

基于因子分析与定序Logistic回归的RIO鸡尾酒消费群体分类与满意度分析

王行修*, 徐美萍#

北京工商大学数学与统计学院, 北京

收稿日期: 2023年5月30日; 录用日期: 2023年6月30日; 发布日期: 2023年7月10日

摘要

当前国内低度果味鸡尾酒市场竞争日趋激烈, RIO鸡尾酒作为行业的龙头代表, 有良好的市场口碑和大量消费群体, 但近几年市场表现的活力有所下滑, 同时该品牌主打的“微醺”赛道在经历了近2年的“爆红”之后, 在2022年快速“降温”, 消费投资遇冷, 融资受到一定影响, RIO鸡尾酒需要新的营销策略去处理这些挑战。本文基于所收集消费者的样本数据进行建模, 首先使用因子分析方法对影响购买的十个因素降维分类, 归纳成为影响人群购买行为的三个不同要素的公共因子: 产品质量与品牌特点因子、产品宣传与方便性因子及产品服务与物流因子。然后基于这三个公共因子对现有消费者进行K-means聚类, 得到三类消费人群: 注重产品质量的参与人群、注重品牌与宣传的热爱人群以及注重方便与售后的忠实人群。最后使用定序Logistic回归模型分析影响消费者购买RIO鸡尾酒时的因素, 发现影响消费者购买的三个显著因素: 价格、便携性以及品牌口碑, 以期为RIO鸡尾酒精准化定位与差异化营销提供一定借鉴意义。

关键词

RIO鸡尾酒, 因子分析, K-Means聚类, 定序Logistic回归

Classification and Satisfaction Analysis of RIO Cocktail Consumer Groups Based on Factor Analysis and Ordered Logistic Regression

Xingxiu Wang*, Meiping Xu#

School of Mathematics and Statistics, Beijing Technology and Business University, Beijing

*第一作者。

#通讯作者。

文章引用: 王行修, 徐美萍. 基于因子分析与定序Logistic回归的RIO鸡尾酒消费群体分类与满意度分析[J]. 数据挖掘, 2023, 13(3): 235-243. DOI: 10.12677/hjdm.2023.133024

Abstract

At present, the competition in the domestic market related to low fruity cocktail is becoming increasingly fierce. RIO cocktail, as the leading representative of the industry, has a good market reputation and a large number of consumer groups. However, in recent years, the vitality of market performance has declined, and after the brand's main "slightly drunk" track experiencing nearly two years of explosion, it got rapid cooling in 2022. consumers' investments were cold, and financing was affected to a certain extent. Therefore, RIO cocktail needs new marketing strategies to deal with these challenges. Based on the collected sample data of consumers, this paper first uses factor analysis method to reduce the dimensionality of the ten factors affecting purchase, and reduces them into three common factors that affect people's purchase behavior, named as product quality and brand characteristics factor, product promotion and convenience factor, and product service and logistics factor. Then, K-means clustering is performed on existing consumers based on the three common factors, and three types of consumer groups are obtained, which are participants who pay attention to product quality, people who pay attention to brand and publicity, and loyal people who pay attention to convenience and after-sales. Finally, ordered Logistic regression model is employed to analyze the factors affecting consumers' purchase of RIO cocktails. The result shows that three significant factors affecting consumers' purchase are price, portability and brand reputation. Our research is intended to provide some reference for the precise positioning and differentiated marketing of RIO cocktails.

