e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

Research Paper

LINEAR MODELS : ENERGY FORECASTING, PERFORMANCE EVALUATION METHODS.

SUGUNA.T
Research Scholar
M.Phil Mathematics
Bharath Institute Of Higher Education And Research
Mail Id: rsugu1971@gmail.com
Guide Name: **Dr. R. DEEPA**
Assistant Professor, Department Of Mathematics
Bharath Institute Of Higher Education And Research

Address for Correspondence

SUGUNA.T
Research Scholar
M.Phil Mathematics
Bharath Institute Of Higher Education And Research
Mail Id: rsugu1971@gmail.com
Guide Name: **Dr. R. DEEPA**
Assistant Professor, Department Of Mathematics
Bharath Institute Of Higher Education And Research

ABSTRACT

Discrete outcome variable is common in public health, behavioral sciences and in many medical applications; the Poisson regression model is useful to analyze discrete random variable. For clustered discrete outcome variable where the observations are correlated among individual subjects, the number of observed discrete is sometimesgreater than the expected frequency of the Poisson distribution and the discrete random variables are over-dispersed. Overdispersion is familiar in discrete random variable models particularly within the area of ecology and biological science because of missing covariates, non-independent, aggregations of data and an excess frequency of zeros. Every cluster levels received a singular level of a random effect that models the extra Poisson variation given within the data, are usually utilized to discuss heterogeneity in discrete random variable. However, studies investigating that the power of cluster level random effects as a way to discrete random variablemodel with over-dispersion is scarce. A situation where the variance of the response variable exceeds the mean, and hence, both over- dispersion and heterogeneity problems occur, in the appropriate imposition of the Poisson model may underestimate the standard error and overestimate the significance of the regression parameters, and so, giving misleading inference about the regression

parameters. Often, because of the hierarchal study design procedure, zero- inflation, over-dispersed and lack of independence may occur simultaneously, which render the standard ZIP model is inadequate. The multilevel ZIP and MZINB regression model are suitable for examining clustered correlated and over- dispersed discrete random variablewith many zeros. In this thesis, derivation of a proposed score test for evaluating the over-dispersed, heterogeneity and zero-inflation parameters in discrete random variableregression models are performed.

INTRODUCTION

BACKGROUNDOF THE STUDY

Modeling is the heart of applied Mathematical and Statistical sciences. Model is the key component in any Mathematical and Statistical analysis. In recent years in all mostall fields of science, several research works have been directed to either the Mathematical models or the Statistical models. Model is a set of structural and functional relationships that can be expressed interms of mathematical equations.A Mathematical model is a set of mathematical equations concerns with two or more variables. By introducing an error random variable or a disturbance term, the mathematical model becomes a stochastic model or statistical model.

Generally, the mathematical model or the statistical model may be specified either in the form of a linear model or in the form of a nonlinear model. The linear modelconsists of a set of linear equations concerns with two or more variables. Linear model has received greatest attention both in theory and in practice. From the theoretical point of view, it is mathematically tractable, and in practical applications of the wide variety, ithas shown itself to be of great value. Further, many non-linear models can often be rearranged to be in a linear form.

Regression method is a statistical technique for investigating and modeling the relationship between the dependent and independent variables. Linear regression analysis establishes an average linear relationship between a dependent variable and a set of independent variables. Applications of linear regression models have a vital role in

analyzing various mathematical and statistical problems on different fields of science such as Economics, Business, Management, Engineering, Agriculture, Medicine, Social science, Biological and Life sciences, Physical sciences and Technology.

STATEMENT OF THE PROBLEMS

Nested data are very common in social sciences, psychology, health science, and others. The classical

linear models assuming that the observations havethe same effects across groups.When the observations have differenteffects across the groups, the linear regression model is not appropriate. Ignoring group membership and focuses exclusively on inter-individual variation and on individual level attributes. This approach has a drawback of ignoring the potential importance of a group-level attribute in influencing an individual-level outcome. In addition, if outcomes for individuals within groups are correlated, the assumption of independence of observation is violated, resulting in an incorrect standard error and inefficient estimate (Diggle PJ et al., 1994).

The method focuses exclusively on inter-group variation and data aggregated to the group level variables. This approach eliminates the non-independence problem mentioned above but has the drawback of ignoring the role of individual-level variables

in shaping the outcome. Both methods essentially collapse all variables to the same level and ignore the multilevel structure.These approaches allow and define separate regression for each group coefficients to differ from group to group, but does not examine how specific group-level properties may affect individual-level outcomes or interact with individual-level variables.

REVIEW OF LITERATURE

The purpose of this chapter is to give a brief summary of concepts and theories about the generalized linear model for clustered discrete random variable, overdispersion, zero-inflation and the power of tests. This chapter also explains the review of the literature on the discreterandom variableregression models.

LINEAR STATISTICAL MODELLING

Regression analysis is collection of statistical techniques for modeling and investigating the relationship between a response variable of interest and a set of predictor variables. In the classical regression model,the response variable y which is ourmain interest, select a sample of size n from our population of interest and observe values y_i, i = 1, . . . , n, then wish to infer properties of the variable y in terms of other observed predictors$x_i = (x_{1i}, \ldots, x_{ki})$. The main use of these predictor variables is to acount for differences in the response variableor to put it another way to explain the variation in y.Consider the classical linear model

$$y_{ij} = X_{ij}\beta + e_i \quad \ldots\ldots (2.1)$$

whereβ is the coefficient of regression for X_{ij}, and $e_i$is a random or error term and assumed that the error terms are identically and independentely distributed a normaldistribution

with mean zero and variance σ^2. Also assumed that the outcome variable y

e ij

follows a normal distribution with mean $X_{ij}\beta$ and variance σ^2. The coeffi$_e$cient of determination

(R^2) is a measure of the amount of variance in the dependent variable explained by the independent variable(s).

SCORE TEST FOR HOMOGENIETY OF GROUPS IN THE MULTILEVEL POISSON MODEL FOR CLUSTERED DISCRETE RANDOM VARIABLE

3..1. INTRODUCTION

Data with a multilevel nature often happens in public health, health service research, behavioral sciences, and in medicine. Examples embrace patients nested within hospitals, residents nested within a geographical area, students nested within a class, class nested within facilities and staff nested within a corporation. For clustered discrete random variable, the observations are usually correlated, there are three main approaches for correlated discrete random variable have been proposed (Perntile, 1988); conditional models, random effects model, generalized estimating equations (GEE`s). Conditional models Rosner (1984) are convenient only for particular cases, such as data with small group sizes or with order structure.

Analysts are progressively aware that the multilevel regression models are an appropriate way to analyze clustered data. Aconsequence of clustering of the groups within clusters or higher level units is that subjects from the same area could have more similar outcomes than subjects came from a different area. Themultilevel regression models incorporate cluster-specific random effects that account for the dependency of the observation by partitioning the total individual variance into variation as a result of the cluster. The likelihood ratio test between the ordinary regression model and mixed effects models may be used as a homogeneity test (Self and Liang 1987).This chapter focuses on a multilevel Poisson regression modelapproach, the distribution of the response variables is modelled conditionally in a group-specific parameter that is itself a random variable (Laird and Ware 1982, Stiratelli, Laird, and Ware 1984), to take of the coefficient of regression and random parameters in Poisson discrete points.

