

# 基于协方差分析模型的东三省居民收入差异研究

宋玥璇, 牟唯嫣

北京建筑大学理学院, 北京

收稿日期: 2023年9月10日; 录用日期: 2023年10月9日; 发布日期: 2023年10月16日

## 摘要

本文以协方差分析作为主要分析方法, 选取人口结构作为协变量, 利用2013~2021年的居民人均可支配收入和人口城乡结构等信息, 运用协方差分析对黑龙江省、辽宁省、吉林省的居民收入差异进行研究, 结果表明东三省居民收入与人口结构密切相关, 可以通过改善人口结构来提高居民收入和缩小居民收入差距。

## 关键词

协方差分析, 居民收入差异, 人口结构, 方差分析

## Research on the Income Difference of Residents in Three Eastern Provinces Based on Covariance Analysis Model

Yuexuan Song, Weiyan Mu

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

Received: Sep. 10<sup>th</sup>, 2023; accepted: Oct. 9<sup>th</sup>, 2023; published: Oct. 16<sup>th</sup>, 2023

## Abstract

This paper uses covariance analysis as the main analysis method, selects population structure as the co-variable, makes use of the per capita disposable income of residents from 2013 to 2021 and the urban-rural structure of population and other information, and uses covariance analysis to study the income difference of residents in Heilongjiang, Liaoning and Jilin provinces. The results show that the income of residents in the three eastern provinces is closely related to population

structure. The income of residents can be increased and the income gap narrowed by improving the population structure.

## Keywords

Analysis of Covariance, Resident Income Difference, Population Structure, Analysis of Variance

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

不同地区的居民收入差异是研究收入分配的重要组成部分之一, 而收入分配作为经济学、社会学乃至政治学的中心论题之一尤为为重要。从宏观上讲, 这种差异的存在及其变化不但关系到社会与政局的稳定, 而且影响国民储蓄与投资。从微观上讲, 对任一产品的市场需求不但取决于人口数量、平均收入水平、价格和消费嗜好, 也取决于国民收入在地区间的分配[1]。除此之外, 区域间收入差异与各地区经济增长和教育水平都密切相关。因此, 研究不同地区的居民收入差异有着十分重要的现实意义。

振兴东北老工业基地战略实施以来, 东三省经济发展快速[2]。近年来, 为响应“东北振兴”战略, 东三省的经济工作持续展开。本文首先搜集了 2013~2021 的居民人均可支配收入和城乡人口数量, 以居民人均可支配收入作为评价指标, 从而能够描述不同省份间的收入差异。同时以人口结构作为协变量, 其中, 选取城镇人口占总人口的比重作为人口结构的衡量标准, 利用协方差分析方法和 SPSS 统计软件对黑龙江省、辽宁省、吉林省的居民收入差异进行分析评价, 发现人口结构对居民收入有显著影响。

## 2. 模型理论

### 2.1. 单因素方差分析理论

方差分析是由统计学家 R. A. Fisher 在进行试验设计时为解释试验数据而首先引入的一种统计方法, 它是从对数据误差来源的分析入手, 从而判断分类型自变量对数值型因变量是否有显著影响[3]。其中, 单因素方差分析是研究一个分类型自变量对一个数值型因变量的影响。

本文中要检验东北三个省份的居民人均可支配收入是否相等, 只涉及省份这一个因素, 因而使用单因素方差分析方法对其进行研究。为方便分析, 用 A 表示省份这一因素, 因素的三个水平分别为黑龙江省、辽宁省、吉林省, 用  $A_i$  表示,  $x_{ij}$  表示第  $i$  个水平下的第  $j$  个观测值,  $n_i$  表示每个水平下的观测次数, 对此数据进行单因素方差分析, 实际上就是检验黑龙江省、辽宁省、吉林省这三个省份的总体均值是否相等。一般来说, 单因素方差分析有三个步骤: 提出假设, 构造检验统计量, 以及统计决策。

