

## APPROACH FOR DESIGNING A SYSTEM THAT ACHIEVES THE DESIRED PERFORMANCE AT A SERVICE DESK

GAREBANGWE VICTORIA MABE-MADISA

Department of Decision Sciences  
University of South Africa  
P O Box 392, Unisa 0003, South Africa  
mabemgv@unisa.ac.za

Received April 2020; accepted July 2020

**ABSTRACT.** *In this paper, the problem of a high rate of abandoned calls at a centralized service desk of a South African telecommunications company is addressed. A queuing methodology is used to design a system that may achieve the desired performance level. A quality standard is created by determining the number of service channels (employees) that can handle the workload of answering all the calls arriving at the service desk. The paper demonstrates how queuing theory may be used to enhance a service desk information system, including the end-users, in order to provide more efficient service delivery.*

**Keywords:** Queuing theory, Service desk, Abandoned calls, Sensitivity analysis

**1. Introduction.** This paper addresses the problem of unacceptably high rate of abandoned calls at the service desk. It uses a queuing methodology to design a system that may achieve the desired performance level. A quality standard is created by determining the number of service channels (employees) that can handle the workload of answering all the calls arriving at the service desk. The paper demonstrates how queuing theory may be used to enhance a service desk information system, including the end-users, in order to provide more efficient service delivery. This approach is unique because it uses confidence intervals for estimating the average number of calls arriving at the service desk, not just the point estimates [1].

The company in which the research took place, is a telecommunication company with its headquarters situated in Pretoria, South Africa. Company employees from all regions in South Africa report events by phoning the centralized service desk. An event is anything that an end-user of the company's information and communication technologies (ICT) infrastructure regards as a problem to be fixed or as a request to be attended to in an information system. For example, the installation of new software, creation of a new e-mail account and setting up a computer on a network, are all requests. The re-installation of software and fixing of computer hardware are faults.

The time required to service a customer varies considerably from call to call because every call has its own problems. Arriving calls seek service from one of several service channels. A service channel is an employee servicing customers. Each call is automatically switched to an open channel. If all channels are busy, arriving calls are denied access to the system. Arrivals occurring when the system is full are blocked and are cleared from the system. These calls are referred to as abandoned calls. The service desk's focus is on customer satisfaction [2], and therefore it must have a performance evaluation system to monitor and encourage service desk employees in order to provide efficient service delivery [3,4].

The objective of this study is to determine the optimal number of service channels that should be used at the service desk in order to minimize the number of unanswered

calls. The research findings may be used to design a system that achieves the desired performance.

The paper is organized as follows: Section 2 discusses the research methodology; experiments and results are presented in Section 3, while Section 4 discusses conclusions and future work.

**2. Research Methods.** Queuing theory is the research method used. The company has a centralized service desk where all calls are reported through a telephone line for all the regions. Events are logged and routed to their respective regions.

**2.1. Data collection and interpretation.** Here, 1) the number of calls, as they entered the telephone system per minute (arrival rate), was recorded. Each call that arrives is displayed on the central screen for everyone to see, and when it is answered or abandoned, it is also shown on the screen. 2) The duration of the service (average service rate at each channel) was recorded. 3) Lastly, the number of channels was monitored. Table 1 shows the arrival rate of calls per day for a month.

TABLE 1. Arrival rate of calls (company's service desk)

Days of the month	3	4	5	6	7	10	11	12	13	14	17	18	19	20	21	24	25	26	27	28
Calls per minute	8	5	3	5	3	11	7	7	5	5	9	9	6	6	6	4	10	8	12	12

Summaries and visualization are tools that facilitate understanding of one's data, which is a necessary pre-requisite for the building of good models [1].

A commonly used starting point in summarizing data is to put the data into classes and then construct a histogram from the data that have been thus grouped. In this study this is done to verify whether the arrivals are Poisson distributed. The data used here are the data gathered from the centralized service desk for all the regions. The data from Table 1 above are used to determine the optimal number of channels that can handle the workload at the service desk in order to reduce the number of abandoned calls. The histogram of the arrival rates of calls in Table 1 above is depicted in Figure 1.