THE MULTILEVEL POISSON REGRESSION MODEL

In the model formulation, when the outcome variable is discrete denoting the number of time that an incident occurred, a Poisson regression model can be accustomed relate the mean number of events to

e-ISSN 2320 –7876 www.ijfans.org
Research Paper
Vol.11, Iss.9, Dec 2022

a group of explanatory variables employing a logarithmic link function. Then, the Poisson regression model will be used as $lo(\mu_i) = X_i\beta$, $Y_i \sim Poisson\ (\mu_i)$ ---------- (3.1)

where X_i denotes a p×1 column matrix of covariates measured with the i^{th} subject, Y_i

denotes the discrete outcome variablemeasured with the i^{th} subject, denotes a 1×p row matrix of the regression coefficients and the parameter μ_i denotes the expected or mean number of events for the i^{th} subject given their set of observed covariates.Consider a two-level random intercept Poison regression model and it is assumed that the intercept is allowed to vary randomly across the groups.

Let Y_{ij} be the responsevariable for the observation jof group i, i =1, 2… k; j = 1, 2… n_i

k

with $N = \sum n_i$, the conditional distributions of the outcome variable $i=1$

$y = (y_{i1}, y_{i2}, ..., y_{in_i})'$, given a set of cluster level random effects u_i and the probability

density functions of Y_{ij} is defined as follows

$$f(Y_{ij}|\mu_{ij}) = \prod_{i=1}^{n_i} \exp\{y_{ij}(x_{ij}\beta + z_{ij}u_i) - \exp(x_{ij}\beta + z_{ij}u_i) - \log(y_{ij}!)\}$$

$$= \exp\left\{\sum_{i=1}^{n_i} \{y_{ij}(x_{ij}\beta + z_{ij}u_i) - \exp(x_{ij}\beta + z_{ij}u_i) - \log(y_{ij}!)\}\right\}$$

$$= \exp\left\{\sum_{i=1}^{n_i} \{\theta_i y_{ij} - g(\theta_{ij})\} + C(y_{ij}, c_i)\right\} \quad \text{.....(3.2)}$$

e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

Research Paper

i=1

e-ISSN 2320 –7876 www.ijfans.org
Research Paper
Vol.11, Iss.9, Dec 2022

where $(y_{ij}, c_i) = \sum^{n_i}$ $_{i=1}log\ (y_{ij}!)$,which$C(y_{ij}, c_i)$ does not depend on the model parameters.The mean and the variance ofY_{ij} are respectively,

$$\mu_{ij} = E(y_{ij}/\alpha_i) = g'(\theta_{ij}) = exp\ -x_{ij}\beta + z_{ij}\ i\ /\ \text{and}$$

$$\sigma^2 = Var(y_{ij}) = c_i g''(\theta_{ij}) = \mu_{ij} = exp\ -x_{ij}\beta + z_{ij}\ /,$$
ij i

where′denotesthe differentiation with respect toparameter θ_{ij}. Consider the mixedeffects modelallowing atleast one regression coefficient to be random is

$\theta_{ij} = lo(\mu_{ij}) =$ ij $= x_{ij}\beta + z_{ij}\alpha_i$ ----------(3.3)

whereβ denotes a $p\times1$ vector of fixed effects with its associated design vector $x_{ij}$and

$\alpha_i$is the scalar random subject with associated covariatesz_{ij}. In this model the mean

$\mu_{ij}$and the variance σ^2 are conditi$_i$o$_j$nally ona$_i$, now to test homogeneity across andwithin groups, consider the random intercept model inwhich z_{ij}=1 for all *i, j*.To test

the homogeneity hypothesis that all of the variables and the correlation among the random effects are zero in a generalized linear mixed model.

The parameter a$_i$ can be written as $\alpha_i = \alpha + D^{1/2}u_i$, where the$u_i$`s areindependently and identically distributed as a normal distribution with zero mean and unit variance. Therefore, $\alpha_i$`s are identical and independently distributed with mean α and variance D. Our interest is to test $H_0: D = 0$ against the alternative H_0 :

$D > 0$. Note that fordiscrete random variablemodels; this is equivalent for testing homogeneity across groups as well as testing homogeneity within groups.

The first and the second partial derivatives of the log likelihood function with respect to the scalar random subject parameterα_i for the multilevel Poisson regression model is given by

e-ISSN 2320 –7876 www.ijfans.org
Research Paper
Vol.11, Iss.9, Dec 2022

$$\left[\frac{\partial}{\partial \alpha_i} \log f_i(\beta, \alpha, , D = 0)\right] = \sum_{j=1}^{n_i} (y_{ij} - \mu_{ij})_j$$

and

$$\frac{\partial^2}{\partial \alpha^2} \log f_{ij}(\beta, \alpha, , D = 0) = \frac{\partial}{\partial \alpha_i} \sum_{j=1}^{n_i} (y_{ij} - \mu_{ij}) z_j = -\sum_{j=1}^{n_i} z^2 \sigma^2_{ij}$$

Using the first and second partial derivation of the log likelihood equationwith respect to a_i, the score statistic is given by

$$S(\beta, \alpha,) = \sum_{i=1}^{k} \{ \sum_{j=1}^{n_i} (y_{ij} - \mu_{ij}) z_{ij} - \sum_{j=1}^{n_i} [z_{ij} \sigma_{ij}] \}^2$$

Then the score test statistic for testing homogeneity H_0: D = 0 for the known nuisance parameters β will be

$$H_{PD} = \frac{S^2_{PD}(\beta, \alpha)}{\beta\beta} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.5)$$

$(I_{DD} - I_{D\beta} I^{-1} I_{\beta D})$

Now, the asymptotic variance function as the group size $k \to \infty$ of $S_{PD}(\beta, a, D)$ under H_0 (Cox and Hinkley, 1974) can be expressed as a function of information matrix. Then the asymptotic variance function for the score test is

$$(\beta) = I_{DD} - I_{D\beta} I^{-1} I_{\beta D \beta\beta}$$

e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

Research Paper

where $I = \sum^{k} \frac{6l_i}{DD_{i=1} E\, 0_{6D} \mid D = 01}$ is a scalar

$$I_{\beta\beta} = \sum_{i=1}^{k} E\, 8\frac{-\partial^2 l}{\partial\beta\partial\beta`}{}^{i}9, \qquad I_{\beta D} = I_{D\beta} = \sum_{i=1}^{k} E\, 8\frac{-\partial^2 l}{\partial\beta\partial D}{}^{i}9$$

e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

Research Paper

PARAMETRIC ESTIMATIONS OF SCORE TEST BASED ON THE MULTILEVEL POISSON REGRESSION MODEL

Now to evaluate the variance of the score function (S) defined as Var(D) =$I_{DD} - I_{D\beta}I^{-1}I_{\beta D}$. The i^{th} summand of I_{DD} can be written as

$$E\left(\frac{\partial^2 l_i}{\partial D^2}\,\middle|\, D=0^*\right) = \frac{1}{2}\sum_{i=1}^{n_i}\sum_{j=1}^{n_i} E\left\{\left(\sum_{j=1}^{n_i}(y_{ij}-\mu_{ij})\right)^2 - \sum_{ij}\sigma^2\right\}$$

