Scientific
Research

System Analysis of Mystery Shopping Data: ISO 9001 for Control of Service Quality

E. G. Aminev[1], Qing Wang[2]

[1]Hi-Tech Psychology Central Institute, HiTPCI (Director), Moscow, Russia
[2]School of Public Administration, University of International Business and Economics, Beijing, China
Email: aminev-eg@mail.ru, russiacn@163.com

Abstract

The dynamics of service staff in advertences in the filling stations network of one of the largest Russian oil company was investigated. Errors of personnel were revealed from mystery shopping questionnaires and it was proved that the distribution of mistakes can be described by refusing theory. Zipf-Mandelbrot's model of the statistical structure of languages and, likewise, Gompertz-Makeham's law of human mortality, are not allowed for approximation. Customer loyalty can be effectively predicted by methods of fast integer linear programming. Relevance factors allow one to authentically judge which components of social behaviour of the personnel have the greatest influence on the benevolence of clients to a company. The personnel are given concrete technologies of training design, statistical methods of detection of incorrect questions and algorithms of estimation of mystery consumer attainment. It is proposed to amend the recommendations of statistical methods to supplement the standard "quality management System" ISO 9001:2008 in the parts concerning with such subjects as personnel, education, training and the modeling of service quality.

1. Introduction

Applying standards of management quality to operation [1] [2] and the invitation of the "Mystery customer" [3] [4] have become an established practice in modern business. However, the interaction of these two formats in practice is not unified and does not lead to building customer loyalty models and algorithms of training design orientated on both personnel and buyer's needs. On the one hand, the application of statistical methods was an actual breakthrough in the development of quality control systems [5] [6]. The global adoption of ISO 9001

Family of quality management standards ISO 9000 [1] [2] has taken place in many countries and the certification process is actively discussed from various positions in financial publications [7] [8].

At the same time, the complex nature of statistical tools used in the standard to maintain and improve the quality of management according to the requirements of ISO 9001:2000 can hardly be considered as complete. In fact, many units of modern computer technologies are not even mentioned. For example, neither integer programming [9] nor neural networks [10] are included. Statistics of recommended statistical methods is about these technologies. It is proposed to use descriptive statistics more often (39), following by sampling (37). Process capability analysis was mentioned 25 times, measurement analysis—23, regression analysis—22, time-series analysis—18, Statistical process control charts—16, reliability analysis—15, hypothesis testing—15, design of experiments—14, statistical tolerance—6, simulation—4, hypothesis testing 3 times.

On the other hand, corporate trainers and psychologists of HR departments must regret that the least developed parts of ISO 9001:2000 are those that are intended to improve the sales process and service. And that is when ISO 9001 is declared a universal system for all kinds of "products" as well as service. The same is with section 6, "Resource management", and subsection 6.2, "Human resources". Statistical methods for most subsections of that important part are not defined at all! And only subsection 6.2.2, which estimates the effectiveness of measures, competence and training processes of the simplest methods such as descriptive statistics and sampling is offered.

This state indicates that the psychological community did not take the leading role in the development of a standard "quality management System" ISO 9001 in parts concerning personnel behaviour. This situation points to a number of negative trends. Industrial psychologists are not interested in the methodology of quality standards, avoid statistical control of training effectiveness (including sales training) and they are not engaged in the development of statistical models and validity checking of stuff control data coming neither from internal company services nor from "mystery shopper" outsourcing organizations. Finally, the choice of training procedures is executed without regard to predictive models of customer loyalty.

The aim of this article is to offer a complete solution to service quality management. It contains computation analysis of employee falls distribution type and determines the weak link in the training of personnel. We also include a sample of building the main service factors that need to be taken into account to make trade set more attractive. Finally, we give recommendations as to how design staff training should be focused on customer loyalty. Dynamics of service staff inadvertences in the filling station network of one of the largest Russian oil company was investigated. However, technology may be easily adapted to other commercial networks and services.

2. Methods

Methods of mathematical and statistical analysis are authoring. Data was obtained from mystery shopping questionnaires and includes a set of service staff inadvertences in the filling stations network of one Russian oil company. Each questionnaire had 44 questions, and in 28 items no remarks was made. The model was based on $N = 903$ observations. In addition the data from 63 inspections of petroleum stations had personnel estimation of satisfaction to quality of service and was used for computing a model of loyalty.

