E-ISSN: 1115 – 8492 P-ISSN: 3027 – 0693
MPRI-JNSAR | INTERNATIONAL JOURNAL OF NATURE AND SCIENCE ADVANCE RES.
VOL. 14 NO. 9 OCTOBER-2026

STATISTICAL MODELLING OF OUTBOUND GLOBAL SYSTEM FOR MOBILE COMMUNICATIONS (GSM) CALL COUNTS USING COUNT REGESSION MODELS

ISMAILA A. BOLARINWA; BUSHIRAT T. BOLARINWA; & ABRAHAM ZAKA

Department of Statistics, The Federal Polytechnic, P.M.B. 55, Bida, Niger State, Nigeria.

DOI: https://doi.org/10.70382/mejnsar.v14i9.0112

Abstract

This article compared Poisson and Negative Binomial Models on outbound Global System for Mobile Communications (GSM) call counts based on primary data collected from mobile telecommunications users among students of The Federal Polytechnic, Bida, Niger State, Nigeria. The data provided for 315 respondents: the outbound call count, gender, level of study, age, distance from respondent's base to Bida and number of siblings. The Poisson regression and Negative Binomial models were utilized for data modeling where outbound call count constituted the dependent variable. Models were estimated through the maximum likelihood method. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were the performance metrics used. The empirical distribution was right skewed with a mean of 4.324 calls and variance of 15.016 and an unconditional ratio of variance to mean of 3.473. The Pearson dispersion statistic from the Poisson regression was 3.175 and significant overdispersion was observed with the Cameron-Trivedi auxiliary test. The Negative Binomial (NB) model estimated a dispersion parameter of 0.383 ($p < .001$), reduced the Pearson dispersion ratio to 1.180, and produced substantially better likelihood-based fit statistics (log-likelihood = −772.562; AIC = 1,563.124; BIC = 1,596.897) than the Poisson model (log-likelihood = −891.820; AIC = 1,799.640; BIC = 1,829.660) and the Zero-inflated Poisson (log-likelihood = −889.107; AIC =1,796.213; BIC = 1,829.986). The preferred NB model found age to be statistically significant at 5% level (ß = 0.030, p = 0.015) yielding Incidence-Rate Ratio (IRR) of 1.030 and indicating that a one-

Keywords: Count data; GSM calls; Poisson regression; Negative Binomial regression; Overdispersion; Nigeria

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

year increase in age is associated with approximately 3.0% increase in the expected number of calls, holding other variables constant. The coefficient of ND2 was negative (ß = -0.246, p = 0.034) but significant. Its IRR of 0.782 indicates that, controlling for the other predictors, ND2 respondents have an expected call count approximately 21.8% lower than the ND1 reference group. Other variables were found to be not significant at 5% level. The study concludes Negative Binomial model as the appropriate for GSM calls data under consideration. The study recommends duplication of the research and extension to staff in other higher institutions of higher learning.

Introduction

The prevalence of mobile telephony in Nigeria has revolutionised how individual people communicate with one another and the use of the mobile phone has become a quantifiable means of communication in the modern era. Individual call frequency needs to be considered within the context of the national telecommunications system. According to the National Bureau of Statistics (2024), Nigeria had 224.7 million active voice subscriptions during the fourth quarter of 2023, and 219.3 million active voice subscriptions in the first quarter of 2024. In 2024, the International Telecommunication Union (ITU) estimated that there were about 165 million mobile-cellular subscriptions in Nigeria and mobile ownership and network coverage suggested high access to mobile communication technologies. These are aggregate data, however, and do not show how these individual characteristics relate to the number of calls that are made. The calls in one observation period may be made by a respondent or not made by a respondent, and the other observation period may be made by several respondents, with the result that one observation is either a zero or a positive value and the other observation is a positive value and can be highly skewed.