To solve the variance of the score function, let us to define

$U_{ij} = y_{ij} - \mu_{ij}$, the i^{th} term of the score test can be written as

$$\frac{\partial l_i}{\partial D}\Big|_{D=0} = \frac{1}{2}\left(\sum_{j=1}^{n_i} U_{ij}\right)^2 - \sum_{j=1}^{n_i}\mu_{ij}$$

Thus

$$E\left(\frac{\partial l_i}{\partial D}\,\middle|\, D=0^*\right)^2 = \frac{1}{4}\left(U_i - \sum_{j=1}^{n_i}\sigma_{ij}\right)^2$$

e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

Research Paper

where$U_i = \sum^{n_i}_{j=1} U_{ij}$, now since $(U_{ij}) = 0$,then

$$(U^2) = E\left[\sum_i U^2\right] = E\left[\sum_{ij} U^2 + \sum_{j\neq j'} \sum U_{ij} U_{ij'}\right] = E\left[\sum_{ij} U^2\right]$$

$$= E\left[\sum_{ij} (_{ij} - \mu_{ij})^2\right] = \sum_{j=1}^{n_i} \sigma^2_{ij}$$

Therefore,

$$E\left(\frac{\partial l}{\partial D}\Big|_{D=0}\right)^2 = \frac{1}{4} E^*U^2 - (U^2)^{+2} = \frac{1}{4} Va(U^2) = \frac{1}{4}(\mu_4 - \mu^2)$$

whereμ_2 and μ_4 are the second and the fourth central moments ofY_{ij}, respectively,which can be expressed as a function of the second and the fourth cumulates K_2 and

e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

Research Paper

K_4 of Y_{ij} (Kendall and Stuart, 1977). $\mu_4 = K_4 + 3K^2_{22} = K_2 = \sigma^2 = \mu_{ij}$, then

$\mu_4 = \mu_{ij} + 3\mu^2$. After simplification the values of I_{DD} is

$$\frac{1}{4}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\mu_{ij} - \mu_{ij} + 2\mu^2_{ij})$$

The variance of the score functions can be derived from the Fisher information.

$$I(\beta) = \begin{pmatrix} I_{\beta\beta} & I_{\beta D} \\ I_{D\beta} & I_{DD} \end{pmatrix} *$$

where

$$\frac{\partial l_i}{\partial \beta} = \sum_{j=1}^{n_i} - \sum_{j=1}^{n_i} \mu_{ij} x_{ij} x_{ij}$$

$$\frac{\partial^2 l}{\partial \beta \partial \beta' \partial \beta} = \frac{\partial l_i}{\partial \beta} > \sum_{j=1}^{n_i} y_{ij} x_{ij} - \sum_{j=1}^{n_i} \mu_{ij} x_{ij} ? = - \sum_{i=1}^{k} \sum_{j=1}^{n_i} \mu_{ij} x_{ij} x_{ij}'$$

Therefore

$$-\frac{\partial^2 l}{}$$

e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

Research Paper

k n_i

$-\partial^2 log f$ $(\beta, \alpha, , D =$

0

I $_{\beta\beta} = \sum E\,8\_\_\_\_^{i}9 = \sum\sum E\,8\,\frac{ij}{\partial\beta\partial\beta`}\,9\ D = 0$

$\partial\beta\partial\beta`$

i=1 i=1 j=1

k k n_i

n

i

$= \sum\sum g''(\theta_{ij})x_{ij}\ x_{ij}' = \sum\sum \mu_{ij}x_{ij}\ x_{ij}'$ k $n_i\ n_i$

i=1 j=1 ∂li=1 j=1

k $-\partial^2 l$

$I_\beta = I_{D\beta} = \sum E\,8\_\_\_\_\_^{i}9 = \sum E\,\frac{i}{\partial\alpha_i} > \sum y_{ij}x_{ij} - \sum \mu_{ij}x_{ij}?$

$\partial\beta\partial D$

i=1 i=1 j=1 j=1

k k n_i

$n\,1$

i

1

$= {}_2\Sigma\Sigma g'''(\theta_{ij})x_{ij} = {}_2\Sigma\Sigma \mu_{ij}x_{ij}$

i=1 j=1 i=1 j=1

where$x_{ij} = (1, x_{1ij}, \ldots . x_{n}$ij)`.

which asymptotically as n→ ∞ has a chi-square distribution with one degree of freedom, now the maximum likelihood estimate of β can be estimated iteratively by using Fisher`s scoring method from the following equations.

e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

Research Paper

MODEL SELECTION

If there are several models to be compared, in order to select the best modelwhich fits the data instead of using the likelihood ratio test, it can be selected by using the Akakie information criteria (AIC) and Bayesian information criteria (BIC).

3.7.1 AKAKIE INFORMATION CRITERIA

AIC is the most common means of identifying the model which fits thedatawell by comparing two or more than two models. The goodness of fit test against the complexity of the model is similar to that of the coefficient of multiple determination(R^2); however, it penalized by the number of parameters included in the complexity of the model. Unlike the R^2, the good model is the one which has the minimum AIC value.It is given by the following formula AIC=−2ℓ+2k,whereℓ is the log likelihood function of a model that will compare with the other models and k is the number of parameters in the model including the intercept (Ismail and Jemain, 2007).

BAYESIAN INFORMATION CRITERIA

Unlike the Akakie information criteria, the Bayesian information criteria take into account the size of the data under consideration. It is given byBIC=−2ℓ+$k$log (n) whereℓ is the log-likelihood of a model that will compare with the other models, n is the sample size of the data and k is the number of parameters in the model including the intercept.

SIMULATION STUDY

In this section,a simulation study is conducted to compare the proposed and the existing models in terms of sizes and powers. For studying the properties of the statistic in terms of empirical size, generatingdiscrete random variable from a Poisson distribution under the null hypothesis of homogeneity and assume that the random effects parameterareonce(z_{ij} = 1) andthe samples are comprised of 10; 20; 50;

100observations and 5; 10; 20; 50 groups with simulating data, the multilevel Poissondistribution is simulated for the distribution of the response variable assuming that under the null hypothesis homogeneity within different number of groups according to the variance D of the distribution of the group-specific random effects of the response variable. Each simulation experiment for level and power was basedon 1000 simulated samples.

