

基于R语言的我国城镇居民消费水平研究

刘 杰

亚信科技(中国)有限公司, 北京

收稿日期: 2024年7月1日; 录用日期: 2024年7月22日; 发布日期: 2024年8月5日

摘 要

本文针对我国城镇居民人均可支配收入及消费支出开展研究,选取了国家2020年的统计年鉴数据,利用RStudio软件对数据进行采集、处理与分析,主要开展了描述性分析及线性回归分析。描述性分析得出全国31个省级区域人均可支配收入和人均消费支出的构成、占比与差异性;回归分析构建了可支配收入与消费支出之间的散点图,分析二者之间的数据关系,并利用一元线性回归模型验证二者之间的拟合度。本文创新性地基于R语言分析了我国城镇居民人均可支配收入与人均消费支出的内生动力及相关性程度,并以数据分析结果为依据,全面量化阐述东西部省份之间消费水平的差异及趋势,所采用的技术路线具备延展性并适用于其它时间段,对国家及地区宏观经济研究具有一定的参考价值。

关键词

R语言, 城镇居民, 消费水平

Research on the Consumption Level of Urban Residents in China Based on R Language

Jie Liu

AsiaInfo Technology (China) Co., Beijing

Received: Jul. 1st, 2024; accepted: Jul. 22nd, 2024; published: Aug. 5th, 2024

Abstract

This paper carries out research on the per capita disposable income and consumption expenditure of urban residents in China, selects the national statistical yearbook data of 2020, uses RStudio software to collect, process and analyze the data, and mainly carries out descriptive analysis and linear

文章引用: 刘杰. 基于 R 语言的我国城镇居民消费水平研究[J]. 统计学与应用, 2024, 13(4): 1027-1032.

DOI: 10.12677/sa.2024.134104

regression analysis. Descriptive analysis results in the composition, proportion and difference of the per capita disposable income and per capita consumption expenditure of the national 31 provincial regions; the regression analysis constructs the scatter plot between disposable income and consumption expenditure, analyzes the data relationship between the two, and verifies the fit between the two using a one-dimensional linear regression model. This paper innovatively analyzes the endogenous dynamics and degree of correlation between per capita disposable income and per capita consumption expenditures of urban residents in China based on R language, and comprehensively quantifies the differences and trends in the consumption level between eastern and western provinces based on the results of the data analysis, and the technical route adopted has the extensibility and applicability to other time periods, which is of a certain reference value to the national and regional macroeconomic research.

Keywords

R Language, Urban Residents, Consumption Level

Copyright © 2024 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

我国幅员辽阔且人口基数大,随着城镇化的快速发展与产业结构调整,出现了人口迁移及产业聚集等现象,全国 31 个省份及直辖市经济发展与城镇化水平逐步拉开了差距,城镇等级也出现了分层,一般可分为一线城市(北京、上海、广州、深圳)、新一线城市(成都、武汉、南京、杭州、天津等)、二线城市(昆明、沈阳、济南、无锡等,以省会城市为主)、三线城市(以地级市为主),各省份对应的各城市级别数量与规模也不尽相同。然而,城镇居民消费水平可直观反映出国家及地区的经济发展状况,对拉动“内循环”及调整产业结构政策具有一定的指导意义[1]。

本文利用 RStudio 软件,采集国家相关统计数据,围绕我国城镇居民消费水平进行研究,绘制相应的可视化图形,探讨各地区城镇居民人均消费支出及构成,并分析潜在的关联关系。

2. 研究意义及方法

2.1. 研究意义与主要内容

我国经济发展主要依靠外贸出口、内部消费以及投资带动,其中城镇居民消费指数则是经济发展水平的“晴雨表”,现阶段我国经济处于结构性调整阶段,出现了消费结构性失衡的现象。因此,研究城镇居民消费的现状,对于促进经济发展、扩大内需具有重要意义[2]。

本报告研究的主要问题包括:1) 我国城镇居民人均可支配收入与消费支出的总体特征;2) 分析我国城镇居民收入水平与消费能力之间的相关性等。

2.2. 研究方法数据来源

本报告使用 2020 年中国部分地区城镇居民人均可支配收入以及人均消费支出构成相关数据,数据主要来源于中国统计年鉴,涉及全国各地区数据;为分析数据内在逻辑,利用 RStudio 软件对数据进行处理、分析,综合运用描述性分析、线性回归等统计方法,以便更直观地展示分析结果[3]。

3. 描述性分析

经过整理分地区城镇居民人均可支配收入及人均消费支出相关数据，形成相应的 Excel 表格，导入已设置好相应软件的 RStudio 软件中，通过运行 RStudio 软件可直接输出相应的结论，相关公式及主要结论如下。

