



兰州大学

新冠的预测与控制

团队成员： 阮黎光 姚忱 李嘉康

所在学院： 信息科学与工程学院

提交时间： 2021 年 6 月 6 日

摘 要

2019年中国湖北省武汉市爆发新型冠状病毒肺炎疫情,随着2020年春运的到来,人员大量流动导致该病毒从武汉市迅速扩散至全国各地。目前,新型冠状病毒的传播已成为全国乃至全球的突发公共卫生事件,严重威胁人类生命安全和社
会稳定。因此对于疫情风险等级评估、发展趋势预测等研究具有重大现实意义。

针对问题一,本文先对数据集进行了预处理,通过对累计确诊人数的计算得到了各城市十四天内新增确诊人数指标,并简化了其他分类指标,忽略了纬度、月份平均温度等指标的影响。最终选择了累计病例、新增病例、人口密度、人均GDP 4个指标作为分类依据。然后分别采用了层次分析法和系统聚类法两种方法对各城市的疫情情况进行了分类,最终结果显示系统聚类法的分类效果并不明显,而层次分析法能够得到较好的量化风险指标,分类效果良好。

针对问题二,本文首先根据数据对城市疫情发展情况进行了定义,采用移动平均法对累计确诊人数进行平滑处理,得到了修正后的疫情发展情况指标。然后从经济、政治、文化、社会环境层面对不同的城市数据进行了讨论,利用了Spearman检验对其和修正后的疫情发展情况指标进行相关性计算,给出了哪些城市数据与疫情发展情况具有较强的相关性。

针对问题三,本文建立了时间序列ARIMA模对温州市的新冠疫情情况进行预测和评估,为了提高预测的准确度,又另外采用了SEIR传染病动力学模型来对比分析。并将二者预测结果与温州市的实际疫情发展情况作对比,给出了疫情防控的一些控制建议。

最后,对本文所建立的模型的优点和缺点进行了客观评价,并给出了一些模型改进的方向。

关键词: 新冠疫情 层次分析法 相关性分析 时间序列模型 SEIR模型

1. 前言

1.1 问题背景

2019 年新型冠状病毒病疫情是一次由严重急性呼吸系统综合征冠状病毒所引发的全球大流行疫情。疾病在 2019 年末月于湖北省武汉市首次被发现，随后在 2020 年初迅速扩散至全球多国，逐渐变成一场全球性大瘟疫。截至 2021 年 6 月 4 日，全球已有个国家和地区累计报告逾 1.72 亿例确诊病例，其中逾 370.1 万人死亡，是人类历史上大规模流行病之一。而中国在疫情防控方面所取得的巨大成就有目共睹，在疫情爆发的第一时间就采取最全面最严格最彻底的防控措施，有效阻断病毒传播链条。在此过程中，如何对一个地区的疫情情况进行科学的评估和预测，并据此提出相应的政策是做好疫情防控工作的重要环节。

1.2 问题重述

➤ 任务一

根据给出的相关数据，对各城市的疫情情况进行分类，标注出哪些城市疫情严重，哪些城市疫情轻微。

➤ 任务二

探讨疫情发展情况与哪些城市数据有关系，数据来源可以来自题中给定的数据文件，也可以自行从网上搜索。

➤ 任务三

根据数据建立 Wenzhou(温州)的疫情模型，并给出疫情的预测与控制建议。

2. 模型准备

2.1 基本假设

以下假设均适用于本文中的所有模型

1. 所给的数据均是真实准确可靠的，不考虑在数据统计时存在偏差。
2. 在统计的时间段内，病毒的传播方式，传播途径，传播能力等没有发生变化。
3. 在统计时间段内，某地区的人数及人口密度保持恒定不变。
4. 所有制定的政策都符合国家法律，考虑社会发展和当地的实际情况。
5. 国家的响应措施及其相关政策的执行力度只会增强不会减弱；
6. 每个个体之间的接触机会均等,每个个体都有被感染的可能性