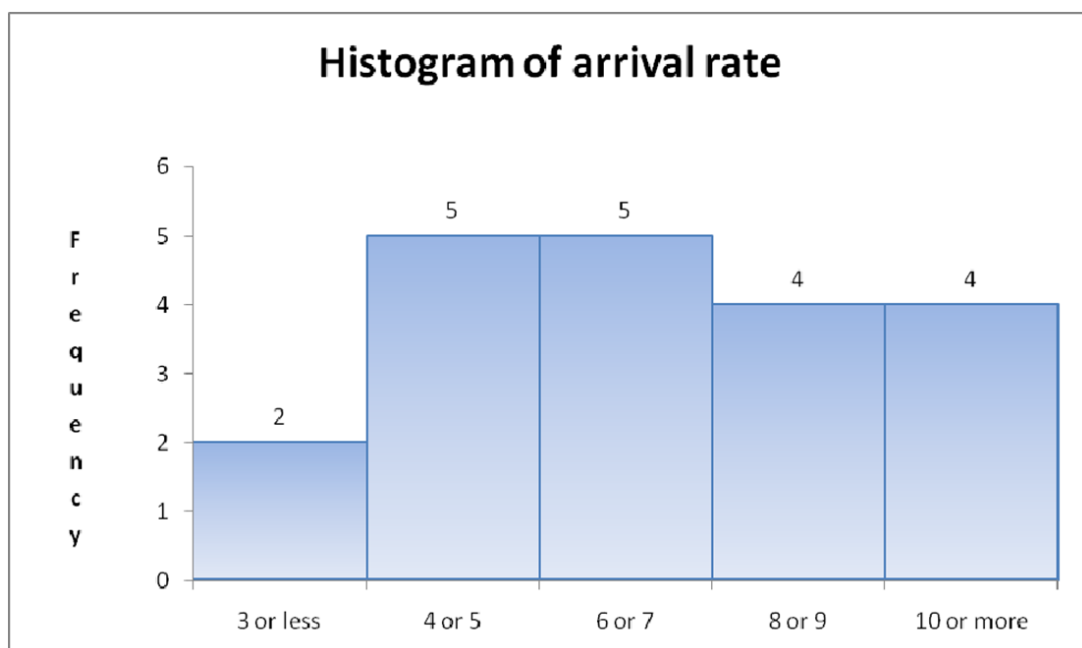


FIGURE 1. Arrival rate of calls

Some distributions are characterized at least partially by functions of their true parameters. Given Figure 1, one can make a fairly accurate guess that the observations point to a Poisson distribution. The mean and variance are almost equal, which confirms the data to be Poisson distributed [1,5,6].

The mean of a data set is simply the arithmetic average of the values in the set, obtained by summing the values and dividing by the number of values. The variance of a data set is the arithmetic average of the squared differences between the values and the mean. If the data set is considered a sample, the sum of the squared differences between the values and the mean is divided by (sample size-1). The standard deviation is the square root of the variance [1,5,6].

**2.2. The queuing model.** This section discusses the behaviour of the study's queuing model. This had to be examined (in order to determine whether the Poisson distribution is approximating the sample data), before a queuing model is developed. This was done in order to determine which assumptions the queuing model had to follow or which variables the queuing model had to use. The basic components of the queuing process are the arrival rate, the queue and the service rate; the researcher actually wants to find out which queuing assumptions have to be followed. In this study the multi-channel, single phase system is used. In this system the service rate does not follow any distribution, but the arrival rate follows a Poisson distribution. In the Poisson probability distribution, the observer records the number of events that occur in a time interval of fixed length. The observer determines the mean and the variance of the data, and if they are equal, then the distribution is Poisson. Also, the chi-square test is used to fit possible Poisson distributions [5-7]. In this study, there is an unlimited or infinite logging of events assumed.

If the average time a customer waits in the queue is denoted by  $W$ , and the average customer arrival in the queue by  $\lambda$ , a generalized equation applying to queuing model is  $L_q = \lambda W$ , where  $L_q$  is the average number of customers in the queue. This is known as Little's Law, as it was discovered by John D.C. Little [5].

The following particular assumptions are used in this model. 1) The queuing environment has an infinite calling population, and has multiple channel facility. 2) The arrival time is unpredictable and described by a Poisson distribution. 3) The service times (processing rate at the servicing facility) are exponential or unpredictable. 4) The queue lengths are infinite. 5) All customers wait in the single queue. 6) Service is on first-come first-served basis. 7) All arriving events enter the queue [5-7].

