

Usage of Different Statistical Methods On a Single Data and Its Implications On Managerial Decision Making

** Satya S.*

*** Dr. R. K. Gopal*

INTRODUCTION

Application of market research can be at a strategic level or at a tactical level in an organization. Market research can assist in strategic decision making such as arriving at demand, forecasting, identifying market segments and positioning strategies for the products. Tactical applications can be in the area of product testing, price research, brand equity and advertising research, research relating to service levels in distribution channels, logistic efficiencies and customer satisfaction.

Both strategic and tactical applications of the market research findings for an organization are critical as it directly affects the economies of the scales of the company and also affects the employees, competition, shareholders, and suppliers to name a few. Hence, market research and its effect on organization's function is crucial. Market research these days is highly dependent on statistical interventions. The advent of statistical software packages has shifted the domain of acumen in statistical analysis from 'computation skill sets' to 'interpretation of output' skill sets.

Interpretation of output is solely done on the output data given by the statistical packages. Based on these output data, managerial implications are deliberated and processes are built in an organization for day-to-day practices. Any given statistical package has more than one method to analyze a data. For example, multiple regression analysis can be done using forward stepwise regression and back ward stepwise regression. It is of immense interest to note if the output generated by the two methods are similar or not. On analyzing the same data, if two methods of regression give two different interpretations, then there arises a need for another research to find out which of the two should be prescribed for managerial practices.

Hence this research's objective is to generate data with primary study by taking a marketing problem. Single data was analyzed using two methods of multiple regressions. Outputs of these data were interpreted and discussed to see how it affects the managerial decision-making.

STATEMENT OF THE PROBLEM

There are a variety of statistical methods in each multivariate technique available in SPSS package. For example, in multiple regressions, both forward and backward regression can be applied on the same data. However, consistency of results is not known and this in turn would affect interpretations for managerial decision making.

OBJECTIVES OF THE STUDY

- To apply forward and backward multiple regressions on a market research data and check the consistency of results.
- To know how these results alter the interpretations of study and affect the implications for decision makers.

TOPIC OF THE STUDY

The Topic of the study was 'Factors Influencing Retail Sales of Medical Stores in Bangalore City'.

LITERATURE REVIEW

Literature review is done to understand the various methods of regression analysis used in research. Though there were no studies where two methods were used on the same data, to check the consistency of results, the variety of methods available in regression can be understood.

Multiple regressions have long been a popular and widely used method in marketing research. Multiple regressions are a common statistician method in marketing studies (Bass and Frank M. 1977). Multiple regression analysis has been called as one of the most popular tools in the marketing researcher's kit (Paul and Tull, 1978). Of all the 'fancy' multivariate procedures, regression analysis is by far the most widely used in marketing research (Lehmann, 1979).

Fonton N H, Palm R (1998)- Performed a Monte Carlo simulation study of the predictions given by multiple linear

** Assistant Professor, MBA Department, R. V. Institute of Management, Jayanagar, Bangalore.*

E-mail : ss.manuscript@gmail.com

*** Professor and HOD, MBA Department, R. V. Institute of Management, Jayanagar, Bangalore.*

regression equations combining 13 methods of variables selection with six methods of estimation of regression coefficient. The increase in precision is only about 3 to 4 percent, with regards to the widely used ordinary least squares estimation combined with the stepwise selection procedure. None of the compared methods leads to regression equation, which are useful for making predictions when the number of predictions is greater than the number of observations.

Roland T Rust (1998) observed that marketing researchers who want to do multiple regressions often have data in which some of the standard regression assumptions are violated. Flexible regression is a new method for performing a non-parametric multiple regression while relaxing several of the standard assumptions of regression. In particular, the assumptions of linearity and normal errors are relaxed. The approach is based on non-parametric density estimation, which results in a more synergistic and less parametrically constrained method of analysis. Necessary conditions are established for the rejection of significance lists that correspond to the major tests of regression.

