0.4	0.6	0.8
10	$S_c(ql)$	43.5	88.4	207.3	376.5	547	46.9	154.5	409	688.2	875	51.3	340.1	820	980.9	998.6
	$S_c(eql)$	48.7	97.7	226.5	402.4	578.9	49.5	158.5	425	702.4	885.8	52.4	340.3	822.5	981.9	998.9
	$S_c(deql)$	49.1	98.4	227.2	404	585.5	49.6	158.9	425.8	703.5	886.8	52.8	341.8	824	982	998.9
20	$S_c(ql)$	44.8	103.2	239.5	420.5	598.6	48.5	171.9	465.5	751.8	910.8	50	376.1	860.8	991	999.1
	$S_c(eql)$	49.2	114.4	266.3	451.7	635.7	50.1	179.3	483.3	765.6	917.9	52.2	379	863.1	991.6	999.1
	$S_c(deql)$	49.5	114.7	266.9	452.9	637.2	50.2	180	484.3	766.7	918.3	52.4	380	863.8	991.6	997.9
50	$S_c(ql)$	44.9	102.7	250.7	444	628.8	48.1	184.7	498.4	782.5	927	52.1	417	889.7	993.6	999.9
	$S_c(eql)$	50.1	114	276	481.3	670.4	50.5	195.4	516.3	796.6	937.1	53.3	419.8	892.9	994.2	999.9
	$S_c(deql)$	50.2	114.1	277	483.1	671.5	50.6	195.8	517.5	797.4	937.3	53.3	420.9	893.3	994.3	999.9

Table 4.3: $10^3\times$ empirical levels: α=0.05; based on 10,000 replications

		m=7			m=15			m=20			m=30			m=40		
		c			c			c			c			c		
$n_1 = n_2 = n_3$	Statistic	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50
10	$S_c(ql)$	44.7	42.5	42.3	48	40.5	43.4	43.2	41.8	40.8	45.8	43.8	37.7	43	40.9	40.2
	$S_c(eql)$	45.4	46.4	44.4	53.2	44.1	45.5	47.9	46.6	46.5	52.8	47.8	42.3	48.3	43.7	43.2
	$S_c(deql)$	45.5	47.1	45.2	53.2	44.2	46.8	48	46.7	47.3	52.8	48.4	43.5	48.3	43.9	44
20	$S_c(ql)$	48.1	42.8	43.9	46.8	50.2	49.3	49.5	45.4	48.2	46.4	47	49.3	48.8	47.3	44
	$S_c(eql)$	49. 6	43.5	44.7	49.9	52.1	48.7	52.1	46	50.7	49.3	48.3	48.9	51.5	48.5	46.2
	$S_c(deql)$	49.6	43.7	46.5	49.9	52.4	49.4	52.1	46.7	51.4	49.3	48.7	49.7	51.5	48.9	47.1
50	$S_c(ql)$	50.4	51.5	49	50.7	50.6	49.2	50	47.4	48.7	48.5	43.3	49.9	46.2	49.4	47.5
	$S_c(eql)$	51.5	52	50.9	51	52.1	50.8	51.2	47.5	48.9	49.8	45.7	50.4	46.5	50	49.2
	$S_c(deql)$	51.5	52.7	51.7	51	52.3	52.2	51.2	47.7	49.7	49.9	46.1	51.4	46.5	50	50

Table 4.4: $10^3\times$ empirical power corresponding to nominal significance level α=0.05; based on 10,000 replications for 3 groups; $m_1 = m,\ m_2 = m + \phi_1,\ m_3 = m + \phi_2,\ \delta_1 = \phi_1/m,\ \delta_2 = \phi_2/m$; c=0.25

		$n_1 = n_2 = n_3 = 10$					$n_1 = n_2 = n_3 = 20$					$n_1 = n_2 = n_3 = 50$				
		δ					δ					δ				
m	Statistic	(0,0)	(0,0.2)	(0.2,0.4)	(0.4, 0.6)	(0.4,0.8)	(0,0)	(0,0.2)	(0.2,0.4)	(0.4, 0.6)	(0.4,0.8)	(0,0)	(0,0.2)	(0.2,0.4)	(0.4, 0.6)	(0.4,0.8)
10	$S_c(ql)$	39.8	86.8	150.5	274.2	418.1	47.9	148.5	319.2	595.3	795.9	50.3	346.9	717	961.6	996
	$S_c(eql)$	42.7	95	163	294.7	447.9	49.4	154.6	327.9	612.8	808.4	51.1	351	723.9	963.5	996.1
	$S_c(deql)$	46.7	96.4	188.9	342.5	511.6	46.7	169	365.7	669.1	860.9	47.7	393.1	784.7	981.4	998.6
20	$S_c(ql)$	41.8	86.5	171.9	314.1	469.3	45.4	162.9	351.2	656.4	844	47.4	389.7	779	980.1	998.6
	$S_c(eql)$	46.6	96	188	341.1	507.5	46	168.4	365.1	667.6	858.4	47.5	392.5	783.9	981.2	998.6
	$S_c(deql)$	46.7	96.4	188.9	342.5	511.6	46.7	169	365.7	669.1	860.9	47.7	393.1	784.7	981.4	998.6
50	$S_c(ql)$	42.3	96	180.5	327.8	506.8	46.5	183.1	398	697.1	866.9	47.9	426.9	821.4	989.7	999.4
	$S_c(eql)$	45.7	105.8	199.8	356.7	541.5	48.4	191.5	412.1	711.3	880.7	48.7	433.1	826.4	989.8	999.5
	$S_c(deql)$	46.2	106.5	201.1	358.2	543.6	48.7	192.1	412.8	712.3	881.2	49	433.7	827.1	989.9	999.5

4.6 Discussion and Conclusion

From the results in table 4.1 we observe that all three statistics $S_c(ql)$, $S_c(eql)$, and $S_c(deql)$ maintain the significance level well. Data departure has no effect on size performances of the statistics. Similar results were observed from table 4.3. For the larger sample size (n=50) all the statistics maintain level quite well. Table 4.2 presents the power performance of all three statistics for two groups. For small to moderate sample size, it is observed that $S_c(ql)$ has lower levels compared to that of other two statistics. However, for large sample size (n=50) it maintains the level well and close to other two statistics. Though these statistics maintain the level well, we observe that in almost all data situations the power estimates of $S_c(deql)$ is higher than that of $S_c(ql)$ and $S_c(eql)$. Table 4.4 also shows that the power performance of $S_c(deql)$ is better than the other two statistics. In both the tables significant increase in power is achieved with the increase in sample size. Based on our simulation study $S_c(deql)$ is recommended. This conclusion is in agreement with that of Saha (2008).

Chapter 5

Effect of missing responses on the $C(\alpha)$ or score tests in One-way Layout of Count Data

5.1 Introduction

Discrete data in the form of counts arise in many health science disciplines such as biology and epidemiology. For examples of discrete count data see Anscombe (1949); Bliss and Fisher (1953); McCaughran and Arnold (1976); Margolin et al. (1981); Böhning et al. (1999), Paul and Deng (2000) and Deng and Paul (2005).

Poisson models are widely used in the regression analysis of count data. The Poisson distribution has a property that the mean and the variance are equal. However, in practice, many count data often display extra-variation or over/under dispersion relative to a Poisson variance. Thus Poisson distribution is not an ideal choice for analysing count data in many applications. One very convenient and common model to accommodate this extra dispersion is the two parameter negative binomial distribution.

For applications of the negative binomial distribution, see, for example, Engel (1984); Breslow (1984); Collings and Margolin (1985); Lawless (1987); Margolin et al. (1989). Different authors have used different parameterizations for the negative binomial distribution. For example, see, Paul and Plackett (1978); Barnwal and Paul (1988); Piegorsch (1990); Paul and Banerjee (1998); Paul and Deng (2000), and Deng and Paul (2005).

One-way layout of count data having over/under dispersion arise in many practical situations. Barnwal and Paul (1988) developed and studied two $C(\alpha)$ statistics (Neyman, 1959) under the negative binomial assumption and three other statistics for testing the equality of means of several groups of count data in presence of a common dispersion parameter and recommend the $C(\alpha)$ statistics. For the same purpose Saha (2008) developed three $C(\alpha)$ type statistics for situations in which the negative binomial assumption might be violated. Of these statistics, they recommend $C(\alpha)$ type statistic based on the double extended quasi-likelihood.

Count data may further be complicated by the existence of missing values. Extensive work has been done on the analysis of continuous response data under normality assumption. See, for example, Rubin (1976), Little and Rubin (1987), Anderson and Taylor (1976), Geweke (1986), Raftery, Madigan, and Hoeting (1997), Chen, Hubbard, and Rubin (2001), Kelly (2007), and Zhang and Huang (2008).

Some work on missing values has also been done on logistic regression analysis of binary data. See , for example , Ibrahim (1990); Lipsitz and Ibrahim (1996), Ibrahim and Lipsitz (1996); Ibrahim, Chen, and Lipsitz (1999); Ibrahim, Chen, and Lipsitz (2001); Sinha and Maiti (2008); Maiti and Pradhan (2009).

Recently, some work on the estimation of parameters of zero-inflated count data model and over-dispersed binomial model with missing responses has been done. See, for example, Mian and Paul (2016), Luo and Paul (2018).

Rubin (1976) and Little and Rubin (1987) discuss various missingness mechanisms. If the missingness does not depend on observed data, then the missing data are called missing completely at random (MCAR). If the missing data mechanism depends only on observed data, then the data are missing at random (MAR). The MAR is also known as ignorable missing, that is, in this case, the missing data mechanism is ignored. If the missing data mechanism depends on both observed and unobserved data, that is, failure to observe a value depends on the value that would have been observed, then the data are said to be missing not at random (MNAR) in which case the missingness is nonignorable. For more detailed discussion on missing data mechanism, see Ibrahim et al. (2005, p333).

In this paper we develop estimation procedures for the parameters involved in the one way layout of count data under different missing data scenarios and study the effect of missingness on the $C(\alpha)$ statistics recommended by Barnwal and Paul (1988) and that by Saha (2008) by Monte Carlo simulation.

In Section 2 we develop the estimation procedure. The score tests are given in Section 3 and a simulation study is conducted in Section 4. A discussion of the findings are given in Section 5.

5.2 Estimation of the Parameters

5.2.1 Maximum likelihood estimates

Let $Y_{ij},\ i = 1, ..., K,\ j = 1, ..., n_i$ be the counts for the j^{th} individual of the i^{th} treatment group. We assume that $Y_{ij} \sim NB(m_i, c)$, which has probability mass function

$$Pr(Y_{ij} = y_{ij}|m_i, c) = \frac{\Gamma(y_{ij}+c^{-1})}{y_{ij}!\Gamma(c^{-1})} \left(\frac{cm_i}{1+cm_i}\right)^{y_{ij}} \left(\frac{1}{1+cm_i}\right)^{c^{-1}} \tag{5.2.1}$$

for $y_{ij} = 0, 1, ...,$ and $m_i > 0$. The mean and variance of Y_{ij} are

$$E(Y_{ij}) = m_i \quad \text{and } \text{var}(Y_{ij}) = m_i(1 + cm_i), \tag{5.2.2}$$

where $c > -1/m_i$. See Paul and Placket (1978) and many other references later on. We are interested in testing $H_0 : m_1 = m_2 = \cdots m_K = m$ against H_1 : not all $m_i^{'}$s are equal for all $c > -1/m$.

Under the null hypothesis, then, the probability mass function is

$$Pr(Y_{ij} = y_{ij}|m, c) = \frac{\Gamma(y_{ij}+c^{-1})}{y_{ij}!\Gamma(c^{-1})} \left(\frac{cm}{1+cm}\right)^{y_{ij}} \left(\frac{1}{1+cm}\right)^{c^{-1}}. \tag{5.2.3}$$

As we are interested in studying as to whether missing values affect the $C(\alpha)$ or $C(\alpha)$-like statistics, we show all calculations for the estimations of the parameters under the null hypothesis of equality of the means. That is, we want to estimate only the common mean m and the common over/under dispersion parameter c. The log-likelihood, under which, is

$$l_0 = \sum_{i=1}^{K}\sum_{j=1}^{n_i} l_{ij}\ , \tag{5.2.4}$$

where

$$l_{ij} = y_{ij}\text{log}m - (y_{ij} + c^{-1})\text{log}(1 + cm) + \sum_{l=1}^{y_{ij}}\text{log}\{1 + c(l - 1)\}. \tag{5.2.5}$$

The maximum likelihood estimator of m is $\hat{m} = \bar{y} = \sum_{i=1}^{K} \frac{\bar{y_{i.}}}{n}$, where $n = \sum_{i=1}^{K} n_i$. The maximum likelihood estimator, denoted by c_M of c under H_0 is obtained as a solution to

$$n\log(1 + c\bar{y}) - \sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{l=1}^{y_{ij}} \left\{ \frac{c}{1 + c(l-1)} \right\} = 0. \tag{5.2.6}$$

5.2.2 Double extended quasi-likelihood estimates

In some situations a full distributional assumption cannot be assumed to be satisfied. In such situations, based on the assumption that a discrete random variable Y satisfies $E(Y) = m$ and $Var(Y) = m(1 + cm)$, where m is the mean and c is the dispersion (over/under) parameter, a few hypbrid likelihood types have been developed. See, for example, quasi-likelihood (Wedderburn, 1974), Extended quasilikelihood (Nelder and Pregibon, 1987), and double extended quasi-likelihood (Lee and Nelder, 2001). Saha (2008) developed the double extended quasi-likelihood estimate of m and c as in what follows.

In a generalized linear model setup Lee and Nelder (2001) developed hierarchical likelihood procedure for the joint estimation of the mean and the variance components. For the situation in which a full distributional assumption is not available, Lee and Nelder (2001) introduced the double extended quasi-likelihood (DEQL) for estimation of the mean and the dispersion parameters of the response variable. For joint estimation of the mean and the dispersion parameters the DEQL has been derived by Paul and Saha (2007). Omitting details of the derivation, the profile DEQL, denoted by Dl_0, is

$$Dl_0 = \sum_{i=1}^{K}\sum_{j=1}^{n_i} dl_{ij} \ , \tag{5.2.7}$$

where

$$\begin{aligned} dl_{ij} = \; & \Big[y_{ij}\ln(m) + \Big(y_{ij} + \tfrac{1}{c}\Big)\ln\Big(\tfrac{1+cy_{ij}}{1+cm}\Big) - \tfrac{1}{2}\ln(1+cy_{ij}) \\ & -\big(y_{ij} + \tfrac{1}{2}\Big)\ln(y_{ij}) + \tfrac{c}{12(1+cy_{ij})} - \tfrac{c}{12} - \tfrac{1}{12y_{ij}} - \tfrac{1}{2}\ln(2\pi)\Big]. \end{aligned} \tag{5.2.8}$$

Again, omitting details, the double extended quasi-likelihood estimate for the parameter m is $\hat{m} = \bar{y}$ and that for c, denoted by c_{deql}, is obtained by solving

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{1}{c^2}\ln\big(\tfrac{1+c\bar{y}}{1+cy_{ij}}\big) + \tfrac{y_{ij}-\bar{y}}{c(1+c\bar{y})} - \tfrac{y_{ij}}{2(1+cy_{ij})} - \tfrac{cy_{ij}(2+cy_{ij})}{12(1+cy_{ij})^2}\Big] = 0. \tag{5.2.9}$$

Note that c_{dql} is $\sqrt{n}$ consistent and efficient (Paul and Saha, 2007).