**3. Computations and Results.** In this study the queuing application involves calls answered from users reporting problems. These users are countrywide in South Africa. Users are people using the computers on a regular basis to perform their duties. They phone the central service desk in Pretoria to report the problems they have with their computers. The major task here is to design a system that achieves the desired performance level. The desired performance level is the number of channels that can handle the workload, thereby optimizing the costs.

**3.1. Test for goodness of fit.** The chi-square test is used to determine how well theoretical distributions, such as the Poisson, as is the case in this study, fit the distribution obtained from the sample data [1]. The data are divided into  $k = 3$  intervals of (0, 1, 2, 3, 4, 5), (6, 7, 8) and (9 and more). The reason for this is that the expected frequency in each of these combined cells is at least 5, so that the chi-square test can be used. The expected frequencies are computed on the basis of a hypothesis  $H_0$ .  $H_0$ : The  $x_i$ 's are random variables with distribution function  $F$  where  $F$  is a Poisson distribution. If, under this hypothesis, the computed value of  $\chi^2$  is smaller than some critical value (such as  $\chi_{0.95}^2$ ) which is the critical value at the 0.05 significance level, the null hypothesis ( $H_0$ ) is not rejected.

The test statistic equation is  $\chi^2 = \sum j(O_j - E_j)^2/E_j$ , where  $\chi^2$  is the chi-square statistic. The  $O_j$ 's are the observed frequencies and  $E_j$ 's are the expected frequencies. The data in Table 1 were used to determine the observed frequencies in Table 2. Similar arrival rates of calls as they occur in a month were grouped together and counted. The total of their counts is their observed frequencies. To calculate the expected frequencies when the arrival rate of calls is 7 calls per minute, the probabilities for when  $x = 3$  or less, 4, . . . , 12 or more, are determined first, and then multiplied by the sample size or sum of the observed frequencies which is equal to 21 in this study. Probabilities are calculated according to the formula, and probabilities cannot be less than 0 or more than 1. Expected values are theoretical results expected according to the rules of probability, given by sample size times probability. For example, for the probability of  $x = 3$  or less  $P(0 \leq x \leq 3)$  and sample size ( $n$ ) the expected frequency is calculated as in Equation (1) and for  $x = 4$  as in Equation (2).

$$\{P(0 \leq x \leq 3)\}n = \{P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3)\}n$$

$$= \left\{ \frac{\sum e^{-\lambda} \lambda^x}{x!} \right\} n = \left\{ \frac{e^{-7} 7^0}{0!} + \frac{e^{-7} 7^1}{1!} + \frac{e^{-7} 7^2}{2!} + \frac{e^{-7} 7^3}{3!} \right\} 21 = 1.7 \quad (1)$$

$$P(x = 4) = \left( \frac{e^{-7} 7^4}{4!} \right) 21 = 1.9 \quad (2)$$

Actually, the mean number of calls per minute varies according to the time of the day, and the day of the week. For the mean number of calls per minute = 7, the observed and expected frequencies are shown in Table 2 below.

TABLE 2. Observed and expected frequencies

$x$	3 or less	4	5	6	7	8	9	10	11	12 or more	
Observed	2	1	4	3	2	2	3	1	1	2	21
Expected	1	1.8	2.6	3.1	3.1	2.8	2.2	1.6	1	0.6	19.8

The number of calls, as they entered the telephone system per minute (arrival rate), was recorded. From these data (Table 1), the observed and expected frequencies for different arrival rates were determined. These frequencies are depicted in Figures 2-8 to show how different arrival rates of calls approximate a Poisson distribution.

