

## 基于次序 Logistic 模型的广西大学生消费水平影响因素分析\*

# Analysis on Consumption Influence Factors of Students from Guangxi University based on Order Logistic Regression Model

陈淑兰, 闫莉

CHEN Shu-lan, YAN Li

(广西大学数学与信息科学学院, 广西南宁 530004)

(School of Mathematics and Information Science, Guangxi University, Nanning, Guangxi, 530004, China)

**摘要:** 采取随机抽样的调查方法对广西 898 名在校大学生月均消费水平进行调查, 并用次序 Logistic 模型对调查结果进行分析. 结果, 以户籍、饮食支出、服饰化妆、抽烟喝酒、娱乐消费、恋爱、话费、兼职收入及家庭收入等因素建立的模型, 回归系数的显著性概率都小 0.05, 通过了 Wald 检验, 且 OR 值均大于 1. 表明 Logistic 模型对上述因素的分析具有显著的统计意义, 即影响大学生月均消费的主要因素即为上述 9 个, 而且在其他因素相同情况下, 各项消费支出越高, 其月均消费就越高, 各项收入越高, 其月均消费也越高. 建议企业从这 9 个方面制定相应的有效营销策略来开发大学生消费市场.

**关键词:** 次序 Logistic 回归 消费水平 影响因素 检验

**中图分类号:** O212.2 **文献标识码:** A **文章编号:** 1002-7378(2011)03-0186-04

**Abstract:** In this article, random sampling method is adopted to investigate the monthly average consumption of students in Guangxi University and about 900 students are randomly sampled. The results from consumption surveys of college students are statistically analyzed by order Logistic model. A reasonable mathematical model is established on the basis of some specific factors. The probability of the significant regression coefficients are all less than 0.05 and pass the Wald test with OR value over 1, which indicates that the model is obviously statistical significance. Therefore, the investigated factors including census register, diet expenditure, dress and make up, smoking and drinking, entertainment, love affair, telephone charges, part time income and family income are main factors that affect the monthly average consumption of students. With the similar other indexes, the higher consumption in these 9 key factors will increase the monthly average consumption. Based on these results, some advises for enterprise developing corresponding strategies for college students market are proposed.

**Key words:** order Logistic regression, consumption level, influence factors, testing

随着社会的发展, 各高等院校招生规模不断扩大, 大学生人数越来越多. 大学生群体追求时尚、消

费意念强烈, 企业有必要关注他们的消费状况, 把握他们的消费特征. 关于大学生消费因素的研究已有不少, 但是大多是从营销的角度分析, 或者是简单的统计分析(频数分析, 条形图, 方差分析等), 这种简单的分析虽然能给出大致的主要影响因素, 但是不能给出主要因素确切的影响程度. 本文应用统计的

收稿日期: 2011-03-01

作者简介: 陈淑兰(1986-), 女, 硕士研究生, 主要从事广义线性模型研究.

\* 国家自然科学基金项目(11061002), 广西自然科学基金项目(0832108)资助.

知识,通过对广西大学生月均消费的影响因素建立次序 Logistic 模型,来具体分析消费因素的影响程度。

## 1 资料来源与研究方法

### 1.1 资料来源

采用随机抽样调查,无记名问卷调查方式,抽取 898 名在校广西大学生,对可能影响其消费水平的性别、年级、户籍、各项支出及收入等 13 个因素进行调查。调查时间为 1 个星期,调查地点为图书馆门前。由于图书馆的学生涉及各学院、各年级,故调查对象达到随机普遍要求。另外,问卷是即填即收,所获的数据比较真实。

### 1.2 研究方法

以月均消费为因变量(用  $Y$  表示),以月均饮食支出、学习花费、聚餐消费等 13 个影响因素为自变量,建立次序 Logistic 回归模型。先对 13 个影响因素进行共线性诊断,再采用极大似然估计,统计软件 SPSS17.0 分析变量间的关系并进行拟合,最后对拟合的结果进行检验和修正。