5.2.3 Estimation of parameters with missing responses

Under MCAR, missingness of the data does not depend upon the observed data and the cases with missing values are deleted before analysis. We can implement the standard methods of CC for the analysis. However, this may result in loss of efficiency due to the reduced sample size comprising of only the complete cases.

For MAR and MNAR, since some of the responses may be missing, we express the response y_{ij} as

$$y_{ij} = \begin{cases} y_{o,ij} & \text{if} \quad y_{ij} \quad \text{is observed,} \\ \\ y_{m,ij} & \text{if} \quad y_{ij} \quad \text{is missing.} \end{cases} \tag{5.2.10}$$

5.2.3.1 Maximum likelihood estimation under MAR

Let Y_o represent the observed responses, Y_m represent the missing responses, and $\theta = (m, c)$.

In MAR the conditional probability of missingness data depends on observed data. Parameters of missingness mechanism are completely separate and distinct from parameters of the model (5.2.3). In likelihood based estimation considering MAR, missingness mechanism can be ignored from the likelihood and the missing data are often known as ignorable missing. However subjects having these missing observations cannot be deleted before analysis. For more details see Little and Rubin (1987), and Ibrahim et. al (2005).

Our purpose is to maximize the following log likelihood with respect to the parameter ($\theta = m, c$)

$$l_0(m, c|Y_o) = \sum_{Y_m} l_0(m, c|Y_o, Y_m). \tag{5.2.11}$$

In more general case where the missing data are not MAR, the missing data mechanism needs to be defined and included in the model. Direct maximization of $l_0(m, c|Y_o)$ is not, in general, straight forward. However, the EM algorithm (Dempster et al., 1977) is a very useful tool for obtaining maximum likelihood estimates when some of the observations in the data are missing. The EM algorithm uses two steps, the expectation and the maximization steps. Following Little and Rubin (1987, 2002, 2014, 2020) the E-step provides the conditional expectation of the log-likelihood $l_0(m, c|Y_o, Y_m)$ given the observed data Y_o and current estimate of parameters ($\theta = m, c$). Suppose A of the n responses are observed and B=n-A responses are missing. Let 's' be an arbitrary number of iterations during maximization of the log-likelihood, then the E-step of the EM algorithm for the ij^{th} missing response for the $(s+1)^{th}$ iteration can be written as

$$\begin{aligned} Q_{ij}(\theta|\theta^{(s)}) = {} & E\big[l_{ij}(\theta^{(s)}|y_{o,ij}, y_{m,ij})|y_{o,ij}, \theta^{(s)}\big] \\ = {} & \textstyle\sum_{y_{m,ij}} l_{ij}(\theta^{(s)}|y_{o,ij}, y_{m,ij}) P(y_{m,ij}|y_{o,ij}, \theta^{(s)}). \end{aligned} \tag{5.2.12}$$

For all the observations, the E-step of the EM algorithm for the $(s+1)^{th}$ iteration is

$$\begin{aligned} Q_0(\theta|\theta^{(s)}) = \ & \textstyle\sum_{ij=1}^{A} l_{ij}(\theta^{(s)}|y_{o,ij}) \\ & +\textstyle\sum_{ij=1}^{B} \sum_{y_{m,ij}} l_{ij}(\theta^{(s)}|y_{o,ij}, y_{m,ij}) P(y_{m,ij}|y_{o,ij}, \theta^{(s)}). \end{aligned} \tag{5.2.13}$$

Note for the situation in which there is no missing response, the EM algorithm requires only maximization of the first term on the right hand side. Here $P(y_{m,ij}|y_{o,ij}, \theta^{(s)})$ is the conditional distribution of the missing response given the observed data and the current (s^{th}) iteration estimate of θ. However in many situations $P(y_{m,ij}|y_{o,ij}, \theta^{(s)})$ may not always be available. Following Ibrahim et. al (2001) and Sahu and Roberts (1999) we can write $P(y_{m,ij}|y_{o,ij}, \theta^{(s)}) \propto P(y_{ij}|\theta^{(s)})$ (the complete data distribution given in equation (5.2.3)). For the $(ij)^{th}$ of the B missing responses we take a sample $a_{ij,1}, a_{ij,2}, ..., a_{ij,m_{ij}}$, from $P(y_{ij}|\theta^{(s)})$ using Gibbs sampler (see Casella and George 1992 for details). Then following Ibrahim et al. (2001), $Q_0(\theta|\theta^{(s)})$ can be written as

$$Q_0(\theta|\theta^{(s)}) = \sum_{ij=1}^{A} l_{ij}(\theta^{(s)}|y_{o,ij}) + \sum_{ij=1}^{B} \frac{1}{m_{ij}} \sum_{k=1}^{m_{ij}} l_{ij}(\theta^{(s)}|a_{ij,k}). \tag{5.2.14}$$

In the M-step of EM algorithm, the $Q_0(\theta|\theta^{(s)})$ is maximized. We denote the resultant estimates by m_{MA} and c_{MA}. Here maximizing $Q_0(\theta|\theta^{(s)})$ is analogous to maximization of complete data log-likelihood where each incomplete observation is replaced by m_{ij} weighted observations.

5.2.3.2 DEQL estimation under MAR

Following section 5.2.3.1 and without going into further detail, the DEQL estimates of the parameters m and c, denoted by dm_{MA} and dc_{MA}, under MAR are obtained by maximizing

$$Q_{D0}(\theta|\theta^{(s)}) = \sum_{ij=1}^{A} dl_{ij}(\theta^{(s)}|y_{o,ij}) + \sum_{ij=1}^{B} \frac{1}{m_{ij}} \sum_{k=1}^{m_{ij}} dl_{ij}(\theta^{(s)}|a_{ij,k}). \tag{5.2.15}$$

5.2.3.3 Estimation under MNAR

Under MNAR, the probability of missing observation in response variable depends on the values of the response that would have been observed. The missing data mechanism cannot be ignored and needs to be incorporated in the likelihood. The missing data that are MNAR are known as non-ignorable missing. A parametric model needs to be specified for this missingness. To incorporate the missing data mechanism into the likelihood we define a random variable r_{ij} as,

$$r_{ij} = \begin{cases} 0 \quad \text{if} \quad y_{ij} \quad \text{is observed,} \\ \\ 1 \quad \text{if} \quad y_{ij} \quad \text{is missing.} \end{cases} \tag{5.2.16}$$

The random variable r_{ij} follows

$$p(r_{ij}|y_{ij}) = [p(r_{ij} = 1)]^{r_{ij}}[1 - p(r_{ij} = 1)]^{(1-r_{ij})}. \tag{5.2.17}$$

See Ibrahim et al. (2001). To model the probability of missing in terms of values of responses that would have been observed, a logit link

$$\log\Big[\frac{p(r_{ij} = 1)}{1 - p(r_{ij} = 1)}\Big] = \alpha_0 + \alpha_1 * y_{ij} \tag{5.2.18}$$

can be used. Here y_{ij} is the responses and the responses that would have been observed. Note that $p(r_{ij}=1)$ can be written as a logistic model

$$p(r_{ij}=1)=\frac{exp(\alpha_0+\alpha_1 * y_{ij})}{1+exp(\alpha_0+\alpha_1 * y_{ij})}. \tag{5.2.19}$$

Then the likelihood function of the parameter $\alpha=(\alpha_0,\alpha_1)$ can be written as

$$l(\alpha|r_{ij},y_{ij})=\sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[r_{ij} * \log\Big[\frac{p(r_{ij}=1)}{1-p(r_{ij}=1)}\Big]+\log(1-p(r_{ij}=1))\Big]. \tag{5.2.20}$$

Following Ibrahim, Lipsitz and Chen (1999), after incorporating the model for missingness mechanism in $l(\alpha|r_{ij},y_{ij})$, the log likelihood for all the parameters involved is

$$\begin{aligned} l_0(\theta,\alpha|Y_o,Y_m,R) = & \sum_{i=1}^{K}\sum_{j=1}^{n_i} l_{ij}+\sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[r_{ij} * \log\Big[\frac{p(r_{ij}=1)}{1-p(r_{ij}=1)}\Big] \\ & +\log(1-p(r_{ij}=1))\Big], \end{aligned} \tag{5.2.21}$$

where l_{ij} is same as in equation 5.2.5.

5.2.3.4 Maximum likelihood estimation under MNAR

As in section 5.2.3.1, for the $(ij)^{th}$ of the B missing responses we take a sample $a_{ij,1},a_{ij,2},...,a_{ij,m_{ij}}$, from $P(y_{ij}|\theta^{(s)})$ using Gibbs sampler (see Casella and George 1992 for details). Then following section 5.2.3.1, the maximum likelihood estimates under MNAR of the parameter m and c are obtained by maximizing

$$\begin{aligned} Q_0(\theta|\theta^{(s)}) = & \textstyle\sum_{ij=1}^{A} l_{ij}(\theta^{(s)}|y_{o,ij})+\sum_{ij=1}^{B}\frac{1}{m_{ij}}\sum_{k=1}^{m_{ij}} l_{ij}(\theta^{(s)}|a_{ij,k}) \\ & +\sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[r_{ij} * \log\Big[\frac{p(r_{ij}=1)}{1-p(r_{ij}=1)}\Big]+\log(1-p(r_{ij}=1))\Big], \end{aligned} \tag{5.2.22}$$

where l_{ij} is same as in equation 5.2.5. We denote the resultant estimates by m_{MN} and c_{MN}.

5.2.3.5 DEQL estimation under MNAR

Following section 5.2.3.1 and without going into further detail, the DEQL estimates of the parameters m and c, denoted by dm_{MN} and dc_{MN}, under MNAR are obtained by maximizing

$$\begin{aligned} Q_{D0}(\theta|\theta^{(s)}) = \ & \textstyle\sum_{ij=1}^{A} Dl_{ij}(\theta^{(s)}|y_{o,ij}) + \sum_{ij=1}^{B} \frac{1}{m_{ij}} \sum_{k=1}^{m_{ij}} Dl_{ij}(\theta^{(s)}|a_{ij,k}) \\ & + \sum_{i=1}^{K}\sum_{j=1}^{n_i} \Big[r_{ij} * \log\Big(\frac{p(r_{ij}=1)}{1-p(r_{ij}=1)}\Big) \\ & + \log(1-p(r_{ij}=1))\Big], \end{aligned} \quad (5.2.23)$$

where Dl_{ij} is same as in equation 5.2.8.

5.3 Test of hypothesis concerning the means in one-way anova with extra-dispersed count data

There is a long history of the development of $C(\alpha)$ (Neyman, 1959) test or score test (Rao, 1947) . For the situation under study here, Barnwal and Paul (1988) developed the $C(\alpha)$ test statistics for testing the equality of the means in one-way layout of count data. Without going into details of derivation which can be found in Barnwal and Paul (1988), the $C(\alpha)$ statistic, using the maximum likelihood estimates of the parameters m and c is

$$S_c(ml) = \sum_{i=1}^{K} \frac{n_i(\bar{y_i.} - \bar{y})^2}{\bar{y}(1+\bar{y}c_{ml})}.$$

Further, the $C(\alpha)$ ($C(\alpha)$-like) statistic developed by Saha (2008) based on the double extended quasi-likelihood is

$$S_c(deql) = \sum_{i=1}^{K} \frac{n_i(\bar{y_{i.}} - \bar{y})^2}{\bar{y}(1 + \bar{y}c_{deql})}.$$

The $C(\alpha)$ statistics using (m_{MC}, c_{MC}), (m_{MA}, c_{MA}), and (m_{MN}, c_{MN}) are

$$S_c(MC) = \sum_{i=1}^{K} \frac{n_i(\bar{y_{i.}} - m_{MC})^2}{m_{MC}(1 + m_{MC}c_{MC})},\ S_c(MA) = \sum_{i=1}^{K} \frac{n_i(\bar{y_{i.}} - m_{MA})^2}{m_{MA}(1 + m_{MA}c_{MA})},$$

and $S_c(MN) = \sum_{i=1}^{K} \frac{n_i(\bar{y_{i.}} - m_{MN})^2}{m_{MN}(1 + m_{MN}c_{MN})}$

respectively.

Similarly, the $C(\alpha)$ statistics using (m_{dMC}, c_{dMC}), (m_{dMA}, c_{dMA}), and (m_{dMN}, c_{dMN}) are

$$S_c(dMC) = \sum_{i=1}^{K} \frac{n_i(\bar{y_{i.}} - m_{dMC})^2}{m_{dMC}(1 + m_{dMC}c_{dMC})},$$

$$S_c(dMA) = \sum_{i=1}^{K} \frac{n_i(\bar{y_{i.}} - m_{dMA})^2}{m_{dMA}(1 + m_{dMN}c_{dMA})},$$

and $S_c(dMN) = \sum_{i=1}^{K} \frac{n_i(\bar{y_{i.}} - m_{dMN})^2}{m_{dMN}(1 + m_{dMN}c_{dMN})}$

respectively.

However, as can be seen from equations 5.2.22 and 5.2.23 each of the likelihoods contain 2 parts, first part involves only the parameters of the count data models and the second part contains the parameters of the missing data mechanism. Therefore, estimates of the parameters m and c and those of the missing data mechanism are independent. Therefore, the statistics $S_c(MA)$ and $S_c(MN)$ are identical and the statistics $S_c(dMA)$ and $S_c(dMN)$ are also identical.

