BENCHMARKING IT SERVICE REGIONS

VICTORIA G MADISA

Hons. B.Sc.

Dissertation submitted in the School of Modelling Sciences of the North-West University in partial fulfilment of the requirements for the degree

MAGISTER SCIENTIA

Supervisor: Prof PD Pretorius

VANDERBIJLPARK
NOVEMBER 2008

ACKNOWLEDGEMENTS

A special word of thanks and appreciation to the following people who made it possible for me to present this study in its final format.

-Prof. P.D. Pretorius, whose comments directed the research, I thank him for his patience, guidance, and motivation.

-Dr J.C Huebsch for the professional assistance in proofreading the study material.

-My husband,William for his loving support and truly believing in me.

-My children, Tebogo, Tebatso, Thabang and Thabiso for their patience and giving me many hours of solitude to work.

-All my friends and relatives for their patience.

Most importantly, my thanks to Almighty God who made this possible.

ABSTRACT

Productivity and efficiency are the tools used in managing performance. This study researches and implements best practices that lead to best performance. A customer quality defined standard has to be created by benchmarking the Information Technology Service Regions which may be used to help decision-makers or management make informed decisions about (1) the effectiveness of service systems, (2) managing the performance of Information Technology Service Regions.

Waiting lines or queues are an everyday occurrence and may take the form of customers waiting in a restaurant to be serviced or telephone calls waiting to be answered. The model of waiting lines is used to help managers evaluate the effectiveness of service systems. It determines precisely the optimal number of employees that must work at the centralised service desk.

A Data Envelopment Analysis (DEA) methodology is used as a benchmarking tool to locate a frontier which is then used to evaluate the efficiency of each of the organizational units responsible for observed output and input quantities. The inefficient units can learn from the best practice frontier situated along the frontier line.

OPSOMMING

Produktiwiteit en effektiwiteit is die gereedskap wat in die bestuur van prestasie gebruik word. Hierdie studie vors die beste praktyke wat na beste prestasie ly na, en implementeer dit. 'n Kliënt kwaliteitsgedefinieerde standard daargestel word om sodoende die diensvelde van die inligtingstegnologie te bepaal. Dit word gedoen om besluitnemers of bestuur te help om ingeligte besluite te neem, eerstens oor die effektiwiteit van die diensstelsels, en tweedens oor die prestasie van die diensvelde van die inligtingstegnologie.

Waglyne of rye is 'n daaglikse gebeurtenis en mag voorkom as 'n ry van klante wat in 'n restaurant vir diens wag of telefoonoproepe wat wag om beantwoord te word. Die model van waglyne word gebruik om bestuurders te help om die effektiwiteit van diensstelsels te evalueer. Dit bepaal presies wat die optimale aantal van werknemers is wat by die gesentraliseerde dienstoonbank moet werk.

'n Data-insluitings analise metode (Data Evelopment Analysis – DEA) word gebruik as 'n maatstaf om 'n grenslyn op te spoor, welke grenslyn dan gebruik word om die effektiwiteit van elk van die organisatoriese eenhede vir die waargenome uitset en inset-hoeveelhede te bepaal. Die ontoereikende eenhede kan dan van die beste praktyk-grenslyn leer.

TABLE OF CONTENTS

CHAPTER 1 OVERVIEW

CHAPTER 2 DATA ANALYSIS

CHAPTER 3 DATA ENVELOPMENT CONCEPTS

CHAPTER 4 RATIO ANALYSIS

CHAPTER 5 LINEAR PROGRAMMING

CHAPTER 6 DATA ENVELOPMENT ANALYSIS

CHAPTER 7 DISCUSSION OF RESULTS

CHAPTER 1

OVERVIEW

1.1 INTRODUCTION

Organisations nowadays have achieved efficiency and quality by emphasizing customer focus and employee participation. (DiBella and Nevis, 1998). By customer focus, it is meant that the company should try by all means to satisfy customer needs. By employee participation, it is meant that all employees must know and share company goals and must team up and do everything possible to achieve these goals.

In order to deliver and maintain quality service, or process efficiency, the company has to engage in continual learning. (DiBella and Nevis, 1998). Employees have to add value. Value-added means those activities or steps that add to or change a service as it goes through a process. These are the activities or steps that clients view as important and necessary.

DiBella and Nevis (1998:7) state as follows. *"Organisations do not operate at peak performance but are or should be in a continual state of becoming something more than or different from what they are at present. The implication is that there are dysfunctional aspects of organisations that limit their effectiveness or performance. The role of organisational learning is to help organisations overcome these limits and become something more."*

Briefly, the ability to learn faster than your competitors or to learn best practices, is the best strategy to keep ahead of them.

1.2 PROBLEM STATEMENT

The company in which the research takes place, is a telecommunication company with its headquarters situated in Pretoria, South Africa. Within this company there are divisions that deal solely with Information Technology services. These divisions are called Information Technology (IT) Service delivery regions and are situated countrywide: in Pretoria, Johannesburg, Durban, Bloemfontein/Kimberly (Bloem/Kby), and Cape Town. These regions are responsible for the execution of their operational responsibilities. In these service delivery regions, the main common function is to support both computer hardware and software. The research concentrates on the problems arising in these divisions.

Firstly, customers from all regions in South Africa report events by phoning a centralised service desk in Pretoria. An event is anything that an end-user finds 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 the network, are all requests. The reinstallation of software and fixing computer hardware are faults. End-users are people using computer services. An Information System is an arrangement of people, data, processes, information presentation, and information technology that interacts to support and prove day to day operations in a business as well as support the problem solving and decision making needs of management and users. (Whitten *et al.*, 2006).

The time required to service the 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 a server servicing customers or 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 percentage of abandoned calls is high.

The second problem deals with resolving the reported problems or logged events. These logged events are routed to their respective regions by the service desk to be attended to. The success in producing as large as possible an output (number of resolved events and satisfied clients) from a given set of inputs (employees or labour) is not achieved. Customers complain that their logged events are not resolved within the specified Service Level Agreement (SLA). The SLA is an agreement on performance System Metrics (Application Availability in Production, Average Request Resolution and Average Fault Resolution). The agreement stands, that a logged fault should be resolved within two days. A logged request should be resolved in four days. Customers wait a longer time before they can work on their computers again. It is imperative that these problems be addressed and precise solutions be found in order to satisfy customers.

1.3 THE RESEARCH GOAL

The purpose of this study is to research and implement best practices that lead to best performance. It researches the queuing methodology that can design a system that achieves the desired performance level by determining the minimum number of service channels that should be used at the service desk in a cost effective way. It finds out more about Data Envelopment Analysis methodology as a benchmarking tool, and as a threefold methodology (ratio analysis, Linear programming and DEA's relative efficiency). It finds out about the relationship between these methodologies. It applies the DEA's ratio analysis methodology, and the DEA's Linear programming methodology to the practical problem. It also applies DEA to evaluate the efficiencies of regions at once. The aim of the three methodologies is to find the best performer.

Benchmarking can be dealt with in many different ways, in marketing, economics and management as examples. This study deals with benchmarking in management. Benchmarking or best practices are ways of carrying out a function that makes a significant difference in the quality of output. This brings down costs, increases customer satisfaction, or improves process. Glen Peters (1994:9) defines benchmarking as following. *"Benchmarking is about improving competitive position, and using best practices to stimulate radical innovation rather than just seeking minor, incremental improvements on historic performance".*

1.4 THE RESEARCH METHODOLOGY

A decision support methodology is used. Queuing theory is used to help employ adequate staff at the service desk. Under the Queuing theory, the chi-square distribution is used to determine whether the arrival rates (observed frequencies) depart significantly from the expected frequencies. Expected frequencies are theoretical results expected according to the rules of probability. Data Envelopment Analysis (DEA) is employed to let data speak for itself, to display the regions where efficiency is attained and those where efficiency is not attained. Linear Programming is used as part of DEA because DEA is Linear-Programming based. According to Charnes *et.al.*, (1994) the DEA model has a Linear Programming (LP) formulation. As any LP, it has two versions, the primal and the dual. In DEA these are known as the ratio formulation and the envelope formulation. Regression Models are used to help effect the solution to the current problem of resolving all the events in order to satisfy clients.

1.4.1 QUEUING THEORY

Queuing Theory had its beginning in the research work of a Danish engineer named A.K. Erlang. The three components of the queuing process are the arrival rate, the queue and the service rate. The arrival rate refers to the rate at which the calls arrive at the service desk. For instance, a call or two calls arriving every minute describes the arrival rate. According to Taha (2007) a queue is created in the following manner. When a customer arrives in the system, he or she joins a waiting line. An employee chooses a customer from the waiting line to begin service. Upon the completion of a service, the process of choosing a new waiting customer is repeated. The service rate refers to how long it takes the server at the servicing channel to service a customer.

If the average time a customer waits in the queue is denoted by W_q, and the average customer arrival rate in the queue by λ, a generalized equation applying to queuing model is $L_q = \lambda W_q$, 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 (Render *et.al.*, 2006).

The following assumptions are used in the queuing model. (1) The queuing environment has either a finite or infinite calling population, and a multiple or single channel facility is used. (2) The arrival time is unpredictable and described by a Poisson distribution, or is predictable. (3) The service times (processing rate at the servicing facility) are unpredictable and exponential or the exact amount of processing time is known. (4) The queue lengths are infinite or finite. (5) All units wait in the single queue. (6) Service is on a first-come first-service basis. (7) All arriving events enter the queue. (Hall, 1993).

1.4.2 CHI-SQUARE DISTRIBUTION

If an experiment has only two outcomes, such as the appearance of a head or a tail in a tossing of a coin, the normal distribution can be used to determine whether the observed frequencies of these two events depart significantly from the expected frequencies. When more than two events occur, the normal distribution can no longer be applied to test for a possible significant difference between the observed and expected frequencies but a chi-square distribution is applied. The chi-square is defined as

$$\chi^2 = \sum_j (O_j - E_j)\text{^}2 / E_j$$

where the O_j's and the E_j's are the observed and the expected frequencies respectively. The closer the agreement between the expected and observed frequencies, the smaller will be the value of χ^2. If $\chi^2 = 0$, each of the terms of the sum in the above formula must be zero, and there is a perfect agreement between the observed and the expected frequencies for all events. (Alder and Roessler, 1975)

1.4.3 LINEAR PROGRAMMING

According to Anderson *et al.*, (2006:15) *"Linear Programming is a problem-solving approach that has been developed for situations involving maximizing or minimizing a linear function subject to linear constraints that limit the degree to which the objective can be pursued".* A Linear Programming Model can be defined as a mathematical model where all the functional relations are linear. For example, a linear function in $x_1, x_2, \ldots, x_n$ is of course a function of the form $a_1x_1 + a_2x_2 + \ldots + a_nx_n$.

Linear Programming (LP) was conceptually developed before World War 2 by the outstanding Soviet mathematician, A.N. Kolmogorov. Linear Programming is a technique that helps in resource allocation decisions. In the past 50 years, LP has been applied extensively to military, industrial, financial, marketing, accounting, and agricultural problems. Even though these applications are diverse, all LP's have four properties in common. (1) Problems seek to maximize or minimize an objective. (2) Constraints limit the degree to which the objective can be obtained. (3) There must be alternatives available. (4) Mathematical relationships are linear (Render *et. al.*, 2006).

1.4.4 DATA ENVELOPMENT ANALYSIS METHODOLOGY

Data Envelopment Analysis (DEA) is a non-parametric estimation method which involves the application of mathematical programming to observed data to locate a frontier which can then be used to evaluate the efficiency of each of the organizational units responsible for observed output and input quantities.

The DEA methodology as discussed by Charnes, Cooper, Lewin and Seiford (1994), is used to evaluate single input, single output production, and the relative efficiency of a set of Decision-making Units (DMU's). This term "DMU's" was coined by Charnes *et.al.*, to describe homogeneous units, each utilising a common set of inputs to produce a common set of outputs. Examples of homogeneos DMU's are a collection of similar firms, departments, group of schools, hospitals and bank branches. A bank branch and a supermarket are not homogeneous units. In this study's perspective, DEA is used to evaluate the efficiency of IT service delivery regions which are denoted as region 1 to region 5, (DMU1 to DMU5), which also are homogeneous with some decision autonomy. Each region consumes one input and produces two outputs. A DEA model is developed that uses these factors (input and outputs) to compute the efficiency degree of a particular region when this region is compared with all the other regions. The regions that are considered efficient, belong to the frontier and, therefore, they can be used as performance benchmarks to study the regions that are operating inefficiently (Charnes *et.al.*, 1994).

1.4.5 REGRESSION ANALYSIS

Regression Analysis is a statistical forecasting model, that is concerned with describing and evaluating the relationship between a given variable (usually called the dependent variable) and one or more other variables (usually called the independent variables). Regression Analysis can predict the outcome of a given key business indicator (dependent variable) based on the interactions of other related business drivers (explanatory variables).

1.5 DEFINITION OF TERMS

A list of terms and concepts used in this study, appears in Appendix A. Tables appear in Appendix B. Articles about this study appear in Appendix C.

1.6 OVERVIEW OF THE CHAPTERS TO FOLLOW

Chapter 2 gives a summary of how data were collected and continues defining the variables used to solve the problem in this study, which the resolution thereof was, to hire more staff. It then applies descriptive statistics to explore data, and to confirm that data collected from the service desk do indeed approximates a Poisson distribution. The data from the service desk are used for analysing queuing of calls at the service desk. Chapter 3 briefly defines and explains the terms and concepts required in the interpretation of the results in the application of DEA methodology. It lays out the theoretical framework in which DEA concepts can be interpreted. In chapter 4, DEA's ratio analysis is used to evaluate the efficiency of Decision-making Units, referred to as "regions" in this study. In this chapter, the single input, single output case was evaluated, and thereafter single input, two outputs case evaluated. Chapter 5 illustrates the solution of a linear program manually and thereafter the solution determined using QM for Windows software.In other words, this chater uses DEA's linear programming analysis to evaluate the efficiency of the regions. Chapter 6 illustrates the relationship between DEA and the Linear Programming methodologies.It applies the DEA methodology to evaluate the efficiency of DMU's (regions). Excel's solver (software) is used in this regard. Discussion of results and final conclusions are discussed in chapter 7.

1.7 CONCLUSION

With this introductory chapter, it is assumed that the reader has now the overview of what is going to be discussed in the forthcoming chapters.

CHAPTER 2

DATA ANALYSIS

2.1 INTRODUCTION

Dorian (1999) views data representation from two perspectives, as data and as a data set. The terms “data” and “data set”, according to him, are used to describe the different ways of looking at the representation. “Data” implies that the variables are to be considered as individual entities and their relationship with other variables are secondary. “Data set” implies that not only variables are considered, but also their interrelationship with other variables.

2.2 DATA COLLECTION

There are two sets of data, namely data collected from the service desk and data collected from the event management system database.

2.2.1 DATA FROM THE SERVICE DESK

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. Here, (1) the number of calls, as they entered the telephone system per minute (arrival rate), were recorded. This was actually easy since 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. Channels refer to employees. This is the data used for the Queuing Model.

2.2.2 DATA FROM THE EVENT MANAGEMENT SYSTEM

From an event management system database, and for each month, in a year, (1) the date on which the event was reported and the date on which the event was resolved, were recorded. These dates were used to determine the ratio of the events resolved to the total number of events logged per month. For example, registering the first of January five times under logged events means five events are logged, and registering the first of January three times under resolved events means that three events are resolved. Determining the ratio will then be 3/5. (2) The average number of events resolved per month in a year, was determined. The data used here, are for twelve months, starting from February 2004 to January 2005.

2.3 VARIABLE DEFINITION

The table below shows the variables used in the practical problem discussed in the next chapters.

Table 2.1 Variables

Variables	**Type**	**Description**
Number of resolved events.	Output	Number of faults and requests that are resolved.
Client Satisfaction.	Output	Ratio of the number of resolved events to the total number of logged events.
Employees.	Input	Number of employees.

The number of resolved events and client satisfaction are regarded as outputs. Employees are regarded as input. Client Satisfaction was determined as the ratio of the number of resolved events to the total number of logged events. The average inputs and outputs per month for a year for the five regions are as given in the following table.

Table 2.2 Inputs and Outputs

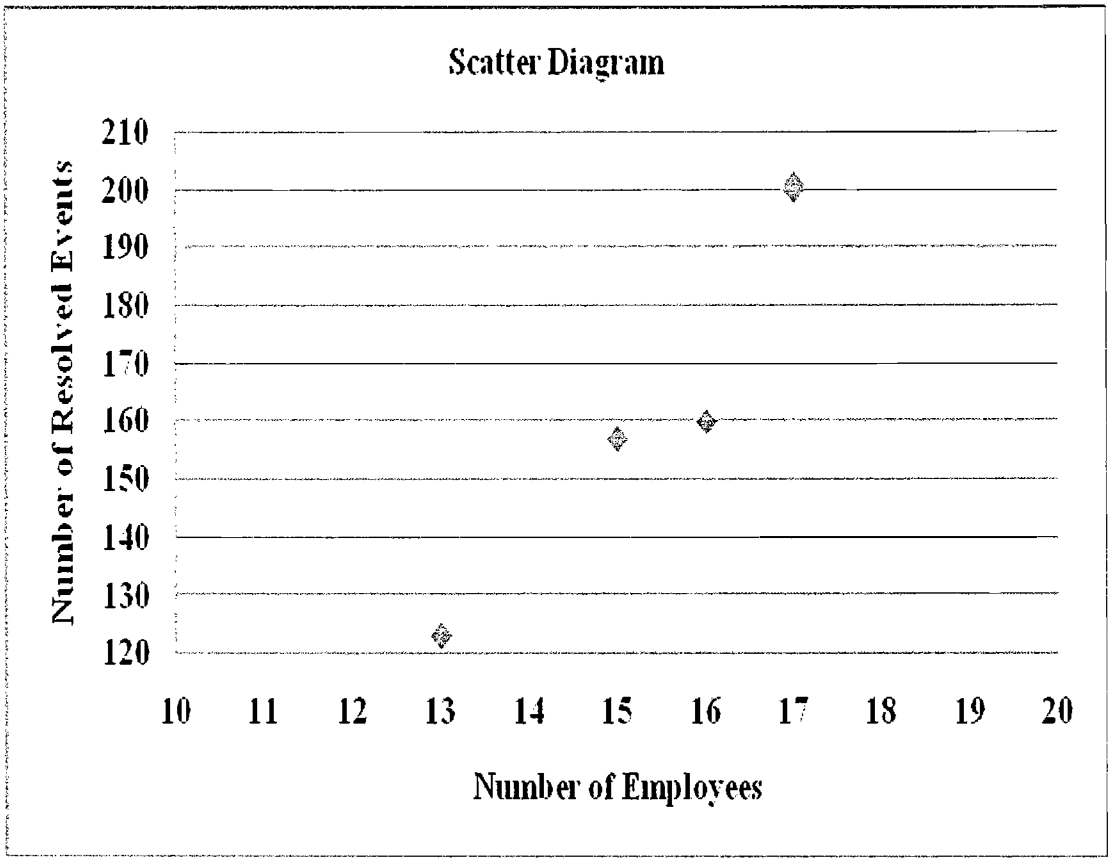
Region	Number of Employees	Number of Resolved Events	Client Satisfaction
	Input	Output	Output
1	17	201	0.66
2	16	160	0.86
3	15	157	0.79
4	17	200	0.67
5	13	123	0.62

Source: (Event Management System, 2004)

2.4 ASSESSING SAMPLE INDEPENDENCE

To make the deduction that the observations are independent or not, scatter diagrams are made use of. The scatter diagram of the observations x_1, $x_2,\ldots,x_n$ is a plot of the pairs (x_i; x_i+1). If the x_i's are independent, one would expect the points (x_i; x_i+1) to be scattered randomly over the area of the plot. If the x_i's are positively correlated, the points (x_i; x_i+1) will tend to lie along a line with a positive slope. If the x_i's are negatively correlated, then the points (x_i; x_i+1) will tend to lie along a line with a negative slope. The graph of the relationship between employees and resolved events is depicted below.

Figure 2.1 Employees and Resolved Events



According to the graph above, there is a linear relationship between the number of employees and the number of resolved events. The correlation coefficient is calculated and is equal to 0.960757. This is a strong positive relationship between the number of employees and the number of resolved events. The relationship is directly proportional. This means that, as the number of employees increase, so does the number of resolved events.

2.5 UNDERSTANDING DATA

Description of Summaries and Visualisation according to Two Crows Corporation (2007) is as following. *"Before you can build good models, you must understand your data. Start by gathering a variety of numerical* *summaries (including descriptive statistics such as averages, standard deviations and so forth) and looking at the distribution of the data.*

Graphing and visualization tools are vital aids in data preparation and their importance to effective data analysis cannot be overemphasized. Data visualization most often provides the "Aha!", leading to new insights and success. Some of the common and very useful graphical displays of data are histograms or box plots that display distributions of values".

The task of hypothesizing a distribution family from observed data, is somewhat unstructured. Three categories are used to aid in making a decision as to what distribution the observed data resembles. These are the the computation of the summary statistics (particularly the mean and the variance), the chi-square test, and the histogram of frequency distribution of calls. Table 2.3 shows the arrival rate of calls for a month, and this was obtained as explained in section 2.2.1.

Table 2.3 Arrival Rate of Calls

Days of the month	3	4	5	6	7	10	11	12	13	14	17	18	19	20	21	24	25	26	27	28	31
Calls per minute	8	5	3	5	3	11	7	7	5	5	9	9	6	6	6	9	10	8	4	12	12

Source: (Service Desk, 2004)

2.5.1 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. 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 be at least 5, so that the chi-square test can be used. The expected frequencies are computed on the basis of a hypothesis $H_{o.}$ H_o: 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 χ^2 is smaller than some critical value (such as $\chi^2_{0.95}$) 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 χ^2 is the chi-square. The O_j's are the observed frequencies and E_j's are the expected frequencies. The data in Table 2.3 above were used to determine the observed frequencies in Table 2.4 below. Similar arrival rates of calls as they occur in a month were grouped together and counted. The total of their counts is their frequencies, 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, but probabilities cannot be more than 1. It should actually be expected values. Expected values are theoretical results expected according to the rules of probability. For example, for x = 3 or less and sample size = 21 the expected frequency is calculated as follows.

$$
\begin{aligned}
P(0 \le x \le 3) &= P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3) \\
&= \sum e^{-\lambda}\lambda^x / x! \\
&= (e^{-7}7^0/0! + e^{-7}7^1/1! + e^{-7}7^2/2! + e^{-7}7^3/3!)\ 21 \\
&= 1
\end{aligned}
$$

And for x = 4 the expected frequency is calculated as follows.

$$P(x = 4) = (e^{-7}7^4/4!)21$$
$$= 1.8$$

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, that is, arrival rate (λ) = 7, the observed and expected frequencies are shown in Table 2.4 below.

Table 2.4 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 observed and expected frequencies for different arrival rates (λ) are depicted in figure 2.2 (a-g) below to show how different arrival rates of calls approximate a Poisson distribution.