在本例中, 使用的数据是 2013~2021 年的居民人均可支配收入, 相当于在  $A_1$ 、 $A_2$ 、 $A_3$  三个水平下进行了 9 次独立试验, 得到 27 个观测值。首先, 建立如下假设:

$$H_0: \mu_1 = \mu_2 = \mu_3$$

$$H_1: \mu_i (i=1,2,3) \text{ 不全相等}$$

其次, 为构造检验统计量, 需要计算三个误差平方和 SSA, SSE, SST, 而数据总的误差主要来自两方面, 一方面是组内误差, 另一方面是组间误差[4]。构造检验统计量  $F = \frac{MSA}{MSE}$ , 由于  $F$  是组间方差与组

内方差的比值,  $F$  的值越大, 意味着三个样本均值之间的差异越大, 越有利于拒绝原假设;  $F$  的值越小, 意味着方差波动主要由随机误差引起, 越有利于接受原假设。假设显著性水平为  $\alpha$ , 若  $F > F_{\alpha}(2, 24)$ , 则拒绝原假设, 表明  $\mu_1, \mu_2, \mu_3$  之间的差异是显著的; 否则, 就接受原假设。

## 2.2. 协方差分析理论

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[5]。它可以帮助我们确定不同组别的修正均值是否存在显著差异, 扣除其他可能干扰结果的混杂因素对分析结果的影响, 从而正确分析自变量对因变量的影响[1]。

本文中研究的是东三省的居民收入是否存在差异, 并且考虑到人口结构对居民收入也有影响, 将人口结构视为协变量, 构建协方差分析模型。在这个问题中, 由于只有省份这一个因素, 它有三个水平: 黑龙江省、辽宁省、吉林省, 故需要构建具有一个协变量的单向分类模型:

$$y_{ij} = \mu + \alpha_i + \gamma Z_{ij} + e_{ij}, \quad i = 1, \dots, a \quad j = 1, \dots, b,$$

这里  $e_{ij} \sim N(0, \sigma^2)$ , 且所有  $e_{ij}$  相互独立。考虑假设:

$$H_1: \alpha_1 = \dots = \alpha_a$$

$$H_0: \gamma = 0$$

根据假设检验步骤, 通过构建  $F$  统计量, 计算统计量的值  $F$  与临界值  $F(\alpha)$  进行比较, 最终得出拒绝或不拒绝原假设的结论。如果结论是拒绝原假设, 则表示省份对居民收入有影响, 反之则没有影响。

## 3. 实证分析

### 3.1. 数据预处理

居民收入是衡量一个地区经济发展状况的重要指标之一。本文收集了东三省 2013~2021 年居民人均可支配收入和城乡人口数, 根据三个省份的居民人均可支配收入, 采用单因素方差分析方法研究东三省的居民收入差异。同时利用协方差分析方法对东三省的居民收入差异进行分析评价, 引入了人口结构这一不可控制的定量变量作为协变量, 为了方便分析, 用城镇人口占总人口的比重作为衡量人口结构的指标。对所得原始数据进行处理后得到表 1 (保留两位小数)。

**Table 1.** Regional per capita income and population structure in the three eastern provinces

**表 1.** 东三省地区居民人均可支配收入与人口结构

年份	黑龙江省		辽宁省		吉林省	
	居民人均可支配收入	人口城乡结构	居民人均可支配收入	人口城乡结构	居民人均可支配收入	人口城乡结构
2013	15,903	0.58	20,818	0.66	15,998	0.56
2014	17,404	0.59	22,820	0.67	17,520	0.57
2015	18,593	0.60	24,576	0.68	18,684	0.58
2016	19,838	0.61	26,040	0.69	19,967	0.59
2017	21,206	0.62	27,835	0.69	21,368	0.60
2018	22,726	0.63	29,701	0.70	22,798	0.61
2019	24,254	0.65	31,820	0.71	24,563	0.62
2020	24,902	0.66	32,738	0.72	25,751	0.63
2021	27,159	0.66	35,112	0.73	27,770	0.63

