Choosing the best estimated regression equation for data subject to geometric distribution (Student data as a case study)

Osanma Abdulazeez kadhim Al-Quraishi

Ministry of Education, Iraq

E-mail: ousamastat@gmail.com

Abstract. The geometric distribution is one of the discrete statistical distributions that is important, especially in the of population statistics, such as studying population growth and death and birth rates. The discrepancy in the quality of the estimated regression models and the inability to use some models of them because they do not possess the characteristics of good estimates, lead to a lack of confidence in their predictive or estimated accuracy, which necessitated the study of the geometric distribution to which the data are subject and the estimation of the distribution regression equation using some of the parameter estimation methods (kernel regression method, Liu method) for data represented by preparing secondary students in the holy Karbala governorate with 50 views, and after the comparison between the results of estimating the two methods using the Standards Comparison (AIC, BIC, MSE) and determining the optimal way to estimate the geometric distribution equation. The kernel regression is the best because it has the smallest values of the Standards Comparison and was used for the purpose of predicting the preparation of secondary students for the time period (2021-2030).

1. Introduction and research methodology

There are many mathematical models, the most important of which are regression equations, especially the transformed equations from statistical distributions that are used to predict. The study of choosing the appropriate distribution for a phenomenon and methods for estimating the regression models converted by logarithmic transformation began in the last century. In 1986, the two researchers (Cameron & Trived) [7] used the Poisson regression model, estimated its parameters and news of those parameters, and applied it in modeling numerical data that represent the number of medical consultations provided. And in (2005) the researcher (CHANG) [8] in a comparative study between the negative binomial regression method and the artificial neural network method (ANN) applied it to road accident data in the state of New York, USA, for the purpose of analyzing these data and predicting with them. Negative binomial regression in prediction as the relationship between variables does not affect the validity of the model. In (2007) the two researchers (WANG & ETIENNE) [14] used the Poisson regression model and the negative binomial model to predict road accidents. In 2019, the two researchers (Shorouk & Adnan) [4] assessed the parameters of the binomial regression model to study the factors affecting the increase in congenital anomalies in Iraq. Also in the same year, two researchers (Abd al-Hussain and Abd al-Amir) [1] used the Poisson regression model estimated in the manner of the greatest possible way to predict road accidents see. So the researcher decided to use the geometric regression model for
the purpose of predicting the preparation of secondary students in the holy governorate of Karbala after
the data are subject to geometric distribution, using non-parametric methods to estimate the regression
equation because these methods deal with high flexibility with random error assumptions and this does
not affect the efficiency of the capabilities, as they are mainly based on the data, the data type explains
the actual shape of the regression curve.

1.1 research problem
The variance in the quality of the estimated regression models and the inability to use some models of
them because they do not have the characteristics of good estimates, which leads to lack of confidence
in their predictive or estimated accuracy.

1.2 The research objective
The research aims to:
1- Estimation of the geometric distribution slope equation using some non-parameter estimation methods
(kernel regression method, leo method).
   2- Differentiating between the results of estimating the two methods using the Standards Comparison
      (AIC, BIC, MSE) and determining the best method for estimating the geometric distribution equation.
   3- Predicting Prepare of secondary students in the holy city of Karbala for the period of time (2021-2030).

1-3: Research Limits:
The limits of the research were temporal, the preparation of secondary students for the period of time
(1969-2019), spatially the number of secondary students for the holy Karbala Governorate.
1-5 Data sources: The data was taken from the Ministry of Planning / the Central Statistical
Organization/the statistical group.