Number of calls will definitely yield count data. Count data are data that come from a process where the response is the number of times an event occurs over a time period or within an observational unit. The outcome is an integer which makes the problem different from standard linear regression, as the conditional distribution is often asymmetric, and the variance is not necessarily constant. In recent methodological literature, Poisson regression is recommended as a natural first choice for count outcomes, and the equidispersion assumption should not be taken for granted and needs to be investigated (Altun, 2019; Green, 2021; Inchausti, 2022). The significant importance of overdispersion is that the lack of knowledge about excess conditional variability can lead to underestimation of standard errors and thereby to wrong statistical inferences (Campbell, 2021; Fathurahman, 2023).

The negative Binomial regression model includes a dispersion parameter that can accommodate cases where there exists heterogeneity, which makes the model more flexible than the standard Binomial regression model (Pagui et al., 2022; Tiara et al., 2023). This methodological problem is particularly relevant for the global system for mobile communications (GSM) call data, because there is a potential for large differences in the calling behaviour of people with similar observable properties. There may be significant differences between the social networks, communication

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

needs, socio-economic status, geographic status and mobile phone usage patterns of two students at the same age and in the same grade. This unobserved heterogeneity can cause more variation in call counts than is allowed by the Poisson model. The current study thus does not assume a priori that using Poisson regression is appropriate but rather considers the frequency of GSM calls an empirical count-distribution problem. The aim of this study is therefore to identify which of the count-data specifications (Poisson regression & Negative Binomial) is suitable for the observed frequency of outbound GSM calls.

THEORETICAL FRAMEWORK

The Normal distribution and the Poisson distribution

Count data are observations that are represented by the non-negative integers Y= 0, 1, 2, … (conventionally). The answer may be a description of what happened during a specific period of time, e.g., the number of phone calls placed in one day. Count responses, in contrast to continuous responses, have a discrete support and often show skewness when the lower limit is set at zero and the upper tail is not bounded, but can be very long. Generalized linear modelling comes in here. Generalised linear modelling (GLM) is a modelling framework in which the random distribution of the response and the systematic relationship between the expected value of the response and explanatory variables are both specified (Fahrmeir et al., 2021; Touchon, 2021). This is important for GSM call counts since the observations in the data set are not continuous, but rather fall between 0 and 29. Thus discrete probability models are the appropriate models for analysing the response. When the empirical distribution fails to follow the Poisson distribution, however, other count distributions such as the Negative Binomial and the Zero-inflated Poisson and more flexible count distributions are also frequently used, with Poisson regression being a traditional first specification (Green, 2021; Inchausti, 2022). The mean and variance of the Poisson distribution is a single positive parameter, λ. Its probability mass function is:

$$P(Y = y) = \frac{e^{-\lambda}\lambda^{y}}{y!}, \qquad y = 0, 1, 2, \ldots$$

with

$$E(Y) = \lambda$$

And

$$Var(Y) = \lambda.$$

The equidispersion property of the Poisson distribution is the equality of the first two moments. The distribution is a right-skewed count distribution which can be used when the mean is relatively small, and its structure is parsimonious, with a single parameter. However, if there is empirical heterogeneity, then the variance is greater than the mean and its simplicity is a disadvantage. Overdispersed counts have thus been the focus of recent research that has proposed and tested

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

alternatives to the strict Poisson mean-variance relationship (Altun, 2019; Cahoy et al., 2021). The central problem in the present application is not whether the response is a count (which is known to be the case), but whether the count-generating process is sufficiently close to the Poisson distribution so as to warrant the equidispersion restriction.

Poisson Regression

The Poisson regression is an extension of a Poisson probability model where the mean of this model can vary with the observed explanatory variables. Suppose that the value of the variable Y_i represents the number of calls made by respondent I and X_i is a vector of the respondent's characteristics. The conditional distribution is given as

$$Y_i|X_i \sim Poisson(\mu_i),$$

where

$$\mu_i = E(Y_i|X_i).$$

The standard log-link specification is

$$\log(\mu_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_7 X_{7i}.$$

Equivalently,

$$\mu_i = \exp(X_i'\beta).$$

Logarithmic link ensures fitted means are positive and transforms the multiplicative association between covariate(s) and expected counts to an additive association on the log scale. This is especially helpful to understand the meaning of regression coefficients – when you exponentiate a coefficient, you get an incidence-rate ratio (IRR). Exp(ß$_i$) is the multiplicative change in the expected count, due to a one-unit change in a continuous predictor, with other predictors fixed. In the case of a binary variable it would be the ratio of the expected frequency of the category coded as 1 and the reference category (Pardoe, 2021; Inchausti, 2022). The equidispersion assumption holds for the Poisson regression model when the explanatory variables are fixed. The condition in question is thus

