DOI:10.19789/j.1004-9398.2017.05.002

对在校大学生对于各种饮品喜爱程度的统计分析

杨姣姣 周 清
(北京邮电大学理学院,北京 100876)

摘 要

本文主要利用 SPSS 软件和 R 语言对大学生对于饮品的喜爱程度进行因子分析,应用多分类 Logistic 回归分析对具体的数据进行了分析,得出大学生的饮食习惯总体处于比较健康的水平.

关键词: 因子分析,对应分析,多分类 Logistic 回归分析.
中图分类号: O01, O213.9

1 因子分析

应用 SPSS 对数据进行因子分析[1],提取碎石图分析结果如下:

表 1.1 KMO 和巴特利特检验

KMO 取样适切性量数		0.729
巴特利特球形度检验	近似卡方	150.970
	自由度	21
	显著性	0.000

由表 1.1 的数据,且 sig 0.0 < 0.05,应该拒绝各变量之间互相独立的假设,得出各个变量之间存在较强的相关性; 且 KMO 为 0.729,大于 0.7,即各变量之间的信息重叠度较高,适合做因子分析.

由表 1.2 可以看到,前三个成分的累计方差贡献率为 63.906%,旋转后的前三个因子的贡献率仍然是 63.906%,并且保持由大到小的顺序.

碎石图[2]将各因子按照特征根由大到小的顺序排列,从而可以从直观上看出各因子的重要性. 由图 1 可以看到选取前两个较合适,但是基于前两个成分的累计方差贡献率刚到 50%,而第三个成分的贡献率和第二个相当,且特征根接近于 1. 因此,我们选取前三个成分进行分析.

表 1.2 总方差解释

成分	初始特征值			提取载荷平方和			旋转载荷平方和		
	总计	方差百分比	累积/%	总计	方差百分比	累积/%	总计	方差百分比	累积/%
1	2.443	34.907	34.907	2.443	34.907	34.907	1.685	24.078	24.078
2	1.055	15.067	49.973	1.055	15.067	49.973	1.439	20.555	44.633
3	0.975	13.933	63.906	0.975	13.933	63.906	1.349	19.273	63.906
4	0.741	10.584	74.491						
5	0.702	10.031	84.522						
6	0.653	9.330	93.851						
7	0.430	6.149	100.000						

提取方法: 主成分分析法。

行方差最大旋转之后,得表 1.3 和 1.4. 设碳酸饮料为 X_1,乳饮料为 X_2,瓶装水为 X_3,咖啡饮料为 X_4,果蔬汁为 X_5,功能饮料为 X_6,茶类饮料为 X_7,由旋转成分矩阵得

表 1.3 旋转后的成分矩阵[a]

	成分		
	1	2	3
功能	0.734	0.185	−0.017
果蔬汁	0.301	0.683	0.378
咖啡	0.660	0.165	0.038
瓶装水	−0.011	−0.048	0.876
乳饮料	0.046	0.875	−0.049
碳酸	0.746	−0.045	0.220
茶类	0.251	0.374	0.621

提取方法: 主成分分析法.
旋转方法: 凯撒正态化最大方差法.
a. 旋转在 4 次迭代后已收敛.

收稿日期: 2016-11-29

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

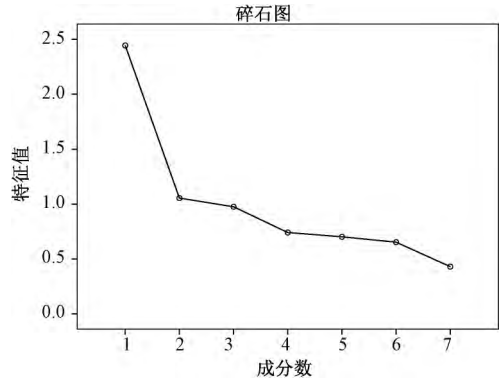


图 1 碎石图

表 1.4 成分得分系数矩阵

	成分		
	1	2	3
功能	0.481	−0.010	−0.164
果蔬汁	−0.003	0.434	0.153
咖啡	0.424	−0.019	−0.102
瓶装水	−0.138	−0.189	0.750
乳饮料	−0.158	0.724	−0.200
碳酸	0.504	−0.245	0.073
茶类	−0.007	0.147	0.420

提取方法: 主成分分析法.
旋转方法: 凯撒正态化最大方差法.