### 3.2. 单因素方差分析

为检验省份不同时居民人均收入是否存在显著差异, 由于这里只有省份这一个因素, 故选用单因素方差分析法。

黑龙江省居民人均可支配收入正态P-P图

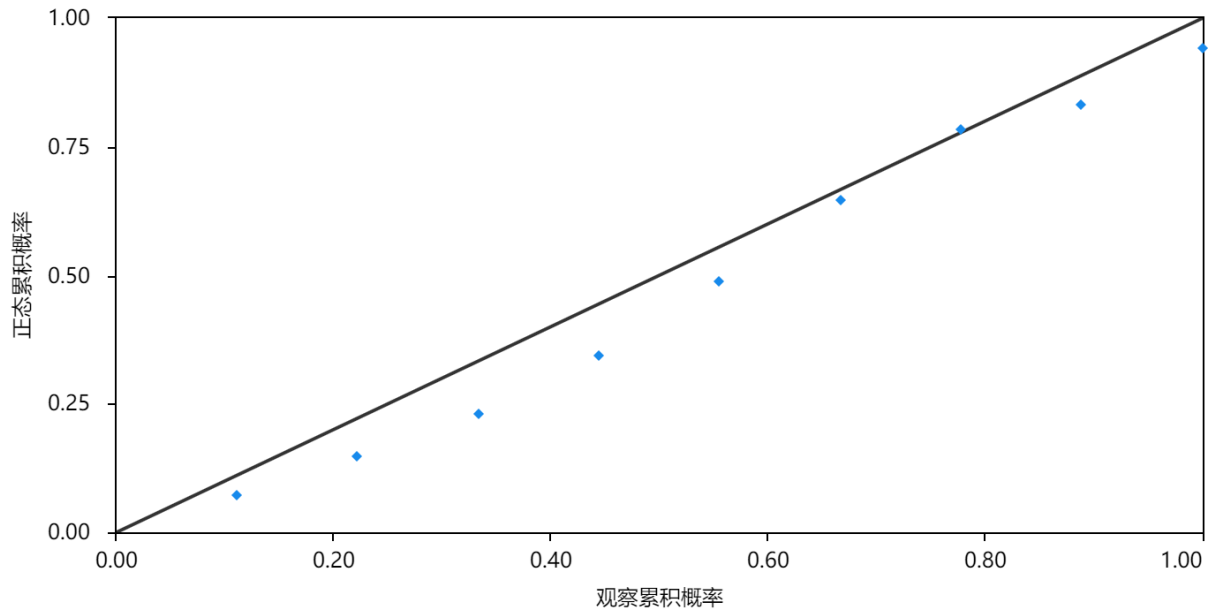


Figure 1. Normal p-p chart of disposable income per capita of all residents of Heilongjiang Province