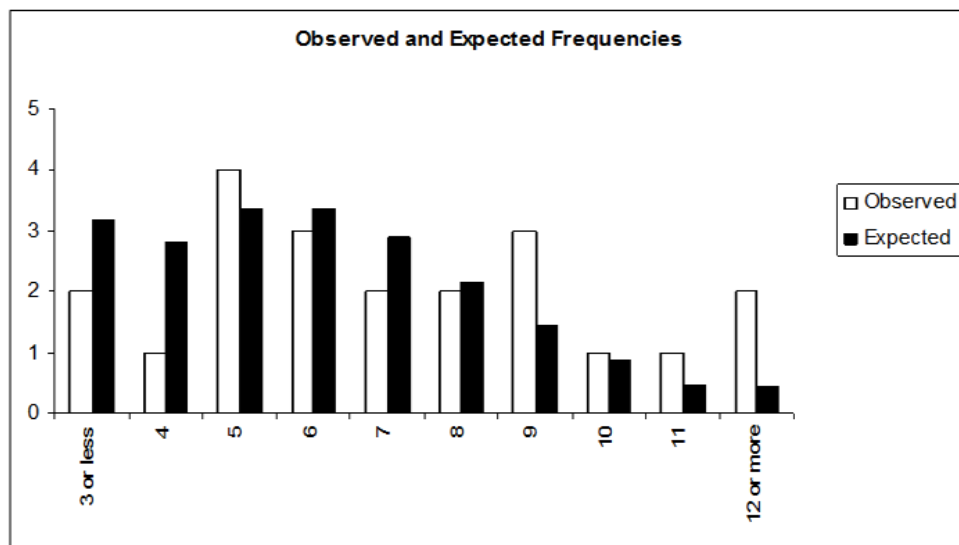


FIGURE 2. Chi-square for  $\lambda = 6$

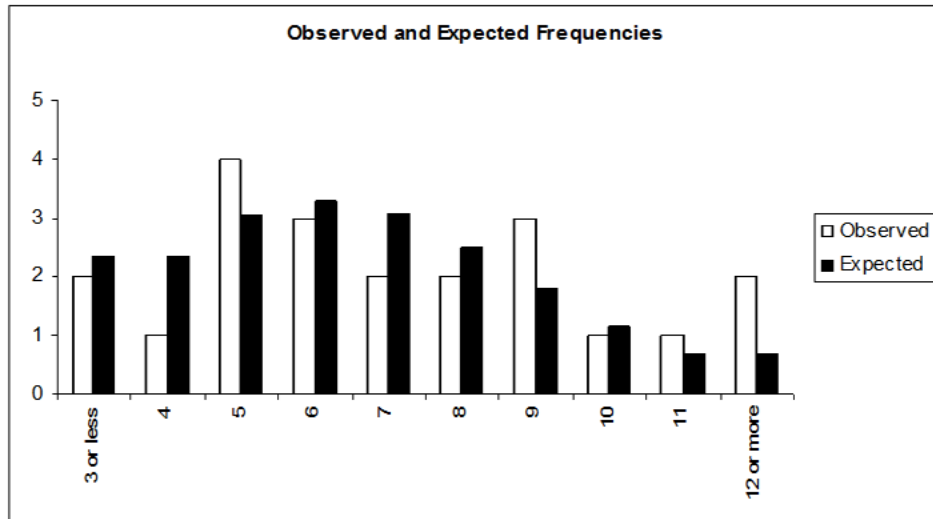


FIGURE 3. Chi-square for  $\lambda = 6.5$

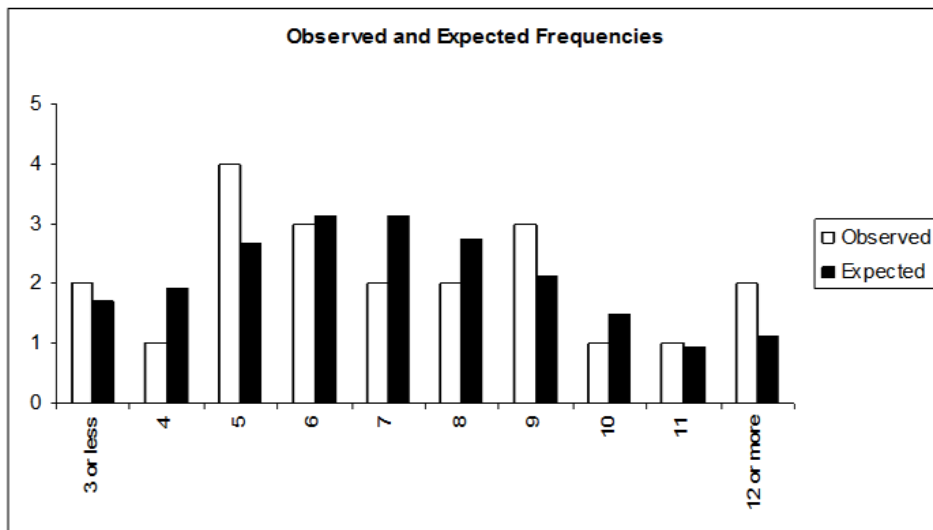


FIGURE 4. Chi-square for  $\lambda = 7$

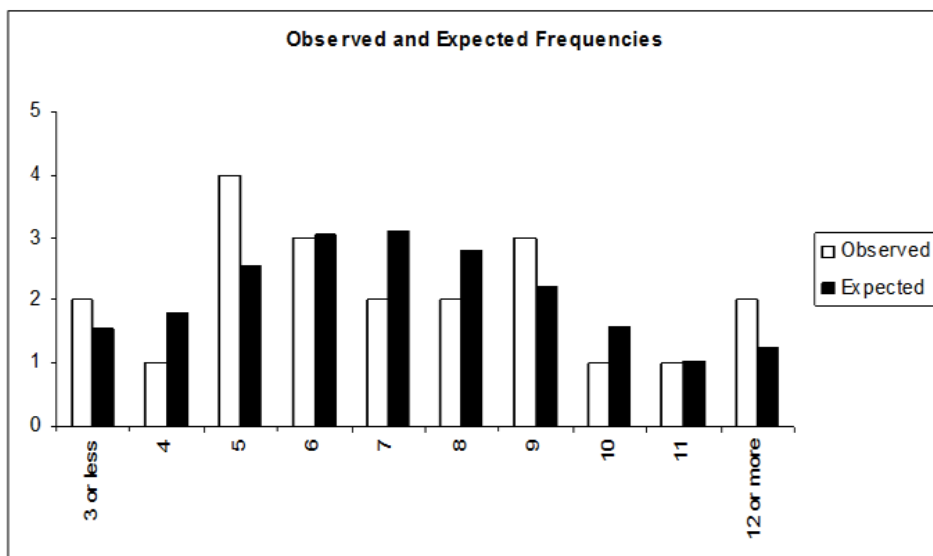


FIGURE 5. Chi-square for  $\lambda = 7.14$

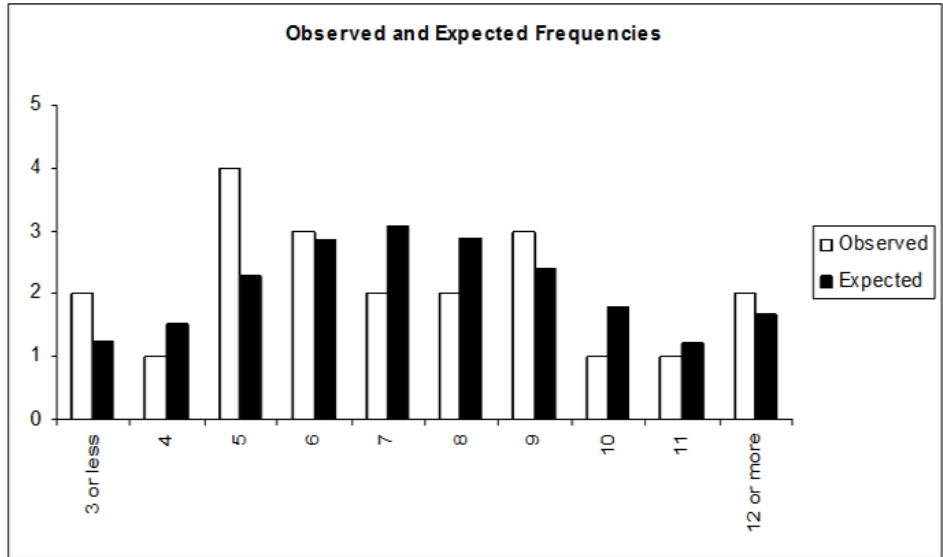


FIGURE 6. Chi-square for  $\lambda = 7.43$

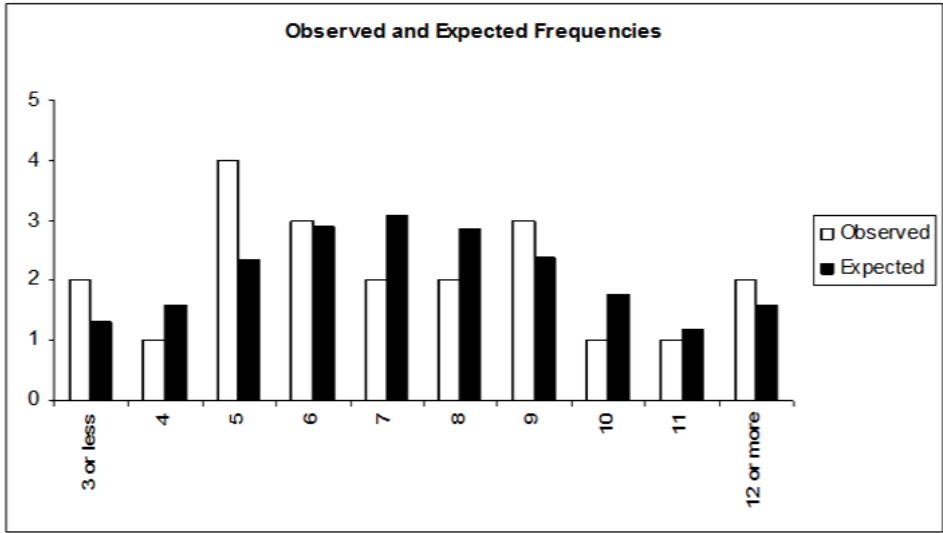


FIGURE 7. Chi-square for  $\lambda = 8$

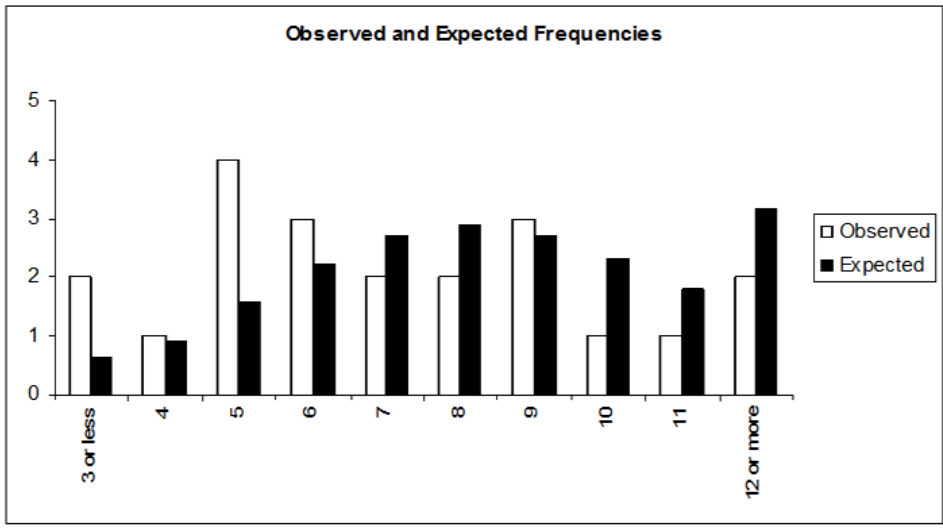


FIGURE 8. Chi-square for  $\lambda = 8.5$

**3.2. Sensitivity analysis.** The different values where the null hypothesis is not rejected at the 5% significant level below are part of a 95% confidence interval, so any of these values are a possibility in the future, given that