$$Var(Y_i|X_i) = E(Y_i|X_i) = \mu_i,$$

instead of comparing the unconditional sample variance with the sample mean. This distinction is crucial, because the observed unconditional overdispersion may be due to the fact that this assumption for the individual observations might be satisfied, but the conditional means are not. On the other hand, it is possible to have a conditionally overdispersed Poisson mean structure.

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

Likelihood-based Estimation

For independent observations, the Poisson log-likelihood is

$$\ell(\beta) = \sum_{i=1}^{n} [y_i \log(\mu_i) - \mu_i - \log(y_i!)],$$

where $\mu_i=\exp(X_i'\beta)$. Maximum likelihood estimation (MLE) chooses the parameter vector which maximises the function. The method does not depend on the conditional mean relationships alone and uses of the full probability distribution enables the use of likelihood-based criteria like the log likelhood, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for comparing competing specifications. Modern regression literature highlights the fact that likelihood-based estimation offers a unified framework for the estimation of non-Gaussian regression models and their goodness of fit (Fahrmeir et al. 2021; Pardoe 2021). Poisson coefficients interpretation is multiplicative and not additive. For instance, if $\beta_j=0.04$, then $\exp(0.04)=1.041$, suggesting that, on average, the predictor variable is associated with an increase of about 4.1% in the response variable, given the other variables in the model are held fixed.

Negative Binomial Regression (NB)

The Negative Binomial model can be obtained as a mixture of the Poisson and gamma models. The outcome is a Poisson distribution of the unobserved rate λ_i and the rate varies between observations as a gamma distribution. The variance will be increased by integrating this latent heterogeneity, resulting in a Negative Binomial distribution with the same conditional mean μ_i, but with a higher variance.

Equidispersion, overdispersion and its consequences

Overdispersion occurs when the conditional variance exceeds the conditional mean:

$$Var(Y_i|X_i) > \mu_i.$$

A useful empirical indicator is the Pearson dispersion statistic,

$$\hat{\phi}_P = \frac{\sum_i (y_i - \hat{\mu}_i)^2/\hat{\mu}_i}{n-p},$$

Let X be a random variable that follows a Negative Binomial distribution, then

$$Var(Y_i|X_i) = \mu_i + \alpha\mu_i^2,$$

where $\alpha \geq 0$ is the dispersion parameter. If α equals 0, the variance becomes the same as that of the Poisson variance. If $\alpha>0$, the variance will be larger than the mean and the higher the excess variance the larger the square of the predicted count (Pagui et al., 2022; Inchausti, 2022). This NB distribution is particularly convenient for data obtained by behaviour counts since it allows for increasingly large variance as count frequency increases.

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

Zero-inflated and Hurdle models

Some of the models incorporate a second process which can produce zeros, known as zero-inflated models. One of these models that assume that an observation may come from a structural-zero process or a Poisson process that can generate both zero and positive counts is called the Zero-Inflated Poisson (ZIP) model. Thus, the probability obtained is zero and is greater than the regular Poisson distribution in the case of a positive inflation component. The Zero-Inflated Negative Binomial (ZINB) model is a model that combines the two processes and adds overdispersion that is allowed by the Negative Binomial distribution (Cummings & Hardin, 2019; Young et al., 2022).

Hurdle models are different in how they model the problem. They split the outcome-generating mechanism into two distinctly different processes: the former a binary process for deciding whether the outcome equals zero or not; the latter a process of the count of the positive outcomes, which is zero truncated. It's not just a matter of the numbers, but the substance. In a zero-inflated model, some observations are considered to be in a "structural-zero" population, and in a Hurdle model, it is assumed that the transition between zero and positive is a separate process (Feng, 2021). Thus, a dataset with zeros should not be used to inform the choice of one or the other of these models. The researcher needs to determine whether the number of observed zeros and/or the type of zero are not consistent with the standard number of zeros.

