

国内旅游消费水平影响因素及其统计研究

胡娜

94 软工专硕 2405 2024303053

摘要：本研究围绕国内旅游消费水平的影响因素展开探讨，基于国内旅游总花费、国民总收入、国内生产总值、居民消费水平等指标，分析其对旅游消费水平的影响。数据来源于国家统计局公开统计数据，采用描述性统计、相关分析、多元回归分析及方差分析等方法，探究各因素的显著性及作用机制。研究结果表明，居民收入水平是影响旅游消费的主要驱动因素，交通条件对消费水平也具有显著影响。本研究为优化旅游消费结构、制定相关政策提供了数据支持，并对推动国内旅游经济发展具有实际意义。

关键词：国内旅游；相关分析；主成分分析；回归分析

1 研究背景

我国旅游业的发展已成为拉动内需和促进经济增长的重要引擎。然而，与旅游人数的快速增长相比，国内旅游消费水平的提升相对滞后，不同地区和群体间的旅游消费存在显著差异。^[1]本文基于近 20 年来的国内旅游消费数据，包括居民可支配收入、教育水平、年龄分布、地区经济差异等指标，采用相关分析、多元回归分析及方差分析方法，分析影响国内旅游消费水平的主要因素，探讨提高居民旅游消费的内在驱动力。

研究我国国内旅游消费水平的影响因素，首先需从政府网站及权威统计资料中获取相关数据，然后通过描述性统计分析揭示旅游消费的基本特征，结合相关分析、多元回归分析及方差分析等方法，建立模型并量化各因素的影响程度。最终，通过对分析结果的检验，明确关键驱动因素，并提出政策建议，以促进国内旅游消费持续增长。

2 统计模型及原理

为了全面分析国内旅游消费水平的影响因素^[2]，本研究采用多种统计方法^[3]，从不同角度探索数据特征与变量关系。

2.1 相关分析

相关分析用于衡量变量之间的线性相关程度，通过计算 Pearson 相关系数构建相关矩阵。其具体步骤如下：

Step 1: 计算均值

设样本为 $X = \{X_1, X_2, \dots, X_p\}$ ，样本量为 n 。

第 j 个变量的均值为 $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, j=1, 2, \dots, p$ 。

Step 2: 计算协方差

变量 X_j 和 X_k 的协方差计算公式为：

$$Cov(X_j, X_k) = \frac{1}{n-1} \sum_{i=1}^n (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k) \quad (1)$$

Step 3: 计算相关系数

变量 X_j 和 X_k 的 Pearson 相关系数为：

$$r_{jk} = \frac{Cov(X_j, X_k)}{\sqrt{Var(X_j) \cdot Var(X_k)}} \quad (2)$$

其中， $Var(X_j) = Cov(X_j, X_j)$ 。

Step 4: 构建相关矩阵

将所有变量的相关系数排列成对称矩阵：

$$R = \begin{bmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \dots & \dots & \dots & \dots \\ r_{p1} & r_{p2} & \dots & 1 \end{bmatrix} \quad (3)$$

