

文章编号: 1001-0920(2007)11-1305-04

基于 Logistic 回归的社区满意度模型

耿金花¹, 高齐圣¹, 任敬喜¹, 张嗣瀛^{1,2}

(1. 青岛大学 复杂性科学研究所, 山东 青岛 266071; 2. 东北大学 信息科学与工程学院, 沈阳 110004)

摘要: 在论述顾客满意度模型是具有二分类因变量的非线性模型的基础上, 提出一种基于 Logistic 回归的联立方程模型. 首先采用因子分析法提取影响满意度和忠诚度的潜在变量; 然后对它们满意与否、忠诚与否进行 Logistic 回归, 建立一个递归联立方程模型; 最后结合社区满意度实例进行了实证研究.

关键词: Logistic 回归; 联立方程; 因子分析; 社区满意度

中图分类号: N945 **文献标识码:** A

Community satisfaction model based on Logistic regression

GENG Jinhua¹, GAO Qisheng¹, REN Jingxi¹, ZHANG Siying^{1,2}

(1. Complexity Science Institute, Qingdao University, Qingdao 266071, China; 2. College of Information Science and Engineering, Northeastern University, Shenyang 110004, China. Correspondent: GAO Qisheng, E-mail: gaoqisheng@qdu.edu.cn)

Abstract: Based on the issue that customer satisfaction model is a nonlinear one with dichotomous dependent variable, a kind of Logistic regression simultaneous equations model is proposed. Latent factors influencing satisfaction and loyalty are extracted by factor analysis. Logistic regression analysis for the satisfaction and loyalty of these latent factors is done, and a recursive simultaneous equation is established. Finally, an example about community satisfaction validates the model.

Key words: Logistic regression; Simultaneous equations; Factor analysis; Community satisfaction

1 引言

产品质量“始于识别顾客的需求, 终于满足顾客的需求”, 顾客满意是当前国内外市场竞争的焦点. 自从美国学者 Cardozo 在 1965 年提出客户满意的概念以来, 已有大量的文献对顾客满意的含义、测评方法以及分析模型进行了研究. 本文旨在研究满意度模型. 考虑到顾客满意度涉及多个影响因素和观测因素, 因此多元统计技术在顾客满意度模型研究中的应用最为广泛. 用于分析顾客满意度数据的多元统计技术包括 3 大类: 关系模型、相关模型和混合模型^[1]. 关系模型根据结果变量的特性和数量又可分为多重回归分析、逻辑回归分析和典型相关分析^[2]等方法. 用于满意度数据分析的相关模型是维度简化技术, 可用来解决多重共线问题. 其中因子分析和主成分分析最为常用. 另外还有几种常用的混

合模型, 如基于主成分回归和偏最小二乘回归的顾客满意度模型^[3-6]以及含潜在变量的结构方程模型^[7-10]等.

以上都是非常通用的线性统计建模技术, 广泛应用于多个学科领域的统计分析. 而对于顾客满意度模型方面的非线性模型, 目前国内外还不多见. 然而大量研究表明: 许多产品或服务的质量特性与顾客满意度之间存在着非线性关系. 在卡诺博士提出的 KANO 模型中, 将顾客的需求分为 3 种: 基本型、期望型和兴奋型, 如图 1 所示. 基本型需求是指顾客认为产品或服务应具有的基本特性, 如果没得到满足, 顾客就会很不满意; 相反, 当完全满足时, 顾客也不会表现出特别满意. 而在市场上顾客经常谈论的是期望型需求, 期望型需求在产品或服务中实现的越多, 顾客就越满意. 兴奋型需求是指令顾客意想不到的产品或服务的特性, 如果产品没有提供这类需

收稿日期: 2006-08-05; 修回日期: 2006-12-31.

基金项目: 国家自然科学基金项目(70571041); 山东省“泰山学者”建设工程项目; 高校博士点专项基金项目(20051065002).

