

APPLICATION OF LINEAR REGRESSION MODEL FOR POISSON DISTRIBUTION IN FORECASTING

Sulaimon Mutiu O.

*Department of Statistics & Mathematics
Moshood Abiola Polytechnic, Abeokuta, Ogun State, Nigeria.*

Abstract - *The underlying idea of modeling relationship of two variables with linear regression involve situation in which there is one independent variable. The problem of this research work is modeling the expected arrival rate of bank customers (whose observed arrival rate assumed to have followed a Poisson distribution) using simple linear regression model. For the successful execution of this research work, primary data on customers' arrival rate from First Bank Nigeria Plc located at Panseke Area of Onikolobo, Abeokuta, Ogun State between 9:00a.m and 2:00p.m in the interval of 5 minutes was employed. Data collected were analyzed electronically using Ms Excel 2007 and SPSS 21.0. Results from the analysis reveal that approximately 31 customers are expected to arrive the banking hall of First Bank Nigeria Plc, Panseke Branch of Abeokuta, Ogun State every 5 minutes. The observed arrival rate has a positive impact on the expected arrival rate and this impact is significant.*

Keywords: *Arrival Rate, Bank, Probability Distribution, Variables*

1. INTRODUCTION

Modeling refers to the development of mathematical expressions that describe in some sense the behavior of a random variable of interest. This variable may be the price of wheat in the world market, the number of deaths from lung cancer, the rate of growth of a particular type of tumor, or the tensile strength of metal wire (John Neter, 2004). In all cases, this variable is called the dependent variable and denoted with Y . A subscript on Y identifies the particular unit from which the observation was taken, the time at which the price was recorded, the county in which the deaths were recorded, the experimental unit on which the tumor growth was recorded, and so forth. Most commonly the modeling is aimed at describing how the mean of the dependent variable $E(Y)$ changes with changing conditions; the variance of the dependent variable is assumed to be unaffected by the changing conditions (John Neter, 2004).

Other variables which are thought to provide information on the behavior of the dependent variable are incorporated into the model as predictor or explanatory variables. These variables are called the independent variables and are denoted by X with subscripts as needed to identify different independent variables. Additional subscripts denote the observational unit from which the data were taken. The X s are

assumed to be known constants. In addition to the X s, all models involve unknown constants, called parameters, which control the behavior of the model. These parameters are denoted by Greek letters and are to be estimated from the data (Christopher J. Nachtsheim, 2004).

The mathematical complexity of the model and the degree to which it is a realistic model depend on how much is known about the process being studied and on the purpose of the modeling exercise. In preliminary studies of a process or in cases where prediction is the primary objective, the models usually fall into the class of models that are linear in the parameters. That is, the parameters enter the model as simple coefficients on the independent variables or functions of the independent variables. Such models are referred to loosely as linear models (Christopher J. Nachtsheim, 2004).

Regressions are a statistical tool used in addressing variety of research hypotheses. It has the potential of predicting a particular outcome. It provides information about the set of data and the contribution of each variable in the analysis. It is usually used as a control tool when exploring the prediction of a model (Tabachnick and Fidell, 1996).

Regression modeling is the process of constructing forecasting models based on the relationship between a dependent variable and independent variables to make the future forecast. Regression modeling is a kind of multifactor forecasting. The basis of regression modeling is the construction of regression models (Tetyana Kuzhda, 2012).

Regression models are used to predict one variable from one or more other variables. Regression models provide the scientist with a powerful tool, allowing predictions about future events to be made with information about past or present events. In order to construct a regression model, both the information which is going to be used to make the prediction and the information which is to be predicted must be obtained from a sample of objects or individuals. The relationship between the two pieces of information is then modeled with a linear transformation. Then in the future, only the first information is necessary, and the regression model is used to transform this information into the predicted. In other words, it is necessary to have information on both variables before the model can be constructed. Regression models are one of the most famous examples

of economic and statistical models used in the forecasting of socio-economic processes (Tetyana Kuzhda, 2012).