其中 $|r_{jk}|$ 越接近 1，两变量之间的线性相关密切程度越高，反之越低。

2.2 主成分分析

主成分分析是一种降维技术，通过线性组合提取原始变量的主要信息，减少数据的冗余性。具体步骤如下：

Step 1: 数据标准化。

标准化处理消除不同变量量纲的影响：

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{s_j} \quad (4)$$

其中 s_j 为变量 X_j 的标准差。

Step 2: 计算相关矩阵，根据标准化后的数据，构建相关矩阵 R 。

Step 3: 计算特征值和特征向量

解方程 $|R - \lambda I| = 0$ ，得到特征值 $\lambda_1, \lambda_2, \dots, \lambda_p$ 及

对应的特征向量 a_1, a_2, \dots, a_p 。

Step4:提取主成分

选择特征值大于 1 或者累计贡献率达到 85%以上

的主成分。每个主成分 F_k 表示为：

$$F_k = \sum_{j=1}^p a_{kj} Z_j \quad (5)$$

Step5:主成分解释

根据主成分在原变量上的载荷确定其实际意义。

2.3 回归分析

回归分析用于研究因变量与自变量间的定性关系，采用最小二乘法估计回归系数。具体步骤如下：

Step 1: 模型建立

设因变量为 Y ，自变量为

$X = [X_1, X_2, \dots, X_p]$ 。线性回归模型表示为：

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$ ，其中

$\beta_0, \beta_1, \dots, \beta_p$ 为待估参数， ε 为随机误差项。

Step 2: 参数估计

使用最小二乘法，求解误差平方和最小的回归系数：

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (6)$$

Step 3: 模型检验

1. 拟合优度检验：

拟合优度 R^2 衡量模型解释因变量变异的能力：

$$R^2 = 1 - \frac{SSR}{SST} \quad (7)$$

其中，

$$SSR = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (9)$$

2. 显著性检验

对回归方程整体显著性进行 F 检验：

$$F = \frac{MSR}{MSE} \quad (10)$$

其中，

$$MSR = \frac{SSR}{p}, MSE = \frac{SSE}{n - p - 1} \quad (11)$$

3. 系数显著性检验

对每个回归系数进行 t 检验：

$$t = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (12)$$

其中， $SE(\hat{\beta}_j)$ 为 $\hat{\beta}_j$ 的标准误差。

Step 4: 多重共线性检验

计算方差膨胀因子 (VIF)：

$$VIF_j = \frac{1}{1 - R_j^2} \quad (13)$$

其中 R_j^2 为解释变量 X_j 对其余变量的拟合优

度。

3 数据收集及预处理

3.1 数据来源

本文数据来源于多种权威渠道，包括国家统计局、文化和旅游部统计年鉴及 EPS 公开数据库^[3-4]。选取了 2004 年至 2023 年间国内与旅游消费相关的主要指标，如国内生产总值 (GDP)、居民人均可支配收入、交通基础设施指数、旅游资源开发水平和国内旅游消费总量等。

这些数据从经济能力、基础设施和消费行为等方面全面反映了影响旅游消费的潜在因素。其中，国内旅游消费总量是衡量居民旅游消费行为的核心指标，也是本研究的因变量；居民人均可支配收入则反映了居民的实际经济能力，是推动旅游消费的关键驱动因素。交通基础设施指数体现了旅游目的地的可达性，直接影响居民的出行决策，而旅游资源开发水平则从景区吸引力的角度影响消费行为。此外，地区 GDP 是评估区域经济能力对旅游消费支持的重要指标。

3.2 数据预处理

为了确保数据的科学性和可用性，需对原始数据进行预处理。首先，针对缺失值和异常值进行了数据清洗。缺失值处理采用均值填补法或插值法，对于关键变量，通过查阅补充数据来源予以完善；异常值则通过箱线图和 Z-score 方法识别，对其进行修正或剔除，确保数据的完整性和准确性。其次，对不同指标进行了标准化处理，以消除量纲差异。

4 基于统计方法的数据分析

4.1 相关分析

利用 SPSS 生成变量的 Pearson 相关矩阵:

表 1 各指标间的 Pearson 相关矩阵

	X_1	X_2	X_3	X_4	X_5	X_6
X_1	1	1.00	0.999	0.999	0.996	0.759
X_2		1	0.999	0.999	0.996	0.756
X_3			1	0.999	0.997	0.772
X_4				1	0.995	0.760
X_5					1	0.772
X_6						1

其中, X_1 表示国民总收入, X_2 表示国内生产总值, X_3 表示居民消费水平, X_4 表示居民人均可支配收入, X_5 表示交通运输、仓储和邮政业增加值、 X_6 表示国内旅游总花费。

结果显示居民消费水平、交通运输、仓储和邮政业增加值与国内旅游总花费呈显著正相关, 相关系数为 0.772, 均在 0.01 水平显著。指标间相关系数都较大, 各指标在 0.01 级别(双尾)相关性都是显著的。

4.2 主成分分析

通过相关分析可知, 变量间相关系数较大(绝对值超过 0.5), 适合进行主成分分析。

具体的主成分分析结果如下:

表 2 主成分分析结果

主成分	特征值	方差贡献率	累计方差贡献率
X_1	5.91869	93.712683%	93.71%
X_2	0.38909	6.160666%	99.87%
X_3	0.00612	0.096997%	99.97%
X_4	0.00153	0.024361%	99.99%
X_5	0.000323	0.005114%	99.99%
X_6	0.000011	0.000179%	100%

本文首先通过特征值大于 1 的标准提取了 2 个主成分, 累计贡献率为 99.87%。主成分如下:

$$F_1 = 0.4194X_1 + 0.4192X_2 + 0.4204X_3 + 0.4194X_4 + 0.4194X_5 + 0.3466X_6$$

$$F_2 = -0.1688X_1 - 0.1747X_2 - 0.1338X_3 - 0.1653X_4 - 0.1315X_5 + 0.9372X_6$$

通过旋转成分矩阵明确主成分含义:

主成分 1(社会经济因素)解释了 93.71%的数据变异, 是最主要的影响因素。它主要由国民总收入、国内生产总值、居民人均可支配收入和居民消费水

平等变量主导, 反映了社会经济发展对国内旅游消费的决定性作用。更高的居民收入水平和经济能力通常会显著提升旅游消费。

主成分 2(行业发展与外部支持因素)解释了 6.16%的数据变异, 主要由交通运输、仓储和邮政业增加值及国内旅游总花费变量主导, 体现了交通和行业发展对旅游消费的支持作用。交通便利性和服务业的提升在一定程度上增强了旅游体验和消费意愿。

