Achieving the Efficiency Frontier in IT Service Delivery

Philip D. Pretorius, North-West University (Vaal Triangle Campus), Vanderbijpark, South Africa, Philip.Pretorius@nwu.ac.za
Victoria G. Madisa, North-West University (Vaal Triangle Campus), Vanderbijlpark, South Africa, vmadisa@sars.gov.za
Jan H. Kroeze, North-West University (Vaal Triangle Campus), Vanderbijlpark, South Africa, jan.kroeze@gmail.com

Abstract

This paper researches and implements the best practices that lead to best performance in Information Technology service delivery. A customer quality defined standard is 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.

Keywords

Efficiency frontier, service delivery, sensitivity analysis

Introduction

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/Kimberley (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).

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 optimised. 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.

The Research Goal

The purpose of this paper is to research and implement best practices that lead to best performance in IT service delivery; 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 (CDQ).

This paper deals with benchmarking in management. Benchmarking and the identification of 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 optimise a process. Glen Peters (1994) defines benchmarking as follow: *"Benchmarking is about improving competitive position, and using best practices to stimulate radical innovation rather than just seeking minor, incremental improvements on historic performance"*.

According to Bowlin (1995), 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.

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 relatively new technique developed in operations research and management science over the last two decades for measuring performance in the public and private sectors (Banker, 1984).

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 (Cooper et al., 1996).

The DEA methodology as discussed by Charnes et al. (1994) is used to evaluate the relative efficiency of a set of Decision-making Units (DMUs). The term was coined by Charnes et al. (1994) to describe homogeneous units, each utilizing a common set of inputs to produce a common set of outputs. Examples of homogeneous DMUs are a collection of similar firms, departments, groups of schools, hospitals and bank branches. A bank branch and a supermarket grouped together are not homogeneous units. 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, can be used as performance benchmarks to study the regions that are operating inefficiently (Charnes et al., 1994).

Data

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 that 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. The average number of events resolved per month in a year was determined (Dorian, 1999). The data used are for twelve months, starting from February 2004 to January 2005. Table 1 explains which variables are used for analysis, their type and a short description of each.

Table 1: Variables

Variables	Type	Description
Number of resolved events	Output	Number of faults and requests resolved
Success rate	Output	Ratio of number of resolved events to total number of logged events (throughput)
Employees	Input	Number of employees

According to Carlson (1999), the number of resolved events and success rate are regarded as outputs. Employees are regarded as input. Success rate was determined as the ratio of number of resolved events to the number of logged events. This is an attempt to combine quantitative and qualitative measures. The average inputs and outputs per month for a year for the five regions are given in Table 2.

Table 2: Inputs and Outputs (Event Management System, 2004)

Region	Number of Employees	Number of Resolved Events	Success Rate
	Input	Output	Output
Pretoria	17	201	0.66
Bloem/Kby	16	160	0.86
Durban	15	157	0.79
Johannesburg	17	200	0.67
Cape Town	13	123	0.62

Efficiency with DEA

In this paper's perspective, DEA is used to evaluate the efficiency of information technology service delivery regions which are denoted as regions 1 to 5 (DMUs 1 to 5 according to Charnes et al., 1994), which also are homogeneous with some decision autonomy (Banker et al., 1981). 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 (Banker et al., 1984).

According to Marcoulides (1998) 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, n.d.).

Customer realionship measurement is also used to evaluate the efficiency of information technology service delivery regions. Measuring client satisfaction is very crucial to customer service. Face-to-face dealings are encouraged in order to be able to detect the happiness or unhappiness of the customer with regard to the service received.

Single Input, Output Measure

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 (Table 3), for instance, Pretoria had 201 resolved events while 17 staff members were employed. In Durban there were 157 resolved events while 15 staff members were employed, etc. These regions are compared and their performance measured by using the data. The output measure is divided by the input measure to get a ratio. For example, 201 is divided by 17 to get 11.80. The following data applies (Table 4):

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 had 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 it to a percentage, resulting in the following (Table 5):

Table 5: Single Input, Output (Resolved Events) Percentages

Region	Relative Efficiency
Pretoria	100%
Bloem/Kby	85%
Durban	89%
Johannesburg	99.60%
Cape Town	80%

The other regions do not compare well with Pretoria, since their performance is weaker 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.