There is recent methodological literature suggesting that handling zeros should consider origin of zero (Blasco-Moreno et al., 2019), test dispersion, evaluate zero inflation and then compare model fit over the zero-inflated models (Feng, 2021). This is especially significant in the light of the current data. It is very possible that a student may have a count of zero, as they may not make any calls in the observation period. This is an observation that will occur for any ordinary count process, and is not likely to imply that the structural group is not capable of making calls. In addition, there are only eight zeros in the data set (2.54% of 315 observations). This proportion is not too high. Therefore, there must be more empirical evidence to support the introduction of a separate zero-generating process than just a few zero observations.

Research design and study population

The proposed methodology includes a cross sectional correlational study which has a quantitative, non-experimental design. Information was gathered using participant observation, the target population was students of The Federal Polytechnic Bida, Niger State, Nigeria who use mobile telecommunication. Total sample size used was 315 made up of 45 students from each of the 7 schools in the Polytechnic. The dependent variable Y is the number of calls made in the previous day; Sunday and Monday were not included in the data collection. Removal of those days was done to minimize the weekend effect.

Variables and Model Specification

The explanatory variables were coded as shown: X_1 is 1 for male and 0 for female (female is the reference category). The level of study was observed at four levels: ND 1, ND 2, HND 1, and HND 2.

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

Using ND 1 as the reference, X_2 = 1 for ND2, 0 otherwise; X_3 = 1 for HND1, 0 otherwise and X_4 = 1 for HND2, 0 otherwise. The age of the respondent is represented by X_5 while X_6 represents number of siblings of respondent; Distance from base of respondent to Bida is X_7. The common linear predictor for both principal models is

$$P(Y_i = y_i|X_i) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})y_i!}\left(\frac{1}{1+\alpha\mu_i}\right)^{\alpha^{-1}}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i}.$$

The conditional mean is

$$E(Y_i|X_i) = \mu_i$$

and the conditional variance is

$$Var(Y_i|X_i) = \mu_i + \alpha\mu_i^2.$$

The log-link relationship remains

$$\log(\mu_i) = X_i'\beta.$$

Using the same mean specification for both models allows for the comparison of the assumed count distribution and variance structure, instead of the differences in the explanatory variables.

Estimation and Diagnostics

Maximum likelihood was used for the estimation of the models. The first model fitted was the Poisson model that served as the standard model. The Pearson dispersion statistic and deviance relative to residual degrees of freedom were used to evaluate the model adequacy. Next, the fitted Poisson model was subjected to an overdispersion test using regression as described by Cameron and Trivedi. The auxiliary variable was

$$\eta_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i}.$$

For the Poisson model,

$$Y_i|X_i \sim Poisson(\mu_i), \qquad \log(\mu_i) = \eta_i.$$

For the Negative Binomial model,

$$Y_i|X_i \sim NB(\mu_i, \alpha), \qquad \log(\mu_i) = \eta_i,$$

with

$$Var(Y_i|X_i) = \mu_i + \alpha\mu_i^2.$$

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

Equidispersion would result in a coefficient on fitted mean of 0. If the statistically significant coefficient is positive, it is an indication of overdispersion. This diagnostic procedure is based on a regression-based logic that has been recently applied in the Cameron-Trivedi procedure (Lopes et al., 2020). The aim is to identify cases of random sampling variation from systematic lack of adherence to the Poisson variance function. A Maximum Likelihood (ML) estimation of the Negative Binomial model was then performed. The estimated dispersion parameter, its significance and standard error have been directly explored. For model comparison the maximised log-likelihood, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used. This is in line with the current applications made by Ademola and Sabri (2021) and Verma et al. (2020) where Poisson, Negative Binomial or zero-inflated specifications were compared.

