

基于京津冀地区居民人均消费水平差异的协方差分析模型

刘程馨, 牟唯嫣, 王欣

北京建筑大学理学院, 北京

收稿日期: 2022年7月9日; 录用日期: 2022年7月20日; 发布日期: 2022年8月2日

摘要

推动京津冀协同发展是一个重要的国家战略。本文基于2014~2020年京津冀地区居民人均消费支出及人均可支配收入数据, 考虑了居民人均可支配收入对居民人均消费支出的影响, 利用统计中的协方差分析方法构建模型, 证实了京津冀三地的人均消费水平存在显著性差异。北京市的人均消费水平显著高于天津市, 天津市的人均消费水平显著高于河北省, 得出的结论较方差分析而言更为准确。本文还分析得到了进一步的结论——除经济发展水平的影响外, 非经济因素对三地居民人均消费水平差异的形成也有显著影响。

关键词

居民人均可支配收入, 居民人均消费支出, 协方差分析模型, 地区差异, 经济发展水平, 非经济因素

The Analysis of Covariance Model Based on the Difference between Per Capita Consumption Level of Residents in Beijing, Tianjin and Hebei

Chengxin Liu, Weiyang Mu, Xin Wang

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

Received: Jul. 9th, 2022; accepted: Jul. 20th, 2022; published: Aug. 2nd, 2022

Abstract

Coordinated development of Beijing-Tianjin-Hebei region is an important national strategy. Based

文章引用: 刘程馨, 牟唯嫣, 王欣. 基于京津冀地区居民人均消费水平差异的协方差分析模型[J]. 统计学与应用, 2022, 11(4): 739-746. DOI: 10.12677/sa.2022.114078

on data of per capita consumption expenditure and per capita disposable income of residents in Beijing-Tianjin-Hebei region from 2014 to 2020, this paper considers the impact of per capita disposable income of residents on per capita consumption expenditure of residents, and uses analysis of covariance method in statistics to build a model. There is a significant difference in the per capita consumption level between Beijing, Tianjin and Hebei. The per capita consumption level in Beijing is significantly higher than that in Tianjin. And the per capita consumption level in Tianjin is significantly higher than that in Hebei. The results of analysis of covariance are more accurate than that of ANOVA. And further conclusions are drawn from the analysis—in addition to the influence of the level of economic development, non-economic factors also have a significant impact on the formation of the difference in the per capita consumption level of residents in the three places.

Keywords

Per Capita Disposable Income of Residents, Per Capita Consumption Expenditure of Residents, Analysis of Covariance Model, Regional Differences, Economic Development Level, Non-Economic Factors

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

推动京津冀协同发展，是党中央在新的历史条件下作出的重大决策部署。目前，京津冀地区的发展取得了阶段性的成果，但发展不平衡的问题仍然存在。而消费对地区发展具有重要影响。因此，研究三个地区居民人均消费水平的差异性及其影响因素对区域经济发展具有指导意义。

居民人均消费水平不仅受经济发展水平的影响，也与文化背景、消费观念及消费环境等非经济因素有关。人均可支配收入是反映区域经济发展和居民生活水平的重要指标[1]。本文以2014~2020年京津冀地区居民人均消费水平为例，以居民人均消费支出和人均可支配收入分别对居民人均消费水平和经济发展水平进行描述。本文将居民人均消费支出作为响应变量，同时考虑了居民人均可支配收入对居民人均消费水平的影响，将居民人均可支配收入视为协变量。采用统计中的协方差分析方法构建了数学模型，证实了北京市、天津市和河北省三地的居民人均消费水平之间存在显著性差异，并且除经济发展水平的影响外，非经济因素对三省(市)居民人均消费水平差异的形成也具有显著影响。

2. 理论方法

协方差分析模型是方差分析模型和线性回归模型的一种“混合”[2]。其中方差分析是主要部分，回归部分是由于协变量无法或很难控制而引入的。协方差分析的目的是略去回归部分，即排除协变量的影响，对得到的纯方差分析模型的方差分析进行适当校正，以分析控制变量对响应变量的影响。

2.1. 数据来源及相关说明

本文选取了北京市、天津市和河北省2014~2020年共7年的数据，居民人均消费支出和居民人均可支配收入数据均来自2021年中国统计年鉴[3]，整理相关数据得表1。居民人均消费支出用于衡量居民人均消费水平，而居民人均可支配收入是衡量经济发展水平的重要指标。因此，本文以居民人均消费支出