Roland Linder, Jeannine Geier, Mathias Kolliker (2004) observed that the most commonly used modeling methods for targeting customers in direct marketing are artificial neural networks, classification trees and logistic regression. The authors investigated the predictive performances of the 3 methods in a competitive test in a simulated direct marketing scenario.

Athanasios Vasilopoulos (2005) conducted a research with the aim of improving the level of comfort with regression methodology. It discusses the theory behind regression methodology and outlines a step- by-step procedure which will allow one to construct a regression forecasting function for both linear and multivariate cases. The model testing for final regression equation is also discussed.

So young Sohn Hong Sikkim (2006) conducted a research using random effects of logistic regression model, and efforts were made to predict the default of funded Small Medium Enterprises of Korean Government based on financial and non-financial factors. Many financial institutions have utilized various default prediction models using logistic regression, multiple discriminant analysis, neural network and multidimensional scaling. These approaches could not explain the situation where the enterprises exhibit different default probability under the same conditions. Random effects logistic regression model also deals with the uncertainty that cannot be explained by financial and non-financial factors.

Kiers and Smilde (2007) said that Regression tends to give unstable and unreliable regression weights when predictors are highly collinear. Several methods have been proposed to counter this problem. The authors have compared six methods- that are Particle Least Squares (PLS), Principle Component Regression (PCR), principle covariates of power regression, reduced rank regression and two variants of power regression. Especially in studies with more real life like simulated data, PLS and PCR give poor recoveries of regression weights in conditions with low noise and co- linearity as far as prediction is concerned. Prediction suffers far less from co linearity than recovery of regression weights.

Nikolopoulos K, P Goodwin, A Patelis, V.Assimakopoulos (2007) through their study tested the accuracy in forecasting the impact of a sports program on Greek TV audience by using multiple linear regression. First bivariate regression model was used and then three different types of artificial neural networking were used followed by three forms of nearest neighbor analysis that were finally compared to human judgment. High accuracy was obtained from forecasts based on simple bivariate regression model, a simple nearest neighbor procedure and from two of the artificial neural networks.

Literature review clearly points out to the variety of regression analysis methods available as statistical tools to analyze marketing data inputs. This study aims in a small way to test the consistency between two most frequently used simple methods of regression-namely forward stepwise regression method and backward stepwise regression method on the same data.

METHOD OF THE STUDY

The topic of the study was “Factors Influencing Retail Sales of Medical Stores in Bangalore City”. The questionnaire was constructed after considering the various factors that affect the retail sales; especially the medical retail sales. After speaking to three medical storeowners, the main factors affecting the sales of medical stores were arrived at. The five main factors under which questions were built are I). Store location (STOLOC) 6 questions, II) Stores Service Quality (STOQUA) 10 questions, III) Competition Prevalent in the Area (COMPRES) 3 questions, IV) Doctor's Reference (DOCREF) 3 questions, and V) Store Promotion (STOPRO) 3 questions. The questions were set to a five point likert's scale, with “strongly agree” denoting 5 points and “strongly disagree” denoting 1 point.

Transform analysis was used to transform the items in the 5 main factors to a composite factor score. So in total, 25 items would study the factors which affect the sales of medical stores. Daily sales of medical stores and the factors affecting the sales are put to the regression model. The same data was analyzed using forward and backward regression methods. Both the outputs were interpreted and the managerial implications are discussed. It is of interest to note whether there is any consistency in the interpretation between the two multiple regression methods.

All medical stores in Bangalore City serve as the universe of the study. Through Area Sampling technique, important commercial, residential, shopping, and institutional areas were included. In the selected geographic areas, medical stores were chosen based on convenience. The data was collected using structured questionnaires. Few shop keepers considered it convenient to answer the questions, as they were orally administered; while others had asked to leave the questionnaire behind to fill it up themselves. In total, about 250 questionnaires were distributed and 200 questionnaires were returned to the researchers before the set dead lines and based on these 200 questionnaires, further analysis was conducted.