作者简介: 耿金花(1979—), 女, 山西平遥人, 讲师, 博士, 从事复杂系统、系统工程等研究; 高齐圣(1966—), 男, 山东昌乐人, 教授, 博士生导师, 从事复杂系统、系统工程等研究.

求,顾客不会不满意;相反,当产品提供了这类需求时,顾客对产品或服务会表现出非常满意。

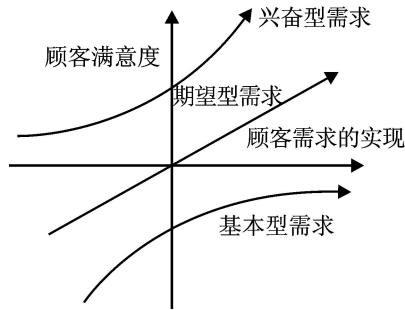


图 1 顾客需求的 KANO 模型

从图 1 可知,这三类需求中,期望型需求与顾客满意度之间呈线性正相关关系,这种关系提供了目前各种顾客满意度测评方法和模型的理论基础。而基本型需求和兴奋型需求与顾客满意度之间则为非线性关系,故使用线性联立方程模型不能给出令人信服的解释。因此有人使用神经网络进行顾客满意度数据的分析^[11]。

表征顾客满意的调查数据往往是二分变量:满意与不满意;表征顾客忠诚的调查数据往往也是二分变量:购买与不购买。对于二分类因变量,线性回归模型已不再适用^[12],应采用非线性模型进行分析。

本文首先对 n 个外生显变量进行因子分析,提取出 m 个外生潜变量,以顾客满意与否和忠诚度与否为因变量,针对顾客满意度模型是二分类因变量的非线性模型,建立潜在变量之间的 Logistic 联立方程模型,如图 2 所示。

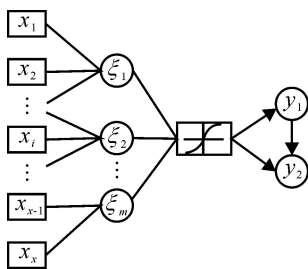


图 2 模型概括

2 含潜在变量的 Logistic 联立方程模型

2.1 因子分析法提取潜在变量

影响满意度的因素很多,由于影响因素的本质特征,它们之间往往存在着某种程度上近似于线性关系的内在联系——多重共线性。当共线性非常明显时,如果还强行实施最小二乘回归,往往会导致很多问题^[12]。因子分析是解决多重共线问题的常用方法之一,它是一种多变量简化技术,可在不丢掉主要信息的前提下避开变量间多重共线性问题。

另外,许多变量往往无法直接观测,需通过一系列观测变量间接反映。通过因子分析,可将这些变量间潜在的结构推导出来,

$$x = x + x. \quad (1)$$

其中: x R 为显变量,即观测变量; R^m 为提取出来的潜在变量(m 的确定一般以主成分累积贡献率达到 80% 以上为准则,也可用特征根大于 1 为标准,通常将两者结合起来综合确定合适的数量); x

$R^{n \times m}$ 为因子分析中的因子负荷矩阵,表示观测变量与潜在变量之间的关系; x R^n 为观测误差或噪声。式(1)需满足 $E(x/) = x$ 。由式(1)可推导出

$$= x + . \quad (2)$$

其中: 为潜在变量与观测变量之间的多元回归系数矩阵,即因子分析中的因子得分矩阵,是 x 的广义逆矩阵; 为相应的回归残差。式(2)也需满足 $E(/ x) = x$ 。通过因子分析,可得 x 和 ,从而将潜在变量用观测变量表示出来。

2.2 Logistic 联立方程模型

由于常规最小二乘法对二分类因变量分析的不适宜性,本文建议使用非线性函数。一个自然的选择便是值域在 $(0, 1)$ 之间有 S 形的曲线,这里采用 Logistic 分布。因为 Logistic 回归模型不仅可解决二分类因变量问题,而且可转换为线性关系,且模型中不存在线性回归模型中的误差项^[12],所以便于处理。