Linear regression is an approach to modeling the relationship between two or more independent variables (X) and a single dependent variable (Y). The case of one explanatory variable is called simple regression model. More than one explanatory variable is multiple regression models (Robert Mills, 2011).

Linear regressions are designed to measure one specific type of relationship between variables: those that take linear form. The theoretical assumption is that for every one-unit change in the independent variable, there will be a consistent and uniform change in the dependent variable. Perhaps one reason why linear regression is so popular is that this is a fairly easy way to conceive of social behavior – if more of one thing is added, the other thing will increase or decrease proportionately. Many relationships do operate this way (Michael H. Kutner, 2004).

2. STATEMENT OF THE PROBLEM

The underlying idea of modeling relationship of two variables with linear regression involve situation in which there is one independent variable. The problem of this research work is modeling the expected arrival rate of bank customers (whose observed arrival rate assumed to have followed a Poisson distribution) using simple linear regression model.

3. AIM AND OBJECTIVES OF THE STUDY

The aim of this study is to model the expected arrival rate of bank customers using simple linear regression model.

The objectives are:

- (i) To determine the strength of the relationship between the observed and expected arrival rate of customers.
- (ii) To determine the proportion of variation in the expected arrival rate that is being explained by the observed arrival rate.
- (iii) To determine the impact of the observed arrival rate on the expected arrival rate.
- (iv) To determine if the observed arrival rate exert significant influence on the expected arrival rate.

4. SCOPE OF THE STUDY

This study covers data on customers' arrival rate from First Bank Nigeria Plc located at Panseke area of Onikolobo, Abeokuta, Ogun State between 9:00a.m and 2:00p.m in the interval of 5 minutes. The data collected is primary in nature.

5. LITERATURE REVIEW

Econometrics is concerned with model building. An intriguing point to begin the inquiry is to consider the question, "What is the model?" The statement of a "model" typically begins with an observation or a proposition that one variable "is caused by" another, or "varies with another," or some qualitative statement about a relationship between a variable and one or more covariates that are expected to be related to the interesting one in question. The model might make a broad statement about behavior, such as the suggestion that individuals' usage of the health care system depends on, for example, perceived health status, demographics such as income, age, and education, and the amount and type of insurance they have. It might come in the form of a verbal proposition, or even a picture such as a flowchart or path diagram that suggests directions of influence. The econometric model rarely springs forth in full bloom as a set of equations. Rather, it begins with an idea of some kind of relationship. The natural next step for the econometrician is to translate that idea into a set of equations, with a notion that some feature of that set of equations will answer interesting questions about the variable of interest. To continue our example, a more definite statement of the relationship between insurance and health care demanded might be able to answer, how does health care system utilization depend on insurance coverage? Specifically, is the relationship "positive"—all else equal, is an insured consumer more likely to "demand more health care," or is it "negative"? And, ultimately, one might be interested in a more precise statement, "how much more (or less)"? From a purely statistical point of view, the researcher might have in mind a variable, y , broadly "demand for health care, H ," and another variable, x , income, I (Greene, 2010).

The bivariate regression model

The bivariate regression model is also known a simple regression model. It is a statistical tool that estimates the relationship between a dependent variable (y) and a single independent variable (x). The dependent variable is a variable which we want to forecast (Burç Ülengin, 2011).

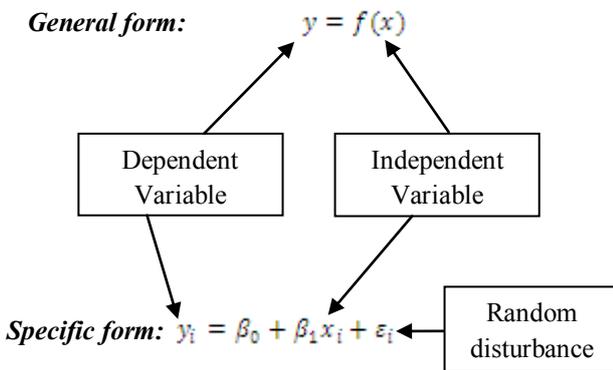
A *simple linear regression model* is a mathematical relationship between two quantitative variables, one of which, y , is the variable we want to predict, using information on the second variable, x , which is assumed to be non-random. Simple linear regression is the most commonly used technique for determining how one variable of interest (the response variable) is affected by changes in another variable (the explanatory variable). The terms "response" and "explanatory" mean the same thing as "dependent" and "independent", but the former terminology is preferred because the "independent" variable may actually be interdependent with many

other variables as well. Simple linear regression is used for three main purposes:

1. To describe the linear dependence of one variable on another.
2. To predict values of one variable from values of another, for which more data are available.
3. To correct for the linear dependence of one variable on another, in order to clarify other features of its variability.

Any line fitted through a cloud of data will deviate from each data point to greater or lesser degree. The vertical distance between a data point and the fitted line is termed a "residual". This distance is a measure of prediction error, in the sense that it is the discrepancy between the actual value of the response variable and the value predicted by the line. Linear regression determines the best-fit line through a scatter plot of data, such that the sum of squared residuals is minimized; equivalently, it minimizes the error variance. The fit is "best" in precisely that sense: the sum of squared errors is as small as possible. That is why it is also termed "**Ordinary Least Squares (OLS)**" regression (Mosteller and Tukey, 1977).

The simple linear regression model, in which there is only one explanatory variable on the right hand side of the regression equation is written as:



The regression model is indeed a line equation. The unknown parameters β_0 and β_1 are the intercept and slope of the regression function. We refer to them as population parameters (Joshua Sherman, 2003).

- y is called the response variable or the dependent variable (because its value depends to some extent on the value of x).
- x is called the predictor (because it is used to predict y) or explanatory variable (because it explains the variation or changes in y). It is also called the independent variable (because its value does not depend on y).
- The parameters of the true regression line are the constants, β_0 and β_1
- β_0 = intercept that tell us the value of y when $x = 0$.
- β_1 = slope coefficient that tell us the rate of change in y per unit change in x .

- ϵ is random disturbance, which causes for given x , y can take different values.

The Objective is to estimate β_0 and β_1 such a way that the fitted values should be as close as possible (Burç Ülengin, 2011).

The classical assumptions

Burç Ülengin (2011) highlighted seven classical assumptions of the simple linear regression model.

1. The regression model is linear in the coefficients, correctly specified, & has an additive error term.
2. $E(\epsilon) = 0$.
3. All explanatory variables are uncorrelated with the error term.
4. Errors corresponding to different observations are uncorrelated with each other.
5. The error term has a constant variance.
6. No explanatory variable is an exact linear function of any other explanatory variable(s).
7. The error term is normally distributed such that: $\epsilon_i \sim iid N(0, \sigma^2)$.

Best fit estimates

In practice, the econometrician will possess a sample of y values corresponding to some fixed x values rather than data from the entire population of values. Therefore the econometrician will never truly know the values of β_0 and β_1 (Joshua Sherman, 2003).

However, we may estimate these parameters. We will denote these estimators as b_0 and b_1 . So how shall we find b_0 and b_1 ? We need a method or rule for how to estimate the population parameters using sample data.

The most widely used rule is the method of least squares, or ordinary least squares (OLS). According to this principle, a line is fitted to the data that renders the sum of the squares of the vertical distances from each data point to the line as small as possible (Joshua Sherman, 2003).

Therefore the fitted line may be written as:

$$\hat{y}_i = b_0 + b_1 x_i \quad \text{_____}(1)$$

The vertical distances from the fitted line to each point are the least squares residuals, ϵ_i . They are given by:

$$\epsilon_i = y_i - \hat{y}_i = y_i - b_0 - b_1 x_i \quad \text{_____}(2)$$

Where \hat{y}_i is the predicted value of y_i

Mathematically, we want to find b_0 and b_1 such that the sum of the squared vertical distances from the data points to the line is minimized.

