

基于多元回归及因子分析的宁波酒店评分的实证分析

严冰悦, 郑 涵, 王 志

宁波工程学院理学院, 浙江 宁波

收稿日期: 2022年5月27日; 录用日期: 2022年6月16日; 发布日期: 2022年6月28日

摘 要

本文选取宁波市近一年在某酒店信息网站上排名前六百名酒店信息, 通过多元回归统计分析, 得出综合评分是由评论数、平均价格、质量等级、是否提供WiFi、环境评级、客房数之间共同形成的线性关系表示。最后通过因子分析, 将不同等级的星级酒店按照酒店服务、酒店规模、酒店口碑三个方面进行分析, 得出星级酒店只在酒店规模有明显的区分, 经济型酒店在酒店服务和酒店口碑方面并不弱于高星级品牌酒店。

关键词

酒店评分研究, 多元回归分析, 因子分析

Empirical Analysis of Ningbo Hotel Ratings Based on Multiple Regression and Factor Analysis

Bingyue Yan, Han Zheng, Zhi Wang

School of Sciences, Ningbo University of Technology, Ningbo Zhejiang

Received: May 27th, 2022; accepted: Jun. 16th, 2022; published: Jun. 28th, 2022

Abstract

This paper selects the information of the top 600 hotels ranked on a hotel information website in Ningbo in recent one year. Through multiple regression statistical analysis, it is concluded that the comprehensive score is expressed by the linear relationship between the number of comments,

average price, quality grade, whether WiFi is provided, environmental rating and the number of rooms. Finally, through factor analysis, the star hotels of different levels are analyzed according to three aspects: hotel service, hotel scale and hotel reputation. It is concluded that star hotels only have obvious distinction in hotel scale, and economy hotels are not inferior to high star brand hotels in hotel service and hotel reputation.

Keywords

Hotel Ratings Research, Multiple Regression Analysis, Factor Analysis

Copyright © 2022 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

近年来,我国酒店行业发展势头强劲,主要归功于酒店行业自身的服务能力不断提高以及旅游业的蓬勃发展,从而推动了酒店行业市场需求不断扩大[1]。网络时代,公开的酒店评分成为大众选择酒店的重要参考依据,提升评分对于吸纳顾客具有重要的意义[2]。本文对比不同星级的酒店评分的显著影响因素及在酒店服务、酒店规模、酒店口碑等方面的区别,分析宁波市酒店行业的发展现状。

通过大数据共享交易市场 iDataAPI 平台获取综合排名为前六百的酒店数据并进行汇总。包括酒店平均价格、品牌名称、早餐信息、经营状况、评论数、地区、环境评级、收藏数、是否提供 WIFI、酒店类型、质量等级、评分、客房数、星级等信息。通过多元回归分析及因子分析等统计分析方法对宁波不同星级的酒店间的规模、服务、口碑等方面进行实证分析。

2. 多元回归模型实证分析

2.1. 多元回归分析意义

初步通过描述性分析得出不同星级的酒店对于该酒店的评分具有较大差异,但不同星级是较为抽象的概念,为深入分析不同星级的酒店之间的具体差异,建立多元回归模型,进一步分析对酒店评分显著因素。