Number of Employees and Success Rate

This time, the output measure is success rate 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 affect efficiency. By increasing or decreasing the number of employees, the optimal output will be reached. Once more, success rate is determined as the ratio of the number of resolved events to the total number of logged events per day. In the data (Table 2), for instance, Pretoria had a ratio of 0.66 success rate while 17 staff members were employed. In Durban there was a ratio of 0.79 success rate while 15 staff members were employed, etc. These regions are compared and their performance measured by using this data. The output measure is divided by the input measure to get a ratio. Hence the following data results (Table 6):

Table 6: Single Input, Output (Success Rate) Ratios

Region	Success Rate 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 success rate 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 it to a percentage, as follow (Table 7):

Table 7: Single Input, Output (Success Rate) Percentages

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

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

Extended Resources

The goal here is to consolidate the single input measure, number of employees, with two output measures, resolved events and success rate in a single tool. Again the five regions are compared.

From Table 2, Durban with 15 employees, for example, 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 8):

Table 8: Efficiency Ratios

Region	Events Resolved	Success Rate
Pretoria	11.8	3.9
Bloem/Kby	10	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 success rate 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 1.11 (10.47/9.46) times as efficient as Cape Town on resolved events and also 1.11 (5.3/4.8) times as efficient as Cape Town on success rate. 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, n.d.).

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 efficiency frontier, derived from the examples of best practice contained in the data considered, represents the performance that the regions that are not on the efficiency frontier, in this case Cape Town, could try to achieve. Hence data envelopment is experienced because the efficiency 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, n.d.).

