

2019年全国各地区肺结核发病率的 影响因素分析

代莲娇, 许传青*

北京建筑大学理学院, 北京

收稿日期: 2023年11月12日; 录用日期: 2024年1月29日; 发布日期: 2024年2月6日

摘要

中国是结核病高负担国家, 为探究影响肺结核发病率的宏观因素, 基于2019年全国各地区肺结核发病率, 提取各地区社会发展、人口构成因素等十个指标, 通过Spearman秩相关性检验、Kendall τ 相关检验, 分别分析各指标体系与肺结核发病率的相关关系; 建立最小二乘估计回归模型, 得到2019年各地区肺结核的发病率与人口的年龄结构显著相关, 且与人均国内生产总值成正相关, 与城镇人口构成比、生活垃圾无害化处理率成负相关。

关键词

肺结核发病率, 相关性检验, 最小二乘回归

Analysis of Factors Influencing the Incidence of Tuberculosis by Region in China, 2019

Lianjiao Dai, Chuanqing Xu*

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

Received: Nov. 12th, 2023; accepted: Jan. 29th, 2024; published: Feb. 6th, 2024

Abstract

China is a country with a high burden of tuberculosis, in order to explore the macro factors affecting the incidence rate of tuberculosis, based on the incidence rate of tuberculosis in all regions of the country in 2019, we extracted ten indicators such as social development and demographic factors in each region, and analyzed the correlation between each indicator system and the inci-

*通讯作者。

dence rate of tuberculosis through the Spearman's rank correlation test, and Kendall's τ correlation test, respectively; The least squares estimation regression model was established to obtain that the incidence rate of tuberculosis in each region in 2019 was significantly correlated with the age structure of the population, positively correlated with the per capita gross domestic product, and negatively correlated with the ratio of the composition of the urban population and the rate of harmless treatment of domestic waste.

Keywords

Number of Tuberculosis Cases, Correlation Tests, Least Squares Regression

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

结核病是由结核杆菌引起的一种慢性传染病。结核杆菌可以侵害人体的各种器官，以肺结核多见，侵入不同部位表现不一。作为全球结核病高负担国家，目前我国结核病防控形势仍不容乐观。“十四五”国民健康规划中要求全面落实结核病防治策略，加强肺结核患者发现和规范化诊疗，实施耐药高危人群筛查，强化基层医疗卫生机构结核病患者健康管理，加大肺结核患者保障力度[1]。

由于我国地域辽阔，各省市自然环境、经济水平、人口组成以及医疗水平差异明显，因此我国结核病发病具有时空分布的特点。尹金凤等人利用贝叶斯时空模型分析了时空分布因素对于 2008~2018 年北京市肺结核报告发病率的影响效果，结果表明 2008~2018 年北京市肺结核报告发病率呈下降趋势，城市发展新区需要受到更多关注[2]。赵明扬等人通过构建普通最小二乘(OLS)模型、地理加权回归(GWR)模型、时空地理加权回归(GTWR)模型描述了 2016~2018 年全国分地区肺结核发病情况，结果表明气象和空气质量因素对肺结核发病情况具有显著影响，且该影响存在时空特异性[3]。Bie 等人利用时空分布模型和 INLA 算法研究了 7 个影响因素及中国大陆结核病相对危险度(RR)的时空分布的影响。结果表明，从空间上看，相对风险最高的两个省份是新疆和贵州；从时间上看，从 2013 年到 2015 年，相对风险逐年下降。每年 2-5 月较高，3 月最为显著。从 6 月到 12 月有所下降[4]。刘家起等人对 2000 年至 2015 年全球结核病患者率的影响因素进行单因素与多因素分析，结果表明全球结核病患者率与人均医疗卫生支出、15~49 岁人口艾滋病病毒感染率等相关，应做好结核病的宣传教育工作[5]。

本文针对中国各地区社会发展、人口构成等因素，建立相关性分析与多元线性回归模型，探究不同地区的宏观因素对 2019 年各地区肺结核发病率的影响。

2. 变量指标与数据的选取

本文选取的被解释变量为 2019 年 31 个省份、直辖市、自治区的肺结核发病率(1/10 万)，将其记为 y ，此数据来源于公共卫生科学数据中心[6]。选取的解释变量有各地区人均国内生产总值(元)、文盲人口占 15 岁以上人口比重(%)、生活垃圾无害化处理率(%)、15~64 岁年龄组构成比(%)、0~14 岁年龄组构成比(%)、城镇人口构成比(%)、男性性别构成比(%)，以及地方一般公共预算支出中的卫生健康支出百分比(%)、节能环保支出百分比(%)、教育支出百分比(%)，以上数据来源于国家统计局[7]。解释变量选取及符号表示见表 1，原始数据见附录表 1。