2- Theoretical aspect:
2-1: Geometric Distribution: [3]
This distribution belongs to the family of discrete distributions and distribution is important in statistical
applications, especially in the population statistics when studying population growth rates and mortality
and birth rates, and the geometric distribution is part of the probability distribution related to Bernoulli
experiments, as we can know the probability density function of the geometric distribution is the number
of attempts failed to try until the first success of that experiment. If the variable (Y) indicates the number
of times the experiment was repeated, (θ) indicates the probability of success of the experiment, (1 - θ)
the probability of the experiment failing, then the probabilistic function of this distribution will be:

\[ P(Y, \theta) = \theta (1 - \theta)^{Y}, \quad 0 < \theta < 1, \quad Y = 0, 1, 2, \ldots \]  

(1)

2-2 Some characteristics of the Geometric distribution:[3]
1- mean of the variable Y is:
\[ M(Y) = EY = \frac{\theta}{(1 - \theta)} \quad \ldots (2) \]

2- mode of the variable Y is:
\[ M_0(Y) = 0 \quad \ldots (3) \]

3- median of the variable Y is:
\[ M_e(Y) = \left[ \frac{-1}{\log_2(1 - \theta)} \right] - 1 \quad \ldots (4) \]

4- variance of the variable Y is:
\[ V(Y) = \frac{\theta}{(1 - \theta)^2} \quad \ldots (5) \]

5- Skewness of the variable Y is:
6- Kurtosis of the variable $Y$ is:
$$S_k(Y) = \frac{2 - \theta}{\sqrt{1 - \theta}} \quad \ldots (6)$$

7- Cumulative function of the variable $Y$ is:
$$F(Y) = 1 - \theta^Y, \quad Y = 0, 1, 2, \ldots n \quad (8)$$

8-Moment generating functions of the variable $Y$ is:
$$M_Y^{(t)} = \frac{\theta}{1 - \theta e^t} \quad \ldots (9)$$

2.3 Regression Analysis: [2]
Regression analysis is one of the methods or statistical tools that are most used in most research in general because it describes the relationship between variables in the form of an equation, and it is defined as a Statistical Tool used to know the relationship between one or more independent variables and dependent variable.
Regression analysis is important because it is used for several important purposes:
1- Data Descriotion: Find the regression equation that summarizes or describes the data available to the researcher.
2- Estimation of Parameters: Estimation of the regression equation parameters used in the description of data to infer the strength and direction of the relationship between the variables.
3- Prediction: Estimating and predicting the values of the dependent variable in the future to benefit planning and decision making.
4- Control: Control of the results of the dependent variable when changing the values of the independent variables. Geometric Regression Mode [13],[2] The geometric regression model is one of the non-linear regression models, and it is converted into one of the Log-Linear Models by taking the natural logarithm of the geometric distribution equation, so it turns into a linear formula. Geometric regression can also be defined as the method by which the dependent variable is modeled when the values of the variable are as Count Data or as Rate Data. The geometric regression equation can be obtained from the geometric distribution as follows:
Let the variable $(Y)$ follow the geometric distribution with the parameter $(\theta)$, and that $(X)$ represents the explanatory variable (time) we can write the geometric distribution equation in the following form:
$$Y = \theta(1 - \theta)^x \quad \ldots (10)$$
Taking the logarithm of both sides of the equation, we get a first degree linear equation:
$$\ln(Y) = \ln(\theta) + X\ln(1 - \theta)$$
Assuming below, we obtain an equation (13) representing the estimated linear regression model for the geometric distribution as follows:
$$\ln(\theta) = b_0 \quad \ln(1 - \theta) = b_1 \quad \ln(y^*) = y^* \quad \ldots (11)$$
In the case of taking the exponential function of both sides of equation (13), we obtain the non-linear discretionary regression model for the geometric distribution as follows:
$$y = e^{b_0 + b_1x_i + e_i} \quad \ldots (12)$$

2.4 Hypothesis of linear regression of random error: [11]
1- Random error $(u_i)$ is normal distributed with an average of zero $(E(u_i) = 0)$ and constant variance equal $(V(u_i) = \sigma^2)$, meaning that:
$$u_i \sim N(0, \sigma^2) \quad \ldots (13)$$
2- Successive errors in the simple linear regression model should not be related to each other, meaning that:
\[ \text{Cov} (u_i, u_j) = 0 \] ...(14)