In the simulated model, we reviewed the literature to define initial parameter values. We presented the result of a small simulation study examining the empirical size and power of the test statistics discussed in this thesis. The following log linear model for the response variable is assumed (see

e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

Research Paper

Jacqmin-Gadda and Commenges (1995)).$\log(\mu_{ij}) = 0.8x_{1ij} + 0.5u_i - 0.5$ (3.8)

$y_{ij} \sim \text{Poisson}(\mu_{ij})$, i =1, 2… k; j = 1, 2… n_i ,

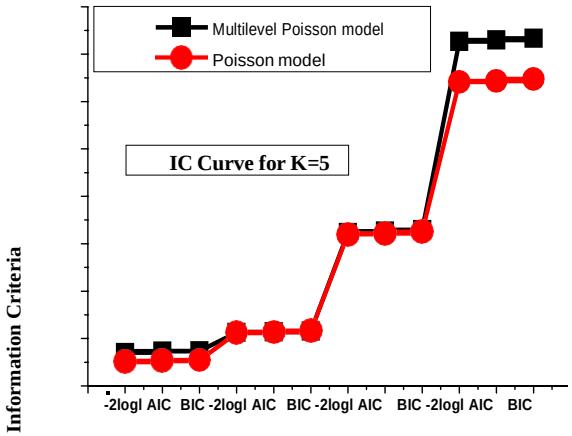
For the variable $x_{1ij}$is a subject-specific effect which is simulated from a uniform distribution with mean zero and unit varianceand u_i is group specificeffects and simulated from a standard normal distribution and generate a set of random numbers from a uniform distribution in the interval (0,1) as the values ofx_{1ij}.For drawing samples and for estimating the maximum likelihood estimates of the regression and homogeneity parameters of interest under the null hypothesis, $a_i = a + D^{1/2}u_i$.

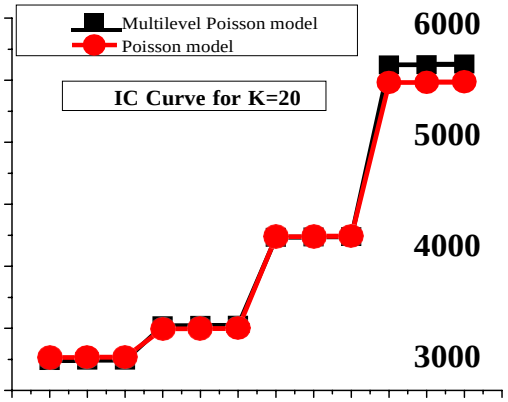
To simulate correlated data, added a group-specific random effect under the hypothesis of homogeneity, that is,D = 0, where the u_i`s are identical and independently distributed with a standard normal distribution with mean zero and unit variance. Therefore, $\alpha_i$`s are identical and independently distributed with mean α and variance D.For each set of generated data, a multilevel Poisson model is fitted for calculating the score test and the existing tests followed by the powers of the score tests. Results from the simulation study are presented in Table- 3.1and 3.2.

In Table 3.1 the results investigated thathow the information criteria perform in the multilevel Poisson regression model selection problems via simulations. Model with smaller AIC is considered to be better.When the number of clusters is large and the number observation is small then the multilevel Poisson regression model is better than the standard Poisson regression model, whereas when the sample size and cluster numbers are small then the Poisson regression model is better than the multilevel Poisson regression model. The simulated data results indicated that the performance of the criteria to select the true model generally involved with an increase of sample size, despite differences in performance among the information criteria. The simulated resultsindicate that the performance of the model depends on the sample size and the number of clusters.

e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

Research Paper



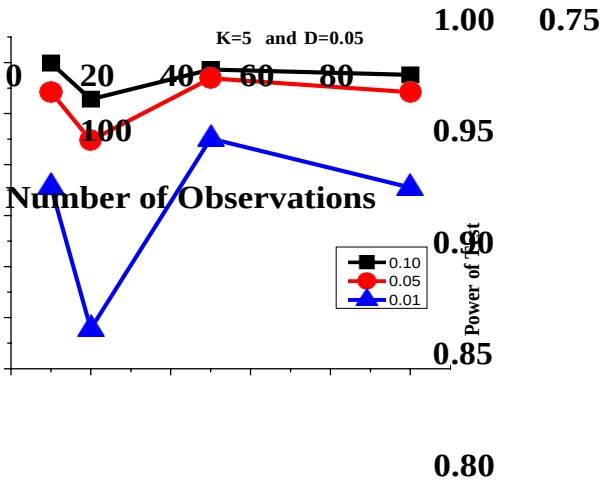


e-ISSN 2320 –7876 www.ijfans.org
Research Paper
Vol.11, Iss.9, Dec 2022

Data are generated from the multilevel Poisson distributio under the nullhypothesis on 1000 replication. In the simulation, thelevels (10%; 5% and 1%), the sample size (10, 20, 50 and 100) and the number of groups (5, 10, 20 and 50) areconsidered.

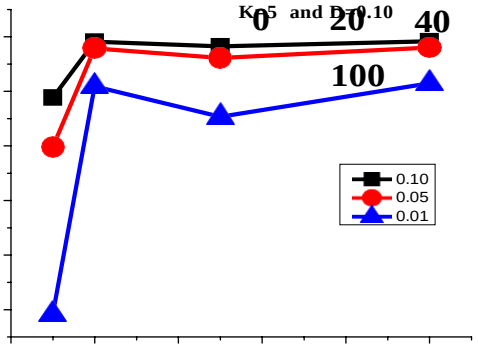
In note that an error probability increases power increases. The power of the tests of the three scenario increases when a increases, and for large sample groups and small variance for the group effect (k = 50, n = 10, D=0.05) the power increase fast and approaches to 1. For small sample groups, when the standard deviations of the group effectsincrease, the power increases slowly, whereas in large sample groups, (k = 50, n = 50, D=0.15 and a =0.1) as standard deviations of the group effect increases, the power increases slowly, however, when the values of D increase from 0.05 to 0.15, the power decreases. Generally, as the number of groups increases, the power is slightly increases. Therefore, the proposed score test is more important for testing and controlling heterogeneity of the group effects by fixing the number of observation and number of groups due to its high power to predict the model.