Table 1. Explanatory variable selection and symbolic representation**表 1.** 解释变量选取及符号表示

解释变量	符号表示	解释变量	符号表示
15~64 岁年龄组构成比(%)	x_1	节能环保支出百分比(%)	x_6
0~14 岁年龄组构成比(%)	x_2	人均国内生产总值(元)	x_7
城镇人口构成比(%)	x_3	教育支出百分比(%)	x_8
男性性别构成比(%)	x_4	文盲人口占 15 岁以上人口比重(%)	x_9
卫生健康支出百分比(%)	x_5	生活垃圾无害化处理率(%)	x_{10}

3. 模型的建立与研究

将所选取的十个指标变量分别与 2019 年各地区结核病发病率进行相关性分析, 采取 Spearman 秩相关性检验、Kendall τ 相关检验两个非参数检验方法, 比较检验结果, 判断指标与结核病发病率是否具有相关性[8]。之后选取具有相关关系的指标与结核病发病率建立多元线性回归模型, 进一步研究各指标对各地区结核病发病率的影响效果。

3.1. 相关性分析

3.1.1. 方法基本原理

1) Spearman 秩相关检验

假设有样本量为 n 的样本为 $(X, Y) = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$, 将观测值在样本中的位置用秩代替, 即用 R_i 表示 X_i 在 (X_1, X_2, \dots, X_n) 中的秩, Q_i 表示 Y_i 在 (Y_1, Y_2, \dots, Y_n) 中的秩。若 X_i 和 Y_i 具有同步性, 则其秩 R_i 和 Q_i 也应具有同步性, 可定义 Spearman 秩相关系数为:

$$r_s = \frac{\sum_i \left[\left(R_i - \frac{1}{n} \sum_i R_i \right) \left(Q_i - \frac{1}{n} \sum_i Q_i \right) \right]}{\sqrt{\sum_i \left(R_i - \frac{1}{n} \sum_i R_i \right)^2 \sum_i \left(Q_i - \frac{1}{n} \sum_i Q_i \right)^2}} \quad (1)$$

并可定义 T 检验统计量进行两变量的相关性分析。

2) Kendall τ 相关检验

对于 n 对观测值 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, 若 $j > i, i, j = 1, 2, \dots, n$, 有 $(x_j - x_i)(y_j - y_i) > 0$, 则说明 (x_i, y_i) 与 (x_j, y_j) 之间具有协同性。Kendall τ 相关检验即从两变量协同一致的角度出发检验其是否具有相关性。Kendall 相关系数统计量定义为:

$$\tau = \frac{N_c - N_d}{n(n-1)/2} \quad (2)$$

其中, N_c 表示同向数对的数目, N_d 表示反向数对的数目, 即对于 n 对观测值, 有 $N_c + N_d = n(n-1)/2$ 。

3.1.2. 相关性检验结果

由于 Spearman 秩相关检验和 Kendall τ 相关检验均为非参数统计方法不受数据分布的影响[9], 不需要对数据进行标准化处理, 故利用原始数据应用上述两种相关性检验方法进行比较检验, 相关性检验结果如表 2, 其中相关系数数值均为绝对值, 可表示各个指标与 2019 年各地区结核病发病率的相关程度。从检验结果可以看出, 15~64 岁年龄组构成比、0~14 岁年龄组构成比、城镇人口构成比、人均国内生产总值与生活垃圾无害化处理率的检验 P 值均小于 0.05, 即以上五个指标与 2019 年各地区肺结核发病率具有

一定的相关关系。而男性性别构成比、文盲人口占 15 岁以上人口比重以及地方一般公共预算支出中的卫生健康支出百分比、节能环保支出百分比、教育支出百分比与 2019 年各地区肺结核发病率不相关。为进一步分析, 接下来将对具有相关关系的五个指标与 2019 年各地区结核病发病率建立回归模型, 具体探究各指标对发病率的影响。

