

The Chi-Square Goodness-Of-Fit Test for a Poisson distribution: Application to the Banking System

¹Sulaimon Mutiu O. ; ²Oyefusi Olutayo A. & ³Ayeleso Timothy O.

^{1,2,3}Department of Statistics & Mathematics Moshood Abiola Polytechnic, Abeokuta, Ogun State, Nigeria.

¹mtsulaimon@gmail.com; ²tayooyefusi@yahoo.com; ³ayelosobiyi@gmail.com

ABSTRACT

This study covers the Chi-square goodness-of-fit test of the Poisson distribution of customers' arrivals rate. A case study of First Bank Plc., Panseke, Abeokuta, Ogun State, Nigeria was used. Data on customers' arrival rate per 5 minutes for duration of 5 hours (9am – 2pm) was collected for the research study. The data for this study was collected via primary method of data collection with the principle of observation method which involves the physical presence of the researchers collecting the data and the subject of enquiry. The Poisson distribution was used to find the probability of a specific number of arrivals per "5 minutes". To determine whether the number of arrivals per 5 minutes follows a Poisson distribution, the Chi-square goodness-of-fit-test was employed. The average arrival rate is estimated to be approximately 31 customers per 5 minutes. The highest and lowest expected frequencies are 4 and 0 respectively while the highest and the lowest observed frequencies are respectively 7 and 0. With respect to the case study employed, scope of the study, research question, research design, method of data collection and technique of analysis, the research results revealed that arrival rate of customers to a banking system per 5 minutes in the duration of 5 hours with average arrival rate of 31 customers does not follow a Poisson distribution.

Keywords: Banking System, Chi-square, Goodness-of-fit Test, Poisson Distribution

INTRODUCTION

Every relationship is a game and banker-customer relationship is not an exception. The corporate objective of any bank which is maximization of shareholders' wealth can only be achieved if customers are retained and satisfied. This is in line with Philip Kotler's (1999) perception that the key to successful marketing of financial services is identification and packaging of customers' needs to their satisfaction. The competition in Nigerian banking sector is getting more intense, partly due to regulatory imperatives of universal banking and also due to customers' awareness of their rights. Bank customers have become increasingly demanding, as they require high quality, low priced and immediate service delivery. They want additional improvement of value from their chosen banks (Olaniyi, 2004).

In Nigeria, a study conducted by Olaniyi (2004) revealed a positive correlation between arrival rates of customers and bank's service rates. He concluded that the potential utilization of the banks service facility was 3.18% efficient and idle 68.2% of the time. However, Ashley (2000) asserted that even if service system can provide

service at a faster rate than customers arrival rate, waiting lines can still form if the arrival and service processes are random.

Often the arrival process of customers can be described by a Poisson process. Poisson process is a viable model when the arrivals originate from a large population of customers (Virtamo, NA).

It is to this effect that this research work intends to examine the goodness-of-fit test of the Poisson distribution of customer's arrival rate at a banking system per 5 minutes.

STATEMENT OF PROBLEM

The problem of this research work is examining the goodness-of-fit of the Poisson distribution of customers' arrival rate at a banking system per 5 minutes.

PURPOSE OF THE STUDY

The purpose of this study is to determine whether customers' arrival rate at a banking system per 5 minutes follows a Poisson distribution.

SCOPE OF THE STUDY

This study covers the Chi-square goodness-of-fit test of the Poisson distribution of customers' arrivals rate at a banking system. A case study of First Bank Plc, Panseke, Abeokuta, Ogun State, Nigeria was employed. Primary data on customers' arrival rate per 5 minutes for duration of 5 hours (9am – 2pm) was collected for the research study.

RESEARCH QUESTION

Does customers' arrival rate at a banking system per 5 minutes follows a Poisson distribution?

RESEARCH HYPOTHESIS

H₀: The arrival of customers at a banking system per 5 minutes follows a Poisson distribution.

H₁: The arrival of customers at a banking system per 5 minutes does not follow a Poisson distribution.

SIGNIFICANCE OF THE STUDY

This research work will be of great significance to the management of banks and other related financial institutions where customers arrival rate is of priority, thereby putting necessary measures in place for possible congestion in the banking system. It will also be of importance to research students with interest in a related research work.