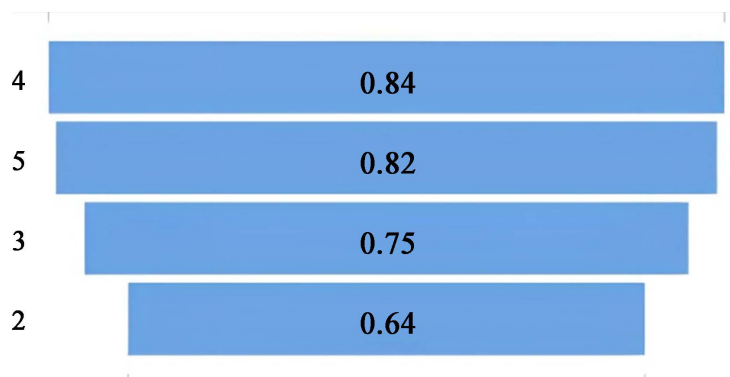


Figure 1. Display of average value of different star ratings

图 1. 不同星级评分平均值展示

由图 1 得, 大数据共享交易市场 iDataAPI 平台获取综合排名为前六百的酒店不同星级评分平均值展示, 纵坐标数字代表酒店星级数。从中发现随着星级的增加, 酒店的总体评分也随之增加, 但各个星级之间的差距并不是很大, 尤其是四星级与五星级酒店, 甚至四星级的平均得分高于五星级酒店。因评分为酒店平台所给标准, 并不清楚其实际的评分规则, 可通过多元回归分析后, 初步探索评分标准, 以便进一步分析对酒店评分显著的因素。

2.2. 模型建立的思路

选择评论数、平均价格、星级、质量等级、是否提供 WiFi、环境评级、优质评论数、收藏数、总共层数和客房数这十个变量作为自变量, 酒店的评分作为因变量初步建立回归模型; 进行显著性检验, 剔除未通过显著性检验的变量; 进行正态性检验所有变量是否或近似符合正态性, 通过逐步回归调整变量, 优化模型的拟合度。

2.3. 实证分析

通过观察选出大致可以用作评分的变量, 分别是评论数、平均价格、星级、质量等级、是否提供 WiFi、环境评级、优质评论数、收藏数、总共层数和客房数这十个相关变量。

Table 1. Estimation of model parameters

表 1. 模型参数估计表

变量	参数估计	P 值
截距	2.415e+00	<2e-16
评论数	1.566e-05	6.34e-07
平均价格	7.481e-05	0.03143
星级	1.478e-02	0.32111
质量等级	2.520e-01	2.49e-07
是否提供 WiFi	9.146e-02	0.00288
环境评级	2.293e-01	2.77e-05
收藏数	-4.151e-06	0.68978
客房数	4.438e-04	0.03696

通过表 1 可以得出, 变量并没有全部通过显著性检验。最显著的变量是评论数、质量等级和环境评级。是否提供 WiFi 对评分有较强的影响。平均价格和客房数对评分要求会有一些影响, 但并不显著。

模型优化 1: 选出显著性变量, 进行分析: 选择评论数、平均价格、质量等级、是否提供 WiFi、环境评级和客房数这六个变量对评分的影响。

模型: $Y = 2.424 + 1.496e - 5X_1 + 9.293e - 5X_2 + 2.526e - 1X_3 + 9.181e - 2X_4 + 2.321e - 1X_5 + 5.266e - 4X_6$ (其中, Y 表示评分, X_1 表示评论数, X_2 表示平均价格, X_3 表示质量等级, X_4 表示是否提供 WiFi, X_5 表示环境评级, X_6 表示客房数)。

从表 2 可知, 删减变量之后的模型的结果显示, 各因变量的 P 值都有所减小, 模型的性能提高了, 说明选择评论数、平均价格、质量等级、是否提供 WiFi、环境评级和客房数这六个变量对评分都有较强的影响。

Table 2. Estimation of optimized model parameters**表 2.** 优化后的模型参数估计表

变量	参数估计	P 值
截距	2.424e+00	<2e-16
评论数	1.496e-05	6.06e-08
平均价格	9.293e-05	0.00265
质量等级	2.526e-01	2.27e-07
是否提供 WiFi	9.181e-02	0.00272
环境评级	2.321e-01	2.13e-05
客房数	5.266e-04	0.00567