和居民人均可支配收入分别对居民人均消费水平和经济发展水平进行描述,并基于 SPSS26.0 进行数据分析。

Table 1. The per capita consumption expenditure and per capita disposable income of residents in the Beijing-Tianjin-Hebei region from 2014 to 2020

表 1. 京津冀地区 2014~2020 年居民人均消费支出和居民人均可支配收入

年份	北京市		天津市		河北省	
	人均消费支出	人均可支配收入	人均消费支出	人均可支配收入	人均消费支出	人均可支配收入
2014	31102.9	44488.6	22343.0	28832.3	11931.5	16647.4
2015	33802.8	48458.0	24162.5	31291.4	13030.7	18118.1
2016	35415.7	52530.4	26129.3	34074.5	14247.5	19725.4
2017	37425.3	57229.8	27841.4	37022.3	15437.0	21484.1
2018	39842.7	62361.2	29902.9	39506.1	16722.0	23445.7
2019	43038.3	67755.9	31853.6	42404.1	17987.2	25664.7
2020	38903.3	69433.5	28461.4	43854.1	18037.0	27135.9

单位:元。

2.2. 模型的构建

考虑到居民人均可支配收入对居民人均消费支出的影响,将前者视为协变量,它是连续数值型变量。控制变量为地区,是分类型变量,它有 3 个水平,即北京市、天津市和河北省。7 年来居民人均消费支出和居民人均可支配收入的数据可看作是重复试验,共重复 7 次[4]。我们来构建具有一个协变量的单因素协方差分析模型,用于分析京津冀三省(市)居民人均消费支出的差异性及其影响因素。假设控制变量有 a 个水平,在各水平下重复观测的次数均为 n 次,则在第 i 水平下第 j 次观测的响应变量观测值可以表示为:

$$y_{ij} = \mu + \alpha_i + \gamma z_{ij} + e_{ij}, i = 1, \dots, a; j = 1, \dots, n,$$

其中 e_{ij} 为随机误差,且假定 $e_{ij} \sim N(0, \sigma^2)$, 诸 e_{ij} 相互独立, μ 为总平均, α_i 为控制变量第 i 水平的效应, γ 为协变量系数, z_{ij} 为第 i 水平下第 j 次观测的协变量观测值。

针对所研究的问题,我们作出如下假设:

原假设 H_0 :居民人均可支配收入对居民人均消费水平无显著影响以及 H_1 :地区对居民人均消费水平无显著影响。

若 y_{ij} 和 z_{ij} 分别为第 i 水平下第 j 次观测的响应变量和协变量观测值;

\bar{y}_i 为控制变量在第 i 个水平下的样本均值;

\bar{z}_i 为协变量在第 i 个水平下的样本均值;

\bar{y} 为所有 n 个响应变量观测值的总均值;

\bar{z} 为所有 n 个协变量观测值的总均值;

$$E_{yy} = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y})^2;$$

$$E_{yz} = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y})(z_{ij} - \bar{z});$$

$$E_{zz} = \sum_{i=1}^a \sum_{j=1}^n (z_{ij} - \bar{z})^2 ;$$

$$A_{yy} = \sum_{i=1}^a \sum_{j=1}^n (\bar{y}_i - \bar{y})^2 ;$$

$$A_{yz} = \sum_{i=1}^a \sum_{j=1}^n (\bar{y}_i - \bar{y})(\bar{z}_i - \bar{z}) ;$$

$$A_{zz} = \sum_{i=1}^a \sum_{j=1}^n (\bar{z}_i - \bar{z})^2 .$$

为检验上述假设，构造的统计量为：

$$\text{对于 } H_0, F_0 = \frac{E_{yz}^2 / E_{zz}}{[(E_{yy} - E_{yz}^2 / E_{zz})] / [(a-1)(n-1) - 1]} ;$$

$$\text{对于 } H_1, F_1 = \frac{[(E_{yy} + A_{yy}) - (E_{yz} + A_{yz})]^2 / (E_{zz} + A_{zz}) - (E_{yy} - E_{yz}^2 / E_{zz})}{(E_{yy} - E_{yz}^2 / E_{zz}) / [(a-1)(n-1) - 1]} .$$

3. 实证分析

3.1. 方差分析

首先我们不考虑协变量的影响，进行方差分析。由于方差分析需满足诸样本相互独立、正态性和方差齐性三个假设条件，我们对此来进行验证。由于观测数据来自不同地区、不同年份，独立性基本可以满足。方差齐性和正态性检验结果见表 2，表 3。