Table 2. Correlation test results
表 2. 相关性检验结果

指标	Spearman		Kendall τ	
	相关系数值	P 值	相关系数值	P 值
15~64 岁年龄组构成比(%)	0.4137097	0.02148	0.3032258	0.01643
0~14 岁年龄组构成比(%)	0.5931452	0.0005577	0.4451613	0.0003084
城镇人口构成比(%)	0.5596774	0.001277	0.4021505	0.001227
男性性别构成比(%)	0.1455645	0.433	0.08387097	0.5215
卫生健康支出百分比(%)	0.2112903	0.2528	0.1483871	0.2501
节能环保支出百分比(%)	0.2814516	0.125	0.1913978	0.1356
人均国内生产总值(元)	0.4016129	0.02592	0.2903226	0.02185
教育支出百分比(%)	0.1016129	0.5851	0.07526882	0.5664
文盲人口占 15 岁以上人口比重(%)	0.2780522	0.1299	0.1722283	0.1739
生活垃圾无害化处理率(%)	0.3611723	0.04591	0.2623406	0.04743

3.2. 多元线性回归模型

3.2.1. 普通最小二乘回归

为避免解释变量量纲的影响, 对相关指标的原始数据进行标准化处理, 标准化后的数据见附录表 2, 并对所选取的 15~64 岁年龄组构成比、0~14 岁年龄组构成比、城镇人口构成比、人均国内生产总值与生活垃圾无害化处理率五个指标建立如下多元线性回归模型:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_7 + \beta_5 x_{10} + \varepsilon$$

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_7 + \beta_5 x_{10}$$
(3)

其中 ε 为随机误差项服从均值为 0, 方差为 σ^2 的正态分布。若记 X 为自变量矩阵, 则最小二乘估计即寻找参数 β_i , $i=1,2,\dots,5$ 的估计值 $\hat{\beta}_i$, 使随机误差

$$Q(\hat{\beta}) = \|y - X\hat{\beta}\|^2$$
(4)

取到最小值。

对多元线性回归方程进行显著性检验, 得到 F 检验统计量值为 9.221, P 值为 4.494×10^{-5} 小于 0.05, 则可知该回归方程显著, 解释变量全体可对被解释变量 2019 年各地区肺结核发病率产生显著线性影响。

接下来对回归模型的基本假定条件进行验证。首先进行正态性检验, 采用 Shapiro-Wilk 检验、Kolmogorov-Smirnov 检验以及 Anderson-Darling 检验[10], 并画出回归模型残差的 Q-Q 图。得到以上三种检验的 P 值分别为 0.5987、0.3656 和 0.8206, 均通过了正态性检验, Q-Q 图见图 1, 表明几乎所有的点都落在了直线附近, 综上可说明该回归方程满足正态性的假定条件。

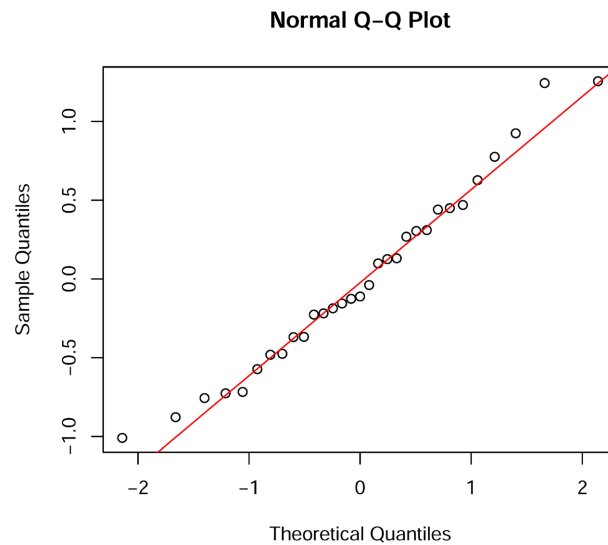


Figure 1. Least squares regression model Q-Q plot
图 1. 最小二乘回归模型 Q-Q 图

检验回归方程随机误差项是否具有同方差性, 其残差图如图 2 所示, 从残差图可以看出残差的变化范围较大, 则需考虑数据是否存在异常值, 是否存在自相关性以及异方差性。