3- The random variable (ui) and the explanatory variable (xi) are independent of each other, meaning that:
\[ \text{Cov} (x_i, u_i) = 0 \] ...(15)

Methods for Estimating the Geometric Regression Model: 2-7

There are many methods for estimating the geometric regression model. In this paragraph, we will address the non-parametric methods that assume that random error is distributed with an average of (0) and a constant variance of \((\sigma^2)\), the fact that non-parametric methods deal with high flexibility with random error assumptions and this does not affect the efficiency of the estimates, since it is mainly based on data and the data type explains the actual shape of the regression curve, and the methods are:

1. Kernel regression method: [9], [10]

The kernel regression method is one of the non-parametric methods that can be used to estimate the regression equation for the geometric distribution, and the kernel regression method was proposed by the scientists Nadaraya and Watson in (1964) by means of the log of the maximum function, it can be rewritten as follows:

\[ \text{LE}(Y_i) = L(\theta_i) = \frac{1}{n} \sum_{i=1}^{n} \left\{ e^{\eta(x_i)} - \eta(x_i) \right\} \] ...(16)

As:

Model Parameters Vector \( \eta(x_i) \):
\[ \eta(x_i) = \ln[\theta(x_i)] \]

The kernel method is summarized in estimate. Using the following linear model:
\[ \eta(x_i) = b_0 + w^T \phi(x_i) \] ...(19)

As:

\[ w^T \phi(x_i) = K(x_i, x_i) \]

\( k(x_i, x_l) \): is one of the kernel functions.

We get the vector of the geometric regression equation parameter by the kernel method \( \eta(x_i) \) by minimizing the logarithm of the Penalized Negative Log-Likelihood.

\[ L(w, b) = \sum_{i=1}^{n} \left[ e^{b_0 + w^T \phi(x_i)} - y_i (b_0 + w^T \phi(x_i)) \right] + \frac{\lambda}{2} \|w\|^2 \] ...(17)

As:

\( \lambda \): Penalty Parameter and its function is to control the balance between Goodness of Fit based on two main factors: data and smoothness \((\text{Smoothness} ||w||^2)\) for the function \( \eta(x_i) \), which The logarithm of the function of the negative penalties is made at the lower end of it, and on it the Vector parameters of the model \( \eta(x_i) \) is written according to the following form:
\[ \eta(x_i) = b_0 + k_i \alpha \] ...(18)

As:

\( \alpha \): fixed border.

\( k_i \): column \( i \) in the kernel matrix of degree \((n * n)\) and its elements \( k(x_i, x_l) \) and \((i, l) = 1, 2, 3, \ldots n \)

\( \alpha \): smoothing parameters vector with degree \((n * 1)\).

Thus, the formula for the logarithm of the negative penalty maximum function is written as follows:

\[ L(b_0, \alpha) = \sum_{i=1}^{n} \left[ e^{b_0 + k_i \alpha} - y_i (b_0 + k_i \alpha) \right] + \frac{\lambda}{2} \|\alpha\|^2 \] ...(19)

To obtain maximum likelihood estimates of the nonlinear regression equations in which the dependent variable \( (Y) \) follows the geometric distribution, estimates can be obtained using numerical estimation methods such as the Newton-Rafson method and the least squares method for multi-stage repetitive weights (Iterative Reweighted Least Squares). The second method (irls) is easier than the first method (Newton - Rafson) in deriving the cross-validation Criterion (cvc), which has a rate of convergence lower than the rate of convergence in the method of Newton - Rafson.
In order to obtain maximum estimates for the unknown parameters \((b_0, \alpha)\), we first find the partial derivatives of the logarithm of the Negative penalty maximum function with respect to the parameter \(\alpha\) and the parameter \((b_0)\), and then equate the derivative to zero Secondly, my agencies:

\[
\begin{align*}
\frac{\partial \log U(O.C.V)}{\partial b_0} &= 0 \quad \text{(20)} \\
\frac{\partial \log U(O.C.V)}{\partial \alpha} &= 0_{n}\quad \text{(20)}
\end{align*}
\]