Table 2. Variance homogeneity test

表 2. 方差齐性检验

	莱文统计	自由度1	自由度2	显著性
居民人均消费支出	基于平均值	0.810	2	0.460
	基于中位数	0.648	2	0.535
	基于中位数并具有调整后自由度	0.648	2	14.942
	基于剪除后平均值	0.814	2	0.459

Table 3. Normal test

表 3. 正态性检验

地区	柯尔莫戈洛夫 - 斯米诺夫(V) ^a			夏皮洛 - 威尔克			
	统计	自由度	显著性	统计	自由度	显著性	
居民人均消费支出	北京市	0.106	7	0.200*	0.994	7	0.998
	天津市	0.144	7	0.200*	0.985	7	0.979
	河北省	0.151	7	0.200*	0.931	7	0.562

*这是真显著性的下限。^a里利氏显著性修正。

方差分析的假设条件满足后, 我们来进行单因素方差分析, 实验结果见表 4。从表中可以看出, 对于给定的显著性水平 $\alpha = 0.05$, 相伴概率值为 0.000, 小于给定的显著性水平 $\alpha = 0.05$, 从而拒绝原假设, 认为不同的地区的居民人均消费水平具有显著性差异[5]。

Table 4. Analysis of variance table

表 4. 方差分析表

	平方和	自由度	均方	F	显著性
组间	1658267152.127	2	829133576.063	76.507	0.000
组内	195073207.443	18	10837400.413		
总计	1853340359.570	20			

我们采用 LSD 检验法进一步进行多重比较, 结果见表 5。在给定的显著性水平 $\alpha = 0.05$ 下, 北京市、天津市和河北省中任意两省(市)之间的相伴概率 p 值均为 0.000, 由此说明, 三省(市)之间的居民人均消费支出差异均是显著的。

Table 5. LSD multiple comparisons

表 5. LSD 多重比较

(I)地区	(J)地区	平均值差值(I-J)	标准错误	显著性	95%置信区间	
					下限	上限
北京市	天津市	9833.8429*	1759.6591	0.000	6136.936	13530.749
	河北省	21734.0143*	1759.6591	0.000	18037.108	25430.921
天津市	北京市	-9833.8429*	1759.6591	0.000	-13530.749	-6136.936
	河北省	11900.1714*	1759.6591	0.000	8203.265	15597.078
河北省	北京市	-21734.0143*	1759.6591	0.000	-25430.921	-18037.108
	天津市	-11900.1714*	1759.6591	0.000	-15597.078	-8203.265

*平均值差值的显著性水平为 0.05。

3.2. 协方差分析

在满足协方差分析基本假设的条件下, 我们构建的协方差模型才具有统计分析价值, 由此得到协方差分析的结果才是可靠的。因此, 对相关假设进行验证是必不可少的, 检验结果见图 1, 表 6。

从图 1 可以看出, 人均可支配收入和人均消费支出呈线性正相关。由表 6 结果可知, 人均可支配收入和人均消费支出之间的交互效应不显著, 满足平行性检验。

以上分析表明, 收集的数据满足协方差分析的基本假设, 故可对收集的数据进行协方差分析, 分析结果见表 7。

由表 7 可知, 协变量居民人均可支配收入对应的 p 值为 0.000, 小于给定的显著性水平 $\alpha = 0.05$, 说明人均可支配收入对人均消费支出的影响是显著的, 进一步说明地区经济发展水平对居民人均消费水平有显著影响[6]。地区对应的相伴概率在 $\alpha = 0.05$ 的显著性水平下也是显著的, 由此我们得到北京市、天津市和河北省三地的人均消费水平存在显著性差异, 且其 F 值比方差分析的 F 值小。