图 1. 黑龙江省全体居民人均可支配收入(元)正态 p-p 图

吉林省居民人均可支配收入正态P-P图

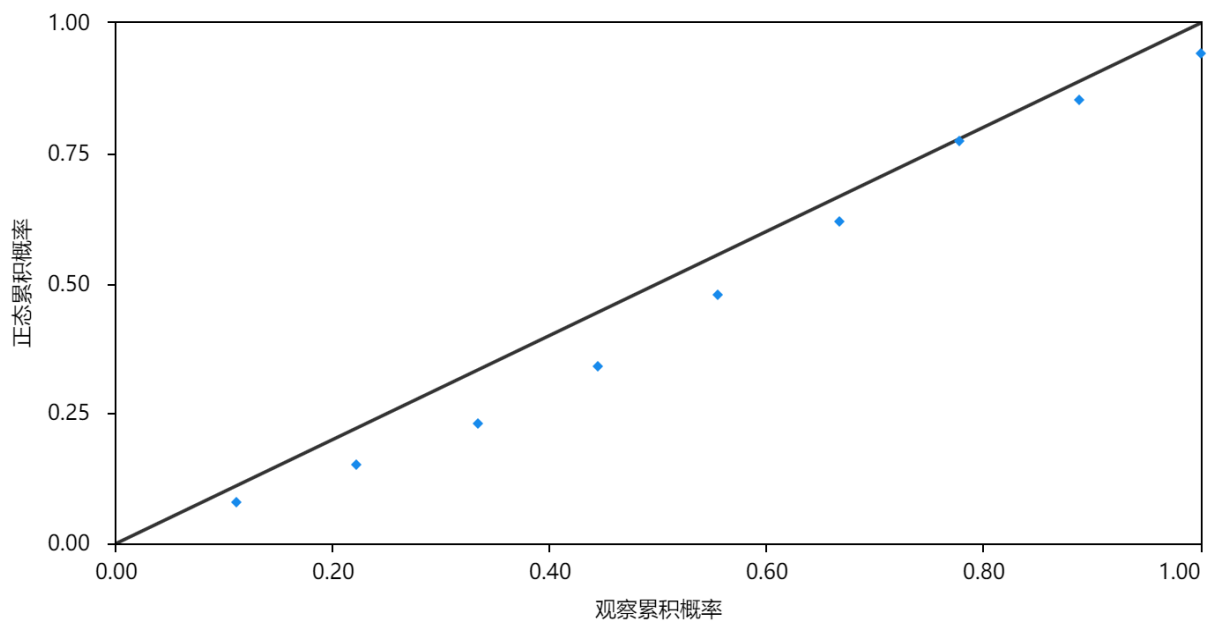


Figure 2. Normal p-p chart of disposable income per capita of all residents of Jilin Province

图 2. 吉林省全体居民人均可支配收入(元)正态 p-p 图

辽宁省居民人均可支配收入正态P-P图

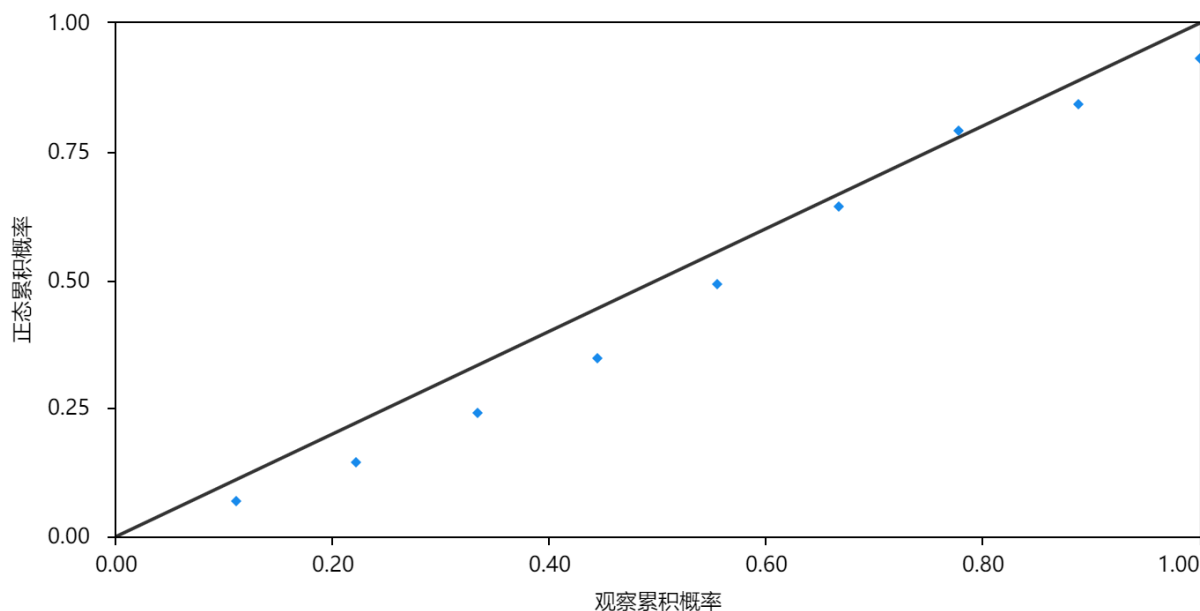


Figure 3. Normal p-p chart of disposable income per capita of all residents of Liaoning Province