As a sensitivity specification model a Zero-Inflated Poisson (ZIP) with an intercept-only inflation component was estimated. Observed number of zeros was compared to that expected by the fitted Poisson and Negative Binomial models. The idea was to know whether the data had a flat region of zero frequency which could be considered as a separate structural-zero process. Based on the limited number of zero observations, a hurdle model was conceptually considered, but would not be well supported by the data. The analysis thus focuses more on the conventional Poisson and Negative Binomial comparison models.

Descriptive characteristics of the GSM call counts

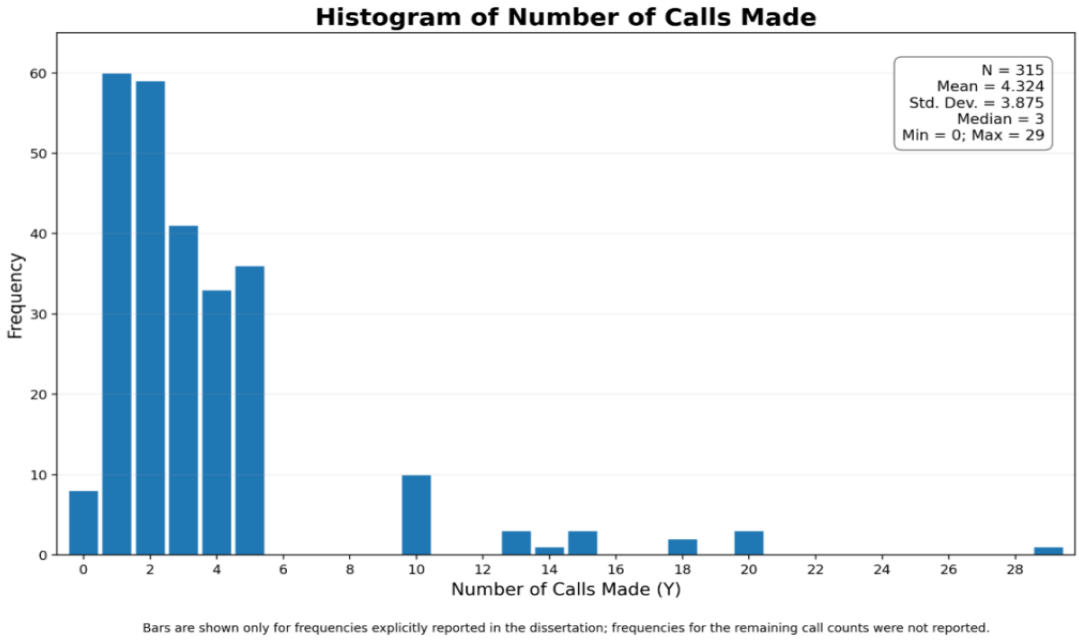


Fig. 1 Histogram of outbound calls

Fig. 1 is the histogram of calls made (outbound) a day prior to day of enumeration. It shows that most respondents made relatively few calls the previous day to date of enumeration, with frequencies concentrated around 1 to 5 calls. The distribution was positively skewed, with fewer

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

respondents making higher numbers of calls. The long right tail, extending to 29 calls, indicates substantial variability and supports the study's finding of overdispersion.

Table 1: Descriptive statistics of the count outcome and continuous predictors

Variable	Mean	SD	Median	Minimum	Maximum
Calls made (Y)	4.324	3.875	3.0	0	29
Age (X_5)	22.241	3.415	22.0	10	45
Siblings (X_6)	4.790	2.771	4.0	0	21
Distance (X_7)	254.552	200.579	190	7	950

Table 1 presents data descriptive statistics. During the observation period the average number of calls placed by the 315 respondents was 4.324, the standard deviation was 3.875 and the median was 3 calls. The lowest number of calls made was 0 and the highest was 29. The variance was 15.016, or over three times the mean. The distribution of the data was strongly skewed to the right (skewness = 2.207) and leptokurtic (excess kurtosis = 7.354). These statistics immediately indicate that the answer is a lot more variable than it would be if the response were truly Poisson distributed with the same variance as its mean. The sample variance: mean ratio was 3.473. It is a descriptive, not a conditional, diagnostic and therefore is inadequate to diagnose Poisson misspecification. However, it does provide a good argument for examining overdispersion, once the available explanatory variables are taken into account.