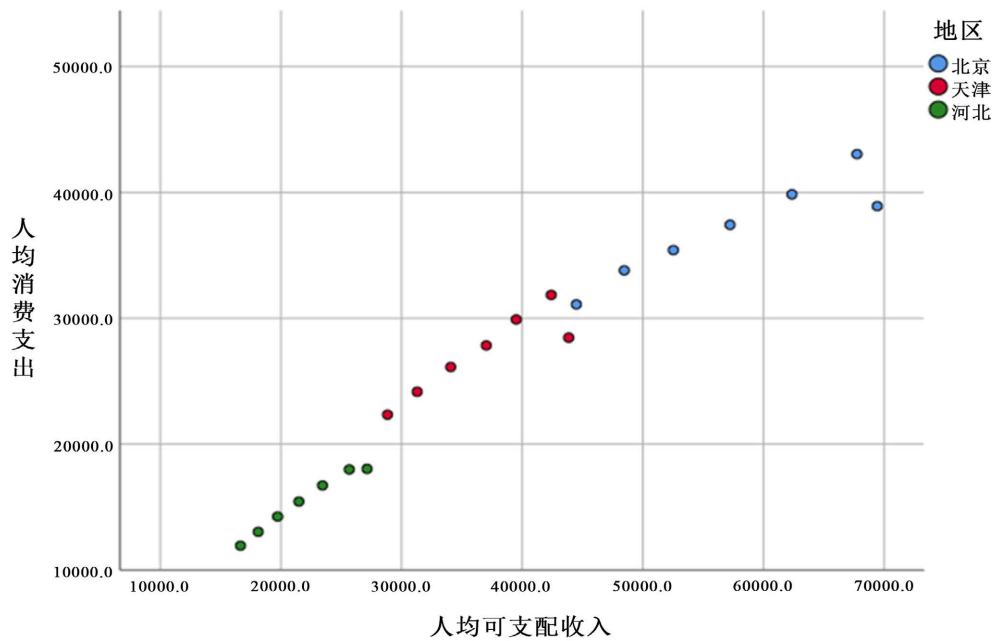


Figure 1. Per capita disposable income and per capita consumption expenditure by region scatter chart
图 1. 人均可支配收入与人均消费支出分地区散点图

Table 6. Test of the effect of the subject (considering the interaction between regions and per capita disposable income)
表 6. 主体间效应检验(考虑地区与人均可支配收入交互作用)

源	III类平方和	自由度	均方	F	显著性	偏Eta平方
修正模型	1827516714.436 ^a	5	365503342.887	212.307	0.000	0.986
截距	32340660.560	1	32340660.560	18.785	0.001	0.556
地区	14036515.798	2	7018257.899	4.077	0.039	0.352
人均可支配收入	128713389.655	1	128713389.655	74.765	0.000	0.833
地区*人均可支配收入	5676638.073	2	2838319.037	1.649	0.225	0.180
误差	25823645.133	15	1721576.342			
总计	16659903879.760	21				
修正后总计	1853340359.570	20				

^aR 方 = 0.986 (调整后 R 方 = 0.981)。

Table 7. Test of the effect of the subject (not considering the interaction between regions and per capita disposable income)
表 7. 主体间效应检验(不考虑地区与人均可支配收入交互作用)

源	III类平方和	自由度	均方	F	显著性	偏Eta平方
修正模型	1821840076.363 ^a	3	607280025.454	327.735	0.000	0.983
截距	47421304.581	1	47421304.581	25.592	0.000	0.601
人均可支配收入	163572924.236	1	163572924.236	88.277	0.000	0.839
地区	53000488.528	2	26500244.264	14.302	0.000	0.627
误差	31500283.206	17	1852957.836			
总计	16659903879.760	21				
修正后总计	1853340359.570	20				

^aR 方 = 0.983 (调整后 R 方 = 0.980)。

进一步通过偏差对比(表 8)我们可以得到,北京(标记为级别 1)、天津(标记为级别 2)与河北(标记为级别 3)之间的居民人均消费支出之间存在显著性差异。

Table 8. Comparison of deviations

表 8. 偏差对比

地区	简单对比 ^a	因变量		
		人均消费支出		
级别1与级别3	对比估算	5851.972		
	假设值	0		
	差值(估算 - 假设)	5851.972		
	标准误差	1840.323		
	显著性	0.005		
	差值的95%置信区间	下限	1969.230	
		上限	9734.714	
级别2与级别3	对比估算	5245.699		
	假设值	0		
	差值(估算 - 假设)	5245.699		
	标准误差	1015.403		
	显著性	0.000		
	差值的95%置信区间	下限	3103.385	
		上限	7388.013	

^a参考类别 = 3。

接下来我们通过协方差分析进行修正,并对消除协变量影响前后京津冀三地的人均消费支出的均值进行比较,结果如表 9 所示。

从表 9 中可以看出,北京市、天津市和河北省三地的人均消费支出依然存在差异,但消除了协变量人均可支配收入的影响后,它们之间的差异明显缩小了。排除了经济发展水平的影响,人均消费支出均值从高到低的排序依然是北京市、天津市、河北省,同方差分析的结果一致,但因排除了协变量的干扰而更加可靠。由图 2 可以看出,京津冀三地修正后的均值之间均存在较大差异。由此可知,京津冀地区居民的平均消费水平在一定程度上因经济发展水平的制约而存在显著差异,但三地间居民的消费观念、消费模式以及社会环境等非经济因素的不同,也致使居民消费意愿有很大不同[7]。