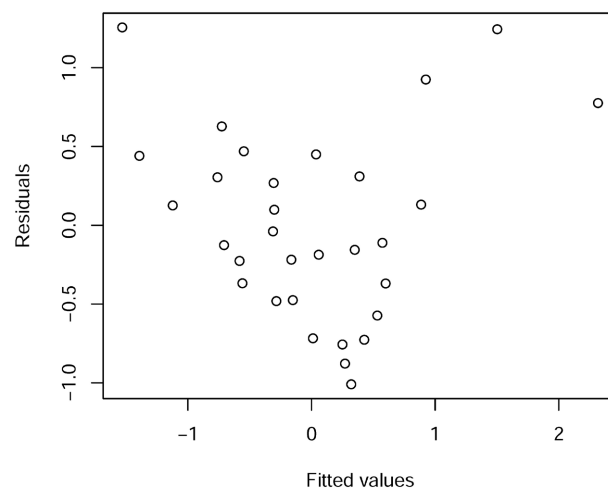


Figure 2. Least squares regression model residual plots
图 2. 最小二乘回归模型残差图

首先利用 D.W 检验判断残差序列是否存在自相关性, 得到其 P 值为 $0.06262 > 0.05$, 可以认为残差序列不存在自相关性。之后应用等级相关系数检验法诊断误差项是否具有异方差性, 得到检验的 P 值为 $0.4044 > 0.05$, 认为残差序列不存在异方差性。利用学生化残差值判断是否存在异常值点, 其观测值的学生化残差值、Cook 距离值和杠杆值见附录表 3, 第 6 个数据的学生化残差为 $SRE_{(6)} = 2.40210507 > 2$, 第 31 个数据的学生化残差为 $SRE_{(31)} = 2.37782078 > 2$, 因而可以诊断第 6 个数据和第 31 个数据为异常值点。由于第 6 个数据和第 31 个数据的 COOK 距离小于 0.5, 其杠杆值也小于 2 倍的杠杆值平均值, 所以第 6 个和第 31 个数据并不是强影响点。通过核实, 本文数据不存在登记错误, 故为消除异常值对建模的影响, 将删除异常值数据, 用其余数据拟合回归方程。

最后检验数据是否具有多重共线性, 采用方差扩大因子法(VIF) [11]和条件数法, 得到的条件数值为

$k = 5.968987 < 10$, 各指标的方差扩大因子值见表 3, 可以看到五个指标的方差扩大因子均小于 10, 故认为数据不存在多重共线性, 满足自变量之间不相关的假定。

Table 3. Variance expansion factor values

表 3. 方差扩大因子值

	15~64 岁年龄组 构成比(%)	0~14 岁年龄组 构成比(%)	城镇人口 构成比(%)	人均国内生产总值(元)	生活垃圾无害化处理率(%)
VIF	2.760155	4.505508	6.568996	3.766278	1.104850

3.2.2. 回归拟合结果

使用标准化后的数据, 将异常值点删除后进行最小二乘估计, 得到的描述性统计量见表 4。

Table 4. Results of descriptive statistics of weighted least squares regression

表 4. 加权最小二乘回归描述性统计结果

解释变量	系数估计值	标准误差	统计量值	P 值
截距	-0.09293	0.10231	-0.908	0.37317
15-64 岁年龄组构成比(%)	0.48379	0.17110	2.827	0.00955
0-14 岁年龄组构成比(%)	0.85589	0.24229	3.533	0.00178
城镇人口构成比(%)	-0.59037	0.26504	-2.228	0.03598
人均国内生产总值(元)	0.45181	0.20438	2.211	0.03728
生活垃圾无害化处理率(%)	-0.32382	0.11020	-2.939	0.00738

由表 4 可以得到, 15~64 岁年龄组构成比、0~14 岁年龄组构成比、城镇人口构成比、人均国内生产总值与生活垃圾无害化处理率的回归系数检验 P 值均小于 0.05, 回归系数显著, 则可以得到删除异常值点后最小二乘估计得到的多元线性回归模型为:

$$y = 0.48379x_1 + 0.85589x_2 - 0.59037x_3 + 0.45181x_7 - 0.32382x_{10} \quad (5)$$