图 3. 辽宁省全体居民人均可支配收入(元)正态 p-p 图

### 3.2.1. 假设条件检验

在进行单因素方差分析前, 需要对方差分析的基本假定进行检验。

首先是正态性检验, 即对于因素的每一个水平, 其观测值应该是来自正态总体分布的简单随机样本。本文采用 p-p 图对东三省全体居民人均可支配收入进行正态性检验, 当数据符合正态分布时, p-p 图中各点应近似呈一条直线。

图 1~图 3 均近似呈现为一条对角直线, 说明数据呈现出正态分布, 此外, 对数据进行 Kolmogorov-Smirnov 检验, Cramer-von Miss 检验等, 都能接受正态性的结论, 故东北三个省份的居民人均可支配收入全部均具备正态性特质。

其次是方差齐性检验, 即各组观察数据应该是从具有相同方差的正态总体中抽取的。如下的表格(表 2)是使用方差齐性检验, 去检验各个组别数据的波动情况是否有明显的差异。结果显示, 如果取显著性水平  $\alpha$  为 0.05, 不同省份样本对于人均可支配收入全部均不会表现出显著性( $p > 0.05$ ), 意味着不同省份样本对于人均可支配收入全部均有着方差齐性。此外还要指出的是, 在实际研究中, 可以对正态分布和方差相等这两个条件适当放宽。

Table 2. Variance chi-square test

表 2. 方差齐性检验

		莱文统计	自由度 1	自由度 2	显著性
人均可支配收入	基于平均数	0.427	2	24	0.657
	基于中位数	0.419	2	24	0.662
	基于中位数并具有调整后自由度	0.419	2	22.958	0.663
	基于剪除后平均值	0.428	2	24	0.656

最后, 观测值应该是独立的。文中东三省近 9 年的居民人均可支配收入和城乡人口数是来自国家统计局 2013 年~2021 年的年度数据, 数据的独立性要求满足。

### 3.2.2. 单因素方差分析

经过上述验证, 数据满足使用方差分析的前提要求, 因此可以使用 SPSS 软件对不同省份的居民人均可支配收入进行方差分析, 得到的方差分析表(表 3)如下:

Table 3. ANOVA table

表 3. 方差分析表

	平方和	自由度	均方	F	显著性
组间	251,736,135.630	2	125,868,067.815	7.174	.004
组内	421,088,726.889	24	17,545,363.620		
总计	672,824,862.519	26			

Table 4. Multiple comparisons

表 4. 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	
黑龙江省	吉林省	-270.444	1974.581	.892	-4345.78	3804.89
	辽宁省	-6608.333*	1974.581	.003	-10683.67	-2533.00
吉林省	黑龙江省	270.444	1974.581	.892	-3804.89	4345.78
	辽宁省	-6337.889*	1974.581	.004	-10413.22	-2262.55
辽宁省	黑龙江省	6608.333*	1974.581	.003	2533.00	10683.67
	吉林省	6337.889*	1974.581	.004	2262.55	10413.22

由方差分析表知,  $p$  值 = 0.004 < 0.05, 这表示不同省份样本对于人均可支配收入呈现出显著性差异, 还可以进一步进行多重比较, 利用 LSD 检验分析具体哪些省份之间有差异, 结果如表 4 所示。根据表 4 可知, 黑龙江省和辽宁省、吉林省和辽宁省的人均可支配收入存在显著差异, 而黑龙江省和吉林省人均可支配收入差异不明显( $p = 0.892 > 0.05$ ), 有着较为明显差异的组别平均值得分对比结果为“辽宁省 > 吉林”, “辽宁省 > 黑龙江省”。

### 3.3. 协方差分析

方差分析没有考虑人口结构对居民收入的影响, 而人口结构对居民收入的影响也不可忽视, 所以在研究东三省居民收入差异时, 不能忽视人口结构的影响。为了排除人口结构这一变量, 选择使用协方差分析方法。