Table 2: Summary of Model Significance and Goodness-of-Fit tests

Statistic	Poisson	Negative Binomial (NB)	Zero-Inflated Poisson
Number of observations	315	315	315
Log-likelihood	-891.820	-772.562	-889.107
AIC	1,799.640	1,563.124	1,796.213
BIC	1,829.660	1,596.897	1,829.986
Pearson dispersion ratio	3.175	1.180	—
Dispersion parameter (α)	—	0.3831	—
SE of α	—	0.0484	—
Significance of α	—	<0.001	—
Inflation probability	—	—	1.48%
Significant predictors at 5%	ND2, Age	ND2, Age	ND2, Age
Overdispersion test	p < 0.001	—	—

The comparison of performance metrics (see Table 2) shows that the Poisson specification displayed considerable overdispersion, whereas the NB specification substantially reduced the Pearson dispersion ratio. NB model recorded a much higher log-likelihood and lower AIC and BIC than both the Poisson and ZIP specifications. Hence, NB has performed better than the other models. Although the AIC slightly improved with ZIP, it is not necessarily the better model, as it has

no theoretical rationale and had a slightly higher BIC value. As observed by Feng (2021), it is not sufficient to assume that a zero inflated model or a hurdle model is appropriate solely because the data contain zero counts and that one should choose one of the two types over any other.

Table 3: Summary of Hypothesis Tests of parameter significance

Predictor	Poisson	Negative Binomial	ZIP Count Component
Gender (X_1)	Not significant (p = 0.275)	Not significant (p = 0.501)	Not significant (p = 0.275)
ND2 (X_2)	Significant (p < 0.001)	Significant (p = 0.034)	Significant (p < 0.001)
HND1 (X_3)	Not significant (p = 0.936)	Not significant (p = 0.690)	Not significant (p = 0.945)
HND2 (X_4)	Not significant (p = 0.056)	Not significant (p = 0.559)	Not significant (p = 0.175)
Age (X_5)	Significant (p < 0.001)	Significant (p = 0.015)	Significant (p < 0.001)
Siblings (X_6)	Not significant (p = 0.392)	Not significant (p = 0.682)	Not significant (p = 0.309)
Distance (X7)	Not significant (p = 0.726)	Not significant (p = 0.971)	Not significant (p = 0.764)

Interestingly, all models mantain same conclusions on significance status of all parameters (see Table 3).

Table 4: Negative Binomial (NB) Regression (Parameter Estimates)

Parameter	ß	Std. Error	Wald χ^2	Sig.	Exp(ß)
(Intercept)	0.807	0.2980	7.327	0.007	2.240
Gender (X_1)	0.063	0.0932	0.453	0.501	1.065
ND2 (X_2)	-0.246	0.1158	4.503	0.034	0.782
HND1 (X_3)	0.056	0.1404	0.159	0.690	1.058
HND2 (X_4)	-0.076	0.1306	0.342	0.559	0.926
Age (X_5)	0.030	0.0123	5.955	0.015	1.030
Siblings (X_6)	0.007	0.0167	0.168	0.682	1.007
Distance (X_7)	0.00001	0.00023	0.001	0.971	1.000
Dispersion (α)	**0.383**	**0.0484**	**62.556**	**<0.001**	—

Table 4 presents the details of parameters and their IRRs for the preferred Negative binomial model. As applicable to the other two models, only dummy variable for gender, X_2 (ND2: Exp(ß) = 0.782) and age, X_5 (Age: Exp(ß) = 1.030) were significant. Holding the other variables constant, ND2 students have an estimated 21.8% lower expected call frequency than ND1 students. A one-year increase in age is associated with an estimated 3.0% increase in expected call frequency, holding the other predictors constant. The dispersion parameter is highly significant (α = 0.383, SE = 0.0484, p < 0.001). This confirms that the data contain significant extra-Poisson variation. Results report a 95% confidence interval of approximately 0.288 to 0.478, which excludes zero. The dispersion parameter α is not interpreted through Exp(ß) as an incidence-rate ratio, so its Exp(ß) cell is left blank. The NB model allows the variance to exceed the conditional mean through this dispersion parameter.