Figure 2.2 (a) Chi-square for λ = 6

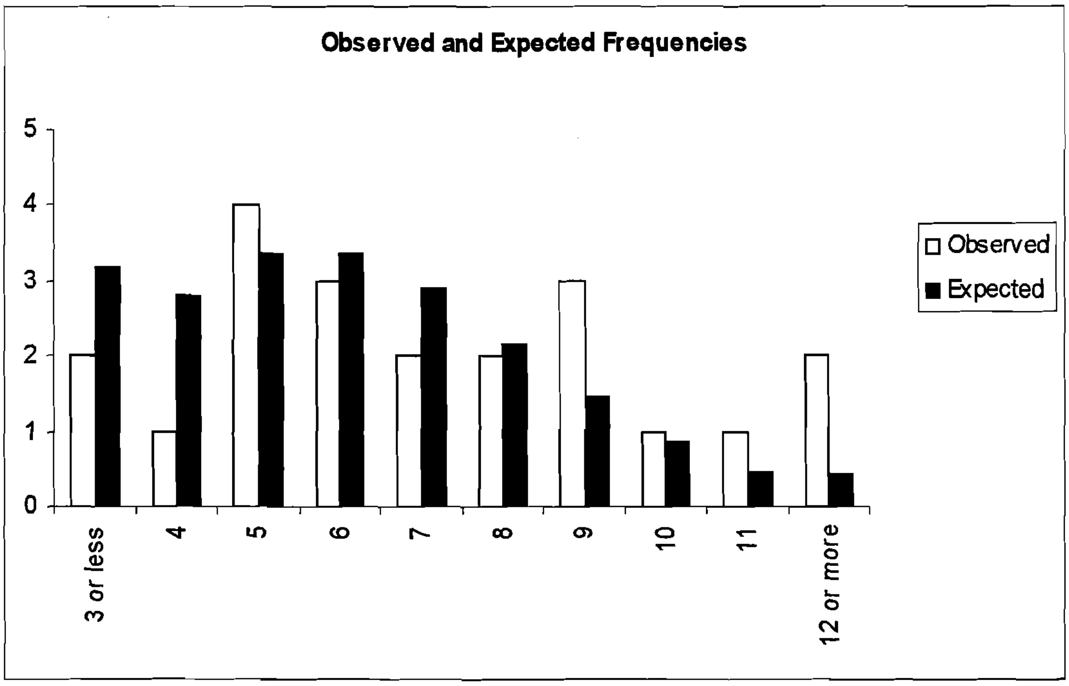


Figure 2.2 (b) Chi-square for λ = 6.5

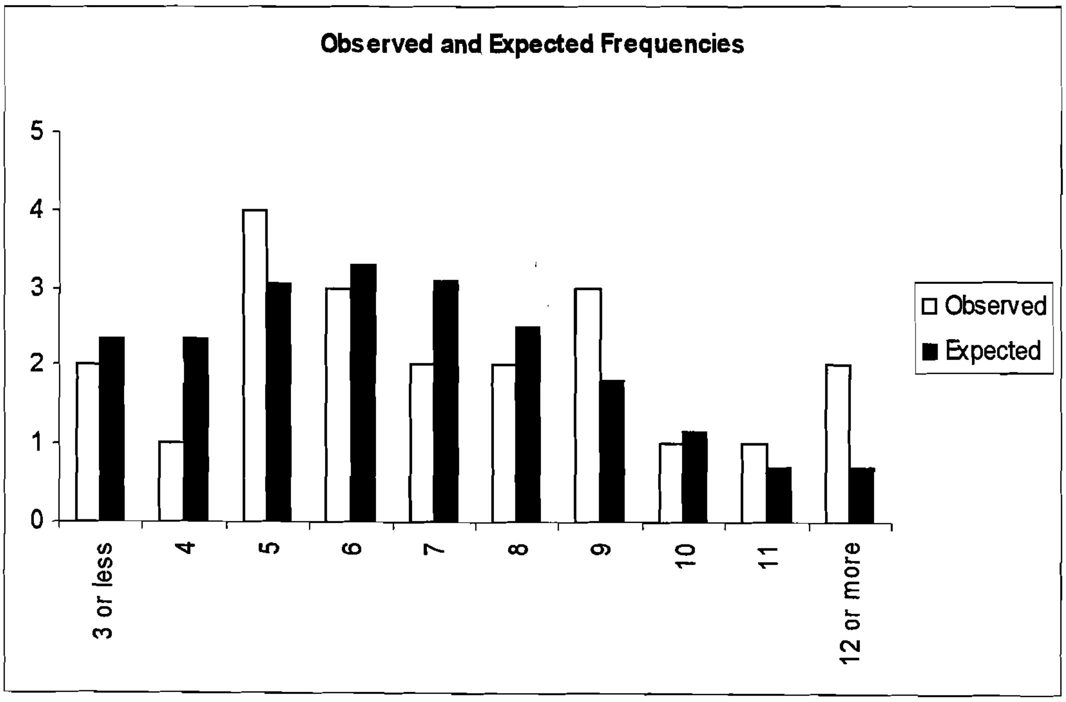


Figure 2.2 (c) Chi-square for λ = 7

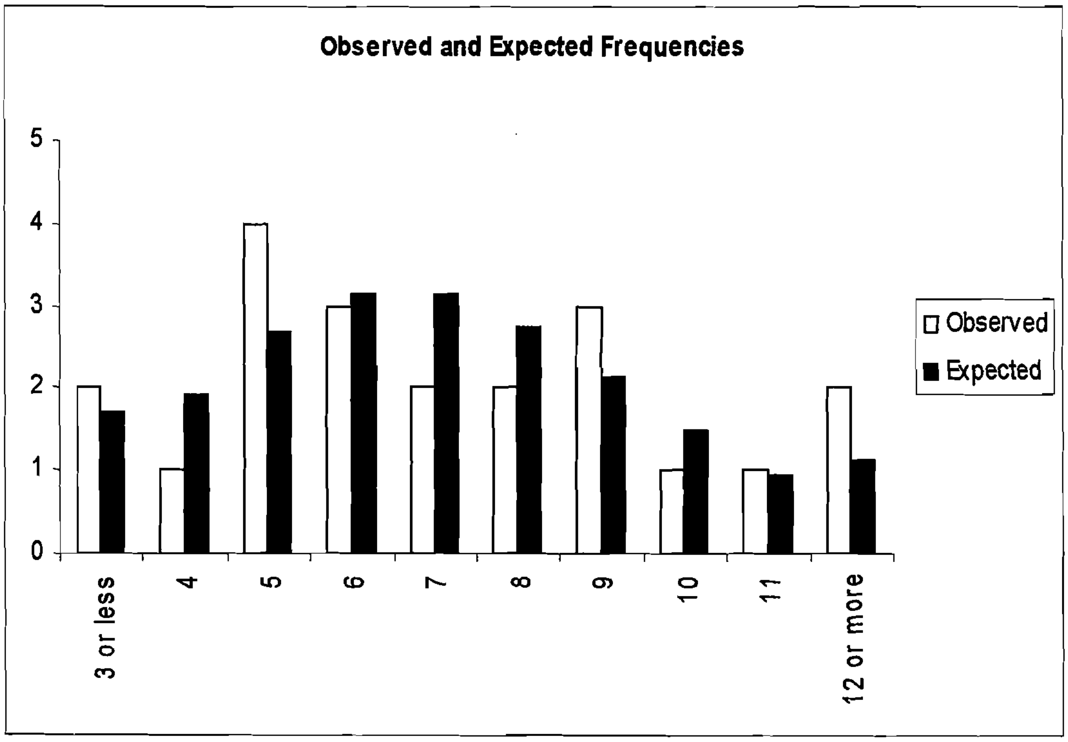


Figure 2.2 (d) Chi-square for λ = 7.14

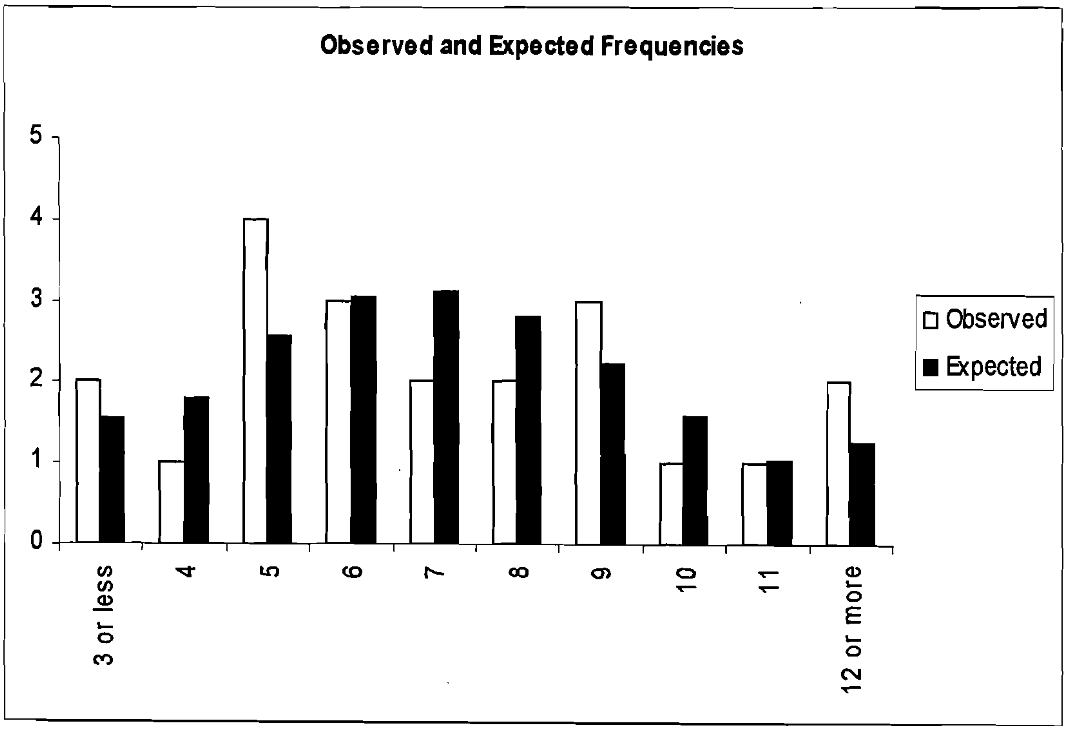


Figure 2.2 (e) Chi-square for $\lambda = 7.43$

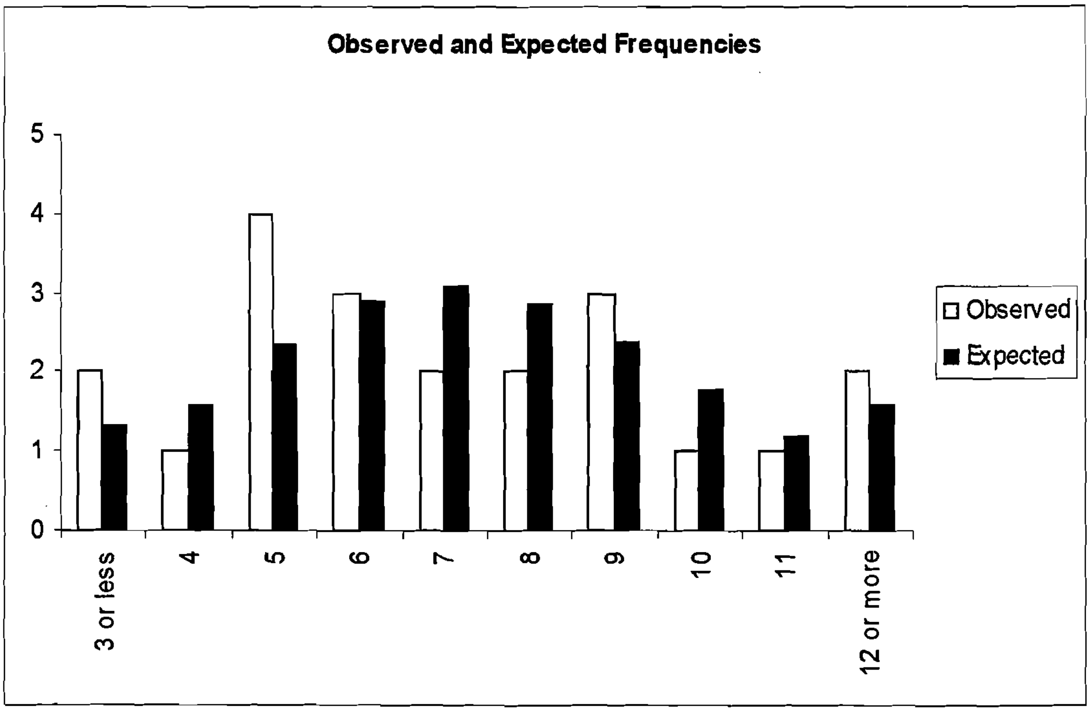


Figure 2.2 (f) Chi-square for $\lambda = 8$

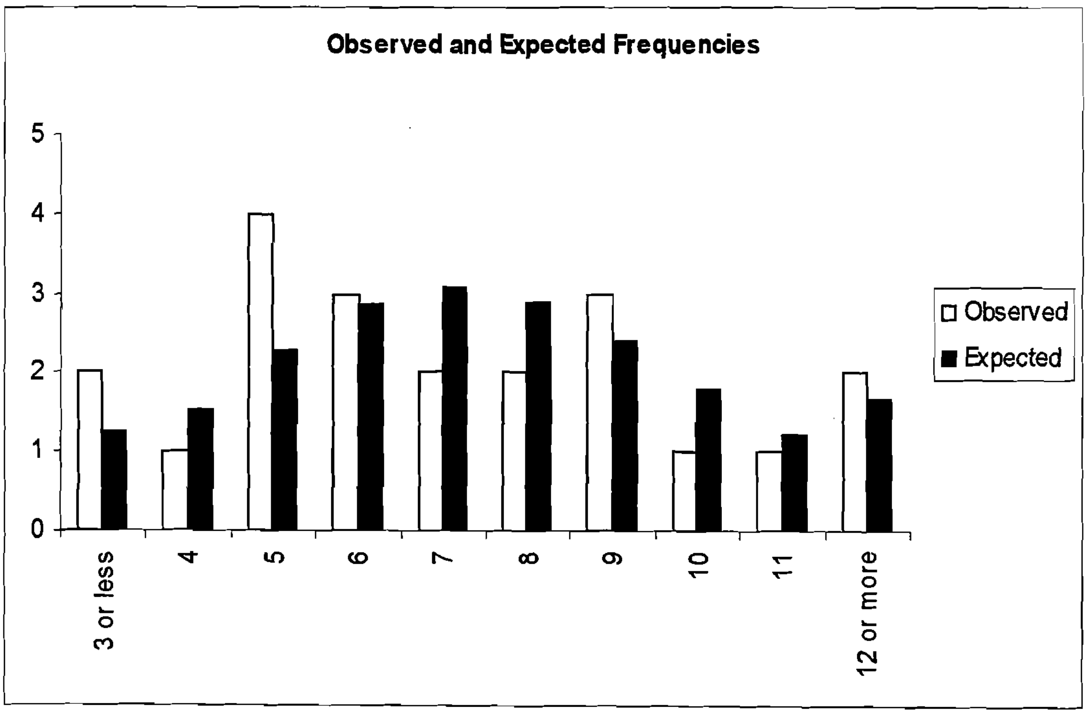
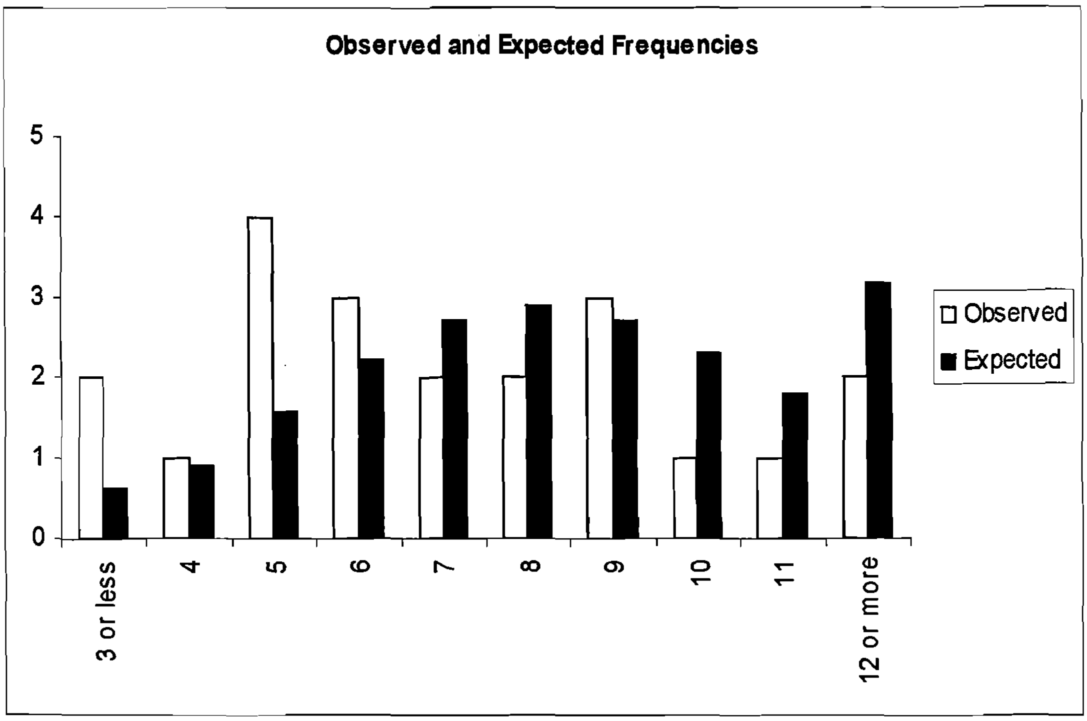


Figure 2.2(g) Chi-square for λ = 8.5



For λ = 6 and λ = 8.5, the null hypothesis is rejected. These do not constitute the 95% confidence interval and even the shape of their histograms manifest this. The test confirmed the hypothesis that the Poisson distribution approximates the sample data at 5% significance level with one degree of freedom. One degree of freedom because, since there are three classes, one degree is missed because λ is estimated and one is lost because of the frequency sum. The chi-square critical value is 3.841.

2.5.2 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 2.5 below shows the computation of the chi-square for λ

=7, where λ is the rounded mean arrival rate.The chi-square values for different lambdas are shown in Table 2.6.

Table 2.5 Chi-square Computation

		Expected	Observed	O-E	(O-E)^2	(O-E)^2/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	**Chi-square**	**0.81928**

Table 2.6 Sensitivity analysis

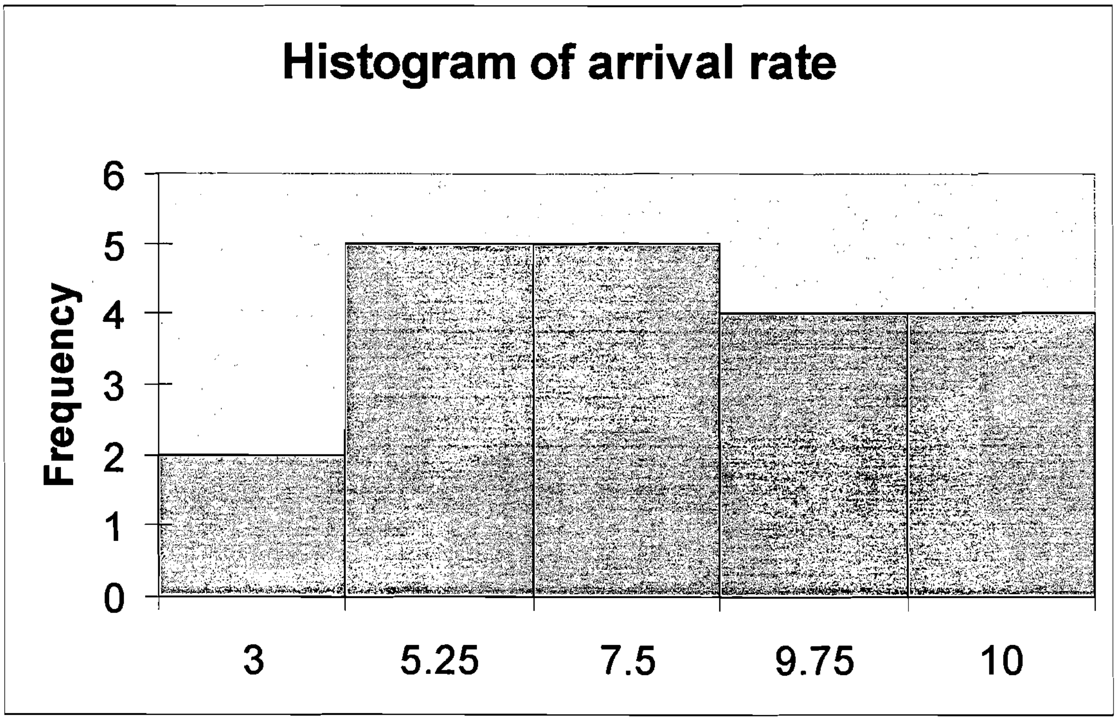
Lambda (λ)	χ^2	**Sensitivity**
6	5.3	The null hypothesis is rejected.
6.5	2.0	The null hypothesis is not rejected.
7	0.8	The null hypothesis is not rejected.
7.14 (the expected value estimate)	0.8	The null hypothesis is not rejected.
7.43 the variance estimate	1.1	The null hypothesis is not rejected.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

The conclusion is that the data do not present sufficient evidence to contradict the hypothesis that F possesses a Poisson distribution.

2.5.3 FREQUENCY DISTRIBUTION

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 2.3, are used to determine the optimal number of channels (employees) 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 2.3 is depicted in Figure 2.3 below.

Figure 2.3 Arrival Rate of Calls Histogram



Some distributions are characterized at least partially by functions of their true parameters. Given the above picture, one can make a fairly accurate guess that the observations point to a Poisson distribution. The mean and the variance computed are almost equal, which confirms the data to be Poisson distributed. 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. The *standard deviation* is the square root of the variance.

2.6 QUEUING THEORY

2.6.1 INTRODUCTION

In this study the queuing application involves calls answered from users reporting problems. Like it was explained in the introductory chapter these users are countrywide in South Afica. They 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 (employees) that can handle the workload, thereby satisfying customers.

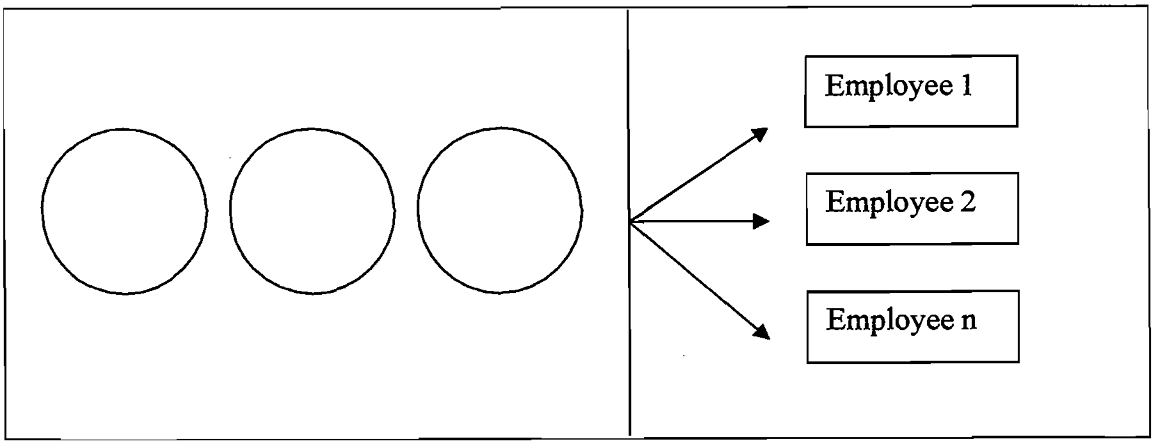
2.6.2 THE QUEUING MODEL

This section discusses the behaviour of the study's queuing model. This had to be examined (as examined in section 2.5, in order to determine whether the Poisson distribution is approximating the sample data), before a queuing model is developed. This was done again 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 multichannel, 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. In this study, there is an unlimited or infinite logging of events.

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. (Hall, 1993). The following diagram depicts the queuing model involved.

Figure 2 .4 Queue (Multi Channel, Single Phase System)



Source: (Render *et al.*, 2006)

2.6.3 OPERATING METHOD

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 λ. The arrivals follow a Poisson probability distribution (as examined in section 2.5). There is an average rate of service μ calls per minute at each channel. Like it was explained in section 1.2, 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.

2.6.4 COMPUTATIONS

The optimal number of employees (channels) is determined by computing a steady state probabilities that j of the k channels will be busy. Formula 2.1

below is used to calculate these percentages (probabilities). The following equation applies.

$$P_j = \frac{(\frac{\lambda}{\mu})^j / j!}{\sum_{i=0}^{k}(\frac{\lambda}{\mu})^i / i!} \qquad (2.1)$$

Source: (Render *et al.*, 2006)

Where λ = the mean arrival rate

μ = the mean service rate for each channel

k = the number of channels

P_j = the probability that j of the k channels are busy for j = 1, 2,…,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 abondened, (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

$$L = \lambda / \mu \, (1 - P_k) \qquad (2.2)$$

Source: (Render *et. al.*, 2006)

Whether arrivals are indeed Poisson distributed, was determined in section 2.5. 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 (λ) = 7. Currently 17 channels are responsible for answering 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 μ = 0.5.

Since there are 17 channels, they cannot handle the workload as there is a high % of abandoned calls daily. Using the Formula 2.1, 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.

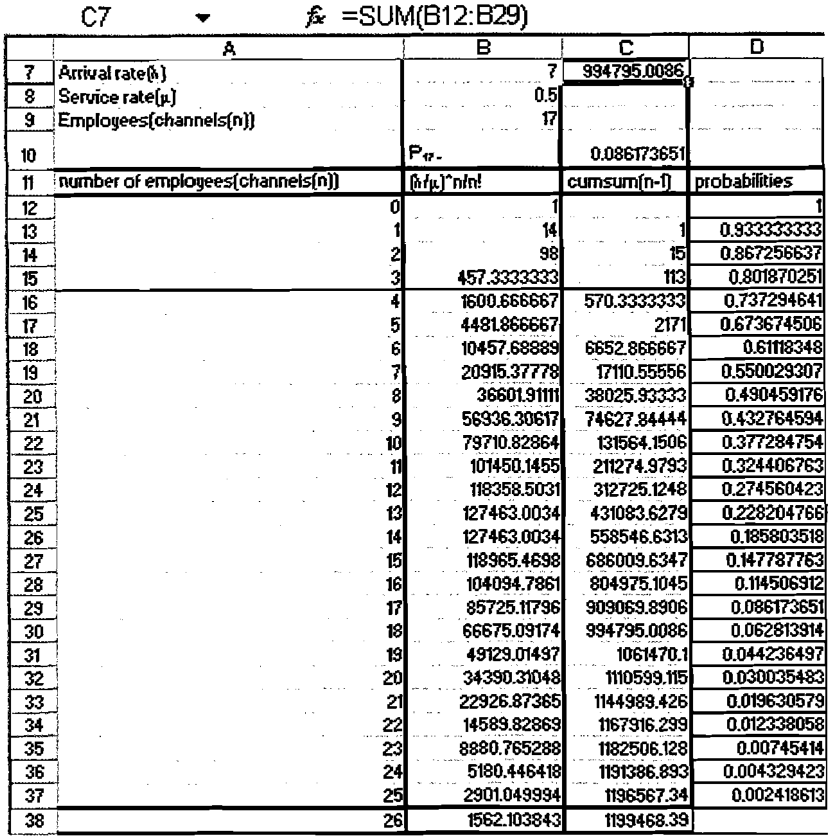
$$P_{17} = P_{\text{abandoned}} = \frac{(7/0.5)^{17}/17!}{[(7/0.5)^0/0! + (7/0.5)^1/1! +,\ldots,+ (7/0.5)^{17}/17!]}$$

$$= 85725.11796 / 994795.009$$

$$= 0.08617365$$

With only 8.61% of calls blocked with 17 channels, 91.39% of calls is 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 2.7 below. In this table, 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. Explanation regarding this spreadsheet model is on Appendix A.

Table 2.7 Abandoned Calls % For Different Number Of Channels



C7 ▾ fx =SUM(B12:B29)

	A	B	C	D
7	Arrival rate(λ)	7	994795.0086	
8	Service rate(μ)	0.5		
9	Employees(channels(n))	17		
10		P_{17}-	0.086173651	
11	number of employees(channels(n))	(λ/μ)^n/n!	cumsum(n-1)	probabilities
12	0	1		1
13	1	14	1	0.933333333
14	2	98	15	0.867256637
15	3	457.3333333	113	0.801870251
16	4	1600.666667	570.3333333	0.737294641
17	5	4481.866667	2171	0.673674506
18	6	10457.68889	6652.866667	0.61118348
19	7	20915.37778	17110.55556	0.550029307
20	8	36601.91111	38025.93333	0.490459176
21	9	56936.30617	74627.84444	0.432764594
22	10	79710.82864	131564.1506	0.377284754
23	11	101450.1455	211274.9793	0.324406763
24	12	118358.5031	312725.1248	0.274560423
25	13	127463.0034	431083.6279	0.228204766
26	14	127463.0034	558546.6313	0.185803518
27	15	118965.4698	686009.6347	0.147787763
28	16	104094.7861	804975.1045	0.114506912
29	17	85725.11796	909069.8906	0.086173651
30	18	66675.09174	994795.0086	0.062813914
31	19	49129.01497	1061470.1	0.044236497
32	20	34390.31048	1110599.115	0.030035483
33	21	22926.87365	1144989.426	0.019630579
34	22	14589.82869	1167916.299	0.012338058
35	23	8880.765288	1182506.128	0.00745414
36	24	5180.446418	1191386.893	0.004329423
37	25	2901.049994	1196567.34	0.002418613
38	26	1562.103843	1199468.39	

As mentioned in section 2.6.4, formula 2.1 was used in a spreadsheet to model the abandoned calls' percentages (probabilities). FACT is factorial and ^ is the index meaning raised to the power of the value in the cell. C10=(B30/C7); B14=($B$7/$B$8)^A14/FACT(A14) and copied to B15 through to B38; C14=SUM(B13;$B$13) and copied to C15 through to B38; D15=(B14/C15) and copied to D16 through to D38; C7= SUM(B13:B30).

Table 2.8 below shows the different abandoned rates of calls with 17 and 25 employees on duty for values of lambda (λ) within the 95% confidence interval.

Table 2.8 Abandoned Rate of Calls Within 95% Confidence Interval

Arrival rate per minute Lambda (λ)	**Abandoned rate% with 17 employees**	**Abandoned rate% with 25 employees**
6.5	6.17	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

2.6.5 CONCLUSION

The conclusions drawn from all the tests done, all the calculations made, recommend 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 costy. 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.

CHAPTER 3

DATA ENVELOPMENT ANALYSIS CONCEPTS

3.1 INTRODUCTION

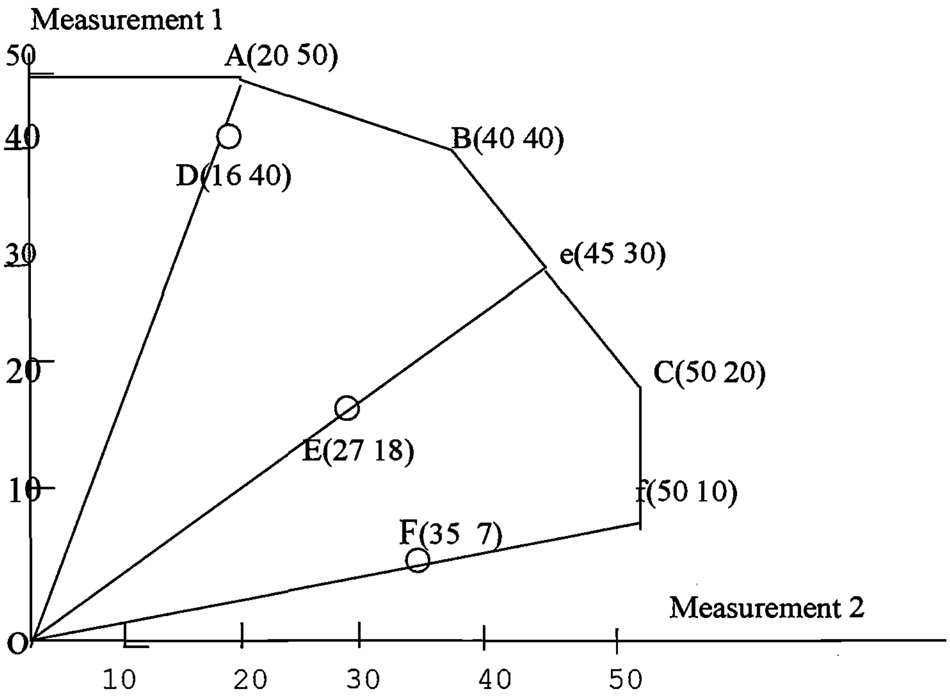
DEA, occasionally called frontier analysis, is a new technique developed in operations research and management science over the last two decades for measuring performance in the public and private sectors. It can also be described as a non-parametric estimation method which involves the application of mathematical programming to observed data to locate a frontier which can then be used to evaluate the efficiency of each of the organizational units responsible for observed output and input quantities. Charnes, Cooper, Lewin and Seiford (1994) give the general description of DEA as the efficiency measure of a Decision - making Unit (DMU), defined by its position relative to the frontier of best performance established mathematically by the ratio of weighted sum of outputs to the weighted sum of inputs.

Since subsequent chapters discuss the application of Data Envelopment Analysis (DEA), it is necessary to explain the terms and concepts which will be required in the interpretation of the results. This chapter can be considered a survey, in the sense that it discusses important contributions to the basic DEA methodology.

3.2 PARETO-OPTIMALITY

Brown, Ellis, Graves & Roman (1987:382) define pareto optimality as following. *"A state of the world A is preferable to a state of the world B if at* *least one person is better off in A and nobody is worse off"*. The best way to explain pareto optimality, is by means of an example as below.

Figure 3.1 Pareto-Optimal Decision-making Units



Source: (Zeleny, 1974)

Figure 3.1. gives an illustration of a Pareto-optimal organization. In this figure there are six Decision-making units designated A,B,C,D,E, and F, with measurement 1 and measurement 2 as coordinates. Decision-making units D, E, and F are not Pareto-optimal because they are not on the efficiency frontier determined by decision-making units A, B, and C. Zeleny (1974) assigns Decision-making units which are on the efficiency frontier, a score of 100. Then the other Decision-making units which are not on the efficiency frontier are assigned a score relative to the score of 100. For example, since both measurements 1 and 2 of D are 0.8 of that of A, the score of D is 80. He further contends that this value is actually 100 multiplied by the ratio of the length OD to the length OA, i.e. (OD/OA)*100. Since B and C are Pareto-optimal, the convex combination of B and C, which is the line segment that connects B and C, should also be Pareto-optimal. If Decision-making unit E is compared to the imaginary Decision-making unit labeled e (see Figure 3.1) a score of (OE/Oe)*100=60 result. Similarly, Decision-making unit F is compared to the imaginary Decision-making unit f (see Figure 3.1) to get a score of (OF/Of)*100=70. The comparison is relative, not absolute, hence the score of a Decision-making unit depends on other Decision-making units being evaluated. When new Decision-making units are added or old ones are deleted, the evaluated score of each Decision-making unit will probably change. However, a non-optimal Decision-making unit will never become optimal when new Decision-making-Units are added for comparison. (Zeleny, 1974).

3.3 WEIGHTS

The measurement of outputs in some organisations like health departments, are qualitative. They cannot be quantified. If these outputs can be defined, they will be denominated in non-homogeneous units. This will make it difficult to form a summary picture of departmental performance. This reflects a lack of appropriate weights. DEA can be used to form a summary picture of departmental operations by generating suitable weights on inputs and outputs. Since DEA is a relative efficiency measure, it computes weights through the comparison of performance. That is, its implementation requires a line structure where each branch is producing the same set of outputs from the same set of inputs (Ganley & Cubbin, 1992).

3.4 THE MEASUREMENT OF EFFICIENCY IN DEA

3.4.1 THE FRACTIONAL DEA PROGRAM

This section discusses the frontier using DEA. The literature on DEA is a collection of programs, both fractional and linear. The fractional program can be thought of as the conceptual DEA model, while the linear program is used in actual computation of the efficiency ratio. To introduce this methodology, the best way is to think of summarizing performance by weighting inputs and outputs in a single ratio. Assume an organization produces outputs y_r, r = 1,...,s from inputs x_i, i = 1,...,m. Then given a set of appropriate weights (u_r, r=1,...,s, v_i =1,...,m) on these variables, it is possible to form the total factor productivity ratio.

$$\frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0}}. \qquad (3.1)$$

Source: (Charnes *et.al.*, 1994)

Consider the performance of departmental branches, each using the same set of inputs to produce the same set of outputs. The total factor efficiency of each branch is the solution of a fractional program. Hence for any branch 0, efficiency can be measured as the maximum of the ratio of weighted outputs to weighted inputs subject to constraints reflecting the performance of the other branches. DEA treats the observed inputs x_i's and outputs y_r's in this ratio as constants and chooses values of the input and output weights to maximize the total factor efficiency of branch 0 relative to the performance of its peers. That is,

$$Max\ h_0 = \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0}}$$

Subject to

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_j x_{ij}} \leq 1; \qquad j = 1, \dots ,n \qquad (3.2)$$

$$u_r \geq 0,\ v_i \geq 0 \qquad r = 1, \dots ,s \quad i = 1, \dots ,m.$$

Source: (Charnes *et. al.,* 1994)

The x_{ij} represents input values for the jth DMU and the outputs are indexed so that the y_{rj} represents the observed amount of each of r=1,…,s outputs obtained for these inputs. Each of the j = 1,…,n DMUs utilizes the same inputs and produces the same outputs in different amounts. The n constraints in the above-mentioned formula ensure that no DMU can achieve an efficiency rating that will exceed unity (Charnes *et.al.*, 1994). DEA proceeds by constructing a frontier composed of best practice performers and then measures efficiency relative to that frontier. Thus the best practice performers are the benchmark on which the performance of others is to be evaluated.

3.4.2 THE LINEAR DEA PROGRAM: PRIMAL FORMULATION

The fractional program is not used for actual computation of the efficiency scores because it has intractable non-linear and non-convex properties. Rather, Charnes *et.al.* (1994) have advocated the use of a transformation to convert the fractional program into an ordinary linear program and this formula will be encountered in the subsequent chapters. The resulting linear program may be constructed to allow either output maximization or input minimization. The former computes the output efficiency ratio of a branch, and the latter its input efficiency ratio. In line with all linear programs, each has two components- a primal and a dual. The linear program for the branch is obtained by setting the denominator in the objective function of the fractional program equal to unity, and the program becomes linear. It contains the weighted sum of inputs to be unity and maximizes the weighted sum of outputs at the branch, choosing appropriate values of inputs and outputs. The less than unity constraints of the fractional program are embodied in the constraints of the primal LP, such that the efficiency score cannot exceed unity (Charnes *et.al.*, 1994).

3.4.3 THE LINEAR DEA PROGRAM: DUAL FORMULATION

Every LP has another LP associated with it, which is called its dual. The first way of stating a linear program, is called the primal of the problem, all the problems formulated, can be viewed as primals. The second way of stating the same problem, is called the dual. The optimal solutions for the primal and the dual are equivalent, but they are derived through alternative procedures.

The dual contains economic information useful to management, and it may also be easier to solve, in terms of less computation, than the primal problem. Generally, if the LP primal involves maximizing a profit function subject to less than or equal to resource constraints, the dual will involve minimizing total opportunity costs subject to greater than or equal to product profit constraints. Formulating the dual problem from a given primal is not excessively complex, and once it is formulated, the solution procedure is exactly the same as for any LP problem (Render *et.al.*, 2006).