5.4 Simulation Study

A simulation study is conducted to examine the comparative behaviour of the test statistics $S_c(ml)$ and $S_c(deql)$ in terms of size and power when some of the responses are missing. We use data under four scenarios: (i) data are observed completely (CC), (ii) some of the responses are missing completely at random

(MCAR), (iii) some responses are missing at random (MAR), and (iv) some responses are missing not at random (MNAR). Empirical significance levels and power of the tests were all based on 10,000 samples from the negative binomial distribution for different values of m and c. When data are observed completely, for all combinations of $m = 7, 15, 20, 30, 40$ and $c = 0.05, 0.25, 0.50$ simulations are run based on 10,000 samples of sizes $n_1 = n_2 = 10, 20$, and 50 for two groups. The values of parameters are same for $n_1 = n_2 = n_3 = 10, 20$, and 50 for three groups. Table 5.1 displays empirical levels based on a nominal significance level of $\alpha = 5\%$ for $S_c(ml)$ and $S_c(deql)$ for complete data for two groups, data generated from the NB distributions. Table 5.2 displays empirical power based on a nominal significance level of $\alpha = 5\%$ for $S_c(ml)$ and $S_c(deql)$ for complete data for two groups, data generated from the NB distributions. For two groups, c = 0.25, $m_1 = m$, $m_2 = m + \phi$, and $\delta = \phi/m$, where $m = 10, 20, 50$ and $\delta = 0.0, 0.2, 0.4, 0.6, 0.8$. The sample sizes considered were $n_1 = n_2 = 10, 20$, and 50. Table 5.3 displays empirical levels based on a nominal significance level of $\alpha = 5\%$ for $S_c(ml)$ and $S_c(deql)$ under MCAR and MAR for two groups. Table 5.4 displays empirical power based on a nominal significance level of $\alpha = 5\%$ for $S_c(ml)$ and $S_c(deql)$ under MCAR and MAR for two groups. Table 5.5 displays empirical levels based on a nominal significance level of $\alpha = 5\%$ for $S_c(ml)$ and $S_c(deql)$ for complete data for three groups. Table 5.6 displays empirical power based on a nominal significance level of $\alpha = 5\%$ for $S_c(ml)$ and $S_c(deql)$ for complete data for three groups, data generated from the NB distributions. For three groups c=0.25; $m_1 = m$, $m_2 = m + \phi_1$, $m_3 = m + \phi_2$, $\delta_1 = \phi_1/m$, $\delta_2 = \phi_2/m$, and $\boldsymbol{\delta} = (0, 0), (0, 0.2), (0.2, 0.4), (0.4, 0.6), (0.4, 0.8)$. Table 5.7 displays empirical levels based on a nominal significance level of $\alpha = 5\%$ for $S_c(ml)$ and $S_c(deql)$ under MCAR and MAR for three groups. For all combinations of $m = 20, 30$

and $c = 0.05, 0.25, 0.50$ simulations are run based on 10,000 samples of sizes $n_1 = n_2 = n_3 = 20$ and 30 . Table 5.8 displays empirical power based on a nominal significance level of $\alpha = 5\%$ for $S_c(ml)$ and $S_c(deql)$ under MCAR and MAR for three groups.

Table 5.1: $10^3\times$ empirical levels: α=0.05; based on 10,000 replications

		m=7			m=15			m=20			m=30			m=40		
		c			c			c			c			c		
$n_1 = n_2$	Statistic	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50
10	$S_c(ml)$	43.5	49.3	46.1	48.1	49.5	42 .2	51.5	49.3	50.3	48.7	49	49.8	54.8	49.1	51.6
	$S_c(deql)$	43.6	49.3	46.6	48.1	49.5	42.5	51.5	49.5	50.5	48.7	49.1	50.1	54.8	49.2	51.7
20	$S_c(ml)$	48.2	51.7	44.2	56.5	52.9	48.5	55	50.2	48.9	54.5	52.6	49	52.9	51.5	59.9
	$S_c(deql)$	48.2	51.8	44.8	50.6	53	48.5	55.1	50.2	49.3	54.5	52.6	49.4	52.8	51.6	48.7
50	$S_c(ml)$	54.1	63.9	56.7	50.4	51.2	47.9	53.4	52.3	51	53.7	50.6	50.1	52.6	53.1	53.7
	$S_c(deql)$	52.9	52.1	49.7	50.4	51.2	48	53.6	52.4	51.1	53.7	50.8	50.3	52.6	53.2	53.8

Table 5.2: $10^3\times$ empirical power corresponding to nominal significance level α=0.05; based on 10,000 replications for 2 groups; $m_1 = m,\ m_2 = m + \phi,\ \delta = \phi/m$; c=0.25

		$n_1 = n_2 = 10$					$n_1 = n_2 = 20$					$n_1 = n_2 = 50$				
		δ					δ					δ				
m	Statistic	0.0	0.2	0.4	0.6	0.8	0.0	0.2	0.4	0.6	0.8	0.0	0.2	0.4	0.6	0.8
10	$S_c(ml)$	48.9	98	226.9	403.5	580.2	49.6	158.8	425.7	703	886.6	52.7	341.3	823.8	982	998.9
	$S_c(deql)$	49.1	98.4	227.2	404	585.5	49.6	158.9	425.8	703.5	886.8	52.8	341.8	824	982	998.9
20	$S_c(ml)$	49.3	114.6	266.7	452.6	636.8	50.2	179.9	484	766.4	918.1	52.3	379.7	863.8	991.6	999.1
	$S_c(deql)$	49.5	114.7	266.9	452.9	637.2	50.2	180	484.3	766.7	918.3	52.4	380	863.8	991.6	997.9
50	$S_c(ml)$	50.2	114.1	276.6	482.4	671.4	50.5	195.8	517.4	797.2	937.3	53.3	420.7	893.3	994.3	999.9
	$S_c(deql)$	50.2	114.1	277	483.1	671.5	50.6	195.8	517.5	797.4	937.3	53.3	420.9	893.3	994.3	999.9

Table 5.3: $10^3\times$ empirical levels: α=0.05; based on 10,000 replications for data under MCAR and MAR

		m=20			m=30		
		c			c		
Statistics	%missing	0.05	0.25	0.50	0.05	0.25	0.50
$n_1 = n_2 = 20$ $S_c(MC)$	0	55	50.2	48.9	54.5	52.6	49
	5	51.9	53	52.9	53	51.8	52.7
	10	52.1	48.1	47.6	50.8	51.1	53.3
	25	53.8	51.3	49	54.7	52.9	50.2
$S_c(dMC)$	0	55.1	50.2	49.3	54.5	52.6	49.4
	5	51.9	53	53.1	53	51.8	52.8
	10	52.1	48.1	47.6	50.9	51.1	53.7
	25	53.9	51.3	49.3	54.7	52.9	50.4
$S_c(MA)$	0	55	50.2	48.9	54.5	52.6	49
	5	52	52.5	52	49.5	48	50.5
	10	59.5	51	50.5	42	48.5	47.5
	25	46	48.5	55.5	46.5	46.5	44.5
$S_c(dMA)$	0	55.1	50.2	49.3	54.5	52.6	49.4
	5	49.5	47	50	55	48	50.5
	10	40	41.5	48	58.5	60	57.5
	25	54	45	49	54	56	43.5
$n_1 = n_2 = 30$ $S_c(MC)$	0	47.5	53	40	47	53	46.5
	5	50.5	52.6	50.7	51.2	53.4	53.9
	10	56.1	55.5	48.7	54.8	52.3	52.7
	25	51.3	51.1	47.6	50.3	53.9	48
$S_c(dMC)$	0	47.5	53.5	41	47	53	46.5
	5	50.5	52.6	51.1	51.1	53.4	54.2
	10	56.1	55.5	49	54.7	52.3	52.7
	25	51.3	51.2	47.7	50.2	53.9	48
$S_c(MA)$	0	47.5	53	40	47	53	46.5
	5	55	49.5	52.5	54	53.5	59.5
	10	55	51	64	54.5	52	54.5
	25	49	51	46.5	50	45	50
$S_c(dMA)$	0	47.5	53.5	41	47	53	46.5
	5	51	56	59.5	49	50.5	47
	10	53.5	53	48	54.5	54	55.5
	25	47.5	58.5	53	43	51	55

Table 5.4: $10^3\times$ empirical power corresponding to nominal significance level α=0.05; based on 10,000 replications for 2 groups (under MCAR and MAR); $m_1 = m,\ m_2 = m + \phi,\ \delta = \phi/m$; c=0.25

		$n_1 = n_2 = 20$					$n_1 = n_2 = 30$				
		δ					δ				
Statistic	%missing	0.0	0.2	0.4	0.6	0.8	0.0	0.2	0.4	0.6	0.8
m=20 $S_c(MC)$	0	50.2	179.9	484	766.4	918.1	53	258.5	669	912.5	986
	5	53	162.2	446.6	730.2	886.5	54.7	224.8	591.7	872.2	972
	10	48.1	148.1	401.5	666.3	853.1	51.1	202.6	562.5	846.4	959.6
	25	51.3	115.4	281.9	510	694.7	52.3	148.7	400.7	665.6	850.5
$S_c(dMC)$	0	50.2	180	484.3	766.7	918.3	53.5	259	669	912.5	986
	5	53	162.4	447.2	730.2	886.7	52.6	227.4	592.2	876.1	972.1
	10	48.1	148.1	401.5	666.5	853.3	55.5	208.1	567.8	842.4	959.6
	25	51.3	115.4	282.3	510.1	695.1	51.2	150.9	397.8	661.2	850.6
$S_c(MA)$	0	50.2	179.9	484	766.4	918.1	53	258.5	669	912.5	986
	5	52.5	180.5	453	773.5	924.5	49.5	233.5	654.5	912.5	984.5
	10	51	162	486.5	756	924	51	244	665.2	899	985.5
	25	48.5	192	493	748.5	920.5	51	257	662.5	908	984
$S_c(dMA)$	0	50.2	180	484.3	766.7	918.3	53.5	259	669	912.5	986
	5	47	182.5	483	753.5	910.5	56	238.5	676.5	905.5	989.5
	10	41.5	172	483	759.8	921.3	53	253	651	905	985
	25	45	177	475.5	775	909.5	58.5	236	679.5	911	978.5
m=30 $S_c(MC)$	0	49.5	170	490.5	788	932	53	258.5	689	920	991
	5	51.8	173	455.6	735.1	904.9	53.4	229.6	612.9	889.6	977.2
	10	51.1	157.6	412.7	695.2	868.1	52.3	226.1	581.5	857.6	968.4
	25	52.9	123.1	298.8	521.7	718	53.9	161.2	424.6	686.1	866.7
$S_c(dMC)$	0	49.5	170	490.5	788	932	53	258.5	689	920	991
	5	51.8	166.5	455.9	739.7	904.9	53.4	229.7	613	889.6	977.2
	10	51.1	160.7	412.9	690.3	868.1	52.3	226.3	581.7	857.6	968.4
	25	52.9	120.6	298.9	510.8	718.3	53.9	161.2	424.9	686.3	866.9
$S_c(MA)$	0	49.5	170	490.5	788	932	53	258.5	689	920	991
	5	48	180	491	766.5	928	53.5	236.4	575.7	798.2	867.8
	10	48.5	186.5	499	786.5	928.5	52	268.5	664	914.5	991
	25	46.5	189.5	493	789.5	928	45	245.5	652.5	935	991
$S_c(dMA)$	0	49.5	170	490.5	788	932	53	258.5	689	920	991
	5	48	190	488.6	775.3	935.5	50.5	258.5	678.5	927.5	990
	10	60	185.5	498	790.5	935.5	54	259	651	905	985
	25	56	183	484	785.5	918	51	264.5	668.5	920.5	988.5

Table 5.5: $10^3\times$ empirical levels: α=0.05; based on 10,000 replications

		m=7			m=15			m=20			m=30			m=40		
		c			c			c			c			c		
$n_1 = n_2 = n_3$	Statistic	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50
10	$S_c(ml)$	45.4	47	44.6	53.2	44.2	46.6	48	46.7	47	52.8	48.3	43	48.3	43.9	43.9
	$S_c(deql)$	45.5	47.1	45.2	53.2	44.2	46.8	48	46.7	47.3	52.8	48.4	43.5	48.3	43.9	44
20	$S_c(ml)$	49.6	43.6	45.7	49.9	52.4	49.2	52.1	46.6	51.2	49.3	48.7	49.4	51.5	48.9	46.8
	$S_c(deql)$	49.6	43.7	46.5	49.9	52.4	49.4	52.1	46.7	51.4	49.3	48.7	49.7	51.5	48.9	47.1
50	$S_c(ml)$	51.5	52.6	51.5	51	52.3	51.6	51.2	47.6	49.7	49.9	45.9	50.9	46.5	50	49.8
	$S_c(deql)$	51.5	52.7	51.7	51	52.3	52.2	51.2	47.7	49.7	49.9	46.1	51.4	46.5	50	50

Table 5.6: $10^3 X\times$ empirical power corresponding to nominal significance level α=0.05; based on 10,000 replications for 3 groups; $m_1 = m,\ m_2 = m + \phi_1,\ m_3 = m + \phi_2,\ \delta_1 = \phi_1/m,\ \delta_2 = \phi_2/m$; c=0.25

		$n_1 = n_2 = n_3 = 10$ δ					$n_1 = n_2 = n_3 = 20$ δ					$n_1 = n_2 = n_3 = 50$ δ				
m	Statistic	(0,0)	(0,0.2)	(0.2,0.4)	(0.4, 0.6)	(0.4,0.8)	(0,0)	(0,0.2)	(0.2,0.4)	(0.4, 0.6)	(0.4,0.8)	(0,0)	(0,0.2)	(0.2,0.4)	(0.4, 0.6)	(0.4,0.8)
10	$S_c(ml)$	43.1	95.6	163.9	296.3	449.7	49.5	155.3	328.7	613.9	809.6	51.5	352.1	724.6	963.7	996.1
	$S_c(deql)$	43.1	95.7	164.1	297	458.1	49.6	155.3	329	614	813.5	51.6	352.3	725.1	963.7	996.2
20	$S_c(ml)$	46.7	96.4	188.8	342.4	508.9	46.6	168.9	365.6	669	859.4	47.6	393	784.5	981.4	998.6
	$S_c(deql)$	46.7	96.4	188.9	342.5	511.6	46.7	169	365.7	669.1	860.9	47.7	393.1	784.7	981.4	998.6
50	$S_c(ml)$	46.1	106.4	201	357.9	542.6	48.5	192	412.6	712.2	881	48.9	433.7	827	989.9	999.5
	$S_c(deql)$	46.2	106.5	201.1	358.2	543.6	48.7	192.1	412.8	712.3	881.2	49	433.7	827.1	989.9	999.5