4. 结论

通过以上研究可得出结论, 15~64 岁年龄组构成比、0~14 岁年龄组构成比与 2019 年各地区肺结核发病率成正相关。说明肺结核的患病率与人口的年龄结构显著相关, 而本文仅研究 2019 年的各地区肺结核发病率, 且未细致区分年龄结构, 故可以进一步应用更多数据, 研究不同年龄组人群肺结核患病率的差异, 并针对不同人群给出不同的防治建议。人均国内生产总值也与 2019 年各地区肺结核发病率成正相关, 说明地区经济发展水平有助于肺结核病的检测, 提供更好的治疗条件。城镇人口构成比、生活垃圾无害化处理率与 2019 年各地区肺结核发病率成负相关关系, 其原因可能由于城镇的建设更完善, 更注重优化卫生环境, 故为了降低结核病的发病率, 我们应该致力于改善人民的生活条件, 推广健康教育和普及结核病预防知识, 并应该重视环境保护问题, 提高垃圾无害化处理率。

基金项目

国家自然科学基金项目(NSFC11901027), 北京建筑大学教学实践类项目(J1703)资助。

参考文献

- [1] 国务院办公厅关于印发“十四五”国民健康规划的通知[EB/OL].

-
- https://www.gov.cn/zhengce/content/2022-05/20/content_5691424.htm, 2022-05-20.
- [2] 尹金凤, 黄若雯, 姜慧, 高志东, 许王莉, 贺晓新, 李卫民. 北京市 2008-2018 年肺结核时空分布及影响因素分析[J]. 中华流行病学杂志, 2021, 42(7): 1240-1245.
- [3] 赵明扬, 周乾宇, 王荣荣, 等. 基于时空地理加权回归模型的中国肺结核发病情况及影响因素研究[J]. 中国全科医学, 2023, 26(5): 583-590.
- [4] Bie, S., Hu, X., Zhang, H., Wang, K. and Dou, Z. (2021) Influential Factors and Spatial-Temporal Distribution of Tuberculosis in Mainland China. *Scientific Reports*, **11**, Article No. 6274. <https://doi.org/10.1038/s41598-021-85781-7>
- [5] 刘家起, 孙德斌, 姜婧, 等. 2000-2015 年全球结核病患病率宏观影响因素分析[J]. 中国公共卫生管理, 2021, 37(2): 261-264.
- [6] 公共卫生科学数据中心. 肺结核数据库[EB/OL]. https://www.phsciencedata.cn/Share/ky_sjml.jsp?id=f90892b6-c000-48fe-a73e-a4c6db172385, 2023-07-22.
- [7] 国家统计局. 地区数据分省年度数据[EB/OL]. <https://data.stats.gov.cn/easyquery.htm?cn=C01>, 2023-07-22.
- [8] 宋德胜, 刘媛媛, 李长平, 崔壮, 胡良平. 如何正确运用 t 检验——相关系数与 0 比较 t 检验及 SAS 实现[J]. 四川精神卫生, 2020, 33(4): 312-316.
- [9] 王星, 褚挺进. 非参数统计[M]. 第 2 版. 北京: 清华大学出版社, 2014.
- [10] 杨斌. 正态性检验的几种方法比较[J]. 统计与决策, 2015(14): 72-74.
- [11] Cheng, J., Sun, J., Yao, K., Xu, M. and Cao, Y. (2021) A Variable Selection Method Based on Mutual Information and Variance Inflation Factor. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **268**, Article ID: 120652. <https://doi.org/10.1016/j.saa.2021.120652>