Solve the following formula using the least squares method for the multi-stage repeated weights (i.r.l.s): \(V'WV + U) \beta = V'Wy \quad \text{(21)}\)

As:

\(\eta_i\) : value of the dependent variable converted to linear formula \(\ln(y : y^*)\)

\(V: \) matrix with degree \((n + 1) \times n \) defined as follows:

\[
V = (1_n, k) \quad \text{(22)}
\]

As:

In: the correct one

\(\beta: \) The kernel matrix contained in the formula \((21)\)

\(W: \) Diagonal matrix with degree \((n * n)\) written as follows:

\[
W = \text{diag} \left \{ \exp(b_0 + k_i \alpha), \ldots, \exp(b_0 + k_n \alpha) \right \} \quad \text{(23)}
\]

\(B: \) The vector parameters to be Estimated are of the same degree \((n+1) \times 1\)

\[
B = (\hat{b}_0, \alpha^T) \quad \text{(23)}
\]

\(U : \) Square matrix of degree \((n + 1) \times (n + 1)\) is written as follows

\[
U = \begin{pmatrix} 0 & 0 \\ 0 & \lambda_k \end{pmatrix} \quad \text{(24)}
\]

As:

Zero vector with degree \((n \times 1)\) \(\Theta_n\)

\(y': \) the vector of the response variable in the kernel regression method \((n \times 1)\) and that \(y_i\) is written as:

\[
y_i = \frac{y_i - \exp(b_0 + k_i \alpha)}{\exp(b_0 + k_i \alpha)} + \exp(b_0 + k_i \alpha) \quad \text{(24)}
\]

The estimation of the parameter vector \((b_0, \alpha)\) when using the Newton-Ravson method, which determines the values of the estimates in the various stages of the method, is:

\[
\left( \begin{array}{c} a^{(t+1)} \\ k^{(t+1)} \end{array} \right) = \left( \begin{array}{c} a^{(t)} \\ k^{(t)} \end{array} \right) + \left( \begin{array}{cc} k w^{(t)} & k + k_e \\ k \mu(i) \mu(i) \end{array} \right)^{-1} \left( \begin{array}{c} k \mu(i) - y + \lambda a^{(t)} \\ 1 - (\mu(i)^2 - y) \end{array} \right) \quad \text{(25)}
\]

That the estimates of the mean function at the input vector \(x_0\) and for the estimated optimum parameter values \((b_0, \alpha)\) are:

\[
\hat{\mu}(x_0) = \exp(\hat{b}_0 + k_0 \alpha) \quad \text{(26)}
\]

As:

\[
k_0 = (k(x_1, x_0), \ldots, k(x_n, x_0))^T \quad \text{(27)}
\]

In order to test the penalty parameter \((\lambda)\) and the parameter of the endodontic function \((\sigma \gamma^2)\) which results in the best preamble, we use for this test forensic (O.C.V) ordinary cross validation.

\[
\frac{1}{n} \sum_{i=1}^{n} \left( \frac{y_i - e^{\hat{\theta}_i(x_i)}}{1-s_{ii}} \right)^2 \quad \text{(29)} \text{O.C.V}
\]

As:

\(s_{ii}: \) the diagonal element of row \(i\) of matrix \(S\), defined as:

\[
S = V(V'WV + U)^{-1}V'W \quad \text{(28)}
\]

When compensating with the average \(s_{ii}\) which equals \(\frac{1}{n} \text{tr} (s)\) in formula (30) we get \((\text{g.c.v})\) Generalized cross validation which is calculated with respect to \(\theta\) of the parameters my agency:

\[
g.c.v(\theta) = \frac{n}{n-\text{tr}(s)} \sum_{i=1}^{n} \left( y_i - e^{\hat{\theta}_i(x_i)} \right)^2 \quad \text{(29)}
\]

As:

\(\hat{\eta}_i(x_i): \) An estimate of the logarithm \(\mu e\), meaning \([\eta_i(x_i) = \log(\mu_e)(x_i)]\) for each parameter after viewing exclusion \((i)\) according to the( Leave-One-Out method)
2.5 Liu Estimators Method: [12],[5]

It is the second nonparametric method and is used treat of inflation estimated variance parameters of the model. Liu laid the foundations for this method in 1993 when he created it for the linear regression model, then going back to the formula for geometric model given in equation (13):

\[ Y^* = b_0 + b_1 X_1 + u_i \]

Liu (And by adopting the following restriction that he set

\[ C' = (S\tilde{\beta} - \tilde{\beta}^*)' (S\tilde{\beta} - \tilde{\beta}^*) \quad \ldots(30) \]

AS:

\[ C'W_C = (Y^* - X\tilde{\beta}^*)' \tilde{W}(Y^* - X\tilde{\beta}^*) + (S\tilde{\beta} - \tilde{\beta}^*)'(S\tilde{\beta} - \tilde{\beta}^*) \]

C': the amount of increase in the mean of the weighted error squares when the estimated parameter vector is replaced by the maximum likelihood method (\( \tilde{\beta} \)) with the estimated parameter vector according to the Liu method (\( \tilde{\beta}^* \)).

\( \tilde{\beta} \) : estimates maximum likelihood when converting the model to a linear model

\( \tilde{\beta}^* \) : Vector sample parameters estimated by the Liu method

S: bias parameter

After the process of converting the model to linear and placing the constraint, the square of the weighted errors of the researched model is taken with the addition of the constraint and then deriving the result with respect to \( \tilde{\beta}^* \) : as follows:

\[ C'W_C = (Y^* - X\tilde{\beta}^*)' \tilde{W}(Y^* - X\tilde{\beta}^*) + (S\tilde{\beta} - \tilde{\beta}^*)'(S\tilde{\beta} - \tilde{\beta}^*) \]

\[ = Y^* \tilde{W}y - 2\tilde{\beta}^* X^* \tilde{W}y + \beta^* X^* \tilde{W}X\tilde{\beta} + S\tilde{\beta}^* S\tilde{\beta} - 2S\tilde{\beta}^* S\tilde{\beta} + \beta^* \beta^* - 2\tilde{\beta}^* 2\tilde{\beta} + 2\tilde{\beta}^* \quad \ldots(31) \]

By deriving equation (33) with respect to the parameter vector (\( \tilde{\beta}^* \)), we get:

\[ \frac{\partial C'W_C}{\partial \tilde{\beta}^*} = -2X^* \tilde{W}y * + 2X^* \tilde{W}x \tilde{\beta} - 2S\tilde{\beta} + 2\tilde{\beta}^* \quad \ldots(32) \]

When the derivation result is equal to zero, we obtain the Liu estimates for the geometric regression model parameters.

\[ \tilde{\beta}_{liu} = (X'\tilde{W}X + 1)^{-1} (X'\tilde{W}y^* + S\tilde{\beta}) \quad \ldots(33) \]

Since the

\[ \tilde{\beta} = \tilde{\beta}_{ML} \quad \ldots(34) \]

\[ \tilde{\beta} = \tilde{\beta}_{ML} = (X'WX)^{-1} (X'W y^*) \quad \ldots(35) \]

Know in numbered form (32) they are estimates maximum likelihood. When converting a geometric regression model to a linear formula

As:

\[ (X'W y^*) = (X'WX) \tilde{\beta}_{ML} \quad \ldots(36) \]

Substituting the value of \( (X'W y^*) \) to its equivalent in the numbered formula (35) we get another form of the liu of the geometric regression model parameters.

\[ \tilde{\beta}_{liu} = [(X'\tilde{W}X + 1)^{-1} (X'\tilde{W}X) \beta_{ML} + S\beta_{ML}] \]

\[ \tilde{\beta}_{liu} = (X'\tilde{W}X + 1)^{-1} (X'\tilde{W}X + S I) \beta_{ML} \quad \ldots(37) \]

Standards Comparison: [6],[15] 2-8:

1. AIC (Akaike Information Criterion) \( AIC = n \ln \sigma_e^2 + 2M \quad \ldots(38) \)

2. SBC (Schwartz Bayesian criterion) \( SBC = n \ln \sigma_e^2 + \ln(n) \quad \ldots(39) \)

3. \( H - Q(M) = \ln(\sigma^2_e) + 2MC \frac{[n \ln(n)]}{n} ; C > 2 \quad \ldots(40) \)

AS :
\[
\sigma^2 = \frac{\sum_{i=1}^{n} (e_i - \hat{e}_i)^2}{n}
\]

n: represents the number of views.
M: number of parameters of the chosen model or model rank.
C: fixed.

3-The applied side.
3-1 Description and definition of the study sample:

The annual secondary student preparation data in the holy governorate of Karbala was relied on for the period of time ((1969-2019) for the geometric regression equation, and the dependent variable (Y) represented the preparation of secondary students during time and the independent variable( X) time (years).

:Appropriate statistical distribution:3-2
To find out the statistical distribution appropriate to the preparation of secondary students, a goodness of fit test was used for the data through the Easy Fit statistical program. The results showed that the data follows the geometric distribution with parameters \( \theta = 2.4920E - 5 \) ana the value (p-value=0.000).

Form (1) probability density function for the geometric distribution
3-3: Statistical analysis of data:
The data were analyzed statistically using the statistical program (Gretl 1.9.11 ).
First: Estimate the geometric regression equation using the kernel regression method and the Leo agency method:

1-The estimated formula for the geometric regression equation using the kernel regression method is:
\[
\ln y = 9.44278 + 0.0368146 x
\]

2-The estimated formula for the geometric regression equation using the Leo method is:
\[
\ln y = 9.29892 + 0.0429313 x
\]
Second: The test of the significance of the putative linear relationship 1-Linear relationship significance test for the estimated engineering regression equation using the kernel method. To find out the significance of the linear relationship, we test the following hypothesis:

\( H_0: B_1 = 0 \)
\( H_1: B_2 \neq 0 \)

Through the numbered analysis of variance table (2) we obtain the value of the test (F) that determines the nature of the relationship between the dependent variable and the independent variable

| Table (1) analysis of variance |
|-------------------------------|
| **Regression** | **Sum of squares** | **df** | **Mean square** |
| Regression     | 14.1122            | 1      | 14.1122         |
| Residual       | 0.27484            | 48     | 0.00572583      |
| Total          | 14.387             | 49     | 0.293613        |
Looking at Table (1), we notice that the test value is \((F = 2464.65)\) and its probability value [p-value = 6.45e-043] It is less than the level of significance (0.05), we accept the alternative hypothesis which states that the time variable exerts its effect on the student numbers variable and the percentage (98%).

2- Test the significance of linear relationship of the estimated geometric regression equation using the Liu method.

To find out the significance of the linear relationship, we test the following hypothesis:

\[ H_0: B_1 = 0 \]
\[ H_1: B_2 \neq 0 \]

Through the numbered analysis of variance table (3) we obtain the value of the test (F) that determines the nature of the relationship between the dependent variable and the independent variable.

### Table (2) analysis of variance

|            | Sum of squares | df  | Mean square |
|------------|----------------|-----|-------------|
| Regression | 19.1912        | 1   | 19.1912     |
| Residual   | 0.835683       | 48  | 0.0174101   |
| Total      | 20.0269        | 49  | 0.408712    |

\[ R^2 = 19.1912 / 20.0269 = 0.958272 \]

\[ F(1, 48) = 19.1912 / 0.0174101 = 1102.31 \] [p-value 9.08e-035]

Looking at Table (2), we notice that the test value is \((F = 1102.31)\) and its probability value [p-value = 9.08e-035] It is less than the level of significance (0.05), we accept the alternative hypothesis which states that the time variable exerts its effect on the variable of student numbers and the percentage (95%).