LITERATURE REVIEW

The chi in Chi-square is the Greek letter "χ" pronounced Ki as in kite. A Chi-square (χ^2) procedure measures the difference between observed and expected frequencies of nominal variables, in which subjects are grouped in categories or cells. (Helen, 2006).

The Chi-square test is a particular useful technique for testing whether observed data are representative of a particular distribution. The Chi-square goodness of fit test is used to find out how the observed value of a given phenomenon is significantly different from the expected value. It determines how well theoretical distribution such as (normal distribution, binomial distribution, or Poisson distribution) fit the empirical distribution.

The Chi-square goodness of fit test compares the expected frequencies of categories from a population distribution to the observed or actual frequencies from a distribution to verify whether

there is difference among the expected and the observed.

The Goodness-of-fit test is applied to a single nominal variable and determines whether the frequencies we observe in n categories fit what we might expect.

The Chi-square statistic is defined as:

$$\chi^2 = \sum_{i=1}^n \left[\frac{(f_o - f_e)^2}{f_e} \right] \quad \text{--- (1)}$$

Where f_o represents the observed frequency (the actual count in a given cell)

f_e represents the expected frequency (a theoretical count for that cell).

The formula read as follows "the value of chi-square equals the sum of $f_o - f_e$ differences squared and divided by f_e ". The more f_o differs from f_e , the larger χ^2 is. When χ^2 exceeds the appropriate critical value, it is declared significant. (Helen C. A., 2006).

The number of degrees of freedom depends on the distribution of the sample, which is $n - p - 1$, where p is the number of parameters estimated from the (sample) data used to generate the hypothesized distribution.

The Poisson distribution

A Poisson random variable is the number of successes that result from a Poisson experiment. A "Poisson distribution" is the probability distribution of a Poisson variable. A discrete random variable X is said to have a Poisson distribution with parameter $\lambda > 0$, if for $X = 0, 1, 2, \dots$

The probability mass function of X is given by:

$$f(x; \lambda) = pr(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{--- (2)}$$

Where e is Euler's number ($e = 2.71828$)

The positive real number λ is equal to the expected value of x and also to its variance.

$$\lambda = E(x) = Var(x) \quad \text{--- (3)}$$

The Poisson distribution can be applied to systems with a large number of possible events each

of which is rare. How many such events will occur during a fixed time interval? Under the right circumstances, this is a random number with a Poisson distribution.

The Poisson distribution can be used to calculate the probabilities of various numbers of “successes” based on the mean number of successes. In order to apply the Poisson distribution, the various events must be independent. Keep in mind that the term “success” does not really mean success in the traditional positive sense. It just means that the outcome in question occurs.

Expectation and variance of the Poisson distribution

The expectation and variance of the Poisson distribution can be derived directly from the definitions which apply to any discrete probability distribution. However, the algebra involved is a little lengthy. Instead we derive them from the binomial distribution from which the Poisson distribution is derived.

One way the mean and variance of the Poisson distribution is estimated is to consider the behavior of the binomial distribution under the following conditions:

- (i) n is large (ii) p is small
- (iii) $np = \lambda$ (a constant)

Recalling that the expectation and variance of the binomial distribution are given by the results:

$$E(X) = np \text{ and } V(X) = np(1-p) = npq \quad \text{--- (4)}$$

It is reasonable to assert that condition (ii) implies since $q = 1 - p$, that q is approximately 1 and so the expectation and variance are given by

$$E(X) = np \text{ and } V(X) = npq \approx np \quad \text{--- (5)}$$

In fact the algebraic deviation of the expectation and variance of the Poisson distribution shows that these results are in fact exact.

METHODOLOGY

Research design

The data for this study was collected by primary method of data collection with the principle of observation method which involves the physical

presence of the researchers collecting the data and the subject of enquiry.

This paper considers the arrival of customers at the First Bank Plc, Panseke branch per “5 minute” for the duration of “5 hours” from 9am to 2pm.

Techniques of data analysis

We used the Poisson distribution to find the probability of a specific number of arrivals, per “5 minutes” interval at First Bank Plc, located at Panseke, Abeokuta, Ogun state, Nigeria.

