

Analysis of Gender Based on Online Shopping Loyalty

Guoshun Lin, Li Zhang, Ye Wang

Dalian Maritime University, Dalian Liaoning
Email: 18342206260@163.com

Received: Jan. 23rd, 2020; accepted: Feb. 17th, 2020; published: Feb. 24th, 2020

Abstract

In the E-market environment, consumers have more options and more free when they choose something. It is particularly important for the business to improve consumer loyalty. Therefore, the business should know how to recognize their own customers so as to improve customers' loyalty and gain more profits when they are in the B2C market. In this paper, a non-parametric test—run test, is adopted to test the gender differences of consumers, and analyze whether the gender will impact consumers' loyalty, so that business can take gender into account when launching appropriate sales strategies to attract customers and achieve greater profits.

Keywords

E-Commerce, Loyalty, Run Test, Difference Analysis

基于网购忠诚度的性别差异分析

林国顺, 张丽, 王野

大连海事大学, 辽宁 大连
Email: 18342206260@163.com

收稿日期: 2020年1月23日; 录用日期: 2020年2月17日; 发布日期: 2020年2月24日

摘要

在网络市场环境下, 消费者选择种类多, 选择自由度加大, 提高消费者的忠诚度对于商家来说显得尤为重要, 因此, B2C商家应当区分自己的客户, 这样才能有针对性地提高顾客忠诚, 获得更大利润。本文采用非参数检验法——游程检验, 对消费者的性别做差异性检验, 分析性别对忠诚度是否存在差异, 以便商家在推出销售策略时, 考虑到性别的差异, 有针对性的吸引客户, 获取更大盈利。

关键词

电子商务, 忠诚度, 游程检验, 差异性分析

Copyright © 2020 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

随着智能网络的快速发展, 各类行业都相继进入“互联网+”的发展时代, 我国网民规模继续保持平稳增长, 电商时代到来, 网络经济产生巨大的网络数据, 大数据时代到来。

消费者对商家的购买行为促成商家的盈利, 而如何更有效地吸引消费者, 降低消费者流失, 提高消费者对商品质量、物流运输、性价比的满意程度, 才能提高回头率, 进而提高消费者对于商家的忠诚度。不同的消费者有不同的属性, 例如性别、年龄、职业、收入等, 属性的差异可能会影响忠诚度, 进而影响商家实施销售策略。本文主要分析消费者的性别差异, 旨在判断性别在忠诚度方面是否产生差异。

2. 研究现状

国内学者韩经纶和韦福祥从影响顾客忠诚度要素的角度出发, 指出顾客忠诚度是由价格、产品、服务特性等各种要素影响下, 顾客长期购买某一品牌产品或服务的行为[1]。陆娟指出服务忠诚由情感忠诚和行为忠诚两维度构成; 服务质量不仅对服务忠诚存在着显著的直接正向影响, 而且通过顾客满意对服务忠诚存在着显著的间接正向影响[2]。美国资深营销专家 Jill Griffin 认为, 顾客忠诚度是指顾客出于对企业或品牌的偏好而经常性重复购买的程度[3]。

国内众多研究者对影响 B2C 顾客忠诚度的各种相关因素进行了深入的研究, 如狄佩佩探讨了 B2C 模式下顾客忠诚度的定义以及影响 B2C 电子商务的各种影响因素, 构建了 B2C 顾客忠诚度的驱动模型[4]; 唐胡鑫提出一种采用回归预测的支持向量机的算法, 旨在应用回归预测机更好的利用用户历史购买行为记录, 并分析用户购买行为记录的背景, 从而提高电子商务用户忠诚度评估的准确度[5]。邓爱民等构建了以信任、在线网站特性、线下物流服务质量、顾客满意度、转换成本为外因潜在变量, 以顾客忠诚度为内因潜在变量的假设模型, 并采用因子分析和结构方程模型揭示网络环境下顾客忠诚度的影响因素和作用机制[6]。李敏等对电子商务顾客忠诚度影响因素进行了全面的分析, 以经典 RFM 客户忠诚度模型为基础, 建立 RFMSA 电子商务客户忠诚度划分模型。通过聚类分析算法对顾客忠诚度进行划分, 以经典聚类分析算法 K-means 为基础, 提出分段确定初始聚类中心的改进算法对顾客忠诚度进行划分, 提高了聚类的准确率[7]。

