Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/isl/120403

Forecasting the Number of Traffic Accidents in Jordan using the Poisson Regression Model

Adeeb Ahmed AL Rahamneh[1,*], *Mohammed Ali Kriem Alzubeidi*[2], *Yaser Ahmad Arabyat*[3] *and Sulieman Ibraheem Shelash Al-Hawary*[4]

[1]Department of Economics, Faculty of Business, Al-Balqa Applied University, Salt, Jordan
[2]Department of Business Administration, Faculty of Business, Al-Balqa Applied University, Salt, Jordan
[3]Department of Economics and Finance, Faculty of Business, Al-Balqa Applied University, Salt, Jordan
[4]Department of Business Administration, School of Business, Al al-Bayt University, Mafraq, Jordan

Received: 9 Jan. 2023, Revised: 2 Feb. 2023, Accepted: 5 Feb. 2023
Published online: 1 Apr. 2023

Abstract: The study aims at forecasting the number of traffic accidents in Jordan for the year 2022, based on the monthly data related to traffic accidents for the period (2017–2021) using the Poisson regression model. SPSS version 26 and Minitab version 19 data analysis programs were used to analyze the collected data. The study concluded that the use of the Poisson regression model is very appropriate to forecast the number of traffic accidents during the next period of time. The Poisson regression method is a useful method for estimating and forecasting. The researchers recommend adoption of this technique in related studies, conducting more extensive studies on the Poisson regression model, and reconsidering the current legislation and the penalties related to traffic accidents.

Keywords: Poisson regression model, traffic accidents, MLE, forecasting, Jordan

1 Introduction

The Poisson distribution is one of the most commonly used probability distributions for modeling enumeration data, and this distribution is one of the important regression models used in studying the effect of one or more independent variables on a dependent variable. The Poisson regression model can be considered a special case of the generalized linear models. in terms of construction machinery.

Many researchers and scholars have been interested in studying this model. In 1998, [1] presented an explanation and clarification of this distribution, and in 2016, [2] presented a scientific paper that dealt with the prediction of football results in Brazil. Through the use of the Poisson regression model, in the same year, [3] presented a scientific paper that dealt with an explanation of the statistical models for counting data. The Poisson regression model was considered the most appropriate model for analyzing enumeration data through the use of the simulation method.

The classical linear regression model assumes that the response variable depends on a set of Explanatory variables, as these variables can be continuous variables or variables it is non-infectious, however, and when the response variable is in the form of non-infectious variables such as the number of traffic accidents. The assumptions of linear regression will not be fulfilled. Therefore, the Poisson regression model was proposed as one of the regression models that are compatible with such cases [4].

The current study has dealt with the Poisson regression model, which is considered one of the most important and most used statistical models among the models that have a countable response variable and is a special case of the generalized linear models [5]. The study aims at forecasting the number of traffic accidents in Jordan for the year 2022 based on the monthly data related to traffic accidents for the period (2017–2021) using the Poisson regression model, where the dependent variable represents the number of traffic accidents and the independent variable represents time (month).

* Corresponding author e-mail: dr. adeeb@bau.edu.jo

2 Poisson's Regression Model

The Poisson regression model is one of the logarithmic linear models, and this name came from taking the natural logarithm in order to convert the formula of the Poisson regression model to the linear one, considering the Poisson regression model a special case of generalized linear models [6]. Resulting from the linear logarithmic relationship between the mean and the linear prediction.

$$g(\lambda) = loglog(\lambda) \tag{1}$$

2.1 The general form of the poisson regression model

We assume that the discrete random variable (Yi) represents a random sample with a Poisson distribution and is conditioned by a vector of explanatory variables (Xi) with a probability distribution function of the variable (Yi) as follows [7]:

$$f(\frac{yi}{xi}) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \tag{2}$$

$$E(\frac{yi}{xi}) = \lambda_i = e^{x_i^t\beta} \tag{3}$$

Such that:
β: Parameters vector of the Poisson Regression.

$$LnLn(\lambda_i) = x_i^t\beta = \beta_0 + \sum_{j=1}^{p}\beta_j x_{ij} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \tag{4}$$

Where: Y: is an $nx1$ vector of observations of the dependent variable, X: is an $nx(p+1)$ known design matrix of independent variables, β: is a $1x(p+1)$ vector of unknown regression coefficients?, n: Sample size, p: Matrix of independent variables.