#### 1.2.1 次序 Logistic 模型

次序反应变量通常按 1,2,3 等整数序列编码,有明确的从低到高排序,但相邻各类的差距未知。

次序 Logistic 模型<sup>[1]</sup>为

$$y = \alpha + \sum_{k=1}^m \beta_k x_k + \epsilon, \quad (1)$$

其中, $y$  是观测现象内在趋势,它并不能被直接测量; $\alpha, \beta_k$  是模型需要标定的系数; $x_k$  是模型中自变量; $\epsilon$  是误差项。当实际观测反应变量有  $j$  种类别时( $j=1,2,\dots,J$ ),相应取值为  $y=1,\dots,y=J$ 。则有  $j-1$  个未知分界点  $u_i (i=1,\dots,j-1)$  将各相邻类别分开。即若  $y \leq u_1$ ,则  $y=1$ ;若  $u_1 < y \leq u_2$ ,则  $y=2$ ;... $u_{j-1} < y$ ,则  $y=j-1$ 。其中  $u_1 < u_2 < \dots < u_{j-1}$ 。累积概率由以下公式预测:

$$p(y^* \leq j | x) = p(y \leq u_j | x) = \frac{\exp\{u_j - (\alpha + \sum_{k=1}^m \beta_k x_k)\}}{1 + \exp\{u_j - (\alpha + \sum_{k=1}^m \beta_k x_k)\}}, \quad (2)$$

其中  $p(y \leq u_j | x)$  表示在  $x$  条件下  $y \leq u_j$  发生的概率。

#### 1.2.2 模型的标定

模型影响因素及变量标定如表 1。

表 1 模型影响因素及变量标定

变量	代码	赋值
月均消费	$Y$	1=300 以下;2=300-400;3=400-600; 4=600-800;5=800 以上
饮食支出	$x_1$	1=200 以下;2=200-300;3=300-400; 4=400-500;5=500 以上
学习花费	$x_2$	1=200 以下;2=200-300;3=300-400; 4=400-600;5=600 以上
聚餐消费	$x_3$	1=0;2=0-50;3=50-100;4=100-150; 5=150-200;6=200 以上
服饰化妆	$x_4$	1=基本没有;2=50 以下;3=50-150; 4=150-200;5=250 以上
抽烟喝酒	$x_5$	1=基本没有;2=50 以下;3=50-100; 4=100-200;5=200 以上
休闲娱乐	$x_6$	1=0;2=0-50;3=50-100;4=100 以上
恋爱花费	$x_7$	1=0;2=100 以下;3=100-200; 4=200-400;5=400 以上
话费支出	$x_8$	1=50 以下;2=50-100;3=100-150; 4=150-200;5=200 以上
兼职收入	$x_9$	1=0;2=0-200;3=400-600;4=600-800; 5=800 以上
家庭收入	$x_{10}$	1=500 以下;2=500-1500;3=1500-3000; 4=3000-5000;5=5000 以上
性别	$x_{11}$	1=男;2=女
户籍	$x_{12}$	1=农村;2=城市
年级	$x_{13}$	1=大一;2=大二;3=大三;4=大四; 5=研一;6=研二

## 2 结果分析

### 2.1 调查结果统计

在所调查的 898 名大学生中,76.6% 的学生月均消费在 400 元以上,其中 46.9% 的学生是男生,53.1% 的学生是女生;32.2% 的学生来自城市;63.2% 的学生月均饮食支出在 400 元以上;49.7% 的学生服饰化妆月均在 50 元以上;37.3% 的学生月均娱乐消费在 100 元以上;37.8% 的学生月均兼职收入在 200 元以上;58.9% 的学生家庭收入在 3000 元以上。

### 2.2 影响因素共线性诊断

对饮食支出,学习花费等 13 个自变量进行相关分析,发现其条件指数小于 30,故可以认为其共线性程度较小。此外,模型中不存在异常点及强影响点。

### 2.3 次序 Logistic 模型分析

模型检验结果显示,原假设使回归模型无效<sup>[2]</sup>,所有回归系数都是 0。由模型检验结果可知,模型似然比检验<sup>[3]</sup>  $\chi^2 = 863.448$ ,显著性概率(Sig) = 0.000,拒绝原假设,认为回归模型有效<sup>[4]</sup>。对月均消费建立关于饮食支出等 13 个影响因素的次序 Logistic 模型,得到参数估计和显著性概率。从表 2 可以看出性别、年级、学习花费、聚餐消费这 4 个变

量的显著性概率均大于 0.05,故认为这 4 个变量不显著.故需要对模型进行修正.