$$X_1 = F_1 * 0.746 + F_2 * (-0.045) + F_3 * 0.220$$

$$X_2 = F_1 * 0.046 + F_2 * (0.875) + F_3 * (-0.049)$$

$$X_3 = F_1 * (-0.011) + F_2 * (-0.048) + F_3 * 0.876$$

$$X_4 = F_1 * 0.660 + F_2 * (0.165) + F_3 * 0.038$$

$$X_5 = F_1 * 0.301 + F_2 * 0.683 + F_3 * 0.378$$

$$X_6 = F_1 * 0.734 + F_2 * 0.185 + F_3 * (-0.017)$$

$$X_7 = F_1 * 0.251 + F_2 * 0.374 + F_3 * 0.621$$

由成分得分系数矩阵得

$$F_1 = 504 * X_1 - 0.158 * X_2 - 0.138 * X_3 + 0.424 * X_4 - 0.003 * X_5 + 0.481 * X_6 - 0.007 * X_7$$

$$F_2 = -0.245 * X_1 + 0.724 * X_2 - 0.189 * X_3 - 0.019 * X_4 + 0.434 * X_5 - 0.010 * X_6 + 0.147 * X_7$$

$$F_3 = 0.037 * X_1 - 0.200 * X_2 + 0.750 * X_3 - 0.102 * X_4 + 0.153 * X_5 - 0.164 * X_6 + 0.420 * X_7$$

综合表 1.3 和 1.4 的数据可得[3]:

(1) 第一公因子在碳酸饮料、咖啡、功能饮料上有较大的载荷. 因为碳酸饮料会造成身体钙的流失,而根据我们查阅相关资料显示咖啡饮料饮用多了会加剧高血压,而功能饮料有可能带来脱水,提高警觉性和失眠概率,使心跳加速,血糖升高等危害,所以我们可以认为成分一主要代表对身体有一定副作用的饮品.

(2) 第二公因子在乳饮料、果蔬汁上较大的载荷. 他们分别为身体的健康发展提供了蛋白质和维生素等营养物质,同时口感也是极佳的,所以我们可以认为成分二主要代表对身体健康发展有益的口感不错的饮品.

(3) 第三公因子在瓶装水、茶饮料上有较大的载荷. 水是身体健康发展不可缺少的物质,而茶饮料有抗氧化,抗衰老等作用,对身体基本没有什么危害,即水和茶饮料均对身体无副作用,但是因为它们总体没有乳饮料和果蔬汁的味道好,所以我们可以认为成分三代表对身体健康发展有益但是口感一般的饮品.

2 对应分析

我们把各个饮料的喜爱程度进行汇总,得如下:

表 2.1 各种饮料的整体喜好程度

	很不喜欢	不喜欢	一般	喜欢	非常喜欢
碳酸	21	39	66	24	5
乳饮料	7	11	50	59	28
瓶装水	2	12	101	28	12
咖啡饮料	8	28	67	41	11
果蔬汁	7	20	53	53	22
功能饮料	14	28	81	27	5
茶饮料	5	8	72	58	12

我们用 R 软件[4]进行因子分析,得到如图 2 的实验结果:

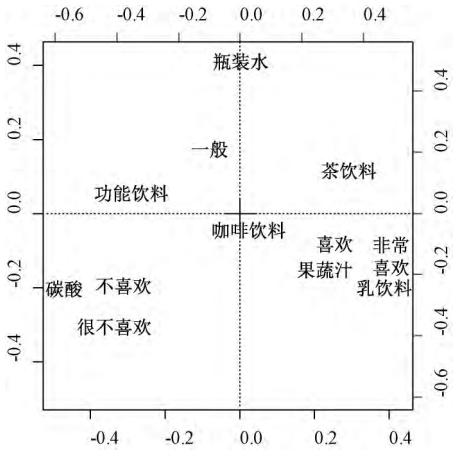


图 2 对应分析图

对应图分析[5]:

由于所有主要信息都在图 2 中反映,各个类别在坐标中的位置和各自之间的距离反映了各个类别之间的关系. 分析如下:

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

根据上图我们发现碳酸饮料和不喜欢、很不喜欢靠得非常近,所以可认为碳酸饮料和不喜欢之间存在关联; 同理,瓶装水、茶饮料、咖啡、功能饮料和一般之间的距离相对较近,而且这四种饮料和别的分类又相对较远,所以可认为这四种饮料和一般之间存在关联; 而果蔬汁、乳饮料则和喜欢,非常喜欢靠得很近,因此认为果蔬汁、乳饮料和喜欢存在关联. 在图形中,相似的类会聚在一起. 靠得很近,因而我们根据两种定性变量(饮品的分类与喜欢程度) 之间的距离,就可以得出两个变量的哪些类相似,从而进行分组.