3-4 Standards Comparison:

After the geometric regression models have been estimated and the linear relationship is significant, the differentiation criteria are used to choose the method that best achieves the agency model:

Table (3) Standards Comparison

|            | AIC    | BIC    | H-Q    |
|------------|--------|--------|--------|
| Liu Method | -58.68260 | -54.85855 | -57.22638 |
| Kernel Regresion Method | -114.2857 | -110.4616 | -112.8294 |

By looking at Table (3), we notice that the kernel regression method is preferable because it has the smallest differentiation criteria, compared to the Leo method.

3-5: Prediction:

The number of secondary students can be predicted for the holy city of Karbala for the future time period (2021-2030) using the estimated geometric regression model according to the Kernel method based on time series data, Table (1).

Table (4) Predictive Values for Preparing Secondary Stage Students for the Duration (2030-2021)

| Years | 2021 | 2022 | 2023 | 2024 | 2025 |
|-------|------|------|------|------|------|
| Predictive Values | 82481 | 85574 | 88783 | 92112 | 95567 |
| Years | 2026 | 2027 | 2028 | 2029 | 2030 |
| Predictive Values | 99150 | 102863 | 106726 | 110729 | 114881 |
4. Result and recommendations:
1- The secondary student preparation data has appeared subject to the geometric distribution.
2- The significance of the estimated regression equation using the kernel regression method.
3- Significance of the estimated regression equation using the Liu method.
4- The preference of the kernel regression method is preferable to the Leo method, since the first method has the lowest Standards Comparison.
5- Depending on the values predicted by the geometric regression model, it was noted that there has been a decrease in the number of students from the last sex previous years.
6- Using other methods to estimate the geometric regression equation to arrive at the best prediction model.
7- Making use of the research findings and adopting them to prepare future plans from the relevant authorities.
8- Studying the decrease in the number of future students.

References
[1] Abdul-Hussein H and Abdul-Amir T 2019 Predicting the number of road accidents in Dhi Qar governorate for the year 2022 using the Poisson model Karbala University Scientific Journal 17(1)
[2] Zahra H et al. 2014 regression analysis. Directorate of Dar Al-Kutub for Printing and Publishing - Basra University / College of Administration and Economics
[3] Amir H 1990 Mathematical Statistics. Directorate of Dar Al Kutub for Printing and Publishing - University of Mosul / College of Administration and Economics
[4] Shorouk A and Adnan F 2019 A comparison of the two least squares and weighted least squares methods to estimate a negative binomial regression model. Scientific Journal of the University of Karbala 17 (1) Scientific
[5] Stephen W R 1988 Educational Applications of Hierarchical Linear Models: A Review Journal of Educational Statistics 13 85-116
[6] Akaike H 1973 Information theory and extension of the maximum likelihood principle In : B.N. petrov and F Csaki eds 2 nd International symposium on Information. Theory Academia Kiado Budapest 267-281
[7] Cameron A C and Trived P R 1986 Regression Analysis of count Data Cambridge University press 1-434
[8] Chang L 2005 Analysis of freeway Accident Frequencies : Negative binomial regression versus artificial neural network . Safety science.
[9] Frome E L 1983 The Analysis of Rates Using Poisson Regression Models Biometrics 39 665-674
[10] Hwang G and Shim J 2008 Semiparametric Kernel Poisson Regression for longitudinal Count Data Communication of the Korean Statistical Society 15 (6) 1003-1011
[11] John O1998 Applied regression analysis Aresearch Tool North Carolina State University, USA
[12] Long J S 1997 Regression Models for Categorical and Limited Independent Variables SAGE Publican Inc, USA
[13] Robert S P and Daniel I R 2000 Econometric models and economic forecasts New York McGraw-Hill Book company
[14] Wang P D and Etienne E K 2007 Why Fuzzy Logic A Spectrum of theoretical and pragmatics Issues studies in fuzziness and soft Computing. Springer Berlin Heidelberg
[15] Schwarz G 1978 EstimatingThe dimension of a model Annuals of Statistics 6 461-464.