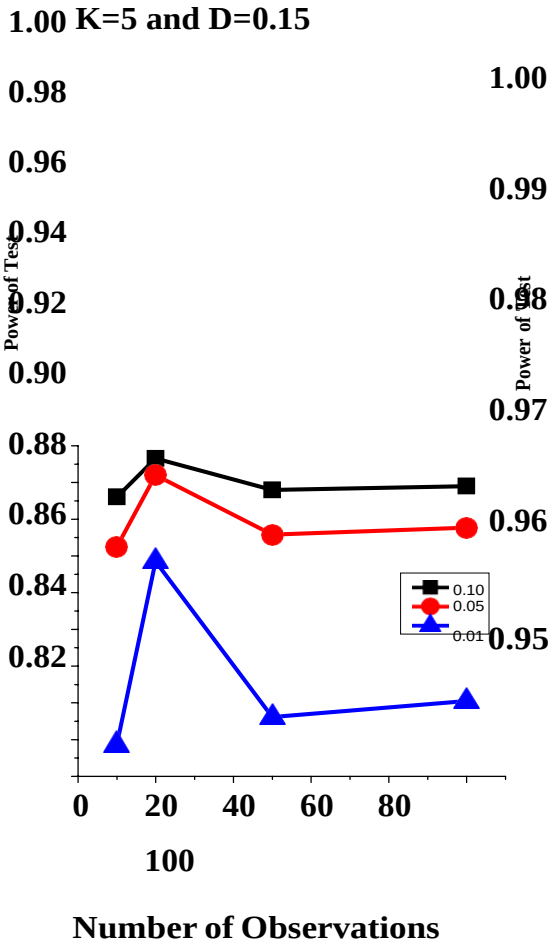
Increasing the sample size will decrease the standard error (and increasepower). Similarly, increase the amount of variance in x will increase power. However, increasing amount of unexplained variance will serve to decrease power. Sample size is not the only factor for power in the multilevel model, effects on power can depend on the parameters (intercept estimate, individual specific intercept, slope estimate, individual specific slope, standard error of the average (between person), slope error, variances of the independent variable, variances of individual slope (multilevel), sample size and cluster size), Sean P. Lane and Erin P. Hennes (2018).

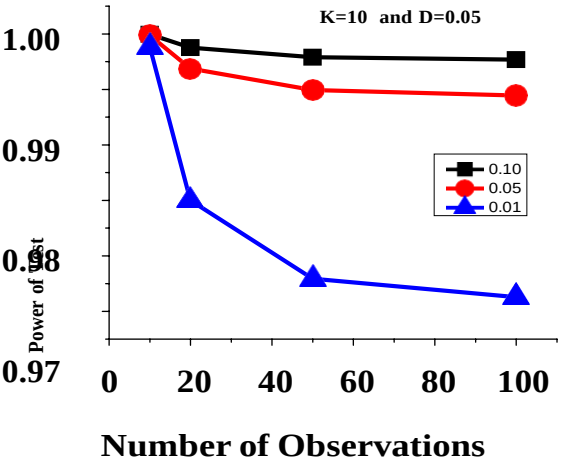


e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

Research Paper

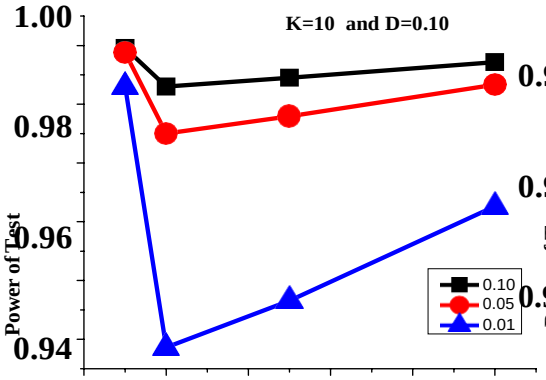


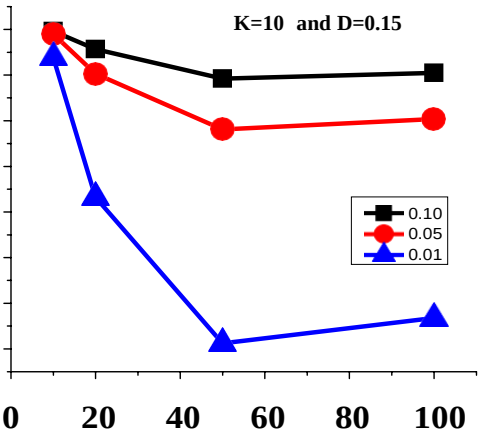


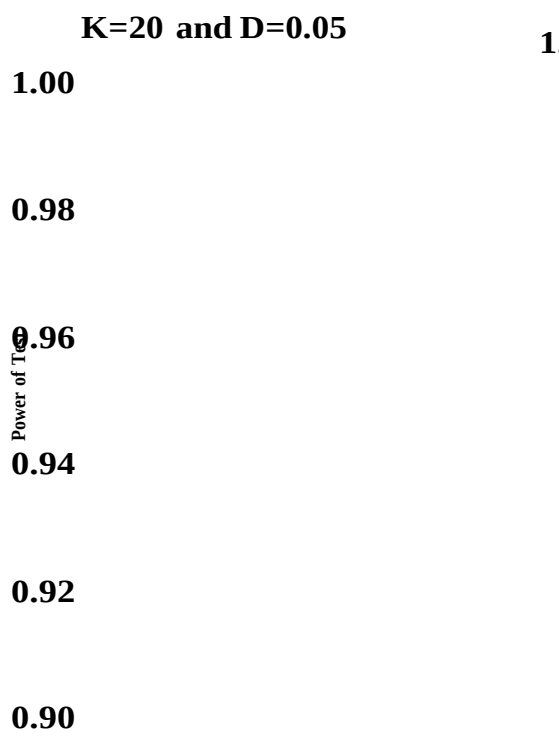


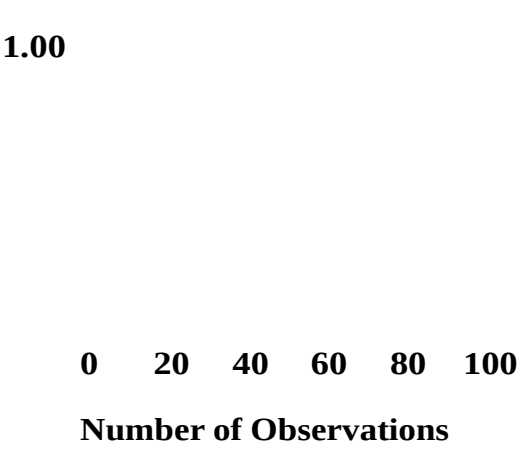
e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