Table 5.7: $10^3\times$ empirical levels: α=0.05;based on 10,000 replications for data under MCAR and MAR

		m=20			m=30		
		c			c		
Statistics	%missing	0.05	0.25	0.50	0.05	0.25	0.50
$n_1 = n_2 = n_3 = 20$	0	52.1	46.6	51.2	49.3	48.7	49.4
	5	44.7	49.2	47.4	49.8	52.3	47.6
	10	49.7	47.9	45.2	51.8	46.7	48.2
$S_c(MC)$	25	46.6	45.9	39.5	49.6	49.7	43.7
	0	52.1	46.7	51.4	49.3	48.7	49.7
	5	44.7	49.2	48	49.7	52.3	50.1
$S_c(dMC)$	10	49.7	47.9	45.8	51.8	46.7	46.3
	25	46.6	46	40	49.6	49.7	43.5
	0	52.1	46.6	51.2	49.3	48.7	49.4
	5	48.5	48.5	45	52	47.5	46.5
$S_c(MA)$	10	51	49	51	51.5	56.5	47.5
	25	51.5	51	55.5	45	49	50.5
	0	52.1	46.7	51.4	49.3	48.7	49.7
	5	45	43	45.5	55	52	50.5
$S_c(dMA)$	10	43.5	64.5	51.5	56	49	55.5
	25	51	60	45	40.5	46.5	54
$n_1 = n_2 = n_3 = 30$	0	46.5	55	47	48.5	50.5	46.5
	5	46.6	53.4	48.6	50.5	54	49.1
	10	49.4	49.9	47	48.5	50.4	50.8
$S_c(MC)$	25	46.2	48.5	47.7	51.5	46.7	48.8
	0	46.5	55.5	47	48.5	50.5	47
	5	48	53.4	48.9	50.5	54.1	49.1
	10	49.3	50	49.5	48.5	50.5	51.2
$S_c(dMC)$	25	46.4	48.5	48.7	51.5	46.7	49.3
	0	46.5	55	47	48.5	50.5	46.5
	5	38.5	54	47	46.5	48.5	44.5
$S_c(MA)$	10	56	58	53	55	47.5	44.5
	25	57	47	42.5	41.5	45	53.5
	0	46.5	55.5	47	48.5	50.5	47
	5	55	51	42	48.5	54	52
$S_c(dMA)$	10	49	49.5	51	54.5	50.5	60
	25	47.5	51	58	52.5	56.5	51

Table 5.8: $10^3\times$ empirical power corresponding to nominal significance level α=0.05; based on 10,000 replications for 3 groups (under MCAR and MAR) $m_1 = m,\ m_2 = m + \phi_1,\ m_3 = m + \phi_2,\ \delta_1 = \phi_1/m,\ \delta_2 = \phi_2/m$; c=0.25

		$n_1 = n_2 = n_3 = 20$					$n_1 = n_2 = n_3 = 30$				
		δ					δ				
Statistic	%missing	(0,0)	(0,0.2)	(0.2, 0.4)	(0.4,0.6)	(0.4,0.8)	(0,0)	(0,0.2)	(0.2, 0.4)	(0.4,0.6)	(0.4,0.8)
m=20	0	46.6	168.9	365.6	669	859.4	55	249.5	543	856	968.5
	5	49.2	156.9	323.9	588.3	790.4	53.4	201.6	452.9	760.7	921.7
	10	47.9	135.6	281.2	497.9	707.1	49.9	196.6	406.2	710.1	884.8
$S_c(MC)$	25	45.9	86.9	156.2	272.4	411.8	48.5	113.7	214.7	408	584.6
	0	46.7	169	365.7	669.1	860.9	55.5	249.5	543	856	968.5
	5	49.2	156.9	324.1	588.5	790.4	53.4	201.7	453.1	760.8	921.8
$S_c(dMC)$	10	47.9	135.7	281.5	498.3	707.4	50	196.8	406.2	710.4	884.8
	25	46	87.1	156.4	273.1	412	48.5	113.8	214.8	408.5	584.7
	0	46.6	168.9	365.6	669	859.4	55	249.5	543	856	968.5
	5	48.5	177.5	383.5	646.5	852	54	239.5	529	859.5	966.5
$S_c(MA)$	10	49	187	388	660.5	858	58	232	561.5	856	968
	25	51	168	372.5	668	838	47	237.5	547.5	863.5	962.5
	0	46.7	169	365.7	669.1	860.9	55.5	249.5	543	856	968.5
	5	43	168.5	365.5	648.5	856	51	229	547.5	855.5	961.5
$S_c(dMA)$	10	64.5	175.5	370	661.5	864.5	49.5	260	534.5	848.5	969.5
	25	60	177	365.5	661	861.5	51	259.5	546	860	961
m=30	0	48.5	193.5	395.5	689	875	50.5	248.5	555	872	964.5
	5	52.3	163.4	340.4	610.4	802.7	54	221.4	473.7	791.9	932.2
	10	46.7	137.2	286.7	519.5	717	50.4	205.4	427.2	733.8	903.9
$S_c(MC)$	25	49.7	91.4	163.5	283.6	434.3	46.7	120.6	235.1	425.2	612.1
	0	48.5	194	396	689	875	50.5	248.5	555.5	872	964.5
	5	52.3	163.6	340.5	611	803	54.1	221.4	473.7	792.2	932.2
$S_c(dMC)$	10	46.7	137.4	287.1	519.6	717.2	50.5	205.4	427.3	733.8	903.9
	25	49.7	91.6	163.5	283.9	434.6	46.7	120.7	235.2	425.4	612.3
	0	48.5	193.5	395.5	689	875	50.5	248.5	555	872	964.5
	5	47.5	189.5	404	695.5	865.5	48.5	252.5	568.5	865.5	978.5
$S_c(MA)$	10	56.5	177	398	670	867.5	47.5	262.5	566	885	975
	25	49	182.5	393	691.5	865.5	45	245.5	550.5	856.5	973
	0	48.5	194	396	689	875	50.5	248.5	555.5	872	964.5
	5	52	174	377	670.5	864.5	54	253	565	871.5	969
$S_c(dMA)$	10	49	173.5	381	696.5	864	50.5	259	579	870	965.5
	25	46.5	186	399.5	684	857.5	56.5	266	559	867	966

5.5 Conclusions from the simulation study

Results in table 5.1 show that both the score tests based on the maximum likelihood estimates ($S_c(ml)$) and those based on the double extended quasi likelihood ($S_c(deql)$) hold level reasonably well irrespective of the value of the mean parameter m (common m) and sample size ($n_1 = n_2 = 10, 20, 50$).

Table 5.2 show the empirical power of these two statistics for increasing $m(m = 10, 20, 50)$ and increasing sample sizes ($n_1 = n_2 = 10, 20, 50$). A general conclusion is that empirical power increases as sample size increases. Power increases as the value of common m increases and also power increases as the difference between two $m's$ increases.

Table 5.3 shows the empirical level of the score tests ($S_c(MC), S_c(dMC)$), ($S_c(MA)$, $S_c(dMA)$) for $n_1 = n_2 = 20$, $n_1 = n_2 = 30$ and for common $m = 20$ and 30 for percentage missing $0\%, 5\%, 10\%,$ and 25%. The result in the table shows that there is virtually no qualitative difference in the empirical levels irrespective of the percentage missing, irrespective of the common m chosen, and irrespective of the common n chosen.

Table 5.4 provides empirical power for two groups where common $m = 20$ and 30 and common $n_1 = n_2 = 20$ and $n_1 = n_2 = 30$. Again the general conclusion is that as m increases (δ increases) power increases. This behaviour is consistent irrespective of the common m (for example $m = 20$ and $m = 30$) or common n ($n_1 = n_2 = 20$ and $n_1 = n_2 = 30$).

Same conclusion holds for table 5.5 where the number of population increases from 2 to 3, where $n_1 = n_2 = n_3 = 10, 20, 50$. Both the statistics ($Sc(ml)$) and ($Sc(deql)$) hold the level reasonably well.

Table 5.6 shows the empirical power of these two statistics when population increases from 2 to 3. As in table 5.3, a general conclusion here is that the empirical power increases with the increase in sample size and also the power increases when the difference in $m's$ increases.

Table 5.7 shows the empirical level of the score tests $(S_c(MC), S_c(dMC))$, $(S_c(MA)$, $S_c(dMA))$ for $n_1 = n_2 = n_3 = 20$, $n_1 = n_2 = n_3 = 30$ and for common $m = 20$ and 30 for percentage missing $0\%, 5\%, 10\%$, and 25%. Similar conclusions hold as in table 5.3.

Table 5.8 provides empirical power for three groups with same parameters and sample sizes as in table 5.4 and again the general conclusion is that power increases when the difference in $m's$ increases (δ increases). This is consistent irrespective of common m ($m = 20$ and $m = 30$) or common n ($n_1 = n_2 = n_3 = 20$ and $n_1 = n_2 = n_3 = 30$).

Under MCAR, the power for both the statistics ($Sc(MC)$ and $Sc(dMC)$) decreases as the percentage missing increases. In general, $Sc(MA)$ has better power than $Sc(MC)$ and $Sc(dMA)$ has better power than that of $Sc(dMC)$.

5.5.1 Illustrative Examples

In this section we present the analysis of data sets collected and published by researchers.

Example 1 (*Toxicological data*): Table 5.9 presents data on embryonic deaths in mice in a control group and two treatment groups. These data have been analyzed by Barnwal and Paul (1988) which shows that the assumption of a common dispersion parameter among the control group and the two treatment groups is reasonable. Based on the assumption of common c, they showed that the means for the treatment groups are not different from that of the control group. The test statistic was based on the maximum likelihood estimates of m and c. Saha (2008), showed that the estimates of m and c based on the double extended quasi likelihood are 0.7667 and 0.5354, respectively. The $C(\alpha)$ statistic based on these estimates is 3.021 (p-value =0.2208) which showed that the means of treatment groups do not differ from that of the control group as in Barnwal and Paul (1988). Table 5.10 shows the values of parameter estimates and the test statistics for complete case and under various missing scenarios. The results show that both statistics $(Sc(ml), Sc(deql))$ based on the maximum likelihood estimates and double extended quasi-likelihood estimates respectively have the p-value > 0.05 for all missing percentage (0 %, 10%, and 25 %) and under all missing data mechanism (MCAR and MAR). This indicates that the means do not differ between the treatment groups and control group. The conclusion is in agreement with those achieved by Barnwal and Paul (1988) and Saha (2008).

Table 5.9: Counts of embryonic deaths in a control group and two treatment groups from Barnwal and Paul(1998)

Number of deaths	Observed frequencies		
	Control group	Dose level 1	Dose level 2
0	7	5	4
1	2	4	2
2	1	0	3
3	0	1	0
4	0	0	1

Table 5.10: Estimates of parameters m and c and score statistics using maximum likelihood and double extended quasi-likelihood for data in table 5.9

Missing %		Estimates of m and c and score statistics							
		$\hat{m}_{ml}$	$\hat{c}_{ml}$	$S_c(ml)$	p-value	$\hat{m}_{deql}$	$\hat{c}_{deql}$	$S_c(deql)$	p-value
Complete Case	0%	0.7667	0.5439	3.0069	0.2224	0.7667	0.5354	3.0209	0.2208
MCAR	10%	0.7778	0.5038	1.8471	0.3970	0.8148	0.5275	3.3062	0.1915
	25%	0.8333	0.7499	3.0154	0.2214	0.7917	0.6978	2.5088	0.2852
MAR	10%	0.7837	0.6222	2.8096	0.2454	0.8090	0.6569	2.6799	0.2618
	25%	0.7253	0.9484	2.710	0.2579	0.7874	0.8287	2.5205	0.2836

Example 2 (*Biological data*): Table 5.11 presents the data on the total number of borers per hill in each plot for a control group and three treatment groups, originally given and analyzed by Bliss and Fisher (1953). In a field experiment of insect pests on the corn borer, four treatments were arranged in 15 randomized blocks. In each plot, eight hills of corn were selected randomly and the borers per hill were recorded at the end of season. Saha (2008) showed that the value of $C(\alpha)$ statistic for testing the assumption of a common dispersion among the groups is 7.5303 (p-value = 0.0568) and the assumption of common c is reasonable. For the data in table 5.11, the values of parameter estimates and the test statistics for complete case and under various missing scenario are given in table 5.12. The results show that both statistics $(Sc(ml), Sc(deql))$ based on the maximum likelihood estimates and double extended quasi-likelihood estimates respectively have the p-value < 0.05 for all missing percentage (0 %, 5%,10%, and 25 %) and under

all missing data mechanisms (MCAR and MAR). This indicates that the means differ among the groups. Saha (2008) showed that the C(deql) statistic for testing equality of means across groups is 68.2764 (p-value $=9.9 \times 10^{-15}$) indicating the difference in means among the groups. Our results for the complete case are in agreement with those achieved by Saha.

Table 5.11: Distribution of corn borers in a field experiment arranged in 15 randomized blocks from Bliss and Fisher (1953)

Borers per hill	Observations for			
	Control (C)	Treatment 1 (T_1)	Treatment 2 (T_2)	Treatment 3 (T_3)
0	19	24	43	47
1	12	16	35	23
2	18	16	17	27
3	18	18	11	9
4	11	15	5	7
5	12	9	4	3
6	7	6	1	1
7	8	5	2	1
8	4	3	2	
9	4	4		
10	1	3		1
11				1
12	1	1		
13	1			
15	1			
17	1			
19	1			
26	1			

Table 5.12: Estimates of parameters m and c and score statistics using maximum likelihood and double extended quasi-likelihood for data in table 5.11

Missing %		Estimates of m and c and score statistics							
		$\hat{m}_{ml}$	$\hat{c}_{ml}$	$S_c(ml)$	p-value	$\hat{m}_{deql}$	$\hat{c}_{deql}$	$S_c(deql)$	p-value
CC	0%	2.5479	0.9239	67.4487	1.5×10^{-15}	2.5479	0.9080	68.2764	9.9×10^{-15}
MCAR	5%	2.5482	0.9708	64.0609	8.0×10^{-14}	2.5614	0.9274	68.7422	7.8×10^{-15}
	10%	2.5602	0.9579	60.1439	5.5×10^{-13}	2.5671	0.8235	55.9171	4.4×10^{-12}
	25%	2.4722	0.9892	52.5694	2.3×10^{-11}	2.6028	0.9410	56.1262	3.9×10^{-12}
MAR	5%	2.5467	0.9761	64.9284	5.2×10^{-14}	2.5451	0.9603	65.7590	3.5×10^{-14}
	10%	2.5593	0.9715	64.6061	6.1×10^{-14}	2.5571	0.9504	65.7172	3.5×10^{-14}
	25%	2.4931	1.0165	65.5773	3.8×10^{-14}	2.4852	0.9884	67.3212	1.5×10^{-14}

Chapter 6

Summary and Plan for Future Research

6.1 Summary

One-way layout of count data often arises in practice. Poison models are widely used in the regression analysis of count data. Poisson model has strong assumption that the mean and variance are equal however in practice count data often exhibit extra-Poisson variation. Among several distributions available in literature, one very convenient and common model to accommodate this extra dispersion is the two parameter negative binomial distribution. For the over/under dispersed count data in one way layout, one may be interested in testing the equality of means of two or more groups.

In chapter 3 we study the two $C(\alpha)$ statistics, $S_c(ml)$ and $S_c(mm)$ recommended by Barnwal and Paul (1988). We studied, through the simulation studies, the performance of the test statistics based on small, moderate and large sample sizes. Performance of the test procedure were compared in terms of size and power.

