

线性回归方法在北京市居民消费水平影响因素分析中的应用

张琪, 牟唯嫣, 刘程馨

北京建筑大学理学院, 北京

收稿日期: 2023年9月17日; 录用日期: 2023年10月8日; 发布日期: 2023年10月23日

摘要

本文应用线性回归方法探讨北京市居民消费水平的影响因素, 通过收集北京市各类数据, 建立多元线性回归模型进行分析, 识别出对北京市居民消费水平具有显著影响的核心因素。结果表明, 北京市居民消费水平与人均可支配收入(元)以及旅游业国内旅游收入(亿元)密切相关。同时, 本文收集到了不同时间段内的影响因素的数据进行分析, 揭示了北京市近年来的居民消费水平发展趋势。通过本文可以更好地认识到北京市经济发展的未来走向, 分析了其相关性和复共线性可以为政府和有关单位提供理论支撑。

关键词

线性回归, 消费水平, 复共线性

Application of Linear Regression Method in the Analysis of Influencing Factors of Beijing Residents' Consumption Level

Qi Zhang, Weiyan Mu, Chengxin Liu

School of Science, Beijing University of Civil Engineering and Architecture, Beijing

Received: Sep. 17th, 2023; accepted: Oct. 8th, 2023; published: Oct. 23rd, 2023

Abstract

In this paper, the linear regression method is applied to explore the factors affecting the consumption level of Beijing residents. Through collecting all kinds of data in Beijing, a multiple linear regression model is established for analysis, and the core factors that have a significant im-

pact on the consumption level of Beijing residents are identified. The results show that the consumption level of Beijing residents is closely related to the per capita disposable income (RMB) and the domestic tourism income (RMB 100 million). At the same time, this paper collected the data of influencing factors in different time periods for analysis, revealing the development trend of Beijing residents' consumption level in recent years. Through this paper, we can better understand the future trend of Beijing's economic development, and analyze its correlation and complex collinearity, which can provide theoretical support for the government and national units.

Keywords

Linear Regression, Consumption Level, Complex Collinearity

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

随着我国经济的发展，人们的消费水平也不断提升，对于国家和政府而言，为了做出更好的决策，需要分析出影响人民的消费水平的因素，北京作为中国的首都和经济中心城市，其居民的消费水平及其相关影响因素一直备受关注。

线性回归是一种广泛应用的数据分析方法，本文通过对北京市居民的消费水平及其影响因素进行详细的实证分析，希望能为政府和国家提供参考和价值，以促进北京市经济的发展。

本文首先介绍了线性回归的理论基础和应用范围，基于收集到的旅游业国内旅游收入(亿元)社会消费品零售总额(亿元)人均可支配收入(元)常住人口(万人)全市各类学校数(个)医疗卫生机构(个)境内道路总里程(公里)居民消费水平(元)一般公共预算收入(亿元)在内的实证数据，我们建立了一个多元线性回归模型，并采用 SPSS 软件对数据进行处理和回归分析。

2. 变量和原始数据

2.1. 解释各变量

见表 1，本文选取的自变量有：旅游业国内旅游收入(亿元)社会消费品零售总额(亿元)人均可支配收入(元)常住人口(万人)全市各类学校数(个)医疗卫生机构(个)境内道路总里程(公里)居民消费水平(元)一般公共预算收入(亿元)，因变量为北京市居民消费水平(元)。

Table 1. Indicators of independent variables

表 1. 自变量指标

自变量	名称	自变量	名称
x_1	旅游业国内旅游收入(亿元)	x_5	全市各类学校数(个)
x_2	社会消费品零售总额(亿元)	x_6	医疗卫生机构(个)
x_3	人均可支配收入(元)	x_7	境内道路总里程(公里)水平(元)
x_4	常住人口(万人)	x_8	一般公共预算收入(亿元)

2.2. 数据来源

见表 2, 本文选取的数据是 2005~2020 年十六年内的数据, 因变量为北京市居民消费水平(元), 自变量为旅游业国内旅游收入(亿元)、社会消费品零售总额(亿元)、人均可支配收入(元)、常住人口(万人)、全市各类学校数(个)、医疗卫生机构(个)、境内道路总里程(公里)、一般公共预算收入(亿元)。数据的来源为《北京统计年鉴 2005~2020》。