3.5 RETURNS TO SCALE

This section discusses some extensions to the original DEA program of Charnes, Cooper, and Rhodes (1979, 1978) according to Charnes *et.al.* (1994). This is about the addition of constraints to the program to permit a greater diversity of scale possibilities in the estimated production surface. These subsequent developments, particularly in Banker, Charnes and Cooper, (1984) and Banker, Charnes, Cooper and Schinnar (1981), according to Charnes *et.al.* (1994) have extended the original Farrell program to allow for a wide range of more general reference technologies.

Returns to scale refers to increasing or decreasing efficiency based on size. For example, a manufacturer can achieve certain economies of scale by producing a thousand circuit boards the same time rather than one at a time. It might be only 100 times as hard as producing one at a time. This is an example of increasing returns to scale (IRS). On the other hand, the manufacturer might find it more than a trillion times as difficult to produce a trillion circuit boards at a time because of storage problems and limits on the world-wide copper supply. This range of production illustrates decreasing returns to scale (DRS). Combining the two extreme ranges, would necessitate the variable returns to scale (VRS).

Constant returns to scale (CRS) means, that the producers are able to linearly scale the inputs and outputs without increasing or decreasing efficiency. This is a significant assumption. The assumptions of CRS may be valid over limited ranges, but its value must be justified. CRS tends to lower the efficiency scores, while VRS tends to raise efficiency scores (Beasley, 2007).

3.6 DATA ENVELOPMENT ANALYSIS CCR AND BCC MODELS

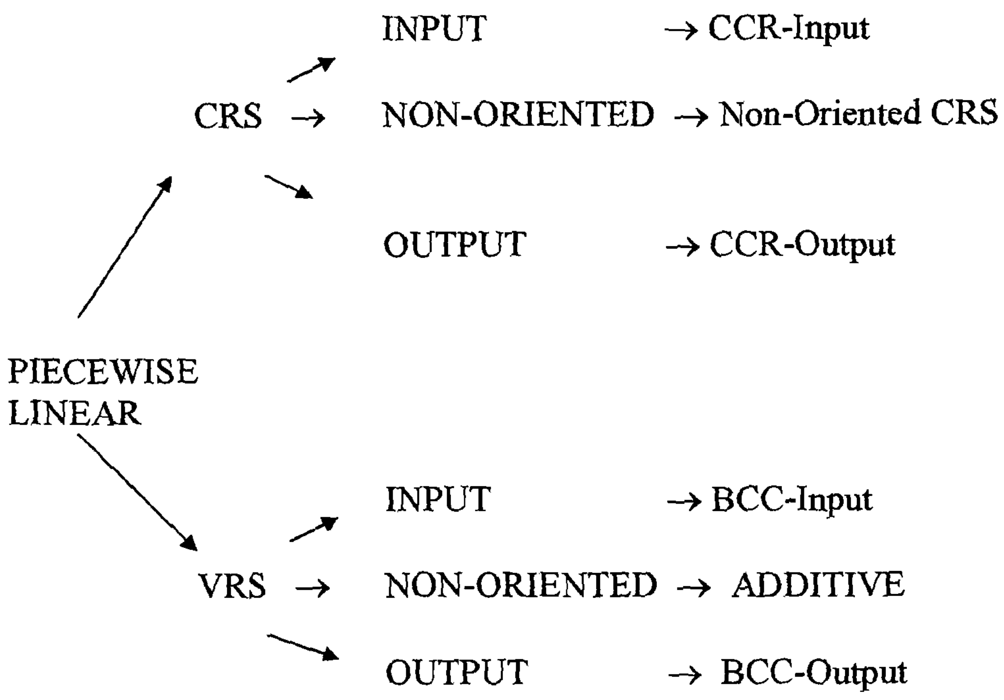
Charnes *et.al.* (1994) contends, that getting started with DEA, involves several issues, the first of which relates to choosing the DEA model to be formulated, either Charnes, Cooper and Rhodes (CCR) or Banker, Charnes and Cooper (BCC). The primary difference between CCR and BCC models is the treatment of returns to scale. The CCR version bases the evaluation on constant returns to scale. The BCC version is more flexible and allows variable returns to scale. For a DMU to be considered BCC efficient, it only needs technical efficiency, and for it to be CCR efficient, it needs both technical and scale efficiencies. The choice of a DEA model can be made by answering two questions. Does the problem formulation justify an assumption of constant returns to scale (CRS) or is the problem formulation oriented toward output maximisation, input minimisation?

3.7 DEA ANALYSIS

DEA analysis presumes the selection of a specific DEA model for analysis.

The models that assume a piecewise linear envelopment surface, can be further classified with respect to the assumed returns to scale, which may be either constant (CRS) or variable (VRS). Further classification is based on orientation, a model may not have any orientation, may be input-orienting, or may be output-orienting (Charnes *et.al.*, 1994). The classification is pictorially in Figure 3.2 below.

Figure 3.2 Classification by Returns to Scale and Orientation



Source :(Charnes *et.al.* 1994)

Determination of whether or not a decision-making unit, DMU_i, for some i, lies on the envelopment surface requires the solution of a mathematical program and that will be encountered in the next chapters when the model analysis is approached.

3.8 CONCLUSION

Now that the terms and concepts used in the practical problem were explained, it will be easy to understand the discussions that are to follow in subsequent chapters.

CHAPTER 4

RATIO ANALYSIS

4.1 INTRODUCTION

As already mentioned in section 1.4, the DEA model has a linear programming (LP) formulation. As any LP, it has two versions, the primal and the dual. In DEA, these are known as the ratio formulation and the envelope formulation.

According to Marcoulides (1998: 121), the need to compare performance with some known number or quantity in order to understand how well the organization performs brought about the increasing popularity of what is known as performance ratios. A commonly used traditional ratio method in DEA, is input-oriented and measures productivity or efficiency as a ratio of output to input (Beasley, 2007). The model in this study is input-oriented and follows a constant returns to scale as explained in section 3.5.

4.2 SINGLE INPUT, OUTPUT MEASURE

4.2.1 NUMBER OF EMPLOYEES AND RESOLVED EVENTS

Suppose that the inputs and outputs for the five regions are as given in Table 2.2. Considering the company's regions, each region has a single output measure (number of resolved events) and a single input measure (number of employees). From Table 2.2 the input, number of employees and the output, number of resolved events are used to compute the ratio. The following apply.

Table 4.1 Single Input, Output (Resolved Events)

Region	Number of Employees	Number of Resolved Events
Pretoria	17	201
Bloem/Kby	16	160
Durban	15	157
Johannesburg	17	200
Cape Town	13	123

In the above data, for instance, Pretoria had 201 resolved events while 17 staff members were employed. In Durban there were 157 resolved events and 15 staff members were employed, etc. These regions are compared and their performance measured by using the data. Some output measure is divided by some input measure to get a ratio. For example, 201 is divided by 17 to get 11.80. The following data applies.

Table 4.2 Single Input, Output (Resolved Events) Ratios

Region	Events Resolved per Employee
Pretoria	201/17 =11.80
Bloem/Kby	160/16 =10.00
Durban	157/15 =10.47
Johannesburg	200/17 =11.76
Cape Town	123/13 = 9.46

According to the above data, Pretoria has the highest ratio of resolved events per staff member, whereas Cape Town has the lowest. Since Pretoria has the highest ratio of 11.80, other regions are compared to it and their relative efficiencies calculated with respect to it. The ratio for any region is divided by the ratio for Pretoria (11.80), multiplied by 100 to convert to a percentage, resulting in the following.

Table 4.3 Single Input, Output (Resolved Events) Percentages

Region	**Relative Efficiency**
Pretoria	100%
Bloem/Kby	85%
Durban	89%
Johannesburg	99.6%
Cape Town	80%

The other regions do not compare with Pretoria, they are performing lower and are relatively less efficient at using their staff (input) to produce output (number of resolved events). Pretoria can be used to set a target for other regions. This is an input target, since it deals with input measure.

4.2.2 NUMBER OF EMPLOYEES AND CLIENT SATISFACTION

This time, output measure is client satisfaction and the input measure remains the number of employees since this ratio method is input oriented. The target is the number of employees. This is the variable that is going to be adjusted to effect efficiency. By increasing or decreasing the number of employees, the optimal output will be reached. Once more, client satisfaction is determined as the ratio of number of resolved events to the total number of logged events per day. From Table 2.2, data again are as follows.

Table 4.4 Single Input, Output (Client Satisfaction)

Region	Employees	Client Satisfaction
Pretoria	17	0.66
Bloem/Kby	16	0.86
Durban	15	0.79
Johannesburg	17	0.67
Cape Town	13	0.62

In the data, for instance, Pretoria had a ratio of 0.66 client satisfaction and 17 staff members were employed. In Durban a ratio of 0.79 events was resolved, while 15 staff members were employed, etc. These regions are compared and their performance measured by using this data. Some output measure is divided by some input measure to get a ratio. Hence the following data results.

Table 4.5 Single Input, Output (Client Satisfaction) Ratios

Region	Client Satisfaction per Employee
Pretoria	0.66/17 =0.039
Bloem/Kby	0.86/16 =0.054
Durban	0.79/15 =0.053
Johannesburg	0.67/17 =0.039
Cape Town	0.62/13 =0.048

According to the above data, Bloem/Kby had the highest ratio of Client Satisfaction per employee, whereas Pretoria had the lowest. Since Bloem/Kby had the highest ratio of 0.054, all other regions are compared to it and their relative efficiency calculated with respect to Bloem/Kby. The ratio for any region is divided by the ratio of Bloem/Kby (0.054) and multiplied by 100 to convert to a percentage, as following.

Table 4.6 Single Input, Output (Client Satisfaction) Percentages

Region	Relative Efficiency
Pretoria	72%
Bloem/Kby	100%
Durban	96%
Johannesburg	72%
Cape Town	87%

The other regions do not compare with Bloem/Kby, they are performing less, they are relatively less efficient at using their staff (input) to produce output (satisfaction). Bloem/Kby could set target for other regions. This is still an input target, since it deals with input measure.

4.3 EXTENDED RESOURCES

Considering a single input measure, number of employees, and two output measures, resolved events and client satisfaction could be resolved at the same time. Again the five regions are compared. From Table 2.2, the data are again as following.

Table 4.7 Extended Resources

Region	Number of Employees	Resolved Events	Client Satisfaction
Pretoria	17	201	66%
Bloem/Kby	16	160	86%
Durban	15	157	79%
Johannesburg	17	200	67%
Cape Town	13	123	62%

Durban, for example, with 15 employees, had an average of 157 events resolved per month and satisfied its clients up to 79 percent. Ratios are still used to compare these regions. Dividing each output measure with the single input (number of employees) gives the following.

Table 4.8 Efficiency Ratios

Region	Events Resolved Per Employee	Client Satisfaction per Employee
Pretoria	11.80	3.9
Bloem/Kby	10.00	5.4
Durban	10.47	5.3
Johannesburg	11.76	3.9
Cape Town	9.46	4.8

Pretoria had the highest ratio of resolved events per employee whereas Bloem/Kby had the highest ratio of client satisfaction per employee. Figure 4.1 in the next section presents the above data.

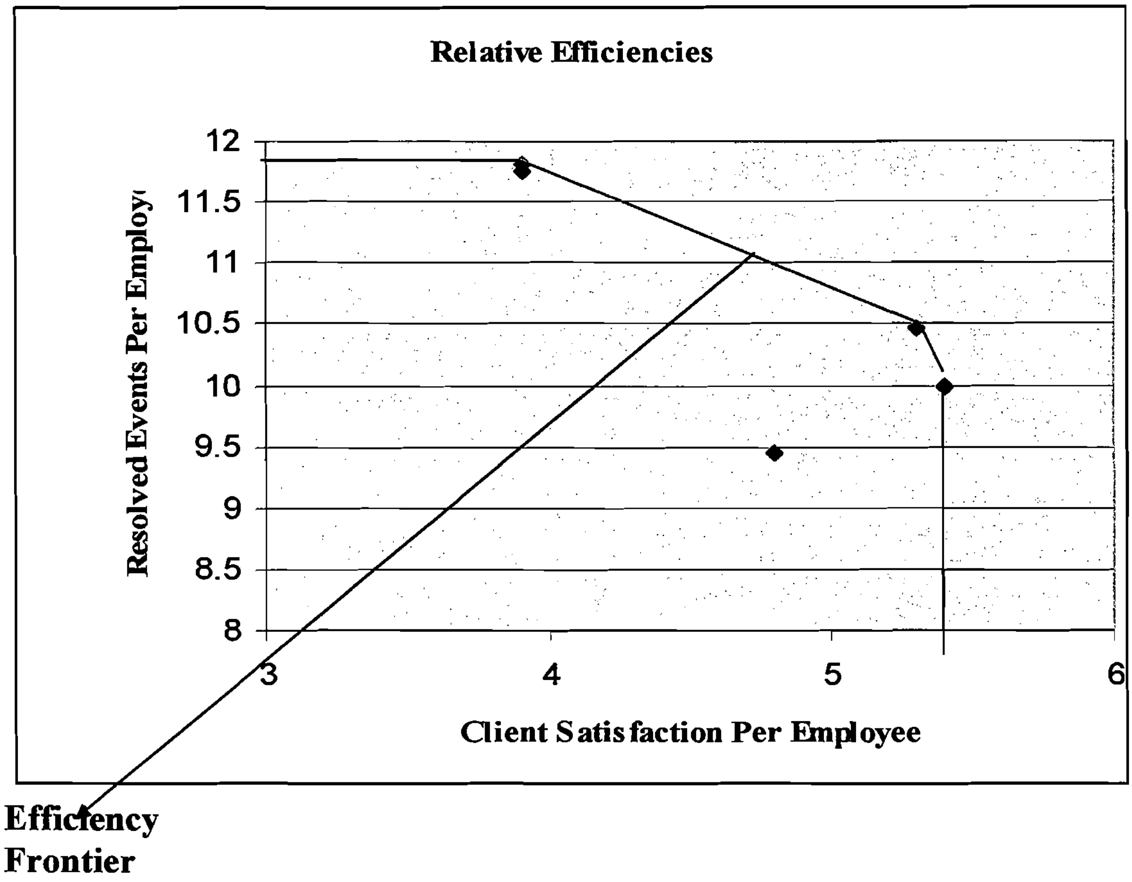
The problem with comparing ratios, is that a different ratio could give a different picture and it becomes difficult to combine these ratios into one ratio, where one could draw one's own judgement. For example, if we consider Durban and Cape Town, Durban gives (10.47/9.46) = 1.11 times as efficient as Cape Town on resolved events and also (5.3/4.8) = 1.11 times as efficient as Cape Town on client satisfaction. It is not easy to combine these ratios into a judgement. This can be more clearly seen if there are more inputs and more outputs (Beasley, 2007).

4.4 GRAPHICAL ANALYSIS

Another way of evaluating the efficiency, at least for problems involving two outputs and a single input, is by graphical analysis, as shown in Figure 4.1 below. In this Figure, all the regions are on the frontier line, except Cape Town. Johannesburg and Pretoria almost make the same data point, since their readings are almost equal (the thickest point in the graph).

Again in Figure 4.1 below, a horizontal line is drawn from the y-axis to Pretoria, from Pretoria to Johannesburg, from Johannesburg to Durban, from Durban to Bloem/Kby. A vertical line is drawn from Bloem/Kby to the x-axis. This line is called the efficiency frontier. The efficient frontier, derived from the examples of best practice contained in the data considered, represents the performance that the regions, in this case Cape Town, not on the efficient frontier could try to achieve. Hence data envelopment because the efficient frontier envelopes (encloses) all data available. All the regions on the frontier are 100% efficient. Therefore, all the regions are efficient except Cape Town (Beasley, 2007).

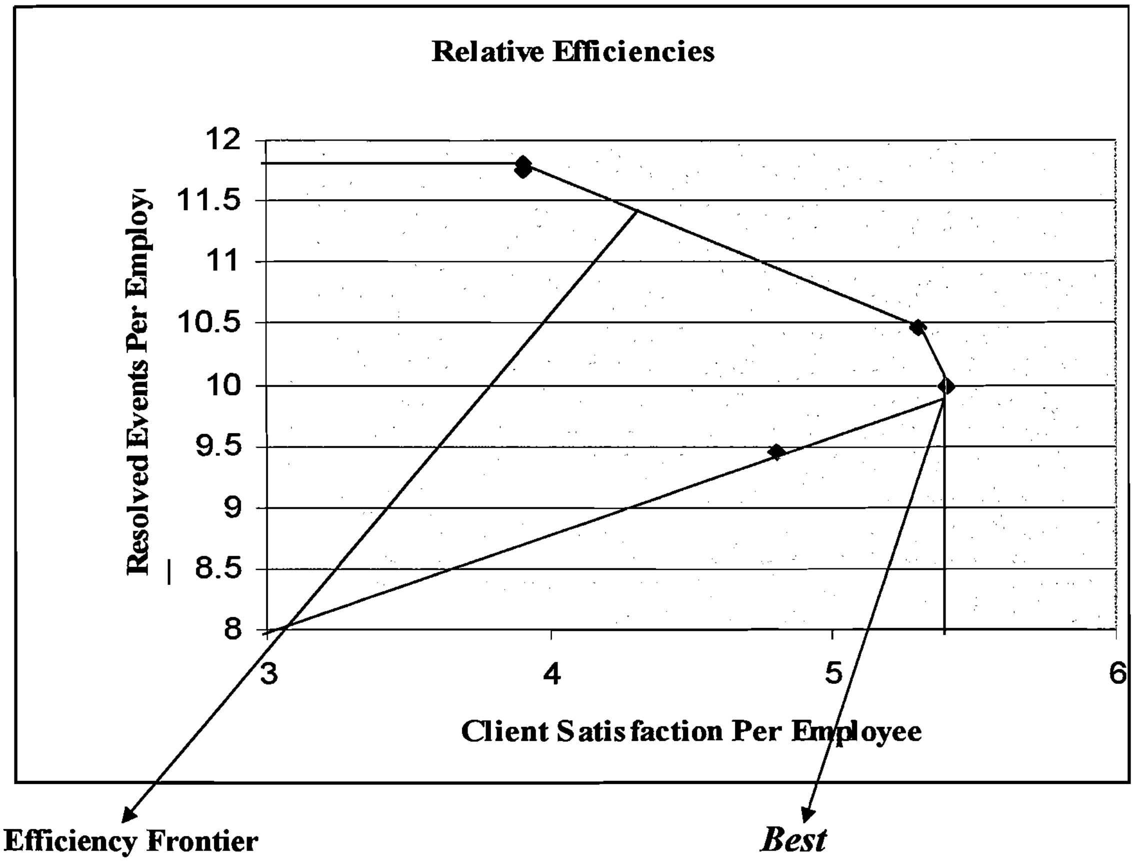
Figure 4.1 Relative Efficiencies



4.5 QUANTIFYING EFFICIENCY SCORE FOR CAPE TOWN

Cape Town is less efficient, but by how much? It has 13 staff members, 123 resolved events, 9.46 resolved events per employee, 62% client satisfaction, and 4.8 client satisfaction per employee. The ratio of resolved events/client satisfaction = (123/62) = 1.98; that is, there are 1.98 resolved events for every percentage of client satisfaction. This ratio is the same as resolved events per employee to client satisfaction per employee.

Figure 4.2 Relative Efficiencies



Considering Figure 4.2, Cape Town is not on the efficiency frontier. A line drawn from the origin through Cape Town to the efficiency frontier line has a slope of 1.98. If Cape Town were to retain this ratio, but to vary the number of staff it employs, its performance, would lie on the line from the origin through its current position as shown above. It might be reasonable to say that the best possible performance that Cape Town could be expected to achieve is labelled Best in the graph. This is the point where the line from the origin through Cape Town meets the efficiency frontier.

According to Beasley (2007), DEA gives only the relative efficiencies; efficiencies relative to the data considered. It does not and cannot give absolute efficiencies.

4.6 SENSITIVITY ANALYSIS

The illustration below shows that it is possible to move beyond the efficient frontier created by DEA ratio analysis. Predicted number of events resolved in Figure 4.3 confirms this.

Figure 4.3 Number of Employees Line Fit Plot

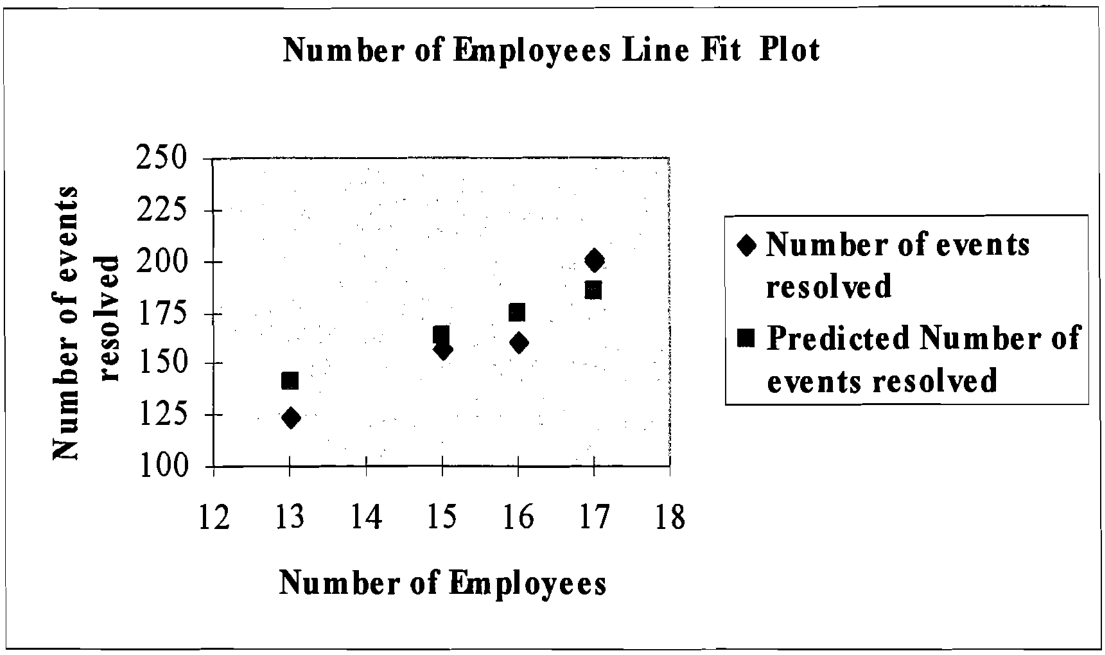


Figure 4.4 and 4.5 below show plot of residuals. This is to assess the model adequacy by checking whether the model assumptions are satisfied. The basic assumption is that the residuals are uncorrelated with zero mean and constant variance.

Figure 4.4 Number of Employees Residual Plot

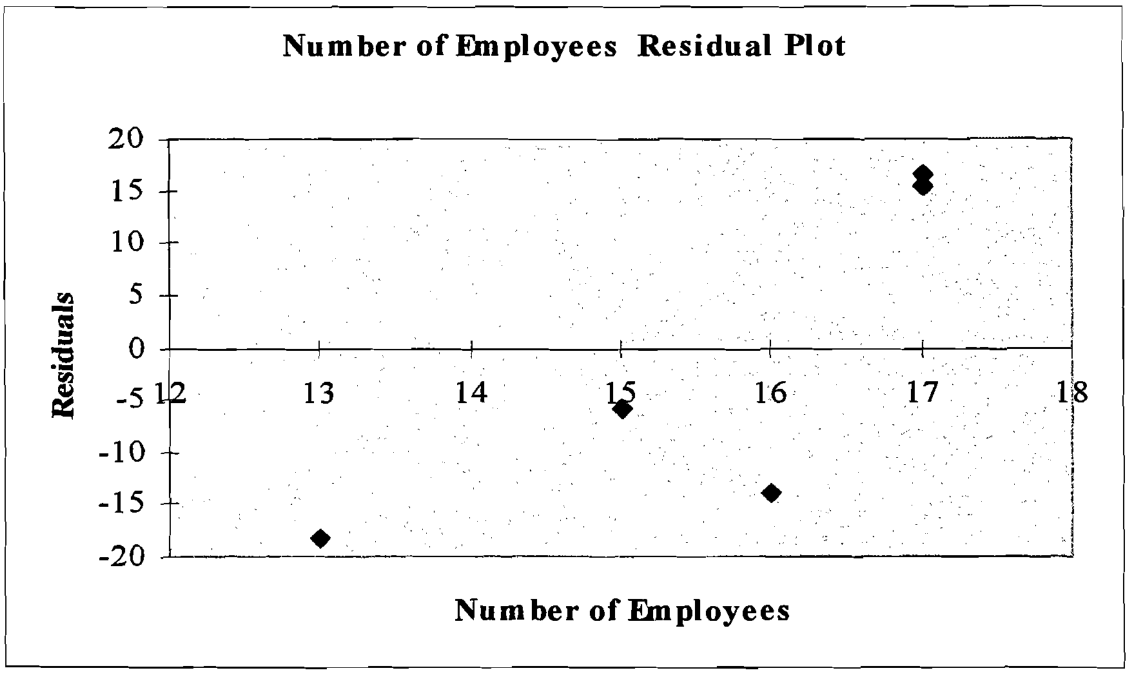
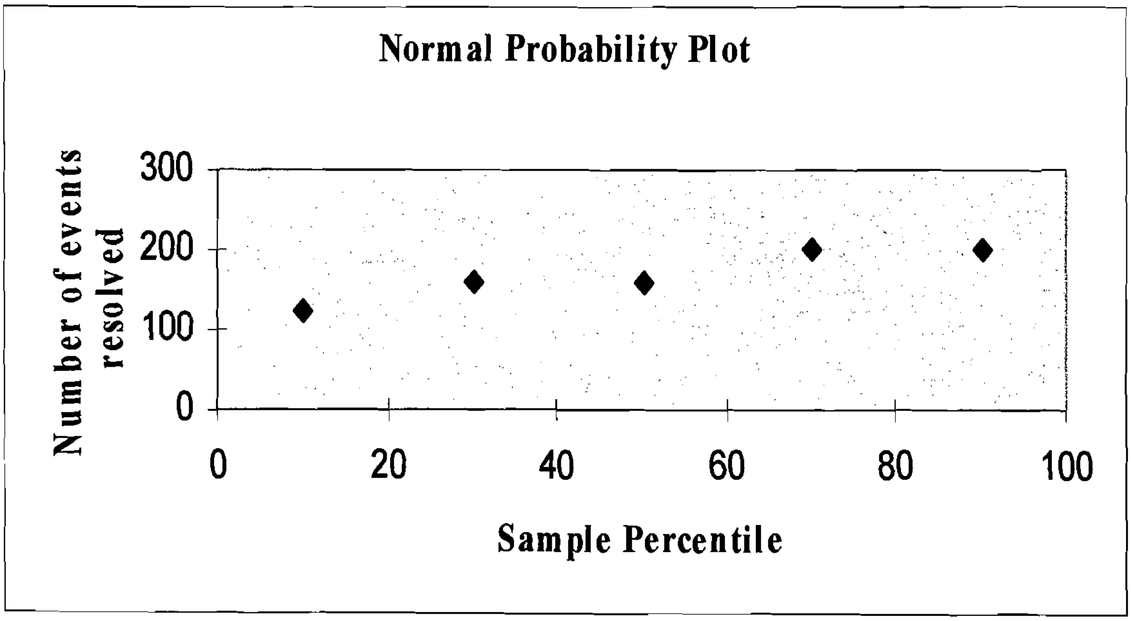


Figure 4.5 Normal Probability Plot



The residuals in Figure 4.4 look evenly distributed with no pattern, meaning that they are uncorrelated. Figure 4.5, the normal probability plot shows no serious deviations from the fourty five degree line. We conclude from the plots that the residuals are normally distributed. The model is adequate.

Table 4.9 Predicted Outputs

	Number of Employees	Predicted Number of resolved events	Predicted Client Satisfaction	Number of logged events
1	17	195	0.64	305
2	16	176	0.94	186
3	15	157	0.79	199
4	17	195	0.65	299
5	13	119	0.60	198
	Regression		Intercept	-126.25
			Number of Employees	18.875
	Number of Employees	Predicted Number of resolved events	Predicted Client Satisfaction	Number of logged events
1	23	308	1.01	305
2	17	195	1.05	186
3	18	214	1.07	199
4	23	308	1.03	299
5	18	214	1.08	198

In Table 4.9 above, the current number of employees could resolve all the logged events. When there are more employees, statistics shows that all logged events could be resolved. This is the number of employees necessary to resolve all the logged events, using the extrapolated extrapolating the regression equation. The assumption is that the extra people could resolve the unresolved events at the same rate as the existing employees, which may not be valid if the unresolved events are more difficult to resolve than the resolved events. This may be a way to move beyond the current efficient frontier and create new quality standards.

4.7 CONCLUSION

While ratios are easy to compute, their interpretation is problematic, especially when they provide conflicting answers. While this may be generally true, statistics can be used to understand this. For example, using the number of resolved events, Pretoria is at the top while using client satisfaction, Pretoria is at the bottom. This may look like conflicting results, but in reality they are not. More employees can resolve more events, if there are events to resolve.

CHAPTER 5

LINEAR PROGRAMMING

5.1 INTRODUCTION

According to section 3.4.1, the efficiency of each region is computed using the transformed linear Formula 3.2, referred to as Formula 5.1 in this chapter. Formula 3.2 is not used for actual computation of efficiency scores since it has intractable non-linear and non-convex properties. For any region, the efficiency can be measured as the maximum of the ratio of weighted outputs to weighted inputs subject to constraints reflecting the performance of other regions. Inputs and outputs are treated as constants and values of input and output weights are chosen to maximize the efficiency of the region relative to the performance of other regions. This model assumes a constant returns to scale (CRS) as discussed in sections 3.5. This chapter illustrates the use of Formula 5.1 in computing efficiency scores in Linear programming.

Charnes *et.al.* (1978) recognized the difficulty in seeking a common set of weights to determine relative efficiency among comparable units and generalized a technique known as DEA first proposed by Farrell (1957). They recognized the legitimacy of the proposal that units might value inputs and outputs differently and therefore adopt different weights, and proposed that each unit (DMU) or region in this study should be allowed to adopt a set of weights that shows it in the most favourable light in comparison to the other units.

As can be seen in Formula 3.2, h_0 (called the technical efficiency of DMU_o) is merely the weighted sum of outputs/inputs, where in this case, the u's and v's are the weights. The big difference is that there is no priori selection of what these weights should be. These are instead merely constrained to be greater than or equal to some small positive quantity (ε) in order to avoid any input or output being totally ignored in determining the efficiency. The constraint merely ensures that the weights are chosen so that no region can have an efficiency of greater than 1. If $h_0 = 1$ then region 0 is efficient relative to others. If h_0 turns out to be less than 1 then some other region is more efficient than region 0.

It is also possible to differentiate an input orientation (such as that in the model in this study) and output orientation. The core practical difference between these two lies in the interpretation of the efficiency score. Input oriented measures quantify the weighted input reduction necessary to become efficient, assuming the output remains constant. For example, an efficiency score of 70% would mean that if the region were efficient, inputs could be reduced by 30% with the same resulting outputs. On the other hand, output oriented measures quantify the relative output possible, assuming 100% efficiency with the inputs remaining constant. For example, an efficiency score of 110% would mean that if the region were efficient, output could be 10% higher with the same inputs. Input oriented models maximize the ratio weighted outputs/weighted inputs and output oriented models minimize the ratio weighted inputs/weighted outputs. (Charnes *et.al.*, 1994).