Both the statistics maintain the significance level well, however the power of $S_c(ml)$ is always higher than those for the $S_c(mm)$. Thus, $S_c(ml)$ is recommended.

In chapter 4, we studied the performance of the $C(\alpha)$ statistics based on the semi-parametric models quasi-likelihood, extended quasi-likelihood, and double extended quasi-likelihood, namely, $S_c(ql)$, $S_c(eql)$ and $S_c(deql)$ in terms of size and power. For small to moderate sample sizes, $S_c(ql)$ has lower levels compared to the other two statistics. However, for large sample size it maintains the level well and close to other two statistics. The power performance of the statistics $S_c(deql)$ is higher than the other two statistics in almost all data situations. Thus based on our simulation study $S_c(deql)$ is recommended.

In chapter 5, through the simulation studies, we studied the effect of missingness on the $C(\alpha)$ statistic based on maximum likelihood and double extended quasi-likelihood.

6.2 Future Research

6.2.1 Effect of Missing Data on the Score Test of Interaction in Two-Way Layout of Count Data Involving Multiple Counts in Each Cell

Standard contingency tables involving fixed factors can be analyzed using log-linear models (e.g., Agresti 1990; Bishop, Fienberg, and Holland 1975; Plackett 1981) or score tests under a Poisson assumption. When multiple counts in each cell occur, particularly when the data are unbalanced such standard analyses will provide misleading conclusions. For the balanced two-way layout of Poisson-distributed data involving two fixed factors, Thall (1992) developed score tests for interaction and main effects that have simple forms. Thall (1992) also developed score tests for the main effects for the balanced two-way layout when one factor is fixed and the other is random. In future research I plan to study how the performance of the score tests developed by Thall (1992) are affected when some of the data in some cells are missing. Here I give a review of the work done by Thall (1992) and provide a plan of research to be done on this topic.

6.2.2 Score Test for Interaction

Let Y_{ijk} denote the k^{th} response in the $(ij)^{th}$ cell, $i = 1, 2, \ldots, a,\ j = 1, 2, \ldots, b,\ k = 1, 2, \ldots, n$. The interaction model for the mean is given by

$$\mu_{ij} = \alpha_i \beta_j (\tau + \phi_{ij}) \tag{6.2.1}$$

with $\alpha_a = \beta_b = 1$ and $\phi_{ib} = \phi_{aj} = 0$ for all i and j. The hypothesis of no interaction is $H_0 : \phi = 0$. The vector of interaction parameter is $\boldsymbol{\phi} = (\phi_{1,1}, \ldots, \phi_{1,b-1}, \ldots, \phi_{a-1,1}, \ldots, \phi_{a-1,b-1})'$ with dimension $(a-1)(b-1)$ and the vector of nuisance parameter is $\theta = (\alpha', \beta', \tau)' = (\alpha_1, \ldots, \alpha_{a-1}, \beta_1, \ldots, \beta_{b-1}, \tau)'$ with dimension $(a+b-1)$. The log-likelihood for testing the interaction when $Y_{ijk} \sim Poisson(\mu_{ij})$, apart from a constant independent of the parameters,takes the form,

$$\begin{aligned} l_1 = \ & \sum_{i=1}^{a-1} Y_{i..} log(\alpha_i) + \sum_{j=1}^{b-1} Y_{.j.} log(\beta_j) + \sum_{i=1}^{a-1} \sum_{j=1}^{b-1} Y_{ij.} log(\tau + \phi_{ij}) \\ & + (Y_{a..} + Y_{.b.} - Y_{ab.}) log(\tau) - n\{\tau + \alpha_. \tau + \beta_. \tau + \textstyle\sum_{i=1}^{a-1} \sum_{j=1}^{b-1} \alpha_i \beta_j (\tau + \phi_{ij})\} \end{aligned} \tag{6.2.2}$$

where, $Y_{i..} = \displaystyle\sum_{j=1}^{b} \sum_{k=1}^{n} Y_{ijk}$, $\quad Y_{.j.} = \displaystyle\sum_{i=1}^{a} \sum_{k=1}^{n} Y_{ijk}$, $\quad Y_{ij.} = \displaystyle\sum_{k=1}^{n} Y_{ijk}$ $\quad \alpha_. = \displaystyle\sum_{i=1}^{a-1} \alpha_i$ and $\beta_. = \sum_{j=1}^{b-1} \beta_j$. Under null hypothesis $\mu_{ij} = \alpha_i \beta_j \tau$. The log-likelihood under the hypothesis of no interaction is,

$$l_0 = \sum_{i=1}^{a-1} Y_{i..} log(\alpha_i) + \sum_{j=1}^{b-1} Y_{.j.} log(\beta_j) + Y_{...} log(\tau) - n\tau\{(\alpha_. + 1)(\beta_. + 1)\}. \quad (6.2.3)$$

In order to compute $T_{C(\alpha)}$, the quantities $\psi = \frac{\partial l_1}{\partial \boldsymbol{\phi}}|_{\phi=0}$, $\boldsymbol{F} = E(-\frac{\partial^2 l_1}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'}|_{\phi=0})$, $\boldsymbol{C} = E(-\frac{\partial^2 l_1}{\partial \boldsymbol{\phi} \partial \boldsymbol{\theta}'}|_{\phi=0})$ and $\boldsymbol{D} = E(-\frac{\partial^2 l_1}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}|_{\phi=0})$ are required. Using the maximum likelihood estimates $\hat{\theta}$ of θ under the null hypothesis, in $\psi, \boldsymbol{F}, \boldsymbol{C}$, and $\boldsymbol{D}$, then the score test for interaction is given by

$$T_{C(\alpha)} = \hat{\psi}'(\hat{\boldsymbol{F}} - \hat{\boldsymbol{C}}\hat{\boldsymbol{D}}^{-1}\hat{\boldsymbol{C}}')^{-1}\hat{\psi}.$$

The maximum likelihood estimates of $\boldsymbol{\theta}$ under H_0 are $\hat{\alpha}_i = Y_{i..}/Y_{a..}$, $\hat{\beta}_j = Y_{.j.}/Y_{.b.}$ and $\hat{\tau} = Y_{a..}Y_{.b.}/nY_{...}$ so that $\hat{\mu}_{ij} = Y_{i..}Y_{.j.}/nY_{...} = \bar{Y}_{i..}\bar{Y}_{.j.}/\bar{Y}_{...}$. For Poisson observations in the $a \times b$ layout with n observations per cell, the $C(\alpha)$ test statistic for the hypothesis of no multiplicative interaction computed using the MLE of $\boldsymbol{\theta}$ under the null hypothesis as given by Thall (1992) is

$$T_{C(\alpha)}(A \times B) = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{(Y_{ij.}Y_{...} - Y_{i..}Y_{.j.})^2}{Y_{i..}Y_{.j.}Y_{...}} \quad (6.2.4)$$

which, asymptotically, as $n \to \infty$, has a chi-squared distribution with $(a-1)(b-1)$ df. The test statistic can be expressed in simple equivalent form as

$$T_{C(\alpha)}(A \times B) = n\sum_{i=1}^{a}\sum_{j=1}^{b} \frac{(\bar{Y}_{ij.} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \quad (6.2.5)$$

6.2.3 Estimation of parameters with missing responses

We asssume $Y_{ijk} \sim$ Poisson (μ_{ij}) which has probability mass function

$$P(Y_{ijk} = y_{ijk}|\mu_{ij}) = \frac{e^{-\mu_{ij}}{\mu_{ij}}^{y_{ijk}}}{(y_{ijk})!}, \tag{6.2.6}$$

where μ_{ij} is the mean parameter. The mean and variance of the Poisson distribution are equal and is μ_{ij}. The interaction model for the mean is given by 6.2.1. Under null hypothesis $\mu_{ij} = \alpha_i\beta_j\tau$. Under MCAR, the cases with missing values are deleted before analysis. For MAR, the response y_{ijk} can be expressed as

$$y_{ijk} = \begin{cases} y_{o,ijk} & \text{if} \quad y_{ijk} \quad \text{is observed,} \\ \\ y_{m,ijk} & \text{if} \quad y_{ijk} \quad \text{is missing.} \end{cases} \tag{6.2.7}$$

6.2.3.1 Maximum likelihood estimation under MAR

Since some of the responses are considered missing, let Y_o represent the observed values and Y_m represent the missing values. We are interested in studying the effect of missing values on the score test of interaction, thus our purpose is to maximize the following log-likelihood under the null hypothesis of no interaction, with respect to the parameter, $\theta = (\alpha', \beta', \tau)'$

$$l_0(\theta|Y_o) = \sum_{Y_m} l_0(\theta|Y_o, Y_m). \tag{6.2.8}$$

The E-step of EM algorithm gives the conditional expectation of the log-likelihood given the observed data Y_o and current parameter estimates. The E-step for $(ijk)^{th}$

missing response for the $(s+1)^{th}$ iteration is given as,

$$\begin{aligned} Q_{ijk}(\theta|\theta^{(s)}) =& \; E\big[l_{ijk}(\theta^{(s)}|y_{o,ijk}, y_{m,ijk})|y_{o,ijk}, \theta^{(s)}\big] \\ =& \; \textstyle\sum_{y_{m,ijk}} l_{ijk}(\theta^{(s)}|y_{o,ijk}, y_{m,ijk})P(y_{m,ijk}|y_{o,ijk}, \theta^{(s)}). \end{aligned} \tag{6.2.9}$$

For all the observations, the E-step of the EM algorithm for the $(s+1)^{th}$ iteration is

$$\begin{aligned} Q(\theta|\theta^{(s)}) =& \sum_{ijk=1}^{O} l_{ijk}(\theta^{(s)}|y_{o,ijk}) \\ &+ \sum_{ijk=1}^{M} \sum_{y_{m,ijk}} l_{ijk}(\theta^{(s)}|y_{o,ijk}, y_{m,ijk})P(y_{m,ijk}|y_{o,ijk}, \theta^{(s)}), \end{aligned} \tag{6.2.10}$$

where O and M represent the number of observed and missing responses respectively. Here $P(y_{m,ijk}|y_{o,ijk}, \theta^{(s)})$ is the conditional distribution of the missing response given the observed data and the current (s^{th}) iteration estimate of θ. Following Ibrahim et. al (2001) we have $P(y_{m,ijk}|y_{o,ijk}, \theta^{(s)}) \propto P(y_{ijk}|\theta^{(s)})$. For each $(ijk)^{th}$ missing responses we generate samples $a_{ijk,1}, a_{ijk,2}, ..., a_{ijk,m_{ijk}}$, from $P(y_{ijk}|\theta^{(s)})$ using Gibbs sampler. Then following Ibrahim et. al (2001), $Q(\theta|\theta^{(s)})$ can be written as

$$Q_(\theta|\theta^{(s)}) = \sum_{ijk=1}^{O} l_{ij}(\theta^{(s)}|y_{o,ijk}) + \sum_{ijk=1}^{M} \frac{1}{m_{ijk}} \sum_{r=1}^{m_{ijk}} l_{ijk}(\theta^{(s)}|a_{ijk,r}). \tag{6.2.11}$$

In the M-step of EM algorithm, the $Q(\theta|\theta^{(s)})$ is maximized which is analogous to maximization of complete data log-likelihood where each incomplete observation is replaced by m_{ijk} weighted observations.

In my future study I will investigate the properties of estimators $\theta = (\alpha', \beta', \tau)'$ by Monte Carlo simulation studies and extend this theory to the two-way layout of count data involving multiple counts in each cell (Paul and Banerjee, 1998).

Appendices

A Derivation of C(ml)

The log-likelihood in terms of reparameterization of $m_i = m + \delta_i$ and $c_i = c$, apart form some constant terms, can be written as

$$\begin{aligned} l = & \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[y_{ij}\ln(m+\delta_i) - \Big(y_{ij}+\tfrac{1}{c}\Big)\ln(1+cm+c\delta_i) \\ & +\sum_{l=1}^{y_{ij}}\log\{1+c(l-1)\}\Big]. \end{aligned} \tag{A.1}$$

Now, define $\delta = (\delta_1, ..., \delta_{K-1})$ and $\nu = (\nu_1, \nu_2)^{'} = (m,\ c)^{'}$. Then,

$$\begin{aligned} \phi_i = & \tfrac{\partial l}{\partial \delta_i}\Big|_{\delta=0} = \sum_{j=1}^{n_i}\Big[\frac{y_{ij}}{(m+\delta_i)} - \frac{(y_{ij}+c^{-1})c}{(1+cm+c\delta_i)}\Big]\Big|_{\delta=0}. \\ & = \sum_{j=1}^{n_i}\frac{y_{ij}-m}{m(1+cm)} \\ & = \tfrac{n_i(\bar{y_i}.-m)}{m(1+cm)}, \end{aligned} \tag{A.2}$$

$$\begin{aligned}
\eta_1 = \ & \tfrac{\partial l}{\partial \nu_1}\Big|_{\delta=0} = \tfrac{\partial l}{\partial m}\Big|_{\delta=0} = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{y_{ij}}{(m+\delta_i)} - \frac{(y_{ij}+c^{-1})c}{(1+cm+c\delta_i)}\Big]\Big|_{\delta=0}. \\
& = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\frac{y_{ij}-m}{m(1+cm)} \\
& = \sum_{i=1}^{K}\frac{n_i(\bar{y_{i.}}-m)}{m(1+cm)},
\end{aligned} \tag{A.3}$$

$$\begin{aligned}
\eta_2 = \ & \tfrac{\partial l}{\partial \nu_2}\Big|_{\delta=0} = \tfrac{\partial l}{\partial c}\Big|_{\delta=0} \\
& = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{-(y_{ij}+c^{-1})m}{(1+cm+c\delta_i)} + ln(1+cm+c\delta_i)c^{-2} + \sum_{l=1}^{y_{ij}}\frac{l-1}{(1+c(l-1))}\Big]\Big|_{\delta=0}. \\
& = -\sum_{i=1}^{K}\sum_{j=1}^{n_i}\frac{my_{ij}}{1+cm} - \sum_{i=1}^{K}\sum_{j=1}^{n_i}\frac{m}{c(1+cm)} + \sum_{i=1}^{K}\sum_{j=1}^{n_i}\frac{ln(1+cm)}{c^2} \\
& + \sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{l=1}^{y_{ij}}\frac{1/c\{1+c(l-1)\}-1/c}{1+c(l-1)} \\
& = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{ln(1+cm)}{c^2} - \sum_{r=0}^{y_{ij}-1}\frac{1}{c(1+cr)}\Big]
\end{aligned} \tag{A.4}$$

$$\Delta_{st} = E\Big(-\tfrac{\partial^2 l}{\partial \delta_s \partial \delta_t}\Big|_{\delta=0}\Big) \tag{A.5}$$

$$\tfrac{\partial l}{\partial \delta_s} = \ \sum_{j=1}^{n_s}\Big[\frac{y_{ij}}{(m+\delta_s)} - \frac{(y_{ij}+c^{-1})c}{(1+cm+c\delta_s)}\Big]$$

$$\tfrac{\partial^2 l}{\partial {\delta_s}^2} = \ \sum_{j=1}^{n_s}\Big[-\frac{y_{ij}}{(m+\delta_s)^2} + \frac{(1+cy_{ij})c}{(1+cm+c\delta_s)^2}\Big]$$

$$\frac{\partial^2 l}{\partial {\delta_s}^2}\Big|_{\delta=0} = \sum_{j=1}^{n_s}\Big[-\frac{y_{ij}}{m^2}+\frac{(1+cy_{ij})c}{(1+cm)^2}\Big]$$

$$E\Big(-\frac{\partial^2 l}{\partial {\delta_s}^2}\Big|_{\delta=0}\Big) = n_s/m(1+cm) \tag{A.6}$$