To determine whether the number of arrival per 5 minutes follows a Poisson distribution, the Chi-square goodness-of-fit-test was employed.

The Poisson distribution has one parameter, its mean λ , and we need to specify its value in the null and alternative hypotheses.

Hence, we estimate “ λ ” using the formula

$$\lambda = \frac{\sum fx}{\sum f} \quad \text{--- (6)}$$

We compute the expected frequency for each number of arrivals by multiplying the appropriate Poisson probability by the sample size “ N ”.

Hence, to solve for probabilities $p(x)$ for the Poisson distribution with the λ , we use

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}; \quad x = 0, 1, 2, 3 \dots \quad \text{--- (7)}$$

The expected frequency is estimated with the formula

$$f_e = N \cdot p(x) \quad \text{--- (8)}$$

Where N is the sample size i.e $N = \sum f$

The chi-square test is then carried out using

$$\chi^2 = \sum_{i=1}^n \left[\frac{(f_o - f_e)^2}{f_e} \right]$$

Hence, the test statistic χ_{cal}^2 follows a chi-square distribution with $n - p - 1$ degree of freedom.

Where n = number of categories or classes remaining after combining classes.

p = number of parameters estimated.

The Decision rule is to reject H_0 if $\chi_{cal}^2 \geq \chi_{tab}^2$ at α significance level. χ_{tab}^2 (the critical value) is the table value of the Chi-square.

RESULTS

Table 1: Customers' arrival rate at First Bank Plc per 5 minutes for 5 hours (9am to 2pm)

S/N	Time	X
1	09:00 - 09:05	22
2	09:05 - 09:10	20
3	09:10 - 09:15	27
4	09:15 - 09:20	29
5	09:20 - 09:25	36
6	09:25 - 09:30	24
7	09:30 - 09:35	29
8	09:35 - 09:40	30
9	09:40 - 09:45	29
10	09:45 - 09:50	21
11	09:50 - 09:55	33
12	09:55 - 10:00	22
13	10:00 - 10:05	29
14	10:05 - 10:10	33
15	10:10 - 10:15	35
16	10:15 - 10:20	26
17	10:20 - 10:25	24
18	10:25 - 10:30	20
19	10:30 - 10:35	30
20	10:35 - 10:40	31
21	10:40 - 10:45	25
22	10:45 - 10:50	18
23	10:50 - 10:55	35
24	10:55 - 11:00	34
25	11:00 - 11:05	24
26	11:05 - 11:10	24
27	11:10 - 11:15	30
28	11:15 - 11:20	34
29	11:20 - 11:25	26
30	11:25 - 11:30	26
31	11:30 - 11:35	29
32	11:35 - 11:40	40
33	11:40 - 11:45	36
34	11:45 - 11:50	32
35	11:50 - 11:55	35
36	11:55 - 12:00	31
37	12:00 - 12:05	31
38	12:05 - 12:10	29
39	12:10 - 12:15	53

40	12:15 - 12:20	22
41	12:20 - 12:25	40
42	12:25 - 12:30	35
43	12:30 - 12:35	27
44	12:35 - 12:40	41
45	12:40 - 12:45	29
46	12:45 - 12:50	30
47	12:50 - 12:55	35
48	12:55 - 01:00	46
49	01:00 - 01:05	23
50	01:05 - 01:10	30
51	01:10 - 01:15	36
52	01:15 - 01:20	32
53	01:20 - 01:25	45
54	01:25 - 01:30	41
55	01:30 - 01:35	30
56	01:35 - 01:40	26
57	01:40 - 01:45	36
58	01:45 - 01:50	23
59	01:50 - 01:55	46
60	01:55 - 02:00	36

Source: Field Survey (2016)

Table 2: Frequency distribution of customers' arrival rate at First Bank Plc per 5 minutes.

x	f	fx
≤ 1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	1	18
19	0	0
20	2	40
21	1	21
22	3	66

23	2	46
24	4	96
25	1	25
26	4	104
27	2	54
28	0	0
29	7	203
30	6	180
31	3	93
32	2	64
33	2	66
34	2	68
35	5	175
36	5	180
37	0	0
38	0	0
39	0	0