the population is stationary. Table 3 below shows the computation of the chi-square for  $\lambda = 7$ , where  $\lambda$  is the rounded mean arrival rate. The chi-squares for different lambdas are shown in Table 4 below.

TABLE 3. Chi-square computation

	Probability	Expected	Observed	O-E	(O-E) <sup>2</sup>	(O-E) <sup>2</sup> /E
0, 1, 2, 3, 4, 5	0.301	6.3	7	0.7	0.4694	0.07433
6, 7, 8	0.428	9	7	-2	3.9842	0.44288
9 and more	0.271	5.7	7	1.3	1.7185	0.30207
	1	21	21	0	<b>Chi-square</b>	<b>0.81928</b>

TABLE 4. Sensitivity analysis

Lambda ( $\lambda$ )	$\chi^2$	Sensitivity
6	5.3	The null hypothesis is rejected.
6.5	2.0	The null hypothesis is not rejected.
7	0.8	The null hypothesis is not rejected.
7.14 (the expected value estimate)	0.8	The null hypothesis is not rejected.
7.43 (the variance estimate)	1.1	The null hypothesis is not rejected.
8	2.7	The null hypothesis is not rejected.
8.5	5.7	The null hypothesis is rejected.

These tests examined the hypothesis that the Poisson distribution approximates the sample data at 5% significance level with one degree of freedom. Since there are three classes, one degree is lost because  $\lambda$  is estimated from the data, and one is lost because of the frequency sum. The chi-square critical value is 3.841, obtained from a table of percentiles of the chi-square distribution with one degree of freedom (nr of classes – nr of parameters estimated – 1 = 3 – 1 – 1 = 1). One may also obtain this value, 3.841 using the CHIINV(0.05,1) function in Microsoft Excel. The conclusion is that the data do not present sufficient evidence to contradict the hypothesis that F possesses a Poisson distribution.