And,

$$E\Big(-\frac{\partial^2 l}{\partial \delta_s \delta_t}\Big|_{\delta=0}\Big) = 0 \tag{A.7}$$

$$\Gamma_{st} = E\Big(-\frac{\partial^2 l}{\partial \delta_s \partial \nu_t}\Big|_{\delta=0}\Big) \tag{A.8}$$

$$\frac{\partial^2 l}{\partial \delta_s \partial \nu_1} = \sum_{j=1}^{n_s}\Big[-\frac{y_{ij}}{(m+\delta_s)^2}+\frac{(1+cy_{ij})c}{(1+cm+c\delta_s)^2}\Big]$$

$$E\Big(-\frac{\partial^2 l}{\partial \delta_s \partial \nu_1}\Big|_{\delta=0}\Big) = n_s/m(1+cm) \tag{A.9}$$

$$\frac{\partial^2 l}{\partial \delta_s \partial \nu_2} = -\sum_{j=1}^{n_s}\Big[\frac{(1+cm+c\delta_s)y_{ij}-(1+cy_{ij})(m+\delta_s)}{(1+cm+c\delta_s)^2}\Big]$$

$$E\Big(-\frac{\partial^2 l}{\partial \delta_s \partial \nu_2}\Big|_{\delta=0}\Big) = 0 \tag{A.10}$$

$$\gamma_{st} = E\Big(-\frac{\partial^2 l}{\partial \nu_s \partial \nu_t}\Big|_{\delta=0}\Big) \tag{A.11}$$

$$\gamma_{11} = E\Big(-\tfrac{\partial^2 l}{\partial m^2}\Big|_{\delta=0}\Big)$$

$$\tfrac{\partial l}{\partial m} = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{y_{ij}}{(m+\delta_i)} - \frac{(y_{ij}+c^{-1})c}{(1+cm+c\delta_i)}\Big]$$

$$\tfrac{\partial^2 l}{\partial m^2} = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[-\frac{y_{ij}}{(m+\delta_i)^2} + \frac{(1+cy_{ij})c}{(1+cm+c\delta_i)^2}\Big]$$

$$\therefore \gamma_{11} = E\Big(-\tfrac{\partial^2 l}{\partial m^2}\Big|_{\delta=0}\Big) = \tfrac{n}{m(1+cm)}$$

$$\gamma_{12} = E\Big(-\tfrac{\partial^2 l}{\partial m \partial c}\Big|_{\delta=0}\Big)$$

$$\tfrac{\partial^2 l}{\partial m \partial c} = -\sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{(1+cm+c\delta_i)y_{ij} - (1+cy_{ij})(m+\delta_i)}{(1+cm+c\delta_i)^2}\Big]$$

$$\therefore \gamma_{12} = E\Big(-\tfrac{\partial^2 l}{\partial m \partial c}\Big|_{\delta=0}\Big) = 0$$

$$\gamma_{22} = E\Big(-\tfrac{\partial^2 l}{\partial c^2}\Big|_{\delta=0}\Big) = b(say)$$

The $C(\alpha)$ test is based on $\lambda_i(\hat{\nu}) = \phi_i(\hat{\nu}) - \beta_{1i}\eta_1(\hat{\nu}) - \beta_{2i}\eta_2(\hat{\nu})$, where β_{1i} and β_{2i} are, respectively, the partial regression coefficient of δ_i on η_1 and δ_i on η_2. The regression coefficients $\beta = (\beta_1, \beta_2)$ with $\beta_1 = (\beta_{11}, ..., \beta_{1K-1})$ and $\beta_2 = (\beta_{21}, ..., \beta_{2K-1})$ are obtained as $\Gamma\gamma^{-1}$.

$$\Gamma\gamma^{-1} = \begin{bmatrix} \frac{n_1}{m(1+cm)} & 0 \\ \frac{n_2}{m(1+cm)} & 0 \\ \vdots & \vdots \\ \frac{n_{(k-1)}}{m(1+cm)} & 0 \end{bmatrix} \begin{bmatrix} \frac{m(1+cm)}{n} & 0 \\ 0 & \frac{1}{b} \end{bmatrix} = \begin{bmatrix} \frac{n_1}{n} & 0 \\ \frac{n_2}{n} & 0 \\ \vdots & \vdots \\ \frac{n_{(k-1)}}{n} & 0 \end{bmatrix} \tag{A.12}$$

$\beta_{1i} = \frac{n_i}{n}$ and $\beta_{2i} = 0$ for $i = 1, 2, ..., K-1$. Substituting the values of $\beta'_{ij}s$ from (A.12) in λ_i , we get,

$$\lambda_i(\hat{\nu}) = \phi_i(\hat{\nu}) - \beta_{1i}\eta_1(\hat{\nu}) \tag{A.13}$$

The variance-covariance of $\Lambda(\nu) = [\lambda_1(\nu), ..., \lambda_{K-1}(\nu)]^{'}$ is

$$V = \Delta - \Gamma\gamma^{-1}\Gamma^{'}$$

Define,

$$d_i = \frac{n_i}{m(1+cm)}, \text{ for i =1,2, ...K}$$

and a vector

$$d' = (d_1, d_2, \cdots, d_{K-1}) \tag{A.14}$$

Then , we get,

$$V = \text{Diag}(d) - \frac{dd'}{1'd + d_K}$$

$$V^{-1} = \text{Diag}(1/d_1, 1/d_2, \cdots, 1/d_{K-1}) + \frac{11'}{d_K} \tag{A.15}$$

The $C(\alpha)$ statistic is obtained as $\Lambda^{'}(\Delta-\Gamma\gamma^{-1}\Gamma^{'})^{-1}\Lambda$, which is approximately distributed as a chi-square distribution with K-1 degrees of freedom.For K=4 we have,

$$C(ml)=\begin{bmatrix}\Lambda_1 & \Lambda_2 & \Lambda_3\end{bmatrix}\begin{bmatrix}\frac{1}{d_1}+\frac{1}{d_4} & \frac{1}{d_4} & \frac{1}{d_4}\\ \frac{1}{d_4} & \frac{1}{d_2}+\frac{1}{d_4} & \frac{1}{d_4}\\ \frac{1}{d_4} & \frac{1}{d_4} & \frac{1}{d_3}+\frac{1}{d_4}\end{bmatrix}\begin{bmatrix}\Lambda_1\\ \Lambda_2\\ \Lambda_3\end{bmatrix} \tag{A.16}$$

$$=\frac{{\Lambda_1}^2}{d_1}+\frac{{\Lambda_2}^2}{d_2}+\frac{{\Lambda_3}^2}{d_3}+\frac{{\Lambda_1}^2}{d_4}+\frac{{\Lambda_2}^2}{d_4}+\frac{{\Lambda_3}^2}{d_4}+\frac{2\Lambda_1\Lambda_2}{d_4}+\frac{2\Lambda_1\Lambda_3}{d_4}+\frac{2\Lambda_2\Lambda_3}{d_4}$$

$$C(ml)=\sum_{i=1}^{3}\frac{{\Lambda_i}^2}{d_i}+\frac{\left(\sum_{i=1}^{3}\Lambda_i\right)^2}{d_4} \tag{A.17}$$

Hence we have,

$$C(ml)=\sum_{i=1}^{K-1}\frac{{\Lambda_i}^2}{d_i}+\frac{\left(\sum_{i=1}^{K-1}\Lambda_i\right)^2}{d_K} \tag{A.18}$$

From (A.2) define,

$$\phi_k=\frac{n_K(\bar{y_{k.}}-m)}{m(1+cm)}$$

And from (A.12) define,

$$\beta_k=\frac{n_k}{n}$$

Then from (A.2), (A.3) and (A.13) we get,

$$\begin{aligned}\sum_{i=1}^{K-1}\Lambda_i = & \sum_{i=1}^{K-1}\phi_i-\eta_1\sum_{i=1}^{K-1}\Lambda_i\beta_{i1}\\ & =-(\phi_k-\beta_{k1}\eta_1)\\ & =-\Lambda_k\end{aligned} \tag{A.19}$$

From (B.18) and (B.19) we get,

$$C(ml) = \sum_{i=1}^{K} \frac{{\Lambda_i}^2}{d_i}. \tag{A.20}$$

Replacing $\hat{m}$ by the maximum likelihood estimate , from (B.3) we get,

$$\eta_1 = \sum_{i=1}^{K} \frac{n_i(\bar{y_{i.}} - m)}{m(1+cm)} = 0$$

This gives,

$$\Lambda_i(\hat{\nu}) = \phi_i(\hat{\nu}) \tag{A.21}$$

Hence from (A.20) and (A.21) after some algebra we get,

$$C(ml) = \sum_{i=1}^{K} \frac{n_i(\bar{y_{i.}} - \hat{m})^2}{\hat{m}(1+\hat{m}\hat{c})} \tag{A.22}$$

B Derivation of C(ql)

After reparameterizing m_i under H_1, the quasi-log-likelihood for the parameters $\delta_1, \delta_2, \cdots, \delta_{K-1}$, m and c is,

$$\begin{aligned} Q = \ & \sum_{i=1}^{K}\sum_{j=1}^{n_i} \Big[(y_{ij} + c^{-1})\text{ln}\Big(\tfrac{1+cy_{ij}}{1+c(m+\delta_i)}\Big) - y_{ij}\text{ln}\Big(\tfrac{y_{ij}}{m+\delta_i}\Big)\Big] \\ & = \sum_{i=1}^{K}\sum_{j=1}^{n_i} \Big[(y_{ij} + c^{-1})\ \text{ln}(1+cy_{ij}) - (y_{ij}+c^{-1})\text{ln}(1+cm+\delta_i) \\ & -y_{ij}\text{ln}y_{ij} + y_{ij}\text{ln}(m+\delta_i)\Big] \end{aligned} \tag{B.1}$$

Given c the unbiased estimating functions for the parameters $\delta_1, \delta_2, \cdots, \delta_{K-1}$ is,

$$U_i = \frac{\partial Q}{\partial \delta_i}$$

$$\begin{aligned} U_i = & \left.\tfrac{\partial Q}{\partial \delta_i}\right|_{\delta=0} = \sum_{j=1}^{n_i} \Big[\frac{y_{ij}}{(m+\delta_i)} - \frac{(y_{ij}+c^{-1})c}{(1+cm+c\delta_i)} \Big]\Big|_{\delta=0}. \\ & = \sum_{j=1}^{n_i} \frac{y_{ij}-m}{m(1+cm)} \\ & = \tfrac{n_i(\bar{y_{i.}}-m)}{m(1+cm)}, \end{aligned} \tag{B.2}$$

$$V_1(\delta_i, m, c) = \frac{\partial Q}{\partial m}$$

$$\begin{aligned} \left.\tfrac{\partial Q}{\partial m}\right|_{\delta=0} = & \sum_{i=1}^{K}\sum_{j=1}^{n_i} \Big[\frac{y_{ij}}{(m+\delta_i)} - \frac{(y_{ij}+c^{-1})c}{(1+cm+c\delta_i)} \Big]\Big|_{\delta=0} \\ & = \sum_{i=1}^{K} \frac{n_i(\bar{y_{i.}}-m)}{m(1+cm)} \end{aligned} \tag{B.3}$$

$$\left.\tfrac{\partial^2 Q}{\partial {\delta_i}^2}\right|_{\delta=0} = \sum_{j=1}^{n_i} \Big[-\frac{y_{ij}}{(m+\delta_i)^2} + \frac{(1+cy_{ij})c}{(1+cm+c\delta_i)^2} \Big]\Big|_{\delta=0}$$

$$E\Big(-\left.\tfrac{\partial^2 Q}{\partial {\delta_i}^2}\right|_{\delta=0} \Big) = n_i/m(1+cm) \text{ for i=j=1,2,...,K-1} \tag{B.4}$$

And,

$$E\Big(-\left.\tfrac{\partial^2 Q}{\partial \delta_i \partial \delta_j}\right|_{\delta=0} \Big) = 0 \text{ otherwise.} \tag{B.5}$$

$$\begin{aligned}\Gamma_{ij} = \quad & E\Big(-\tfrac{\partial^2 Q}{\partial \delta_i \partial \nu_j}\Big|_{\delta=0}\Big) \\ & = E\Big(-\tfrac{\partial^2 Q}{\partial \delta_i \partial m}\Big|_{\delta=0}\Big) \text{ for j=1}\end{aligned} \tag{B.6}$$

$$\tfrac{\partial^2 Q}{\partial \delta_i \nu_1}\Big|_{\delta=0} = \quad \sum_{j=1}^{n_i}\Big[-\frac{y_{ij}}{(m+\delta_i)^2} + \frac{(1+cy_{ij})c}{(1+cm+c\delta_i)^2}\Big]\Big|_{\delta=0}$$

$$\therefore E\Big(-\tfrac{\partial^2 Q}{\partial \delta_i \partial \nu_1}\Big|_{\delta=0}\Big) = \quad n_i/m(1+cm) \tag{B.7}$$

$$\tfrac{\partial^2 Q}{\partial \delta_i \nu_2}\Big|_{\delta=0} = - \quad \sum_{j=1}^{n_i}\Big[\frac{(1+cm+c\delta_i)y_{ij} - (1+cy_{ij})(m+\delta_i)}{(1+cm+c\delta_i)^2}\Big]\Big|_{\delta=0}$$

$$\therefore E\Big(-\tfrac{\partial^2 Q}{\partial \delta_i \nu_2}\Big|_{\delta=0}\Big) = 0 \tag{B.8}$$

$$\gamma_{ij} = E\Big(-\tfrac{\partial^2 Q}{\partial \nu_i \partial \nu_j}\Big|_{\delta=0}\Big) \tag{B.9}$$

$$\tfrac{\partial^2 Q}{\partial m^2}\Big|_{\delta=0} = \quad \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[-\frac{y_{ij}}{(m+\delta_i)^2} + \frac{(1+cy_{ij})c}{(1+cm+c\delta_i)^2}\Big]\Big|_{\delta=0}$$

$$\therefore \gamma_{11} = E\Big(-\tfrac{\partial^2 Q}{\partial m^2}\Big|_{\delta=0}\Big) = \tfrac{n}{m(1+cm)}$$

$$\gamma_{12} = E\Big(-\tfrac{\partial^2 l}{\partial m \partial c}\Big|_{\delta=0}\Big)$$

$$\frac{\partial^2 Q}{\partial m \partial c}\Big|_{\delta=0} = \quad -\sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{(1+cm+c\delta_i)y_{ij}-(1+cy_{ij})(m+\delta_i)}{(1+cm+c\delta_i)^2}\Big]\Big|_{\delta=0}$$

$$\gamma_{12} = E\Big(-\frac{\partial^2 Q}{\partial m \partial c}\Big|_{\delta=0}\Big) = 0$$

$$\gamma_{22} = E\Big(-\frac{\partial^2 Q}{\partial c^2}\Big|_{\delta=0}\Big) = b(say)$$