Square of the residuals gives

$$\varepsilon_i^2 = (y_i - b_0 - b_1 x_i)^2 \quad \text{---(3)}$$

The sum of the square of the residuals gives

$$\sum \varepsilon_i^2 = \sum (y_i - b_0 - b_1 x_i)^2 \quad \text{---(4)}$$

Let S represents the sum of square of the residuals, so that

$$S = \sum \varepsilon_i^2$$

$$\implies S = \sum (y_i - b_0 - b_1 x_i)^2 \quad \text{---(5)}$$

We are to minimize S with respect to b_0 and b_1 . This is achieved by differentiating S and equate the derivative to zero (0).

$$\text{Thus } \frac{dS}{db_0} = -2 \sum (y_i - b_0 - b_1 x_i) = 0 \quad \text{---(6)}$$

$$\text{Or } -2 \sum (y_i - b_0 - b_1 x_i) = 0$$

$$\sum (y_i - b_0 - b_1 x_i) = 0$$

$$\sum y_i - \sum b_0 - \sum b_1 x_i = 0$$

$$\sum y_i - \sum b_0 - b_1 \sum x_i = 0$$

$$\sum y_i - nb_0 - b_1 \sum x_i = 0$$

$$nb_0 + b_1 \sum x_i = \sum y_i$$

$$nb_0 = \sum y_i - b_1 \sum x_i$$

$$b_0 = \frac{\sum y_i}{n} - b_1 \frac{\sum x_i}{n}$$

$$b_0 = \bar{y} - b_1 \bar{x} \quad \text{---(7)}$$

Also

$$\frac{dS}{db_1} = -2 \sum (y_i - b_0 - b_1 x_i) x_i = 0 \quad \text{---(8)}$$

$$\text{Or } -2 \sum (x_i y_i - b_0 x_i - b_1 x_i^2) = 0$$

$$\sum (x_i y_i - b_0 x_i - b_1 x_i^2) = 0$$

$$\sum x_i y_i - \sum b_0 x_i - \sum b_1 x_i^2 = 0$$

$$\sum x_i y_i - \sum b_0 x_i - b_1 \sum x_i^2 = 0$$

$$\sum b_0 x_i + b_1 \sum x_i^2 = \sum x_i y_i$$

$$b_1 \sum x_i^2 = \sum x_i y_i - \sum b_0 x_i$$

$$b_1 \sum x_i^2 = \sum x_i y_i - \sum b_0 x_i$$

$$b_1 \sum x_i^2 = \sum x_i y_i - b_0 \sum x_i \quad \text{---(9)}$$

On substituting (7) into (9) we have

$$b_1 \sum x_i^2 = \sum x_i y_i - (\bar{y} - b_1 \bar{x}) \sum x_i$$

$$b_1 \sum x_i^2 = \sum x_i y_i - \bar{y} \sum x_i + \bar{x} b_1 \sum x_i$$

$$b_1 \sum x_i^2 - \bar{x} b_1 \sum x_i = \sum x_i y_i - \bar{y} \sum x_i$$

$$b_1 (\sum x_i^2 - \bar{x} \sum x_i) = \sum x_i y_i - \bar{y} \sum x_i$$

$$b_1 = \frac{\sum x_i y_i - \bar{y} \sum x_i}{\sum x_i^2 - \bar{x} \sum x_i}$$

$$= \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad \text{---(10)}$$

Where $\bar{y} = \frac{\sum y_i}{n}$ and $\bar{x} = \frac{\sum x_i}{n}$ are the sample means of the observations on y and x .

OLS and the true parameter values

The OLS estimators b_0 and b_1 are related to β_0 and β_1 thus (Joshua Sherman, 2003):

1. If assumptions 1 and 2 from earlier hold, then $E(b_0) = \beta_0$ and $E(b_1) = \beta_1$. That is, if we were able to take repeated samples, the expected value of the estimators b_0 and b_1 would equal the true parameter values β_0 and β_1 .
2. When the expected value of any estimator of a parameter equals the true parameter value, then that estimator is unbiased.