Keywords

RIO Cocktail, Factor Analysis, K-Means Clustering, Ordered Logistic Regression

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

预调鸡尾酒起源于欧美国家, 长期以来, 国内预调鸡尾酒一直由欧美国家的产品占据, 如爵士、冰锐等品牌。2003 年, 国内品牌锐澳(RIO)预调鸡尾酒正式诞生, 率先将消费主力定位为年轻人, 2005 年全面开拓市场, RIO 预调鸡尾酒迅速在都市年轻人中广为认知, 开启了中国预调鸡尾酒的时尚风潮。在 2011 年以前, 年轻人的酒饮选择非常有限, 除了白酒、啤酒等传统酒饮外, 仅有为数不多的几个鸡尾酒品牌。而在 2012 年初, 年轻人的酒饮选项变得空前丰富。RIO 鸡尾酒在 2012 年被植入《爱情公寓》从而打开知名度, 2013 年邀请周迅代言, 在各媒体上大规模投放, 走进了家家户户的电视里, 2014 年产品覆盖 200 个主要城市, RIO 大营销布局全面启动, 成为了年轻人的时尚潮流饮品。

从 2021 年开始, 伴随着近两年的热潮, 为了争抢年轻酒饮市场, 各类新品牌加大发力, 新品层出。虽然 RIO 面临着巨大挑战, 但品牌并没有坐以待毙, 而采用了跨界思维的营销手段, 推出了各类产品, 不断创新, 百润方面在 2022 年年报中披露, 强爽系列和清爽系列在 2022 年销售量上表现良好, 有着不错的复购率和复购周期。

关于低度鸡尾酒市场营销, 夏雨等(2017) [1]对于通过问卷收集, 使用因子分析与线性回归等方法进

行分析消费者看重的因素等, 对于 RIO 鸡尾酒市场的营销策略进行总结, 并在品牌宣传、产品价格、产品质量、品牌形象等方面提供营销建议。刘亮亮等人(2017) [2]对于 RIO 鸡尾酒天津分公司管理架构进行分析与提出问题, 并使用波特五力模型、PEST 模型等方法对天津市场分析后依据结果提出营销方法优化设计方案: 品牌定位于推广需要优化调整、产品策略定位需要更加精准、营销渠道有待进一步改进以及营销人员架构等问题需要进一步优化与加强。Tamara Bucher 等人(2018) [3]分析了低酒精饮料消费量持续增长背后的原因, 研究认为, 这是对酒精相关控制政策的一种回应, 并且还体现了消费者和整个社会健康意识的加强, 因为低酒精饮料被看作为一种更健康的选择。年轻消费者对低度酒饮料的需求不断增加, 低度酒行业有很大的进步发展空间。

关于消费群体分类, 徐吉等人(2020) [4]在奢侈品消费群体划分研究中, 根据内部因素(年龄、性别)与外部因素(地区生产力等)考虑, 从职业特征入手, 对奢侈品消费人群进行划分, 并基于划分结果提出营销策略建议。贾文帅等人(2021) [5]在中超联赛球迷消费动机的细分研究中, 对于消费人群使用多维度的划分方法, 根据消费特征进行划分。金华珍等人(2022) [6]在客户细分的研究中, 使用聚类算法对于客户进行划分, 并基于客户画像使用 STP 模型分析市场, 最终针对性提出未来的营销策略。

综上所述, 针对 RIO 鸡尾酒的发展, 诸多学者通常从宏观角度出发, 经过整体市场研究分析后给出进一步的营销策略建议, 具有较强的指导意义, 但就我们所知, 少有文献针对消费群体进行分类与分析消费者满意度从而提供营销策略。本文则从 RIO 鸡尾酒消费群体角度出发, 针对消费群体进行分类和具体画像, 同时也对消费因素进行满意度分析, 以期为 RIO 鸡尾酒未来市场发展提供一定借鉴意义。

2. 建模思路与所用方法介绍

2.1. 研究思路

与其他研究文献不同的是, 如图 1 研究思路所示, 本文获取到数据以及进行预处理后, 首先使用因子分析方法进行降维, 再基于其结果使用 K-means 聚类, 得到消费人群的不同画像结果并进行差异化分析。同时根据定序 Logistic 回归模型, 得到对消费者满意度影响显著的因素并结合市场需求进行分析。