3. 模型的建立与求解

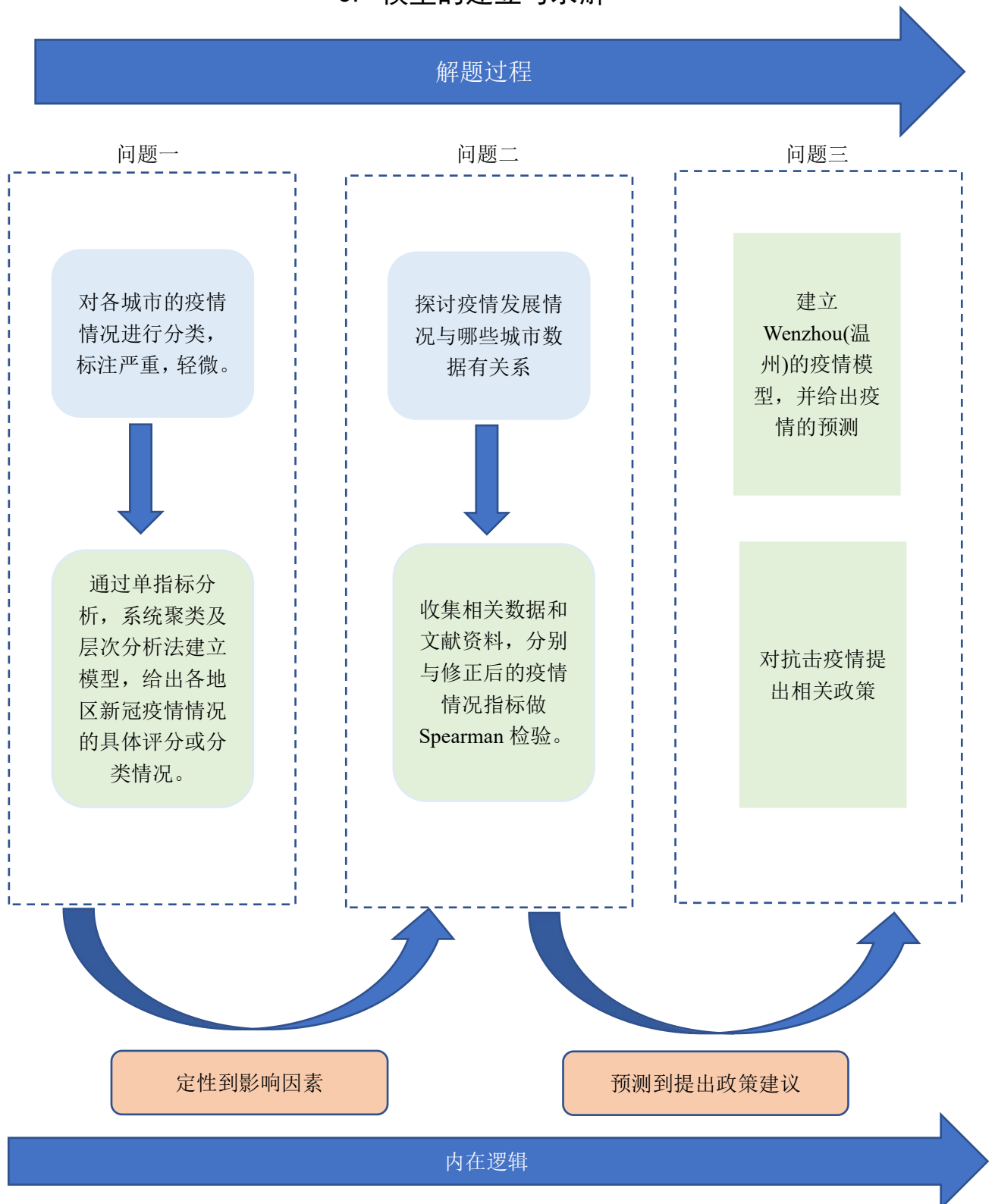


图 1.问题求解思路

3.1 问题一：新冠疫情等级评估模型的建立与求解

3.1.1 问题一分析

由题可知，我们需对数据集中的各个城市疫情情况进行分类，简而言之就是根据不同城市的不同指标对其进行分类，判断哪些城市疫情严重，哪些城市疫情轻微。

早在 2020 年 2 月 17 日，国务院联防联控机制就印发了《关于科学防治、精准施策、分区分级、做好新冠肺炎疫情防控工作的指导意见》[1]，同时颁布了疫情风险评价标准，即单指标评价方法。我国各省份也结合实际情况给出了各个地区的风险等级评估标准和结果。该评价方法直接通过目前最近的确诊病例、是否发生聚集性疫情等指标进行评判，存在一定的局限性，因此进一步探讨新的风险综合评价方法具有非常重要的实际应用价值和指导意义。