附录

Table 1. Raw data
附表 1. 原始数据

地区	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
北京市	32.22076007	78.12312	10.42242	87.3516	50.39002	7.213812	4.168438	161776	15.35028	1.73	100
天津市	24.98078524	77.64365	10.27697	84.33213	55.18809	5.564601	6.813999	101557	13.15154	1.72	100
河北省	39.24146715	68.0285	18.82817	58.77535	51.04338	8.365236	6.047597	47036	18.49895	2.54	99.43391
山西省	33.88054438	73.63505	15.3695	61.2811	50.57218	7.783833	4.800656	48469	14.78055	2	100
内蒙古	49.86228381	76.77269	13.02674	66.45963	51.61568	6.316087	3.028526	71170	11.958	4.15	99.81033
辽宁省	52.75074705	73.88166	10.19527	71.21814	50.17751	6.345246	2.258058	58019	12.22569	1.29	99.41647
吉林省	39.97787275	74.9415	11.76752	61.64216	50.47997	7.161543	3.761795	47554	12.72501	2.25	90.24263
黑龙江省	59.56538315	76.24841	9.973609	64.60829	50.68376	6.273966	4.211442	41156	11.0769	2.58	95.48578
上海市	26.71617823	73.70242	10.0346	89.23821	51.89247	6.032766	2.250487	153299	12.17338	2.49	100
江苏省	31.25070801	71.1342	13.78727	72.46428	50.83158	7.205697	2.964735	116650	17.60708	4.75	100
浙江省	45.79047724	72.91061	13.05592	71.57647	51.8406	7.317308	2.681265	98770	17.5538	4.52	100
安徽省	48.65426605	67.38429	18.64735	57.02561	50.52661	9.298405	4.222208	60561	16.53371	6.43	100
福建省	43.86183153	73.13162	16.87129	67.87527	51.80926	9.211703	3.535819	102722	19.07361	6.81	99.94582
江西省	70.80011227	69.64315	20.20761	59.07883	51.05182	9.879636	3.04195	54640	17.98246	2.95	100
山东省	27.66930707	66.49174	17.67282	61.86424	50.47499	8.492442	2.853841	69901	20.07621	7.33	99.94156
河南省	51.79907	67.33789	21.0609	54.01475	50.50142	9.708666	3.46605	54356	17.81505	4.45	99.7
湖北省	61.88442703	71.3542	15.5745	61.83567	50.35899	7.550839	3.539834	76712	14.39237	4.84	100
湖南省	82.07188398	66.86784	19.98849	57.45482	50.9517	8.234319	3.020535	60104	15.80731	2.71	100
广东省	58.40030391	75.45539	15.96574	72.64793	54.13904	9.131778	4.320992	86956	18.56017	2.99	100
广西	81.12474191	68.00248	21.83013	52.97069	51.68551	9.661425	1.706395	42778	17.33937	2.73	100
海南省	90.21501549	71.18367	19.5102	59.39698	51.53741	9.136026	3.496331	53929	14.71562	3.84	100
重庆市	72.20996406	67.87097	16.77927	68.22459	50.0412	7.906016	3.56773	74337	15.02281	3.15	88.82224
四川省	60.34885226	67.84762	16.41423	55.35864	49.16909	9.115296	2.580226	55619	15.25759	6.81	99.81666
贵州省	102.5111481	66.171	22.24819	51.48129	51.7897	8.989756	3.16924	43727	17.94694	10.19	96.59305
云南省	61.78479805	72.07847	18.03876	48.66356	51.01389	8.988099	3.030355	49323	15.80262	7.31	99.77232
西藏	182.3822894	67.91834	26.06635	34.62604	50.05468	5.624644	1.869785	47491	12.03334	33.11	98.34122
陕西省	55.55081713	73.40341	14.55004	61.28296	50.73944	8.154122	4.29279	65506	16.6342	3.97	99.70824
甘肃省	36.18137615	71.37874	17.10106	50.69749	50.54258	8.260175	2.695516	34707	16.09606	10.48	100
青海省	134.5292595	71.96281	19.5225	58.81356	51.65857	7.953682	3.722906	49976	11.87813	10.6	96.28407
宁夏	36.86045976	69.76314	20.72539	63.59833	51.44338	7.404048	3.757152	52537	12.46808	8.39	99.89037
新疆	169.0512627	67.79247	24.14039	55.5295	50.43235	5.688322	1.652978	53542	16.23686	3.71	96.26037