假定顾客不满意取值为 1,满意为 0,有 m 个因素影响顾客满意与否。共有 n 个样本, p_i 为第 i 个顾客不满意的概率,即

$$p_i = \frac{1}{1 + e^{-(\sum_j \beta_{ij} x_{ij})}} = \frac{e^{+\sum_j \beta_{ij} x_{ij}}}{1 + e^{+\sum_j \beta_{ij} x_{ij}}} \quad (3)$$

其中: 1 j $m, 1$ i n, β_{ij} 为 j 个影响因素的第 i 个样本值。 $1 - p_i = \frac{1}{1 + e^{+\sum_j \beta_{ij} x_{ij}}}$ 为该顾客满意的概率,顾客不满意的

发生比为 $\frac{p_i}{1 - p_i} = e^{+\sum_j \beta_{ij} x_{ij}}$ 。这个比值一定为正,没有上界,并且取对数之后为线性函数,即

$$\ln\left(\frac{p_i}{1 - p_i}\right) = + \sum_j \beta_{ij} x_{ij} \quad (4)$$

在顾客满意度和忠诚度数据分析中,选择满意与否和忠诚与否两个二分类因变量,且它们之间有一定相关关系。文献[13,14]认为满意与忠诚之间存在着正向关系,满意是忠诚的必要条件。文献[15]则提出忠诚度与满意度之间为非线性关系。国内也有不少学者研究顾客满意度与忠诚度之间的关系,如文献[16]。因此,要构造的模型不能是单一方程模型,而应是一个联立方程模型。一般而言,满意与否

对于忠诚与否有一定的影响,而忠诚与否对于满意与否则没有影响,因此该联立方程模型应是递归模型.令顾客不满意对数发生比为 $\ln(\frac{p_i}{1-p_i}) = \eta_1$,顾客不忠诚对数发生比为 $\ln(\frac{p_i}{1-p_i}) = \eta_2$.以 η_1 和 η_2 为模型的因变量,可将 Logistic 模型转化为线性模型,其中有 m 个影响它们的潜在变量,而且 η_2 也受到 η_1 的影响.则模型如下:

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2m} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ \dots \\ m \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} \quad (5)$$

3 实例研究

3.1 问题描述

居民对社区的总体满意度是一项综合性指标,它是人们长期以来对社区各种微观感受的一种累计效应^[17],存在众多影响因素.而社区忠诚与否是指居民如果有条件,是否愿意在该社区居住,是否愿意推荐该社区给自己的亲朋好友,反映了居民对于社区的情感和归属感.忠诚有两个主要的维度:情感的和理性的^[1].其中情感的是主观偏好的体现,而理性的则是对社区条件的满意程度.可见,居民对社区的满意度将影响忠诚度.

本文将影响居民对社区满意与否的诸多因素归纳为 16 个方面: x_1 为居民委员会工作, x_2 为房屋建筑质量, x_3 为物业管理, x_4 为便民服务, x_5 为环境卫生, x_6 为绿化情况, x_7 为治安情况, x_8 为文化娱乐活动情况, x_9 为配套设施, x_{10} 为附近交通便利程度, x_{11} 为外出就餐便利程度, x_{12} 为附近购物便利程度, x_{13} 为就医便利程度, x_{14} 为托儿便利程度, x_{15} 为孩子上学便利程度, x_{16} 为居民在社区的居住时间.另外选择两个内生变量: y_1 为居民对社区满意与否, y_2 为居民对社区忠诚与否.通过青岛市市情民意调查中心对若干社区的 300 户居民进行问卷调查,得到相关数据.

3.2 模型建立

3.2.1 潜在变量的提取

经检验,影响社区居民满意与否和忠诚与否的 16 个自变量存在多重共线性,不能直接参与回归分析.经因子分析,根据实际情况提取出 4 个公因子,共解释总变异的 82.134%.因子分析的结果如表 1 所示,该结果直观地反映了各潜在变量是由哪些显著变量提供信息以及各自的回归系数.