针对本题，我们主要采用了层次分析法和系统聚类法两种主流的风险综合评价方法进行分析，并针对其优缺点进行了不同程度的讨论。

3.1.2 单指标评价方法

根据所查询的国务院的评价标准，我们将疫情情况分级的流程进行了总结，如右图所示：

但由于传染病流行的“风险”不仅与目前的流行状况有关，也与该流行地区的社会经济、人口密度等因素有关，所以该评价方法考虑较为片面。再者，这种评价方式的主观性太强。例如，考虑连续“14 天”无新增确诊病例这一指标，主要依据 14 天为潜伏期时间，但迄今为止潜伏期还没有一个确切的结论。综上，在本文中我们主要采用了系统聚类和综合评价法来对不同城市的疫情情况进行评估。

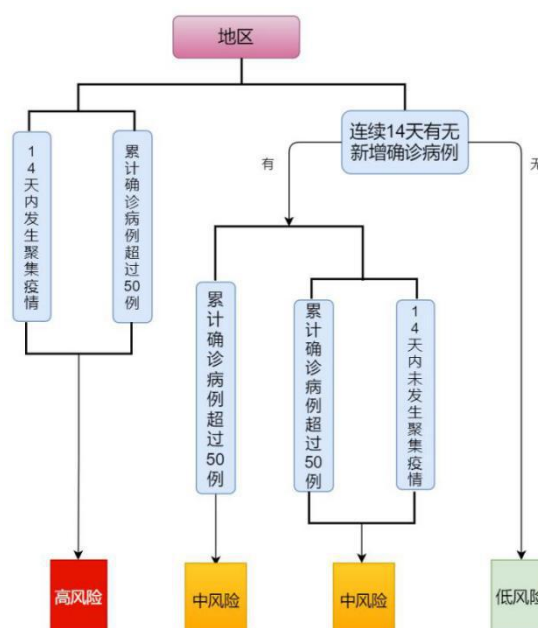


图 2.单指标评价法流程图

3.1.3 系统聚类法

该方法是首先将每个样本单独看成一类，在规定类间距离的条件下，选择距离最小的一对合并成一个新类，并计算新类与其他类之间的距离，再将距离最近的两类合并，这样每次会减少一个类，直到所有的样本合为一类为止。

(1) 数据预处理

由于月份平均温度、城市所处纬度、老人百分比等因素对疫情风险等级的评估影响较小，故在后文中我们并未考虑这些因素。根据单指标评价法中 14 天内

新增确诊病例对疫情风险等级的影响，我们对数据集进行了处理，选取了每座城市最新一天的前 14 天内新增病例作为一个指标，而对于每个地区的累计确诊病例数、前 14 天新增病例数、人口密度和人均 GDP 指标视为一个风险“向量”，比如温州的风险向量为 $\alpha = (504, 140, 767396, 64931.46)$ 。

再将不同城市的风险向量合并列成如下矩阵形式：

$$\begin{array}{c}
 \text{城市1} \\
 \text{城市2} \\
 \dots \\
 \text{城市200}
 \end{array}
 \begin{array}{c}
 \text{累计病例} \quad \text{新增病例} \quad \text{人口密度} \quad \text{人均GDP} \\
 \left[\begin{array}{cccc}
 X_{1,1} & X_{1,2} & X_{1,3} & X_{1,4} \\
 X_{2,1} & X_{2,2} & X_{2,3} & X_{2,4} \\
 \dots & \dots & \dots & \dots \\
 X_{200,1} & X_{200,2} & X_{200,3} & X_{200,4}
 \end{array} \right]
 \end{array}$$