2.4. 正态性检验

在用线性回归模型拟合数据之前，首先要求数据应符合或近似符合正态分布，否则得到的拟合函数不正确。

对因变量进行正态性检验：

Table 3. Test table for normality of dependent variables**表 3.** 因变量正态性检验表

指标	酒店评分
P 值	4.53e-14

从表 3 可知因变量“评分”是正态性。

对自变量进行正态性检验：

Table 4. Test table for normality of independent variables**表 4.** 自变量正态性检验表

自变量	P 值
评论数	<2.2e-16
平均价格	<2.2e-16
质量等级	<2.2e-16
是否提供 WiFi	<2.2e-16
客房数	<2.2e-16
环境评级	<2.2e-16

从表 4 可知所有的自变量都是正态性。

对残差进行正态性检验：

Table 5. Test table for normality of residuals**表 5.** 残差正态性检验表

指标	残差
P 值	4.708e-12

从表 5 可知残差也呈现正态性。

模型优化 2: 通过逐步回归, 进行分析: 从可用变量中选择对因变量有较大影响的自变量以建立方程式, 并且对因变量没有较大影响的自变量不包括在方程中。

Table 6. Stepwise regression results

表 6. 逐步回归结果表

指标	AIC 值
-客房数	-1412.1
-是否提供 WiFi	-1410.8
-平均价格	-1410.7
-环境评级	-1401.5
-质量等级	-1392.7
-评论数	-1390.1

从表 6 的程序运行的结果可以得知, 当用全部变量作回归时, AIC 值最小。接下来显示的数据说明, 如果分别去掉其中一个自变量, AIC 值都会发生变化, 但是所有变化的 AIC 的值都大于原先全部变量回归时的 AIC 值。

分析一下计算结果, 提取相关信息:

Table 7. Parameter estimation after stepwise regression

表 7. 逐步回归后参数估计表

变量	参数估计	P 值
截距	2.424e+00	<2e-16
评论数	1.496e-05	6.06e-08
平均价格	9.293e-05	0.00265
质量等级	2.526e-01	2.27e-07
是否提供 WiFi	9.181e-02	0.00272
环境等级	2.321e-01	2.31e-05
客房数	5.266e-04	0.00567

从表 7 的运行结果来看, 回归系数的检验的显著性水平变化不大, 说明所选择的自变量都是具有较强的显著性。把全部变量作回归时, 不仅 AIC 值最小, 残差的平方和也是最小。拟合越好的方程, 残差的平方和越小。模型的 R 方值为 0.4023, 相比于之前的模型提高了。

逐步回归后的模型:

$Y = 2.424 + 1.496e - 5X_1 + 9.293e - 5X_2 + 2.526e - 1X_3 + 9.181e - 2X_4 + 2.321e - 1X_5 + 5.266e - 4X_6$ (其中, Y 表示评分, X_1 表示评论数, X_2 表示平均价格, X_3 表示质量等级, X_4 表示是否提供 WiFi, X_5 表示环境评级, X_6 表示客房数)。

3. 因子分析实证分析

3.1. 因子分析意义

因子分析是将多个实测变量转换成少数及格综合指标，体现的也是一种降维思想。通过降维将相关性高的变量聚类在一块，从而减少待分析变量的数量，进而减少问题分析的复杂程度[3]。