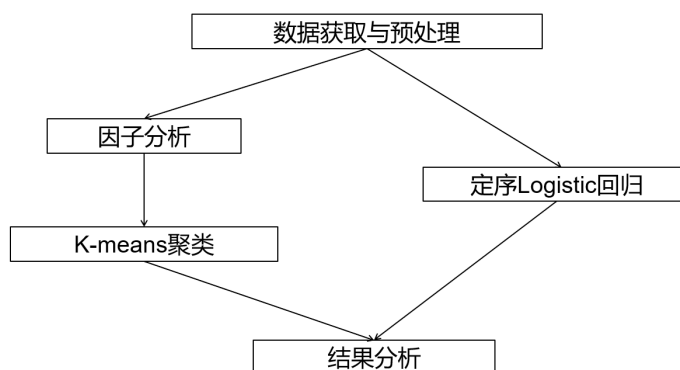


Figure 1. Research idea map
图 1. 研究思路图

2.2. 方法介绍

2.2.1. 因子分析

因子分析方法常用于综合评价中, 是通过降维的方法, 将原始指标按照关联程度进行分组, 组内指标相互之间的关联度更大, 并且每一组都会存在唯一的公共因子[7]。其优点在于: 因子之间不相关, 每

个因子能反映变量之间的关系, 有助于更客观地描述多个变量对样品的联系性影响, 因子方差贡献率生成的信息量权数, 比人为确定权数工作量少些, 有助于保证客观性。

2.2.2. K-Means 聚类

K-means 聚类是一种经典的基于划分的聚类方法[8]。其主要思想是: 在给定的数据集中, 首先随机选择 K 个初始聚类中心, 利用欧氏距离计算各个样本之间的距离, 根据距离最近原则将样本分为 K 个簇; 然后求出每一簇的平均位置, 重新分配聚类中心, 当迭代达到规定次数或者聚类中心不再发生变化, 则聚类结束并输出结果。

2.2.3. 定序 Logistic 回归

逻辑回归(Logistic Regression)是机器学习中最常见的一种用于二分类的算法模型。定序 logistic 回归模型中响应变量通常按照 1, 2, 3 等整数序列编码, 有明确的从高到低排序, 但相邻各类的差距未知。

定序 Logistic 模型为式(1):

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

其中, y 是观测现象内的趋势, 不能直接被测量; α 和 β_k 是模型需要标定的系数; x_k 是模型中自变量; ε 是模型的误差项。当实际观测反应变量有 n 种类别时($n=1, 2, \dots, N$), 相应取值为 $y=1, \dots, y=N$ 。则有 $N-1$ 个未知分界点 u_i ($i=1, \dots, N-1$)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 $y=2$; \dots ; 若 $u_{n-1} < y$, 则有 $y=n-1$ 。其中 $u_1 < u_2 < \dots < u_{n-1}$ 。累计概率由以下公式(2)预测:

$$P(y^* \leq n|x) = p(y \leq u_i | x) = \frac{\exp\left\{u_i - \left(\alpha + \sum_{k=1}^m \beta_k x_k\right)\right\}}{1 + \exp\left\{u_i - \left(\alpha + \sum_{k=1}^m \beta_k x_k\right)\right\}} \quad (2)$$

其中, $P(y \leq u_i | x)$ 表示在 x 的条件下 $y \leq u_i$ 发生的概率。

3. 实证分析

3.1. 数据获取与预处理

本文数据来自互联网调查问卷收集与电商平台数据, 并且使用 R 语言对数据进行剔除无效数据、转换虚拟变量等预处理, 便于下文建模分析。

3.2. 因子分析

3.2.1. 评价指标因子提取

首先对于原始变量进行信效度检验, 得到结果 KMO 值为 $0.915 > 0.8$, 巴特利特球形检验 sig 值 < 0.05 , 说明样本的 10 个变量特征具有很强的显著性, 可以进行因子分析。