40	2	80
41	2	82
42	0	0
43	0	0
44	0	0
45	1	45
46	2	92
47	0	0
48	0	0
49	0	0
50	0	0
51	0	0
52	0	0
53	1	53
>53	0	0
	$\sum f = 60$	$\sum fx = 1851$

Table3: Chi-square test of the Poisson distribution of customers' arrival rate

$\lambda = \frac{\sum fx}{\sum f} = 30.85, \quad N = \sum f = 60, \quad P(x) = \frac{(30.85)^x \cdot e^{-(30.85)}}{x!}, \quad f_e = 60 \cdot p(x)$							
POISSON DISTRIBUTION				CHI-SQUARE TEST			
x	f	fx	$p(x)$	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
≤1	0	0	1.2E-12	7.4E-11	-7.403E-11	5.481E-21	7.4032E-11
2	0	0	1.9E-11	1.14E-09	-1.142E-09	1.304E-18	1.1419E-09
3	0	0	2E-10	1.17E-08	-1.174E-08	1.379E-16	1.1743E-08
4	0	0	1.5E-09	9.06E-08	-9.057E-08	8.203E-15	9.0568E-08
5	0	0	9.3E-09	5.59E-07	-5.588E-07	3.123E-13	5.5881E-07
6	0	0	4.8E-08	2.87E-06	-2.873E-06	8.255E-12	2.8732E-06
7	0	0	2.1E-07	1.27E-05	-1.266E-05	1.603E-10	1.2663E-05
8	0	0	8.1E-07	4.88E-05	-4.883E-05	2.384E-09	4.883E-05
9	0	0	2.8E-06	0.000167	-0.0001674	2.802E-08	0.00016738
10	0	0	8.6E-06	0.000516	-0.0005164	2.666E-07	0.00051636
11	0	0	2.4E-05	0.001448	-0.0014482	2.097E-06	0.00144816
12	0	0	6.2E-05	0.003723	-0.003723	1.386E-05	0.00372299
13	0	0	0.00015	0.008835	-0.0088349	7.806E-05	0.00883493
14	0	0	0.00032	0.019468	-0.0194684	0.000379	0.01946841
15	0	0	0.00067	0.04004	-0.04004	0.0016032	0.04004003



16	0	0	0.00129	0.077202	-0.0772022	0.0059602	0.07720218
17	0	0	0.00233	0.140099	-0.1400993	0.0196278	0.14009926
18	1	18	0.004	0.240115	0.7598854	0.5774259	2.4047933
19	0	0	0.0065	0.38987	-0.3898702	0.1519988	0.38987022
20	2	40	0.01002	0.601375	1.3986252	1.9561524	3.25280071
21	1	21	0.01472	0.883448	0.1165518	0.0135843	0.01537647
22	3	66	0.02065	1.238835	1.7611646	3.1017009	2.50372321
23	2	46	0.02769	1.661655	0.3383447	0.1144772	0.06889345
24	4	96	0.0356	2.135919	1.8640806	3.4747966	1.62683884
25	1	25	0.04393	2.635725	-1.6357245	2.6755947	1.01512683
26	4	104	0.05212	3.127388	0.8726115	0.7614508	0.24347817
27	2	54	0.05956	3.573331	-1.5733309	2.4753702	0.69273467
28	0	0	0.06562	3.937045	-3.937045	15.500323	3.93704497
29	7	203	0.0698	4.188201	2.8117987	7.906212	1.88773448
30	6	180	0.07178	4.306867	1.693133	2.8666994	0.66561131
31	3	93	0.07143	4.286027	-1.2860273	1.6538663	0.38587394
32	2	64	0.06887	4.131998	-2.1319982	4.5454164	1.10005284
33	2	66	0.06438	3.862792	-1.8627923	3.469995	0.89831261
34	2	68	0.05842	3.504916	-1.5049159	2.2647719	0.64617012
35	5	175	0.05149	3.089333	1.910667	3.6506483	1.18169463
36	5	180	0.04412	2.647387	2.3526132	5.534789	2.09066125
37	0	0	0.03679	2.207348	-2.2073482	4.8723859	2.20734817
38	0	0	0.02987	1.792018	-1.7920182	3.2113292	1.79201818
39	0	0	0.02363	1.417532	-1.4175323	2.0093979	1.41753233
40	2	80	0.01822	1.093272	0.9067282	0.822156	0.75201428
41	2	82	0.01371	0.82262	1.1773796	1.3862228	1.68513062
42	0	0	0.01007	0.604234	-0.6042343	0.365099	0.60423425
43	0	0	0.00723	0.433503	-0.4335029	0.1879248	0.43350295
44	0	0	0.00507	0.303945	-0.3039447	0.0923824	0.30394468
45	1	45	0.00347	0.208371	0.791629	0.6266765	3.00750413
46	2	92	0.00233	0.139744	1.8602556	3.4605508	24.7634236
47	0	0	0.00153	0.091726	-0.0917259	0.0084136	0.09172587
48	0	0	0.00098	0.058953	-0.058953	0.0034755	0.05895298
49	0	0	0.00062	0.037116	-0.0371163	0.0013776	0.03711632
50	0	0	0.00038	0.022901	-0.0229008	0.0005244	0.02290077
51	0	0	0.00023	0.013853	-0.0138527	0.0001919	0.01385272