3.2 实证分析

1) 确定因子个数

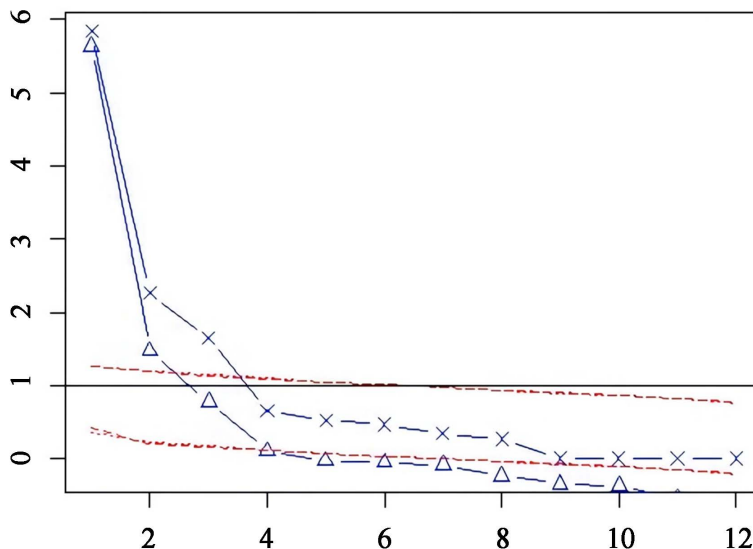


Figure 2. Gravel diagram
图 2. 碎石图

图 2 中横坐标表示因子数，纵坐标表示特征值。图 2 为帮助我们确定主成分合适个数的视觉工具，将特征值从大到小排序，选取一个拐点对应的序号，若此序号后的特征值全部较小且彼此大小差异不大，这样选取的序号作为主成分个数，依据上述图利用决定提取三个主成分。

2) 计算因子成分与贡献率

选择三个主因子，第一公共因子为酒店规模管理，第二公共因子为酒店服务管理，第三公共因子为酒店口碑管理。如下图 3 所示：

因子贡献及贡献率如表 8 所示：

由表 8 可以看出选择 3 个主成分达到 81% 的累计贡献率，能够提取到数据中的有效信息，其他主成分部分之间的贡献率较小且差异不大，可选择舍弃，因子载荷矩阵见表 9。

3) 计算因子得分

计算因子得分，并进行总得分综合。

用回归方法求各个因子得分函数：

$$F_j = \beta_{j1}x_1 + \beta_{j2}x_2 + \beta_{j3}x_3 + \beta_{j4}x_4$$

记每个样本点对因子 F_j 综合得分的估计值为： $F = \frac{30.9F_1 + 30.7F_2 + 23.9F_3}{85.5}$ 。

利用计算所得各因子指标与综合得分指标结果作图展示如下图：

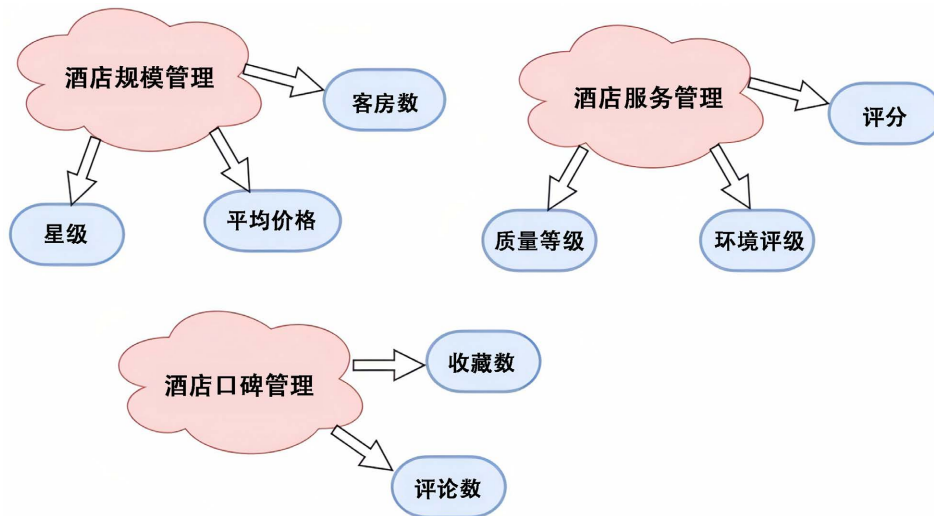


Figure 3. Diagram of three main factors
图 3. 三大主因子图

Table 8. Factor contribution rate
表 8. 因子贡献率表

因子	贡献	贡献率	累积贡献率
酒店规模管理	5.15	43.2	43.2
酒店服务管理	2.43	20.3	63.5
酒店口碑管理	2.18	17.5	81

Table 9. Factor load matrix
表 9. 因子载荷矩阵表

指标	酒店规模管理	酒店服务管理	酒店口碑管理
平均价格	0.83	0.07	-0.1
评论数	0.22	0.11	0.84
环境评级	0.03	0.90	0.01
收藏数	-0.12	0.03	0.88
质量等级	0.05	0.90	0.02
评分	0.31	0.68	0.24
客房数	0.75	0.11	0.37
星级	0.87	0.13	0.00