**3.3. The queuing model computations.** This queuing model involves a system in which no waiting is allowed. There are multiple service channels. Customers log events calling a telephone line. The calls arrive at the telephone system at an average rate of  $\lambda$ . The arrivals follow a Poisson probability distribution. There is an average rate of service  $\mu$  calls per minute at each channel. Arriving calls seek service from one of several service channels or each call is automatically switched to an open channel. If all channels are busy, arriving calls are denied access to the system. In waiting-line terminology, arrivals occurring when the system is full, are blocked and are cleared from the system. These calls are abandoned.

The optimal number of employees (channels) is determined by computing a steady state probabilities that  $j$  of the  $k$  channels will be busy. Formula (3) below is used to calculate these percentages (probabilities). The following equation applies.

$$P_j = \frac{\left(\frac{\lambda}{\mu}\right)^j / j!}{\sum_{i=0}^k \left(\frac{\lambda}{\mu}\right)^i / i!} \tag{3}$$

where  $\lambda$  = the mean arrival rate,  $\mu$  = the mean service rate for each channel and  $k$  = the number of channels.

$P_j$  = the probability that  $j$  of the  $k$  channels are busy for  $j = 1, 2, \dots, k$ . The important issues to determine here, are 1) the probability  $P_k$ , which is the probability that all the channels are busy. On a percentage basis,  $P_k$  indicates the percentage of arrivals that are blocked and abandoned; 2) and the average number of events in the system: this is the same as the average number of channels in use. If  $L$  denotes the average number of events in the system, then

$$L = \lambda/\mu(1 - P_k) \quad (4)$$

Whether the arrivals are indeed Poisson distributed, is determined (see above). There is an average arrival rate of 3360 calls per day. A day has 8 working hours; therefore, the rate is  $3360/8 = 420$  calls per hour. An hour has 60 minutes; therefore, the rate is  $420/60 = 7$  calls per minute. This means the arrival rate  $\lambda = 7$ . Currently 17 channels are responsible for answering or logging the calls. Each channel is expected to handle about 240 calls per day. A day has 8 working hours; therefore, the service rate is  $240/8 = 30$  calls per hour. An hour has 60 minutes; therefore, the service rate is  $30/60 = 0.5$  calls per minute, which is one call in two minutes. This means the service rate  $\mu = 0.5$ .

Since there are 17 channels, they cannot handle the workload as there is a high percentage (%) of abandoned calls daily. Using the above Formula (1), the probability that  $j$  of the  $k$  channels are busy (the percentage of abandoned calls) is calculated when seventeen channels are used as set out below: With  $\lambda = 7$  and  $\mu = 0.5$  we calculate the percentage of abandoned calls.

$$\begin{aligned} P_{17} = P_{abandoned} &= \frac{(7/0.5)^{17}/17!}{[(7/0.5)^0/0! + (7/0.5)^1/1! + \dots + (7/0.5)^{17}/17!]} \\ &= 85725.11796/994795.009 = 0.08617365. \end{aligned}$$

With only 8.61% of calls blocked with 17 channels, 91.39% of calls are answered. The service is then modelled with a different number of channels, but management has to select only from 17 channels upwards, to find out how many additional channels can be used. The percentages (probabilities) of abandoned calls are calculated with the mean arrival rate ( $\lambda = 7$ ) for a different number of channels, in Table 5. In Table 5, it is shown that when the number of employees (channels) increases, the probability (percentage) of abandoned calls decreases. For example, with 22 employees (channels), 1.23% of calls is abandoned, and with 25 employees (channels), 0.24% is abandoned.

As mentioned above, Formula (3) was used in a spreadsheet to model the abandoned calls' percentages (probabilities). The function FACT is the factorial function, for example,  $\text{FACT}(4) = 4 * 3 * 2 * 1 = 24$  and  $\text{FACT}(2) = 2 * 1 = 2$ . The symbol  $\wedge$  is the index meaning raised to the power of the value in the cell.  $\text{C10} = (\text{B30}/\text{C7})$ ;  $\text{B14} = (\text{\$B\$7}/\text{\$B\$8})^\wedge \text{A14}/\text{FACT}(\text{A14})$  and copied to B15 through to B38;  $\text{C14} = \text{SUM}(\text{B13};\text{C13})$  and copied to C15 through to C38;  $\text{D15} = (\text{B14}/\text{C15})$  and copied to D16 through to D38;  $\text{C7} = \text{SUM}(\text{B12};\text{B29})$ .