#### 3.3.1. 假设条件检验

采用协方差分析方法前仍然需要检验数据是否满足协方差分析的条件。首先, 数据来自国家统计局, 并且根据 p-p 图分析散点图近似呈一条直线, 故独立性和正态性条件满足, 并且对数据进行方差齐性检验的  $p$  值大于 0.05, 各样本的方差基本相同; 其次, 进行协方差分析的因变量和协变量之间应该存在线性关系, 且线性关系一致, 如图 4, 各个省份下人口结构与人均收入存在线性关系, 且线性趋势基本一致; 此外, 还应满足处理因素和协变量之间没有交互作用, 如图 5 所示, 对省份和人口结构作交互作用图, 发现三条折线两两平行, 故交互作用不显著。

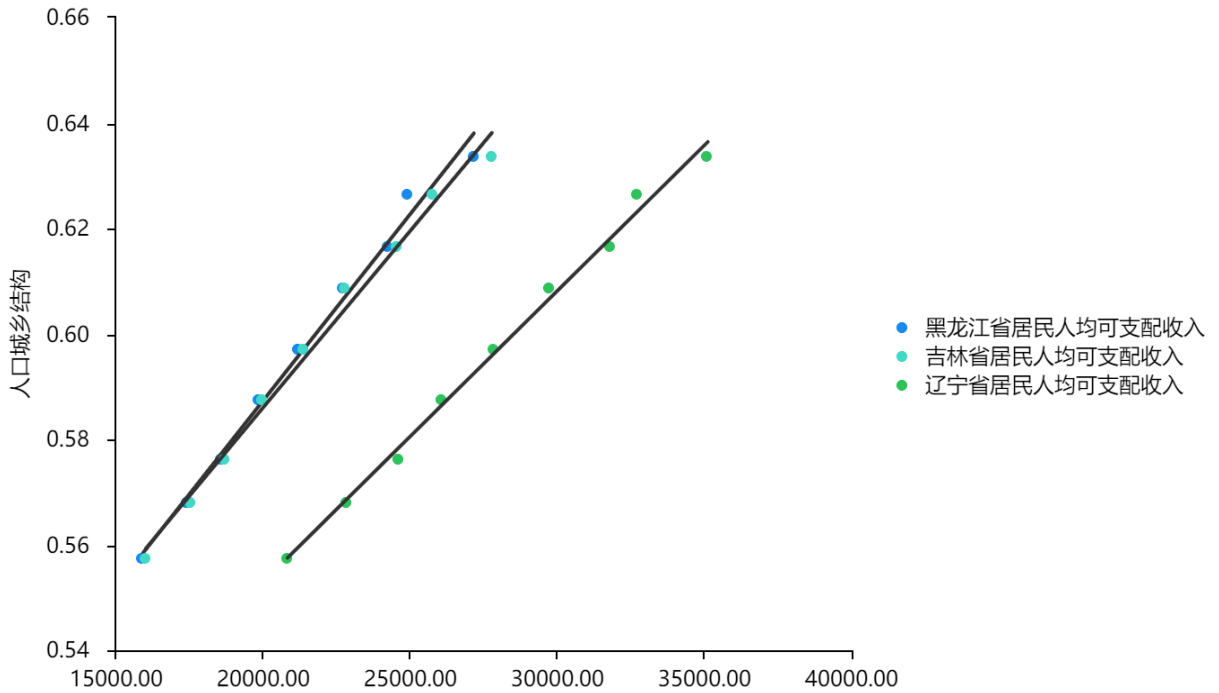


Figure 4. Trend comparison chart

图 4. 趋势对比图

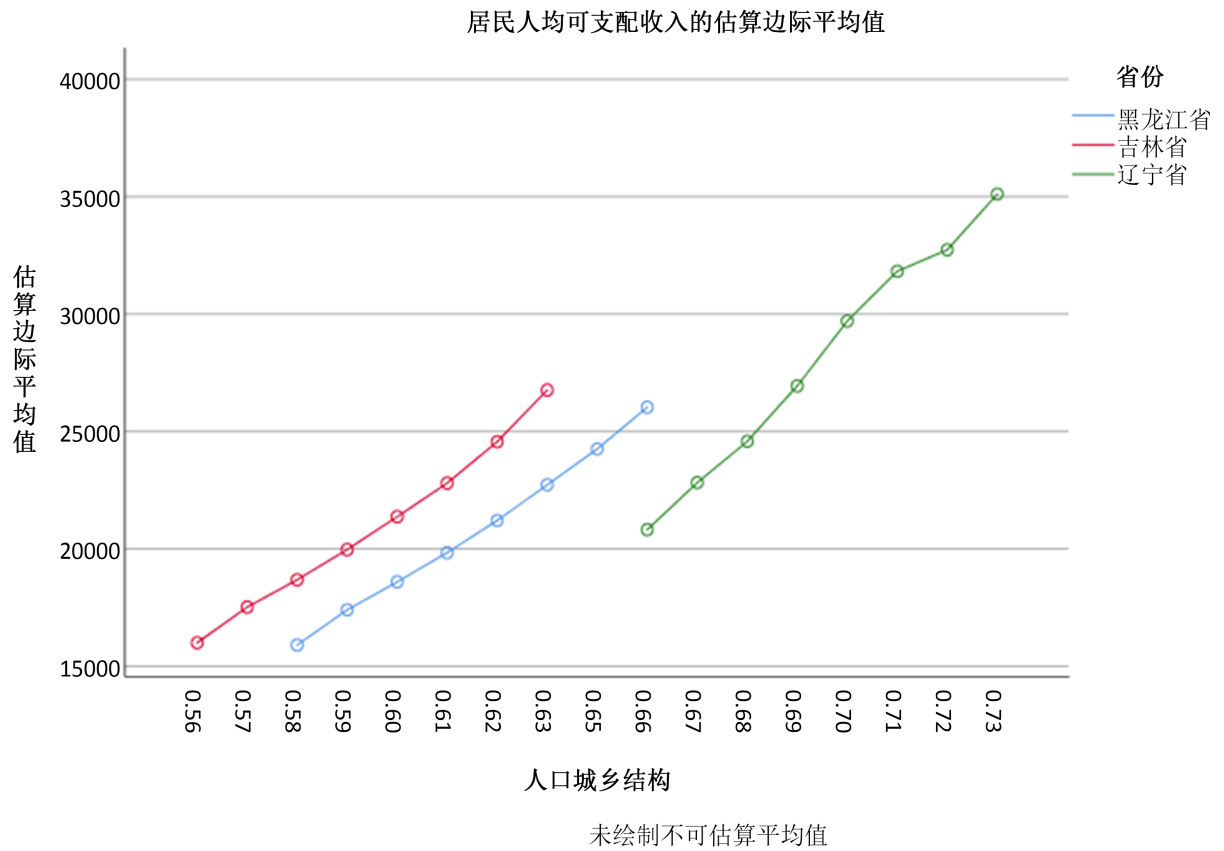


Figure 5. Interactive diagram

图 5. 交互作用图

### 3.3.2. 协方差分析

假设条件均满足, 故可以使用协方差分析方法对数据进行分析。由表 5 可知, 排除了人口结构造成的混杂干扰之后, 三个省份间的居民收入差距增大, 此时的居民收入顺序从高到低为黑龙江省、吉林省、辽宁省; 根据表 6 可知, 黑龙江省和吉林省的  $p$  值小于 0.05, 此时三个省份的收入差异均显著, 显然协方差分析结果更符合实际。

**Table 5.** Comparison of residents' income before and after revision

**表 5.** 修正前后居民收入对比

省份	修正前居民人均可支配收入	修正后居民人均可支配收入
黑龙江省	21,331.67	23,885.639
吉林省	21,602.11	28,214.81
辽宁省	27,940	18,773.33

**Table 6.** Pairwise comparison

**表 6.** 成对比较

	(I)地区 (J)地区	平均值差值(I-J)	标准误差	显著性	插值的 95% 置信区间	
					下限	上限
黑龙江省	吉林省	-4329.171*	509.525	0.000	-5644.774	-3013.568
	辽宁省	5112.309*	747.332	0.000	3182.684	7041.934
吉林省	黑龙江省	4329.171*	509.525	0.000	3013.568	5644.774
	辽宁省	9441.480*	913.477	0.000	7082.866	11,800.094
辽宁省	黑龙江省	-5112.309*	747.332	0.000	-7041.934	-3182.684
	吉林省	-9441.480*	913.477	0.000	-11,800.094	-7082.866

