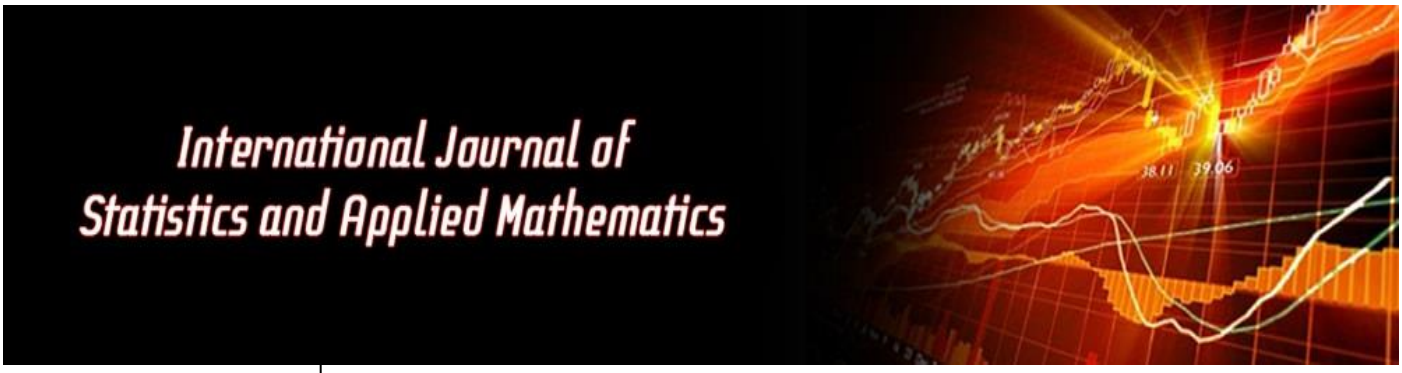
International Journal of Statistics and Applied Mathematics 2018; 3(1): 476-480



ISSN: 2456-1452
Maths 2018; 3(1): 476-480

www.mathsjournal.com
Received: 22-11-2017
Accepted: 24-12-2017

G Satyanarayana Reddy
Department of Statistics, Govt College for Men (A), Kadapa, Andhra Pradesh, India

N Viswam
HOD, Department of Statistics, Hindhu College, Guntur, Andhra Pradesh, India

Correspondence
G Satyanarayana Reddy
Department of Statistics, Govt College for Men (A), Kadapa, Andhra Pradesh, India

Logistic regression model applications in retail industry

G Satyanarayana Reddy and N Viswam

Abstract
The Indian retail industry has emerged as one of the most dynamic and fast-paced industries due to the entry of several new players. It accounts for over 10 per cent of the country's Gross Domestic Product (GDP) and around 8 per cent of the employment. India is the world's fifth-largest global destination in the retail space. The Boston Consulting Group and Retailers Association of India published a report titled, 'Retail 2020: Retrospect, Reinvent, Rewrite', highlighting that India's retail market is expected to nearly double to US$ 1 trillion by 2020 from US$ 600 billion in 2015, driven by income growth, urbanization and attitudinal shifts. Retail spending in the top seven Indian cities amounted to Rs.3.58 trillion (US$ 53.7 billion), with organized retail penetration at 19 per cent as of 2014.

Keywords: Logistic regression, model applications, retail industry

Introduction
Promotions can be categorized based on the intended coverage of a single promotional message. For example, a single television advertisement for a major event, an in store promotion or a promotion on a newspaper could be seen by hundreds of customers at the same time. Such mass promotion, intended to reach as many people as possible, has been a mainstay of Retailers' promotional efforts for a long time. Unfortunately, while mass promotions are delivered to a large number of people, the actual number that fall within the marketer's target market may be small. Because of this, many who use mass promotion techniques find it to be an inefficient way to reach desired customers. Instead, today's marketers are turning to newer techniques designed to focus promotional delivery to only those with a high probability of being in the marketer's target market. In so doing, the marketing focus shifts away from the breadth of customer base to the depth of each customer's needs. The performance metric changes from market share to so-called "wallet share".
Data mining is used to construct six types of techniques aimed at solving business problems: classification, regression, time series, clustering, association analysis, and sequence discovery. The first two, classification and regression, are used to make predictions, while association and sequence discovery are used to describe behavior. Clustering can be used for either forecasting or description.
This article examines the performance of the mass promotions Vs target promotion here we use logistic regression technique with qualitative and quantitative variable to predict the response of a customer to a particular promotion and we also propose how can we use the customer segmentations to know a customer at a 360^0 view and how to use these segmentations to select correct populations for any target promotions. Finally, we draw conclusions based on the discussion.