从该矩阵中我们可以明显的看出每座城市所对应的数据大小，但由于各指标的绝对数相差较大，因此我们采用了标准差标准化对数据进行处理，其理论公式如下：

$$X' = \frac{X - \mu}{\sigma}$$

其中，人均 GDP 由于其与风险等级指数呈负相关，故其标准化之后需在其前面加一个负号。 μ 为该组数据所对应的均值， σ 为该组数据所对应的标准差，利用该标准化方法能够将不同属性的数据统一归一化至同等量级，便于后续进行处理。

(2) 聚类模型建立与求解

在本文题中，我们选取组间联接的方法进行聚类，所测量距离用平方欧式距离表示，并将聚类数设置为 3 类，即疫情情况严重、疫情情况轻微和疫情风险等级中等。

根据不同城市基本数据的差异性可知，最终分类结果 3 为疫情情况严重，分类结果 2 为疫情风险等级中等，分类结果 1 为疫情情况轻微，最终聚类结果为 1 所城市疫情严重，3 所城市疫情情况中等，其余情况轻微其部分结果如下图所示：

聚类成员	
个案	3 个聚类
1:Wenzhou	1
2:Shanghai	1
3:Chongqing	1
4:Arizona-Maricopa	1
5:Arizona-Pima	1
6:Arizona-Pinal	1
7:Arkansas-Jefferson	1
8:California-Alameda	1
9:California-Contra Costa	1
10:California-Fresno	1
11:California-Humboldt	1
12:California-Los Angeles	1
13:California-Madera	1
109:New Mexico-Santa Fe	1
110:New Mexico-Socorro	1
111:New York-Nassau	2
112:New York-New York	3
113:New York-Rockland	1
114:New York-Saratoga	1
115:New York-Suffolk	2
116:New York-Ulster	1
117:New York-Westchester	2
118:North Carolina-Chatham	1
119:North Carolina-Forsyth	1

由图可知，New York-New York 的疫情情况严重，仅有 New York-Nassau 等三地被评为了疫情风险等级中等，温州、上海等 197 地被评为了疫情情况轻微，评价效果较差。

图 3.单指标评价法流程图

3.1.4 层次分析法

层次分析法是基于多指标的综合评价方法。首先，通过查阅相关文献，得到针对 4 个评价指标的判断矩阵，经过矩阵的一致性检验后，计算出 4 个指标的权重向量。采用线性变换方法，得到累计确诊病例数、最近 14 天累计新增病例、人口密度和 GDP 指标的风险评分矩阵，乘以权重向量得到各个地区的综合总评分。采用层次分析方法可以得到风险评分，从而可以对各个地区的风险程度进行综合排序，确定各地区的风险分类。

(1) 数据预处理

与系统聚类法相类似，由于月份平均温度、城市所处纬度、老人百分比等因素对疫情风险等级的评估影响较小，故在后文中我们并未考虑这些因素。根据单指标评价法中 14 天内新增确诊病例对疫情风险等级的影响，我们对数据集进行了处理，选取了每座城市最新一天的前 14 天内新增病例作为一个指标，然后将其与其他有效的城市数据列成如下矩阵形式：

$$\begin{array}{c} \text{城市1} \\ \text{城市2} \\ \dots \\ \text{城市200} \end{array} \begin{array}{c} \text{累计病例} \quad \text{新增病例} \quad \text{人口密度} \quad \text{人均GDP} \\ \left[\begin{array}{cccc} X_{1,1} & X_{1,2} & X_{1,3} & X_{1,4} \\ X_{2,1} & X_{2,2} & X_{2,3} & X_{2,4} \\ \dots & \dots & \dots & \dots \\ X_{200,1} & X_{200,2} & X_{200,3} & X_{200,4} \end{array} \right] \end{array}$$

从该矩阵中我们可以明显的看出每座城市所对应的数据大小，但由于各指标的绝对数相差较大，因此我们采用了标准差标准化对数据进行处理，其理论公式如下：

$$X' = \frac{X - \mu}{\sigma}$$

其中， μ 为该组数据所对应的均值， σ 为该组数据所对应的标准差，利用该标准化方法能够将不同属性的数据统一归一化至同等量级，便于后续进行处理。另外，由于人均 GDP 与风险值成反相关，所以在计算人均 GDP 标准化值时，需要改变符号。