52	0	0	0.00014	0.008218	-0.0082184	6.754E-05	0.00821839
53	1	53	8E-05	0.004784	0.9952163	0.9904554	207.046929
>53	0	0	0.0001	0.0060678	-0.0060678	3.6818E-05	269.550774
	$\sum_{i=1}^{54} f = 60$	$\sum_{i=1}^{54} fx = 1851$	1	60			269.550774

NOTE that $f = f_0$

Therefore, Chi-square Goodness of fit test value is:

$$\chi_{cal}^2 = \sum_{i=1}^{54} \left[\frac{(f_o - f_e)^2}{f_e} \right] = 269.550774 \approx 269.551$$

DISCUSSION OF RESULTS

The average arrival rate is estimated to be approximately $31(\lambda = 30.85)$ customers per 5 minutes.. The highest and lowest expected frequencies are 4 and 0 respectively while the highest and the lowest observed frequencies are respectively 7 and 0.

In testing for the hypothesis, the decision rule is to reject H_0 if $\chi_{cal}^2 \geq \chi_{tab}^2$, otherwise accept H_0 . The critical value is estimated as χ_{tab}^2 with $n - p - 1$ degrees of freedom at α (0.05) level of significance.

$$\begin{aligned} \text{That is } \chi_{tab,0.05}^2 &= \chi_{n-p-1,0.05}^2 \\ &= \chi_{54-1-1,0.05}^2 = \chi_{52,0.05}^2 = 69.832 \end{aligned}$$

Therefore, since $\chi_{cal}^2 > \chi_{tab}^2$ i.e. (269.551 > 69.832), we reject H_0 to accept H_1 .

CONCLUSION

From the analysis of the research study, it can be concluded that the arrival rate of customers' at First Bank Plc, Panseke, Abeokuta, Ogun state per 5 minute interval in the duration of 5hours (9am to 2pm) with average arrival rate of 31 customers does not follow a Poisson distribution.

In general, the arrival of customers at a banking system per 5 minutes does not follow a Poisson distribution.

RECOMMENDATIONS

1. The estimated high arrival rate of 31 customers per 5 minutes indicates a possible congestion in the banking system. The bank management should try as much as possible to employ the optimal use of division labour in order to increase their service rate.
2. The bank should focus on a comparable average service rate of 31 or more customers per 5 minutes in other to avoid overcrowding due to the arrival rate.

BIBLIOGRAPHY

- [1] Ashley, D. W. (2000). "Introduction to waiting lime models" Google <http://www.com>
- [2] Helen, C. A. (2006). Studied the Relationship Between Predominant Leadership Style and



Educational Philosophy of Administrators in Christian Colleges and Universities.

[3] Philip, K. (1999) *Marketing Management*, the New Millenium Edition.

[4] Olaniyi, A. (2004) "Relationship between arrival rates of customers and bank's service rate's", seminar paper series, University of Ife.

[5] Virtamo, J. (NA). Queueing Theory/Poisson process

[6] https://en.wikipedia.org/wiki/poisson_distribution

[7] <http://www.stattrek.com/probability-distributions/poisson.aspx>