表 1 因子分析结果

	日常生活	建设管理	服务休闲	社区情感
居委会工作	-0.145	0.043	-0.027	0.778
建筑质量	0.045	0.879	-0.054	-0.032
物业管理	0.017	0.898	-0.157	0.006
便民服务	-0.033	0.257	0.790	-0.022
环境卫生	0.066	0.959	-0.130	-0.023
绿化	0.019	0.912	-0.126	0.004
治安	0.047	0.944	-0.129	-0.066
文化娱乐	-0.006	0.225	0.904	-0.010
配套设施	0.045	0.165	0.879	0.045
交通状况	0.969	-0.041	0.000	0.022
就餐	0.931	-0.019	-0.034	0.088
购物	0.915	0.003	0.013	-0.020
就医	0.925	-0.038	0.051	-0.002
托儿	0.904	-0.045	-0.016	0.050
小学	0.930	-0.033	-0.002	0.074
居住时间	-0.104	0.100	-0.001	0.778

3.2.2 Logistic 联立方程模型

以居民对社区不满意的对数发生比 η_1 和不忠诚对数发生比 η_2 为因变量,4 个影响它们的潜在变量 η_1, η_2, η_3 和 η_4 为自变量,建立如下 Logistic 联立方程模型(注意 η_1 对 η_2 有影响):

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 0.794 & 1.929 & 0.529 & 0 \\ 0 & 0 & 0 & 0.859 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0.356 & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} \quad (6)$$

3.3 模型检验和参数检验

经检验,第 1 个方程预测敏感度为 81.9%,指定度为 78.6%,总的正确率为 80.3%.选择进入方程的 3 个变量都是统计显著的.第 2 个方程预测敏感度为 62.4%,指定度为 76.6%,总的正确率为 67.7%.选择进入方程的变量也都是统计显著的.

3.4 结果解释

对 Logistic 回归系数进行解释时,很难具体把握以对数单位测量的作用幅度,所以通常转换为对应的发生比

$$\frac{p}{1-p} = e^{e^{1x_1} e^{2x_2} \dots e^{nx_n}}$$

式(6)的模型可改写为

$$\begin{cases} e^{\eta_1} = e^{0.974_1} e^{1.929_2} e^{0.529_3} \\ e^{\eta_2} = e^{0.859_4} e^{0.356_1} \end{cases} \quad (7)$$

当 η_1 增加一个单位时,不满意的发生比率增加

$e^{0.974} = 2.65$ 倍; $_2$ 增加一个单位时, 不满意的发生比率增加 6.88 倍; $_3$ 增加一个单位时, 不满意的发生比率增加 1.70 倍; $_4$ 增加一个单位时, 不忠诚的发生比率增加 2.36 倍; 不满意发生比增加一个单位时, 不忠诚发生比率增加 1.43 倍。

4 结 语

本文首先对影响社区满意与否和忠诚与否的 16 个因素进行因子分析, 提取出 4 个潜在变量; 然后建立了它们与居民对社区满意与否和忠诚与否的 Logistic 联立方程模型, 并对结果作了检验和解释。

根据所得模型可知, 潜在外生变量社区管理和建设对于社区居民满意与否的影响最大, 而它主要包括居民对建筑质量、物业管理、环境卫生、绿化以及治安方面的满意程度。潜在外生变量社区情感对社区忠诚与否有较大影响, 它包括居民在社区的居住时间和对居委会工作的满意程度。社区管理者可通过对这几个主要方面的改善获得最大程度的居民满意度和忠诚度。

参考文献 (References)