The $C(\alpha)$ test is based on $\lambda_i(\hat{\nu}) = U_i(\hat{\nu}) - \beta_{1i}V_1(\hat{\nu}) - \beta_{2i}V_2(\hat{\nu})$, where β_{1i} and β_{2i} are, respectively, the partial regression coefficient of δ_i on V_1 and δ_i on V_2.The regression coefficients $\beta = (\beta_1, \beta_2)$ with $\beta_1 = (\beta_{11}, ..., \beta_{1K-1})$ and $\beta_2 = (\beta_{21}, ..., \beta_{2K-1})$ are obtained as $\Gamma\gamma^{-1}$. Following the procedure above in appendix A, the $C(\alpha)$statistic based on the Quasi-likelihood , C(ql) is obtained as,

$$C(ql) = \sum_{i=1}^{K}\frac{n_i(\bar{y_{i.}}-\hat{m})^2}{\hat{m}(1+\hat{m}\hat{c_{mm}})}$$

C Derivation of C(eql)

Using the the parameters $\delta_1, \delta_2, \cdots, \delta_{K-1}$, m and c, the modified extended quasi-likelihood, apart from a constant term is,

$$
\begin{aligned}
Q^{+*} = \; & \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{1}{2}\ln\{1+cy_{ij}+\tfrac{c}{6}\} - \tfrac{1}{2}\ln\{(y_{ij}+\tfrac{1}{6})(1+cy_{ij})^2(1+\tfrac{c}{6})\} + (y_{ij}+\tfrac{1}{c})\ln\Big(\tfrac{1+cy_{ij}}{1+c(m+\delta_i)}\Big) \\
& -y_{ij}\ln\big(\tfrac{y_{ij}}{m+\delta_i}\big)\big)\Big] \\
& = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{1}{2}\ln\{1+cy_{ij}+\tfrac{c}{6}\} - \tfrac{1}{2}\ln(y_{ij}+\tfrac{1}{6}) - \ln(1+cy_{ij}) - \ln\big(1+\tfrac{c}{6}\big) \\
& +\big(y_{ij}+c^{-1}\big)\ln(1+cy_{ij}) - \big(y_{ij}+c^{-1}\big)\ln(1+cm+c\delta_i) - y_{ij}\ln y_{ij} + y_{ij}\ln(m+\delta_i)\Big]
\end{aligned}
$$

$$
\begin{aligned}
\phi_i = \; & \tfrac{\partial Q^{+*}}{\partial \delta_i}\Big|_{\delta=0} = \sum_{j=1}^{n_i}\Big[\frac{y_{ij}}{(m+\delta_i)} - \frac{(y_{ij}+c^{-1})c}{(1+cm+c\delta_i)}\Big]\Big|_{\delta=0} \\
& = \sum_{j=1}^{n_i}\frac{y_{ij}-m}{m(1+cm)} \qquad \text{(C.1)} \\
& = \tfrac{n_i(\bar{y_i}.-m)}{m(1+cm)},
\end{aligned}
$$

$$
\begin{aligned}
\eta_1 = \; & \tfrac{\partial Q^{+*}}{\partial \nu_1}\Big|_{\delta=0} = \tfrac{\partial Q^{+*}}{\partial m}\Big|_{\delta=0} = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{y_{ij}}{(m+\delta_i)} - \frac{(y_{ij}+c^{-1})c}{(1+cm+c\delta_i)}\Big]\Big|_{\delta=0} \\
& = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\frac{y_{ij}-m}{m(1+cm)} \qquad \text{(C.2)} \\
& = \sum_{i=1}^{K}\frac{n_i(\bar{y_i}.-m)}{m(1+cm)},
\end{aligned}
$$

$$\begin{aligned}
\eta_2 = \quad & \frac{\partial Q^{+*}}{\partial \nu_2}\Big|_{\delta=0} = \frac{\partial Q^{+*}}{\partial c}\Big|_{\delta=0} \\
= & \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{y_{ij+1/6}}{2(1+cy_{ij}+c/6)} - \frac{y_{ij}}{(1+cy_{ij})} - \frac{1/6}{2(1+c/6)} \\
& + \tfrac{y_{ij}(y_{ij}+1/c)}{(1+cy_{ij})} - c^{-2}\ln(1+cy_{ij}) - \tfrac{(y_{ij}+c^{-1})(m+\delta_i)}{1+cm+c\delta_i} + c^{-2}\ln(1+cm+c\delta_i)\Big]\Big|_{\delta=0} \\
= & \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{(1+6y_{ij})}{2(6+6cy_{ij}+c)} - \frac{y_{ij}}{(1+cy_{ij})} - \frac{1}{2(c+6)} + \frac{y_{ij}}{c} - \frac{ln(1+cy_{ij})}{c^2} \\
& - \tfrac{m(1+cy_{ij})}{c(1+cm)} + \tfrac{ln(1+cm)}{c^2}\Big] \\
= & \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{y_{ij}-m}{c(1+cm)} + c^{-2}\ln\big(\tfrac{1+cm}{1+cy_{ij}}\big) + \tfrac{1+6y_{ij}}{2(c+6+6cy_{ij})} - \tfrac{y_{ij}}{1+cy_{ij}} - \tfrac{1}{2(c+6)}\Big]
\end{aligned} \tag{C.3}$$

$$\frac{\partial Q^{+*}}{\partial \delta_s}\Big|_{\delta=0} = \sum_{j=1}^{n_s}\Big[\frac{y_{ij}}{(m+\delta_i)} - \frac{(y_{ij}+c^{-1})c}{(1+cm+c\delta_i)}\Big]\Big|_{\delta=0} \tag{C.4}$$

$$\frac{\partial^2 Q^{+*}}{\partial {\delta_s}^2}\Big|_{\delta=0} = \sum_{j=1}^{n_s}\Big[-\frac{y_{ij}}{(m+\delta_i)^2} + \frac{(1+cy_{ij})c}{(1+cm+c\delta_s)^2}\Big]\Big|_{\delta=0}$$

$$E\Big(-\frac{\partial^2 Q^{+*}}{\partial {\delta_s}^2}\Big|_{\delta=0}\Big) = n_s/m(1+cm) \text{ for s=t=1,2,...,K-1} \tag{C.5}$$

And,

$$E\Big(-\frac{\partial^2 Q^{+*}}{\partial \delta_s \partial \delta_t}\Big|_{\delta=0}\Big) = 0 \text{ otherwise.} \tag{C.6}$$

$$\Gamma_{st} = E\Big(-\frac{\partial^2 Q^{+*}}{\partial \delta_s \partial \nu_t}\Big|_{\delta=0}\Big) \tag{C.7}$$

$$\frac{\partial^2 Q^{+*}}{\partial\delta_s\partial\nu_1} = \sum_{j=1}^{n_s}\Big[-\frac{y_{ij}}{(m+\delta_s)^2}+\frac{(1+cy_{ij})c}{(1+cm+c\delta_s)^2}\Big]$$

$$E\Big(-\frac{\partial^2 Q^{+*}}{\partial\delta_s\partial\nu_1}\Big|_{\delta=0}\Big) = n_s/m(1+cm) \tag{C.8}$$

$$\frac{\partial^2 Q^{+*}}{\partial\delta_s\partial\nu_2} = \sum_{j=1}^{n_s}\Big[\frac{(1+cm+c\delta_i)y_{ij}-(1+cy_{ij})(m+\delta_i)}{(1+cm+c\delta_i)^2}\Big]$$

$$E\Big(-\frac{\partial^2 Q^{+*}}{\partial\delta_s\partial\nu_2}\Big|_{\delta=0}\Big) = 0 \tag{C.9}$$

$$\gamma_{st} = E\Big(-\frac{\partial^2 Q^{+*}}{\partial\nu_s\partial\nu_t}\Big|_{\delta=0}\Big) \tag{C.10}$$

$$\gamma_{11} = E\Big(-\frac{\partial^2 Q^{+*}}{\partial m^2}\Big|_{\delta=0}\Big)$$

$$\frac{\partial Q^{+*}}{\partial m} = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{y_{ij}}{(m+\delta_i)}-\frac{(y_{ij}+c^{-1})c}{(1+cm+c\delta_i)}\Big]$$

$$\frac{\partial^2 Q^{+*}}{\partial m^2} = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[-\frac{y_{ij}}{(m+\delta_i)^2}+\frac{(1+cy_{ij})c}{(1+cm+c\delta_i)^2}\Big]$$

$$\therefore \gamma_{11} = E\Big(-\tfrac{\partial^2 Q^{+*}}{\partial m^2}\Big|_{\delta=0} \Big) = \tfrac{n}{m(1+cm)}$$

$$\gamma_{12} = E\Big(-\tfrac{\partial^2 Q^{+*}}{\partial m \partial c}\Big|_{\delta=0} \Big)$$

$$\tfrac{\partial^2 Q^{+*}}{\partial m \partial c} = \; -\sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{(1+cm+c\delta_i)y_{ij}-(1+cy_{ij})(m+\delta_i)}{(1+cm+c\delta_i)^2}\Big]$$

$$\therefore \gamma_{12} = E\Big(-\tfrac{\partial^2 Q^{+*}}{\partial m \partial c}\Big|_{\delta=0} \Big) = 0$$

$$\gamma_{22} = E\Big(-\tfrac{\partial^2 Q^{+*}}{\partial c^2}\Big|_{\delta=0} \Big) = b(say)$$

The $C(\alpha)$ test is based on $\lambda_i(\hat{\nu}) = \phi_i(\hat{\nu}) - \beta_{1i}\eta_1(\hat{\nu}) - \beta_{2i}\eta_2(\hat{\nu})$, where β_{1i} and β_{2i} are, respectively, the partial regression coefficient of δ_i on η_1 and δ_i on η_2.The regression coefficients $\beta = (\beta_1, \beta_2)$ with $\beta_1 = (\beta_{11}, ..., \beta_{1K-1})$ and $\beta_2 = (\beta_{21}, ..., \beta_{2K-1})$ are obtained as $\Gamma\gamma^{-1}$. Following the procedure above in appendix A, the $C(\alpha)$statistic based on the Extended Quasi-likelihood , $S_c(eql)$ is obtained as,

$$S_c(eql) = \sum_{i=1}^{K}\frac{n_i(\bar{y_{i.}} - \hat{m})^2}{\hat{m}(1+\hat{m}\hat{c_{eql}})}$$

where $\hat{c}_{eql}$ is the maximum extended quasi-likelihood estimate of c, under H_0 , obtained by solving

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{y_{ij}-\hat{m}}{c(1+c\hat{m})} + c^{-2}\ln\Big(\tfrac{1+c\hat{m}}{1+cy_{ij}}\Big) + \tfrac{1+6y_{ij}}{2(c+6+6cy_{ij})} - \tfrac{y_{ij}}{1+cy_{ij}}\Big] = \tfrac{n}{2(c+6)}.$$

D Derivation of C(deql)

The double extended quasi-likelihood excluding constant term, using the reparameterization of m_i under H_1, can be written as

$$\begin{aligned} p_v{}^*(DEQ) = \;& \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[y_{ij}\ln(m+\delta_i) + \Big(y_{ij}+\tfrac{1}{c}\Big)\ln\Big(\tfrac{1+cy_{ij}}{1+c(m+\delta_i)}\Big) - \tfrac{1}{2}\ln(1+cy_{ij}) \\ & + \tfrac{c}{12(1+cy_{ij})} - \tfrac{c}{12} - \tfrac{1}{12y_{ij}}\Big] \end{aligned} \tag{D.1}$$

Now, define $\delta = (\delta_1, ..., \delta_{K-1})$ and $\nu = (\nu_1, \nu_2)^{'} = (m,\ c)^{'}$. Then,

$$\begin{aligned} \phi_i = \;& \tfrac{\partial p_v{}^*(DEQ)}{\partial \delta_i}\Big|_{\delta=0} = \sum_{j=1}^{n_i}\Big[\frac{y_{ij}}{(m+\delta_i)} - \frac{(y_{ij}+c^{-1})c}{(1+cm+c\delta_i)}\Big]\Big|_{\delta=0}. \\ & = \sum_{j=1}^{n_i}\frac{y_{ij}-m}{m(1+cm)} \\ & = \tfrac{n_i(\bar{y_i}.-m)}{m(1+cm)}, \end{aligned} \tag{D.2}$$

$$\begin{aligned} \eta_1 = \;& \tfrac{\partial p_v{}^*(DEQ)}{\partial \nu_1}\Big|_{\delta=0} = \tfrac{\partial p_v{}^*(DEQ)}{\partial m}\Big|_{\delta=0} = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{y_{ij}}{(m+\delta_i)} - \frac{(y_{ij}+c^{-1})c}{(1+cm+c\delta_i)}\Big]\Big|_{\delta=0}. \\ & = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\frac{y_{ij}-m}{m(1+cm)} \\ & = \sum_{i=1}^{K}\frac{n_i(\bar{y_i}.-m)}{m(1+cm)}, \end{aligned} \tag{D.3}$$

$$\begin{aligned} \eta_2 = \;& \tfrac{\partial p_v{}^*(DEQ)}{\partial \nu_2}\Big|_{\delta=0} = \tfrac{\partial p_v{}^*(DEQ)}{\partial c}\Big|_{\delta=0} \\ & = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{(y_{ij}+\frac{1}{c})y_{ij}}{(1+cy_{ij})} - \frac{ln(1+cy_{ij})}{c^2} - \frac{(y_{ij}+\frac{1}{c})(m+\delta_i)}{(1+cm+c\delta_i)} + \frac{ln(1+cm+c\delta_i)}{c^2} \\ & - \tfrac{y_{ij}}{2(1+cy_{ij})} + \tfrac{1}{12(1+cy_{ij})^2} - \tfrac{1}{12}\Big]\Big|_{\delta=0} \\ & = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{y_{ij}-m}{c(1+cm)} + \frac{1}{c^2}\ln\big(\tfrac{1+cm}{1+cy_{ij}}\big) - \tfrac{y_{ij}}{2(1+cy_{ij})} - \tfrac{cy_{ij}(2+cy_{ij})}{12(1+cy_{ij})^2}\Big] \end{aligned} \tag{D.4}$$

$$\Delta_{st} = E\Big(-\tfrac{\partial^2 p_v{}^*(DEQ)}{\partial\delta_s\partial\delta_t}\Big|_{\delta=0}\Big) \tag{D.5}$$

$$\tfrac{\partial p_v{}^*(DEQ)}{\partial\delta_s} = \sum_{j=1}^{n_s}\Big[\frac{y_{ij}}{(m+\delta_s)} - \frac{(y_{ij}+c^{-1})c}{(1+cm+c\delta_s)}\Big]$$

$$\tfrac{\partial^2 p_v{}^*(DEQ)}{\partial{\delta_s}^2} = \sum_{j=1}^{n_s}\Big[-\frac{y_{ij}}{(m+\delta_s)^2} + \frac{(1+cy_{ij})c}{(1+cm+c\delta_s)^2}\Big]$$

$$\tfrac{\partial^2 p_v{}^*(DEQ)}{\partial{\delta_s}^2}\Big|_{\delta=0} = \sum_{j=1}^{n_s}\Big[-\frac{y_{ij}}{m^2} + \frac{(1+cy_{ij})c}{(1+cm)^2}\Big]$$

$$E\Big(-\tfrac{\partial^2 p_v{}^*(DEQ)}{\partial{\delta_s}^2}\Big|_{\delta=0}\Big) = n_s/m(1+cm) \tag{D.6}$$