Table 9. The means before and after correction

表 9. 校正前后的均值

地区	北京市	天津市	河北省
未校正的值	37075.857	27242.014	15341.843
已校正的值	29784.981	28266.893	25634.624

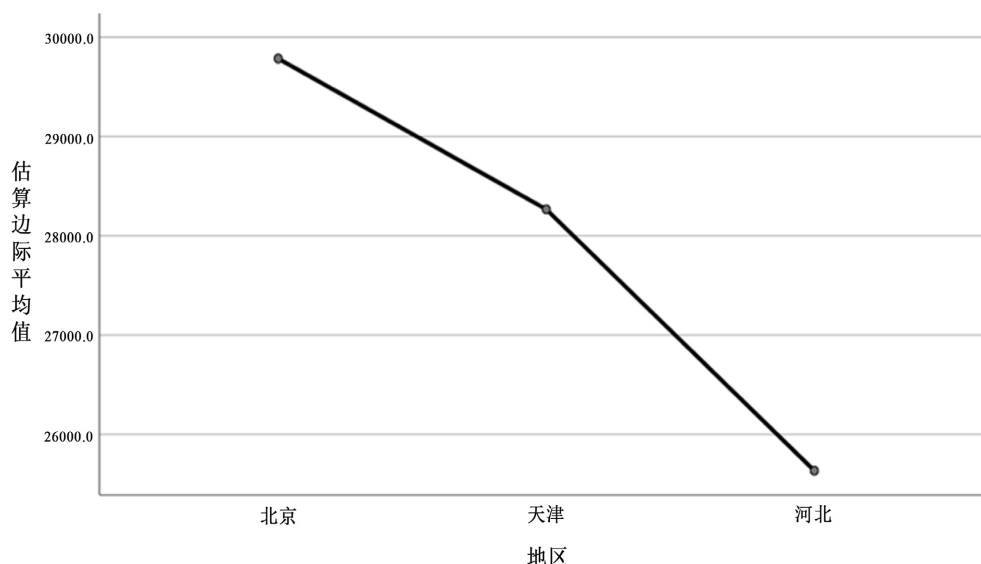


Figure 2. Line chart of the estimated marginal average of per capita consumer expenditure
图 2. 人均消费支出的估算边际平均值折线图

4. 结论

本文针对 2014~2020 年京津冀地区居民平均消费水平的差异,应用协方差分析方法构建了数学模型,进行了统计分析,相较于方差分析方法,得到了更为准确的结果。从结果中可以发现,有效消除了反映经济发展水平的协变量——居民人均可支配收入的影响后,北京市、天津市和河北省之间的居民人均消费水平的差距虽有缩小,但仍显著存在,居民人均消费水平从高到低的排序依然是:北京市、天津市和河北省。这表明京津冀三省(市)的消费差异不仅受经济发展水平的影响,也与社会背景、消费观念等非经济因素的显著影响有关。因此,国家在大力推动地区经济发展的同时,也要促进居民消费价值观的正向更新,积极营造良好的社会环境及消费环境,从而进一步推动京津冀协同发展。本文仅从消费的角度研究了京津冀地区的差异性,给出了相关建议,具有一定的局限性,而后的研究工作可从其他视角对京津冀地区的发展展开进一步研究。

基金项目

特别感谢北京建筑大学 2022 年度研究生创新项目(项目编号: PG2022138)对本文的资助。

参考文献

- [1] 朱婷婷, 马小雯, 孙玉会, 徐永芳, 周生伟. 基于空间统计分析甘肃省人均可支配收入发展研究[J]. 甘肃科技, 2019, 35(4): 66-68+16.
- [2] 王松桂, 史建红, 尹素菊, 吴密霞, 编著. 线性模型引论[M]. 北京: 科学出版社, 2004.
- [3] 国家统计局. 《中国统计年鉴 2021》[EB/OL]. <http://www.stats.gov.cn/tjsj/ndsj/2021/indexch.htm>, 2021-10-23.
- [4] 赵淳. 江浙沪区域内社会消费品零售总额协方差分析[J]. 中国市场, 2011(2): 170-171.
- [5] 薛薇. 统计分析与 SPSS 的应用[M]. 第四版. 北京: 中国人民大学出版社, 2014: 8.
- [6] 杜强, 贾丽艳, 编著. SPSS 统计分析从入门到精通[M]. 北京: 人民邮电出版社, 2011.
- [7] 马海锋, 朱琴华. 基于试验设计的我国城镇居民消费区域差异分析[J]. 中国市场, 2010(27): 6-8.