本文所指的忠诚度, 是指顾客出于对企业或品牌的偏好而经常性重复购买的程度。主要根据消费者的行为数据建立量化模型, 进行差异性分析, 通过游程检验法判断性别是否为影响忠诚度差异的因素。以便商家在推出销售策略时, 考虑到性别的差异, 有针对性的吸引客户, 获取更大盈利。

3. 模型构建

3.1. 忠诚度量模型构建

为进行忠诚度差异性分析, 需要建立忠诚度量模型, 即顾客对店铺的重复浏览, 加购、收藏、购

买行为的综合量化的指标。

$$\text{忠诚度 } F_i = (\text{Browse } i., \text{ Add to basket } i., \text{ Collect } i., \text{ Purchase } i.) \tag{1}$$

建立忠诚度模型 F

F 的四个分量应满足以下条件：

- 1) 非负性
- 2) 分量越大，忠诚度越高

本文引出忠诚度模型旨在将忠诚度量化，后引出忠诚度指数模型，文章主要将忠诚度指数模型作为依据判断性别差异是否引起忠诚度的差异。

忠诚度 F 取决于四个忠诚度指数：

Browse_{ij} 表示 i 顾客对 j 店铺的浏览量(简写 B_{ij})

Add to basket_{ij} 表示 i 顾客对 j 店铺的加购量(简写 A_{ij})

Collect_{ij} 表示 i 顾客对 j 店铺的收藏量(简写 C_{ij})

Purchase_{ij} 表示 i 顾客对 j 店铺的购买量(简写 P_{ij})

$$i \text{ 顾客浏览忠诚度指数 } \text{Browse}_i = \sum_j B_{ij}^2 \tag{2}$$

$$i \text{ 顾客加购忠诚度指数 } \text{Add to basket}_i = \sum_j A_{ij}^2 \tag{3}$$

$$i \text{ 顾客收藏忠诚度指数 } \text{Collect}_i = \sum_j C_{ij}^2 \tag{4}$$

$$i \text{ 顾客购买忠诚度指数 } \text{Purchase}_i = \sum_j P_{ij}^2 \tag{5}$$

通过构建这样的模型，将忠诚度量化，忠诚度越高，函数值越大。而四个忠诚度之间的关系如下图 1 所示：

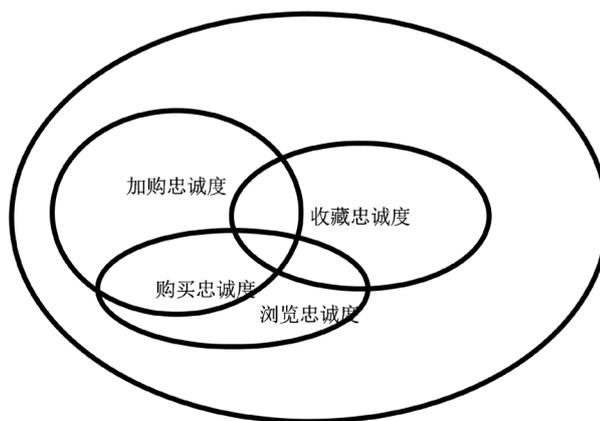


Figure 1. Schematic diagram of the relationship between the four indicators
图 1. 四个指标相互关系示意图

发生加购或购买或收藏行为以浏览行为为前提，所以浏览忠诚度包含加购忠诚度，收藏忠诚度和购买忠诚度，而加购、收藏、购买之间可能相关可能独立，这之间的忠诚度为相交关系。忠诚度指数是根据忠诚度得到的，所以浏览忠诚度指数影响加购、收藏和购买的忠诚度指数，加购、收藏、购买这三个