2.2 The link function

The response variable (yi) takes countable values and considering the Poisson regression such as (Generalized Linear Model) and by finding the equation g (λ) that is constrained to be positive (λ) values so that the linear prediction is any value on the real number line, then the natural selection of a function Poisson regression correlation represents a logarithmic correlation that takes counting numbers as input and converts them to a value on the real number line [8,9].

$$g(\lambda) = LnLn(\lambda) = X\beta \tag{5}$$

$$\lambda = e^{X\beta} \tag{6}$$

2.3 Estimating the parameters of the poisson regression model using (MLE)

The method will be used (MLE) which is considered one of the most important methods of estimating its good characteristics, and the main principle on which this method is based is to find an estimate of the unknown community parameter so that it makes (MLE) at its maximization end, it gives sufficient, unbiased, and least variance estimations assuming that the average (λ). It has a non-linear relationship to the independent variables that takes the form of equation (6), and the relationship between the expected value of the random variable (yi) a linear forecast can be written as [10,11]:

$$\lambda_i = exp(\sum_{j=0}^{p}\beta_j x_j) \tag{7}$$

Where: $y_i \sim Po(\lambda_i), i = 0,1,2,k$

To estimate the parameters of the Poisson regression model, the Maximum Likelihood method will be used for the sample measurements, it is defined as the joint distribution of those measurements, and so, we symbolized the Maximum Likelihood method with the symbol (L) then:

$$L(\beta) = \prod_{i=1}^{k} P(Yi = yi) = \prod_{i=1}^{k} \frac{\lambda_i^{yi\beta - \lambda_i}}{\prod_{i=1}^{k} yi!} \tag{8}$$

Such that: λ_i is a function from $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_p)$

MLE= $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_p)$ are the values of the parameters that maximize maximum likelihood function $L(\beta)$ to simplify the calculation process, by taking the logarithm of the maximum likelihood function $L(\beta)$,*weget* [3,12]:

$$L(\beta) = \sum_{i=1}^{k} loglog(yi!) + \sum_{i=1}^{k} yiloglog(\lambda_i) - \sum_{i=1}^{k} \lambda_i \tag{9}$$

$$\frac{dL}{d\beta} = \sum_{i=1}^{k} xi(yi - e^{x^t\beta}) = Zero \tag{10}$$

Equation 10 does not generate correct and final solutions because it leads to a set of non-linear equations and therefore it must be solved using numerical methods that represent the numerical algorithm to find the regression equations that achieve the Maximum Likelihood method [13,14,15].

Among the most common iterative methods for solving this equation that has been used is the Newton-Raphson method for obtaining the parameter, which takes the following form:

$$\hat{\beta} = \sum_{i=1}^{n} xi(yi - e^{x^t\beta}) = Zero \tag{11}$$

2.4 Model Assumptions [16,17,18]

(1) The conditional probability function of the dependent variable (yi) when the parameter of the (λ) distribution is known follows the Poisson distribution with a parameter of (λ) as shown in equation (1).
(2) The parameter of the distribution in the model is equal to:

$$\lambda_i = e^{x_i^t\beta} \tag{12}$$

Such that: x_i^t: represents row (i) of the matrix of independent variables.
(3) The ordered pairs of the variables (xi,yi) are independent.
Properties of the MLE [19].
Asymptotically unbiased estimators.
It has a distribution approaching the normal distribution.
Asymptotically efficient estimators.
It is possible to express these properties through the following equations [20,21]:

$$\sqrt{n}(\underline{\hat{\beta}}_{ML} - \beta_0) \sim N(O, I(\underline{\beta_0})^{-1}) \tag{13}$$

Where: $\underline{\beta_0}$: original parameters vector, $\underline{\hat{\beta}}_{ML}$: Maximum likelihood estimators vector for model parameters, $I(\underline{\beta_0})$: Fisher Information Matrix.

$$I(: -E\left[H_n(\underline{\beta_0})\right] = E\left[exp(\underline{X}^t\underline{\beta_0})\underline{X}^t\underline{X}\right] \tag{14}$$

The variance of these estimators equals the inverse of the information matrix fisher.
As for the approximate distribution for the maximum likelihood estimators, it can be written as follows [22,23]:

$$\underline{\hat{\beta}}_{ML} \sim app(\underline{\beta_0}, [nI(\beta_0)]^{-1}) \tag{15}$$

Where: $[nI(\beta_0)]^{-1}$: Represents the variance matrix for the estimator's parameters vector, ($\underline{\hat{\beta}}$)
Which means:

$$Var.(\underline{\hat{\beta}}) = [nI(\beta_0)]^{-1} \tag{16}$$

The consistent estimate of the variance matrix is given as follows:

$$V\hat{a}r.(\underline{\hat{\beta}}) = \left[nI(\hat{\beta}_0)\right]^{-1} \quad (17)$$