根据以上解释可将样本点和变量分为三组:

第一组,变量: 碳酸饮料

样品: 不喜欢、很不喜欢

第二组,变量: 茶饮料,咖啡饮料,功能饮料,瓶装水

样品: 一般

第三组,变量: 果蔬汁,乳饮料

样品: 喜欢,非常喜欢

根据分组情况,我们可以看出大家对碳酸饮料基本是不喜欢或者很不喜欢,对于茶饮料、咖啡饮料、功能饮料和瓶装水基本是一般的态度,而对于果蔬汁和乳饮料大家则是喜欢或者很喜欢.

根据这些初步了解,就商场、超市完成了初步的市场定位. 市场,超市在进购饮品时,乳饮料和果蔬汁方面可以占相对较大的比例,而茶饮料、咖啡饮料、功能饮料、瓶装水可以适当的进购,而于碳酸饮料而言,则不宜进购太多,这样既可以让大家身体健康,同时也可以减少不必要的资金消耗.

同时,建议饮料生产商可以发明更多的口味接近果蔬汁、乳饮料,同时对人体健康也有益的饮品.

我们对在校大学生对各种饮料的喜爱程度进行了问卷调查: 1 代表非常不喜欢; 2 代表不喜欢; 3 代表一般; 4 代表喜欢; 5 代表非常喜欢.

性别: 2 代表女生,1 代表男生.

3　多变量 Logistic 回归分析

由于男生、女生对不同饮料的喜爱程度不同,而且对饮料的喜爱程度分类有五类,可以看成多类别数据,为了得到更具体的概率数据,分别从男女生角度对饮料的喜爱程度进行分析,接下来应用多变量 Logistic 回归分析的方法[6] 进行分析.

对功能饮料进行分析,应用 SPSS[7] 进行分析[8]: 分析→回归→多项 Logistic. 得到如下重要表格:

表 3.1　参数估算值

	功能[a]	B	标准误差	瓦尔德	自由度	显著性	Exp(B)	Exp(B) 的 95% 置信区间	
								下限	上限
1	截距	-1.540	0.318	23.451	1	0.000			
	[性别 = 1]	-0.985	0.801	1.514	1	0.219	0.373	0.078	1.793
	[性别 = 2]	0[b]	—	—	0	—	—	—	—
2	截距	-0.934	0.252	13.788	1	0.000			
	[性别 = 1]	-0.493	0.520	0.900	1	0.343	0.611	0.221	1.691
	[性别 = 2]	0[b]	—	—	0	—	—	—	—
4	截距	-1.081	0.265	16.575	1	0.000			
	[性别 = 1]	-0.059	0.485	0.015	1	0.904	0.943	0.364	2.441
	[性别 = 2]	0[b]	—	—	0	—	—	—	—
5	截距	-21.790	0.490	1978.407	1	0.000			
	[性别 = 1]	20.181	0.000	—	1	—	581 357 440.800	581 357 440.800	581 357 440.800
	[性别 = 2]	0[b]	—	—	0	—	—	—	—

a. 参考类别为: ^1.

b. 此参数冗余,因此设置为零.

建立五个 Logit 模型的值,根据上表的统计量可以计算得:

LnP(功能 1/功能 3) = -1.540 - 0.985 = -2.525;

LnP(功能 2/功能 3) = -0.934 - 0.493 = -1.427;

LnP(功能 4/功能 3) = -1.081 - 0.059 = -1.14;

LnP(功能 5/功能 3) = -21.790 + 20.181 = -1.609;

P(功能 1) + P(功能 2) + P(功能 3) + P(功能 4) + P(功能 5) = 1.

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

解上述方程得到:

对功能饮料的喜爱程度为 1 的概率: $P1 = 0.4$;

对功能饮料的喜爱程度为 2 的概率: $P2 = 0.13$;

对功能饮料的喜爱程度为 3 的概率是: $P3 = 0.54$;

对功能饮料的喜爱程度为 4 的概率是: $P4 = 0.17$;

对功能饮料的喜爱程度为 5 的概率是: $P5 = 0.11$

由上述数据可见,对功能饮料的整体喜欢程度为一般的概率远大于其余程度的概率,这和我们之前应用对应分析和聚类分析得到的结果相同. 而且由上述表格的 Exp(B) 值可以看出: 在非常讨厌和不太喜欢方面来看,男生分别是女生的 30%,60%(相对一般喜欢来说); 在比较喜欢方面来看,男女相当; 男生对功能饮料的非常喜欢程度(相对一般喜欢来说) 远远大于女生.