So the idea behind OLS is that if we are dealing with an instance in which certain assumptions hold, the expected value of the estimators b_0 and b_1 will equal the true parameter values β_0 and β_1 (Joshua Sherman, 2003).

Poisson process

A Poisson process is a specific counting process. Let $N(t)$ be a counting process. That is, $N(t)$ is the number of occurrences (or arrivals, or events) of some process over the time interval $[0, t]$. $N(t)$ looks like a step function.

Examples: $N(t)$ could be any of the following.

- (a) Cars entering a shopping center (time).
- (b) Defects on a wire (length).
- (c) Raisins in cookie dough (volume).

Let $\lambda > 0$ be the average number of occurrences per unit time (or length or volume). In the above examples, we might have: (a) $\lambda = 10/\text{min}$. (b) $\lambda = 0.5/\text{ft}$. (c) $\lambda = 4/\text{in}^3$.

First, some notation: $o(h)$ is a generic function that goes to zero faster than h goes to zero.

A Poisson process is one that satisfies the following assumptions:

1. There is a short enough interval of time, say of length h , such that, for all t ,

$$\Pr(N(t+h) - N(t) = 0) = 1 - \lambda h + o(h)$$

$$\Pr(N(t+h) - N(t) = 1) = \lambda h + o(h)$$

$$\Pr(N(t+h) - N(t) \geq 2) = o(h)$$

That is arrivals basically occur one-at-a-time, and then at rate λ /unit time. (We must make sure that λ doesn't change overtime.)

2. If $t_1 < t_2 < t_3 < t_4$, then $N(t_4) - N(t_3)$ and $N(t_2) - N(t_1)$ are independent random variables. That is the numbers of arrivals in two disjoint time intervals are independent.

The Poisson distribution

Many experimental situations occur in which we observe the counts of events within a set unit of time, area, volume, length etc. For example,

- The number of cases of a disease in different towns
- The number of mutations in set sized regions of a chromosome
- The number of dolphin pod sightings along a flight path through a region
- The number of particles emitted by a radioactive source in a given time
- The number of births per hour during a given day

In such situations we are often interested in whether the events occur randomly in time or space. The Poisson distribution plays a key role in modelling such problems. The Poisson distribution is a discrete probability distribution for the counts of events that occur randomly in a given interval of time or space (Jonathan Marchini, 2008).

Suppose we are given an interval (this could be time, length, area or volume) and we are interested in the number of "successes" in that interval. Assume that the interval can be divided into very small subintervals such that:

1. the probability of more than one success in any subinterval is zero;
2. the probability of one success in a subinterval is constant for all subintervals and is proportional to its length;
3. subintervals are independent of each other.

If we let X = The number of events in a given interval, Then, if the mean number of events per interval is λ The probability of observing x events in a given interval is given by

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad x = 0, 1, 2, 3, \dots \quad \text{_____}(11)$$

If the probabilities of X are distributed in this way, we write

$$X \sim P_o(\lambda) \quad \text{_____}(12)$$

λ is the **parameter** of the distribution. We say X follows a Poisson distribution with parameter λ .

Using the Poisson to approximate the Binomial

The Binomial and Poisson distributions are both discrete probability distributions. In some circumstances the distributions are very similar (Jonathan Marchini, 2008).

In general, If n is large (say > 50) and p is small (say < 0.1) then a $\text{Bin}(n, p)$ can be approximated with a $P_o(\lambda)$ where $\lambda = np$.

The idea of using one distribution to approximate another is widespread throughout statistics. In many situations it is extremely difficult to use the exact distribution and so approximations are very useful.

6. MATERIALS & METHODS

Research design

This research was designed to model the expected arrival rate of bank customers (whose observed arrival rate assumed to have followed a Poisson distribution) using simple linear regression model.

For the successful execution of this research work, primary data on customers' arrival rate from First Bank Nigeria Plc located at Panseke area of Onikolobo, Abeokuta, Ogun State between 9:00a.m and 2:00p.m in the interval of 5 minutes was employed. Data collected were analyzed electronically using Ms Excel 2007 and SPSS 21.0.