Research Paper

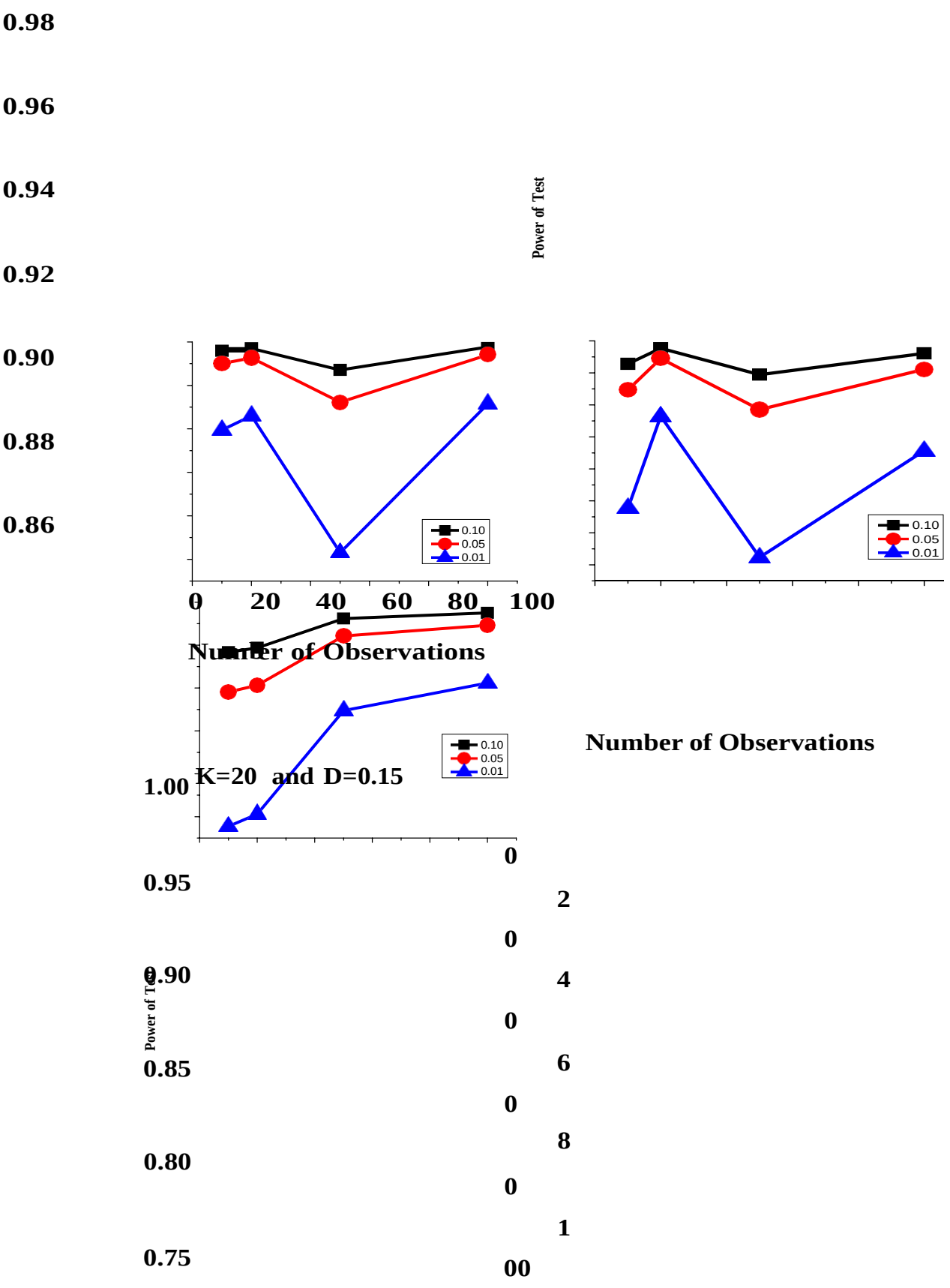






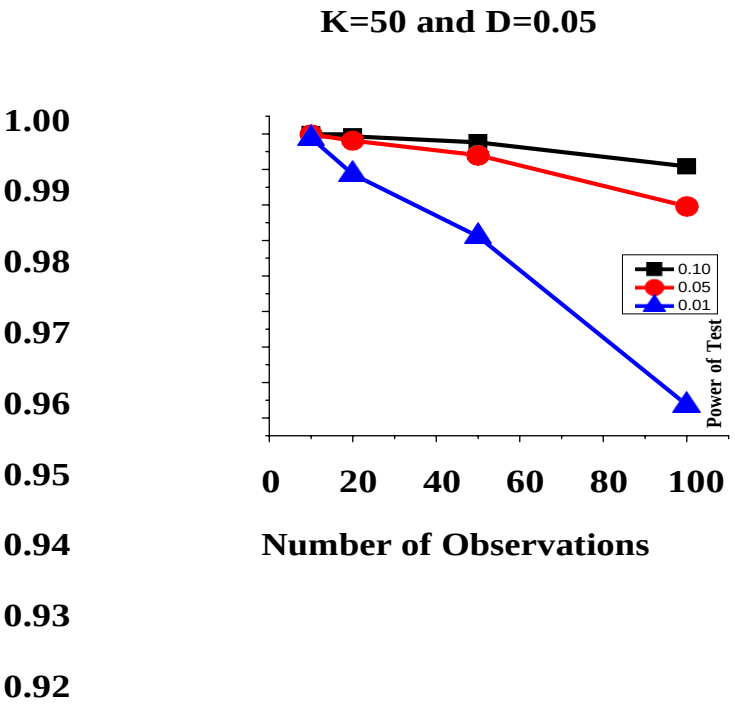


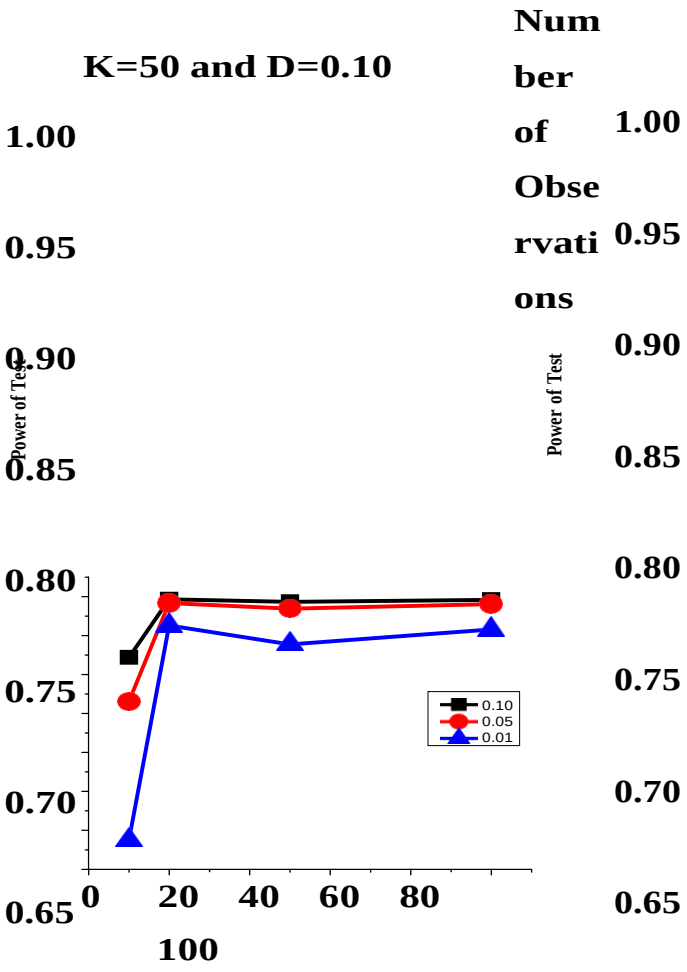
e-ISSN 2320 –7876 www.ijfans.org
Research Paper
Vol.11, Iss.9, Dec 2022