忠诚度指数之间相互影响，具体关系示意图如图 2 所示：

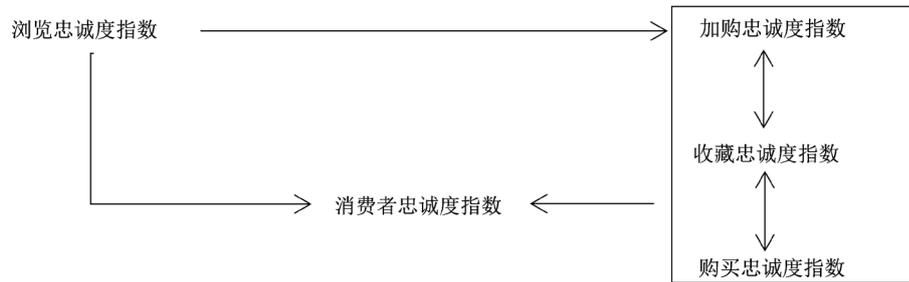


Figure 2. Schematic diagram of the relationship between the loyalty indexes
图 2. 忠诚度指数相互关系示意图

3.2. 数据介绍以及模型分析

本文数据采用阿里巴巴天池大赛的数据。

数据集包含匿名用户 2017 年 5 月到 11 月总共 6 个月内的购物日志。

数据集字段描述表如下表 1：

Table 1. Field description table

表 1. 字段描述表

User_id	Gender	Seller_id	Action_type	Time_stamp
用户	性别	商户	行为种类	时间戳

现从中抽取 8 个消费者对不同商家的浏览次数，以及计算得到的这 8 个消费者的浏览忠诚度指数。

通过分析浏览的集中程度判断忠诚指模型构建的准确性，如表 2 所示：

Table 2. Example table of consumer visits and browsing loyalty index for different merchants

表 2. 消费者对不同商家的浏览次数及浏览忠诚度指数示例表

消费者	商家 ID	浏览次数	忠诚度指数	总浏览量
*157073	131	10	685	81
*157073	749	5	685	81
*157073	1158	11	685	81
*157073	1462	6	685	81
*157073	1885	12	685	81
*157073	2014	3	685	81
*157073	2064	4	685	81
*157073	2869	10	685	81
*157073	3214	3	685	81
*157073	3308	10	685	81
*157073	3845	3	685	81
*157073	3926	4	685	81

Continued

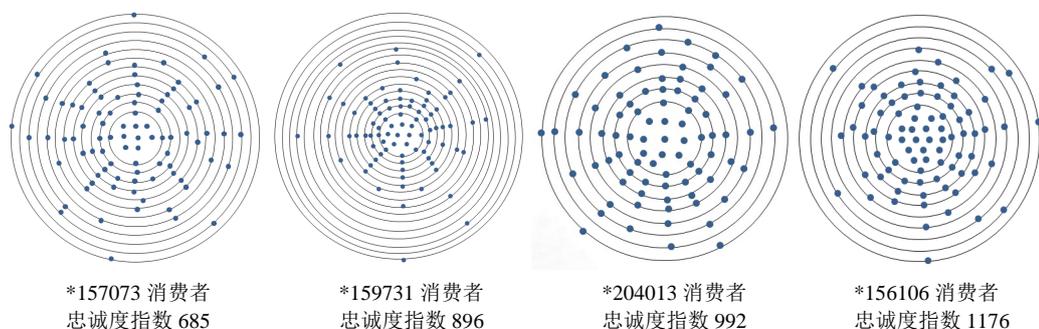
*100258	1125	4	2144	86
*100258	1186	12	2144	86
*100258	1253	7	2144	86
*100258	1495	5	2144	86
*100258	1640	2	2144	86
*100258	2492	4	2144	86
*100258	2781	43	2144	86
*100258	3293	5	2144	86
*100258	4131	4	2144	86
*156106	749	2	1176	82
*156106	1010	6	1176	82
*156106	1249	14	1176	82
*156106	1472	16	1176	82
*156106	1574	4	1176	82
*156106	2811	11	1176	82
*156106	2940	3	1176	82
*156106	3484	23	1176	82
*156106	4536	3	1176	82
*160139	142	5	3320	86
*160139	343	1	3320	86
*160139	1617	2	3320	86
*160139	1802	3	3320	86
*160139	1860	10	3320	86
*160139	1875	56	3320	86
*160139	2387	6	3320	86
*160139	2576	3	3320	86
*159731	270	14	896	84
*159731	840	18	896	84
*159731	882	9	896	84
*159731	899	5	896	84
*159731	913	4	896	84
*159731	964	0	896	84
*159731	1000	4	896	84
*159731	1236	10	896	84
*159731	1805	11	896	84
*159731	1839	1	896	84