Table 2. Data of influencing factors of Beijing residents' consumption level (yuan) from 2005 to 2020

表 2. 2005~2020 年北京市居民消费水平(元)影响因素数据

年份	北京市居民消费水平(元)	旅游业国内旅游收入(亿元)	社会消费品零售总额(亿元)	人均可支配收入(元)	常住人口(万人)	全市各类学校数(个)	医疗卫生机构(个)	境内道路总里程(公里)	一般公共预算收入(亿元)
2005	16636	1300.0	3221	16853	1538.0	3782	7536	19015	919.21
2006	19003	1482.7	3673	19296	1601.0	3751	7640	25377	1117.15
2007	21769	1753.6	4307	21458	1676.0	3593	9023	25765	1492.64
2008	24005	1907.0	5258	24371	1771.0	3508	9647	26921	1837.32
2009	26353	2144.5	6140	26571	1860.0	3425	9717	27436	2026.81
2010	29483	2425.1	7273	29228	1961.9	3330	9511	27907	2353.93
2011	32960	2864.3	8335	33176	2023.8	3367	9699	28446	3006.28
2012	36076	3301.3	9440	36817	2077.5	3314	9974	28585	3314.93
2013	40313	3666.3	10383	40830	2125.4	3439	10141	28808	3661.11
2014	42918	3997.0	11354	44489	2171.1	3437	10265	29209	4027.16
2015	46423	4320.0	12272	48458	2188.3	3454	10425	29069	4723.86
2016	50195	4683.0	13135	52530	2195.4	3524	10637	29282	5081.26
2017	54965	5122.4	13934	57230	2194.4	3556	10986	29463	5430.79
2018	61458	5556.2	14422	62361	2191.7	3585	11100	29429	5785.92
2019	66348	5866.2	15064	67756	2190.1	3640	11340	29531	5817.10
2020	62842	2880.9	13716	69434	2189.0	3799	11211	29537	5483.89

2.3. 标准化处理数据

标准化处理数据(见表 3)的优势有:

1) 消除量纲影响: 不同量纲单位会带来很多不便, 如果一个变量的数值范围很大, 另一个变量的数值范围很小, 那么这种不同量纲的变量就不能直接参与数据分析和比较。通过标准化让所有变量都处在一个数量级上, 增加了计算的准确性, 能更好的分析预测结果。

2) 增强稳定性: 通过数据的标准化能够增强模型的稳定性, 还能降低离群点的影响。如果不进行标准化处理, 在有异常值的情况下, 模型会受到异常值的干扰, 使得运算结果不可靠。

3) 提高效率: 数据标准化处理后会大幅提高计算机的运行效率, 标准化之后数据都变成了 0 附近的数值, 可以提升算法的收敛速度, 从而更快地建模。

4) 方便对比不同变量: 标准化处理数据方便后期的评价和比较。

Table 3. Standardized raw data
表 3. 标准化原始数据

年份	北京市居民 消费水平 (元)	旅游业国内 旅游收入 (亿元)	社会消费品 零售总额 (亿元)	人均可支配 收入(元)	常住人口 (万人)	全市各类 学校数(个)	医疗卫生 机构(个)	境内道路 总里程 (公里)	一般公共 预算收入 (亿元)
2005	-1.40149	-1.3754	-1.52571	-1.36332	-1.96074	1.63544	-2.11618	-3.26059	-1.48541
2006	-1.2563	-1.25158	-1.4158	-1.22353	-1.69171	1.43305	-2.02418	-0.88205	-1.3717
2007	-1.08664	-1.06798	-1.26164	-1.09982	-1.37144	0.40151	-0.80078	-0.73698	-1.156
2008	-0.94948	-0.96401	-1.03039	-0.93314	-0.96577	-0.15342	-0.24879	-0.30479	-0.95799
2009	-0.80546	-0.80305	-0.81592	-0.80725	-0.58572	-0.69531	-0.18687	-0.11225	-0.84914
2010	-0.61347	-0.61288	-0.54042	-0.65521	-0.15058	-1.31553	-0.3691	0.06384	-0.66122
2011	-0.40019	-0.31522	-0.28218	-0.42931	0.11375	-1.07397	-0.20279	0.26535	-0.28647
2012	-0.20906	-0.01905	-0.01348	-0.22097	0.34306	-1.41999	0.04047	0.31732	-0.10916
2013	0.05084	0.22832	0.21582	0.00866	0.54761	-0.6039	0.1882	0.40069	0.0897
2014	0.21063	0.45245	0.45193	0.21803	0.74276	-0.61696	0.29789	0.55061	0.29998
2015	0.42562	0.67136	0.67516	0.44514	0.8162	-0.50597	0.43942	0.49827	0.70021
2016	0.65699	0.91738	0.88501	0.67815	0.84652	-0.04897	0.62696	0.5779	0.90552
2017	0.94958	1.21517	1.0793	0.94708	0.84225	0.15995	0.93568	0.64557	1.10631
2018	1.34786	1.50917	1.19796	1.24068	0.83072	0.34928	1.03653	0.63286	1.31032
2019	1.64781	1.71927	1.35407	1.54939	0.82389	0.70836	1.24883	0.671	1.32823
2020	1.43275	-0.30397	1.02629	1.64541	0.81919	1.74642	1.13472	0.67324	1.13681