e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

Research Paper





e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

Research Paper

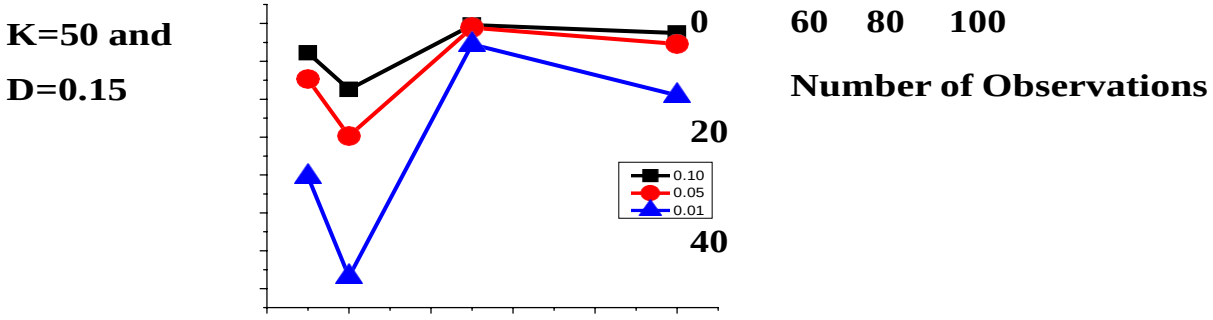


Figure 3.2.Simulated curve for power and goodness of fit test of the models

e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

Research Paper

In this article, the power of the multilevel Poisson regression models viasimulated data results arepresented. A simulation and application data are used to illustrate our method. The results revealed that the proposed score testis more preferable than the existing model. Based on the information criteria, AIC, the

proposed model is better than the standard Poisson model. The results in Table3.8,revealed that deaths of children varied among regions. In addition, all covariates and dummy explanatory variables were found to be significant differencein the deaths of children between regions. The AIC values of the empty model with random intercept are larger than that of the random intercept and fixed coefficient model, which implies that the random intercept model is better than the standard Poisson model and also on the predicted probability, the multilevel Poisson regression model is better than a Poisson regression model.

CONCLUSION

In clustered discrete random variable, when the responses of each observation are correlated, familiar ANOVA and regression models do not give optimal analysis. The standard multilevel models yield correct inference for clustered normally distributed data. Generalized linear models specifically the Poisson and negative binomial regression models give correct inference for non clustered data. In this thesis, developed the multilevel discrete random variablemodelsillustrated withexamples and simulation study, presentingsome score test statistics foranalyzing overdispersion, zero-inflated and heterogeneity in clustered discrete random variable and compare the existing model with its alternative model tests and identified the best test statistic for testing coefficients of regression, overdispersion, zero inflation, and heterogeneity parameters of clustered discrete random variable interms of its power and size.

we developa proposed score test based on the multilevel Poisson model for testing heterogeneous parameterin the equidispersed clustered discrete random variableand analyzed the Poisson regression model with the assumption that it is used for model fit under the null hypothesis. Furthermore, the likelihood ratio tests are used as an alternative test to select the best model.Fromthe simulation and application study results shown that when the dataset has heterogeneous groups indiscrete random variable, the multilevel Poisson regression model gives a good and correct result in the analysis while Poisson regression is clearly not adequate for handling heterogeneous data. However, when

the data has homogenous between groups, the Poisson regression model is more reliable. From the simulation study, for fixed values of sample size, when the coefficient of regression and heterogeneity parameter are increasing the power of the scores are increasing. On the other hand, forfixed values of the coefficient of regression and heterogeneous parameters, when the sample size increasing the power of the score test is increasing. For large values of thesample size and coefficient of the regression and heterogonous parameters, then the difference among different tests become trivial in terms of its power. For other cases, the proposed score test is more appropriate for general use because of its high Power.

REFERENCES

- Abramowitz, M., and Stegun,I.A. (1972). Handbook of Mathematical Functions, New York: Dover Publications, Inc.
- Agresti, A. (1996). An introduction to categorical data analysis, John Wiley and Sons, New York.
- Alker, H. R. (1969). A typology of ecological fallacies, in Dogan, M. and Rokkan,S.(eds.), Quantitative Ecological Analysis in the Social Sciences. Cambridge, MA: MIT Press.
- Anseombe, FJ. (1967). Topics in the Investigation of Linear RelationsFitted by the Method of Least Squares, *J.R.S.S., Series*, 29, 1-32.
- Bohning,D. (1998). Zero inflated models and C.A.MAN: a tutorial collection of evidence, *Biometrical*, 40, 833-843.
- Bolger, N., Zuckerman, A., & Kessler, R. C. (2000). Invisible support and adjustment to stress.Journal of Personality and Social Psychology, 79, 953– 961.Sage Publications, New York
- Bolker BM, Brooks ME, Clark CJ, Geanges SW, Poulsens MHH, WhiteJSS. (2009). The generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology &Evolution* ,24, 127-135.
- Breslow, N. E. (1999). Tests of hypotheses in over dispersed Poisson regression and other quasi-likelihood models, *Journal of the American Statistical Association*, Vol. 85, 565-571.
- Bryk, A.S., & Raudenbush, S. W. (1992). Hierarchical linear models (applicationsand data analysis methods). Sage Publications, New York.
- Cameron, A. Colin &Trivedi, Pravin K., (1990). Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics, Elsevier*, vol.46(3), 347-364.
- Diggle, P.J., Liang, K.Y. and Zeger, S.L. (1994). The Analysis of Longitudinal Data.

Oxford. Clarendon.

- Draper, N.R. and Smith. If. (1998). Applied Regression Analysis. Third Edition, John Wileg&Sons. New York.
- Efroyrnson. M.A. (1960). Multiple Regression Analysis. In A.Ralston andH.S. Wile (Eds.) Mathematical Methods for Digital Computers,John Wiley, New York
- Fisher, R. A. (1941). Average excess and average effect of a gene substitution. *Annals of Eugenics* 11, 53–63.
- Fitzmaurice GM, Laird NM, Ware JH(2004). Applied longitudinal analysis.New Jersey: John Wiley & Sons Inc
- Gardner, W., Mulvey, E.P., and Shaw, E.C (1995). Regression Analyses of Discretes and Rates: Poisson, Over-dispersed Poisson, and Negative Binomial Models, *Psychological Bulletin*, 118, 392-404.
- Hilbe JM.(2011). Negative binomial regression (2^{nd} edition). Cambridge:Cambridge University Press.
- Hinde, J. P. and Demetrio, C. G. B. (1998). Overdispersion: models andestimation, *Computational Statistics and Data Analysis*, Vol. 27, 151-170.
- Hocking, R.R. (1976). The Analysis and Selection of Variables in LinearRegression, *Biometrics*,32, 1-49.
- Jacqmin-Gadda, H., and Commenges, D.(1995). Tests of homogeneity for generalized linear models. *Journal of the American Statistical Association*, 90, 1237 – 1246.
- Jansaku, N. and Hinde, J.P.(2004). Linear mean variance negative binomial models for analysis orange tissue culture. *Journal of Science and Technology*,Vol. 26, No.5, 683 – 689.
- JahsakulN.and Hinde,J.P.(2002). Score Tests for Zero Inflated Poisson Models, *Comput. Data Anal*, 40,75-96.
- Jahsakul,N., and Hinde, J.P.(2009). Score Tests for extra-Zero models in zero inflated negative binomial models, *commun. Statist. Simul. Comput*, 38, 92-108.
- Kendall, M.G. and Stuart,A. (1977). The Advanced theory of statistics. Macmillan, New York, Vol. 1, 168.
- Kibria, B.M.G., Kristofer, M. and Shukur, G. (2013). Some Ridge Regression Estimations for zero inflated Poisson Model. Journal Applied Statistics.40(4),721-735.
- Leckie, G. and Goldstein, H. (2009). The limitations of using school league tables to inform school choice, *Journal of the Royal Statistical Society: Series* A, 172, 835-851.