表 2 参数估计

项目	B	OR 值	Std. error	Wald Chi-Square	Sig.
月消费=5	-13.774		0.697	390.508	0.000
月消费=4	-11.259		0.627	322.787	0.000
月消费=3	-7.558		0.549	189.796	0.000
月消费=2	-4.463		0.544	67.377	0.000
性别	0.234	0.79	0.203	1.336	0.248
户籍	-0.552	1.74	0.182	9.221	0.002
年级	-0.026	1.03	0.059	0.195	0.659
饮食支出	-1.354	3.87	0.091	219.298	0.000
学习花费	-0.077	1.08	0.060	1.613	0.204
聚餐消费	0.034	1.03	0.033	1.017	0.313
服饰化妆	-0.331	1.40	0.091	13.345	0.000
抽烟喝酒	-0.308	1.36	0.138	5.003	0.025
娱乐消费	-0.626	1.87	0.134	21.748	0.000
恋爱	-0.289	1.34	0.093	9.681	0.002
话费	-0.671	1.96	0.130	26.701	0.000
兼职收入	-0.126	1.13	0.034	13.704	0.000
家庭收入	-0.606	1.83	0.081	56.455	0.000

2.4 修正后的模型回归分析

剔除性别、年级、学习花费、聚餐消费这 4 个变量,对月均消费建立关于户籍、饮食支出等 9 个影响因素的次序 Logistic 模型.得到参数估计及 OR 值发生比率见表 3.从表 3 各个参数估计的  $\chi^2$  检验和显著性概率可看出,这 9 个变量均通过检验.

这时,以  $p_1, p_2, p_3, p_4, p_5$  分别表示月均消费在 300 以下,300 ~ 400,400 ~ 600,600 ~ 800,800 以上的概率,则由表 3 可以得到次序 Logistic 回归方程<sup>[5]</sup>为:

$$\ln \frac{p_1 + p_2 + p_3 + p_4}{p_5} = 14.083 - 0.519x_{12} - 1.375x_1 - 0.282x_4 - 0.342x_5 - 0.670x_6 - 0.308x_7 - 0.665x_8 - 0.120x_9 - 0.619x_{10},$$

$$\ln \frac{p_1 + p_2 + p_3}{p_4 + p_5} = 11.564 - 0.519x_{12} - 1.375x_1 - 0.282x_4 - 0.342x_5 - 0.670x_6 - 0.308x_7 - 0.665x_8 - 0.120x_9 - 0.619x_{10},$$

$$\ln \frac{p_1 + p_2}{p_3 + p_4 + p_5} = 7.877 - 0.519x_{12} - 1.375x_1 - 0.282x_4 - 0.342x_5 - 0.670x_6 - 0.308x_7 - 0.665x_8 - 0.120x_9 - 0.619x_{10},$$

$$\ln \frac{p_1}{p_2 + p_3 + p_4 + p_5} = 4.811 - 0.519x_{12} - 1.375x_1 - 0.282x_4 - 0.342x_5 - 0.670x_6 - 0.308x_7 - 0.665x_8 - 0.120x_9 - 0.619x_{10}.$$

此时模型的显著性概率(Sig)=0.000,模型有效.再由表 3 知回归系数的显著性概率都小 0.05,通过了 Wald 检验,且 OR 值均大于 1 故模型具有显著的统计意义.所以由模型分析结果知,影响大学生月均消费的主要因素有户籍、饮食支出、服饰化妆、抽烟喝酒、娱乐消费、恋爱、话费、兼职收入和家庭收入 9 大因素.且其他指标相同情况下,各项消费支出越高,其月均消费就越高;各项收入越高,其月均消费也越高.