Information matrix fisher

$$(\hat{\beta}_0) = \frac{1}{n}\sum_{i=1}^{n} exp(\underline{X_i^t\hat{\beta}})\underline{X_i^t X_j} \quad (18)$$

$$\Rightarrow V\hat{a}r.(\underline{\hat{\beta}}) = \left[\sum_{i=1}^{n} exp(\underline{X_i^t\hat{\beta}})\underline{X_i^t X_j}\right]^{-1} \quad (19)$$

3 Application Side

The data of monthly traffic accidents that occurred in the Hashemite Kingdom of Jordan for the period (2017-2021), was relied upon, where the dependent variable represents monthly traffic accidents, and the independent variable represents time (month), The number of traffic accidents for the year 2017 reached a total of (10446), while in the year 2018 it amounted to (10431) accidents, and in the year 2019 the number of traffic accidents reached (10,857) as for the year 2020, the number of traffic accidents reached a total of (8451), and in the year 2021, the number of traffic accidents increased to (11,241) traffic accidents, meaning that the average annual number of traffic accidents for the last five years is (10285) accidents. Table 1 shows the number of monthly traffic accidents for the period (2017-2021).

Table 1: Number of monthly traffic accidents for the period (2017-2021).

2017		2018		2019		2020		2021	
Month	Traffic accidents	Month	Traffic accidents	Month	Traffic accidents	Month	Traffic accidents	Month	Traffic accidents
Jan.	791	Jan.	739	Jan.	744	Jan.	747	Jan.	815
Feb.	708	Feb.	712	Feb.	755	Feb.	782	Feb.	768
Mar.	833	Mar.	921	Mar.	849	Mar.	719	Mar.	764
Apr.	834	Apr.	958	Apr.	943	Apr.	866	Apr.	844
May	950	May	889	May	1003	May	784	May	1068
Jun.	939	Jun.	941	Jun.	952	Jun.	814	Jun.	970
Jul.	958	Jul.	947	Jul.	1079	Jul.	956	Jul.	1091
Aug.	982	Aug.	988	Aug.	931	Aug.	876	Aug.	1039
Sep.	905	Sep.	899	Sep.	859	Sep.	842	Sep.	1048
Oct.	966	Oct.	878	Oct.	919	Oct.	764	Oct.	981
Nov.	842	Nov.	759	Nov.	967	Nov.	589	Nov.	985
Dec.	738	Dec.	800	Dec.	856	Dec.	712	Dec.	868
Total	**10446**	**Total**	**10431**	**Total**	**10857**	**Total**	**8451**	**Total**	**11241**

Ref.

–Traffic accidents in Jordan (2017), Jordan traffic institute, p25
–Traffic accidents in Jordan (2018), Jordan traffic institute, p27
–Traffic accidents in Jordan (2019), Jordan traffic institute, p25
–Traffic accidents in Jordan (2020), Jordan traffic institute, p25
–Traffic accidents in Jordan (2021), Jordan traffic institute, p24

4 Estimation of The Poisson Regression Equation Using The MLE

The equation was The equation was estimated using Minitab version 19, they are as follows:
$\hat{yi} = e^{6.74927+0.000769(month)+\varepsilon_i}$
In order to know the significance of the estimated model, the statistic of the greatest possibility ratio referred to in the equation was calculated.

It was equal to (9.30), while the probability p-value amounted to (0.002), which confirms the significance of the model. Table (2) shows the expected number of traffic accidents for the year 2022,the number of actual traffic accidents for the first seven months of 2022, and the difference between the number of actual and expected traffic accidents for those months.

Table 2: Expected number of traffic accidents for the year (2022

2022			
Month	The expected number of traffic accidents	Actual traffic accidents for the first seven months	Difference between actual and expected traffic accidents
Jan.	894	881	-13
Feb.	895	897	2
Mar.	896	887	-9
Apr.	896	878	-18
May	897	901	4
Jun.	898	867	-30
Jul.	899	902	3

5 Conclusions

By forecasting the estimated Poisson regression model based on the maximum likelihood method, it was found that the average number of traffic accidents expected for 2022 (10779) accidents, and this average exceeds the average number of traffic accidents for the last five years (2017-2021) by only (494) accidents. The Poisson regression method is a useful method for estimating and forecasting.