Table 6 shows the different abandoned rates of calls with 17 and 25 employees on duty for values of lambda ( $\lambda$ ) within the 95% confidence interval.

**4. Conclusion.** The results of all the tests done and all the calculations made, suggest that more staff be hired for the service desk in order to improve the service by managing the workload, thereby satisfying customers. To provide an excellent customer service, with seldom more than one or two customers in a queue means retaining a large staff which may be costly. An unlimited number of employees cannot therefore be appointed since this would not be cost effective. Managers must deal with the trade-off between the cost of providing excellent service and customer satisfaction.



TABLE 5. Abandoned calls % for different number of channels

C7 $f_x = \text{SUM}(B12:B29)$			
	A	B	C
7	Arrival rate( $\lambda$ )	7	994795.0086
8	Service rate( $\mu$ )	0.5	
9	Employees(channels( $n$ ))	17	
10		$P_{17}$	0.086173651
11	number of employees(channels( $n$ ))	$(\lambda/\mu)^n/n!$	cumsum( $n-1$ )
12	0	1	1
13	1	14	0.933333333
14	2	98	0.867256637
15	3	457.3333333	0.801870251
16	4	1600.666667	0.737294641
17	5	4481.866667	0.673674506
18	6	10457.68889	0.61118348
19	7	20915.37778	0.550029307
20	8	36601.91111	0.490459176
21	9	56936.30617	0.432764594
22	10	79710.82864	0.377284754
23	11	101450.1455	0.324406763
24	12	118358.5031	0.274560423
25	13	127463.0034	0.228204766
26	14	127463.0034	0.185803518
27	15	118965.4698	0.147787763
28	16	104094.7861	0.114506912
29	17	85725.11796	0.086173651
30	18	66675.09174	0.062813914
31	19	49129.01497	0.044236497
32	20	34390.31048	0.030035483
33	21	22926.87365	0.019630579
34	22	14589.82869	0.012338058
35	23	8880.765288	0.00745414
36	24	5180.446418	0.004329423
37	25	2901.049994	0.002418613
38	26	1562.103843	1199468.39

TABLE 6. Abandoned rate of calls

Arrival rate per minute Lambda ( $\lambda$ )	Abandoned rate % with 17 employees	Abandoned rate % with 25 employees
6.5	6.1	0.10
7	8.61	0.24
7.14 (the expected value estimate)	9.35	0.30
7.43 (the variance estimate)	10.92	0.45
8	14.16	0.93

As future work, more complex service systems will be analyzed using simulation. How simulation may help to show relationship between quality of service and customer satisfaction, as well as how these are connected to staffing issues or to workload forecasting and handling will be investigated.

**Acknowledgements.** This work was supported by the University of South Africa, College of Economic and Management Sciences Research Department.

**REFERENCES**

- [1] J. A. Rice, *Mathematical Statistics and Data Analysis*, 3rd Edition, Thomson/Brooks/Cole, Belmont, 2007.
- [2] X. J. Tang and Y. Todo, A study of service desk setup in implementing IT service management in enterprises, *Technology and Investment*, vol.4, pp.190-196, DOI: 10.4236/ti.2013.43022, 2013.
- [3] O. O. Sawyerr, S. Srinivas and S. Wang, Call center employee personality factors and service performance, *Journal of Services Marketing*, vol.23, no.5, pp.301-317, DOI: 10.1108/08876040910973413, 2009.
- [4] J. Hogan and B. Holland, Using theory to evaluate personality and job-performance relations: A socioanalytic perspective, *Journal of Applied Psychology*, vol.88, no.1, pp.100-112, DOI: 10.1037/0021-9010.88.1.100, 2003.
- [5] M. E. Hanna, B. Render, R. M. Stair and T. S. Hale, *Quantitative Analysis for Management*, 13th Edition, Pearson Education Limited, Edinburgh, England, 2018.
- [6] H. A. Taha, *Operations Research: An Introduction*, 8th Edition, Macmillan, New York, 2007.
- [7] J. van Ommeren, A. Al Hanbali and R. Boucherie, Analysis of polling models with a self-ruling server, *Queueing Systems*, DOI: 10.1007/s11134-019-09639-6, 2019.