表 3 参数估计

项目	B	OR 值	Std. error	Wald Chi-Square	Sig.
月消费=5	-14.083		0.611	531.814	0.000
月消费=4	-11.564		0.532	472.985	0.000
月消费=3	-7.877		0.433	330.887	0.000
月消费=2	-4.811		0.418	132.529	0.000
户籍	-0.519	1.68	0.180	8.356	0.004
饮食支出	-1.375	3.96	0.090	234.169	0.000
服饰化妆	-0.282	1.33	0.068	17.242	0.000
抽烟喝酒	-0.342	1.41	0.135	6.462	0.011
娱乐消费	-0.670	1.95	0.132	25.961	0.000
恋爱	-0.308	1.36	0.092	11.118	0.001
话费	-0.665	1.94	0.130	26.289	0.000
兼职收入	-0.120	1.13	0.033	12.927	0.000
家庭收入	-0.619	1.86	0.080	59.714	0.000

2.5 因素影响程度分析

2.5.1 户籍因素

户籍因素的 OR 值是 1.68.在其他消费指标相同情况下,来自城镇的学生相对月均消费比  $k = \frac{p(Y \geq j)}{p(Y \leq j)}$ ,其中  $j = 300, 400, 600, 800$ .是来自农村学生的 1.68 倍.例如某位来自农村的学生,其他消费支出均处于最低,收入也处于最低,即他的各指标为  $x_{12} = 1, x_1 = 1, x_4 = 1, x_5 = 1, x_6 = 1, x_7 = 1, x_8 = 1, x_9 = 1, x_{10} = 1$ .根据模型,他月消费小于 300 元的概率是 0.478,月消费小于 400 元的概率是 0.953.若这位学生来自城镇,其他指标均不变,即  $x_{12} = 2, x_1 = 1, x_4 = 1, x_5 = 1, x_6 = 1, x_7 = 1, x_8 = 1, x_9 = 1, x_{10} = 1$ .则他月消费小于 300 元的概率是 0.353,月消费小于 400 元的概率是 0.921,概率分别减少了 0.125 和 0.032.说明相同条件下,户籍是城镇的大学生消费水平高于户籍是农村的大学生.这是因为城乡的环境不同,消费观念存在差异.一般来说,城镇学生的消费观超前于农村学生的消费观.

2.5.2 饮食支出因素

饮食支出因素的 OR 值是 3.96.某位大学生饮食支出  $x_1$  取 3,其他指标均取 1.根据模型,他月均

消费小于 300 元的概率是 0.055, 月均消费小于 400 元的概率是 0.556. 月消费小于 300 元和小于 400 元的概率分别减少了 0.423 和 0.397. 另外, 由调查结果知, 76.6% 的学生月均饮食支出在 400 元以上. 说明饮食支出占总消费的较大比例, 因此大学生这一群体的恩格尔系数较高.

### 2.5.3 服饰化妆及抽烟喝酒因素

服饰化妆因素的 OR 值是 1.33, 抽烟喝酒因素的 OR 值是 1.41. 这两项消费和性别有关, 服饰化妆的女生居多, 抽烟喝酒以男生为主. 两项的 OR 值相差不多, 即这两个因素对月消费的影响程度大致相同. 可能被调查对象中的男生和女生数目大致相同是一个原因.

### 2.5.4 娱乐因素

娱乐消费因素的 OR 值是 1.95. 即学生在校的娱乐消费每增加一个级别, 其月均消费比  $k$  就增加到原来的 1.95 倍. 这是因为大学生的娱乐生活越来越丰富, 他们在这方面的消费支出比例也越来越大.

### 2.5.5 恋爱因素

恋爱消费因素的 OR 值是 1.36. 恋爱消费每增加一个级别, 其月均消费比  $k$  就增加到原来的 1.36 倍. 这是因为大学校园中, 热恋的情侣随处可见, 由恋爱产生的一系列花费势必占据月消费的一定比例.