Logistic Distribution
A random variable X is said t have Logistic distribution with mean μ and variance σ^2, if it has a cumulative distribution function,

$$F(x,\mu,\sigma) = \left\{1+\exp\left[\frac{-\pi(x-\mu)}{\sigma\sqrt{3}}\right]\right\}^{-1} \quad -\infty < x < \infty, \quad -\infty<\mu<\infty \quad \sigma>0 \tag{2.1}$$

and probability density function related to its distribution function,

$$f(x,\ \mu,\ \sigma) = \frac{\pi}{\sigma\sqrt{3}} F(x,\mu,\sigma)\,[1-F(x,\mu,\sigma)] \tag{2.2}$$

Alternatively, these functions may be expressed as

$$F(x,\mu,\sigma) = \frac{1}{2}\left\{1+\tanh\left[\frac{\pi(x-\mu)}{2(\sigma\sqrt{3})}\right]\right\} \tag{2.3}$$

and $$F(x,\mu,\sigma) = \frac{\pi}{4\sigma\sqrt{3}}\operatorname{sech}^2\left[\frac{\pi}{2}\frac{(x-\mu)}{(\sigma\sqrt{3})}\right] \tag{2.4}$$

Properties of Logistic Distribution

1. The density f (x, μ, σ) is bell shaped and symmetrical with heavier tails than a normal density with the same mean and variance.
2. The canonical form of the logistic distribution, which corresponds to the random variable z, with mean μ = 0 and variance $\sigma^2 = \frac{\pi^2}{3}$ and has cumulative distribution function and probability density functions,

$$G(z) = \frac{1}{1+e^{-z}} \tag{2.5}$$

$$\text{and } g(z) = G(z)\,(1-G(z)) \tag{2.6}$$

Equation (2.6) and therefore equation (2.2) characterizes the logistic distribution and it is equivalent to the linearity of the transformation (known as logit)

$$\text{Log}\left[\frac{G(z)}{1-G(z)}\right] = z \tag{2.7}$$

1. The most popular application of the logistic distributions is the logit in the content of modeling Quantal response data and performing Logistic regression.
2. The distribution function of the standardized random variable $z/(\pi\sqrt{3})$ is very close to the standard normal distribution.
3. The sum of independent logistic random variables is not a logistic random variable.
4.
5. The characteristic function of z is given by

$$\phi_z(t) = \Gamma(1-it)\,\Gamma(1+it) = \prod_{j=1}^{\infty}\left(1-\frac{t^2}{j^1}\right)^{-1} \tag{2.8}$$

6. The absolute moments are given by

$$E|z|^k = 2\,\Gamma(k+1)\left[1-\frac{1}{2^{k+1}}\ \xi(k)\right] \tag{2.9}$$

7. The Logistic distribution can be obtained from a mixture of the extreme value distributions and the exponential distribution. The Logistic distribution is infinitely divisible.

8. If a random variable Y is uniformly distributed on [0,1] then the logit transformation of Y, say $\log\left[\frac{y}{(1-y)}\right]$ has the logistic distribution function G.

Specification of Logistic Regression Model

Generally linear regression model is used to approximate the relationship between a continuous response variable and a set of predictor variables. However the response variable is often categorical rather than continuous variable, for such cases, linear regression is not appropriate, but the biostatistician can turn to an analogous method, Logistic regression, which is similar to linear regression in many ways.