Formula 3.2 is difficult to solve because the objective function is nonlinear and fractional. Charnes *et.al.* (1994) transformed this formula into a linear one as follows.

$$Max\ h_0 = \sum_{r=1}^{s} u_r y_{r0}$$

Subject to

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0$$

$$\sum_{i=1}^{m} v_i x_{i0} = 1$$

$$u_r, v_i \geq 0; for\ i=1,2,\ldots,m;\ r=1,2,\ldots,s;\ and j=1,2,\ldots,n. \qquad (5.1)$$

A Linear Programming (LP) problem may be defined as the problem of maximizing or minimizing a linear function subject to linear constraints. The constraints may be equalities or inequalities. Firstly a Linear Programming model for region 1 only is developed using Formula 5.1 and the LP's Big M method (simplex algorithm) is applied to solve the mathematical model manually. Secondly, Linear Programming models for all the regions are developed and solved using Quantitative Methods (QM) for windows software for optimal objective function.

5.2. MANUAL SOLUTION

In this study's problems, there are three unknowns, and six constraints. All the constraints are equalities and inequalities and they are all linear in the sense that each involves an inequality in some linear function of the variables. The constraints, $u_1, u_2, v_1 \geq 0$ are called non negativity constraints. The other constraints are called the main constraints. The function to be maximized (or minimized) is called the objective function. The objective function of region 1 is 201 u_1+0.66 u_2. Table 2.2 is again used in the formulation of this program.

Max z = 201 u_1+0.66 u_2

$$\begin{aligned} \text{Such that} \quad & 201\,u_1+0.66\,u_2 -17\,v_1 \leq 0 \\ & 160\,u_1+0.86\,u_2 -16\,v_1 \leq 0 \\ & 157\,u_1+0.79\,u_2 -15\,v_1 \leq 0 \\ & 200\,u_1+0.67\,u_2 -17\,v_1 \leq 0 \\ & 123\,u_1+0.62\,u_2 -13\,v_1 \leq 0 \\ & 17\,v_1 = 1 \\ & u_1, u_2, v_1 \geq 0 \end{aligned}$$

Adding the artificial variable R_1, and the slack variables, s_1,s_2,s_3,s_4,s_5, the equations in standard form are as below. The objective equation for region 1 becomes:

$17\,v_1 + R_1 = 1 \rightarrow R_1 = (1 - 17\,v_1)$; $z = 201u_1+0.66\,u_2 - M(1-17\,v_1)$.

Max z = 201 u_1+0.66 u_2 -M R_1

$$\begin{aligned} \text{s.t.} \quad & 201u_1+0.66\,u_2 -17\,v_1 + s_1 = 0 \\ & 160\,u_1+0.86\,u_2 -16\,v_1 + s_2 = 0 \\ & 157\,u_1+0.79\,u_2 -15\,v_1 + s_3 = 0 \\ & 200\,u_1+0.67\,u_2 -17\,v_1 + s_4 = 0 \\ & 123\,u_1+0.62\,u_2 -13\,v_1 + s_5 = 0 \\ & 17\,v_1 + R_1 = 1 \\ & u_1,u_2,v_1,s_1,s_2,s_3,s_4,s_5,R_1, \geq 0 \end{aligned}$$

The initial simplex table and hence the solution is as following.

Table 5.1 Iteration 1

		201	0.66	0	0	0	0	0	0	-M
Basis	Quantity	u_1	u_2	v_1	s_1	s_2	s_3	s_4	s_5	R_1
s_1	0	201	0.66	-17	1	0	0	0	0	0
s_2	0	160	0.86	-16	0	1	0	0	0	0
s_3	0	157	0.79	-15	0	0	1	0	0	0
s_4	0	200	0.67	-17	0	0	0	1	0	0
s_5	0	123	0.62	-13	0	0	0	0	1	0
R_1	1			17	0	0	0	0	0	1
		-201	-0.66	0	0	0	0	0	0	0
		0	0	-17M	0	0	0	0	0	M

Table 5.2 Iteration 2

		201	0.66	0	0	0	0	0	0
Basis	Quantity	u_1	u_2	v_1	s_1	s_2	s_3	s_4	s_5
s_1	1	201	0.66	0	0	0	0	0	0
s_2	0.9412	160	0.86	0	0	1	0	0	0
s_3	0.8824	157	0.79	0	0	0	1	0	0
s_4	1	200	0.67	0	0	0	0	1	0
s_5	0.7647	123	0.62	0	0	0	0	0	1
v_1	0.0588			1	0	0	0	0	0
	0	-201	-0.66	0	0	0	0	0	0

The maximum negative value in the last row is under the $u_1$column, therefore u_1 enters the basis. The leaving variable is determined as the minimum of the ratios between columns two and three.

Min (1/201; 0.9412/160;0.8824/157;1/200;0.7647/123 =1/201

The minimum value is on the first row which indicates that s_1 leaves the basis.The last iteration is as below.

Table 5.3 Iteration 3

		201	0.66	0	0	0	0	0	0
Basis	Quantity	u_1	u_2	v_1	s_1	s_2	s_3	s_4	s_5
u_1	0.005	1	0.0033	0	0	0	0	0	0
s_2	0.1452	0	0.3346	0	0	1	0	0	0
s_3	0.1013	0	0.2745	0	0	0	1	0	0
s_4	0.005	0	0.0133	0	0	0	0	1	0
s_5	0.1528	0	0.2161	0	0	0	0	0	1
v_1	0.0588		0	1	0	0	0	0	0
	1	0	0	0	1	0	0	0	0

All the bottom row's values are ≥ 0, therefore the solution is optimal.

$z = 1$, Basic ($u_1 = 0.005$, $s_2 = 0.1452$, $s_3 = 0.1013$, $s_4 = 0.005$, $s_5 = 0.1528$, $v_1 = 0.0588$) and Nonbasic ($s_1, u_2 = 0$).

5.3 SIMPLEX METHOD APPLICATION

The entering variable = the minimum value in the row that has the minimum negative value. The leaving variable = the minimum of the ratios of the quantity column and the entering variable column.

The column under the entering variable is identified as the entering column. The row associated with the leaving variable is the pivot equation and the element at the intersection of the entering column and the pivot equation is the pivot element.

Pivot equation

Type 1 New pivot equation = old pivot equation divide by pivot element.

Type 2 All other equations including z.

New equation = old equation – (its entering column coefficient) x (new pivot equation).

Type 1 computations make the pivot element equal to 1 in the new pivot equation, whereas type 2 computations create zero coefficient everywhere else in the entering column.

$s_1,\ldots,s_5$ are slack variables. A slack variable is a variable which is added to a constraint to turn the inequality into an equation. This is required to turn an inequality into an equality where a linear combination of variables is less than or equal to a given constant in the former. If the constraints represent the limit on the usage of a resource, $s_1,\ldots,s_5$ will represent the unused amounts of the resources.

The artificial variable calls for adding a nonnegative variable to the left side of the equation that has no obvious starting basic variable. The added variable plays the same role as that of the slack variable in providing a starting variable. Since the artificial variable has no physical meaning from the standpoint of the original problem, the procedure will only be valid if these variables are forced to be zero when the optimum is reached. In other words, they are used only to start the solution and must subsequently force them to be zero in the final solution, otherwise the solution will not be feasible Taha (2007).

5.4 LINEAR PROGRAMMING FORMULATIONS

The LP formulations for each region are as below. A number $\varepsilon = 0.0001$, is introduced to ensure that all the observed inputs and outputs have positive weights. These are constraint to be greater than or equal to this number $\varepsilon =$

0.0001 in order to avoid any input or output being totally ignored in determining the efficiency.

Region 1 LP:

$$\begin{array}{rll}
\text{Max } z = & 201\,u_1+0.66\,u_2 & \\
\text{Subject to} & 201\,u_1+0.66\,u_2-17\,v_1 & \le 0\\
& 160\,u_1+0.86\,u_2-16\,v_1 & \le 0\\
& 157\,u_1+0.79\,u_2-15\,v_1 & \le 0\\
& 200\,u_1+0.67\,u_2-17\,v_1 & \le 0\\
& 123\,u_1+0.62\,u_2-13\,v_1 & \le 0\\
& 17\,v_1 & = 1\\
& u_1 & \ge 0.0001\\
& u_2 & \ge 0.0001\\
& v_1 & \ge 0.0001
\end{array}$$

Region 2 LP:

$$\begin{array}{rll}
\text{Max } z = & 160\,u_1+0.86\,u_2 & \\
\text{s.t.} & 201u_1+0.66\,u_2-17\,v_1 & \le 0\\
& 160\,u_1+0.86\,u_2-16\,v_1 & \le 0\\
& 157\,u_1+0.79\,u_2-15\,v_1 & \le 0\\
& 200\,u_1+0.67\,u_2-17\,v_1 & \le 0\\
& 123\,u_1+0.62\,u_2-13\,v_1 & \le 0\\
& 16\,v_1 & = 1\\
& u_1 & \ge 0.0001\\
& u_2 & \ge 0.0001\\
& v_1 & \ge 0.0001
\end{array}$$

Region 3 LP:

$$\begin{array}{rll}
\text{Max } z = & 157\,u_1+0.79\,u_2 & \\
\text{s.t.} & 201\,u_1+0.66\,u_2-17\,v_1 & \le 0\\
& 160\,u_1+0.86\,u_2-16\,v_1 & \le 0\\
& 157\,u_1+0.79\,u_2-15\,v_1 & \le 0\\
& 200\,u_1+0.67\,u_2-17\,v_1 & \le 0\\
& 123\,u_1+0.62\,u_2-13\,v_1 & \le 0\\
& 15\,v_1 & = 1\\
& u_1 & \ge 0.0001\\
& u_2 & \ge 0.0001\\
& v_1 & \ge 0.0001
\end{array}$$

Region 4 LP: $\text{Max } z = 200u_1+0.67\,u_2$

$$\begin{array}{lll}
\text{s.t} & 201u_1+0.66\,u_2-17\,v_1 \le 0 & \\
 & 160\,u_1+0.86\,u_2-16\,v_1 \le 0 & \\
 & 157\,u_1+0.79\,u_2-15\,v_1 \le 0 & \\
 & 200\,u_1+0.67\,u_2-17\,v_1 \le 0 & \\
 & 123\,u_1+0.62\,u_2-13\,v_1 \le 0 & \\
 & 17\,v_1 = 1 & \\
 & u_1 & \ge 0.0001 \\
 & u_2 & \ge 0.0001 \\
 & v_1 & \ge 0.0001
\end{array}$$

Region 5 LP: $\text{Max } z = 123u_1+0.62\,u_2$

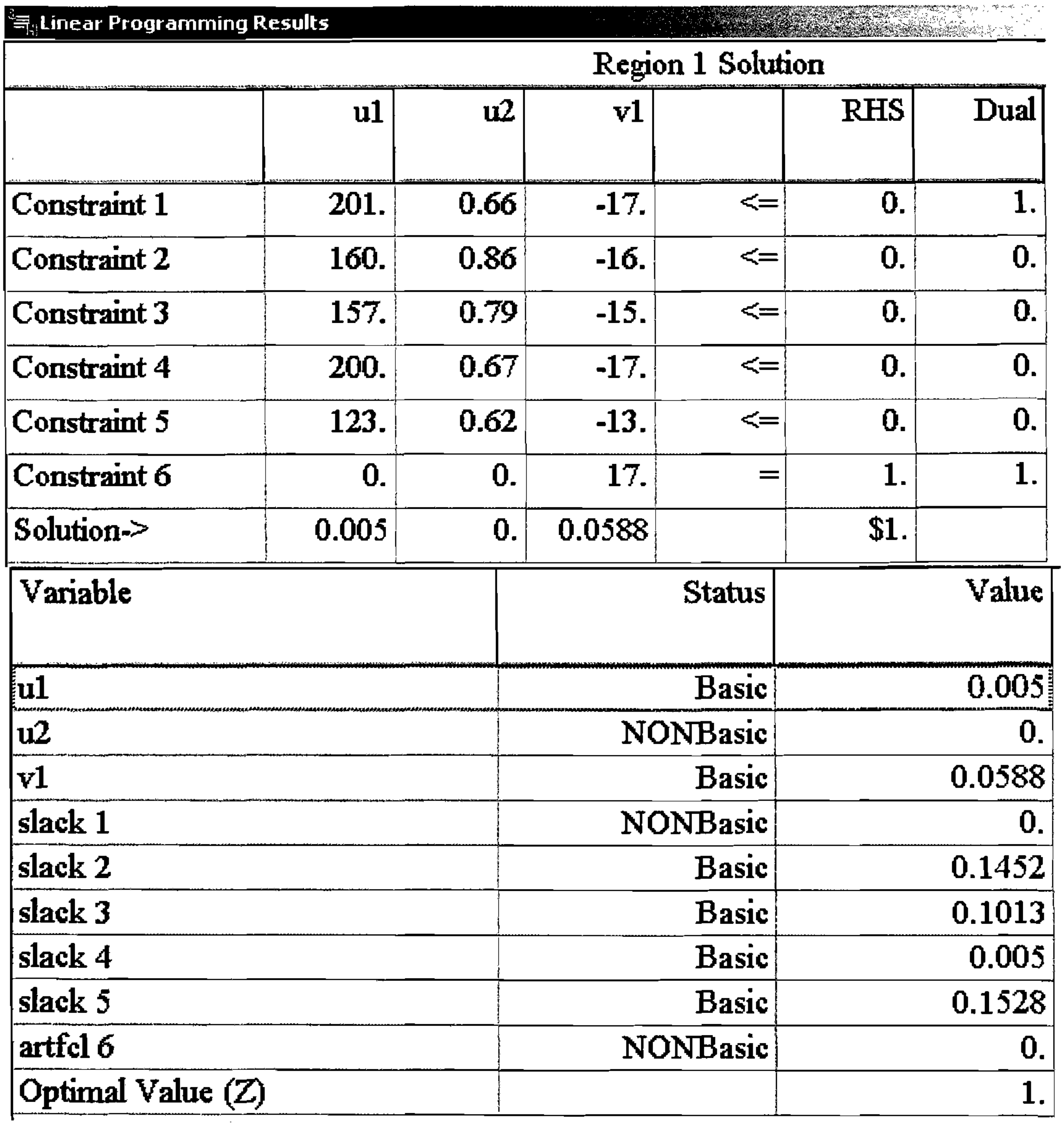
$$\begin{array}{lll}
\text{s.t} & 201\,u_1+0.66\,u_2-17v_1 \le 0 & \\
 & 160\,u_1+0.86\,u_2-16\,v_1 \le 0 & \\
 & 157\,u_1+0.79\,u_2-15\,v_1 \le 0 & \\
 & 200\,u_1+0.67\,u_2-17\,v_1 \le 0 & \\
 & 123\,u_1+0.62\,u_2-13\,v_1 \le 0 & \\
 & 13\,v_1 = 1 & \\
 & u_1 & \ge 0.0001 \\
 & u_2 & \ge 0.0001 \\
 & v_1 & \ge 0.0001
\end{array}$$

The linear programming formulations above are derived using Formula 5.1. Since it is cumbersome to solve each one of these formulations manually, the software application is used to do the solutions.

5.5 SOFTWARE APPLICATION

QM for windows is used to determine the optimal solutions of the individual regions. The procedure used in software is exactly the same as the one used in the manual solution, that is the simplex algorithm's Big M method. The efficiency scores of the regions are depicted in Figure 5.1 to 5.5 below.

Figure 5.1 LP Solution for Region 1



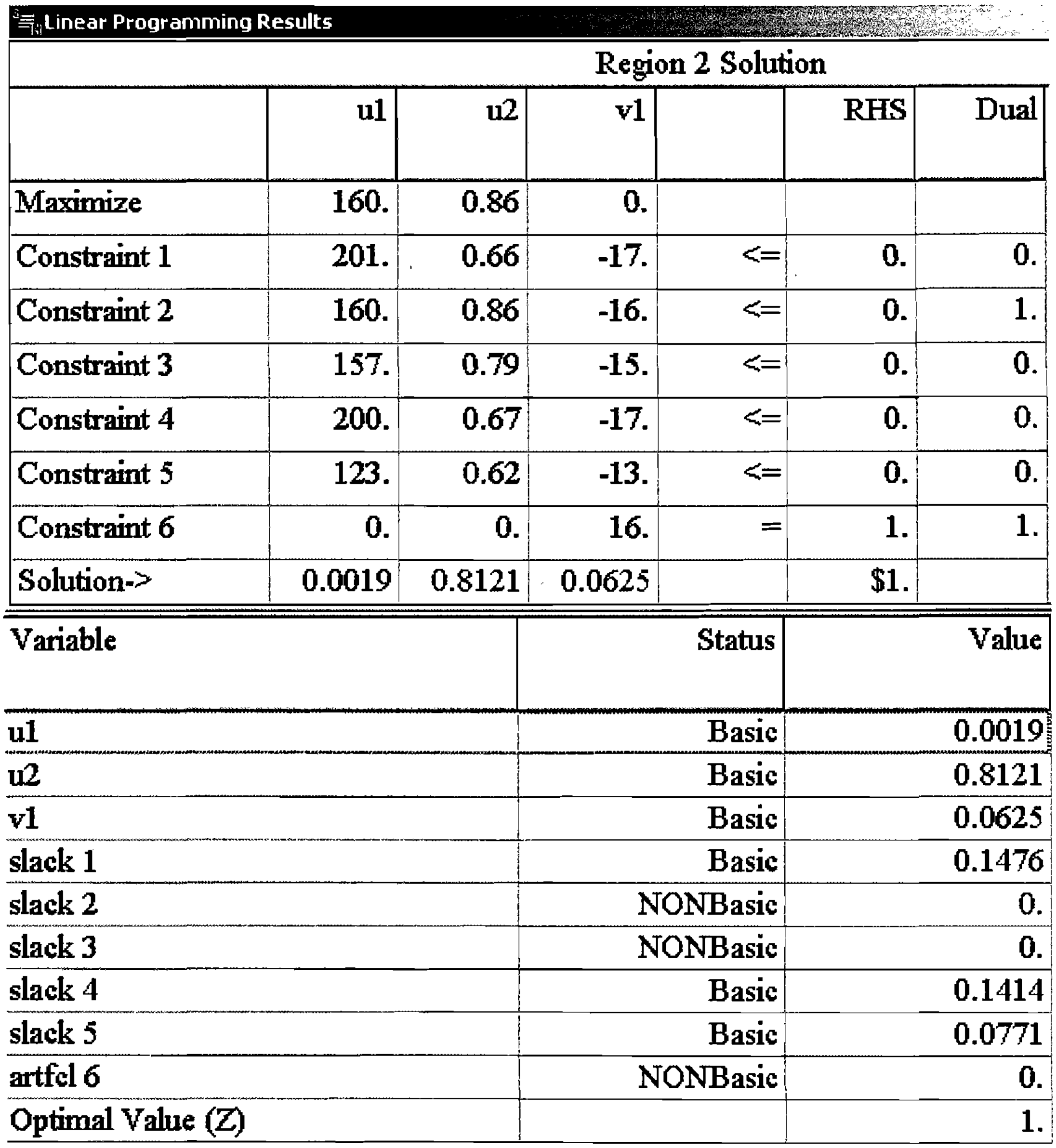
Region 1 Solution

	ul	u2	vl		RHS	Dual
Constraint 1	201.	0.66	-17.	<=	0.	1.
Constraint 2	160.	0.86	-16.	<=	0.	0.
Constraint 3	157.	0.79	-15.	<=	0.	0.
Constraint 4	200.	0.67	-17.	<=	0.	0.
Constraint 5	123.	0.62	-13.	<=	0.	0.
Constraint 6	0.	0.	17.	=	1.	1.
Solution->	0.005	0.	0.0588		$1.	

Variable	Status	Value
ul	Basic	0.005
u2	NONBasic	0.
vl	Basic	0.0588
slack 1	NONBasic	0.
slack 2	Basic	0.1452
slack 3	Basic	0.1013
slack 4	Basic	0.005
slack 5	Basic	0.1528
artfcl 6	NONBasic	0.
Optimal Value (Z)		1.

Region1 (Pretoria) has an optimal value of 1, which is an efficiency score. This region is 100% efficient.

Figure 5.2 LP Solution for Region 2

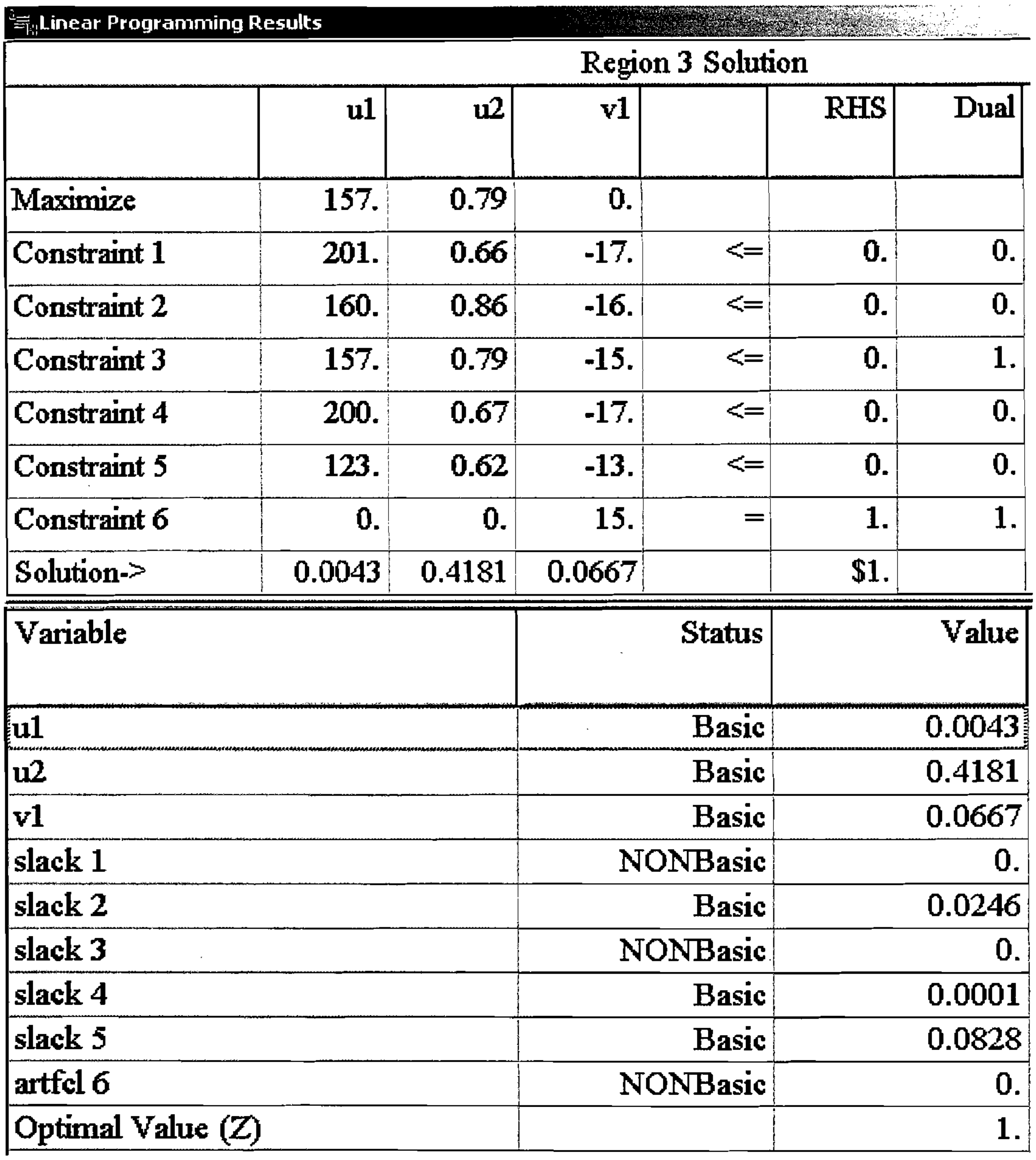


Region 2 Solution						
	ul	u2	vl		RHS	Dual
Maximize	160.	0.86	0.			
Constraint 1	201.	0.66	-17.	<=	0.	0.
Constraint 2	160.	0.86	-16.	<=	0.	1.
Constraint 3	157.	0.79	-15.	<=	0.	0.
Constraint 4	200.	0.67	-17.	<=	0.	0.
Constraint 5	123.	0.62	-13.	<=	0.	0.
Constraint 6	0.	0.	16.	=	1.	1.
Solution->	0.0019	0.8121	0.0625		$1.	

Variable	Status	Value
ul	Basic	0.0019
u2	Basic	0.8121
vl	Basic	0.0625
slack 1	Basic	0.1476
slack 2	NONBasic	0.
slack 3	NONBasic	0.
slack 4	Basic	0.1414
slack 5	Basic	0.0771
artfcl 6	NONBasic	0.
Optimal Value (Z)		1.

Region 2 (Bloem/Kby) has an optimal value of 1, which is an efficiency score. This region is 100% efficient.

Figure 5.3 LP Solution for Region 3



Linear Programming Results

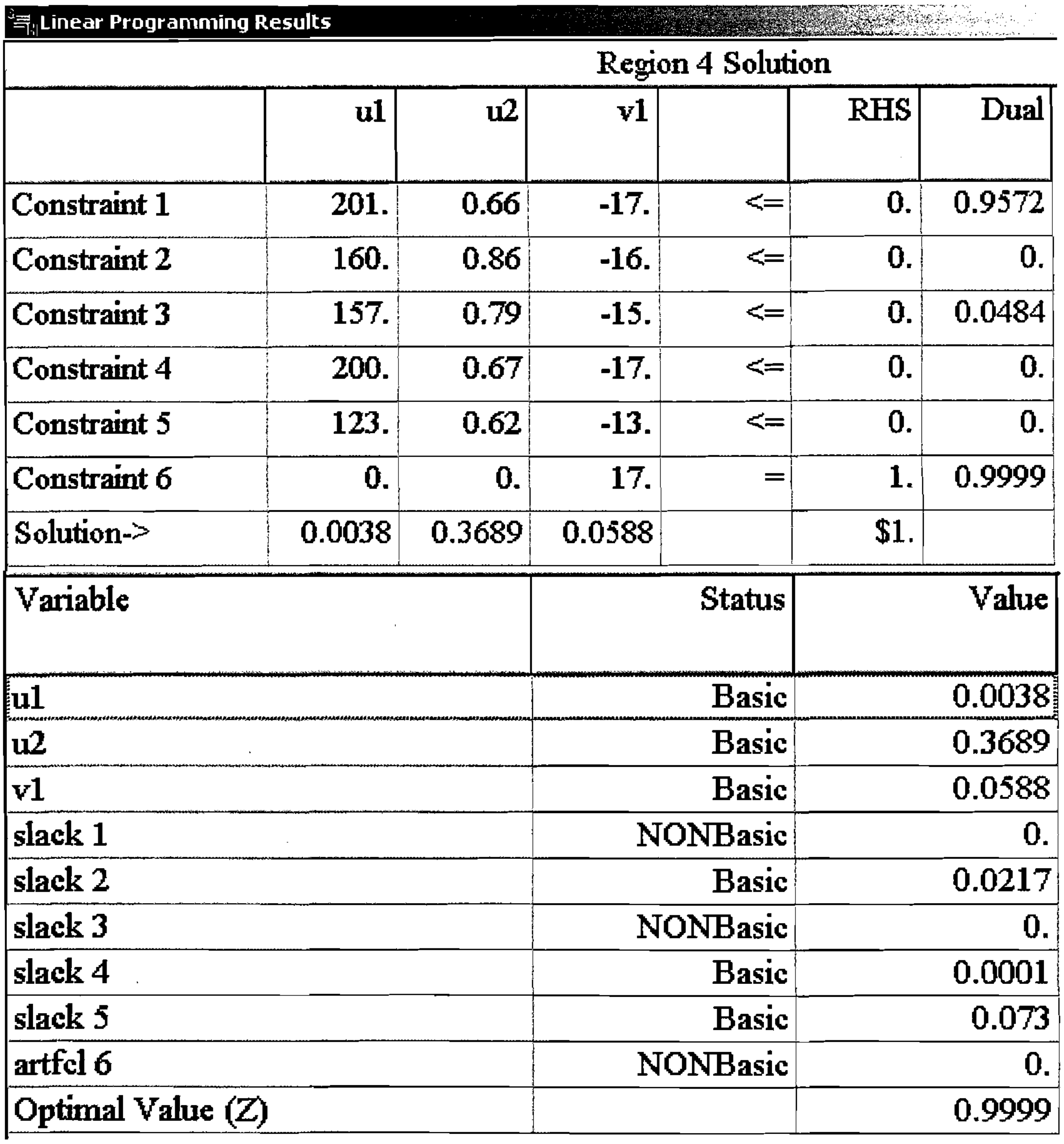
Region 3 Solution

	u1	u2	v1		RHS	Dual
Maximize	157.	0.79	0.			
Constraint 1	201.	0.66	-17.	<=	0.	0.
Constraint 2	160.	0.86	-16.	<=	0.	0.
Constraint 3	157.	0.79	-15.	<=	0.	1.
Constraint 4	200.	0.67	-17.	<=	0.	0.
Constraint 5	123.	0.62	-13.	<=	0.	0.
Constraint 6	0.	0.	15.	=	1.	1.
Solution->	0.0043	0.4181	0.0667		$1.	

Variable	Status	Value
u1	Basic	0.0043
u2	Basic	0.4181
v1	Basic	0.0667
slack 1	NONBasic	0.
slack 2	Basic	0.0246
slack 3	NONBasic	0.
slack 4	Basic	0.0001
slack 5	Basic	0.0828
artfcl 6	NONBasic	0.
Optimal Value (Z)		1.

Region 3 (Durban) has an optimal value of 1, which is an efficiency score. This region is 100% efficient.