运用主成分分析法, 根据碎石图提取三个特征值大于 1 的公共因子, 累计总方差载荷平方和达 81.063%, 可以反映原始变量中的大部分信息, 因子分析结果较好。

3.2.2. 因子命名

表 1 是因子分析的旋转成分矩阵, 其中公共因子 F1 在基酒口感、口味种类、价格、包装水准以及品牌口碑五个变量特征上具有较大载荷, 该因子描述产品质量与品牌特点, 因此我们将其命名为产品质量与品牌特点因子。

公共因子 F2 在文案质量、便携性与广告这三个变量特征上具有较大载荷, 该因子描述产品宣传与方便性特点, 因此我们将其命名为产品宣传与方便性因子。

公共因子 F3 在客服服务态度与效率、发货物流与售后这两个变量特征上具有较大载荷, 该因子描述产品服务与物流特点, 因此我们将其命名为产品服务与物流因子。

Table 1. Rotated component matrix
表 1. 旋转成分矩阵

	Component		
	1	2	3
Base wine taste	0.813	0.173	0.307
Types of taste	0.751	0.120	0.400
Price	0.696	0.522	0.038
Packaging level	0.736	0.363	0.225
Paperwork quality	0.222	0.728	0.433
Portability	0.284	0.756	0.191
Advertisement	0.226	0.824	0.241
Brand reputation	0.626	0.419	0.379
Customer service attitude and efficiency	0.273	0.368	0.822
Physical distribution and after-sales	0.439	0.275	0.770

3.3. 基于因子分析的消费人群聚类分析

我们根据降维后的公共因子, 对数据进行归一化后, 对消费人群进行聚类分析, 结果如下图 2 与表 2 所示。

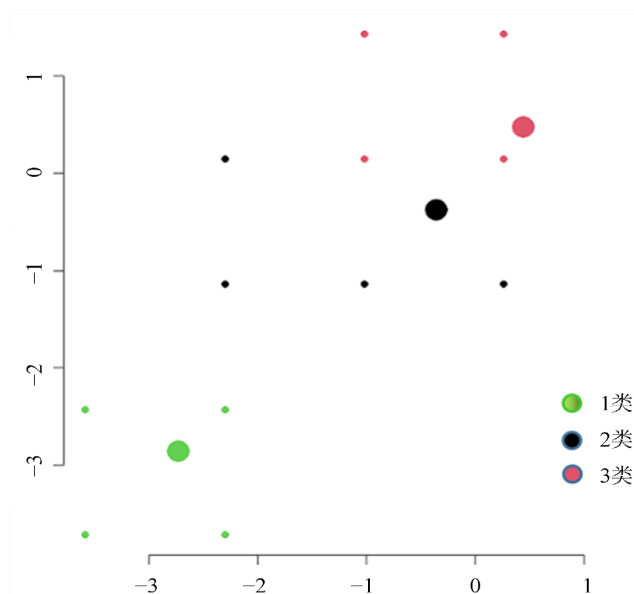


Figure 2. The normalized clustering result distribution map
图 2. 归一化后聚类结果分布图

由图 2 可知, 该三种聚类中心明显可辨, 不同类别消费人群可以较为容易区分开来, 具有良好分类效果。接下来, 本文针对这三种人群, 分别计算各类人群在 F1、F2、F3 这三种不同公共因子上的因子得分并比较其类间平均值, 结果如下表 2 所示。

Table 2. The average value of factors among the three common factors of all kinds of core population

表 2. 各类消费人群在三种公共因子上的因子类间平均值

Case clustering number		F1 score	F2 score	F3 score
Participants focusing on product quality	Average value	0.8710	-1.1160	0.0901
People who pay attention to brand and publicity	Average value	-0.8894	1.2131	-0.8322
Loyal crowd focusing on convenience and after-sales	Average value	-0.7695	0.1983	0.2906