ANALYSIS OF DATA

In this research, forward stepwise regression and backward step wise regression was applied on the same data to find out whether both the methods give the same significant variables as output.

FORWARD STEPWISE MULTIPLE REGRESSION

In forward multiple regression, the algorithm adds one independent variable at a time, starting with the one which explains most of the variation in sales turnover, and adding one more variable to it, rechecking the model to see that both variables form a good model, then adding a third variable if it still adds to the explanation of the dependent variable. In the output of forward stepwise regression, the regression ends with two independent variables out of five independent variables remaining in the regression model. The two variables in the model are Store Location and Doctor's References.

Table 1: Variables In The Regression Model

Variable	B	SE B	Beta	T	Sig T
STOLOC	0.313655	0.074336	0.592073	4.219	0.1012
DOCREF	6.114345	2.074735	0.413532	2.947	0.0122
(CONSTANT)	7.771172	3.094059		2.512	0.0273

Table 2: Analysis of Variance (ANOVA)

	DF	Sum of squares	Mean Square
Regression	2	1395.54630	697.77315
Residual	12	120.18703	10.01559
F=69.66873		Signif F= .0000	

RESULTS

It is noticed that only one significant variable, (with sig T<0.05) at more than 95 % confidence level is “doctor's references” (sig T 0.0122). However, “store location” is significant at 0.1012 which is also close to 90% confidence level. The F test for the model indicates that it is highly significant = 69.66873.at sig F = .0000 and R square value for the model is 0.92071.If we use this equation, the model would be

$$\text{SALTUR} = 7.771172 + 6.114345 (\text{DOCREF}) + 0.31(\text{STOLOC})$$

BACKWARD STEPWISE REGRESSION

This is performed on the same set of five independent variables. This regression model starts with all the five independent variables and gradually eliminates the variables one after the other (those which do not explain much of the variation in dependent variable), until it ends with an optimal mix of independent variables according to the preset criteria for the exit of variables, resulting in a model with only one independent variable “Doctor's reference” DOCREF, which was significant at 95% confidence level. It is observed that the ANOVA output for the backward regression is the same as for the forward regression.

Table 3: Variables in the Regression Model

Variable	B	SE B	Beta	T	Sig T
DOCREF	7.124346	2.074735	0.413532	2.947	0.011
(CONSTANT)	8.771172	3.094059		2.512	0.0273

$$\text{SALTUR} = 8.771172 + 7.124346(\text{DOCREF})$$

(Contd. on page 56)

- 2) The impact of corporate image on quality, customer satisfaction and loyalty for customers with varying degrees of service expertise", *International Journal of Service Industry Management*, Volume 9, Number 1, 1998, pp. 7-23(17).
- 3) A. K. Chakrabarty, 'The Efficiency of LIC during the Post-Liberalization and Globalization period A Case Study of West Bengal', *Business Perspective*, Vol. 8, No. 2, July Dec. 2006, pp.25-39.
- 4) Citizen's charter, India Post, <http://www.indiapost.gov.in>.
- 5) Haemoon Oh, 'Service quality, customer satisfaction, and customer value: A holistic perspective', *International Journal of Hospitality Management*, Volume 18, Issue 1, March 1999, pp. 67-82.
- 6) Henning Gebert, Malte Geib, Lutz Kolbe, Gerold Riempp, 'Towards Customer Knowledge Management: Integrating Customer Relationship Management and Knowledge Management Concepts', Paper presented at the Second *International Conference on Electronic Business*, Taipei, Taiwan, Dec. 10-13, 2002.
- 7) Jane E. Fountain, 'Paradoxes of Public Sector Customer Service', *Governance*, Volume 14 Issue 1, Dec. 2002, pp. 55-73.
- 8) Ms. Sonia Riyat, 'Globalisation of Indian Postal Service - A Study', *IndiaMBA.Com*.
- 9) T. Ramasamy, 'Customer Service at Arrupukottai Head Post Office', *Southern Economist*, Vol. 45, No. 1, May 1, 2006, pp. 67-70.