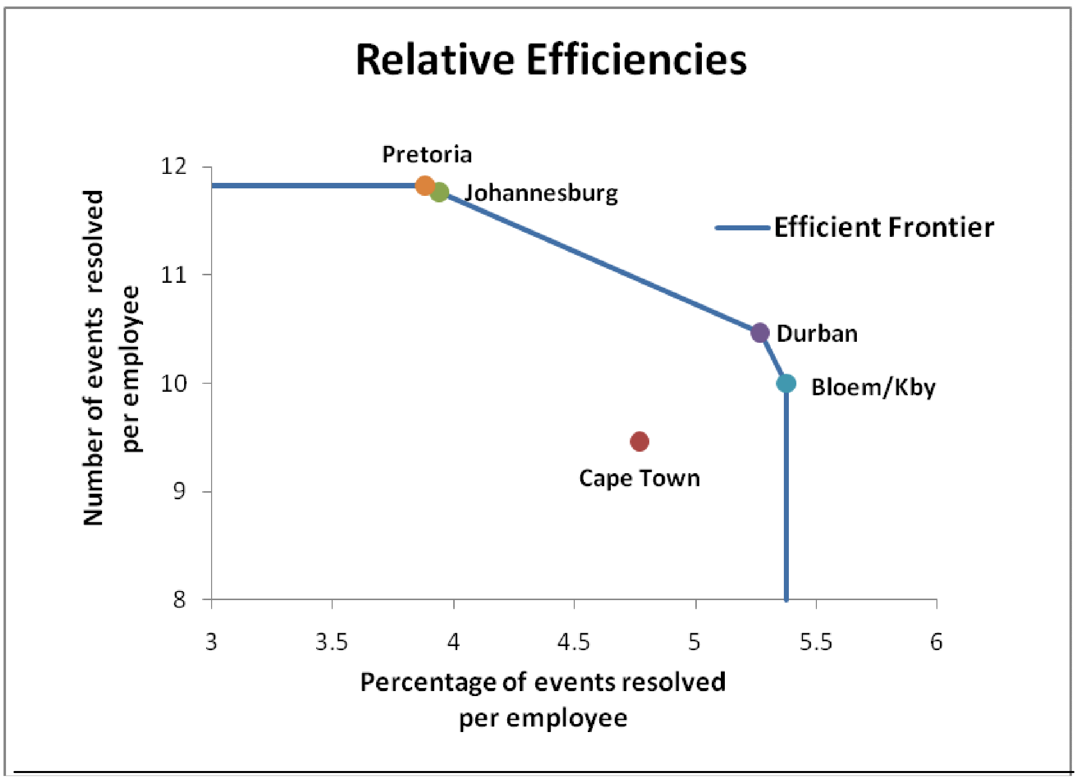
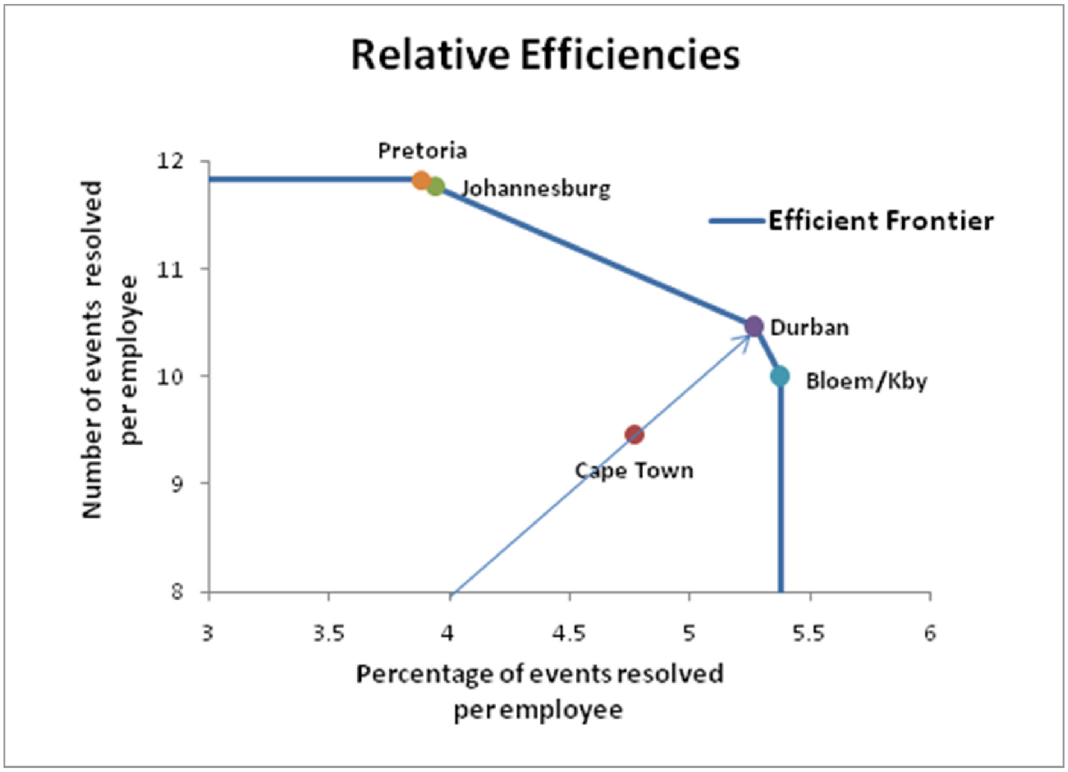


Figure 1. Relative Efficiencies Highlighted by the Efficiency Frontier

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% success rate, and 4.8 success rate per employee. The ratio of resolved events/success rate = (123/62) = 1.98; that is, there are 1.98 resolved events for every percentage of success rate. This ratio is the same as resolved events per employee to success rate per employee.

Considering Figure 2 below, 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.



Best

Figure 2. Relative Efficiencies

According to Beasley (n.d.), DEA gives only the relative efficiencies; efficiencies relative to the data considered. It does not and cannot give absolute efficiencies. The data provided does not specify the weights of logged events. For instance, the weight of logged events is not the same. Some events are more difficult to resolve than the others. Some events take longer to resolve than the others. They are directly related to the success rate and the number of working employees respectively. These kinds of events may reduce the number of resolved events per day since all employees are working for the logged events. There is no reinforcement.