Thus, the mathematical and statistical analysis consisted of two stages. The first stage "Situation vision" represents analysis of failure rate of obligatory company personnel behavior rules. The distribution curve was checked for compliance with Zipf-Mandelbrot law or the principle of relative frequency in language [11]-[13]. Also it was compared with Gompertz-Makeham's function of human mortality [14] and with models of refusing theory [15]. Items that did not fit the identified model were determined, and proposals to modify and reduce the questionnaire were made. The aim of the second stage "Case regulation" was to calculate the factors influencing the positive dynamics of customer loyalty. For this purpose functional relation between satisfaction index and aggregated variables describing total estimation of carelessness in 10 groups of questions was computed. Regression analysis and fast integer linear programming methods were applied [9] [16].

3. Results

3.1. Situation Vision: System Sustainability Resource

System sustainability resource is lost by portions. The weak link fails first and shows the main vulnerability in

the mechanism of situation-behavior (*locus minoris resistentiae*). Finding this base defect and remedying it leads to rapid success. For this we must look at the situation as a whole, and determine the law describing the distribution of service inadvertences. Mistakes of personnel were ranked in order of their frequency. Communication errors (which are most characteristic for operators at the moment) received first place. Last place went to the inadvertences that were detected infrequently. In the analysis of $N = 903$ observations, we have $F = 166$ times when operators made incorrect responses to prohibited acts. This item examines the vigilance of personnel in situations when the client contravenes the safety regulations (i.e. repairing a car near a petrol pump or fuelling into a plastic can). In $F = 143$ cases, mystery customers reported that service station attendants didn't say goodbye to them. In $F = 125$ cases, operators had not given a prescribed response to a question about gasoline quality. We have $F = 122$ questionnaires that indicate that staff did not smile, and, in $F = 99$ cases, mystery customers were not greeted by station attendants. Comments about untidy appearances and the uncleanliness of uniforms however, were extremely rare. These observations were only identified only in $F = 2$ cases and shared in row positions 22 and 23. Finally, in order in places 24-26, were comments in relation to the attendants such as "hands in pockets", " slang words used in speech" and "relaxed poses" $(F = 1)$.

For each rank I, $\ln F_i$ and the ratio F_i/F_{i+1} were calculated. It was found that the dependence between frequency F and rank i can't be approximated by neither biolinguistic law of Zipf-Mandelbrot nor by Gompertz-Makeham's law of mortality. Linearization was achieved only in mono-logarithmic scale, i.e. the failure rate of component schemes of social behavior operator obeys a simple exponential function, and this is the simplest model of refusing theory (1).

$$f_i = f_0 \cdot \exp(-k \cdot i), \tag{1}$$

where i is the number or rank, f the probability of this type of failure and f_0 and k coefficients of the model. We obtained the following values of coefficients $f_0 = 5.20$, $k = -0.21$.
Therefore, the logarithm of Equation (1) results in a simple linear dependence of the failure rate of quality service.

$$\ln f_i = 5.20 \cdot \exp(-0.21 \cdot i). \tag{2}$$

From Equation (2) follows that the attenuation rate of failure has value $k = \exp 0.21 = 1.24$. Thus, the coefficient of resource losses F_i/F_{i+1} has an almost constant value of approximately 1.24. This dependence may be demonstrated on a graph as very close to a straight line (see **Figure 1**).

As we see from **Figure 1**, point 7 "friendly eyes" does not fit the general curve. This point is sitting too low $f_7 = 37$. For this reason we have a jump in the coefficient of resource loss in the previous point $f_6 = 85/37 = 2.3$. It should be noted that this item provoked puzzling indignant questions from service station attendants: What does it mean to have "friendly eyes", "how can it be measured?!" So we can assume that the law of the constancy of resource losses remains in force and gives us an opportunity to eliminate incorrect questions.