- 均值：均值主要采用算数平均数，公式为： $\bar{x} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum(X)}{n}$ 。
- 方差：总体方差用 σ^2 表示，方差越大意味着数据间离散程度越大，公式为： $s^2 = \frac{\sum(X - \bar{X})^2}{n-1}$ 。
- 偏度：偏度是对分布倾斜方向及程度的测量，公式为： $a_3 = \frac{\sum f(X - \bar{X})}{\sigma^3 \sum f}$ ，在公式中， a_3 为正，表示分布为右偏态分布，也称正偏峰分布，它的倾斜方向是向右的； a_3 为负，则表示分布为左偏，也称负偏态分布，倾斜方向是向左的。
- 峰度：峰度是频数分布曲线与正态分布相比较，顶端的尖峭程度，公式为： $a_4 = \frac{\sum f(X - \bar{X})^4}{\sigma^4 \sum f}$ ，当 $a_4 = 3$ 时，分布曲线为正态分布；当 $a_4 < 3$ 时，分布曲线为平峰分布；当 $a_4 > 3$ 时，分布曲线为尖峰分布。

3.1. 人均可支配收入描述性分析

3.1.1. 构成及占比情况

2020 年，我国 31 个省级区域城镇居民人均可支配收入为 43,834 元，主要包括工资性收入、经营净收入、财产净收入与转移净收入，各部分占比分别为 60.2%、10.7%、10.6%、18.5%，可知工资性收入占据主导地位，也从侧面反映出城镇居民大多数为工薪族，需要有稳定的工作以保障家庭收入来源；而对比 2019 年公布的人均可支配收入 42,359 元，同比增长 3.5%，而扣除价格因素后实际增长 1.2% [4]。

3.1.2. 省份差异性分析

2020 年，人均可支配收入最高的地区为上海，达到了 76437.26 元，其工资性占比为 57.3%；而人均可支配收入最少的地区为黑龙江，达到了 31114.75 元，其工资性占比为 56.4%；二者的差值为 45322.51 元，即上海的人均可支配收入为黑龙江的 2.5 倍，且工资性占比更高。结合数据分析可以得出，上海作为全国乃至全球经济中心城市，具有较高的薪资水平，城镇居民能够获得更高的劳动报酬；而黑龙江地处东北，缺乏高薪酬的工作岗位，城镇居民难以获得高报酬的工作；通过描述统计分析可知，人均可支配收入平均值为 42252.67，标准差为 11129.23，偏度系数与峰度系数分别为 1.92 与 2.98，表明各省人均可支配收入存在明显差异[5]。

3.2. 人均消费支出描述性分析

3.2.1. 构成及占比情况

2020 年，全国 31 个省级区域城镇居民人均消费支出为 27,007 元，包括食品烟酒、衣着、居住、生活用品及服务、交通通信、教育文化、医疗保健及其它用品，占比分别为 29.2%、6.1%、25.8%、6.1%、12.9%、9.6%、8.04%、2.4%，可知食品烟酒与居住为主要消费支出，城镇居民的消费支出主要为满足基本生活需求；而对比 2019 年公布的人均消费支出 28,063 元，同比下降 3.8%，扣除价格因素，实际下降 6.0%，可知 2020 年受到新冠疫情的影响，城镇居民的消费水平出现大幅下降[6]。

3.2.2. 省份差异性分析

2020 年, 人均消费支出最高的地区为上海, 达到了 44839.32 元, 其食品烟酒及居住支出占比高达 62.4%; 而人均消费支出最少的地区为山西, 达到了 20331.85 元, 其食品烟酒及居住支出占比为 47.99%; 二者的差值为 24507.47 元, 即上海的人均消费支出为山西的 2.2 倍, 且食品烟酒及居住支出占比为山西的约 1.3 倍。结合数据分析可知, 上海作为国际化大都市, 消费成本对比山西较高, 城镇居民的消费能力也更强; 而从另一个方面也得出上海城镇居民在满足日常基本需求方面需要花费更多的资金, 生活压力对比山西更大; 通过描述统计分析可知, 人均消费支出的数据平均值为 26080.76 元, 偏度系数与峰度系数分别为 1.60 与 1.95, 表明山西消费能力与上海存在较大差距且山西的人均消费支出低于均值, 各省人均可支配收入存在明显差异[7]。

3.3. 小结

总之, 2020 年全国 31 个省级区域中, 上海城镇居民在人均可支配收入及消费支出方面均领先于其它地区, 体现了上海作为全国经济中心的绝对地位, 而黑龙江虽然人均可支配收入最低, 但是人均消费支出却高于山西, 说明一定程度上黑龙江城镇居民的消费意愿更强烈; 人均可支配收入与人均可消费支出超过对应中位数的地区分别只有 8 个、10 个, 占比均不到 50%, 说明大部分地区的城镇居民整体生活质量不高; 而 2020 年受到新冠疫情的影响, 城镇居民的人均收入水平及消费能力均同比下降, 对我国经济内循环造成一定程度的影响[8]。