Human resources management has to take informed decisions as to whether to employ additional regular employees or employ temporary employees to assist meeting business demands yet allowing the employer to avoid the cost of hiring a regular employee.

Sensitivity Analysis

Sensitivity and post-optimality analyses guide the analyst on how the solution to the problem will be modified when the input variables or the model changes. The analysis is crucial when the input variables or model has not been specified suitably. A solution is sensitive when the outcome to the problem is modified by changes to the input data. The solution or model is not sensitive when there is no significant change due to alteration of input data. It is a crucial requirement that the input data and the model itself be thoroughly tested for accuracy and consistency with the problem statement (Anderson et al., 2006).

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 success rate, Pretoria is almost 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 results look conflicting because the measurement does not use common weights. DEA allows for the selection of the most advantageous weights in calculating their efficiency scores. In this case the weights were not attached to the inputs and outputs (Makuil et al., 2008). The traditional ratio method was used. The illustration below (Figure 3) shows that it is possible to move beyond the efficiency frontier created by DEA previously.

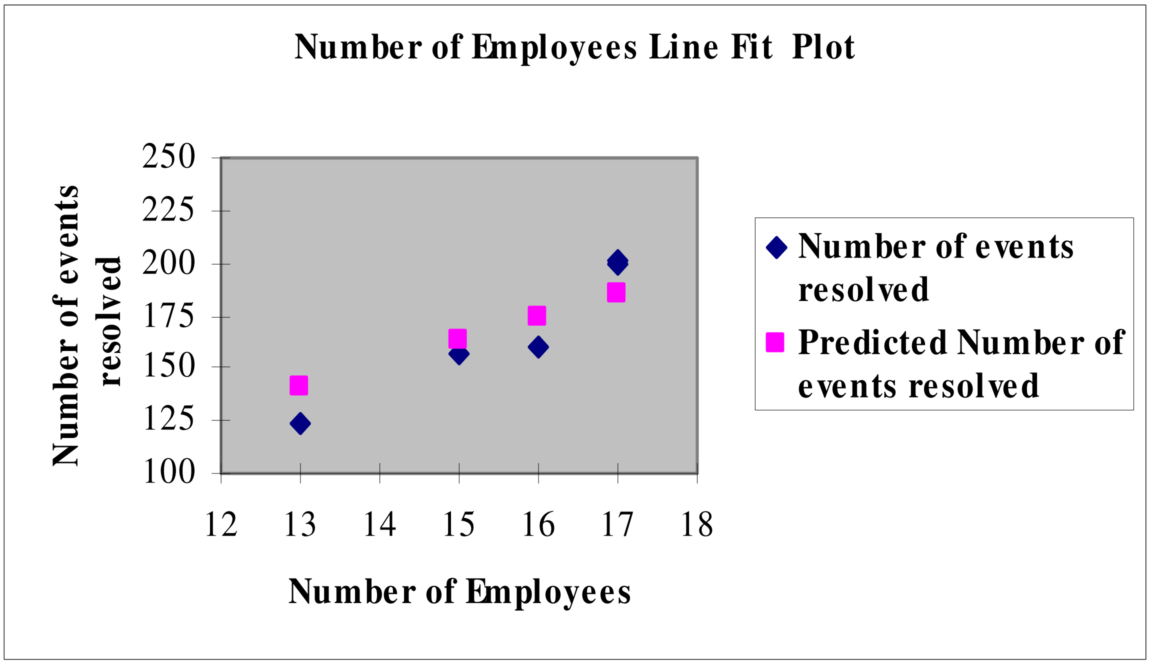


Figure 3. Number of Employees Line Fit Plot

Residuals are defined as the departure or deviation from the overall trend. In other words, it is the difference between the actual and the estimated values. 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, representing 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.