3. 模型的建立与研究[1]

含有 $p - 1$ 个自变量的多元线性回归模型的一般形式为:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{p-1} x_{p-1} + e$$

本文选取了 8 个影响因素作为自变量, 即本文的线性回归模型为:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + e$$

其中 y 为——北京市居民消费水平(元), β_0 为常数变量, $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9$ 为回归系数, x_1 ——旅游业国内旅游收入(亿元), x_2 ——社会消费品零售总额(亿元), x_3 ——人均可支配收入(元), x_4 ——常住人口(万人), x_5 ——全市各类学校数(个), x_6 ——医疗卫生机构(个), x_7 ——境内道路总里程(公里), x_8 ——一般公共预算收入(亿元), e 为其他因素产生的不可控的误差。

3.1. 线性模型的数据检验

1) 正态性检验

见图 1, 几乎所有的点都在直线附近, 这表明基本符合正态性假设。

2) 独立性检验

德宾 - 沃森指数(Durbin-Watson Statistic)的取值范围在 0 到 4 之间。当德宾 - 沃森指数的值为 2 时, 表示样本数据不存在自相关性, 即样本数据是相互独立的。当德宾 - 沃森指数小于 2 时, 表明样本数据存在正自相关性; 当德宾 - 沃森指数大于 2 时, 表明样本数据存在负自相关性。德宾 - 沃森指数值在 1.5