### 2.5.6 话费因素

话费因素的 OR 值是 1.94. 话费支出每增加一个级别, 其月均消费比  $k$  就增加到原来的 1.94 倍. 这是因为大学生社会交际逐步广泛, 由社交产生的通讯费也会逐步增加.

### 2.5.7 兼职收入及家庭收入因素

兼职收入和家庭收入的 OR 值分别为 1.13 和 1.86. 兼职收入的 OR 值小于家庭收入的 OR 值, 是因为校园中只有部分大学生参加兼职工作, 主要的消费支出还是靠家里提供. 但是不管是兼职收入, 还是家庭收入, 总是收入越高, 月消费比就越高.

## 3 结束语

大学生消费市场是一个规模不断增大、非常有前景的市场, 其市场价值巨大, 回报率极高, 企业不应该忽视它. 由以上的影响因素分析, 建议企业对大学生开发相应饮食、服饰化妆、烟酒、娱乐、通讯以及家教辅导班市场, 制定有效的营销策略<sup>[6,7]</sup>, 来满足这个这个特殊群体的消费需要.

(1) 饮食. 对于多数大学生, 餐饮消费是主消费

因素. 大学生餐饮不再像是中学时代那样, 被动地在指定饭堂就餐. 他们就餐自主选择权较大, 传统的单一化模式将被淘汰. 建议大学校园内或者附近餐饮店, 根据学生的个性化需求, 推崇多元化的经营模式.

(2) 服饰化妆. 大学生从这个年龄段开始已经很注意自己的仪容仪表, 尤其是女大学生, 更是喜欢打扮. 他们追求时尚, 喜欢新潮, 然而他们的经济能力有限. 建议企业以款式时尚、新潮, 价格实惠为主打去吸引大学生这个巨大的群体.

(3) 娱乐. 大学生的娱乐消费将趋近多元化, 他们越来越不满足看电视剧、逛街和进影院等传统休闲娱乐, 而是追求刺激、出奇、有品位的活动, 蹦极、漂流、潜水、探险等刺激活动, 泡沙龙、旅游观光、体育健身也成为主流. 建议企业从这些方面为大学生开放相应的市场, 制定相应的营销策略.

(4) 通讯. 大学生越来越注重交际, 由社会交往产生的通讯费势必会增加. 此外, 有相当部分的大学生都处于恋爱时期. 所以, 为大学生制定一个经济的交际通讯套餐, 情侣套餐, 提供一个可靠实惠的交友平台, 将会受到他们青睐.

(5) 兼职. 大学生消费超前, 有一定的工作能力, 受时间的约束, 却渴望独立. 建议企业为其提供一些兼职岗位, 如家教, 辅导及其他适合大学生兼职的岗位. 这样既可以为企业带来丰厚利润, 促进经济发展. 又可以给大学生提供接触社会的平台, 增加大学生的收入, 促进消费.

### 参考文献:

- [1] 李世民, 孙明玲, 关宏志. 基于累积 Logistic 模型的交通事故严重程度预测模型[J]. 交通与安全, 2009, 190/192:170-171.
- [2] 何晓群, 刘文卿. 应用回归分析(二)[M]. 北京: 中国人民大学出版社, 2007:241-266.
- [3] 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, 2000:237-262.
- [4] 李艳华, 付丽丽, 邸海峰. 累积 Logistic 回归在后发企业技术能力提升研究中的应用[J]. 数学的实践与认识, 2009, 39(10):107-108.
- [5] 牛田华, 孟庆跃, 宋涛, 等. 农村老年人心理健康影响因素的累积 Logistic 回归分析[J]. 中国健康心理学杂志, 2009, 17(11):1324-1325.
- [6] 赵小剑. 大学生消费现状及发展趋势的社会学研究[J]. 网络财富·理论探讨, 2009(2):184-185.
- [7] 周建. 当代大学生消费市场的开发及营销策略[J]. 商场现代化, 2008, 543:168-169.

(责任编辑:尹 闯)