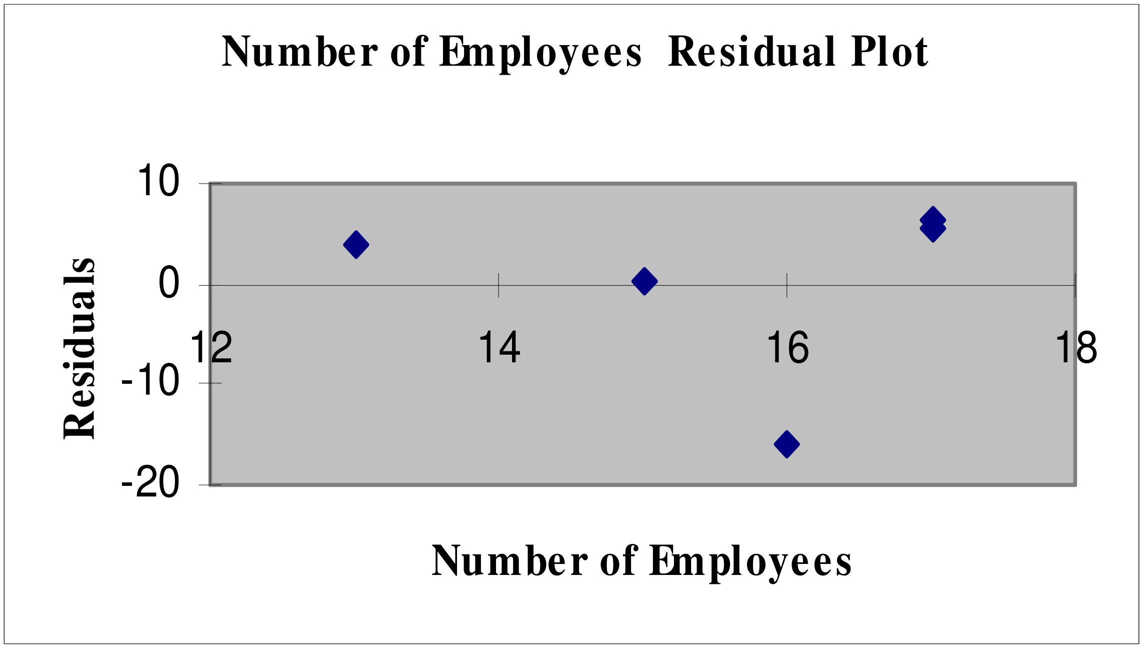


Figure 4. Number of Employees Residual Plot

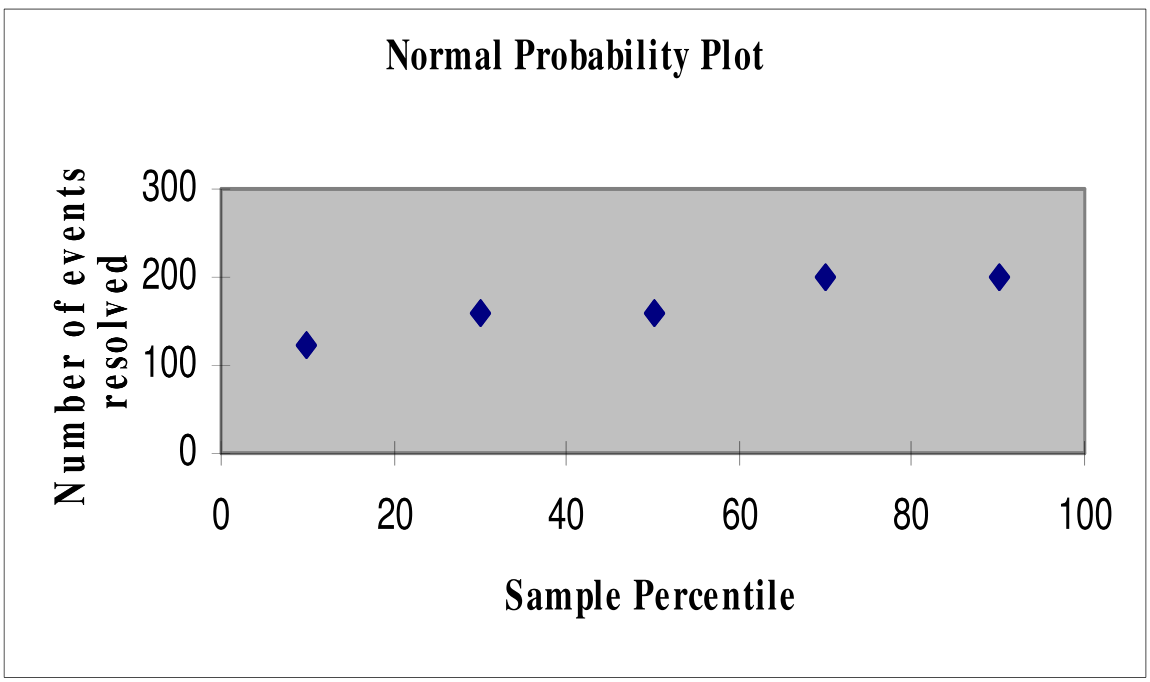


Figure 5. Normal Probability Plot

Table 9: Predicted Outputs

	Number of Employees	**Predicted Number of Resolved Events**	**Predicted Success Rate**	**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 Success Rate**	**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 9 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 regression equation in order to perform extrapolation. 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 efficiency frontier and create new quality standards.

Conclusion

While ratios are easy to compute, their interpretation is problematic, especially when they provide conflicting answers. For example, using the number of resolved events, Pretoria is at the top, while when using success rate, Pretoria is almost 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. Also more skilled, experienced employees can resolve more events.

In this paper, the success in producing as large as possible an output (number of resolved events and success rate) from a given set of inputs employees (labour) was optimised. Customers complained that their logged events were not resolved within the specified Service Level Agreement (SLA). Data Envelopment Analysis determined the relative performance efficiencies of the homogeneous regions. The excellence of the service could be determined by determining the success rate ratio.

It was demonstrated that the ratio analysis indicated Cape Town as being inefficient. The best practice regions could be identified and used as benchmarks for improving the efficiency, quality and effectiveness of the inefficient region, in this case Cape Town.