到 2.5 之间时, 可以认为数据是独立的。见表 4 可知德宾 - 沃森指数值为 2.465, 认为样本间相互独立, 数据分析有效。

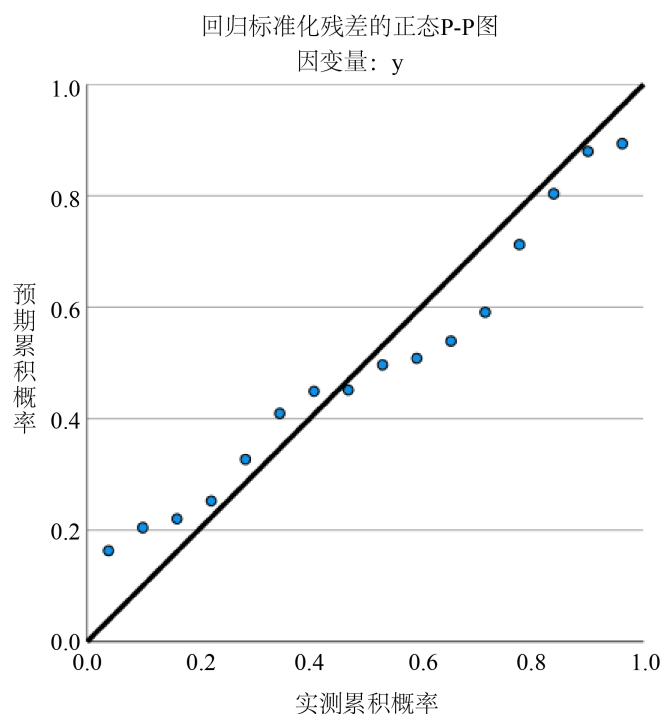


Figure 1. Results of least squares estimation

图 1. 最小二乘估计结果

Table 4. Independence test results

表 4. 独立性检验结果

模型	R	R 方	调整后 R 方	标准估算的错误	德宾 - 沃森
1	1.000 ^a	1	1	0.02117	2.465

R 方是一种模型拟合程度的指标, 也是回归模型的拟合优度评估指标。R 方取值范围为 0 到 1, 数值越接近 1, 模型的可解释能力就越好, 对样本数据的拟合程度越高; 反之, 值越接近 0, 则表示模型拟合预测效果越差。

表 4 中 R 方为 1 表示接近 1, 意味着这个模型能够对因变量(北京市生产总值)有很高的解释。

3) 残差分布检验

见图 2 可以看出, 残差图呈现出一种近似随机分布, 这些残差可以被认为是服从 $N(0,1)$ 的一组简单随机样本, 最少有 95% 的点落在 $[-2, 2]$ 中, 才能保证 Y 与残差 e 不相关。在残差图中, 我们可以发现点都分布在 $[-2, 2]$ 区域内, 且整体呈现随机性, 所以我们认为假设 $e \sim N(0, \sigma^2 I)$ 基本上是合理的, 线性模型可以较好地捕捉了数据变化。

3.2. 线性模型的显著性检验

见表 5 可知线性回归的显著性远小于 0.05, 故建立线性回归是合理的。也就是说存在至少一个自变量是可以有效拟合因变量的。

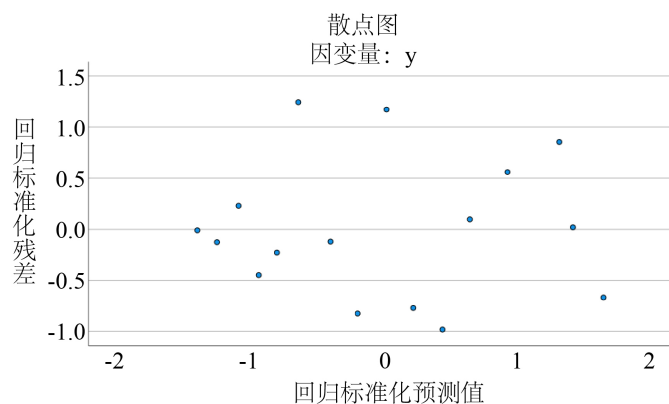


Figure 2. Residual distribution

图 2. 残差分布

Table 5. Significance test

表 5. 显著性检验

模型	平方和	自由度	均方	F	显著性
1 回归	14.997	8	1.875	4182.297	0.000 ^b
残差	0.003	7	0		
总计	15	15			

3.3. 相关系数矩阵

相关系数矩阵中的每个位置的数值都在-1 到 1 之间，每个位置的元素表示该行和该列元素之间的相关系数；当数值为负时，表示对应的两个变量呈现负相关关系，当数值为正时，表示对应的两个变量呈现正相关关系；当数值为 0 时，表示对应的两个变量没有线性相关关系，见表 6 可知大部分的数据都在 0.8 和 0.9 附近。

Table 6. Correlation matrix

表 6. 相关性矩阵

相关系数	y	x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈
y	1	0.887	0.979	0.998	0.884	0.106	0.914	0.727	0.986
x ₁	0.887	1	0.929	0.859	0.864	-0.132	0.835	0.702	0.921
x ₂	0.979	0.929	1	0.973	0.95	-0.048	0.927	0.779	0.996
x ₃	0.998	0.859	0.973	1	0.877	0.138	0.907	0.718	0.982
x ₄	0.884	0.864	0.95	0.877	1	-0.323	0.926	0.872	0.928
x ₅	0.106	-0.132	-0.048	0.138	-0.323	1	-0.203	-0.432	0.009
x ₆	0.914	0.835	0.927	0.907	0.926	-0.203	1	0.866	0.924
x ₇	0.727	0.702	0.779	0.718	0.872	-0.432	0.866	1	0.761
x ₈	0.986	0.921	0.996	0.982	0.928	0.009	0.924	0.761	1

3.4. 方差膨胀因子(Vif)