主成分分析表明, 社会经济因素和交通与资源因素是影响国内旅游消费的主要维度。

4.3 回归分析

4.3.1 回归模型的建立

以国民总收入、国内生产总值、居民消费水平、居民人均可支配收入、交通运产业增值作为自变量, 国内旅游总花费作为因变量, 设置显著性水平为 0.05, 用 SPSS 软件进行线性回归分析, 得到相关系数 $R=0.872$, 因此可以初步判断该模型的拟合效果较好, 可以接受。

4.3.2 最小二乘法估计回归参数

以通过主成分分析提取得到的主成分作为自变量, 我国国内旅游总花费标准化后的数据作为因变量, 建立回归模型。

通过最小二乘法, 得到回归参数如表 3 所示。

表 3 回归参数

自变量	回归参数
常量值	-5088.27
主成分 F_1	3.12
主成分 F_2	-3.05

4.3.3 模型检验

基于 SPSS 的计算, 模型拟合优度为 87.2%, 拟合度较好, 提取出的主成分能较大程度地描述我国国内居民旅游总花费的变化。

计算出的模型的 F 值及其显著性如表 4 所示。

表 4 ANOVA 计算结果

	平方和	自由度	均方	F 值	显著性
SR	13.340	4	3.335	8.839	0.000
SE	5.660	15	0.377		
总计	19.000	19			

根据表 4, 模型的显著性为 0.000, 因此拒绝原假设, 认为我国国内居民旅游总花费关于各主成分存在显著的线性回归关系。

4.3.4 系数检验

通过 SPSS 计算回归参数的显著性如表 5 所示。

根据表 5，回归参数在显著水平 $\alpha = 0.05$ 下，除常数项外，其余各回归参数均不显著为零。

表 5 回归参数的显著性

自变量	显著性
常量	1.00
主成分 F_1	0.00
主成分 F_2	0.00

4.3.4 残差分析

使用 SPSS 软件对残差图进行正态性检验，结果如下图所示：

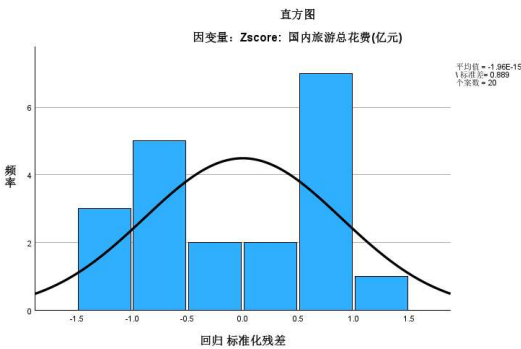


图 1 残差直方图

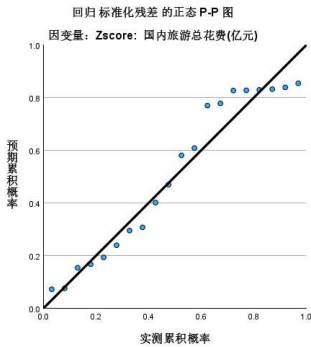


图 2 实测累积概率图

首先，从直方图来看，残差大致服从正态分布；其次，从 P-P 图来看，点基本上都围绕线的周围，于是可以认为残差是服从正态分布的。

5 结论与建议

5.1 研究结论

统计结果综合分析显示，居民人均可支配收入、国民总收入和国内生产总值对旅游消费的贡献最为显著。经济发展和收入水平的提高，直接决定了居民旅游消费的能力和意愿。这表明，提升居民收入是促进国内旅游消费增长的根本手段。

研究结果表明，社会经济发展和行业基础设施建设是驱动国内旅游消费水平提高的两个核心动力。未来政策的重点应放在促进居民收入增长、优化交通基础设施、提升旅游服务质量以及加强区域

协调发展，以推动国内旅游消费的可持续增长。这些措施将不仅促进旅游行业的发展，还将成为拉动内需、稳定经济增长的重要力量。

5.2 建议

基于研究结果，提出以下政策建议以促进国内旅游消费：

第一、提升居民可支配收入。通过优化收入分配机制，提高低收入群体的消费能力，以扩大旅游市场规模。

第二、加快交通基础设施建设。继续推进高速铁路、机场等重点交通项目，提升旅游目的地的可达性，为旅游消费创造便利条件。

第三、发展旅游服务业。加强旅游资源开发，提高景区服务质量和多样性，吸引更多消费者。

第四、增强政策支持力度。针对旅游相关行业提供财政补贴与税收优惠，尤其是促进偏远地区的旅游发展

参考文献

- [1] 刁宗广，张涛. 中国城乡居民国内旅游消费水平和消费结构比较研究[J]. 人文地理，2010，25(2)：158-160, 143.
- [2] 吴晋峰，石晓腾，贾薇. 中国城市居民旅游活跃度研究[J]. Tourism Tribune/Lvyou Xuekan, 2024, 39(7).
- [3] 王传美，王展青. 应用数理统计[M]. 武汉理工大学出版社，2021.
- [4] EPS 数据库 <http://olap.epsnet.com.cn/> 2018.12.1.
- [5] 国家统计局 <http://data.stats.gov.cn/> 2018.12.1