The model fit statistics are: log-likelihood = −772.562; AIC = 1,563.124 and BIC = 1,596.897.

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

A much better statistical fit was achieved with the use of the Negative Binomial model. The log-likelihood for it was –772.562 while for the Poisson was –891.820. The corresponding AIC declined from 1,799.640 to 1,563.124, a reduction of 236.515 points. BIC declined from 1,829.660 to 1,596.897. The conditional variance of the response variable, Y is related to mean through: $Var(Y_i|X_i) = \mu_i + 0.383\mu_i^2$.

The Pearson dispersion ratio was also decreased from 3.175 to an extent of around 1.180 in the Negative Binomial model. This does not mean that the model captures all the features of the data exactly, but rather it is a significant decrease in the amount of dispersion in the data not accounted for by the model. Empirical studies in recent years also have shown that Negative Binomial regression can be used for materially better fits when the Poisson variance restriction is violated (Ademola & Sabri, 2021; Tiara et al., 2023; Winata, 2023). This argument is now applied to call counts of GSM in a population of Nigerian students.

CONCLUSION

In this study, Poisson and Negative Binomial regression models were investigated for distributional suitability in modelling individual outbound GSM calls by 315 students of The Federal Polytechnic, Bida, Niger State. The call-count distribution is quite overdispersed (It has a mean value of 4.324 and variance of 15.016). The fitted Poisson model has a Pearson dispersion ratio of 3.175 and the regression-based overdispersion test suggests that the model is overdispersed. The results of this study suggest that the Poisson model is not a good model for the observed data because it violates the mean-to-variance relationship. The Negative Binomial model offers a more realistic description of the data. The dispersion parameter's estimated value is positive and statistically significant, and all its AIC, BIC and log-likelihoods exhibit significant improvements when compared to those of Poisson. Similarly, the Pearson dispersion ratio is also around 1.00. The preferred Negative Binomial model suggests that holding the other variables constant, ND2 students have an estimated 21.8% lower expected call frequency than ND1 students; one-year increase in age is associated with an estimated 3.0% increase in expected call frequency, holding the other predictors constant. Other models also found the two variables significant.

Small numbers of zeros (only 8) and the lack of any improvement in the ZIP model over Poisson regression or the drop in model fit compared to NB regression provide little support for the use of the zero-inflated or hurdle modelling. The methodological contribution of this study is thus the proof of an approach to GSM call-count modelling that takes into consideration the distribution sensitivity. The proper approach would be to determine the nature of the count response (or responses, if there were multiple), look at the empirical distribution of the response, test for equidispersion, estimate other possible variance structures if necessary, compare models using likelihood-based methods, and consider zero inflation only if it is justified by the process. In this instance, on this particular data, Negative Binomial regression is the best of the models in contention. The conclusion is based on a distributional evidence and diagnostic analysis of the models compared, not on the assumption that all count outcomes have to be analysed using the

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

Poisson regression model. It is recommended that this research is conducted in other higher institutions and extended to staff.

REFERENCES

Ademola, A. A., & Sabri, S. R. M. (2021). Modelling claim frequency in insurance using count models. *Asian Journal of Probability and Statistics, 14*(4), 14–20. https://doi.org/10.9734/ajpas/2021/v14i430334

Altun, E. (2019). A new model for over-dispersed count data: Poisson quasi-Lindley regression model. *Mathematical Sciences, 13*, 241–247. https://doi.org/10.1007/s40096-019-0293-5

Blasco-Moreno, A., Pérez-Casany, M., Puig, P., Morante, M., & Castells, E. (2019). What does a zero mean? Understanding false, random and structural zeros in ecology. *Methods in Ecology and Evolution, 10*, 949–959. https://doi.org/10.1111/2041-210X.13185