方差膨胀因子是检验复共线性的重要参考因素。在线性回归分析中，如果复共线性值较大，说明模型自变量之间存在有较高的相关性。也就是说存在有至少一个自变量间有较高的相关性，而较高的复共线性会严重干扰模型的判断，而我们的标准为：当 $0 < VIF < 10$ ，不存在多重共线性；当 $10 \leq VIF < 100$ ，存在较强的多重共线性；当 $VIF \geq 100$ ，存在严重多重共线性。

见表 7 我们可以知道前三个自变量对北京市居民消费水平存在显著影响，但其存在严重的复共线性，由此我们引出以下估计。

Table 7. Least square linear estimation

表 7. 最小二乘线性估计

模型		未标准化系数		标准化系数	t	显著性	共线性统计	
		B	标准错误	Beta			容差	VIF
1	(常量)	7.96E-07	0.005		0	1		
	x_1	0.302	0.057	0.302	5.345	0.001	0.009	106.898
	x_2	-0.887	0.304	-0.887	-2.918	0.022	0	3092.554
	x_3	1.386	0.1	1.386	13.832	0	0.003	336.193
	x_4	0.235	0.125	0.235	1.877	0.103	0.002	524.83
	x_5	-0.032	0.027	-0.032	-1.148	0.289	0.04	25.205
	x_6	-0.038	0.027	-0.038	-1.398	0.205	0.04	24.983
	x_7	-0.025	0.017	-0.025	-1.494	0.179	0.106	9.453
	x_8	0.067	0.112	0.067	0.598	0.569	0.002	421.163

4. 最小二乘估计[2]

最小二乘估计(Least Square Estimation)是数据分析中一种常见的参数估计方法。它的基本思想是：对于我们所收集数据，构建一个数学模型，使得线性模型的预测值和真实数据之间的误差平方和最小。这个过程就称为“最小二乘法”。

下面介绍最小二乘法的原理和方法：

假如我们有一组收集到数据 $(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)$ 。

我们希望找到一个函数 $f(x; \theta)$ ，其中 θ 是参数向量，使得对于每一个预测值 (x_i, y_i) 有 $y_i \approx f(x_i; \theta)$ 。

同时定义误差 e_i 为观测值 y_i 与预测值 $f(x_i; \theta)$ 之间的差，即 $e_i = y_i - f(x_i; \theta)$ 。

最小二乘法的基本思想就是通过寻找一组参数 θ ，使所有误差的平方和 $S(\theta) = \sum_{i=1}^n e_i^2$ 最小。也就是说需要解决以下的优化问题：

$$\hat{\theta} = \arg \min_{\theta} S(\theta) = \arg \min_{\theta} \sum_{i=1}^n (y_i - f(x_i; \theta))^2$$

为了解决上述问题，我们可以采取最大似然估计的思想。假设我们的误差 e_i 独立同分布且服从正态分布 $N(0, \sigma^2)$ ，则误差平方和 $S(\theta)$ 的最小化等价于对数似然函数 $\ln L(\theta)$ 的最大化，其中：

$$\ln L(\theta) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n e_i^2$$

最大化似然函数 $\ln L(\theta)$ 可以通过求其梯度为零的算法得到最优参数 $\hat{\theta} = (x^T x)^{-1} x^T y$ 。

其中, x 是 $n \times m$ 的设计矩阵, 其中的第 i 行代表第 i 个原始数据的各个特征(如 $x_i^0 = 1, x_i^1, x_i^2, \dots$), y 是 $n \times 1$ 的向量, 代表各个原始数据的输出值。

在得到最优参数 $\hat{\theta}$ 后, 我们可以来预测未来数据观测值。如果我们采集到一个新的样本数据 x_{new} , 则其预测输出为 $\hat{y}_{new} = f(x_{new}; \hat{\theta})$ 。

由表 7 可以建立最小二乘模型:

$$y = 0.302x_1 - 0.887x_2 + 1.386x_3 + 0.235x_4 - 0.032x_5 - 0.038x_6 - 0.025x_7 + 0.067x_8$$