Table 2. Post-standardization data
附表 2. 标准化后数据

地区	y	x_1	x_2	x_3	x_7	x_{10}
北京市	-0.80911496	1.93123254	-1.459714628	2.18192459	2.99052564	0.4715744
天津市	-0.99724299	1.79567525	-1.493198139	1.91547611	1.05221341	0.4715734
河北省	-0.62668453	-0.92274872	0.475436567	-0.33974431	-0.7026932	0.2674173
山西省	-0.76598609	0.66235041	-0.320808975	-0.11862793	-0.6565682	0.4715744
内蒙古	-0.35070659	1.54943382	-0.860153331	0.33834369	0.0741252	0.4031715
辽宁省	-0.27565096	0.73207248	-1.51200868	0.75825188	-0.34917549	0.2611273
吉林省	-0.6075493	1.03171347	-1.150049462	-0.08676683	-0.68601996	-3.0473908
黑龙江省	-0.0985752	1.40120912	-1.563037983	0.17497561	-0.89195699	-1.156466
上海市	-0.95214946	0.68139875	-1.548996389	2.34840589	2.71767035	0.4715744
江苏省	-0.8343214	-0.04469594	-0.685066689	0.86821547	1.53802265	0.4715744
浙江省	-0.45651096	0.45753593	-0.853435538	0.78987191	0.96250624	0.4715744
安徽省	-0.38209648	-1.10488276	0.433809135	-0.49414738	-0.26735431	0.4715744
福建省	-0.50662609	0.52002013	0.024929962	0.46326514	1.0897121	0.4520345
江西省	0.19335502	-0.46624973	0.793008244	-0.31296373	-0.45793779	0.4715744
山东省	-0.92738276	-1.35722636	0.209455331	-0.06716952	0.03327899	0.4504988
河南省	-0.30037993	-1.11800062	0.989451032	-0.75983636	-0.4670791	0.3633804
湖北省	-0.03831572	0.01750279	-0.273615031	-0.06969079	0.25250954	0.4715744
湖南省	0.48624777	-1.25089423	0.74256211	-0.45627221	-0.28206409	0.4715744
广东省	-0.12884935	1.17700263	-0.183545031	0.88442127	0.58224053	0.4715744
广 西	0.46163664	-0.93010424	1.166540535	-0.85196714	-0.83974851	0.4715737
海南省	0.69784399	-0.03070912	0.632453067	-0.28488869	-0.48082326	0.4715744
重庆市	0.22998949	-0.96728529	0.003744235	0.49409039	0.17606371	-3.559649
四川省	-0.07821705	-0.97388883	-0.08029515	-0.64124646	-0.42642601	0.4054544
贵州省	1.01735438	-1.44790551	1.262784324	-0.98339751	-0.80920236	-0.7571318
云南省	-0.04090454	0.22227053	0.293700904	-1.23204422	-0.62907989	0.3894611
西 藏	3.09277599	-0.95389426	2.141792789	-2.47076401	-0.68804779	-0.1266599
陕西省	-0.20289219	0.59686081	-0.509464347	-0.1184635	-0.10818604	0.3663508
甘肃省	-0.70619984	0.02444132	0.077825028	-1.05256277	-1.09953559	0.4715744
青海省	1.84933298	0.18957181	0.635284166	-0.33637222	-0.60806131	-0.868564
宁夏	-0.68855411	-0.43232687	0.912209815	0.08585244	-0.52562856	0.4320382
新疆	2.74637427	-0.98947933	1.698402129	-0.62616882	-0.4932799	-0.8771133

Table 3. Diagnostic analysis of outliers
附表 3. 异常值的诊断分析

序号	学生化残差	cook 距离	杠杆值
1	-0.43992049	0.0214491	0.391555
2	0.21574128	0.0023763	0.227588
3	-0.76730361	0.0117137	0.105074
4	-0.77807579	0.0121691	0.106107
5	-0.06370156	0.000118	0.143502
6	2.40210507	0.2397379	0.228904
7	-1.89737305	0.4258252	0.439309
8	1.14363247	0.0830255	0.278277
9	0.81306292	0.0512613	0.314575
10	-0.20975185	0.0015776	0.171445
11	0.48165688	0.0036064	0.082911
12	-0.34972035	0.0025034	0.105945
13	-1.25477749	0.0343718	0.118161
14	-0.24697076	0.0010168	0.087813
15	-0.61787547	0.0141802	0.178543
16	-1.17587827	0.0198114	0.080278
17	0.42069041	0.0022853	0.069704
18	0.72383061	0.0102642	0.103386
19	-0.31296722	0.0038143	0.183819
20	-0.17736102	0.0005911	0.097775
21	0.49315812	0.0042878	0.093036
22	-0.86266186	0.1656811	0.569362
23	0.79713728	0.0252187	0.190057
24	0.21057164	0.0010581	0.121035
25	-0.97840522	0.037017	0.18806
26	1.78780775	0.5196669	0.514849
27	0.15472912	0.0003488	0.077493
28	-1.20072483	0.0378446	0.138138
29	1.58025036	0.0636271	0.139439
30	-1.84698909	0.1490769	0.223295
31	2.37782078	0.2380612	0.230567