- [1] Derek Allen, Tanniru Tao. Analysis of customer satisfaction data [M]. Dalian: Dongbei University of Finance and Economics Press, 2005.
- [2] Mohammad R Azaranga, Graciela Gonzalez, Lawrie Reavill. An empirical investigation of the relationship between quality improvement techniques and performance — A mexican case [J]. J of Quality Management, 1998, 3(2): 265-292.
- [3] Lars Nilsson, Michael D Johnson, Anders Gustafsson. The impact of quality practices on customer satisfaction and business results: Product versus service organizations[J]. J of Quality Management, 2001, 6(1): 5-27.
- [4] 林盛, 刘金兰, 韩文秀. 基于 PLS-结构方程的顾客满意度评价方法[J]. 系统工程学报, 2005, 20(6): 653-656. (Lin Sheng, Liu Jin-lan, Han Wen-xiu. Method on customer satisfaction assessment based on the partial least square for structural equation [J]. J of System Engineering, 2005, 20(6): 653-656.)
- [5] 张新安, 田澎. OLS 与 PLS 方法在顾客满意度建模中的模拟比较[J]. 生产力研究, 2004, 19(6): 81-83. (Zhang Xin-an, Tian Peng. The simulation compare between OLS and PLS method on customer satisfaction modeling [J]. Productivity Research, 2004, 19(6): 81-83.)
- [6] 张新安, 田澎. 顾客满意度线性建模中多重共线性处理方法的模拟研究[J]. 工业工程与管理, 2004, 11(4): 73-77. (Zhang Xin-an, Tian Peng. A simulation study on the methods coping with multi-collinearity in customer satisfaction modeling [J]. Industrial Engineering and Management, 2004, 11(4): 73-77.)
- [7] Robin L Snipes, Sharon L Oswald, Michael La Tour, et al. The effects of specific job satisfaction facets on customer perceptions of service quality: An employee-level analysis [J]. J of Business Research, 2005, 58(10): 1330-1339.
- [8] Tae Hyup Roh, Cheol Kyung Ahn, Ingoo Han. The priority factor model for customer relationship management system success [J]. Expert Systems with Applications, 2005, 28(4): 641-654.
- [9] Chinho Lin, Wing S Chow, Christian N Madu, et al. A structural equation model of supply chain quality management and organizational performance [J]. Int J of Production Economics, 2005, 96(3): 355-365.
- [10] 廖颖林. 结构方程模型及其在顾客满意度研究中的应用 [J]. 统计与决策, 2005, 21(17): 24-26. (Liao Ying-lin. Structural equation model and it's application on customer satisfaction research [J]. Statistics and Decision, 2005, 21(17): 24-26.)
- [11] 曾凤章, 王元华. 神经网络在顾客满意度测评中的应用 [J]. 北京理工大学学报, 2005, 7(1): 45-47. (Zeng Feng-zhang, Wang Yuan-hua. The application of NNT in measuring customer satisfaction degree [J]. J of Beijing Institute of Technology, 2005, 7(1): 45-47.)
- [12] 王济川, 郭志刚. Logistic 回归模型——方法与应用 [M]. 北京: 高等教育出版社, 2001. (Wang Ji-chuan, Guo Zhi-gang. Logistic regression models: Methods and application [M]. Beijing: Higher Education Press, 2001.)
- [13] Shoemaker S, Lewis R C. Customer loyalty: The future of hospitality marketing [J]. Hospitality Management, 1999, 18(4): 345-370.
- [14] Cronin J J, Taylor S A. Measuring service quality: A reexamination and extension [J]. J of Marketing, 1992, 56(3): 55-68.
- [15] Oliver T A, Oliver R L, MacMillan I. A catastrophe model for developing service satisfaction strategies [J]. J of Marketing, 1992, 56(3): 83-95.
- [16] 张新安, 田澎, 张烈平. 顾客满意度测评模型 [J]. 系统工程理论方法与应用, 2002, 11(3): 248-252. (Zhang Xin-an, Tian Peng, Zhang Lie-ping. Study on the model of customer satisfaction measurement [J]. System Engineering Theory Methodology Applications, 2002, 11(3): 248-252.)
- [17] 单菁菁. 社区情感与社区建设 [M]. 北京: 社会科学文献出版社, 2005. (Shan Jing-jing. The community attachment and community construction [M]. Beijing: Social Sciences Academic Press, 2005.)