图 4 为酒店星级与各因子得分平均值的交叉分析，由上图可以看出，只有在酒店规模管理方面呈现明显上升趋势，而在服务管理方面与口碑方面并没有明显的区分，由此可以看出根据星级判定只是酒店规模方面有明显的区分，而并没有对酒店进行综合性区分，通过对酒店根据各因子指标与综合指标进行聚类分析显示酒店被明显区分为 3 类，且并无模棱两可的酒店类型，区分良好，以下根据上述聚类类别对酒店各因子指标与星级进行分析：

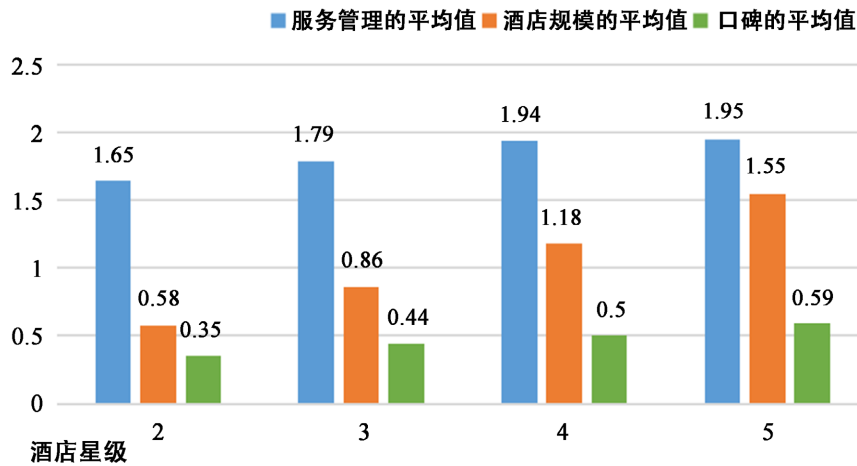


Figure 4. Factor analysis
图4. 因子分析

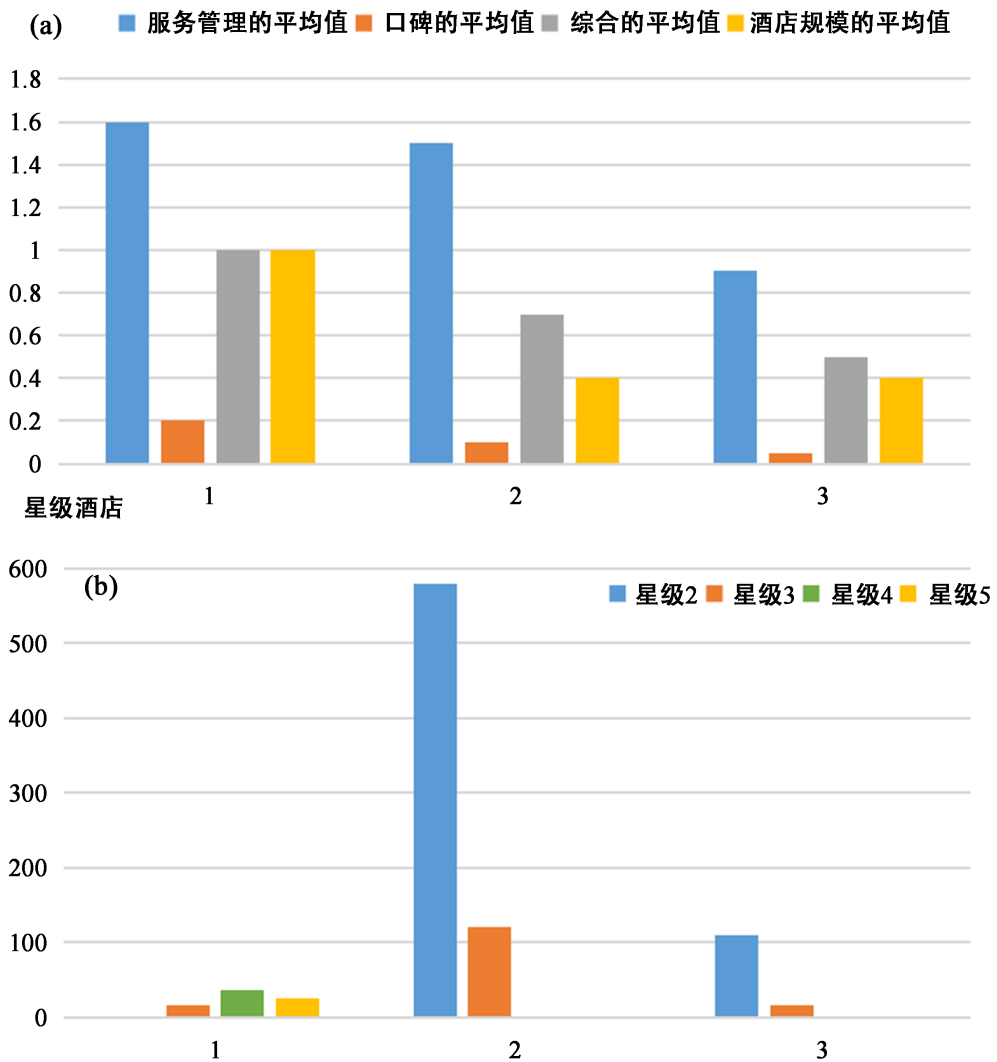


Figure 5. Relationship between stars, categories and factors
图5. 星级、类别和因子关系图

图 5(a)为酒店聚类类别与各因子指标之间的关系,可以看出,类型 1 不管从酒店服务管理方面、口碑方面、酒店规模方面还是总体综合方面都表现最优秀,而类别 2 的酒店次之,类别 3 的酒店最差。