The main idea of disclosed law is that we have a gradual loss of resources. Note must be made that resources here are understood to be from positions of human capital theory [17] and includes individual competences such as health, working capacity, adaptive possibilities, interpersonal skills, attention, etc. We can say that each employee spends a defined portion of his resources to comply with every item of requirement. Therefore, fulfilling each new requirement expends new resources, and so we have gradually less and less resources with every step leading to an increase in mistakes. This portion of resource loss explains why we have an exponential law. Another conclusion is that, if we identify the weak link and eliminate the main failure, the loss of resources in fact does not stop. Tiredness, fatigue, natural disturbances and stress does not disappear. Professional sustainability again needs to feed from new resources and new "pieces" and thus the attendant himself is shaken by the next components of his situational behavior. Hence the main task of training and psychobiological support includes ways of increasing resources and reducing the costs of complying with requirements. On the other hand, the law shows the importance of minimizing the set of requirements according to the level of their real significance to buyers. In this case, resources would be distributed over a smaller number of items, and would ensure their implementation.

3.2. Case Regulation: Design of Customer Loyalty Oriented Training

As shown in the preceding paragraph, the employee allocates his resources sequentially between the items of
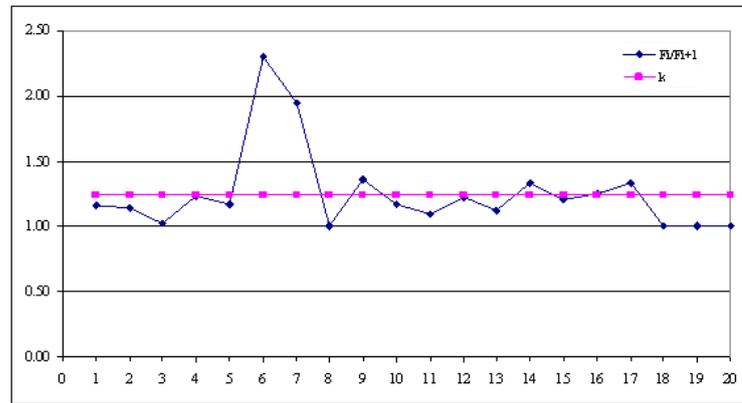
Figure 1. The dynamics of the ratio of the number of failures of quality service F_i/F_{i+1}. F_i is frequency of observations of this type of mistake in mystery customer reports, i is the rank of error in the row of personnel inadvertences.

corporate requirements as prescribed in mystery customer questionnaires. Therefore, complete minimization of the requirements list according to the weight of items in relation to the positive dynamics of customer loyalty is a problem of great importance. We also took values of customer loyalty from mystery shopping questionnaires as the form had a special paragraph for this purpose. Supervisory pilots put on the list their own index of loyalty or service satisfaction score. This score varied from zero to ten, where zero means that satisfaction level was very low, and ten written if the station produced an excellent impression. We adjusted this index to a five-point scale:

9 - 10 scores (+2)—service is excellent, want to recommend it to my friends

7 - 8 scores (+1)—service is rather good, I'm going to use this station again

5 - 6 scores (0)—service is average, I'll come only by the way

3 - 4 scores (−1)—I did not like the service and I will not come again

0 - 2 scores (−2)—I was very disappointed with the station, I will alert friends to never use it

As a reminder, the questionnaire contained 44 items. The form has 10 unequal groups of questions: I. Cleanness (environment); II. General personnel requirements (greeting, farewell); III. Communication policy (attentiveness, listening ability); IV. Response to prohibited acts; V. Standard answers to the control questions (fuel quality?); VI. Service process (estimation of station attendant activities); VII. Communication style (communication distance with customers); VIII. Clothing and appearance of employees (uniform, untidy appearance); IX. Body posture, gestures, facial expressions (smile, friendly eyes); X. Other requirements (no conversations between employees). Here we also present some questions of each group (in brackets). We computed aggregated variables $X_1, X_2, \cdots, X_{10}$ which estimate the degree of carelessness in 10 groups of questions. Then polar groups were marked out with the highest and lowest scores. The groups were then used to work out a Loyalty model via a fast integer linear programming method. For each of the aggregated parameters X_i, relevance coefficients a_i were calculated. The coefficients of relevance include the following graduations: 5 - extremely significant, 4-very significant, 3-significant, 2-less important, 1. of little importance. The sign of the relevance coefficient may even be negative if the requirement is not very important and blind adherence may even annoy the customers. In our case, we have the following equation:

$$L_{\text{theor}} = 0.37 \times \left[27.2 - \left(5X_1 + 3X_7 + 2X_2 + 2X_8 + X_3 + X_4 + X_5 + X_9 - X_{10} \right) \right] \tag{3}$$