由表 7 可知前三个自变量对北京市居民消费水平存在显著影响, 其余自变量影响不显著, 所以我们剔除了显著性大于 0.05 的自变量。

调整后的普通最小二乘模型为:

$$y = 0.302x_1 - 0.887x_2 + 1.386x_3$$

5. 逐步回归法: [3]

逐步回归法是一种常见的变量选择方法, 它适用于在给定的自变量集合中, 确定具体有哪些自变量对于因变量的预测有重要影响。逐步回归法的本质是通过逐步添加和减少某些自变量, 并通过比较每个步骤生成的模型的性能指标作为模型的选择的依据。

这是逐步回归法的基本步骤:

步骤一: 设置一个初始模型

首先建立一个初始的空模型。

步骤二: 自变量的选择

对于选择的自变量, 通过一定的选取准则(如 p 值等)进行评估, 筛选有显著影响的自变量加入到模型中。

步骤三: 逐步添加和删除

每个步骤当中, 根据我们所定义的准则, 逐步添加和减少某些自变量, 本文采取的是步进的方法, 其本质是通过前进法(每步只添加一个自变量)或后退法(每步删除一个自变量)筛选变量。

步骤四: 模型修正与评估

对于每次增加或删除自变量后拟合出来的模型, 通过比较模型的性能指标(R^2 、AIC/BIC 等)选择最优模型。

步骤五: 终止准则

我们设定了一个终止准则, 当达到终止准则时停止。例如当没有更多自变量可加入或删除时停止筛选。

逐步回归法能帮我们选择出对因变量有重要影响的自变量, 建立有较高预测能力的回归模型, 通过添加和减少自变量, 提高模型的拟合效果, 然而逐步回归法也在一些问题, 如果有复共线性较高的自变量, 会降低模型的稳定性。

见表 8 可以建立逐步回归法模型为

$$y = 0.196x_1 - 0.340x_2 + 1.167x_3 - 0.046x_5$$

6. 主成分回归[4]

主成分回归的本质是将主成分分析与线性回归结合, 对于高度相关的自变量的多元回归问题很有优势。

Table 8. Stepwise regression method
表 8. 逐步回归法回归

模型		未标准化系数		标准化系数	t	显著性	共线性统计	
		B	标准错误	Beta			容差	VIF
6	(常量)	3.21E-07	0.006		0	1		
	x_3	1.167	0.056	1.167	21.003	0	0.012	82.021
	x_1	0.196	0.021	0.196	9.176	0	0.083	12.094
	x_5	-0.046	0.011	-0.046	-4.082	0.002	0.302	3.31
	x_2	-0.34	0.068	-0.34	-4.97	0	0.008	124.601

主成分分析(PCA):

首先对自变量进行主成分分析,降低自变量的维数,从而得到较小的主成分个数。

主成分分析的本质是对原始变量重新进行线性组合,对选取的主成分进行分析。

线性回归模型:

其次,主成分作为因变量,利用线性回归建立模型。

对于选定的因变量,使用 PCR 通过选择适当的主成分个数,构建新的模型。

PCR 的步骤如下:

数据标准化:

对原始数据进行标准化处理,保证其均值为 0,方差为 1。

主成分分析:

对标准化后的自变量进行降维,提取主成分。

主成分选择:

选择解释程度最高的主成分数量。

回归模型建立:

筛选出来的主成分作为因变量,建立线性回归模型。

模型评估:

选择最佳的主成分数目,同时对 PCR 模型进行评估。

主成分回归的优势是:可以用来解决复共线性问题以及高维数据的回归问题,通过减少或者增加自变量的维度消除共线性,提高了模型的预测效果。

见图 3 显示前两个主成分的特征值接近 1 以上,说明取前两个主成分(见表 9)。

Table 9. Principal component coefficients
表 9. 主成分系数

	成分	
	1	2
x_1	0.921	0.061
x_2	0.986	0.135
x_3	0.946	0.306
x_4	0.972	-0.154

Continued

x_5	-0.171	0.973
x_6	0.964	-0.05
x_7	0.863	-0.348
x_8	0.979	0.189

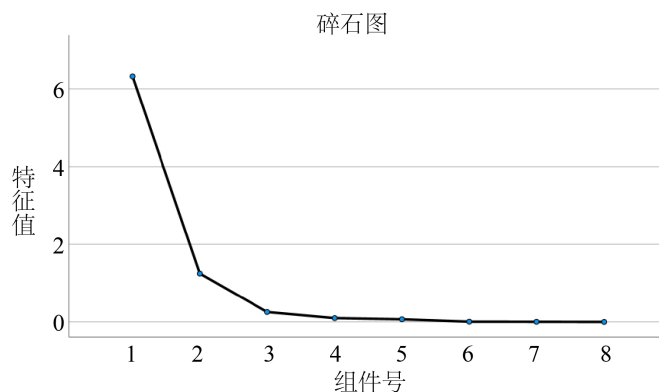


Figure 3. Lithotripsy and principal component coefficients
图 3. 碎石图和主成分系数

两个主成分展示:

$$Z_1 = 0.921x_1 + 0.986x_2 + 0.946x_3 + 0.972x_4 - 0.171x_5 + 0.964x_6 + 0.863x_7 + 0.979x_8$$

$$Z_2 = 0.061x_1 + 0.135x_2 + 0.306x_3 - 0.154x_4 + 0.973x_5 - 0.050x_6 - 0.348x_7 + 0.189x_8$$

其中

$$y = 0.956Z_1 + 0.279Z_2$$

故而, 主成分分析法得出的模型为:

$$y = 0.897x_1 + 0.980x_2 + 0.990x_3 + 0.886x_4 + 0.108x_5 + 0.908x_6 + 0.728x_7 + 0.989x_8$$

7. 分析与结论

经过对北京市居民消费水平以及影响因素的分析, 使用最小二乘回归模型和逐步回归法得出了旅游业国内旅游收入与人均可支配收入对北京市居民消费水平存在显著正相关性的结论, 且与人均可支配收入显著正相关, 这与常识相吻合, 但社会消费品零售总额与北京市居民消费水平呈负相关, 与常识不相吻合, 我认为是本文中建模不够准确, 所使用的数据量较小造成的失误, 还可以进一步改善。

我们发现最小二乘线性运算简单, 但其回归拥有较高的复共线性以及有多余的弱相关的自变量, 逐步回归法选取了高相关的自变量, 但仍然有着一定的复共线性, 综合三个模型我们发现, 人均可支配收入与旅游业国内收入出现在所有模型当中, 故我们可以认为人均可支配收入、旅游业国内旅游收入与北京市居民消费水平呈正相关。

旅游业国内旅游收入与北京市居民的消费水平呈正相关。我们知道北京市的旅游业非常发达, 有完善的旅游配套设施, 吸引了很多人选择在北京市进行旅游和观光, 如果政府可以大力继续发展北京市的旅游业, 相信会推动北京市居民的消费水平提升。

人均可支配收入与居民消费水平呈现显著正相关。人均可支配收入是比较家庭经济能力的重要指标，会直接影响居民的消费水平。人均可支配收入越高，居民的消费水平也会提升，人均可支配收入越低，居民的消费水平也会降低，因此，如果政府可以提升人均可支配收入也会导致北京市居民的消费水平提升。

综上所述，旅游业国内旅游收入和人均可支配收入对北京市居民的消费水平都有积极的影响，从而影响提高北京市居民消费水平，这一结论对于了解北京市居民的消费行为具有重要意义。

参考文献

- [1] 王燕,戴力辉,曹晓婧. 失能老年人护理模式需求影响因素的多层线性模型研究[J]. 卫生软科学, 2023, 37(4): 44-47.
- [2] 张欣,赵俊龙. 高维线性模型的影响点诊断[J]. 北京师范大学学报(自然科学版), 2023, 59(2): 313-318.
- [3] 席宇飞,康玫,杨英,等. 基于逐步回归方法的心外科住院抗菌药物用药指标趋势预测模型研究[J]. 实用药物与临床, 2021, 24(9): 804-808.
- [4] 梁钰梅,李可,林籽汐,等. 基于主成分分析法的9个品种蓝莓品质评价[J/OL]. 食品工业科技: 1-16. <https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C45S0n9fL2suRadTyEVl2pW9UrhTDCdPD65Gwj5UxcG6fj7Gt0y4dSExRqk-zhDE77GtkmJf5fmBfuRPULX8aOuQ&uniplatform=NZKPT>, 2023-09-18.