Logistic regression refers to techniques for describing the relationship between a categorical response variable and a set of predictor variables. In other words the good of a logistic regression analysis is to find the best fitting and most parsimonious, yet biologically reasonable, model to describe the relationship between a response variable and a set of predictor or independent variables. Generally the response variable in the Logistic regression model is categorical and usually Binary or bichotonous variable.

There are two main reasons for choosing the Logistic regression model by Biostatistician. These are:

(i) In the mathematical point of view, it is an extremely flexible and easily used function;

(ii) It trends itself to a biologically meaningful interpretation.

Suppose that $\pi(x) = E[y/x]$ be the conditional mean of y given x. The logistic regression model is given by

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \tag{3.1}$$

The logit transformation g(x) in terms of π (x) is given by

$$g(x) = \lin\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x \tag{3.2}$$

The logit g(x) has many of the desirable properties of linear regression model. The logit g(x) is linear in its parameters, may be continuous and may range from -∞ to +∞ depending on the range of x.

In the linear regression analysis, one may assume, that an observation of the outcome variable may be expressed as

$$y = E[y/x] + \in \tag{3.3}$$

Such that the error variable ∈ follows a normal distribution with mean zero and some constant variance That is constant across the levels of the independent variable. This implies that the conditional distribution of the dependent variable (y) given x is normal with mean E [y/x] and constant variance.

In the case of Logistic regression model, one may express the value of the response variable given x as

$$Y = \pi(x) + \in \tag{3.4}$$

Here, ∈ may assume one of are two possible values. If y = 1 then ∈ = 1 π (x) with probability π (x) and if y = 0 then ∈ = -π (x) with probability 1-π (x). i.e., ∈ has a distribution with mean zero and variance equal to $\pi(x)[1-\pi(x)]$.

Thus, the conditional distribution of the response variable follows a Binomial distribution with probability given by the conditional mean π(x).

A simple example of logistic regression is as follows: suppose that medical researchers are interested in exploring the relationship between patient age (x) and the presence (1) or absence (0) of a particular disease (y). For a number of patients, generally, this relationship shows logistic regression.

Estimating the Logistic Regression Model

Consider (x_i, y_i), i = 1,2,……., n be a sample of n independent pairs of observations on two variables X and Y, where y_i denotes the value of a dichotomous response variable and x_i denotes the value of the independent variable for the i^{th} subject.

Assume that Y has been coded as 0 or 1 representing the absence or presence of the characteristic respectively.

Write the logistic regression model as

$$\pi(x) = \frac{\exp(\beta_o + \beta_1 x)}{1 + \exp(\beta_o + \beta_1 x)} \tag{4.1}$$

To estimate the logistic regression model, the maximum likelihood method of estimation may be used to obtain optimum estimators for β_o and β_1.

Suppose that Y is coded as 0 or 1, then one may have

$$\pi(x) = P(y=1/x) \text{ and } 1-\pi(x) = P(y=0/x)$$

For the pairs (x_i, y_i), where $y_i = 1$ and $y_i = 0$, the likelihood function may be expressed as

$$L(\beta_o, \beta_1) = \prod_{i=1}^{n} \pi(x_i)^{y_i} \left[1-\pi(x_i)\right]^{1-y_i} \quad (4.2)$$

$$\Rightarrow \ln L(\beta_o, \beta_1) = \sum \left\{ y_i \ln\left[\pi(x_i)\right] + (1-y_i)\ln\left[1-\pi(x_i)\right]\right\} \quad (4.3)$$

Under the method of maximum likelihood estimation, one may obtain the maximum likelihood equations as

$$\sum_{i-1}^{n}\left[y_i - \pi(x_i)\right] = 0 \quad (4.4)$$

$$\text{and } \sum_{i-1}^{n} x_i \left[y_i - \pi(x_i)\right] = 0 \quad (4.5)$$

Since, the maximum likelihood equations are nonlinear in parameters, Iterative methods may be used to obtain the maximum likelihood estimators $\hat{\beta}_o$ and $\hat{\beta}_1$ for β_o and β_1 respectively. In particular, the solutions to the equations (4.4) and (4.5) may be obtained by using a generalized weighted least squares estimation.