从表 2 中我们对于 RIO 消费人群进行划分, 如图 3 所示。

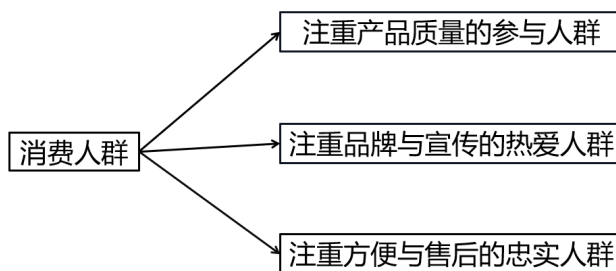


Figure 3. The division of consumer groups of RIO cocktails

图 3. RIO 鸡尾酒的消费人群划分

第一类消费人群归类为注重产品质量的参与人群。其在 F1 上得分较高, 故该人群更易受产品质量的影响。这类人群对于鸡尾酒原料的选取与质量高低有着很大看重。在 F3 上得分适中, 表明该类群体对于产品服务、物流等方面有所看重, 在 F2 上得分最低, 表明该类人群对于产品宣传并不看重。

第二类消费人群归类为注重品牌与宣传的热爱人群。其在 F1、F2 上得分均较高, 故该类人群对于 RIO 鸡尾酒品牌、质量与产品的宣传方式比较看重, 在 F2 上得分最大, 表明该类人群对于 RIO 广告宣传方式非常重视, 易受到广告宣传的影响进行消费。在 F3 上得分为负数表明该类人群对于售后服务并无需求。

第三类消费人群归类为注重方便与售后的重视人群, 其在 F2、F3 上得分较高, 表明该类人群更易受到 RIO 产品的便携性与售后服务影响。在 F1 上得分最小, 表明该类群体在消费时品牌意识淡薄。

3.4. 基于定序 Logistic 回归模型的满意度影响因素分析

3.4.1. 评价指标因子提取

本文基于定序 Logistic 回归模型, 利用所收集到的样本中消费者的数据, 即消费或饮用过 RIO 鸡尾酒的用户数据, 以消费者对 RIO 鸡尾酒十个方面的满意度打分作为自变量, 将“每月用于购买 RIO 鸡尾酒的花费”这种多元选择行为分为四个等级, 作为判断用户购买力的因素, 当选择“20 元以下”时, 取值为 1; 选择“20~50 元”时, 取值为 2; 选择“50~150 元”时, 取值为 3; 选择“150 元以上”时, 取值为 4, 以此作为因变量构建关于 RIO 鸡尾酒消费影响因素的序次 Logistic 回归模型。根据问卷调查结果, 从中选择模型变量如下表 3 所示。

Table 3. Ordered Logistic model variable table
表 3. 定序 Logistic 模型变量表

Attribute of variable	Variable name	Variable definition	
Dependent variable	The monthly cost of purchasing RIO cocktails	1) Under 20 yuan 2) 20~50 yuan 3) 50~150 yuan 4) Above 150 yuan	
Independent variable	Core explanatory variables	Base wine	
		Flavor	
		Price	
		Packing	
		Copywriting	1) Highly dissatisfied 2) Dissatisfied 3) General 4) Satisfied 5) Highly satisfied
		Portability	
		Ad	
		Brand	
		Attitude	
		Logistics	
Control variable	Profession	1, Student group 2, Enterprises, institutions 3, Servicing business 4, Production industry 5, Special profession	
		Income	1) Under 1000 yuan 2) 1000~1999 yuan 3) 2000~2999 yuan 4) 3000~4999 yuan 5) Above 5000 yuan

本文构建的定序 Logistic 回归模型如式(3)所示:

$$y = \beta_0 + \beta_1 \text{Basewine} + \beta_2 \text{Flavor} + \beta_3 \text{Price} + \beta_4 \text{Packing} + \beta_5 \text{Copywriting} + \beta_6 \text{portability} + \beta_7 \text{Ad} + \beta_8 \text{Brand} + \beta_9 \text{Attitude} + \beta_{10} \text{Logistics} + \beta_{11} \text{Profession} + \beta_{12} \text{Income} + \varepsilon \quad (3)$$