Figure 5.4 LP Solution for Region 4



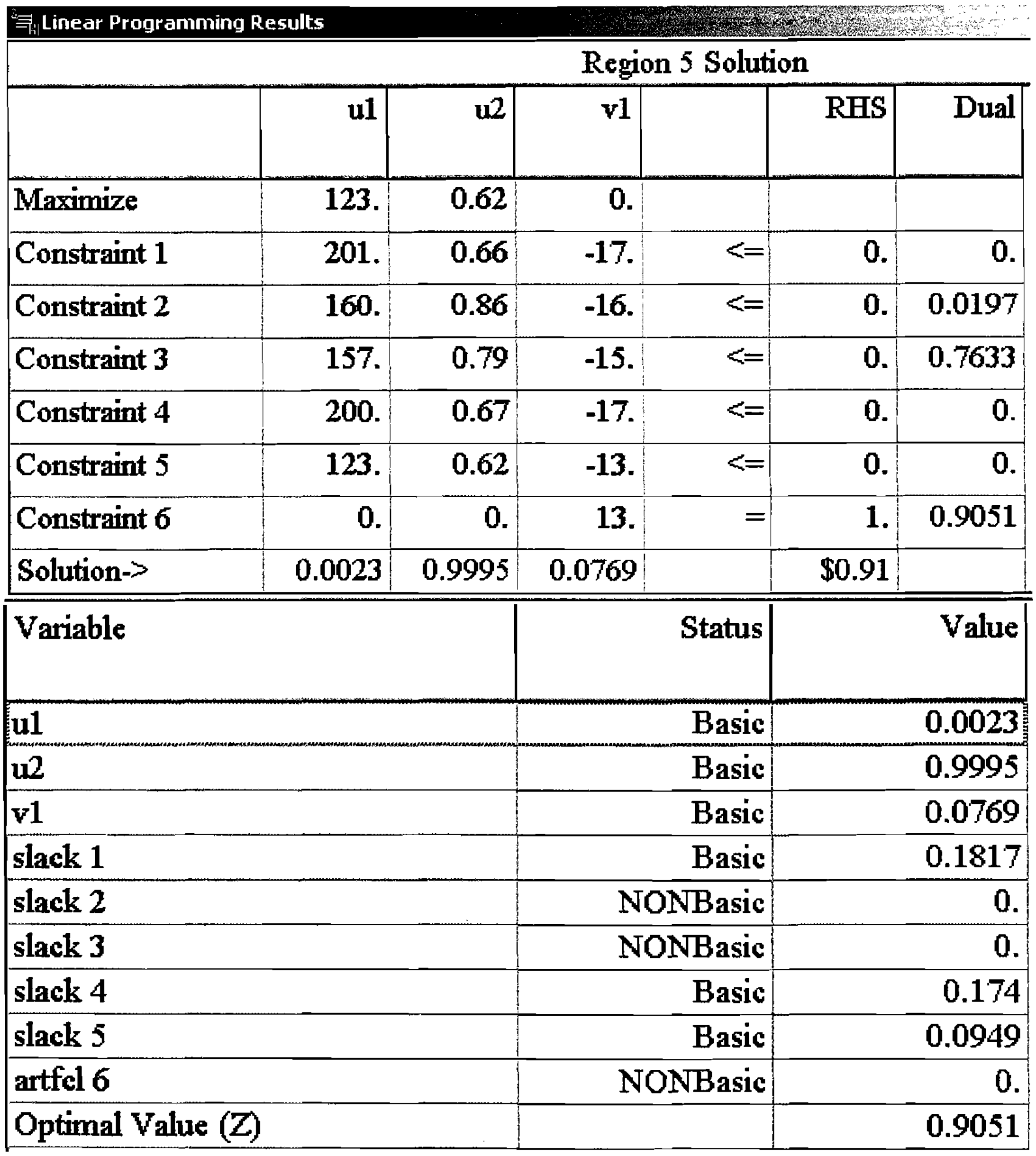
Region 4 Solution

	u1	u2	v1		RHS	Dual
Constraint 1	201.	0.66	-17.	<=	0.	0.9572
Constraint 2	160.	0.86	-16.	<=	0.	0.
Constraint 3	157.	0.79	-15.	<=	0.	0.0484
Constraint 4	200.	0.67	-17.	<=	0.	0.
Constraint 5	123.	0.62	-13.	<=	0.	0.
Constraint 6	0.	0.	17.	=	1.	0.9999
Solution->	0.0038	0.3689	0.0588		$1.	

Variable	Status	Value
u1	Basic	0.0038
u2	Basic	0.3689
v1	Basic	0.0588
slack 1	NONBasic	0.
slack 2	Basic	0.0217
slack 3	NONBasic	0.
slack 4	Basic	0.0001
slack 5	Basic	0.073
artfcl 6	NONBasic	0.
Optimal Value (Z)		0.9999

Region 4 (Johannesburg) has an optimal value of 0.9999 ≈ 1, which is an efficiency score. This region is almost 100% efficient.

Figure 5.5 LP Solution for Region 5



Region 5 Solution

	u1	u2	v1		RHS	Dual
Maximize	123.	0.62	0.			
Constraint 1	201.	0.66	-17.	<=	0.	0.
Constraint 2	160.	0.86	-16.	<=	0.	0.0197
Constraint 3	157.	0.79	-15.	<=	0.	0.7633
Constraint 4	200.	0.67	-17.	<=	0.	0.
Constraint 5	123.	0.62	-13.	<=	0.	0.
Constraint 6	0.	0.	13.	=	1.	0.9051
Solution->	0.0023	0.9995	0.0769		$0.91	

Variable	Status	Value
u1	Basic	0.0023
u2	Basic	0.9995
v1	Basic	0.0769
slack 1	Basic	0.1817
slack 2	NONBasic	0.
slack 3	NONBasic	0.
slack 4	Basic	0.174
slack 5	Basic	0.0949
artfcl 6	NONBasic	0.
Optimal Value (Z)		0.9051

Region 5 (Cape Town) has an optimal value of 0.9051, which is an efficiency score. This region is not 100% efficient.

5.6 SENSITIVITY ANALYSIS

With LP model, the status of the resources are secured directly from the optimum tableau by observing the values of the slack variables. In the case of region 5, which is inefficient, slacks 2 and 3, and the artificial variable are equal to zero, meaning they are scarce resources while all other variables are abundant. Increasing the abundant variable would not effect any change on the objective function, but increasing the scarce resources would.

5.7 CONCLUSION

The efficiency score or optimal solution determined manually using the simplex method for region 1 and that determined by software, is the same. The optimal solution (z) on both is 1. This is a 100% efficiency with output remaining constant. This means that inputs could be reduced by 0% with the same resulting output. To locate the best practicing region, the CCR primal model (discussed in section 3.5) which estimates a piecewise linear envelopment surface with constant returns to scale was selected. The determined scores for the regions, just like with ratio analysis still manifest that Cape Town is inefficient as compared to the other regions. Cape Town with an efficieny score of 0.9051 means that if this region were efficient, inputs could be reduced by 9.49% with the same resulting output. (Charnes *et. al.*, 1994).

CHAPTER 6

DATA ENVELOPMENT ANALYSIS

6.1 INTRODUCTION

Linear programming solution is follwed by applying Data Envelopment Analysis methodology to determine all efficiency scores at once. This chapter still illustrates the relationship between DEA and Linear Programming. But first using Table 2.2, and Formula 5.1 again, the solution is illustrated manually when s=1 and m=1, one input, output measure. The weights v_1 and u_2 and slack variables $s_1,\ldots,s_5$, are determined manually. This merely illustrates how weights and slacks are determined manually as these are extremely important in DEA.

Max $z = 0.66\ u_2$

Subject to
$$
\begin{aligned}
0.66\ u_2 - 17\ v_1 &\le 0\\
0.86\ u_2 - 16\ v_1 &\le 0\\
0.79\ u_2 - 15\ v_1 &\le 0\\
0.67\ u_2 - 17\ v_1 &\le 0\\
0.62\ u_2 - 13\ v_1 &\le 0 \qquad (1)\\
17\ v_1 &= 1
\end{aligned}
$$

Therefore $v_1 = 1/17$ (2)

From (1) & (2)

s.t.
$$
\begin{aligned}
0.66\ u_2 &\le 17\ (1/17) = 1\\
0.86\ u_2 &\le 16\ (1/17) = 16/17\\
0.79\ u_2 &\le 15\ (1/17) = 15/17\\
0.67\ u_2 &\le 17\ (1/17) = 1\\
0.62\ u_2 &\le 13\ (1/17) = 13/17 \qquad (3)
\end{aligned}
$$

From (3)

$$
\begin{aligned}
u_2 &\le 1.5151\\
u_2 &\le 1.0944\\
u_2 &\le 1.1160\\
u_2 &\le 1.4900\\
u_2 &\le 1.2330
\end{aligned}
$$

Therefore $u_2 \leq 1.0944$

Max z $= 0.66(16/17)(1/0.86)$

$u_2 = 0.7223$

LP in standard form:

Max z = $0.66\ u_2$

$$
\begin{aligned}
\text{s.t.}\quad & 0.66\ u_2 - 17\ v_1 + s_1 = 0 \\
& 0.86\ u_2 - 16\ v_1 + s_2 = 0 \\
& 0.79\ u_2 - 15\ v_1 + s_3 = 0 \\
& 0.67\ u_2 - 17\ v_1 + s_4 = 0 \\
& 0.62\ u_2 - 13\ v_1 + s_5 = 0 \\
& 17\ v_1 = 1
\end{aligned} \qquad (4)
$$

Therefore $v_1 = 1/17 = 0.0588$

$u_2 = (0.66/0.86)(16/17) = 0.7223$

Substituting v_1 and u_2 in (4)

$$
\begin{aligned}
s_1 &= 17\ v_1 - 0.66\ u_2 = 17(0.0588)-0.66(0.7223)=0.5229 \\
s_2 &= 16\ v_1 - 0.86\ u_2 = 16(0.0588)-0.86(0.7223)=0.3196 \\
s_3 &= 15\ v_1 - 0.79\ u_2 = 15(0.0588)-0.79(0.7223)=0.3114 \\
s_4 &= 17\ v_1 - 0.67\ u_2 = 17(0.0588)-0.67(0.7223)=0.5229 \\
s_5 &= 13\ v_1 - 0.62\ u_2 = 13(0.0588)-0.62(0.7223)=0.3166
\end{aligned}
$$

$s_1,\ldots,s_5$ are slack variables. Slack variables were explained in chapter 5. The minimum requirements on slack are met exactly if s = 0. All the regions have slack variables greater than 0, meaning all the input resources are not used up completely. Increasing these resources (employees) will have no effect on the optimal value since already an excess number of employees exists.

6.2 WEIGHTING

In data envelopment analysis, "weighting" allows a level of control over the efficiency scores. In the above section, weights v_1 and u_2 were calculated manually.Weighting ensures that at least some part of the efficiency score of every unit takes account of a particular input/output variable. For example, in this study, the two outputs, resolved events and client satisfaction are measured differently. Without weighting, a unit may be judged as efficient because it is doing well on resolved events per member of staff and another unit be judged as efficient on client satisfaction. But if weights v_i's are added to inputs, and weights u_r's are added to outputs then every unit is judged at least to some extent on something that is critical to business. (Anderson *et. al.*, 2006).

To use DEA to measure the relative efficiencies of all regions, the Linear Programming model is used to construct a hypothetical composite region based on the outputs and inputs for the regions in the problem. The heart of the analysis lies in finding the best virtual producer for each real producer, while the procedure for finding the best virtual producer is formulated as a linear program. With the DEA, the Linear Programming models can be either solved for each region for optimal efficiency or the relative efficiencies of all the regions determined at the same time. In this chapter, the relative efficiencies of all the regions are determined at the same time. (Anderson *et.al.*, 2006)

6.3 DATA ENVELOPMENT ANALYSIS SOLUTION

DEA is used to evaluate the efficiency of information technology service delivery regions denoted as regions 1 to 5, (DMU's 1 to 5, according to Charnes *et.al.* (1994)), which also are homogeneous with some decision autonomy. Each region is characterized by "consuming" one input and by producing two outputs. A DEA model is formulated that uses these factors (inputs and outputs) to compute the efficiency degree of a particular region when this region is compared with all the other regions. Like the ratio analysis, the regions that are considered efficient relative to the other regions belong to the frontier and therefore, can be used as performance benchmarks to study the regions that are operating inefficiently. Regions that are inefficient do not belong to the frontier.

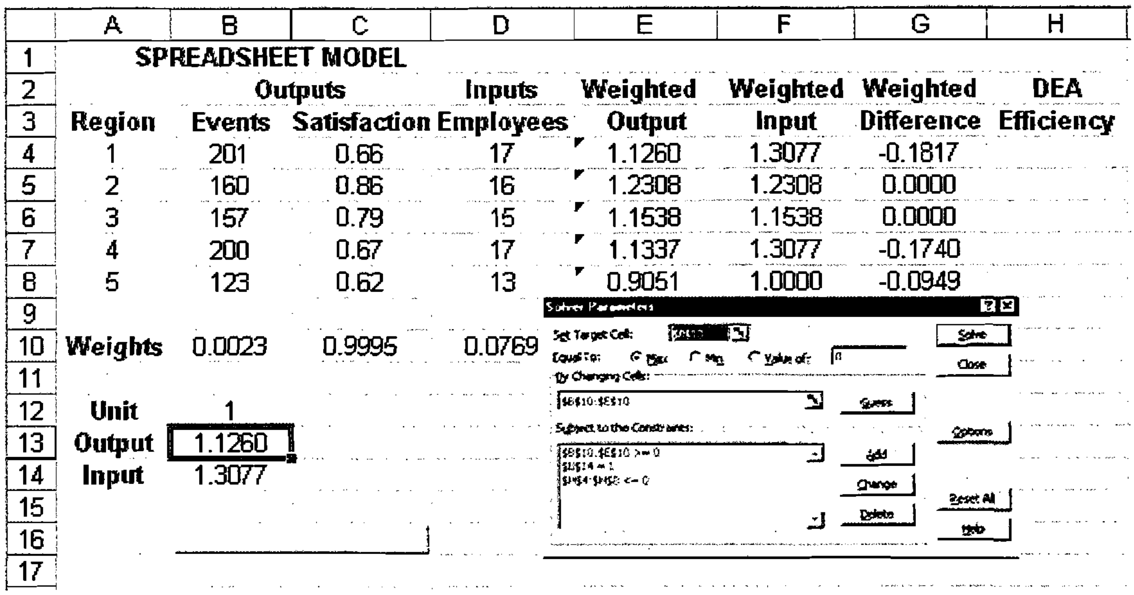
Using the Linear Programming model, a hypothetical composite is constructed or a composite region, based on the outputs and inputs for all operating regions with the same goals. For each of the five regions' output measures, the output for the composite region is determined by computing a linear combination of the corresponding outputs for all five regions. Again, for each of the input measures, the input for the composite region is determined by using the same weights to compute a linear combination of the corresponding inputs for all five regions. Constraints in the Linear Programming model require all outputs for the composite region to be greater than or equal to the outputs of region 1, the region being evaluated. If the inputs for the composite region can be shown to be less than the inputs for region 1, the composite region is shown to have the same or more output for less input. In this case, the model shows that the composite region is more efficient than region 1. In other words, the region being evaluated, is less efficient than the composite region. Because the composite region is based on all five regions, the region being evaluated can be judged relatively efficient when compared to the other regions in the group.

To illustrate how DEA works, let's consider the five regions. Each region converts one input into two outputs. The only input used by each region is the employees. The two outputs produced by each region are, the number of resolved events and client satisfaction. The inputs and outputs for the five regions are as previously given in Table 2.2.

The DEA approach applies the following point of view to determine if the region is efficient. That no region can be more than 100% efficient. Thus the efficiency of each region must be less than or equal to 1. According to Anderson *et.al.* (2006) as each LP is solved, the region under investigation, cannot select weights for itself that would cause the efficiency for any region (including itself) to be greater than 100%. Thus, for each individual region, the weighted sum of the region's outputs is required to be less than or equal to the weighted sum of its inputs (so the ratio of weighted outputs to weighted inputs does not exceed 100%). The excel's Solver is first used to determine the efficiency score of region 1 (Figure 6.1), and thereafter the efficiency scores of all the regions determined at the same time.

The efficiency of region 1 is shown enclosed in a rectangle in Figure 6.1. The value in cell B12 is 1, meaning region 1, therefore the efficiency score of each region can be determined by changing the value in B12 to be 2, 3, 4 or 5. But doing this is rather cumbersome, therefore the efficiencies of all the regions are determined at once by writing a macro-in-excel to carry out this process.

Figure 6.1 Optimal DEA Solution for Region 1



	A	B	C	D	E	F	G	H
1	SPREADSHEET MODEL							
2		Outputs		Inputs	Weighted	Weighted	Weighted	DEA
3	Region	Events	Satisfaction	Employees	Output	Input	Difference	Efficiency
4	1	201	0.66	17	1.1260	1.3077	-0.1817	
5	2	160	0.86	16	1.2308	1.2308	0.0000	
6	3	157	0.79	15	1.1538	1.1538	0.0000	
7	4	200	0.67	17	1.1337	1.3077	-0.1740	
8	5	123	0.62	13	0.9051	1.0000	-0.0949	
9								
10	Weights	0.0023	0.9995	0.0769				
11								
12	Unit	1						
13	Output	1.1260						
14	Input	1.3077						
15								
16								
17								

Solver Parameters

Set Target Cell: [illegible]

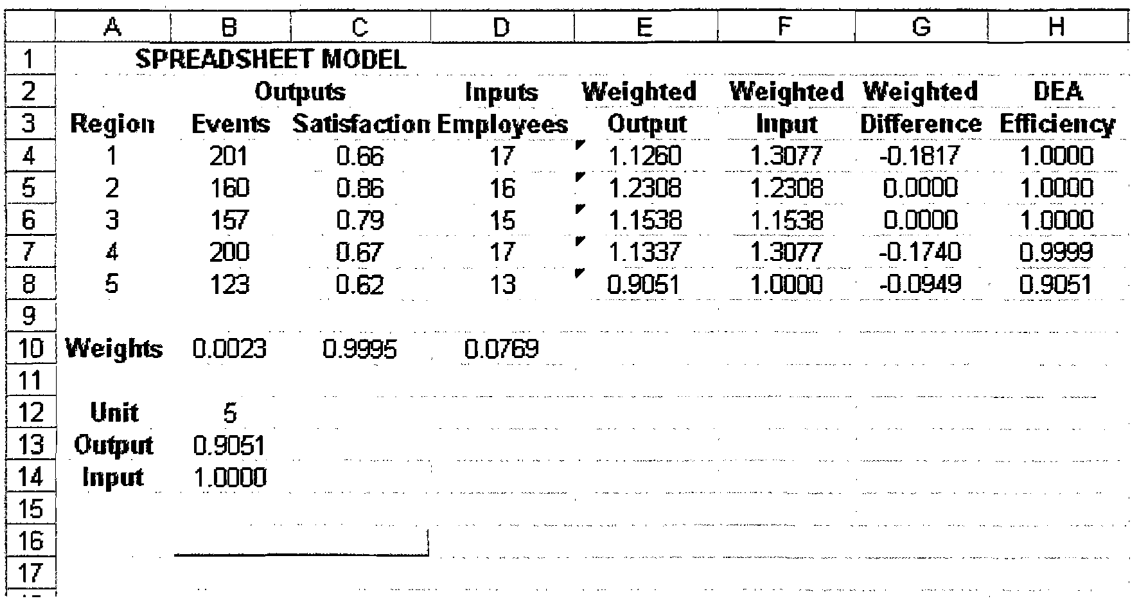
Equal To: ● Max ○ Min ○ Value of: 0

By Changing Cells: $B$10:$E$10

Subject to the Constraints:

$B$10:$E$10 >= 0
$B$14 = 1
$G$4:$G$8 <= 0

Figure 6.2 Efficiency Scores for all the Regions



	A	B	C	D	E	F	G	H
1	SPREADSHEET MODEL							
2		Outputs		Inputs	Weighted	Weighted	Weighted	DEA
3	Region	Events	Satisfaction	Employees	Output	Input	Difference	Efficiency
4	1	201	0.66	17	1.1260	1.3077	-0.1817	1.0000
5	2	160	0.86	16	1.2308	1.2308	0.0000	1.0000
6	3	157	0.79	15	1.1538	1.1538	0.0000	1.0000
7	4	200	0.67	17	1.1337	1.3077	-0.1740	0.9999
8	5	123	0.62	13	0.9051	1.0000	-0.0949	0.9051
9								
10	Weights	0.0023	0.9995	0.0769				
11								
12	Unit	5						
13	Output	0.9051						
14	Input	1.0000						
15								
16								
17								

The efficiency scores for all the regions are shown in the last column (H) in Figure 6.2. These efficiency scores are exactly the same as those determined by the quantitative methods for windows software in chapter 5. This confirms that DEA is Linear Programming based.

6.4 THE MODEL

In Figure 6.1, cells B10 through D10 represent the weights for each of the input and output variables. The weighted output for each region is computed in column E as follows.

Formula for cell E4=SUM PRODUCT(B4:C4,$B$10:$C$10) and copied to E5 through to E8. Similarly, the weighted input for each region is computed in column F as =SUMPRODUCT(D4:D4,$D$10:$D$10) and copied to F5 through to F8. The differences between the weighted outputs and weighted inputs are computed in column G. Solver was instructed to constrain these values to be less than or equal to 0. Formula for cell G4 =E4-F4 and copied to G5 through to G8.

The weighted output for region 1 (computed in cell E4) implements the appropriate objective function and is used as the set cell for Solver in this problem. Similarly, the weighted input for region 1 is computed in cell F4 and is constrained to equal 1 (as specified by the input constraint for region 1 above). However, because the need was to solve a separate LP problem for each of the five regions, it was more convenient to handle the objective function and input constraint in a slightly different manner. Therefore, cell B12 was reserved to indicate the region number under investigation. Cell B13 contains a formula that returns the weighted output for this region from the list of weighted outputs in column E. Formula for cell B13 = INDEX (E4:E8,B12,1). In general, the function INDEX (range, row number, column number) returns the value in the specified row number and column number of the given range. Because cell B12 contains number 1, the previous formula returns the value in the first row and first column of the range E4:E8 or the value in E4. Therefore, as long as the value of cell B12 represents a valid region number from 1 to 5, the value in cell B13 will represent the appropriate objective function for the DEA model for that region. Similarly, the input constraint requiring the weighted inputs for the region in question to equal 1, is implemented in cell B14 as follows. Formula for cell B14 = INDEX (F4:F8,B12,1). Therefore, for whatever region number listed in cell B12, cell B13 represents the appropriate objective function to be maximized and cell B14 represents the weighted input that must be constrained to equal 1. This actually simplifies the process of solving the required series of DEA models.

6.5 SOLVING THE MODEL

To solve the model, the set cells, variable cells, and constraints are specified as in Figure 6.1. Exactly the same Solver settings would be used to find the optimal DEA weights for any other region. To do this, the value in Cell B12 is changed manually to 2, 3,...,5 and Solver is used to re- optimize the worksheet for each region and record their efficiency scores in column H. But dealing with many regions, this method could be quite cumbersome, therefore a macro was written in excel to carry out this process as following.

Turned on the control tool box command bar and placed a command button on the worksheet.

1.Clicked View, Toolbars, Control Toolbox.

2.Clicked the Command Button icon on the Control Toolbox.

3.Clicked and dragged on the worksheet to draw a command button.

A few properties of the newly created command button were changed as follows.

1. Clicked the Command Button to make sure it was selected.
2. Clicked the Properties icon on the Control Toolbox.

These actions cause the Properties window to drop down the menu. This window lists several properties (or attributes) of the command button that can be changed to customize its appearance and behaviour. In this problem, the command button's property values were changed as following.

Table 6.1 Command Button Property Values

Property	**New Value**
(Name)	DEA
Caption	Run DEA
TakeFocusOnClick	False

Then, double-clicked the command button to launch the Excel's Visual Basic Editor and brought up the code window for the command button's click event. Initially, the click event will not have any commands in it. The following statements were inserted.

```
Private Sub DEA_Click()
For unit = 1 To 5
Range("B12") = unit
SolverSolve UserFinish:=True
Range("H" & 3 + unit) = Range("B13")
Next unit

End Sub
```

In the above code, the For and next statements define a loop of code that will be repeated 5 times. During the first execution of the loop, the variable "unit" will equal 1. During the second execution of the loop, the variable "unit" will equal 2, and so on. During each execution of the loop, the following operations take place.

Table 6.2 Loop Execution Operations

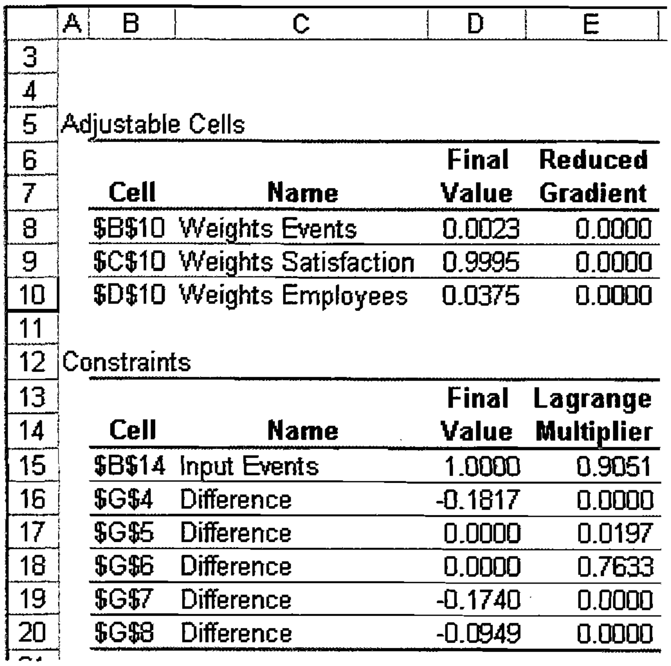
Macro Statement	**Purpose**
Range("B12") = unit	Places the current value of "unit"(the number 1,2,3,..., 5) into cell B12 on the worksheet.
SolverSolve UserFinish:=True	Tells Solver to solve the problem without displaying the usual Solver results dialog box.
Range("H" & 3 + unit) = Range("B13")	Takes the optimal objective function value in cell B13 and places it in row "3 + unit" (that is, row 4,5,..., or 8) in column H.

Note: Toggle back and forth between Excel and the Visual Basic editor by pressing Alt+F11.

6.6 SENSITIVITY ANALYSIS

In Figure 6.3 below, the absolute value of the shadow prices for the "difference" constraints (cell E16 through E20) are the weights that should create a composite region that is more efficient than the region in question; that is, region 5. In other words, a linear combination of 1.97% of region 2, plus 76.33% of region 3 produce a hypothetical composite region with outputs greater than or equal to those of region 5 and requiring less input than region 5. The assumption in DEA is, that region 5 should have been able to achieve this same level of performance (Anderson *et.al.*, 2006).

Figure 6.3 Sensitivity Analysis



	A	B	C	D	E
3					
4					
5	Adjustable Cells				
6				**Final**	**Reduced**
7		**Cell**	**Name**	**Value**	**Gradient**
8		$B$10	Weights Events	0.0023	0.0000
9		$C$10	Weights Satisfaction	0.9995	0.0000
10		$D$10	Weights Employees	0.0375	0.0000
11					
12	Constraints				
13				**Final**	**Lagrange**
14		**Cell**	**Name**	**Value**	**Multiplier**
15		$B$14	Input Events	1.0000	0.9051
16		$G$4	Difference	-0.1817	0.0000
17		$G$5	Difference	0.0000	0.0197
18		$G$6	Difference	0.0000	0.7633
19		$G$7	Difference	-0.1740	0.0000
20		$G$8	Difference	-0.0949	0.0000

Sensitivity analysis and post optimality analysis allow the decision maker to determine how the final solution to the problem will change when the input data or the model changes. This type of analysis is very important when the input data or model has not been specified properly. A sensitive solution is one in which the results of the solution to the problem will change drastically or by a large amount with small changes in the data or in the model. When the model is not sensitive, the results or solutions to the model will not change significantly with changes in the input data or in the model. Models that are very sensitive require that the input data and the model itself be thoroughly tested to make sure that both are very accurate and consistent with the problem statement (Taha, 2007).

6.7 CONCLUSION

The efficiency scores obtained with DEA are exactly the same as those obtained with LP analysis in chapter 5. The reason being that DEA is LP based. To use DEA to measure the relative efficiency of regions, a Linear programming model was used to construct a hypothetical composite region based on the outputs and inputs for the five regions in the problem.

This study demonstrated how to use Linear programming to assist in the decision-making process. It was illustrated how Linear programming can be applied to problems and DEA. DEA is therefore a Linear programming application used to measure the relative efficiency of operating units with the same goals and objectives. With LP the individual scores are determined one at a time but with DEA all the scores are determined at the same time.

CHAPTER 7

DISCUSSION OF RESULTS

7.1 INTRODUCTION

This chapter discusses the findings from the practical problems, and also interprets the results. It also discusses the research methodology

7.2 STATISTICAL TECHNIQUES

7.2.1 THE SCATTER DIAGRAM

Figure 2.1 (Scatter diagram) shows a positive linear relationship between employees and the number of resolved events. Since this is a positive relationship, it means more people do a greater amount of work than fewer people.

7.2.2 THE CHI-SQUARE DISTRIBUTION

The null hypothesis could not be rejected since for lambda =6.5 to 8 (arrival rate), the chi-square values are less than the critical chi-square value of 3.481. In other words the different values where the null hypothesis is not rejected at the 5% significant level are part of a 95 % confidence interval, so any of these values are a possibility in the future, given that the population is stationary. Figure 2.2 (a-g) and the χ^2 values within the 95% confidence interval therefore confirm that the data do not present sufficient evidence to contradict the hypothesis that F possesses a Poisson distribution.

7.2.3 THE FREQUENCY DISTRIBUTION

Figure 2.3 depicts the presentation of data, simply how data are distributed, which then is still a Poisson distribution. Its purpose was to confirm the distribution of the sample data. Likewise, the chi-square test's purpose was to fit possible Poisson distributions. The data collected from the service desk does indeed approximate a Poisson distribution.

7.3 QUEUING THEORY

The conclusion drawn from the queuing theory, recommends that more staff be hired, but management has to decide how many with informed recommendation from the queuing model, particularly considering the trade-off between the cost and customer satisfaction. From the computations of the abandoned calls rate, the following conclusions are drawn. With 17 employees (channels), 91.39% of calls is answered and 8.61% abandoned. With 23 employees (channels), 0.74% of calls is abandoned and with 25 employees (channels), 0.24% is abandoned.

7.4 THE RATIO ANALYSIS

7.4.1 SINGLE INPUT, OUTPUT MEASURE

Using the traditional ratio analysis, single input, single output measure, region 1 which is Pretoria, outperformed the other regions. On the other hand, a different conclusion was reached with regard to the output client satisfaction and input employees where region 2, which is Bloem/Kby, outperformed the other regions. Pretoria can be used as a benchmark to set the target for the other regions with respect to the output number of resolved events. Bloem/Kby also can be used as a benchmark to set the target for the other regions with respect to the output client satisfaction.

These conflicting results with ratio analysis is probably due to the fact that "weighting' was not applied. "Weighting" allows a level of control over the efficiency scores. It ensures that at least some part of the efficiency score of every unit takes account of a particular input/output variable. For example, in this study, the two outputs, resolved events and client satisfaction are measured differently. Without weighting, a unit may be judged as efficient because it is doing well on resolved events per member of staff and another unit be judged as efficient on client satisfaction. But if weights are added to inputs, and outputs then every unit is judged at least to some extent on something that is critical to business (Anderson *et.al.*, 2006).

7.4.2 EXTENDED RESOURCES

Analysis with one input and two outputs could be achieved by the graphical method in chapter 4. The efficiency frontier line is the maximum combination of outputs that can be produced for a given set of inputs. The efficient frontier, derived from the examples of best practice contained in the data considered, represents the performance that the region, in this case Cape Town, not on the efficient frontier, could try to achieve. Hence data envelopment, because the efficient frontier envelopes (encloses) all the data. All the regions on the efficient frontier are 100% efficient. Therefore, all the regions are efficient except Cape Town, which is inefficient.

7.5 THE LINEAR PROGRAMMING SOLUTION

When the LP models for the regions are evaluated to obtain the optimal objective values, being the efficiencies of the regions, regions 1, 2,3 all have the optimal values of 1, while region 4 has the optimal value of 0.9999 ≅ 1. Region 5 has the optimal value of 0.9051 (Figures 5.1 to 5.5). The linear programming, like the ratio analysis, confirms that Cape Town is inefficient.