The understanding and satisfaction of 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 made a contribution to the body of knowledge of Operations Research through the search for quality improvement techniques.

References

Anderson, D., Sweeney, D. and Williams, T. (2006), Quantitative methods for business, 8th ed., South-Western College Publishers, Cincinnati, Ohio.

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. and Cooper, W. (1984) 'Some models for estimating technical and scale efficiency in data envelopment analysis,' *Management Science*, 30(9), 1078-1092.

Banker, R., Charnes, A., Cooper, W. and Schinnar, A. (1981), 'A bi-extremal principle for frontier estimation and efficiency evaluations,' *Management Science,* 27, 1370-1382.

Beasley, J.E. (n.d.) 'Data envelopment analysis'. [Online] [Retrieved August 1, 2007], http://people.brunel.ac.uk/~mastjjb/jeb/or/contents.html

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

Carlson, S. (1999), The pure theory of production, P.S. King & Son, London.

Charnes, A., Cooper, W.W., Lewin, A.Y. and Seiford, L.M. (eds) (1994), Data envelopment analysis: theory, methodology and applications, Kluwer Academic Publishers, Dordrecht, Holland.

Cooper, W.W., Thompson, R.G. and 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, Kaufmann Publishers, San Francisco, CA.

Makuil A., Alinezhad A., Kiani R. and Zohrehbandian, M.M. (2008), 'A goal programming method for finding common weights in DEA with an improved discriminating power for efficiency'. [Online]. [Retrieved October 5, 2010], http://www.jise.info/issues/volume1no4/21.pdf

Marcoulides, G.A. (ed) (1998), Modern methods for business research, Lawrence Erlbaum Associates, Mahwah, New Jersey.

Peters, G. (1994), Benchmarking customer service, Pitman Publishers, London.

Whitten, J., Bentley, L. and Dittman, K. (2006), System analysis and design methods, 6th ed., McGraw-Hill, New York.