6 Recommendations

The researcher recommends the following:

1. Adopting the Poisson regression method as an appropriate technique in analyzing data related to traffic accidents.
2. Conducting more extensive studies on the Poisson regression model.
3. Reconsidering the legislation in force and the penalties related to traffic accidents.

References

[1] A. Cameron and P. Trivedi. *Regression Analysis of count Data*. Cambridge University Press, Cambridge UK (1998).
[2] E, Saraiva, A. Suzuki, A. FilhoCiro and F. Louzada. predicting football scores via Poisson regression model: applications to the National Football League. *Communications for Statistical Applications and Methods*,**23**, 297-319 (2016).
[3] S. Miaou, P. Hu, T. Wright, A. Rathi and S. Davis. Relationship between truck accidents and highway geometric design: a Poisson regression approach. *Transportation Research Record*, **1376**, 1-6 (1992).
[4] Z. Al-Jammal and G. Abdullah. Variable selection in Poisson regression model using invasive weed optimization algorithm. *Iraqi Journal of Statistical Sciences*,**16**, 39-54 (2019).
[5] S. Hossain and E. Ahmed. Shrinkage and penalty estimators of a Poisson regression model. *Australian & New Zealand Journal of Statistics*,**54**, 359–373 (2012).
[6] E. Avcı, S. Altürk and E. Soylu. Comparison count regression models for overdispersed alga data. *International Journal of Research and Reviews in Applied Sciences*,**25**, 1-5 (2015).
[7] M. David and D. Jemna. Modeling the frequency of auto insurance claims by means of poisson and negative binomial models. *Analele Ştiinţifice ale Universităţii "Alexandru Ioan Cuza" din Iaşi. Ştiinţe economice*,**62**, 151-168 (2015).
[8] P. Nilsson and S. Nilsson. *Application of Poisson Regression on Traffic Safety*. M.A. thesis, KTH Royal Institute of Technology, Sweden (2015).
[9] A. Arshad, M. Azam, M. Aslam, C. Jun. A EWMA Control Chart based on Repetitive Sampling to Monitor Process Mean with Geometric Poisson Characteristics. *Industrial Engineering & Management Systems*,**16**, 186-194 (2017).
[10] R. Tajuddin, N. Ismail and K. Ibrahim. Comparison of estimation methods for one-inflated positive Poisson distribution. *Journal of Taibah University for Science*,**15**, 869-881 (2021).
[11] R. Yeh and Y. Lin. Analysis of Revenue-Sharing Contracts for Service Facilities. *Industrial Engineering and Management Systems*,**8**, 221-227 (2009).
[12] J. Spinelli, R. Lockhart and M. Stephens. Tests for the response distribution in a Poisson regression model. *Journal of statistical planning and inference*,**108**, 137-154 (2002).
[13] A. Nicholson. Understanding the stochastic nature of accident occurrence. *Australian road research*,**21**, 30-39 (1991).
[14] D. Smith and M. Faddy. Mean and variance modeling of under-and overdispersed count data. *Journal of Statistical Software*,**69**, 1-23 (2016).
[15] A. Alinezhad, A. Mahmoudi and V. Hajipour. Multi-Objective Soft Computing-Based Approaches to Optimize Inventory-Queuing-Pricing Problem under Fuzzy Considerations. *Industrial Engineering and Management Systems*,**15**, 354-363 (2016).
[16] A. Al–Nasir and D. Rashid. *Statistical Inference*. Higher Education Printing Press, Baghdad University, Iraq (1988).

[17] W. Chendi. Modified Poisson estimators for grouped and right-censored counts. *Communications in Statistics - Theory and Methods*,**51**, 1588-1604 (2022).
[18] J. Subramani and S. Balamurali. A Modified Single Sampling Plan for the Inspection of Attribute Quality Characteristics. *Industrial Engineering & Management Systems*,**15**, 41-48 (2016).
[19] A. Chau, E. Lo, M. Wong and C. Chu. Interpreting Poisson Regression Models in Dental Caries Studies. *Caries Research*,**52**, 339–345 (2018).
[20] D. Koletsi and N. Pandis. Poisson regression. *American journal of orthodontics and dentofacial orthopedics*,**152**, 284-285 (2017).
[21] L. Ma and K. Goulias. Application of Poisson regression models to activity frequency analysis and prediction. *Transportation Research Record*,**1676**, 86-94 (1999).
[22] G. Rodriguez. *Generalized Linear Models*. University Princeton, And Revised September (2007).
[23] R. Yeh and Y. Lin. Analysis of Revenue-Sharing Contracts for Service Facilities. *Industrial Engineering and Management Systems*,**8**, 221-227 (2009).