7.6 THE DATA ENVELOPMENT ANALYSIS SOLUTION

The solutions shown in Figures 6.1 and 6.2, indicate that regions 1, 2, 3 are on the envelope (frontier), or operate at 100% efficiency (in the DEA sense), which means that they are efficient and they are benchmarks to other regions. Region 4 has a score of 0.9999, which is almost 100% and therefore, can be regarded as also efficient. Region 5, on the other hand, is inefficient with a score of 0.9051 or 90.51%. An efficiency of 100% does not necessarily mean that the region is operating in the best possible way. It simply means, that no linear combination of the other regions in the study results in a composite region that produces at least as much output by using the same or less input. On the other hand, for regions that are DEA efficient, there exists a linear combination of efficient regions that results in a composite region that produces at least as much output by using the same or less input than the inefficient region. The idea is, that an inefficient region should be able to operate as efficiently as this hypothetical composite region formed from a linear combination of the efficient regions. (Anderson et.al., 2006) For instance, region 5 has an efficiency score of 0.9051 and is, therefore, inefficient relative to other regions.

7.7 SENSITIVITY ANALYSIS

Ratio analysis shows that customer satisfaction can be improved by appointing more capable staff. Figure 4.3 and Table 4.9 show that it is possible to move beyond the efficient frontier created by DEA ratio analysis. Figures 4.4 and 4.5 are the plots of residuals. This was to assess the model adequacy by checking whether the model assumptions are satisfied. The basic assumption is that the residuals are uncorrelated with zero mean and constant variance.

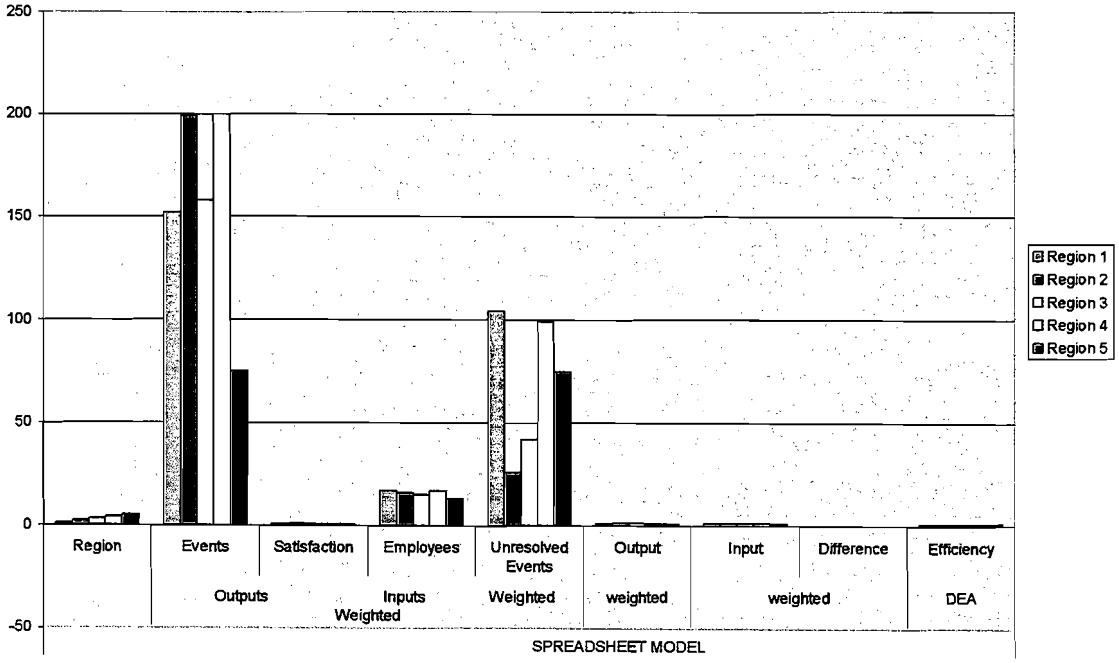
The residuals in figure 4.4 look evenly distributed with no pattern, meaning that they are uncorrelated. Figure 4.5, the normal probability plot shows no serious deviations from the fourty five degree line. We conclude from the plots that the residuals are normally distributed. The regression model is adequate.

In Table 4.9, the current number of employees cannot resolve all the logged events. When there are more employees, statistics shows that all logged events can be resolved. This is the number of employees necessary to resolve all the logged events, using the extrapolated extrapolating the regression equation. The assumption is that the extra people can resolve the unresolved events at the same rate as the existing employees, which may not be valid if the unresolved events are more difficult to resolve than the resolved events. This may be a way to move beyond the current efficient frontier and create new quality standards.

With LP model, the status of the resources are secured directly from the optimum tableau by observing the values of the slack variables. In the case of region 5, which is inefficient, slacks 2 and 3, and the artificial variable are equal to zero, meaning they are scarce resources while all other variables are abundant. Increasing the abundant variable would not effect any change on the objective function, but increasing the scarce resources would do.

The DEA model, Figure 6.3 shows, that a linear combination of 1.97% of region 2, plus 76.33% of region 3, produces a hypothetical composite region with outputs greater than or equal to those of region 5 and requiring less input than region 5. The assumption in DEA is that region 5 should have been able to achieve this same level of performance. Like Ratio Analysis and Linear Programming, the DEA model also confirms Cape Town to be inefficient. This was expected since the three are a family. Briefly all the three methodologies confirm Cape Town to be inefficient.

7.8 SUMMARY ANALYSIS OF THE ACTUAL PERFORMANCE



From this bar graph, the number of resolved events is almost equal to the number of unresolved events in region 5. This is not the case with other regions. This manifests that region 5 is inefficient relative to others.

7.9 THE RESEACH METHODOLOGY

The DEA's ratio analysis methodology (the primal formulation according to Charnes *et.al.* (1994)) cannot be used when there are many inputs and outputs, but the dual formulation simplifies the calculations when there are many inputs and outputs. Since in this study there are few inputs and outputs, the primal formulation was used. The Linear programming methodology is used for optimizing a single objective only. In this study it was illustrated how to determine the efficiency of one region (DMU) at a time. DEA was used to illustrate how to determine the efficiencies of all the regions (DMU's) at once. These methodologies proved to be reliable since they all gave identical results. This illustrated that these methodologies are a family.

7.10 VERIFICATION OF RESULTS

According to Marcoulides (1998: 136), the total number of inputs and outputs depends on the total number of DMU's to be compared. If the number of variables chosen is more than the number of DMU's the model's discriminating power suffers, that is the model cannot discriminate efficient from inefficient. A rule of thumb is that the total number of units (DMU's) available must exceed the product of the number of inputs and outputs. For example if there are two inputs and two outputs, like in this study's case, more than four DMU's would be required for the model to adequately discriminate efficient from inefficient DMUs.

Most importantly, according to Bowlin (1995:19), choosing correct inputs and outputs in DEA is important for the effective interpretation, use and acceptance of the results by management.

(1) There has to be some basis for believing that relationship exists between inputs and outputs. (Section 2.4 Assessing Sample Independence).

(2) It is advisable to stay close to the kind of input and output measures currently used by management for performance evaluation, since they are already familiar with this kind of measures and they should fully measure the activities of the organization under evaluation. (Section 2.2 Data Collection).

7.11 SUMMARISED OUTCOMES OF THE STUDY

The results of these solutions tell us that region 5 (Cape Town), is less efficient when compared to other regions. Secondly, the employees working at the central service desk, cannot handle the workload. Management has to decide how many additional staff members to hire from the informed queuing model's recommendation and also determine the organizational changes that an inefficient region will need to become efficient.

Now a natural question to ask, is how Cape Town can improve its performance in practice. There is a natural obvious answer to this question. Obviously it should use less input to produce its current output, or produce more output using the current level of input or simply a combination of both. But the most appealing answer is, that a region can in fact learn from the best practice frontier situated along the frontier line. The learning can take place simply by contacting those regions or the analyst can initiate seminars where the best practice regions explain the way they do their things.

7.12 LESSONS LEARNED

The two models used, Queuing and DEA, give different insights. The queuing model tells us how many staff members to hire to manage the workload. Data Envelopment Analysis determines the relative performance efficiencies of the homogeneous regions. The excellency of the service could be determined by determining client satisfaction ratio.

There is a relationship between DEA efficiency scores and ratio analysis. In this paper, it was demonstrated, that DEA can augment the traditional ratio analysis. The traditional ratio analysis selected Cape Town as being inefficient. DEA also selected Cape Town as being inefficient. Since the DEA model is linear programming based, it was also expected of them to give the similar results.

Managing data, the regular collection and reporting of information for performance measurement (i.e. quality, effectiveness, and efficiency) is one of the most important aspects of competitiveness. Benchmarking which basically is the comparison of services, work processes and products against "best practice" or best performers, where the overall aim is to improve the performances of regions, is also a tool for competitiveness.

DEA, as a tool for benchmarking, has been found to be particularly suitable in solving the following three basic performance questions that any service delivery region was faced with.

- How well we are doing relative to others doing the same as we do?
- What do we need to improve?
- Who are the best-in-class performers for benchmarking purposes?

7.13 CONTRIBUTION TO THE ORGANIZATION

DEA allows "data to speak for itself". DEA solutions will be used to guide any managerial action, for example, goal setting. It is important to recognize that the calculated improvements in inputs and/or outputs are indicative of potential performance increases by the region located below the efficiency frontier. The DEA solutions can be used to direct management attention towards developing a deeper understanding of why some regions are located on the frontier and others located below the frontier. Briefly why some are efficient and some are inefficient. Once management has this understanding, it is up to management to identify formal structures, assignable causes, or other organizational factors that could account for these observed differences. The objective is to assign organizational meaning to these observed differences in performance and to determine the organizational changes that an inefficient region (Cape Town) will need to become efficient and thereby satisfy customer-defined quality.

7.14 CONTRIBUTION TO THE OPERATIONS RESEARCH

The understanding and satisfying customer-defined quality, emerged in the late 1990 with the goal of covering all the customer requirements in the design of service. Very little of this methodology has been done and tested in the service industry. This study will make a significant contribution to the body of knowledge among the Operations Researchers, in their search for quality improvement techniques through the involvement of the missing link between the service provider and the customer, i.e. "the understanding of customer requirements", referred to (in this study) as SLA.

7.15 CONCLUDING REMARKS

DEA gives the same results given by ratio analysis and linear programming, that, Cape Town is inefficient. The goal of DEA is to identify regions that are relatively inefficient. The method does not necessarily identify the operating regions that are efficient. Just because the efficiency is 1, it cannot be concluded that the region being analysed, is relatively efficient. Any region that has the largest output in respect of any one of the output measures cannot be judged relatively inefficient. DEA constructs a frontier or an envelope composed of best practice performers and then measures efficiency relative to that frontier. Therefore DEA, is an efficiency tool.

BIBLIOGRAPHY

ALDER, H & ROESSLER, E. 1975. Introduction to probability and statistics. 6th ed. San Francisco, W.H. Freeman and Company. 426p.

ANDERSON, D., SWEENEY, D. & WILLIAMS, T. 2006. Quantitative methods for business. 8th ed. Cincinnati, Oh.: South-Western College. 822 p.

BANKER, R. 1984. Estimating the most productive scale size using Data Envelopment Analysis. *European journal of operations research,* 17:35-44.

BANKER, R., CHARNES, A. & COOPER, W. 1984. Some models for estimating technical and scale efficiency in Data Envelopment Analysis. *Management science,* 30:1078-1092.

BANKER, R., CHARNES, A., COOPER, W. & SCHINNAR, A. 1981. A bi-extremal principle for frontier estimation and efficiency evaluations. *Management science,* 27:1370-1382.

BEASLEY J.E. 2007. Data envelopment analysis. http://people.brunel.ac.uk/~mastjjb/jeb/or/contents.html Date of access: 01 Aug. 2007.

BOWLIN, W.F. 1995. Measuring performance: an introduction to data envelopment analysis (DEA). Ceder Falls, Ia: University of Northern Iowa, Department of Accounting. 27 p. (Technical report.)

BROWN, G., ELLIS, J., GRAVES, W. & ROMAN, D. 1987. Real time: wide area dispatch of mobile tank trucks. 107-120, 17 January-February.

CARLSON, S. 1999. The pure theory of production. London: King.

CHARNES, A., COOPER, W.W., LEWIN, A.Y. & SEIFORD, L.M., *eds*. 1994. Data envelopment analysis: theory, methodology and applications. Dordrecth, Holland: Kluwer.

COOPER, W.W., THOMPSON, R.G. & THRALL, R.M. 1996. Extensions and new developments in DEA. *Annals of operations research*, 66:3-45.

DIBELLA, A.J. & NEVIS, E.C. 1998. How organizations learn. San Francisco, Calif.: Jossey-Bass. 216 p.

DORIAN, P. 1999. Data preparation for data mining. New York: Academic Press. 540 p.

FARELL, M. 1957. The measurement of productive efficiency. *Journal of the Royal Statistical Society, Series A (General),* 120:253-281.

GANLEY, J.A. & CUBBIN, J.S. 1992. Public sector efficiency measurement: applications of data envelopment analysis. Amsterdam: Elsevier Science. 180 p.

HALL, O.P. 1993. Computer models for operations management. Redwood City, Calif.: Addison-Wesley. 197 p.

JOHNSTON, J. 1960. Statistical cost analysis. New York: McGraw-Hill.

MARCOULIDES, G.A. 1998. Modern methods for business research. London: Erlbaum.

PETERS, G. 1994. Benchmarking customer service. London: Pitman.

RENDER, B., STAIR, R.M. & HANNA, M.E. 2006. Quantitative analysis for management. 8th ed. Upper Saddle River, N.J.: Pearson/ Prentice Hall. 726 p.

RICHMOND, J. 1974. Estimating the efficiency of production. *International economic review,* 15:515-521.

TAHA, H.A. 2007. Operations research: an introduction. 8th ed. New York: Macmillan. 813 p.

TWO CROWS CORPORATION. 2007. Introduction to data mining and knowledge discovery. 2nd ed. Falls Road, Potomac. http://www.twocrows.com Date of access: 03 June 2007.

WHITTEN, J., BENTLEY, L. & DITTMAN, K. 2006. System analysis and design methods. 6th ed. New York: McGraw-Hill. 724 p.

WINSTON, W. 2004. Operations research: applications and algorithms. 4th ed. Pacific Grove, Calif.: Thomson-Brooks/Cole. 1318 p.

ZELENY, M. 1974. Linear multi objective programming. Vienna: Springer-Verlag. 729 p.

APPENDIX A

LIST OF TERMS AND CONCEPTS

Allocative efficiency

The possible reduction in cost by using appropriate input mixes.

System users

The people who use or are affected by the information system on a regular basis-capturing, validating, entering, responding to, storing and exchanging data and information. A common synonym is client slope of the line. (Whitten, Bentley & Dittman, 2006:8, 11).

Information system

An arrangement of people, data, processes, information presentation, and information technology that interact to support and prove day to day operations in a business as well as support the problem solving and decision making needs of management and users. (Whitten, Bentley & Dittman, 2006:8, 11).

Decision Making Unit

Term coined by Charnes *et.al.*, in order to avoid strictures that long usage has accorded such terms as "plant" and "firm" as organizations concerned with input and output decisions in the literature of economics.

Dual

Consisting of two parts or elements

Efficiency

Success in producing as large as possible an ouput from a given set of inputs. (Farrel, 1957).

Event

An instateneous occurrence that changes the state of the information system.

Events resolved in time	According to service level agreement (SLA), a fault has to be resolved within 2 days,a request in four days..
Envelope	A curve tangent to each of a family of curves or a surface tangent to each of a family of surfaces.
Frontier	The farthermost limits.
Isoquant’	Various combinations of two factors that a perfectly efficicient DMU might use to produce unit output.
Linear program	A mathematical method of solving problems by means of linear functions where the variables involved are subject to constraints.
Optimal	Most desirable or the best possible outcome.
Overall efficiency	The measure that indicates the possible reduction in cost due to a change from observed input quantities to cost minimizing input quantities.
Parametric	An arbitrary contant whose value characterizes a member of a system or a quantity that describes a statistical population.
Productivity	The ratio of outputs to inputs.
Technical efficiency	The measure of the proportion of inputs actually required to produce required outputs or relative efficiency of a DMU in using less inputs to generate same amount of output when compared with other DMUs which use more.

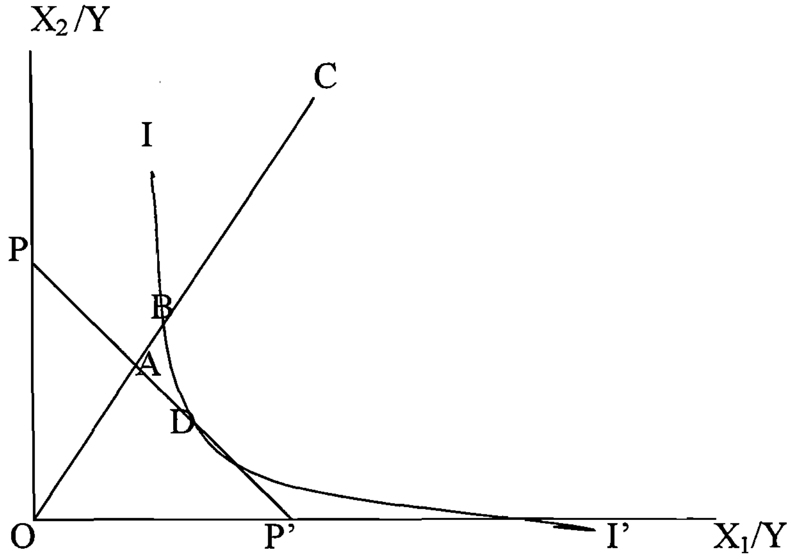
STATISTICAL (PARAMETRICAL) FRONTIER

Frontier performance comparisons flow directly from the definition of the production function itself. Johnston (1960: 4) defines the production function as: *"the relationship describing the maximum flow of output per unit of time achievable for any given rates of flow of input services per unit of time."* In other words, production is a process of transformation in which inputs are combined to generate outputs. An equivalent interpretation holds for the cost function. Duality theory establishes the relationship between production and costs. For a given factor prices, the cost function must be interpreted as a frontier function because it is impossible to achieve costs lower than the minimum input requirements implied by the production frontier. The word "frontier" is applied in either case because the function sets a bound on the range of possible observations. Production may occur below the frontier but not above it; similarly, costs will be above the cost frontier but not below it. The amount by which an organization lies below its production frontier or the amount by which it lies above its cost frontier, is regarded as measures of relative efficiency (Ganley & Cubbin, 1992).

Because the first empirical treatment of a production function as a frontier is in Farell (1957), frontier efficiency is treated the same as the Farell efficiency measurement. Farell sees overall efficiency as being composed of two multiplicative components: OE = TE.AE, where TE is technical and AE is allocative efficiency. Each of these can be defined in terms of a production frontier as the ratio of potential and actual performance. Consider for example an organization consuming two inputs x_1 and x_2, producing an output y. It has a production function $y = f(x_1, x_2)$ which Farell assumed, exhibits constant returns to scale. Accordingly, the production function may be written

$1 = f(x_1/y, x_2/y)$ so that the frontier technology can be characterized by the unit isoquant II' in figure 3.2 below. In this figure, an organisation produces unit output at point C. Its technical efficiency (TE) is the ratio of potential to actual input consumption. This is the radial measure OB/OC, which is in this case, less than unity.

Figurel Farell Eficiency Measurement



Source: (Ganley & Cubbin, 1992)

TE =OB/OC AE=OA/OB TE x AE=OA/OC

Potential or "maximal" performance is defined along the frontier. As observed performance worsens, the distance of an observation from the frontier increases so that the technical efficiency (TE) ratio falls towards zero. Likewise, as performance improves, the efficiency ratio rises in value to unity. In general, then: $0 \leq TE \leq 1$.

Farrel (1957) also included an allocative efficiency (AE) ratio within his frontier framework. Like technical efficiency, the allocative component is a radial measure which lies between zero and unity. At a point such as B in Figure 3.2. AE=OA/OB where PP' is the isocost line defined by the ratio of factor prices. Allocative efficiency is significant in that it emphasises that boundary production per se is not sufficient to minimise costs. Full efficiency (i.e. OE=1) requires simultaneous technical and allocative efficiency, that is AE=TE=1, which is obtained at D (Figure 1).

To fix ideas, it is useful to show how the technical efficiency ratio in Figure.1 can be defined directly in terms of the production and cost function. If inefficiency is possible, the production function may be written as an inequality: $Y_r \leq f(X_i ; \beta)$, where Y_r is the observed output at establishment r, and X_i is a vector of inputs and β a vector of parameters which describes the transformation process. f(.) is the production function and has the interpretation of a frontier, or Y_{max}. At inefficient operations, potential output (Y_{max}) will exceed observed performance (Y_r). Hence, technical inefficiency implies that (Y_r–Y_{max}) is negative. The difference between observed and potential performance can be treated as a residual in the production, function which is equivalent to the technical efficiency ratio. If these residuals are denoted ξ_j, then in terms of the production function, the technical ratio can be written as $\xi_j = Y_r/f(X_i;\beta)$.

To preserve the frontier interpretation of f(.),the ξ_j, are always non-positive. This ensures that the observed output cannot exceed the potential, and that the distribution of residuals is one-sided. The addition of the efficiency residuals balances the production function above.

$Y_r = f(X_i; \beta)-\xi_j \qquad \xi_j \le 0 \qquad$ for all i

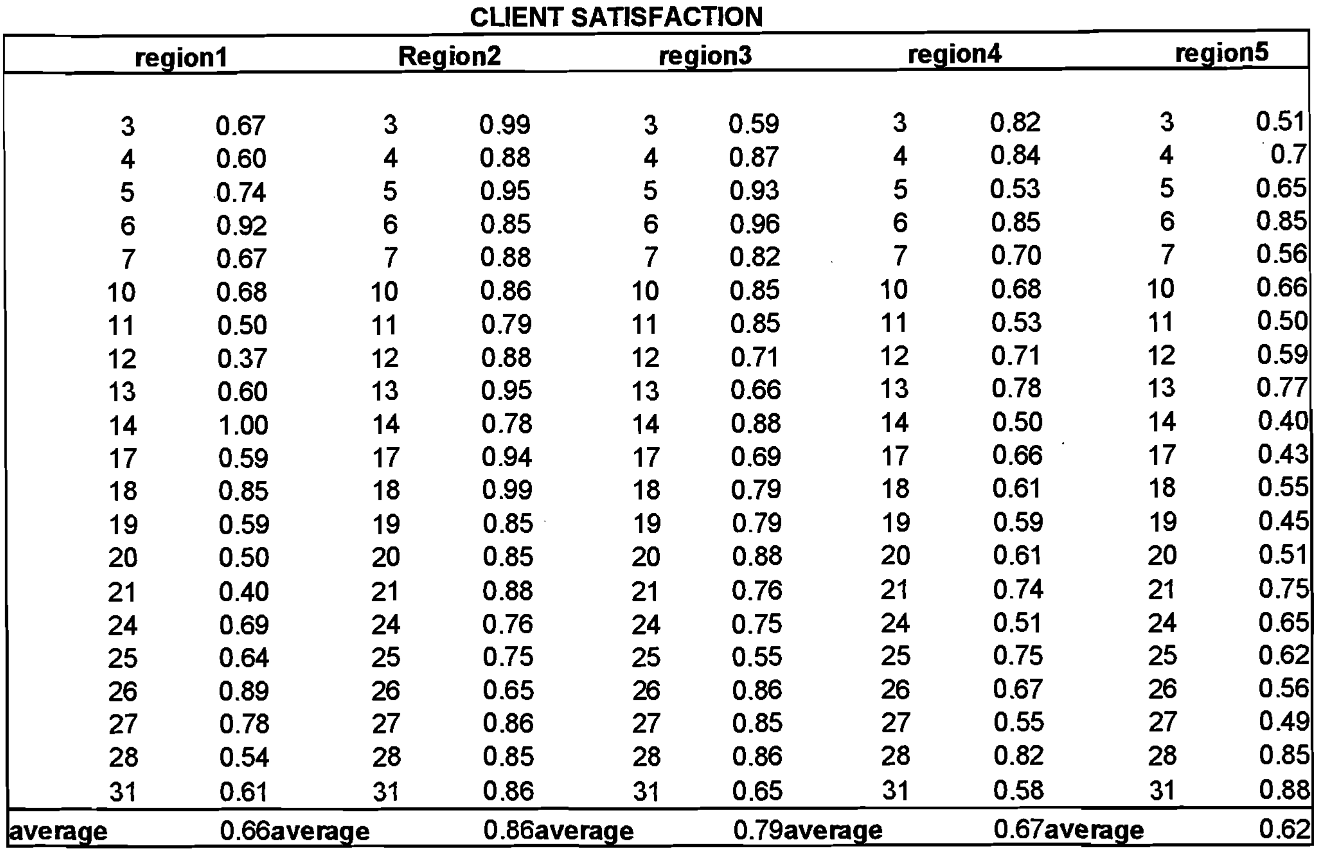
The technical efficiency ratios, ξ_j can be estimated econometrically (Richmond, 1974). This requires the choice of a specific one-sided distribution for technical efficiency, negative half-normal or negative exponential distribution being the most common assumptions (Aigner, Lovell & Schmidt, 2000).

Unlike conventional ordinary least squares residuals, the efficiency distribution must be one-sided in order to ensure that actual output does not exceed potential, i.e $Y_r > Y_{max}$ is not possible. Hence all the efficiency residuals in the production function are non-positive and truncated at zero so that deviations are possible only below the production frontier (Ganley & Cubbin, 1992).

APPENDIX B

TABLES

Table 1 Client Satisfaction



CLIENT SATISFACTION

region1		Region2		region3		region4		region5	
3	0.67	3	0.99	3	0.59	3	0.82	3	0.51
4	0.60	4	0.88	4	0.87	4	0.84	4	0.7
5	0.74	5	0.95	5	0.93	5	0.53	5	0.65
6	0.92	6	0.85	6	0.96	6	0.85	6	0.85
7	0.67	7	0.88	7	0.82	7	0.70	7	0.56
10	0.68	10	0.86	10	0.85	10	0.68	10	0.66
11	0.50	11	0.79	11	0.85	11	0.53	11	0.50
12	0.37	12	0.88	12	0.71	12	0.71	12	0.59
13	0.60	13	0.95	13	0.66	13	0.78	13	0.77
14	1.00	14	0.78	14	0.88	14	0.50	14	0.40
17	0.59	17	0.94	17	0.69	17	0.66	17	0.43
18	0.85	18	0.99	18	0.79	18	0.61	18	0.55
19	0.59	19	0.85	19	0.79	19	0.59	19	0.45
20	0.50	20	0.85	20	0.88	20	0.61	20	0.51
21	0.40	21	0.88	21	0.76	21	0.74	21	0.75
24	0.69	24	0.76	24	0.75	24	0.51	24	0.65
25	0.64	25	0.75	25	0.55	25	0.75	25	0.62
26	0.89	26	0.65	26	0.86	26	0.67	26	0.56
27	0.78	27	0.86	27	0.85	27	0.55	27	0.49
28	0.54	28	0.85	28	0.86	28	0.82	28	0.85
31	0.61	31	0.86	31	0.65	31	0.58	31	0.88
average	0.66	average	0.86	average	0.79	average	0.67	average	0.62

Table 2 Inputs and Outputs

Region	Employees	Number of Resolved Events	Client Satisfaction
	Input	Output	Output
1	17	201	0.66
2	16	160	0.86
3	15	157	0.79
4	17	200	0.67
5	13	123	0.62

Table 3 Arrival Rate Of Calls

Days of the month	3	4	5	6	7	10	11	12	13	14	17	18	19	20	21	24	25	26	27	28	31
Calls per minute	8	5	3	5	3	11	7	7	5	5	9	9	6	6	6	9	10	8	4	12	12

APPENDIX C

Frontier Analysis

Abstract

This study researches and implements the best practices that lead to best performance. A customer quality defined standard has to be created by benchmarking the Information Technology Service Regions. The Data Envelopment Analysis (DEA) methodology is used as a benchmarking tool to locate a frontier which is then used to evaluate the efficiency of each of the organizational units responsible for observed output and input quantities. The inefficient units can learn from the best practice frontier situated along the frontier line.

1. Problem Statement

The company in which the research takes place, is a telecommunication company with its headquarters situated in Pretoria, South Africa. Within this company there are divisions that deal solely with Information Technology services. These divisions are called Information Technology (IT) Service delivery regions and are situated countrywide: in Pretoria, Johannesburg, Durban, Bloemfontein/Kimberly (Bloem/Kby), and Cape Town. These regions are responsible for the execution of their operational responsibilities. In these service delivery regions, the main common function is to repair both computer hardware and software. The research concentrates on the problems arising in these divisions.

Firstly, customers from all regions in South Africa report events by phoning a centralized service desk in Pretoria. An event is anything that an end-user finds 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 of a computer on the network, are all requests. The reinstallation of software and fixing computer hardware, are faults. End-users are people using computer services. An Information System is an arrangement of people, data, processes, information presentation, and information technology that interact to support and prove day-to-day operations in a business, as well as support the problem-solving and decision-making needs of management and users (Whitten *et.al.*, 2006).

Secondly, these logged events are routed to their respective regions by the service desk to be attended to. The success in producing as large as possible an output (number of resolved events and satisfied clients) from a given set of inputs employees (labour) is not achieved. Customers complain that their logged events are not resolved within the specified Service Level Agreement (SLA). The SLA is an agreement on performance System Metrics (Application Availability in Production, Average Request Resolution and Average Fault Resolution). The agreement is, that a logged fault should be resolved within two days. A logged request should be resolved in four days. Customers wait a longer time before they can work on their computers again.

2. The Research Goal

The purpose of this study, is to research and implement best practices that lead to best performance. Also to find out more about the Data Envelopment Analysis methodology as a benchmarking tool and the evaluation of efficiencies of regions in order to satisfy customer defined quality (SLA).

This study deals with benchmarking in management. Benchmarking or best practices are ways of carrying out a function that makes a significant difference in the quality of output. They bring down costs, increase customer satisfaction, or reduce a process. Glen Peters (1994:9) defines benchmarking as follows. *"Benchmarking is about improving competitive position, and using best practices to stimulate radical innovation rather than just seeking minor, incremental improvements on historic performance".*

The major task is to measure the performances of service delivery regions and to evaluate their efficiencies. By making comparisons between the regions, an expectation exists that best practice regions can be identified and used as benchmarks for improving the efficiency, quality and effectiveness of other regions.

3. Data Envelopment Analysis Methodology

A benchmarking methodology, Data Envelopment Analysis (DEA), is used to research and implement best practices. DEA, occasionally called frontier analysis, is a new technique developed in operations research and management science over the last two decades for measuring performance in the public and private sectors.

It can also be described as a non-parametric estimation method which involves the application of mathematical programming to observed data to locate a frontier which can then be used to evaluate the efficiency of each of the organizational units responsible for observed output and input quantities.

The DEA methodology as discussed by Charnes, Cooper, Lewin and Seiford (1994), is used to evaluate the relative efficiency of a set of Decision-making Units (DMU's). This term "DMU's" was coined by Charnes *et al.* to describe homogeneous units, each utilizing a common set of inputs to produce a common set of outputs. Examples of homogeneous DMU's are a collection of similar firms, departments, group of schools, hospitals and bank branches. A bank branch and a supermarket are not homogeneous units. In this study's perspective, DEA is used to evaluate the efficiency of IT service delivery regions which are denoted as region 1 to region 5, (DMU1 to DMU5) which also are homogeneous with some decision autonomy. Each region consumes one input and produces two outputs. A DEA model is developed that uses these factors (input and outputs) to compute the efficiency degree of a particular region when this region is compared with all the other regions. The regions that are considered efficient belong to the frontier and, therefore, they can be used as performance benchmarks to study the regions that are operating inefficiently (Charnes *et.al.*, 1994).