综上所述,基本得出结论: 对功能饮料的喜爱重视程度,男生总体上大于女生. 这也和日常生活中的现象相符合,因为男生整体上喜欢运动,做一些激烈的运动,因此需要补充能量,功能饮料正好具备此作用. 所以可以建议在男寝附近的超市可以适当多地购进功能饮料.

同理对果蔬汁饮料进行分析得到结果是: 女生整体上比男生更喜欢,所以可以建议靠近女寝的超市在果蔬汁的购进上可以适当多些;

对咖啡饮料分析结果是: 女生整体上比男生更喜欢,由此我们可以建议在靠近女寝或者女生比较多的地方,可以适当多的购进咖啡.

对瓶装水分析结果是: 男女生对其的喜欢程度相当,所以建议对瓶装水的购进可以不区分男女的影响.

对乳饮料分析结果是: 女生相对而言更喜欢些,可以建议在女生出现几率大的超市里相对多地购进乳饮料.

对碳酸饮料分析结果是: 男生相对女生会略偏向于喜欢,或者说没有女生讨厌的程度深. 但是因为碳酸饮料对身体的健康影响不好,所以希望超市可以根据公民的身体健康问题,结合销售情况,适当少的购进碳酸饮料.

对茶类分析结果是: 男生相对女生会略偏向于喜欢,或者说没有女生讨厌的程度深. 所以可以在男寝附近或者男生出现多的地方,相对多的购进茶饮料.

4 结 论

大学生对各种饮品的喜爱程度能够从一个侧面反映出大学生的饮食习惯是否健康,同时调查结果也可以为市场,超市提供市场定位. 本文主要叙述用 SPSS 软件和 R 语言对上述饮品的喜爱程度进行因子分析,对应分析,多分类 Logistic 回归分析,其中因子分析主要对其旋转成分矩阵和系数得分矩阵进行了评价和总结; 对应分析对变量和样本进行了综合分析,从而把相似的类聚在了一起; 多分类 Logistic 回归分析则是对大学生对不同的饮料的喜欢程度的具体概率和男女生对不同饮料的喜爱程度的差异进行了分析.

分析结果显示,大学生对乳饮料和果蔬汁饮料的喜爱程度最大,可以得出大家总体的饮食喜好是比较健康的,但是对瓶装水的喜爱程度是一般的水平,然而水是生命必不可少的物质,在这一点上我们建议大家应该引起注意,多喝饮用水.

另外,根据多分类 Logistic 回归分析的结果表明,大家对于碳酸饮料的喜爱程度是一般,而碳酸饮料对身体的危害很大,所以建议大家还是尽量少饮用碳酸饮料为好.

参 考 文 献

[1] 何晓群. 多元统计分析 [M]. 北京: 中国人民大学出版社,2012: 143 – 154,176 – 183.

[2] 何飞. 黄冈在武汉城市圈中的现状及发展对策——基于 SPSS 因子分析方法 [J]. 区域经济,2012,17: 98 – 100.

[3] 高素英,金浩. 因子分析在市场问卷调查分析中的应用 [J]. 河北工业大学学报,2003,5: 44 – 47.

[4] 王斌会. 多元统计分析及 R 语言建模 [M]. 广州: 暨南大学出版社,2011: 236 – 244.

[5] 李映红. 对应分析理论及其在简化抽样调查问卷中的应用 [D]. 吉林: 吉林大学,2004.

[6] 何晓群. 应用回归分析 [M]. 北京: 中国人民大学出版社,2011: 235 – 253.

[7] 张文彤,董伟. SPSS 统计分析高级教程 [M]. 北京: 高等教育出版,2004: 195 – 199.

[8] 张森. Logit 回归模型在医学中的应用 [J]. 高校科技,2013,5: 105.

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

The Statistical Analysis of the College Students´ Liking for a Variety of Drinks

Yang Jiaojiao　Zhou Qing

(School of Science, Beijing University of Posts and Telecommunication, Beijing　100876)

Abstract

In this paper, SPSS software and R language are used to analyze the degree of college students´ liking for drinks. Multivariate Logistic regression analysis is used to analyze the specific data. The results show that the eating habits of college students are in a relatively healthy level.

Key words: factor analysis, correspondence analysis, multivariate Logistic regression analysis.

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net