4. 回归分析

为分析 2020 年我国城镇居民人均可支配收入与人均消费支出之间的相关性, 利用 RStudio 软件, 采用一元线性回归分析法得出相关系数, 作为参考依据。线性回归方程公式是用来描述自变量和因变量之间的线性关系的数学表达式, 公式一般可表示为: $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$, 其中 y 是因变量(预测目标), $x_1, x_2, \dots, x_n$ 是自变量(特征), $\beta_0, \beta_1, \dots, \beta_n$ 是回归系数(模型的参数), ε 是误差项代表不能被自变量所解释的随机误差。

本文采用一元线性回归, 遵循线性回归方程公式, 因变量为消费支出, 自变量主要为可支配收入, 而回归系统由代码运算所得; 而皮尔逊相关系数是最常用的相关系数之一, 它衡量了自变量和因变量之间的线性关系程度, 取值范围在 -1 到 1 之间, 相关系数的绝对值越接近 1, 表示自变量与因变量之间的线性关系越强。

$$y = a + bx, \text{ 其中: } a = \frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i}{n}, \quad b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}。$$

4.1. 散点图绘制

通过图 1 的散点图可知, 2020 年我国城镇居民人均可支配收入越多, 人均消费支出也越多, 二者的观测点基本上呈线性关系, 具备正相关性; 通过计算二者的相关系数, 可得出二者的相关系数值为 0.9686, 具备高度正相关关系; 而从散点图的布局情况来看, 绝大部分地区城镇居民人均可支配收入集中在 30,000 元至 50,000 元的区间, 而消费支出集中在 20,000 元至 35,000 元的区间。

4.2. 计算相关系数及线性回归

确定相关关系后, 将 2020 年全国 31 个地区城镇居民人均可支配收入设置为解释变量 x , 以人均消费支出作为被解释变量 y , 构建一元线性模型为: $y = \alpha_0 + \alpha_1 x + \beta$, 从图 2 的线性回归图结果中可看出对应的方程为: $y = 3855.42395 + 0.52601 \times \text{人均可支配收入}$, 模型 R 方为 0.9382, 调整 R 方为 0.9361, 模型

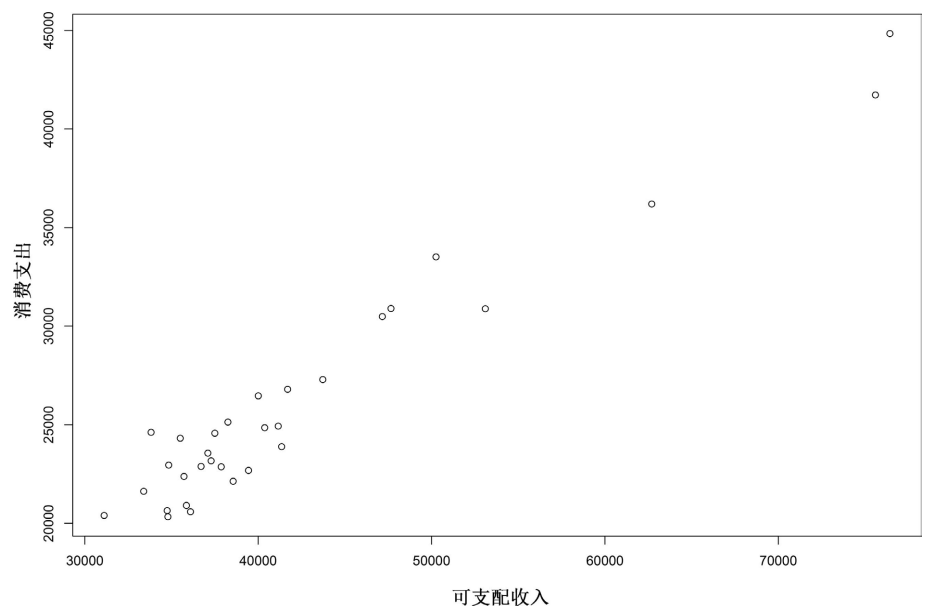


Figure 1. Scatter plot between disposable income and consumption expenditure
图 1. 可支配收入与消费支出之间的散点图