Cahoy, D., Di Nardo, E., & Polito, F. (2021). Flexible models for overdispersed and underdispersed count data. *Statistical Papers, 62*, 2969–2990. https://doi.org/10.1007/s00362-021-01222-7

Campbell, H. (2021). The consequences of checking for zero-inflation and overdispersion in the analysis of count data. *Methods in Ecology and Evolution, 12*(4), 665–680. https://doi.org/10.1111/2041-210X.13559

Cummings, T. H., & Hardin, J. W. (2019). Modeling count data with marginalized zero-inflated distributions. *Stata Journal, 19*(3), 589–616. https://doi.org/10.1177/1536867X19874209

Fahrmeir, L., Kneib, T., Lang, S., & Marx, B. D. (2021). *Regression: Models, methods and applications* (2nd ed.). Springer. https://doi.org/10.1007/978-3-662-63882-8

Fathurahman, M. (2023). Modeling count data with over-dispersion using generalized Poisson regression: A case study of low birth weight in Indonesia. *Jurnal Statistika Universitas Muhammadiyah Semarang, 11*(1).

Feng, C. X. (2021). A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. *Journal of Statistical Distributions and Applications, 8*, Article 8. https://doi.org/10.1186/s40488-021-00121-4

Green, J. A. (2021). Too many zeros and/or highly skewed? A tutorial on modelling health behaviour as count data with Poisson and negative binomial regression. *Health Psychology and Behavioral Medicine, 9*(1), 436–455. https://doi.org/10.1080/21642850.2021.1920416

Inchausti, P. (2022). *Statistical modelling with R: A dual frequentist and Bayesian approach for life scientists*. Oxford University Press. https://doi.org/10.1093/oso/9780192859013.001.0001

International Telecommunication Union. (2024). *Facts and figures 2024*. ITU. https://www.itu.int/

Lopes, L. P. L. F., Belfiore, P., Santos, M. A., & Souza, R. F. (2020). Overdisp: A Stata (and Mata) package for direct detection of overdispersion in Poisson and negative binomial regression models. *Statistics, Optimization & Information Computing, 8*(3), 773–789. https://doi.org/10.19139/soic-2310-5070-557

National Bureau of Statistics (2024). *Telecoms data: Active voice and internet per state, porting and tariff information (Q1 2024)*. National Bureau of Statistics. nigerianstat.gov.ng

Pagui, E. C. K., Salvan, A., & Sartori, N. (2022). Improved estimation in negative binomial regression. *Statistics in Medicine, 41*(13), 2403–2416. https://doi.org/10.1002/sim.9361

Pardoe, I. (2021). *Applied regression modelling* (3rd ed.). Wiley. https://doi.org/10.1002/9781119615941

Tiara, Y., Aidi, M. N., Erfiani, & Rachmawati, R. (2023). Overdispersion handling in Poisson regression model by applying negative binomial regression. *BAREKENG: Journal of Mathematics and Its Applications, 17*(1), 417–426. https://doi.org/10.30598/barekengvol17iss1pp0417-0426

Touchon, J. C. (2021). Generalized linear models. In *Applied statistics with R: A practical guide for the life sciences* (pp. 181–208). Oxford University Press. https://doi.org/10.1093/oso/9780198869979.003.0007

Verma, P., Swain, P. K., Singh, K. K., & Khetan, M. (2020). Count data regression modelling: An application to spontaneous abortion. *Reproductive Health, 17*, Article 106. https://doi.org/10.1186/s12978-020-00955-2

Winata, H. M. (2023). Mengatasi overdispersi dengan regresi binomial negative pada angka kematian ibu di Kota Bandung. *Jurnal Gaussian, 11*(4), 616–622. https://doi.org/10.14710/j.gauss.11.4.616-622

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

Young, D. S., Roemmele, E. S., & Yeh, P. (2022). Zero-inflated modelling part I: Traditional zero-inflated count regression models, their applications, and computational tools. *WIREs Computational Statistics, 14*(1), e1541. https://doi.org/10.1002/wics.1541

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]