Continued

*159731	1952	1	896	84
*159731	2267	2	896	84
*159731	2682	3	896	84
*159731	3242	1	896	84
*159731	4989	1	896	84
*234191	286	6	4547	81
*234191	3063	67	4547	81
*234191	3586	2	4547	81
*234191	3647	3	4547	81
*234191	4368	3	4547	81
*274376	880	14	2670	80
*274376	1403	11	2670	80
*274376	2456	48	2670	80
*274376	3248	7	2670	80
*204013	72	12	992	80
*204013	1583	9	992	80
*204013	1887	9	992	80
*204013	2099	3	992	80
*204013	2791	16	992	80
*204013	2886	18	992	80
*204013	3226	4	992	80
*204013	3663	9	992	80

抽取 8 位消费者对不同商家的浏览量进行统计, 其中这八位消费者总浏览数量均为 80~90, 将消费者对不同商家的浏览量放在不同的圆上, 浏览量高的放在最中间的圆, 向外依次递减, 如图 3 所示。

得到以下示意图, 根据浏览总量大致相同, 不同忠诚度的消费者对商家浏览的分布情况判断此次消费者是否具有较高的浏览集中程度, 以此可判断忠诚度指数模型有效。



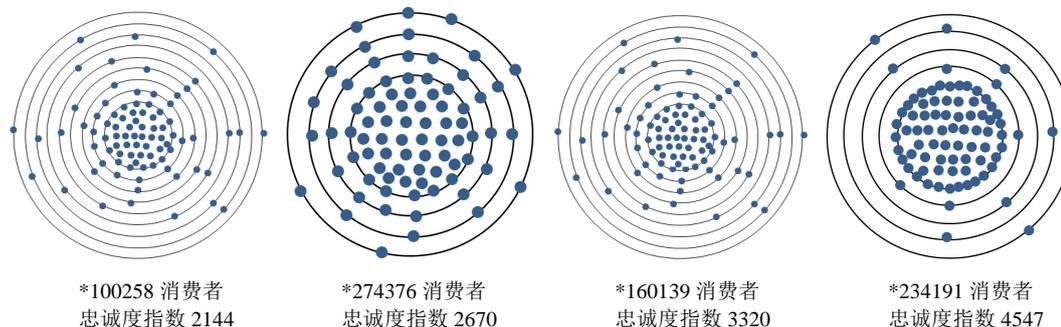


Figure 3. Diagram of the concentrative degree that show consumers' page views to online shops
图 3. 消费者对商家浏览量集中程度示意图

结论：由图可知，ID 为 157073 的消费者忠诚度指数为 685，大致均匀分布；ID 为 159731 的消费者忠诚度指数为 896，相对上一个消费者浏览程度较集中；可得在总浏览量大致相同的情况下，忠诚度指数越高，示意图中心的点越集中，忠诚度指数越低，点会越分散，会均匀分布在圆环上，可知，忠诚度指数模型有效。

4. 性别差异性分析

4.1. 游程检验的方法介绍

游程检验法通常用于检验随机样本的独立性。

假定有一个容量 $n = 10$ 的样本，对于它的每一个观测值，按照一定的标准区分为两种符号，设这两种符号分别为 A 和 B，如果我们把这些符号按照抽样时的先后顺序排序，就可能出现以下情况：