由表 7 可知人口城乡结构结果显著( $p$  值  $< 0.05$ ), 表明人口城乡结构能够显著影响人均收入, 且其偏 Eta 平方达到了 94.6%, 说明人口结构是影响人均收入的主要因素, 而省份虽然也显著, 但其偏 Eta 平方只有 83.5%, 说明东三省人均收入的影响受人口结构影响更大, 所以, 通过人口结构来对东三省居民人均收入差异进行评价具有可靠性。

**Table 7.** Tests for between-subjects effects

**表 7.** 主体间效应检验

源	III 类平方和	自由度	均方	$F$	显著性	偏 Eta 平方
修正模型	650,169,674.646a	3	216,723,224.882	220.022	0.000	0.966
截距	234,413,842.564	1	234,413,842.564	237.982	0.000	0.912
人口城乡结构	398,433,539.016	1	398,433,539.016	404.498	0.000	0.946
省份	114,801,300.041	2	57,400,650.020	58.274	0.000	0.835
误差	22,655,187.873	23	985,008.168			
总计	15,742,101,992.000	27				
修正后总计	672,824,862.519	26				



## 4. 结论与建议

东三省居民收入在整体水平提高的同时, 还存在着收入差异的问题。刘桂玲等人采用了标准差、变异系数和极差等简单的数学方法, 借助 ARCGIS 中的分级色彩设置的方法, 分析得出城镇居民收入水平排序为“辽宁省 > 吉林省 > 黑龙江省”且省间差距发展并不稳定[6]。郭永学等人通过分析东三省收入差距总体呈逐渐拉大趋势的事实, 得出应处理好发展经济与增加居民收入的关系, 以求逐步缩小收入差距[7]。关慧等人通过分析东北三省经济增长缓慢原因, 给出提高东北地区经济的建议[8]。

本文根据黑龙江省、辽宁省、吉林省 2013~2021 年居民人均可支配收入和人口城乡结构, 对东三省居民收入差异进行分析。首先运用单因素方差分析方法分析得出三个省份的居民收入存在显著差异, 其次考虑到人口结构对三个省份居民收入的影响, 运用协方差分析方法剔除这一协变量的影响, 得到了更为可靠的结论。分析表明, 修正后三个省份的居民收入差异更为显著(吉林省和辽宁省差异高达 9441.480), 其中辽宁省的居民人均可支配收入较修正前降低了 9166.67, 变化最为显著, 且低于黑龙江省和吉林省的居民收入, 所以辽宁省的居民人均收入受人口结构影响相较于黑龙江省和吉林省更大。因此, 国家在出台关于提高居民收入和缩小居民收入差距等的相关政策时, 要重视人口结构的变化对居民收入的影响, 改善人口结构、引导人口合理分布。

## 参考文献

- [1] 卢谢峰, 韩立敏. 中介变量、调节变量与协变量——概念、统计检验及其比较[J]. 心理学, 2007(4): 934-936.
- [2] 万广华. 中国农村区域间居民收入差异及其变化的实证分析[J]. 经济研究, 1998(5): 37-42+50.
- [3] 高卓. 单因素方差分析应用的实证研究[J]. 赤峰学院学报(自然科学版), 2014, 30(6): 4-6.
- [4] 王琳. 当代中国共产党人共同富裕思想研究[D]: [博士学位论文]. 北京: 北京交通大学, 2014.
- [5] 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.
- [6] 刘桂玲. 东北三省城镇居民收入水平空间分异研究[D]: [硕士学位论文]. 长春: 东北师范大学, 2012.
- [7] 郭永学, 谢博文. 东北地区城镇居民收入省际差距探析[J]. 长白学刊, 2013(3): 101-105.
- [8] 关慧, 段景超, 刘欣雨. 东北三省消费低迷阻碍经济增长原因及对策分析[J]. 全国流通经济, 2023(12): 156-159.