3.4.2. 模型参数估计及检验

首先, 对模型参数估计结果进行平行线检验和整体显著性检验, 检验结果如下表 4 所示。对于初步构建的回归方程, 平行线检验结果表明, 显著性为 $0.121 > 0.05$, 说明此模型设定的因变量所定义的四个级别的花费金额是符合等差条件的, 即构建定序 Logistic 回归模型被证明是能够真实反映自变量对于因变量的影响的。

Table 4. Parallel line test results
表 4. 平行检验结果

Model	-2 Log likelihood	Chi-square	Degree of freedom	Significance
Null hypothesis	627.210			
Routine	594.974b	32.236c	24	0.121

对该模型的整体性进行检验, 模型中使用似然比检验中的 χ^2 统计量为 48.200, p 值为 0.002, 远小于

0.1 的显著性水平, 可以认为解释变量的整体显著性良好。同时, 我们对于模型自变量进行相关分析, 发现 $VIF < 10$, 即不存在多重共线性问题。此外, 模型中不含有异常点与强影响点。

对于模型回归结果, 在 0.05 的显著性水平下, 删除 p 值大于 0.05 的变量, 经过筛选, 得到剩余三个自变量的回归系数、标准误差、Wald 值以及 p 值, 如下表 5 所示。

Table 5. Regression results of ordinal logistic model

表 5. 定序 Logistic 模型回归结果

Independent variable	Regression coefficient	Standard error	wald	Significance
Price	0.432	0.192	5.076	0.024
Portability	0.493	0.198	6.167	0.013
Brand reputation	-0.749	0.240	9.777	0.002

3.4.3. 模型结果分析

从模型回归结果可以看出, 价格、便携性和品牌口碑三个变量显著, 对于 RIO 鸡尾酒的消费选择影响显著。其中“价格”的回归系数为 0.432, 因此可知, RIO 鸡尾酒价格能显著影响消费者的购买频率, 消费者们越认为 RIO 鸡尾酒定价是合理的, 消费频率越高, 这也符合消费者追求“物美价廉”的正常心理。因此, 在价格制定上 RIO 鸡尾酒应该参考市场标准, 在保障品质的基础上适度降低价格, 以吸引消费者购买。

“便携性”这一变量的回归系数为 0.493, 即 RIO 包装越便携、越方便消费者, 对于 RIO 鸡尾酒的销量越有益处。由此, RIO 鸡尾酒可以改进包装设计, 使产品更轻便易携, 满足消费者不同场景下, 携带该商品的需求, 提高消费者购买欲望, 以促进销售。

“品牌口碑”的回归系数为-0.749, 这表示消费者对各种鸡尾酒的品种区分较小, 了解度不够, 即品牌口碑对消费者的购买频率影响不大。因此, RIO 鸡尾酒品牌应加强品牌宣传, 树立独家优良口碑, 使消费者得以区分各个品牌间的不同, 在购买同类产品时, 偏向选择 RIO 鸡尾酒品牌, 从而提高市场份额, 刺激消费。

4. 营销策略建议

4.1. 针对不同人群偏好, 把握饮用场景需求

互联网时代, 消费者的社交需求更为显著, 也因此催生出更显著的小聚饮酒文化。RIO 鸡尾酒品牌可通过绑定佐餐场景, 如吃火锅、川菜时等, 同时强调聚会氛围, 从而打造优质的社交小酒, 发掘出更多的产品宣传创意, 同时也要根据不同人群看重的购买特征, 在不同侧重点出不断创新, 满足各类消费人群的需求, 进一步拓宽改善营销渠道, 扩大市场份额。