(Contd. from page 33)

DISCUSSIONS

It is observed from forward regression output that "Doctor's reference" plays a critical role in the daily sales turn over of the outlet (>95% confidence level). "Store location" has also emerged as a significant factor (close to 90% confidence level). More than "over the counter sales" drugs, "prescription drugs" play a critical role in the daily sales turnover of medical stores. The backward regression output also indicates that the doctor's reference is crucial (>95 % confidence level). This reiterates that "prescription drug" sales are affected mainly by the "doctor's reference" factor. Those retailers whose shop is located in close proximity to a doctors' clinic and who have established good contacts with the doctor in the area do better business than those retailers who have not established good contacts with the nearby doctor in the area.

MANAGERIAL IMPLICATIONS

Even though "store location" has emerged as a factor in the forward regression output; while comparing results from forward and backward regression outputs, the channel managers need not give much attention to the "store's location". Even if a medical store is strategically located but if the owner of the shop has not established good networking with the doctors in the area of proximity, it is quite likely that he may not be affecting good sales turnover. In other words, the owner of the store should establish a good network with the doctors situated in proximity to the store and link his retail efforts to the doctor. The retail stores can print their contacts and other details and provide the doctor with stationeries such as prescription pads, note pads, which would also be maintained in records by the patients. This would act as ready recon for the next purchase. This communication would prove to be very effective for the retail owners, as it is extremely cost effective and would increase the rapport between the doctors in proximity and the medical stores.

CONCLUSION

Even though different statistical methods lead to different results to facilitate decision-making processes, the research consultants should interpret the outputs of the sponsoring company in such a way that it will not lead to wrong implementation in the market area. In this research, even though both "doctor's reference" and "store location" emerged as significant factors, "doctor's references" influence the sales turnover than the "store's location" as is evident from both regression method outputs.

SCOPE FOR FURTHER RESEARCH

Further research can be carried out to find out how different statistical methods employed may have an impact on decision-making. For example, in certain data, both parametric and non-parametric analysis can be done, and it will be of interest to note if it affects the interpretation and thereby decision making in the market place.

BIBLIOGRAPHY

1. Bass, Frank M (1977) Application of Regression Models in Marketing: Testing v/s forecasting in Multivariate Methods for Market and Survey Research.
2. Green Paul E and Donald S. Tull (1978) *Research for Marketing Decisions* Englewood Cliffs, Prentice-Hall, Inc.
3. Lehmann, Donald R (1979) *Market Research and Analysis* Homewood, IL: Richard D Irwin, Inc.
4. Fonton N H, Palm R (1998) *Empirical Comparison of Prediction Method in Multiple Linneare regression* Review Statistics Application 46(3).
5. Roland T Rust (1998) *Flexible Regression* Journal of Marketing research. 25(1).
6. Roland Linder, Jeannine Geier, Mathias Kolliker (2004) *Artificial Neural Networks, Classification Trees and Regression: Which method for which customer base?* Journal of Database Marketing and Consumer Strategy Management 11(4).
7. Athanasios Vasilopoulos (2005) *Regression Analysis Revisited* Review of Business (26) 3.
8. So Young Sohn, Hong Sikkim (2006) *Random Effects Logistic Regression Model for Default Prediction of Technology Credit Guarantee Fund* European Journal of Operational Research 183(1).
9. Henk A L Kiers and Age K Smilole (2007) *A Comparison of Various Methods for Multivariate Regression with Highly Collinear Variables* Statistical Methods and Applications 16(2).
10. Nikolopoulos K, P Goodwin, A Patelis, V. Assimakopoulos (2007) *Forecasting with One Information: A comparison of multiple regression with alternative forecasting approaches*. European Journal of Operational Research 180(1).