4. Data

(1) From an event management system database, and for each month in a year, the date the event was reported and the date the event was resolved, were recorded. These dates were used to determine the ratio of the resolved events to the total number of logged events per month. For example, registering the first of January five times under logged events, means five events are logged, and registering the first of January three times under resolved events, means three events are resolved. Determining the ratio will then be 3/5. (2) The average number of events resolved per month in a year was determined. The data used here are for twelve months, starting from February 2004 to January 2005.

Table 1 Variables

Variables	Type	Description
Number of resolved events.	Output	Number of faults and requests resolved.
Client Satisfaction.	Output	Ratio of number of resolved events to total number of logged events.
Employees.	Input	Number of employees.

The number of resolved events and client satisfaction are regarded as outputs. Employees are regarded as input. Client Satisfaction was determined as the ratio of number of resolved events to the number of logged events. The average inputs and outputs per month for a year for the five regions are as given in the following table.

Table 2 Inputs and Outputs

Region	Number of Employees	Number of Resolved Events	Client Satisfaction
	Input	Output	Output
1	17	201	0.66
2	16	160	0.86
3	15	157	0.79
4	17	200	0.67
5	13	123	0.62

Source:(Event Management System, 2004)

5. Efficiency with DEA

In this study's perspective, DEA is used to evaluate the efficiency of Information technology service delivery regions which are denoted as regions 1 to 5, (DMU's 1 to 5 according to Charnes *et.al.* (1994), which also are homogeneous with some decision autonomy. Each region uses one input to produce two outputs. A DEA model is formulated that uses these factors to compute the efficiency degree of a particular region when this region is compared with all the other regions. The regions that are considered efficient relative to the other regions, belong to the frontier and, therefore, can be used as performance benchmarks to study the regions that are operating inefficiently. Regions that are inefficient, do not belong to the frontier.

According to Marcoulides (1998: 121) the need to compare performance with some known number or quantity in order to understand how well the organization performs brought about the increasing popularity of what is known as performance ratios.

A commonly used traditional ratio method in DEA, is input-oriented and measures productivity or efficiency as a ratio of output to input (Beasley, 2007).

6. Single Input, Output Measure

6.1 Number of Employees and Resolved Events

Table 3 Single Input, Output (Resolved Events)

Region	Number of Employees	Number of Resolved Events
Pretoria	17	201
Bloem/Kby	16	160
Durban	15	157
Johannesburg	17	200
Cape Town	13	123

In the above data, for instance, Pretoria had 201 resolved events while 17 staff members were employed. In Durban there were 157 resolved events and 15 staff members were employed, etc. These regions are compared and their performance measured by using the data. Some output measure is divided by some input measure to get a ratio. For example, 201 is divided by 17 to get 11.80. The following data applies.

Table 4 Single Input, Output (Resolved Events) Ratios

Region	Events Resolved per Employee
Pretoria	201/17 =11.80
Bloem/Kby	160/16 =10.00
Durban	157/15 =10.47
Johannesburg	200/17 =11.76
Cape Town	123/13 = 9.46

According to the above data, Pretoria has the highest ratio of resolved events per staff member, whereas Cape Town has the lowest. Since Pretoria has the highest ratio of 11.80, other regions are compared to it and their relative efficiencies calculated with respect to it. The ratio for any region is divided by the ratio for Pretoria (11.80), multiplied by 100 to convert to a percentage, resulting in the following.

Table 5 Single Input, Output (Resolved Events) Percentages

Region	Relative Efficiency
Pretoria	100%
Bloem/Kby	85%
Durban	89%
Johannesburg	99.6%
Cape Town	80%

The other regions do not compare with Pretoria, they are performing lower and are relatively less efficient at using their staff (input) to produce output (number of resolved events). Pretoria can be used to set a target for other regions. This is an input target, since it deals with input measure.

6.2 Number Of Employees And Client Satisfaction

This time, output measure is client satisfaction and the input measure remains the number of employees since this ratio method is input oriented. The target is the number of employees. This is the variable that is going to be adjusted to effect efficiency. By increasing or decreasing the number of employees, the optimal output will be reached. Once more, client satisfaction is determined as the ratio of number of resolved events to the total number of logged events per day. From Table 2, data again are as follows.

Table 6 Single Input, Output (Client Satisfaction)

Region	Employees	Client Satisfaction
Pretoria	17	0.66
Bloem/Kby	16	0.86
Durban	15	0.79
Johannesburg	17	0.67
Cape Town	13	0.62

In the data, for instance, Pretoria had a ratio of 0.66 client satisfaction and 17 staff members were employed. In Durban a ratio of 0.79 events was resolved, while 15 staff members were employed, etc. These regions are compared and their performance measured by using this data. Some output measure is divided by some input measure to get a ratio. Hence the following data results.

Table 7 Single Input, Output (Client Satisfaction) Ratios

Region	Client Satisfaction per Employee
Pretoria	0.66/17 =0.039
Bloem/Kby	0.86/16 =0.054
Durban	0.79/15 =0.053
Johannesburg	0.67/17 =0.039
Cape Town	0.62/13 =0.048

According to the above data, Bloem/Kby had the highest ratio of Client Satisfaction per employee, whereas Pretoria had the lowest. Since Bloem/Kby had the highest ratio of 0.054, all other regions are compared to it and their relative efficiency calculated with respect to Bloem/Kby. The ratio for any region is divided by the ratio of Bloem/Kby (0.054) and multiplied by 100 to convert to a percentage, as following.

Table 8 Single Input, Output (Client Satisfaction) Percentages

Region	**Relative Efficiency**
Pretoria	72%
Bloem/Kby	100%
Durban	96%
Johannesburg	72%
Cape Town	87%

The other regions do not compare with Bloem/Kby, they are performing less, they are relatively less efficient at using their staff (input) to produce output (satisfaction). Bloem/Kby could set target for other regions. This is still an input target, since it deals with input measure.

7. Extended Resources

Considering a single input measure, number of employees, and two output measures, resolved events and client satisfaction could be resolved at the same time. Again the five regions are compared. From Table 2, the data are again as following.

Table 9 Extended Resources

Region	Number of Employees	Resolved Events	Client Satisfaction
Pretoria	17	201	66%
Bloem/Kby	16	160	86%
Durban	15	157	79%
Johannesburg	17	200	67%
Cape Town	13	123	62%

Durban, for example, with 15 employees, had an average of 157 events resolved per month and satisfied its clients up to 79 percent. Ratios are still used to compare these regions. Dividing each output measure with the single input (number of employees) gives the following.

Table 10 Efficiency Ratios

Region	Events Resolved	Client Satisfaction
	Per Employee	per Employee
Pretoria	11.80	3.9
Bloem/Kby	10.00	5.4
Durban	10.47	5.3
Johannesburg	11.76	3.9
Cape Town	9.46	4.8

Pretoria had the highest ratio of resolved events per employee whereas Bloem/Kby had the highest ratio of client satisfaction per employee. Figure 1 in the next section presents the above data.

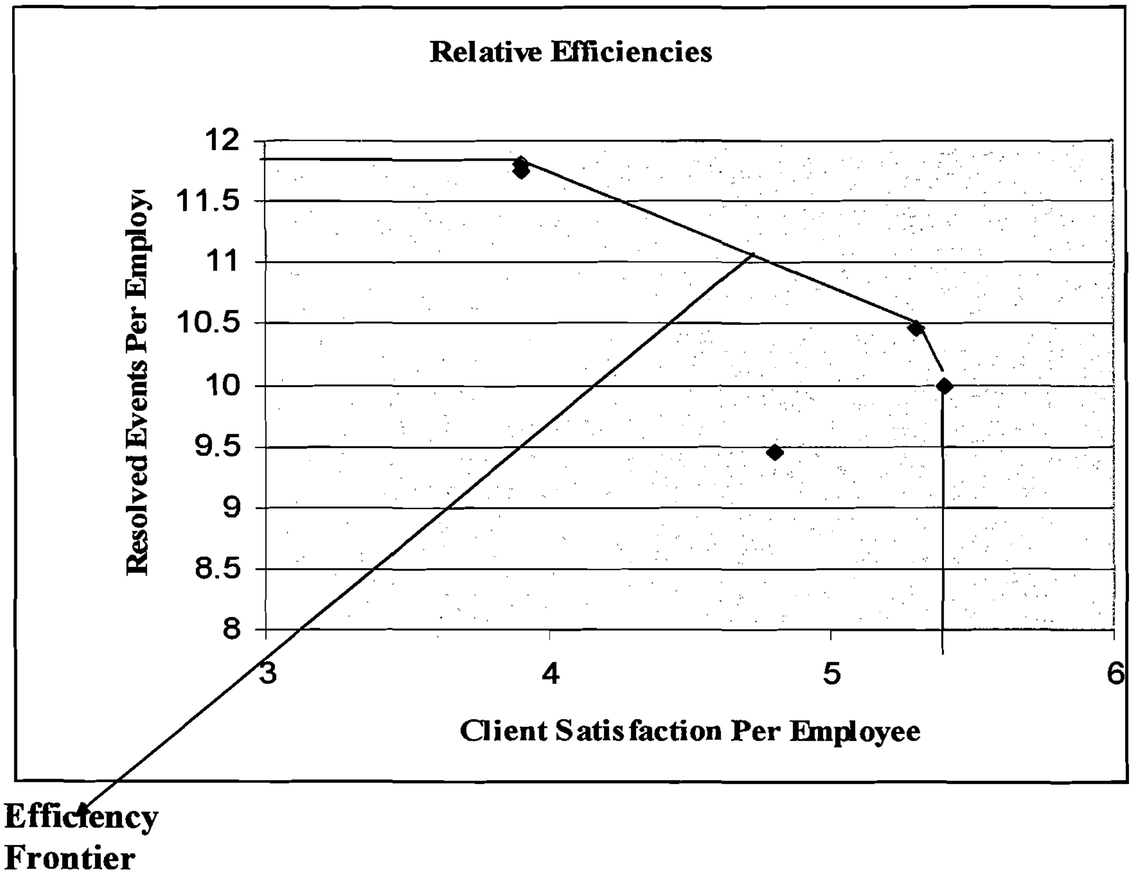
The problem with comparing ratios, is that a different ratio could give a different picture and it becomes difficult to combine these ratios into one ratio, where one could draw one's own judgement. For example, if we consider Durban and Cape Town, Durban gives (10.47/9.46) = 1.11 times as efficient as Cape Town on resolved events and also (5.3/4.8) = 1.11 times as efficient as Cape Town on client satisfaction. It is not easy to combine these ratios into a judgement. This can be more clearly seen if there are more inputs and more outputs (Beasley, 2007).

8 Graphical Analysis

Another way of evaluating the efficiency, at least for problems involving two outputs and a single input, is by graphical analysis, as shown in Figure 1 below. In this Figure, all the regions are on the frontier line, except Cape Town. Johannesburg and Pretoria almost make the same data point, since their readings are almost equal (the thickest point in the graph).

Again in Figure 1 below, a horizontal line is drawn from the y-axis to Pretoria, from Pretoria to Johannesburg, from Johannesburg to Durban, from Durban to Bloem/Kby. A vertical line is drawn from Bloem/Kby to the x-axis. This line is called the efficiency frontier. The efficient frontier, derived from the examples of best practice contained in the data considered, represents the performance that the regions, in this case Cape Town, not on the efficient frontier could try to achieve. Hence data envelopment because the efficient frontier envelopes (encloses) all data available. All the regions on the frontier are 100% efficient. Therefore, all the regions are efficient except Cape Town (Beasley, 2007).

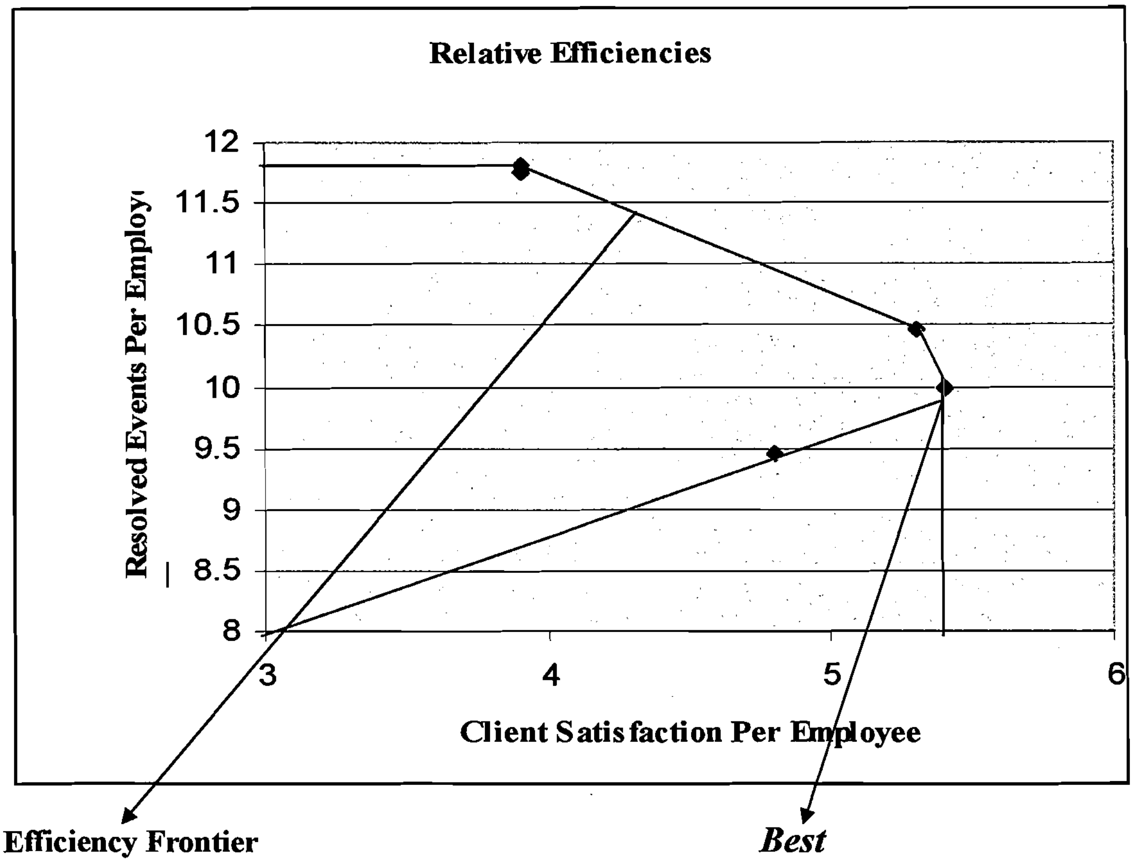
Figure 1 Relative Efficiencies



9 Quantifying Efficiency Score For cape Town

Cape Town is less efficient, but by how much? It has 13 staff members, 123 resolved events, 9.46 resolved events per employee, 62% client satisfaction, and 4.8 client satisfaction per employee. The ratio of resolved events/client satisfaction = (123/62) = 1.98; that is, there are 1.98 resolved events for every percentage of client satisfaction. This ratio is the same as resolved events per employee to client satisfaction per employee.

Figure 2 Relative Efficiencies



Considering Figure 2, Cape Town is not on the efficiency frontier. A line drawn from the origin through Cape Town to the efficiency frontier line has a slope of 1.98. If Cape Town were to retain this ratio, but to vary the number of staff it employs, its performance, would lie on the line from the origin through its current position as shown above. It might be reasonable to say that the best possible performance that Cape Town could be expected to achieve is labelled Best in the graph. This is the point where the line from the origin through Cape Town meets the efficiency frontier.

According to Beasley (2007), DEA gives only the relative efficiencies; efficiencies relative to the data considered. It does not and cannot give absolute efficiencies.

10. Sensitivity analysis

Sensitivity analysis and post optimality analysis allow the decision maker to determine how the final solution to the problem will change when the input data or the model change. This type of analysis is very important when the input data or model has not been specified properly. A sensitive solution is one in which the results of the solution to the problem will change drastically or by a large amount with small changes in the data or in the model. When the model is not sensitive, the results or solutions to the model will not change significantly with changes in the input data or in the model. Models that are very sensitive require that the input data and the model itself be thoroughly tested to make sure that both are very accurate and consistent with the problem statement.

While ratios are easy to compute, their interpretation is problematic, especially when they provide conflicting answers. While this may be generally true, statistics can be used to understand this. For example, using the number of resolved events, Pretoria is at the top while using client satisfaction, Pretoria is at the bottom. This may look like conflicting results, but in reality they are not. More employees can resolve more events, if there are events to resolve. The illustration below, Figure 3 shows that it is possible to move beyond the efficient frontier created by DEA previously.

Figure 3 Number of Employees Line Fit Plot

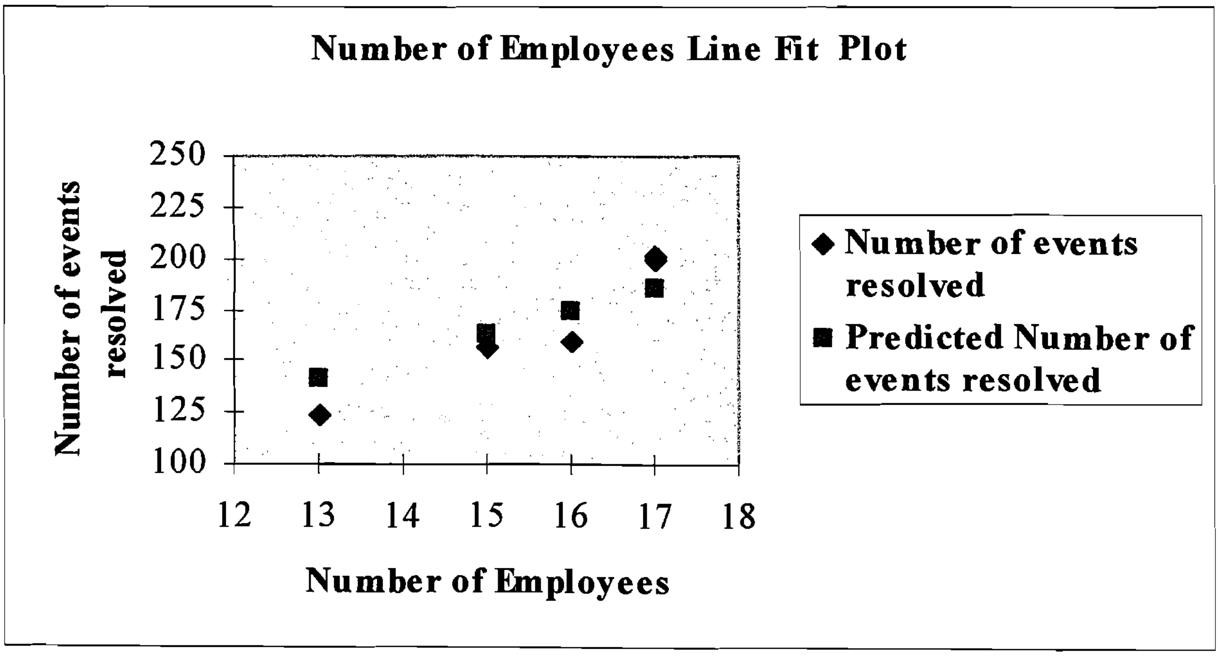


Figure 4 below shows a plot of residuals. This is to assess the model adequacy by checking whether the model assumptions are satisfied. The basic assumption is that the residuals are uncorrelated with zero mean and constant variance. The residuals in Figure 4 look evenly distributed with no pattern, meaning that they are uncorrelated. Figure 5, the normal probability plot shows no serious deviations from the fourty five degree line. We conclude from the plots that the residuals are normally distributed. The regression model is adequate.

Figure 4 Number Of Employees Residual Plot

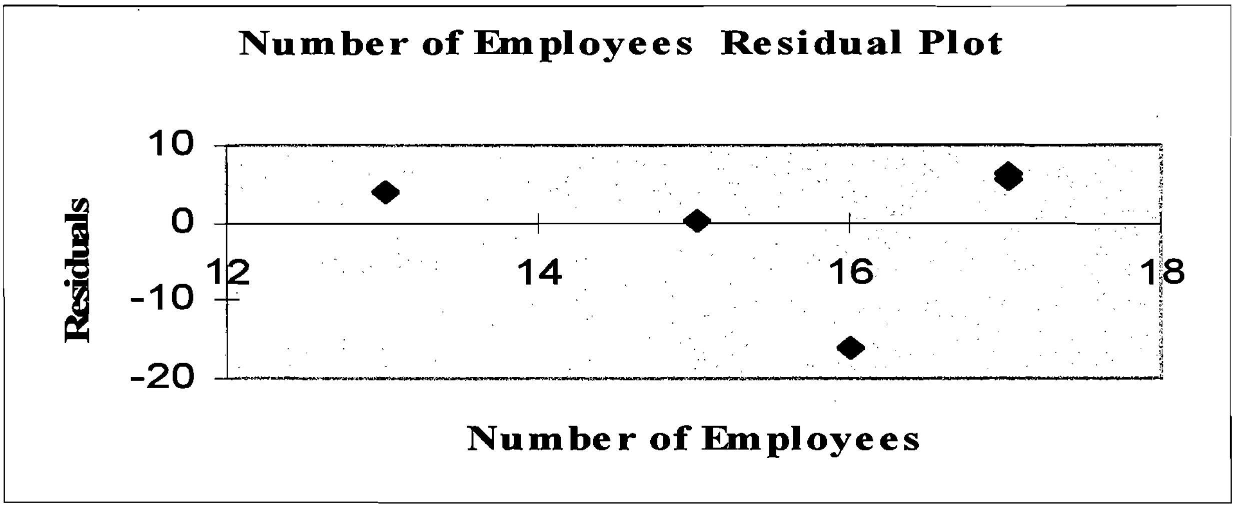


Figure 5 Normal Probability Plot

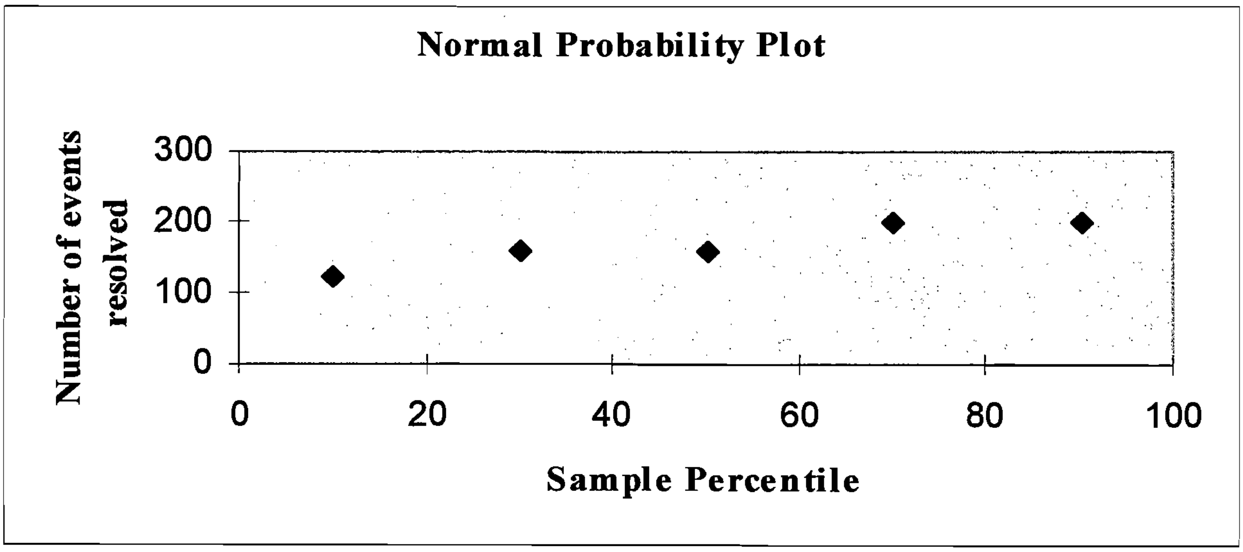


Table 11 Predicted Outputs

	Number of Employees	Predicted Number of resolved events	Predicted Client Satisfaction	Number of logged events
1	17	195	0.64	305
2	16	176	0.94	186
3	15	157	0.79	199
4	17	195	0.65	299
5	13	119	0.60	198
	Regression		Intercept	-126.25
			Number of Employees	18.875
	Number of Employees	Predicted Number of resolved events	Predicted Client Satisfaction	Number of logged events
1	23	308	1.01	305
2	17	195	1.05	186
3	18	214	1.07	199
4	23	308	1.03	299
5	18	214	1.08	198

In Table 11 above, the current number of employees cannot resolve all the logged events. When there are more employees, statistics shows that all logged events can be resolved. This is the number of employees necessary to resolve all the logged events, using the extrapolated extrapolating the regression equation. The assumption is that the extra people can resolve the unresolved events at the same rate as the existing employees, which may not be valid if the unresolved events are more difficult to resolve than the resolved events. This may be a way to move beyond the current efficient frontier and create new quality standards.

11. Conclusion

While ratios are easy to compute, their interpretation is problematic, especially when they provide conflicting answers. While this may be generally true, statistics can be used to understand this. For example, using the number of resolved events, Pretoria is at the top while using client satisfaction, Pretoria is at the bottom. This may look like conflicting results, but in reality they are not. More employees can resolve more events, if there are events to resolve.

REFERENCE

ANDERSON, D., SWEENEY, D. & WILLIAMS, T. 2006. Quantitative methods for business. 8th ed. Cincinnati, Oh.: South-Western College. 822 p.

BANKER, R. 1984. Estimating the most productive scale size using Data Envelopment Analysis. *European journal of operations research,* 17:35-44.

BANKER, R., CHARNES, A. & COOPER, W. 1984. Some models for estimating technical and scale efficiency in Data Envelopment Analysis. *Management science,* 30:1078-1092.

BANKER, R., CHARNES, A., COOPER, W. & SCHINNAR, A. 1981. A bi-extremal principle for frontier estimation and efficiency evaluations. *Management science,* 27:1370-1382.

BEASLEY J.E. 2007. Data envelopment analysis. http://people.brunel.ac.uk/~mastjjb/jeb/or/contents.html Date of access: 01 Aug. 2007.

BOWLIN, W.F. 1995. Measuring performance: an introduction to data envelopment analysis (DEA). Ceder Falls, Ia: University of Northern Iowa, Department of Accounting. 27 p. (Technical report.)

CARLSON, S. 1999. The pure theory of production. London: King.

CHARNES, A., COOPER, W.W., LEWIN, A.Y. & SEIFORD, L.M., *eds.* 1994. Data envelopment analysis: theory, methodology and applications. Dordrecth, Holland: Kluwer.

COOPER, W.W., THOMPSON, R.G. & THRALL, R.M. 1996. Extensions and new developments in DEA. *Annals of operations research,* 66:3-45.

DORIAN, P. 1999. Data preparation for data mining. New York: Academic Press. 540 p.

MARCOULIDES, G.A. 1998. Modern methods for business research. London:

PETERS, G. 1994. Benchmarking customer service. London: Pitman.

WHITTEN, J., BENTLEY, L. & DITTMAN, K. 2006. System analysis and design methods. 6th ed. New York: McGraw-Hill. 724 p.

System Design

Abstract

This study researches the queuing methodology. A system has to be designed that achieves the desired performance level. A quality standard has to be created by determining the number of service channels (employees) that can to handle the workload of answering all the calls arriving at the service. Abandoned or unanswered calls have to be reduced.

1. Problem Statement

The company in which the research takes place, is a telecommunication company with its headquarters situated in Pretoria, South Africa. It has a central service desk situated at its headquarters in Pretoria. Customers from all regions in South Africa report events by phoning a centralised service desk in Pretoria. An event is anything that an end-user finds 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 the network, are all requests. The reinstallation of software and fixing computer hardware are faults. End-users are people using computer services.

An Information System is an arrangement of people, data, processes, information presentation, and information technology that interacts to support and prove day to day operations in a business as well as support the problem solving and decision making needs of management and users. (Whitten *et.al.*, 2006).

The time required to service the customer varies considerably from call to call because every call has its own problems. Arriving calls seek service from one of several service channels (employees). A service channel is a server 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 percentage of abandoned calls is high.

2. The Research Goal

The objective of this study is to research the queuing methodology. The methodology that can design a system that achieves the desired performance. The optimal number of service channels that can be used at the service desk has to be determined.

3. The Research Methodology

A Queuing model is employed in order to find out whether it solves the problem.

4. The Queuing Theory

The Queuing Theory had its beginning in the research work of a Danish engineer named A.K. Erlang. The three components of the queuing process are the arrival rate, the queue and the service rate. The arrival rate refers to the rate at which the calls arrive at the service desk. For instance, a call or two calls arriving every minute describes the arrival rate. According to Taha (2007) a queue is created in the following manner. When a customer arrives in the system, he or she joins a waiting line. An employee chooses a customer from the waiting line to begin service. Upon the completion of a service, the process of choosing a new waiting customer is repeated. The service rate refers to how long it takes the server at the servicing channel to service a customer.

If the average time a customer waits in the queue is denoted by W, and the average customer arrival in the queue by λ, 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 (Render *et.al.*, 2006).

The following assumptions are used in every queuing model: (1) The queuing environment has either a finite or infinite calling population, and a multiple or single channel facility is used. (2) The arrival time is unpredictable and described by a Poisson distribution, or is predictable. (3) The service times (processing rate at the servicing facility) are exponential or unpredictable or the exact amount of processing time is known. (4) The queue lengths are infinite or finite. (5) All units wait in the single queue. (6)

Service is on a first-come first-service basis. (7) All arriving events enter the queue. (Hall, 1993).

5. Data Collection

The company has a centralized service desk were all calls are reported through a telephone line for all the regions. Events are logged and routed to their respective regions. Here, (1) the number of calls, as they entered the telephone system per minute (arrival rate), were recorded. This was actually easy since 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. Channels refer to employees answering the calls.

6. Understanding Data

Description of Summaries and Visualisation according to Two Crows Corporation (2007). *"Before you can build good models, you must understand your data. Start by gathering a variety of numerical summaries (including descriptive statistics such as averages, standard deviations and so forth) and looking at the distribution of the data.*

Graphing and visualization tools are vital aids in data preparation and their importance to effective data analysis cannot be overemphasized. Data visualization most often provides the "Aha!", leading to new insights and success. Some of the common and very useful graphical displays of data are histograms or box plots that display distributions of values"

The task of hypothesizing a distribution family from observed data, is somewhat unstructured. Three categories are used to aid in making a decision as to what distribution the observed data resembles. These are the chi-square test, and the histogram of frequency distribution of calls. Table 1 shows the arrival rate of calls for a month, and this was determined as explained in section 5.