图 5(b)可以看出一个有趣的现象,类别 1 中包含少部分三星级酒店和所有四星级、五星级酒店,说明此类酒店确实是不论从服务方面、口碑方面还是酒店规模方面都表现突出。而类别 2 中包含大部分二星级酒店与部分三星级酒店,从服务管理、口碑与酒店规模方面都表现良好,且体现出目前一部分二星级酒店可以与三星级酒店媲美,除规模方面有所区别外,服务方面与口碑方面都不亚于三星级酒店,而第 3 类则为一部分为二星级酒店与三星级酒店,应为一部分旧型经济型酒店,配备设施等并未跟随潮流或不能表现出一定的舒适性。

4. 结论

结果显示,宁波酒店应在评论数、平均价格、质量等级、是否提供 WIFI、环境评级以及客房数等方面提升,以获取较高的评分,增强自身的竞争力。

基金项目

宁波市自然科学基金(2019A610041); 国家级大学生创新创业项目(202111058033)。

参考文献

- [1] 中国宁波酒店行业调查分析及市场前景预测报告(2019~2025 年) [R]. 宁波: 中国产业调研网, 2019.
- [2] 博思数据研究中心. 2018~2023 年中国高星级酒店连锁行业市场运营状况分析及趋势预测分析报告[R]. 博思数据, 2019-07-31.
- [3] 王健华, 丁小清. 西北地区上市公司经营绩效评价——基于因子分析和聚类分析[J]. 财务与会计, 2021(1): 32-38.