By substituting $\hat{\beta}_o$ and $\hat{\beta}_1$, the maximum likelihood estimate of $\pi(x_i)$ is obtained as $\hat{\pi}(x_i)$.

Hence, $\hat{\pi}(x_i)$ gives the fitted or predicted value for y_i.

Since, $P\left(y=1 \middle/ x\right) = \pi(x_i)$

From (4.4), one may obtain

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{\pi}(x_i) \quad (4.6)$$

Sum of the observed values of y_i = sum of the predicted values of y_i.

Conclusion

In the Growing retail industry promotions and customer retention has become one of the major challenge. They need to focus on using the statistical methodologies like predictive modelling, segmentation, cluster analysis, factor analysis, logistic regression and decision theory to understand their customers well and they need to plan all their promotional and customer retention policies based on the statistical results. The latest software technologies will help them in visualizing and analyzing the vast amount of data they have.

References

1. Camm J, Cochran J, Fry M, Ohlmann J, Anderson D. A Categorization of Analytical Methods and Models. In Essentials of Business Analytics, 2014, 5-7.
2. Cox DR, SNELL EJ. The Analysis of Binary Data, 2nd ed. London: Chapman and Hall, 1989.
3. Fisher D, De Line R, Czerwinski M, Drucker S. Interactions with big data analytics. Interactions. 2012; 19(3):50-59.
4. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer, 2009.
5. Hosmer DW, Lemeshow S. 'A goodness-of-fit test for the multiple logistic regression model', Communications in Statistics. 1980; A10:1043-1069
6. Hosmer DW, Lemeshow S. Applied Logistic Regression, Wiley, New York, 1989
7. Information Discovery, Inc. A characterization of data mining technologies and processes: an Information Discovery, Inc. White Paper
8. Lal Rajiv, David E Bell. The Impact of Frequent Shopper Programs in Grocery Retailing, Quantitative Marketing and Economics. 2003; 1(2):179-202.
9. Lam, Shun Y, Mark Vandenbosch, John Hulland, Michael Pearce. Evaluating Promotions in Shopping Environments: Decomposing Sales Response into Attraction, Conversion,and Spending Effects, Marketing Science. 2001; 20(2):194-215.

10. Leenheer, Jorna, Tammo HA, Bijmolt. Adoption and Effectiveness of Loyalty Programs: The Retailer's Perspective, Research Report No. 03-124. Cambridge, MA: Marketing Science Institute, 2003,
11. Maddala GS. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge University Press, 1983.
12. Peter SH, Leeflang, Dick R Wittink. The Estimation of Pre- and Postpromotion Dips with Store-Level Scanner Data, Journal of Marketing Research. 2000; 37:383-95.
13. Petrissans A. Customer relationship management: the changing economics of customer relationships. White Paper prepared by Cap Gemini and International Data Corporation, 1999.
14. Rice J. ed. Mathematical Statistics and Data Analysis. Cengage Learning, 2006.
15. Roland T Rust, Katherine N Lemon. The Customer Pyramid: Creating and Serving Profitable Customers, California Management Review. 2001; 43(4):118-42.
16. Rossi, Peter E, Greg M. Allenby A Bayesian Approach to Estimating Household Parameters, Journal of Marketing Research. 1993; 30:171-82.
17. Subrahmanyan S. Using Quantitative Models for Setting Retail Prices. Journal of Product & Brand Management. 2000; 9(5)
18. Talluri KT, Van Ryzin GJ. The Theory and Practice of Revenue Management. Springer, 2005.
19. Winer R. A Price Vector Model of Demand for Consumer Durables: Preliminary Developments Marketing Science, 1985.
20. Wolverson R. High and Low: Online Flash Sales Go Beyond Fashion to Survive. Time Magazine, 2012.