Techniques of data analysis

The data analysis techniques employed are the Simple Linear Regression, Correlation (R) and Coefficient of Determination (R^2).

Method of data analysis

In analyzing the data for simple linear regression, customers' observed arrival rate represents the independent variable (O_f) while customers' expected arrival rate represents the dependent variable (E_f). Data for the independent variable was collected primarily while data for the dependent variable was estimated as:

$$E_f = N \cdot P(x) \quad \text{_____}(13)$$

$$\text{where } N = \sum O_f$$

N is the number of intervals of t periods (5 minutes) or total observed frequency.

x is the number of arrivals per period (i.e during the 1st, 2nd, ..., N^{th} intervals).

$P(x)$ is the probability of x arrivals during the 1st, 2nd, ..., N^{th} intervals. x is assumed to follow a Poisson distribution and defined as:

$$P(x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad x = 0, 1, 2, 3, \dots$$

λ is the average arrival per period and defined as:

$$\lambda = \frac{\sum O_f x}{\sum O_f} \quad \text{_____ (14)}$$

$\sum O_f x$ is the total number of arrivals for the N periods.

The expected arrival rate (E_f) is then regressed on the observed arrival rate (O_f) as:

$$E_f = a + b O_f \quad \text{_____ (15)}$$

a is the intercept, which implies the expected arrivals when $O_f = 0$

b is the slope, which implies that for every single increase in observed arrival rate the expected arrival rate will increase by b .

7. RESULT

TABLE 1: Estimation of Expected Arrival Rate

x	O_f	$O_f x$	$P(x)$	E_f
1	0	0	1E-12	0.00
2	0	0	2E-11	0.00
3	0	0	2E-10	0.00
4	0	0	2E-09	0.00
5	0	0	9E-09	0.00
6	0	0	5E-08	0.00
7	0	0	2E-07	0.00
8	0	0	8E-07	0.00
9	0	0	3E-06	0.00
10	0	0	9E-06	0.00
11	0	0	2E-05	0.00
12	0	0	6E-05	0.00
13	0	0	1E-04	0.01
14	0	0	3E-04	0.02
15	0	0	7E-04	0.04
16	0	0	0.001	0.08
17	0	0	0.002	0.14
18	1	18	0.004	0.24
19	0	0	0.006	0.39
20	2	40	0.010	0.60
21	1	21	0.015	0.88

22	3	66	0.021	1.24
23	2	46	0.028	1.66
24	4	96	0.036	2.14
25	1	25	0.044	2.64
26	4	104	0.052	3.13
27	2	54	0.060	3.57
28	0	0	0.066	3.94
29	7	203	0.070	4.19
30	6	180	0.072	4.31
31	3	93	0.071	4.29
32	2	64	0.069	4.13
33	2	66	0.064	3.86
34	2	68	0.058	3.50
35	5	175	0.051	3.09
36	5	180	0.044	2.65
37	0	0	0.037	2.21
38	0	0	0.030	1.79
39	0	0	0.024	1.42
40	2	80	0.018	1.09
41	2	82	0.014	0.82
42	0	0	0.010	0.60
43	0	0	0.007	0.43
44	0	0	0.005	0.30
45	1	45	0.003	0.21
46	2	92	0.002	0.14
47	0	0	0.002	0.09
48	0	0	1E-03	0.06
49	0	0	6E-04	0.04
50	0	0	4E-04	0.02
51	0	0	2E-04	0.01
52	0	0	1E-04	0.01
53	1	53	8E-05	0.00
>53	0	0	1E-04	0.01
	$N = 60$	$\sum O_f x = 1851$		

$$\lambda = \frac{1851}{60} = 30.85$$

$$P(x) = e^{-30.85} \frac{(30.85)^x}{x!}$$

$$E_f = 60 \cdot P(x)$$

TABLE 2: Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.716 ^a	.513	.503	1.04535

a. Predictors: (Constant), Observed Arrival Rate

b. Dependent Variable: Expected Arrival Rate

TABLE 3: ANOVA^a

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	59.803	1	59.803	54.726	.000 ^b
1 Residual	56.824	52	1.093		
Total	116.627	53			

a. Dependent Variable: Expected Arrival Rate
b. Predictors: (Constant), Observed Arrival Rate