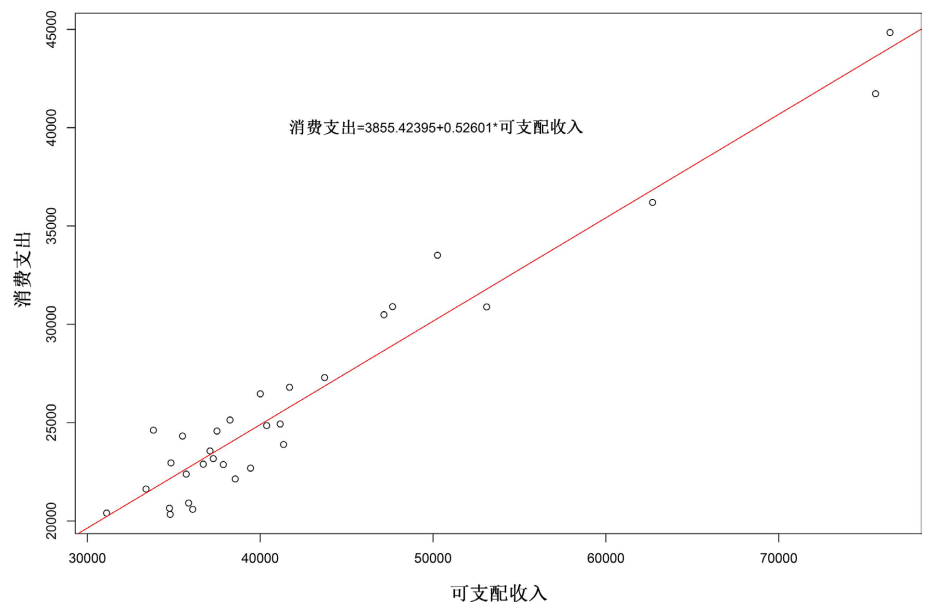


Figure 2. Linear regression plot between disposable income and consumption expenditure
图 2. 可支配收入与消费支出之间的线性回归图

拟合效果较好，说明分析结果具备较好的事实基础与说服力；另外，从观测点在线上线下的布局来看，线上的省份为 15 个，线下的省份为 16 个，上海的消费能力强于北京，东南沿海地区的消费能力要强于东北地区。

4.3. 小结

我国城镇居民的人均可支配收入与人均消费支出总体上呈正线性关系，收入越高对应的消费能力也越强，体现出大宗商品的购买力也越强，然而总体发展水平明显不均衡，各省份经济发展水平差异呈逐

步扩大态势；而对应可在支配收入区间 30,000 元至 45,000 元间的省份居多，占比接近 80%，说明我国城镇居民总体收入水平与消费能力还不够高，也体现出我国仍旧处于发展中国家行业，国家需进一步大力发展经济以提升人民的收入水平，进而促进消费能力升级，新旧动能转换以刺激经济内生动力[9]。

5. 研究主要结论

本文研究对象为 2020 年我国城镇居民人均可支配收入与消费支出的内在逻辑及相关性分析，充分利用 RStudio 软件采集对应数据资源，通过描述性分析与线性回归分析，得出我国各地区城镇居民人均收入与支出的基本情况、地区差异、相关系数等，所采用的技术路线具备延展性并适用于其它时间段，分析结果总体上客观真实，整体上折射出我国地区之间人均可支配收入与消费支出之间存在较大差异，二者相关性突出，也从侧面验证了我国各地区经济发展不均衡，东南沿海地区的优势非常明显。随着我国经济结构化转型不断深化发展，后续研究可基于各时间维度进一步挖掘城镇居民消费可支配收入明细与各类消费支出之间的相关性[10]。

参考文献

- [1] 袁原, 王新国. 新疆城镇居民收入与消费行为关系研究[J]. 北方经贸, 2011(4): 66-68.
- [2] 吴蓓蓓, 陈永福, 于法稳. 基于收入分层 QUAIDS 模型的广东省城镇居民家庭食品消费行为分析[J]. 中国农村观察, 2012(4): 59-69+94-95.
- [3] 杨丹. 我国货币供应量与经济增长关系实证分析[J]. 现代商贸工业, 2009, 21(9): 153-154.
- [4] 王玲玲, 马骁. 我国民用汽车需求的计量经济模型及分析[J]. 山西经济管理干部学院学报, 2006, 14(1): 16-18.
- [5] 朱振亚, 张小青. 经济转型期城乡居民消费地区间非均衡性研究[J]. 经济问题探索, 2011(4): 12-16.
- [6] 辜予薇, 易昆南. 居民消费的计量经济模型实证分析[J]. 数学理论与应用, 2008, 28(4): 28-31.
- [7] 王攀娜. 我国农村居民消费支出地区差异实证研究——关于聚类分析和因子分析的应用[J]. 商业时代, 2013(36): 48-50.
- [8] 刘昱清. 我国城镇居民消费支出结构与水平的因子分析[J]. 济宁学院学报, 2014, 35(4): 80-83.
- [9] 刘思嘉, 赵金楼. 我国城镇居民消费结构的聚类分析及政策建议[J]. 现代管理科学, 2010(1): 69-71.
- [10] 罗世超, 谢蕊霞. 影响我国城镇居民消费水平的主要因素分析[J]. 商场现代化, 2012(26): 26-27.