- Lee, A.H., Wang, K.,Yau, K.K.W., Carrivick, P.J.W., and Stevenson, M.R. (2005). Modelling bivariate discrete series with excess zeros. *Mathematical Biosciences* 196, 226 –237.
- McCulloch, C. E., Searle, S. R. &Neuhaus, J. M. (2008). Generalized, Linear, and Mixed Models. Hoboken, New Jersey: John Wiley & Sons, Inc., 2nd ed.
- McGilchrist, C.A. (1994). Estimation in generalized mixed models, *J.Roy. Statist.Soc. Ser.*B 56, 61-69
- McLachlan G. J.,and Krishan, T. (1997). The EM Algorithm and extensions, Wiley, New York.
- Neyman, J. (1959). Optimal asymptotic tests for composite hypothesis, *In Probability and Statistics*, The Harold Cramer, 213-34, U. Grenander (ed), New York, Wiley.
- Neyman, J. and Scott, E. L. (1966). On the use of (α) Optimal tests of composite hypothesis, *In Bulletins of the Institute of the International Statistics*, Vol. 41, 477-497.
- Osgood, D. W., and Chambers, J. M. (2000). Social disorganization outside the metropolis: An analysis of rural youth violence. *Criminology* 38: 81–115.
- Paul, S. R. and Banerjee, T. (1998). Analysis of two-way layout of Discrete random variable involving multiple discretes in each cell.*Journal of the American Statistical Association*, Vol. 93, 1419-1429.
- PAUL, S.R., andAZAD, K.(2011). Testing Homogeneity in Clustered (Longitudinal) Discrete random variable Regression Model with Over-Dispersion department of Mathematics & Statistics, University of Windsor, Windsor, Canada.
- Pearson, K.(1914). Life and Latters and Labours of Francis Galton, University Press, Cambridge.
- Pettiquanti. M. and Neqaraja, N. (2017). Power analysis for negative binomial models with application to multiple Sclerosis clinical trials, *Journal of binomial Biopharm Statistics*, Vol. 22, No.2.
- Rabe-Hesketh, S, and Skrondal, A. (2005).Multilevel and Longitudinal Modelling using Stata, Stata Press, Stata Corp, College Station, Texas.
- Rabe-Hesketh, S., Skrondal, A. and Zheng, X. (2012). Generalized multilevel structural equation modeling. In Hoyle, R. (Ed.). *Handbook of Structural Equation Modelling*. Guilford Press, 512-531.
- Rao, C. R. (1947). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation, *Proceedings of the Cambridge Philosophical*

Society, Vol. 44, 50–57.

- Saleh, A.K.Md.E., Arashi, M., and Kibria, B.M.G.(2019). Theory of Ridge Regression Estimation with Applications. Wiley, New York.
- Sean P. Lane and Erin P. Hennes (2018). Power struggles: Estimating sample size for multilevel relationships research. Journal of Social andPersonal Relationships, Vol. 35(1) 7–31. Purdue University, USA
- Thall, P. F. (1992). Score tests in the two-way layout of discretes, Communications in Statistics, *Theory and Methods*Vol. 21, 3017-3036.
- Trivedi,P.K. (1990). Memorandum of Understanding: An Approach to Improving Public Enterprise Performance, International Management Publishers.
- Van den Broek J. (1995). A score test for zero-inflation in a Poisson distribution, *Biometrics*, 51, 738 –743.
- Verbeke, G., and Molenberghs, G. (2000). Linear mixed models for longitudinal data, Springer Series in Statistics, Springer-Verlag, New-York.
- Vonesh, E.F. and Chinchilli, V.M. (1997). Linear and Nonlinear Models for Analysis of Repeated Measurements. Marcel Dekker, New York.
- Wang ,K.,Yau, K.K.W., and Lee, A.H. (2002). A zero inflated Poisson mixed model to analyze diagnosis related groups with majority of same day hospital stays, *Comput. Meth. Programs Biomed*, 68, 195 – 203.
- Wang, P., Puterman, M. L., Cockburn, I. and Le, N. (1996). Mixed Poisson regression models with covariate dependent rates, *Journal of Biometrics*, Vol. 52, 381-400.
- Xiang, L., Lee, A.H., Yau,K.K.W., and McLachlan, G.J. (2006). A score test for zero inflation in correlated Discrete random variable, *Statist.Med*, 25, 1660-1671.
- Xiang, L., Lee, A.H., Yau, K.K.W., and McLachlan,G.J.(2007). A score test for overdispersion in zero inflated Poisson mixed regression model, *Statist.Med*, 26, 1608-1622.
- Xie, M., He, B., and Goh, T. N. (2000). Zero-inflated Poisson model in statistical process control, *Computational Statistics and Data Analysis*, Vol. 38, 191-201.
- Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden P. A., Heath, A. C., Martin, N. G., Montgomery, G.W., Goddard, M. E., &Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, 42, 565–569.
- Yau,K.K.W., and Lee, A.H.(2001). Zero inflated Poisson regression with random effects to evaluation an occupation injury prevention programs, *Statist. Med*, 20, 2907-2920.

- Yau, K.K.,Lee, A.H. and Ng, A.S.K. (2003). Finite mixture regression model with random effects: Application to neonatal hospital length of stay, *Computational Statistics & Data Analysis* 41, 359 – 366.
- Zhao, Y.,James,H., Cheryl, L. (2009). Score tests for over dispersion in zero-inflated Poisson mixed models, Epidemiology and Biostatistics,Arnold School of Public Health, University of South Carolina, SC 29208, USA.
- Zhu, H. And Zhang, H. (2006). Generalized score test of homogeneity for mixed effects models. *The Annals of Statistics, 34, 1545 – 1569.*
- Zuur,AF., leno EN, Walker.NJ.,Saveliev, AA., Smith. GM. (2009). Mixed effects models and extensions in ecology with R. New York: Springer.