TABLE 4: Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	.430	.169		2.539	.014
1 Observed Arrival Rate	.613	.083	.716	7.398	.000

a. Dependent Variable: Expected Arrival Rate

The regression model is thus:

$$E_f = 0.430 + 0.613O_f$$

8. DISCUSSION OF RESULTS

The λ value of 30.85 implies that we expect approximately 31 customers to arrive the banking hall every 5 minutes. The minimum and maximum observed frequencies are respectively 0 and 7 while the minimum and maximum expected frequencies are respectively 0 and 4.

The correlation (R) value of 0.716 indicates that there is a strong positive but imperfect correlation between the observed and expected arrival rate of customers. The coefficient of determination (R^2) value of 0.513 indicates that approximately 51.3% of the variation in the expected arrival rate is being explained by observed arrival rate.

The regression coefficients $a = 0.430$ indicates that when the observed arrival rate is zero, the expected arrival rate is 0.43 which is approximately also equal zero and $b = 0.613$ implies that for every single observed arrival rate, the expected arrival rate will increase by 0.613 which is approximately 1. The observed arrival rate has a positive impact on the expected arrival rate with a positive t -value of 7.398. The sig -value of 0.000 indicates that the observed arrival rate exert significant influence on the expected arrival rate.

9. CONCLUSIONS

We expect approximately 31 customers to arrive the banking hall of First Bank Nigeria Plc, Panseke Branch of Abeokuta, Ogun State every 5 minutes. The relationship between the observed and expected customers' arrival rate is positive and strong. An increase in the number of observed arrival rate will leads to corresponding equal increase in expected arrival rate. The observed arrival rate contributes to approximately half of the variation in the expected arrival rate of the customers.

10. RECOMMENDATIONS

1. An average arrival rate of 31 customers per 5 minutes is high. The bank should therefore target a corresponding average service rate of 31 or more customers every 5 minutes in other to avoid over congestion from the arrival rate.
2. In case of an anticipated congestion, the bank management should try as much as possible to decentralize their services to increase their service rate.

REFERENCES

- [1] Burç Ülengin (2011). Forecasting With Regression Models. ITU Management Engineering Faculty.
- [2] Greene (2010). The Linear Regression Model.
- [3] Jonathan Marchini (2008). The Poisson Distribution.
- [4] Joshua Sherman (2003). The Linear Regression Model.
- [5] Kuzhda, T. (2012). Retail Sales Forecasting with Application. The multiple regression. *Sotsial'no ekonomichni problemy i derzhava-Socio-Economic Problems and the State* [online]. 6 (1), p.91-101. [Accessed May 2012]. Available from: <http://sepd.tntu.edu.ua/images/stories/pdf/2012/12k_tibrm.pdf>.
- [6] Michael H. Kutner, John Neter, Christopher J. Nachtsheim, William Wasserman (2004). Applied Linear Statistical Models.
- [7] Mosteller, F. and J. W. Tukey (1997). Data Analysis

- and Regression. 588 pp., *Addison-Wesley*, 1977.
[8] Robert M., (2011). Regression Analysis Appliation in Litigation. *Microeconomics, Inc.*
[9] Tabachnick and Fidell (1996). Multiple Linear Regression.

APPENDIX

TABLE 1: Data on the number of arrivals (x) per period and observed frequency (O_f)

x	O_f
18	1
20	2
21	1
22	3
23	2
24	4
25	1
26	4
27	2
29	7
30	6
31	3
32	2
33	2
34	2
35	5
36	5
40	2
41	2
45	1
46	2
53	1

Source: Field Survey (2015)