Then, the equation obtained by polar groups we had applied to all forms and theoretical values of loyalty indexes was calculated. We compared prognostic value of used method and results of regression analysis as recommended in ISO 9001:2008 [2]. Surely, regression analysis loses a wager and so this material is not considered here. From **Figure 2** we see that the model obtained by the fast integer programming predicts well the real loyalty indexes from mystery shopping forms.

At the same time, the deterioration of service quality follows with a significant increase in fluctuation of va-
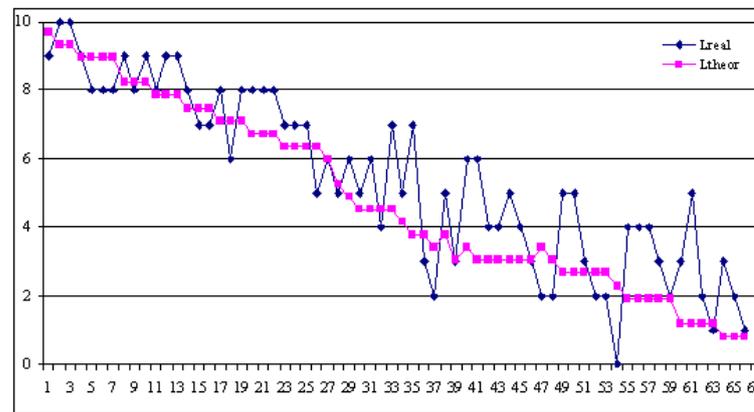
Figure 2. The dynamics of theoretical and empirical values of customer loyalty index. L_{theor} -squares, L_{real} -rhombs.

riance between model (theoretical) and empirical (mystery customer) values. This result indicates a lack of preparation in the supervisory pilots themselves. We call this stage of preparation-mystery customer calibration training. Supervisors who filled forms in the same way should produce similar estimates of service attractiveness. Also, at this stage, we can identify unnecessary and missing items on the questionnaire from the point of impact on custom loyalty

However, the main result is the following: Relevance coefficients significantly show which components of an employee's social behavior has the greatest impact on the loyalty index or on a clients' favorable attitude. In our case, we see that company's image lost more if personnel permit untidiness, a lack of station territory and cleanliness (relevance coefficient $a_1 = 5$, extremely significant). Therefore, in this case, common trade and communication training are definitely not paramount. Sellers with happy expressions, smiles and "friendly eyes" joyfully invited visitors but, if the counter-desks were dirty, the customers would not be attracted and would not trust the products. That's why we insist that computer behavioral training and among them training aimed at reinforcement of tidiness instinct are of great importance.

Group number seven received second place in the model ($a_7 = 3$, significant). This group's name is "Communication style" (requirements: communication distance with customers no less than 1 - 1.5 m, attendant stands behind the client, etc.). Groups number two and eight were less important ($a_2 = a_8 = 1$). Aggregated variable X_2 shows the performance quality of rules of subparagraph "General personnel requirements" (greeting, farewell). X_8 indicates "Communication style" (communication distance with customers). Parameters X_3, X_4, X_5, X_9 are at the end of row with relevance coefficient equal to one (of little importance). Relevance coefficient values show the actual order of aims for personal training, and this must be a strict rule for corporate trainers and psychologists in their work with employees. With the achievement of positive results, relevance coefficients must be recalculated and new line of priorities found.

4. Discussion

Dynamics of personnel inadvertences identified by mystery customer forms are not described by Zipf-Mandelbrot's distribution of words in speech nor by Gompertz-Makeham's law of mortality. A frequency curve is formed in accordance with refusing theory and obeys exponential distribution. The main feature of the revealed law purports the gradual loss of employee durability resources. Hence, the problem of training consists of an increase of human resources and decrease of psychological expenses for personnel. On the other hand, the law shows the importance in minimizing the number of requirements according to their importance for customers. This way we give an employee the ability to distribute his individual resources on a smaller number of points to provide their performance. The analysis of strong deviations provides a method to control the correctness of questions in the mystery shopping form.