第一种情况：BBBBBAAAAA

第二种情况：ABABABABAB

第三种情况：BBAAABAABB

等等。下面，我们给游程(RUN)的定义：一个游程就是指某序列中位于一种之前或者之后的另一种符号持续的最大子序列，或者说，一个游程是指某序列中同类元素的一个最大持续子集[8]。

示例游程分析如下表 3 所示：

Table 3. The analysis of procedure checking
表 3. 流程检验分析

序号	消费者	性别	浏览宽度指数	游程数
1	*13603	0	13622466	1
2	*354236	0	12024796	1
3	*88181	0	9639455	1
4	*289172	0	8350740	1
5	*97807	1	8114343	2
6	*101790	0	6642960	3
7	*325548	1	6193209	4
8	*132509	0	5928619	5
9	*337277	0	5706122	5

浏览忠诚度指数是按照从高到低排列,不同消费者有性别差异,可以得到游程数。其中 0 代表女性,1 代表男性,前 4 个人都是女性第 5 个人性别改变,加一个游程。以此类推,此样本的游程数为 5。

4.2. 性别差异性计算过程

4.2.1. 选取全部数据进行性别差异性分析

选择全部数据,女性 2158,男性 1130,一共 3288 人。

在本文中 $N = 3288$,其中 $N_1 = 2158$, $N_2 = 1130$,游程数 $r = 1479$,取置信度水平 $\alpha = 0.05$,查正态分布表得 $Z_{\alpha/2} = 1.96$

$$E_r = \frac{2n \ln 2}{n_1 + n_2} + \frac{1}{2} = \frac{2 \times 2158 \times 1130}{2158 + 1130} + \frac{1}{2} = 1483.8 \quad (6)$$

$$D_r = \frac{2n \ln 2 (2n \ln 2 - n_1 - n_2)}{n_1 + n_2^2 (n_1 + n_2 - 1)} \quad (7)$$

$$= \frac{2 \times 2158 \times 1130 (2 \times 2158 \times 1130 - 2158 - 1130)}{2158 + 1130^2 (2158 + 1130 - 1)} = 668.9$$

$$Z_r = \frac{|r - \mu_r|}{\sigma_r} = \frac{|1479 - 1483.8|}{\sqrt{668.9}} = 0.18 < 1.96 \quad (8)$$

说明浏览忠诚度对性别是无差异的。

同样对加购,收藏,购买做差异性分析,得到如下统计结果,如表 4 所示。

Table 4. Statistical results of difference-analysis to the whole data

表 4. 全部数据差异性分析统计结果

	游程数	统计指标(Zr)	显著性
浏览忠诚度差异分析	1479	0.18	无关
加购忠诚度差异分析	1487	0.12	无关
收藏度差异分析	1434	1.93	无关
购买忠诚度差异分析	602	34.09	显著

根据结果显示,购买忠诚度对性别有明显差异,浏览,加购,收藏的忠诚度对性别没有差异。

4.2.2. 抽取 50%的数据做性别差异性分析

剔除浏览量低的数据,本文抽取浏览量大的 50%的数据,即浏览量大于 36 次的样本,共计 2313 人,其中女性 1592 人,男性 721 人。

计算过程如 4.2.1。

得到以下汇总表,如表 5 所示。

根据结果显示,购买忠诚度对性别有明显差异,加购忠诚度对性别存在差异,浏览与收藏忠诚度对性别没有差异。

4.2.3. 抽取 30%的数据做性别差异性分析

继续剔除浏览量低的数据,选择浏览量大于 60 次的数据,其中女性 836 人,男性 328 人。

计算过程同 4.2.1。

Table 5. Statistical results of difference-analysis to 50% data
表 5. 50%数据差异性分析统计结果