Table 1 Arrival Rate of Calls

Days of the month	3	4	5	6	7	10	11	12	13	14	17	18	19	20	21	24	25	26	27	28	31
Calls per minute	8	5	3	5	3	11	7	7	5	5	9	9	6	6	6	9	10	8	4	12	12

Source: (Service Desk, 2004)

7. 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. 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 be at least 5, so that the chi-square test can be used. The expected frequencies are computed on the basis of a hypothesis $H_{o.}$ H_o: 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 χ^2 is smaller than some critical value (such as $\chi^2_{0.95}$) 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 χ^2 is the chi-square. The O_j's are the observed frequencies and E_j's are the expected frequencies. The data in Table 1 above were used to determine the observed frequencies in Table 2 below. Similar arrival rates of calls as they occur in a month were grouped together and counted. The total of their counts is their frequencies, 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, but probabilities cannot be more than 1. It should actually be expected values. Expected values are theoretical results expected according to the rules of probability. For example, for x = 3 or less and sample size = 21 the expected frequency is calculated as follows.

$$\begin{aligned}P(0 \leq x \leq 3) &= P(x=0) + P(x=1) + P(x=2) + P(x=3)\\ &= \sum e^{-\lambda}\lambda^{x} / x!\\ &= (e^{-7}7^{0}/0! + e^{-7}7^{1}/1! + e^{-7}7^{2}/2! + e^{-7}7^{3}/3!)\,21\\ &= 1\end{aligned}$$

And for x=4 the expected frequency is calculated as follows.

$$\begin{aligned}P(x=4) &= (e^{-7}7^{4}/4!)21\\ &= 1.8\end{aligned}$$

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 observed and expected frequencies for different arrival rates are depicted in Figure 1 (a-g) below to show how different arrival rates of calls approximate a poisson distribution.

Figure 1(a) Chi-square for λ = 6

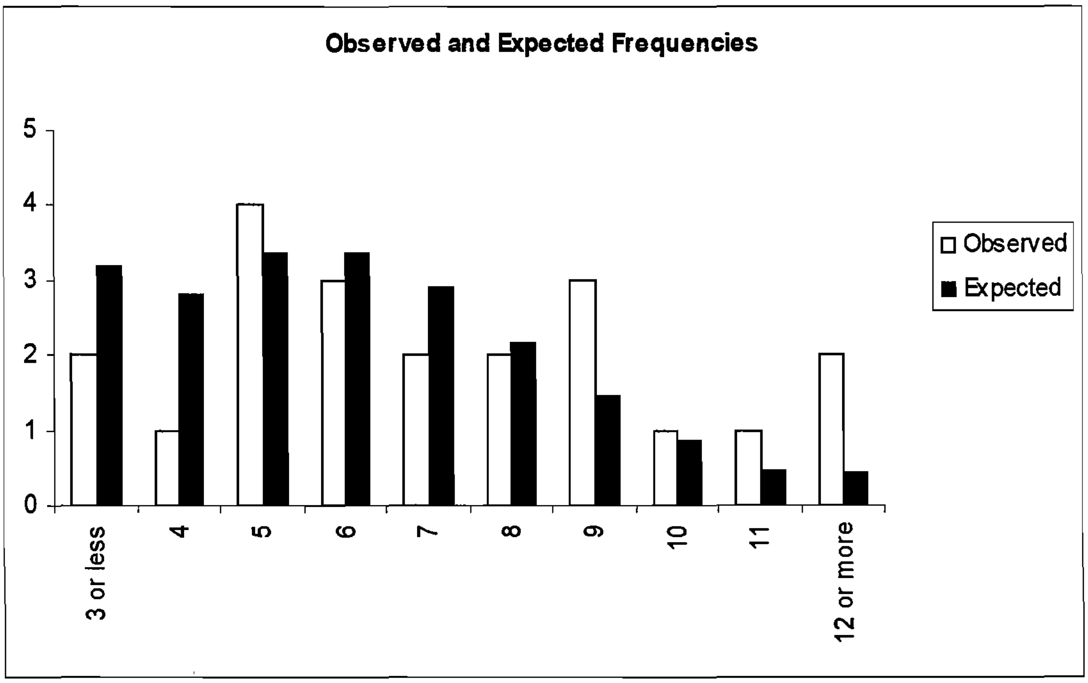


Figure 1(b) Chi-square for λ = 6.5

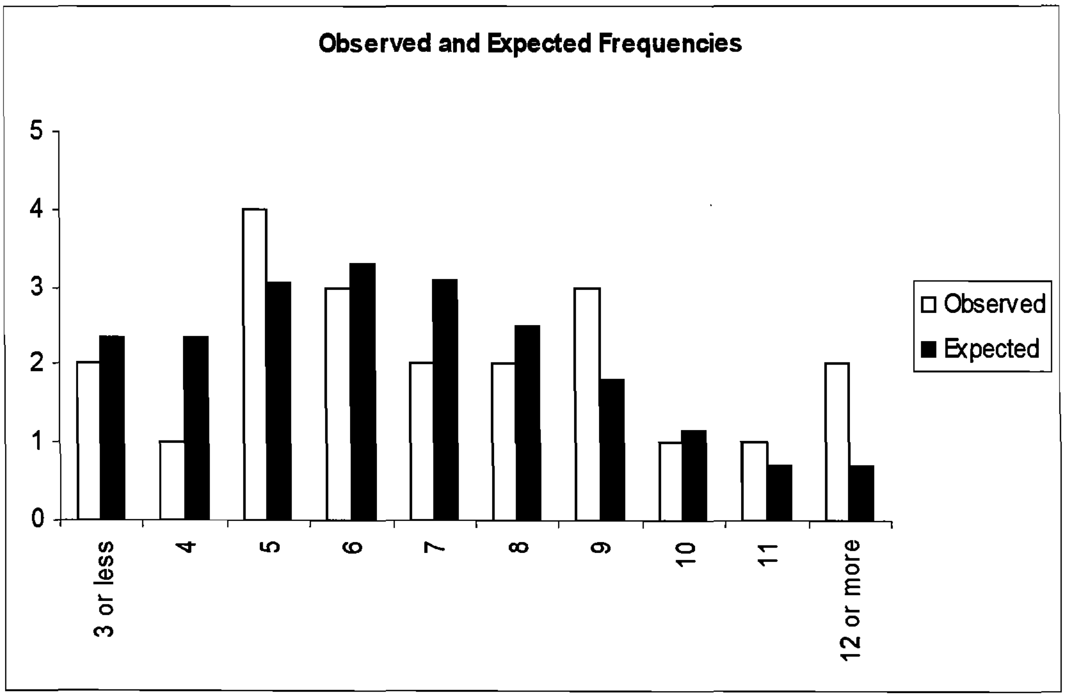


Figure 1(c) Chi-square for λ = 7

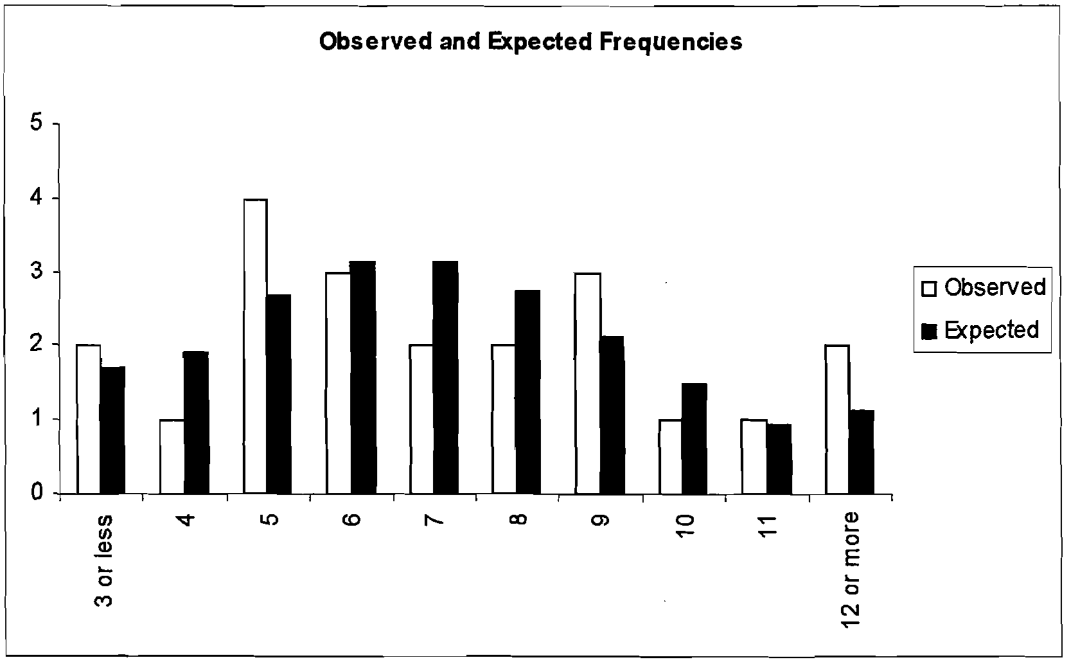


Figure 1(d) Chi-square for λ = 7.14

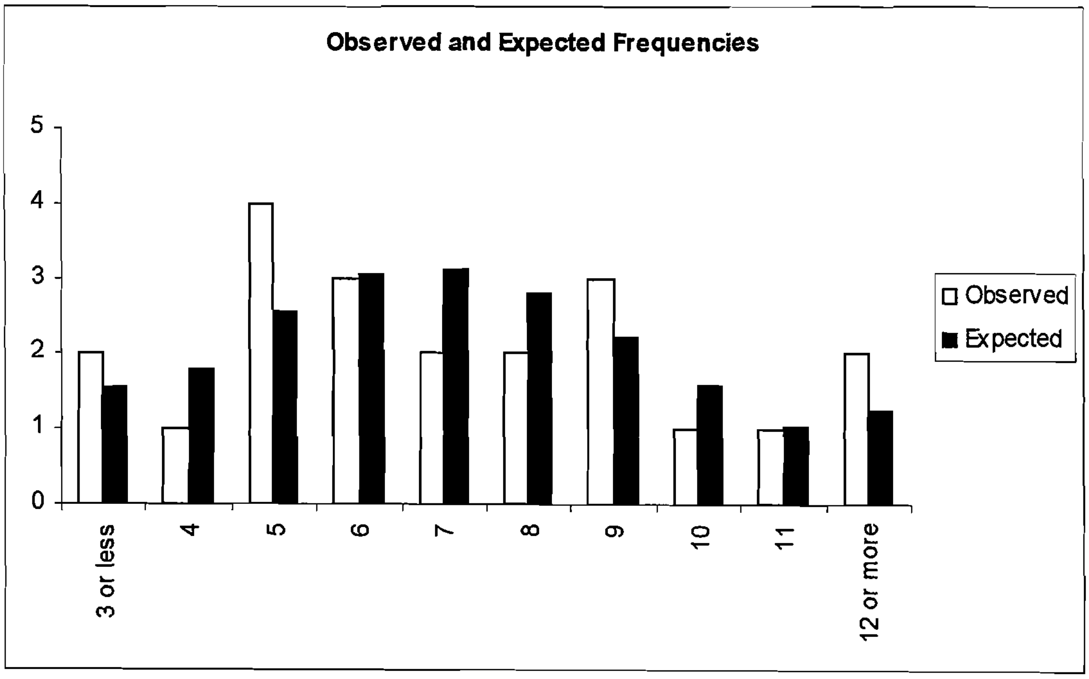


Figure 1(e) Chi-square for λ = 7.43

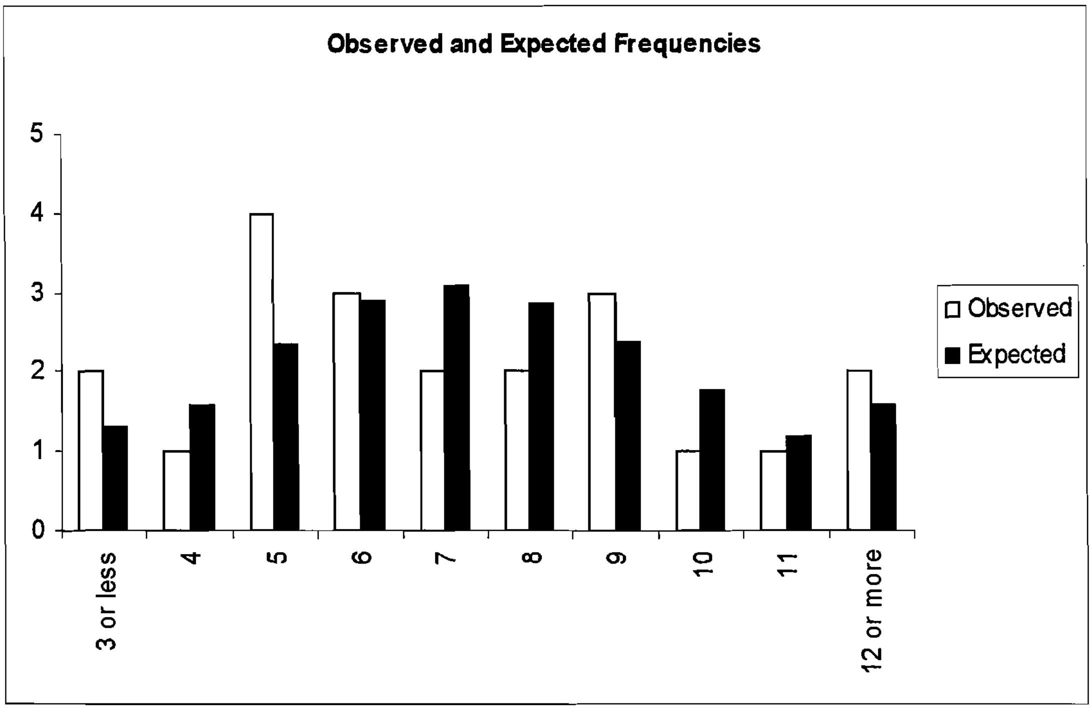


Figure 1(f) Chi-square for λ = 8

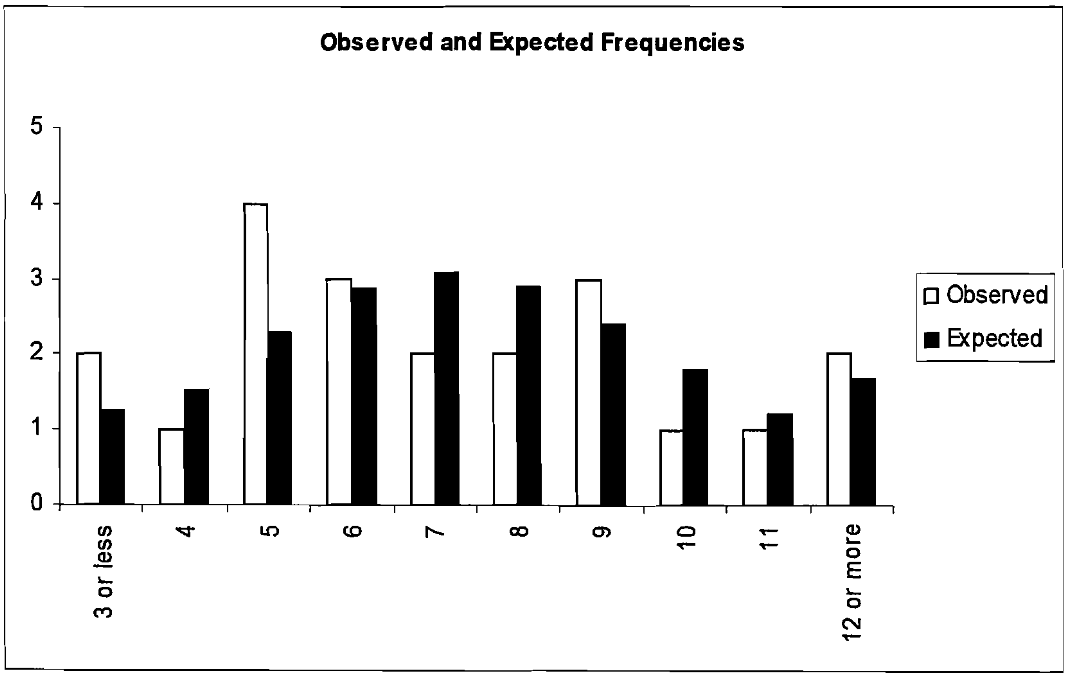
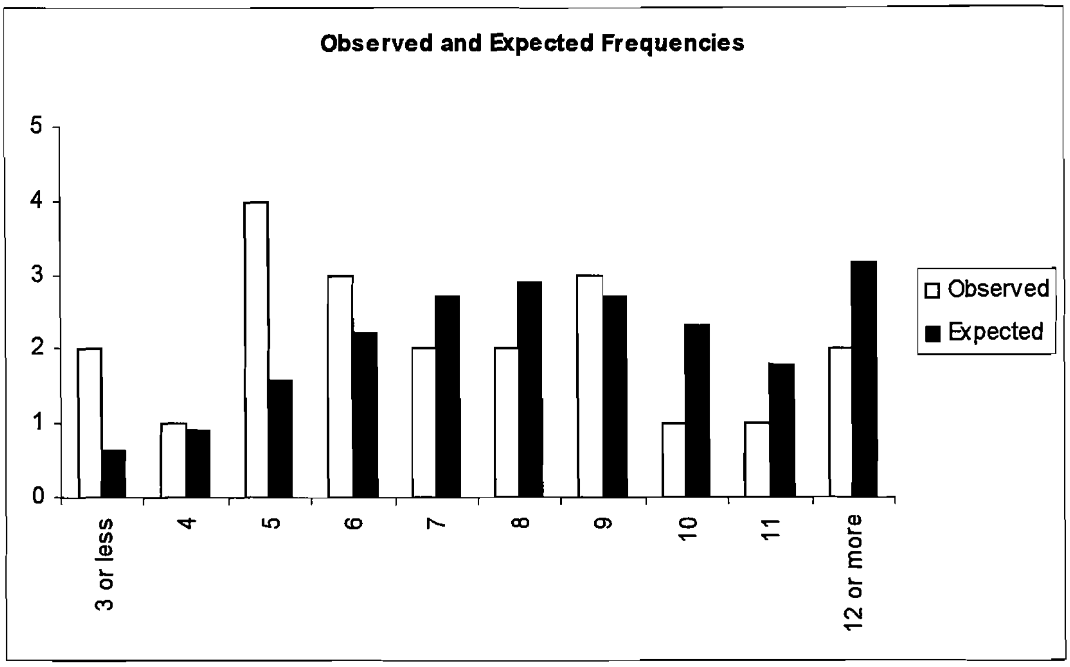


Figure 1(g) Chi-square for λ = 8.5



The test tested the hypothesis that the Poisson distribution approximates the sample data at 5% significance level with one degree of freedom. One degree of freedom because, since there are three classes, one degree is missed because λ is estimated and one is lost because of the frequency sum. The chi-square critical value is 3.841.

8. 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 λ =7, where λ is the rounded mean arrival rate. The chi-squares for different lambdas are shown in table 4 below.

Table 3 Chi-square Computation

		Expected	Observed	O-E	(O-E)^2	(O-E)^2/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	**Chi-square**	**0.81928**

Table 4 Sensitivity analysis

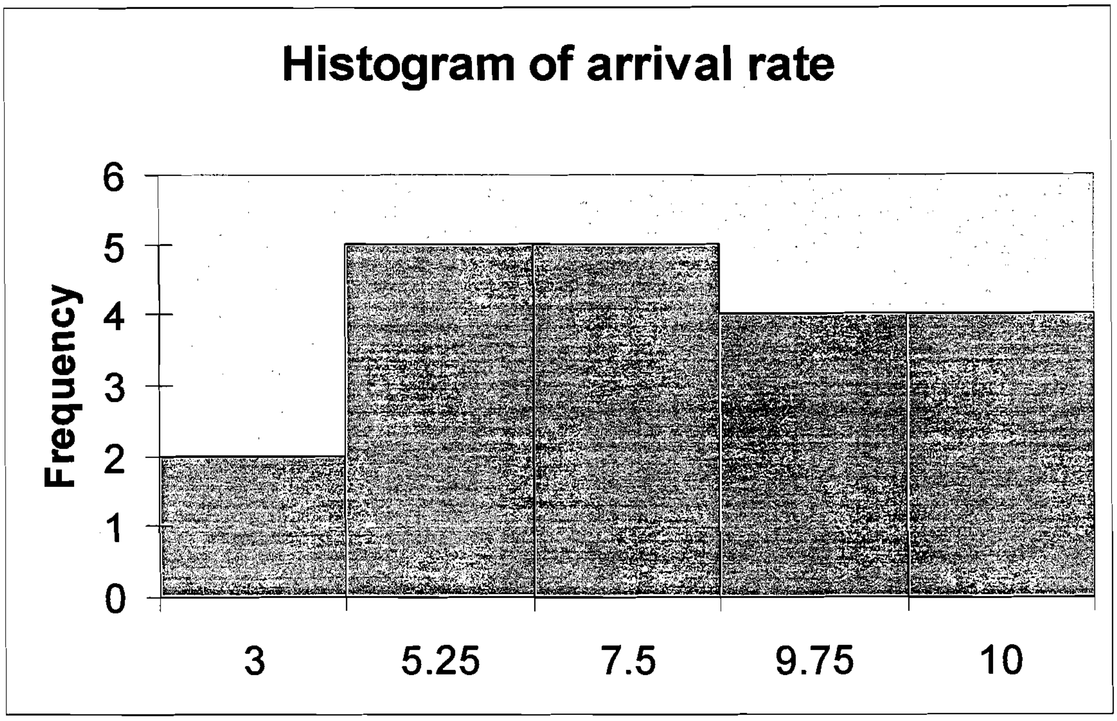
Lambda (λ)	χ^2	Sensitivity
6	5.3	The null hypothesis is rejected.
6.5	2.0	The null hypothesis is not rejected.
7	0.8	The null hypothesis is not rejected.
7.14 (the expected value estimate)	0.8	The null hypothesis is not rejected.
7.43 the variance estimate	1.1	The null hypothesis is not rejected.
8	2.7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

The conclusion is that the data do not present sufficient evidence to contradict the hypothesis that F possesses a Poisson distribution.

9. Frequency Distribution

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. This data from Table 1 above, are used to determine the optimal number of channels (employees) 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 2 below.

Figure 2 Arrival Rate of calls Histogram



Some distributions are characterized at least partially by functions of their true parameters. Given the above picture, 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. 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. The *standard deviation* is the square root of the variance.

10 Queuing Theory

10.1 Introduction

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 (employees) that can handle the workload, thereby optimising (minimising) the costs.

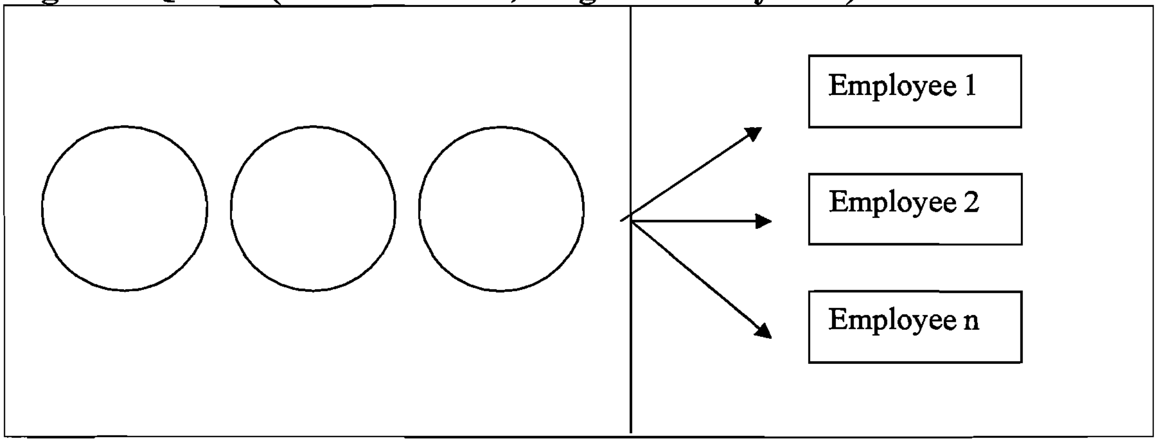
10.2 The Queuing Model

This section discusses the behaviour of the study's queuing model. This had to be examined (as examined in sections 7 and 9, 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. In this study, there is an unlimited or infinite logging of events.

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. (Hall, 1993). The following diagram depicts the queuing model involved.

Figure 3 Queue (Multi Channel, Single Phase System)



Source: (Render *et. al.*, 2006)

10.3 Operating Method

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 λ. The arrivals follow a Poisson probability distribution (as examined in sections 7 and 9). There is an average rate of service μ 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.

10.4 Computations

The optimal number of employees (channels) is determined by computing a steady state probabilities that j of the k channels will be busy. Formula 1 below is used to calculate these percentages (probabilities). The following equation applies.

$$P_j = \frac{(\frac{\lambda}{\mu})^j / j!}{\sum_{i=0}^{k} (\frac{\lambda}{\mu})^i / i!} \qquad (1)$$

Source: (Render *et.al.*, 2006)

Where λ = the mean arrival rate

μ = the mean service rate for each channel

k = the number of channels

P_j = the probability that j of the k channels are busy for j = 1, 2,...,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

$$L = \lambda / \mu \, (1 - P_k) \qquad (2)$$

Source: (Charnes *et.al.*, 1994)

Whether the arrivals are indeed Poisson distributed, is determined. Section 2.5 confirms this. 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 λ = 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 μ = 0.5.

Since there are 17 channels, they cannot handle the workload as there is a high % 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 λ = 7 and μ = 0.5 we calculate the percentage of abandoned calls.

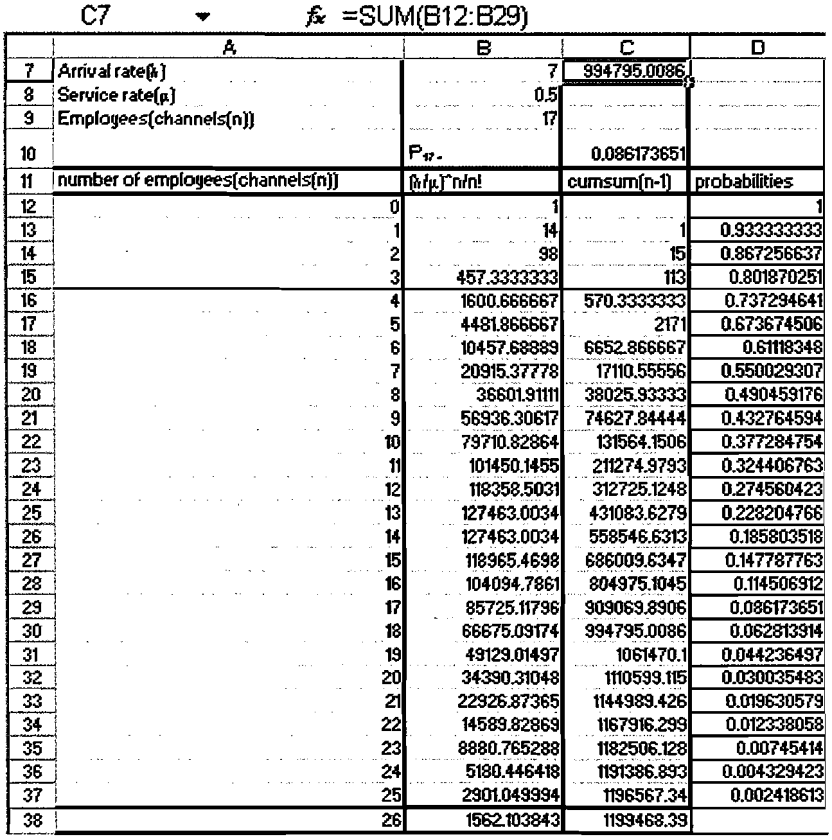
$$P_{17} = P_{\text{abandoned}} = \frac{(7/0.5)^{17}/17!}{[(7/0.5)^0/0! + (7/0.5)^1/1! +, \dots, + (7/0.5)^{17}/17!]}$$

$$= 85725.11796 / 994795.009$$

$$= 0.08617365$$

With only 8.61% of calls blocked with 17 channels, 91.39% of calls is 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 (λ=7) for a different number of channels, in table 2.8 below In this table, it is shown that when the number of employees (channels) increase, the probability (percentage) of abandoned calls decrease. For example, with 22 employees (channels), 1.23% of calls is abandoned, and with 25 employees (channels), 0.24% is abandoned.

Table 5 Abandoned Calls % For Different Number Of Channels



C7 ▾ fx =SUM(B12:B29)

	A	B	C	D
7	Arrival rate(λ)	7	994795.0086	
8	Service rate(μ)	0.5		
9	Employees(channels(n))	17		
10		P₁₇-	0.086173651	
11	number of employees(channels(n))	(λ/μ)^n/n!	cumsum(n-1)	probabilities
12	0	1		1
13	1	14	1	0.933333333
14	2	98	15	0.867256637
15	3	457.3333333	113	0.801870251
16	4	1600.666667	570.3333333	0.737294641
17	5	4481.866667	2171	0.673674506
18	6	10457.68889	6652.866667	0.61118348
19	7	20915.37778	17110.55556	0.550029307
20	8	36601.91111	38025.93333	0.490459176
21	9	56936.30617	74627.84444	0.432764594
22	10	79710.82864	131564.1506	0.377284754
23	11	101450.1455	211274.9793	0.324406763
24	12	118358.5031	312725.1248	0.274560423
25	13	127463.0034	431083.6279	0.228204766
26	14	127463.0034	558546.6313	0.185803518
27	15	118965.4698	686009.6347	0.147787763
28	16	104094.7861	804975.1045	0.114506912
29	17	85725.11796	909069.8906	0.086173651
30	18	66675.09174	994795.0086	0.062813914
31	19	49129.01497	1061470.1	0.044236497
32	20	34390.31048	1110599.115	0.030035483
33	21	22926.87365	1144989.426	0.019630579
34	22	14589.82869	1167916.299	0.012338058
35	23	8880.765288	1182506.128	0.00745414
36	24	5180.446418	1191386.893	0.004329423
37	25	2901.049994	1196567.34	0.002418613
38	26	1562.103843	1199468.39	

As mentioned in section 10.4, Formula 1 was used in a spreadsheet to model the abandoned calls' percentages (probabilities). FACT is factorial and ^ is the index meaning raised to the power of the value in the cell. C10=(B30/C7); B14=($B$7/$B$8)^A14/FACT(A14) and copied to B15 through to B38; C14=SUM(B13;$B$13) and copied to C15 through to B38; D15=(B14/C15) and copied to D16 through to D38; C7= SUM(B13:B30).

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

Table 6 Abandoned Rate Of Calls Within 95% Confidence Interval

Arrival rate per minute Lambda (λ)	Abandoned rate% with 17 employees	Abandoned rate% with 25 employees
6.5	6.17	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

10.5 Conclusion

The conclusions drawn from all the tests done, all the calculations made, recommend 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 costy. 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.

REFERENCE

ANDERSON, D., SWEENEY, D. & WILLIAMS, T. 2006. Quantitative methods for

DORIAN, P. 1999. Data preparation for data mining. New York: Academic Press. 540 p.

HALL, O.P. 1993. Computer models for operations management. Redwood City, Calif.: Addison-Wesley. 197 p.

MARCOULIDES, G.A. 1998. Modern methods for business research. London: Erlbaum.

PETERS, G. 1994. Benchmarking customer service. London: Pitman.

RENDER, B., STAIR, R.M. & HANNA, M.E. 2006. Quantitative analysis for management. 8th ed. Upper Saddle River, N.J.: Pearson/ Prentice Hall. 726 p.

TAHA, H.A. 2007. Operations research: an introduction. 8th ed. New York: Macmillan. 813 p.

TWO CROWS CORPORATION. 2007. Introduction to data mining and knowledge discovery. 2nd ed. Falls Road, Potomac. http://www.twocrows.com Date of access: 03 June 2007.

WHITTEN, J., BENTLEY, L. & DITTMAN, K. 2006. System analysis and design methods. 6th ed. New York: McGraw-Hill. 724 p.

WINSTON, W. 2004. Operations research: applications and algorithms. 4th ed. Pacific Grove, Calif.: Thomson-Brooks/Cole. 1318 p.