And,

$$E\Big(-\tfrac{\partial^2 p_v{}^*(DEQ)}{\partial\delta_s\delta_t}\Big|_{\delta=0}\Big) = 0 \tag{D.7}$$

$$\Gamma_{st} = E\Big(-\tfrac{\partial^2 p_v{}^*(DEQ)}{\partial\delta_s\partial\nu_t}\Big|_{\delta=0}\Big) \tag{D.8}$$

$$\tfrac{\partial^2 p_v{}^*(DEQ)}{\partial\delta_s\partial\nu_1} = \sum_{j=1}^{n_s}\Big[-\frac{y_{ij}}{(m+\delta_s)^2} + \frac{(1+cy_{ij})c}{(1+cm+c\delta_s)^2}\Big]$$

$$E\Big(-\frac{\partial^2 p_v{}^*(DEQ)}{\partial\delta_s\partial\nu_1}\Big|_{\delta=0}\Big) = \quad n_s/m(1+cm) \tag{D.9}$$

$$\frac{\partial^2 p_v{}^*(DEQ)}{\partial\delta_s\partial\nu_2} = - \quad \sum_{j=1}^{n_s}\Big[\frac{(1+cm+c\delta_s)y_{ij}-(1+cy_{ij})(m+\delta_s)}{(1+cm+c\delta_s)^2}\Big]$$

$$E\Big(-\frac{\partial^2 p_v{}^*(DEQ)}{\partial\delta_s\partial\nu_2}\Big|_{\delta=0}\Big) = 0 \tag{D.10}$$

$$\gamma_{st} = E\Big(-\frac{\partial^2 p_v{}^*(DEQ)}{\partial\nu_s\partial\nu_t}\Big|_{\delta=0}\Big) \tag{D.11}$$

$$\gamma_{11} = E\Big(-\frac{\partial^2 p_v{}^*(DEQ)}{\partial m^2}\Big|_{\delta=0}\Big)$$

$$\frac{\partial p_v{}^*(DEQ)}{\partial m} = \quad \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{y_{ij}}{(m+\delta_i)} - \frac{(y_{ij}+c^{-1})c}{(1+cm+c\delta_i)}\Big]$$

$$\frac{\partial^2 p_v{}^*(DEQ)}{\partial m^2} = \quad \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[-\frac{y_{ij}}{(m+\delta_i)^2} + \frac{(1+cy_{ij})c}{(1+cm+c\delta_i)^2}\Big]$$

$$\therefore \gamma_{11} = E\Big(-\frac{\partial^2 l}{\partial m^2}\Big|_{\delta=0}\Big) = \frac{n}{m(1+cm)}$$

$$\gamma_{12} = E\Big(-\frac{\partial^2 p_v{}^*(DEQ)}{\partial m\partial c}\Big|_{\delta=0}\Big)$$

$$\tfrac{\partial^2 p_v{}^*(DEQ)}{\partial m \partial c} \; = \; -\sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{(1+cm+c\delta_i)y_{ij}-(1+cy_{ij})(m+\delta_i)}{(1+cm+c\delta_i)^2}\Big]$$

$$\therefore \gamma_{12} = E\Big(-\tfrac{\partial^2 p_v{}^*(DEQ)}{\partial m \partial c}\Big|_{\delta=0}\Big) = 0$$

$$\gamma_{22} = E\Big(-\tfrac{\partial^2 p_v{}^*(DEQ)}{\partial c^2}\Big|_{\delta=0}\Big) = d(say)$$

The $C(\alpha)$ test is based on $\lambda_i(\hat{\nu}) = \phi_i(\hat{\nu}) - \beta_{1i}\eta_1(\hat{\nu}) - \beta_{2i}\eta_2(\hat{\nu})$, where β_{1i} and β_{2i} are, respectively, the partial regression coefficient of δ_i on η_1 and δ_i on η_2.The regression coefficients $\beta = (\beta_1, \beta_2)$ with $\beta_1 = (\beta_{11}, ..., \beta_{1K-1})$ and $\beta_2 = (\beta_{21}, ..., \beta_{2K-1})$ are obtained as $\Gamma\gamma^{-1}$. Following the procedure above in appendix A, the $C(\alpha)$statistic based on the Double Extended Quasi-likelihood , $S_c(deql)$ is obtained as,

$$S_c(deql) = \sum_{i=1}^{K}\frac{n_i(\bar{y_{i.}}-\hat{m})^2}{\hat{m}(1+\hat{m}\hat{c_{deql}})}$$

where $\hat{c}_{deql}$ is the maximum double extended quasi-likelihood estimate of c, under H_0 , obtained by solving

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i}\Big[\frac{y_{ij}-\hat{m}}{c(1+c\hat{m})} + \frac{1}{c^2}\ln\big(\tfrac{1+c\hat{m}}{1+cy_{ij}}\big) - \tfrac{y_{ij}}{2(1+cy_{ij})} - \tfrac{cy_{ij}(2+cy_{ij})}{12(1+cy_{ij})^2}\Big] = 0 \qquad \text{(D.12)}$$

Bibliography

Agresti, A. (1990). *Categorical Data Analysis.* New York: Wiley.

Allison, P.D. (2001). *Missing Data.* Thousand Oak, CA: Sage.

Anderson, T., & Taylor, J. B. (1976). Strong consistency of least squares estimates in normal linear regression. *The Annals of Statistics*, 788-790.

Anscombe, F. (1949). The statistical analysis of insect counts based on the negative binomial distribution. *Biometrics, 5*(2), 165-173.

Barnwal, R., & Paul, S. (1988). Analysis of one-way layout of count data with negative binomial variation. *Biometrika, 75*(2), 215-222.

Beal, J.M. (1939). Cytological Studies in Relation to the Classification of the Genus Calochortus. *Botanical Gazette, 100*(3), 528-547.

Bishop, Y. V. V., Fienberg, S. E., and Holland, P. W. (1975), *Discrete Multivariate Analysis.* Cambridge, MA: MIT Press.

Bliss, C. I., & Fisher, R. A. (1953). Fitting the negative binomial distribution to biological data. *Biometrics, 9* (2), 176-200.

Böhning, D., Dietz, E., Schlattmann, P., Mendonca, L., & Kirchner, U. (1999). The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 162*(2), 195-209.

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 33*(1), 38-44.

Buck, S.F. (1960). A method of estimation of missing values in multivariate data suitable for use with an electronic computer. *Journal of the Royal Statistical Society, Series B, 22*(2), 302-306.

Casella, G., & George, E.I. (1992). Explaining the Gibbs sampler. *The American Statistician, 46*(3), 167-174.

Chen, J., Hubbard, S., & Rubin, Y. (2001). Estimating the hydraulic conductivity at the south oyster site from geophysical tomographic data using Bayesian techniques based on the normal linear regression model. *Water Resources Research, 37*(6), 1603-1613.

Collings, B. J., & Margolin, B. H. (1985). Testing goodness of fit for the Poisson assumption when observations are not identically distributed. *Journal of the American Statistical Association, 80*(390), 411-418.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological), 39*(1), 1-22.

Deng, D., & Paul, S. R. (2000). Score tests for zero-inflation in generalized linear models. *The Canadian Journal of Statistics, 28*(3), 563-570.

Deng, D., & Paul, S. R. (2005). Score tests for zero-inflation and overdispersion in generalized linear models. *Statistica Sinica, 15*, 257-276.

Efron, B., & Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected fisher information. *Biometrika, 65*(3), 457-483.

Engel, J. (1984). Models for response data showing extra-Poisson variation. *Statistica Neerlandica, 38*(3), 159-167.

Geweke, J. (1986). Exact inference in the inequality constrained normal linear regression model. *Journal of Applied econometrics, 1*(2), 127-141.

Hutto, R.L., Pletschet, S.M., & Hendricks, P. (1986). A fixed-radius point count method for nonbreeding and breeding season use. *The Auk, 103*(3), 593-602.

Ibrahim, J. G. (1990). Incomplete data in generalized linear models. *Journal of the American Statistical Association, 85*(411), 765-769.

Ibrahim, J. G., Chen, M.-H., & Lipsitz, S. R. (1999). Monte Carlo EM for missing covariates in parametric regression models. *Biometrics, 55*(2), 591-596.

Ibrahim, J. G., Chen, M.-H., & Lipsitz, S. R. (2001). Missing responses in generalised linear mixed models when the missing data mechanism is nonignorable. *Biometrika, 88*(2), 551-564.

Ibrahim, J. G., Chen, M.-H., Lipsitz, S. R., & Herring, A. H. (2005). Missing-data methods for generalized linear models: A comparative review. *Journal of the American Statistical Association, 100*(469), 332-346.

Ibrahim, J. G. & Lipsitz, S. R. (1996). Parameter estimation from incomplete data in binomial regression when the missing data mechanism is nonignorable. *Biometrics*, 1071-1078.

Kelly, B. C. (2007). Some aspects of measurement error in linear regression of astronomical data. *The Astrophysical Journal, 665*(2), 1489.

Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics, 15*(3), 209-225.

Lee, Y., & Nelder, J. A. (2001). Hierarchical generalised linear models: a synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika, 88*(4), 987-1006.

Lipsitz, S. R., & Ibrahim, J. G. (1996). A conditional model for incomplete covariates in parametric regression models. *Biometrika, 83*(4), 916-922.

Little, R. J., & Rubin, D. B. (1987, 2002, 2014, 2020). *Statistical analysis with missing data.* John Wiley & Sons.

Little, T. D., Jorgensen, T. D., Lang, K. M., & Moore, E. W. G. (2014). On the joys of missing data. *Journal of pediatric psychology, 39*(2), 151-162.

Luo, R., & Paul, S. (2018). Estimation for zero-inflated beta-binomial regression model with missing response data. *Statistics in Medicine, 37*(26), 3789-3813.

Maiti, T., & Pradhan, V. (2009). Bias reduction and a solution for separation of logistic regression with missing covariates. *Biometrics, 65*(4), 1262-1269.

Margolin, B. H., Kaplan, N., & Zeiger, E. (1981). Statistical analysis of the Ames Salmonella microsome test. *Proceedings of the National Academy of Sciences, 78*(6), 3779-3783.

McCaughran, D. A., & Arnold, D. W. (1976). Statistical models for numbers of implantation sites and embryonic deaths in mice. *Toxicology and Applied Pharmacology, 38*(2), 325-333.

Mian, R., & Paul, S. (2016). Estimation for zero-inflated over-dispersed count data model with missing response. *Statistics in medicine, 35*(30), 5603-5624.

Nakai, M., & Ke, W. (2011). Review of the methods for handling missing data in longitudinal data analysis. *International Journal of Mathematical Analysis, 5*(1), 1-13.

Nelder, J. A., & Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika, 74*(2), 221-232.

Neyman, J. (1959). Optimal asymptotic tests of composite hypotheses. *Probability and statsitics*, 213-234.

Paul, S., & Banerjee, T. (1998). Analysis of two-way layout of count data involving multiple counts in each cell. *Journal of the American Statistical Association, 93*(444), 1419-1429.

Paul, S., & Plackett, R. (1978). Inference sensitivity for Poisson mixtures. *Biometrika, 65*(3), 591-602.

Paul, S., & Saha, K.K. (2007). The generalized linear model and extension: a review and some biological and environmental applications.*Environmetrics, 18*(4), 421-443.

Paul, S. R.,& Deng, D. (2000). Goodness of fit of generalized linear models to sparse data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 62*(2), 323-333.

Piegorsch, W. W. (1990). Maximum likelihood estimation for the negative binomial dispersion parameter. *Biometrics*, 863-867.

Plackett, R. L. (1981). *The analysis of Categorical Data.* London:Griffin.

Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association, 92*(437), 179-191.

Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *In Mathematical Proceedings of the Cambridge Philosophical Society*, 44, 50-57. Cambridge University Press.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581-592.

Saha, K. K. (2008). Analysis of one-way layout of count data in the presence of over or under dispersion. *Journal of statistical planning and inference, 138*(7), 2067-2081.

Sahu, S. K., & Roberts, G. O. (1999). On convergence of the EM algorithm and the Gibbs sampler. *Statistics and Computing, 9*(1), 55-64.

Sinha, S., & Maiti, T. (2008). Analysis of matched case-control data in presence of nonignorable missing exposure. *Biometrics, 64*(1), 106-114.

Thall, P. F. (1992). Score tests in the two-way layout of counts. *Communications in Statistics - Theory and Methods, 21*(10), 3017-3036.

Wedderburn, R. W. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika, 61*(3), 439-447.

Zhang, C.-H., & Huang, J. (2008). The sparsity and bias of the lasso selection in high-dimensional linear regression. *The Annals of Statistics, 36*(4), 1567-1594.

Vita Auctoris

Poonam Shrestha Malakar was born in 1982 in Nepal . She obtained her Masters degree in Statistics from Tribhuvan University of Kathmandu, Nepal in 2007, then she worked as a lecturer of Statistics and Reserch methodology in various reputed colleges, universities in Kathmandu. She is currently a candidate for a Ph.D. in Statistics at the University of Windsor and will graduate in Fall 2022.