	游程数	统计指标(Zr)	显著性
浏览忠诚度差异分析	685	0.94	无关
加购忠诚度差异分析	625	2.70	相关
收藏度差异分析	651	1.12	无关
购买忠诚度差异分析	267	24.39	显著

得到如下汇总表，如表 6 所示。

Table 6. Statistical results of difference-analysis to 30% data
表 6. 30%数据差异性分析统计结果

	游程数	统计指标(Zr)	显著性
浏览忠诚度差异分析	481	0.96	无关
加购忠诚度差异分析	437	3.55	显著
收藏度差异分析	439	3.35	显著
购买忠诚度差异分析	199	27.97	显著

通过剔除低浏览量的数据，得到以上结果，由结果分析可知，加购，收藏，购买忠诚度对性别存在显著差异，浏览忠诚度对性别没有差异。

通过筛选高浏览量数据，剔除一些异常数据，可使得结果更具有准确性。

4.3. 结论

方案 A 代表用全部数据做差异性分析，方案 B 抽取 50%的数据做差异性分析，方案 C 抽取 30%的数据做差异性分析，得到如下汇总表，如表 7 所示。

Table 7. The summary results of proposal
表 7. 方案汇总结果

方案	浏览忠诚度指数(Zr)	加购忠诚度指数(Zr)	收藏忠诚度指数(Zr)	购买忠诚度指数(Zr)
A	无关(0.18)	无关(0.12)	无关(1.93)	显著(34.09)
B	无关(0.94)	相关(2.70)	无关(1.12)	显著(24.39)
C	无关(0.96)	显著(3.55)	显著(3.35)	显著(27.97)

根据表格可知，方案 C 有 3 个指标出现显著差异，方案 A 只有一个指标出现差异性，而方案 B 有 2 个指标存在差异，所以选取高浏览量的数据进行分析，得到的结果更加明显。由结论可知，性别在影响忠诚度方面，是有差异性的。

全部数据存在异常数据，而选择高浏览量的数据作为样本，可以减少异常数据，使得分析的结果更为合理。

数据显示在差异性显著的条件下，女性的忠诚度普遍大于男性。

例如，在 A 方案购买忠诚度下 Z_r 为 34.09，性别差异性显著，其中男性购买忠诚度指数的平均值为 20.53，女性购买忠诚度指数的平均值为 28.29，女性购买忠诚度高出男性购买忠诚度 38%，所以商家在

推出销售策略时, 应考虑到性别的差异, 有针对性地吸引客户, 以获取更大盈利。

利用游程检验法同样可以对年龄进行差异性分析, 作因素的相关性分析, 可以应用于时间序列、股票市场、流行性疾病等方面。同时, 随着电商时代到来, 应利用好网络经济产生的网络数据, 进行分析研究, 谋取更大效益。

参考文献

- [1] 韩经纶, 韦福祥. 顾客满意与顾客忠诚互动关系研究[J]. 南开管理评论, 2001, 4(6): 8-10, 29.
- [2] 陆娟, 芦艳. 服务质量与服务忠诚的多维度关系研究[J]. 财贸研究, 2006(6): 80-87.
- [3] 顾客忠诚度[EB/OL]. <https://baike.so.com/doc/7555106-7829199.html>, 2016-10-19.
- [4] 狄佩佩. B2C 模式下顾客忠诚度影响因素研究[J]. 现代商业, 2017(14): 40-41.
- [5] 唐胡鑫. 关于电商用户忠诚度评估方法的研究[J]. 计算机仿真, 2017, 34(1): 348-351.
- [6] 邓爱民, 陶宝, 马莹莹. 网络购物顾客忠诚度影响因素的实证研究[J]. 中国管理科学, 2014, 22(6): 94-102.
- [7] 李敏, 张玉莹, 于欣雨. 基于聚类分析的电子商务客户忠诚度研究[J]. 哈尔滨商业大学学报(自然科学版), 2017, 33(2): 175-179.
- [8] 林国顺, 李向文. 交通运输模拟建模技术[M]. 大连: 大连海事大学出版社, 2004: 63-67.