4.2. 进一步完善定价与促销方式

RIO 品牌在定价时需要考虑消费者购买力现状, 针对不同消费能力群体特点定制大众化或高端化产品策略。在促销方式方面, 可以推出原装赠品、凑单优惠等多种的促销方式, 不断完善延伸产品服务赠送, 进一步吸引顾客, 刺激消费, 从而占领更多市场。

4.3. 改进产品包装形式, 提升品质保障口碑

品牌要做到在提升产品品质的同时, 不断的创新研制出更好的新产品, 满足当代多元消费者的多样化的需求。所以, 品牌应聘请更多专业研究人员, 提升整体科研水平, 集中公司的财力、物力和人力等

优势资源进行研发,在保障产品品质同时,改进产品包装形式,增加产品的易携带性,促进产品的流通,从而增加品牌影响力。同时,随着大数据时代发展,消费者的自主话语权不断加强,把握住顾客需求,在产品的度数、包装、营销方式多维度的创新,是未来发展的关键。未来在产业政策指导下,打造差异性,从同类产品中脱颖而出,让消费者对于 RIO 品牌更加喜爱与信赖。

5. 总结

本文将所收集数据进行预处理后,通过因子分析进行降维,得到三个公共因子,并且基于公共因子进行 K-means 聚类,得到 RIO 鸡尾酒消费人群可以划分为三类人群:注重产品质量的参与人群、注重品牌与宣传的热爱人群、注重方便与售后的忠实人群。同时使用定序 Logistic 回归模型并对模型不断检验,在保证模型拟合程度优良条件下,针对消费者的 RIO 鸡尾酒满意因素进行分析,得到三个重要特征:价格、便携性与品牌口碑。并根据所得结果对营销新方向进行分析,给出相应营销策略建议,以期为 RIO 鸡尾酒精准性定位与差异性营销策略提供借鉴意义,也为其他低度酒精饮品品牌发展新方向提供一定参考。

基金项目

北京工商大学 2023 年大学生科学研究与创业行动计划市级项目(编号 B008)。

参考文献

- [1] 夏雨. RIO 鸡尾酒市场调查与营销策略研究[D]: [硕士学位论文]. 南昌: 江西师范大学, 2017.
- [2] 刘亮亮. 天津地区锐澳预调鸡尾酒市场营销策略研究[D]: [硕士学位论文]. 天津: 天津大学, 2017.
- [3] Bucher, T., Deroover, K. and Stockley, C. (2018) Low-Alcohol Wine: A Narrative Review on Consumer Perception and Behaviour. *Beverages*, 4, 82. <https://doi.org/10.3390/beverages4040082>
- [4] 徐吉, 侯玲. 奢侈品消费群体分类分析[J]. 纳税, 2020, 14(3): 174+177.
- [5] 贾文帅, 张铁民. 中超联赛球迷消费动机的细分研究——基于多维度球迷群体划分方法[C]//国家体育总局体育文化发展中心, 中国体育科学学会体育史分会. 2021 年“一带一路”体育文化学术大会论文摘要集. 2021: 26-27. <https://doi.org/10.26914/c.cnkihy.2021.065793>
- [6] 金华珍. 基于消费者画像的客户细分与精准营销研究[D]: [硕士学位论文]. 延边: 延边大学, 2022. <https://doi.org/10.27439/d.cnki.gybd.2022.000116>
- [7] 董雨桐, 李瑞锋, 朱文涛, 王薇, 李梦玲, 王玉伟. 基于灰色关联法和因子分析法对北京市不同圈层卫生资源配置水平的综合评价研究[J]. 中国社会医学杂志, 2023, 40(1): 101-105.
- [8] 廖纪勇, 吴晟, 刘爱莲. 基于相异性度量选取初始聚类中心改进的 K-means 聚类算法[J]. 控制与决策, 2021, 36(12): 3083-3090. <https://doi.org/10.13195/j.kzyjc.2020.0554>