Customer loyalty may be effectively predicted by methods of fast integer linear programming. Relevance factors allow one to judge authentically what components of the social behaviour of the personnel have the greatest influence on benevolence of clients to a company and define perspective directions of training and psy-

chobiological support. The high urgency of the computer behavioral coaching was proved (especially via programs aimed at reinforcing base instincts). A synergistic approach provides a greater chance to succeed in retaining a client's loyalty than by pure trade communication training. Fluctuation of variance between mystery customer appraisal and prognostic models indicates a lack of preparation in supervisory pilots and the necessity of calibration training to improve them.

The psychological community did not take the lead in the development of the standard "quality management System" ISO 9001. As a consequence, standards of statistical methods used by calculating, commissioning, support and practical improvement of a system of quality management do not contain recommendations about Section 6.2., "Human resources". Recommendations are only offered for subsection 6.2.2 to estimate the effectiveness of measures, competence and training processes for the simplest methods such as descriptive statistics and sampling. Therefore, it's necessary to organize an international group to improve system ISO 9001 in the section of work relating to the personnel and training. It is offered to hold a conference to discuss the given initiative.

References

[1]   ISO 9001:2008, Quality Management Systems—Requirement.

[2]   ISO/TR 10017:2003, Guidance on Statistical Techniques for ISO 9001:2000.

[3]   Leeds, B. (1995) Mystery Shopping: From Novelty to Necessity. *Bank Marketing*, **27**, 17-23.

[4]   Wilson, A.M. (1998) The Role of Mystery Shopping in the Measurement of Service Performance. *Managing Service Quality*, **8**, 414-420. http://dx.doi.org/10.1108/09604529810235123

[5]   Edwards, D.W. (1986) Out of the Crisis. MIT Press, Cambridge, 489 p.

[6]   Mangelsdorf, D. (1999) Evolution from Quality Management to an Integrative Management System Based on TQM and Its Impact on the Profession of Quality Managers in Industry. *The TQM Magazine*, **11**, 419-425. http://dx.doi.org/10.1108/09544789910287737

[7]   Clare, C.-C., Mark, G. and Boon, W.T. (2003) Does ISO 9000 Certification Improve Business Performance. *International Journal of Quality & Reliability Management*, **20**, 936. http://dx.doi.org/10.1108/02656710310493643

[8]   Sharma, D.S. (2005) The Association between ISO 9000 Certification and Financial Performance. *The International Journal of Accounting*, **40**, 151-172. http://dx.doi.org/10.1016/j.intacc.2005.01.011

[9]   Chen D.-S., Batson, R.G. and Dang, Y. (2010). Applied Integer Programming: Modeling and Solution. John Wiley and Sons, New York, 488 p.

[10] Murray, S. (1993) Neural Networks for Statistical Modeling. Van Nostrand Reinhold, 235 p.

[11] Zipf G.K. (1935) Selective Studies and the Principle of Relative Frequency in Language. Houghton-Mifflin, MIT Press, 125 p.

[12] Mandelbrot, B.B. (1953) An Informational Theory of the Statistical Structure of Languages. In: Jackson, W., Ed., *Communication Theory*, Academic Press, Princeton, 486-502.

[13] Li, W.T. (1992) Random Texts Exhibit Zipf's-Law-Like Word Frequency Distribution. *IEEE Transactions on Information Theory*, **38**, 1842-1845.

[14] Gompertz, B. (1825) On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of the Royal Society of London*, **115**, 513-585. http://dx.doi.org/10.1098/rstl.1825.0026

[15] Rausand, M. and Hoyland, A. (2004) System Reliability Theory; Models, Statistical methods, and Applications. John Wuiley & Sons, New York.

[16] Schrijver, A. (1986) Theory of Linear and Integer Programming. Wiley, Tilburg University, 470 p.

[17] Becker, G.S. (1994) Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education. University of Chicago Press, Chicago, 412 p.