(2) 层次分析模型的建立和求解

第一步：建立层次结构模型，如下图所示

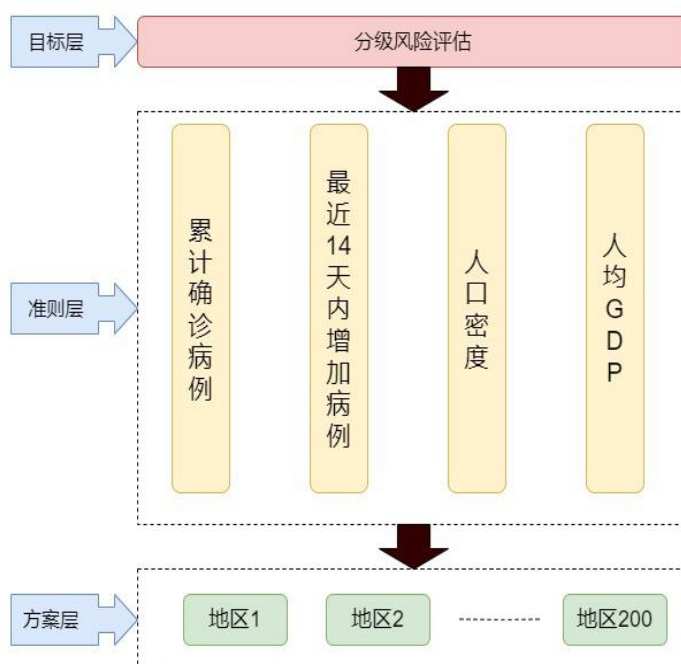


图 4.层次分析法结构

第二步：构造判断矩阵。通过查阅相关文献数据[2]，对层次结构模型准则层的4个风险影响因素构造判断矩阵：

$$\begin{vmatrix} 1 & 1/3 & 7 & 7 \\ 3 & 1 & 9 & 9 \\ 1/7 & 1/9 & 1 & 3 \\ 1/7 & 1/9 & 1/3 & 1 \end{vmatrix}$$

第三步：计算权重向量 W

将判断矩阵按列规范化，即对判断矩阵 A 每一列归一化：

$$\bar{a}_{ij} = \frac{a_{ij}}{\sum_{i=1}^n a_{ij}}$$

按行相加得和向量： $W_i = \sum_{j=1}^n \bar{a}_{ij}$

将得到的和向量正规化，可得权重向量： $\bar{W}_i = \frac{W_i}{\sum_{i=1}^n W_i}$

在本题中，计算出权重向量为：[0.3006 0.5885 0.0706 0.0403]

第四步：判断矩阵和权重向量的一致性检验。

一致性检验是指对判断矩阵 A 确定不一致的允许范围。 n 阶一致阵的唯一非零特征根为 n ， n 阶正互反阵 A 的最大特征根 $\lambda_{\max} \geq n$ 时， A 为非一致矩阵， λ_{\max} 比 n 大的越多， A 的不一致性越严重；当且仅当 $\lambda_{\max} = n$ 时， A 为一致矩阵。因此可由 λ_{\max} 是否等于 n 来检验判断矩阵 A 是否为一致矩阵。

具体的一致性指标用 CI 计算， CI 越小，说明一致性越大。 $CI = 0$ ，有完全的一致性； CI 接近于 0，有满意的一致性； CI 越大，不一致越严重。

$$CI = \frac{\lambda_{\max} - n}{n - 1}$$

考虑到一致性的偏离可能是由于随机原因造成的，因此引入随机一致性指标 RI 衡量随机因素所造成的一致性偏离的大小：

$$RI = \frac{CI_1 + CI_2 + \cdots + CI_n}{n}$$

随机一致性指标 RI 和判断矩阵的阶数有关，一般情况下，矩阵阶数越大，则出现一致性随机偏离的可能性也越大， RI 指标通过查表获得：

矩阵阶数 n	1	2	3	4	5	6	7	9	10
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.45	1.49

在本题中由于判断矩阵 A 为四阶矩阵，所以取 $RI=0.90$ 。

最终使用的检验统计量为检验系数 CR ，公式如下：

$$CR = \frac{CI}{RI}$$

所以在本题中需要计算判断矩阵 A 的随机一致性比率 CI ，当 $CI < 0.1$ 时，判定判断矩阵 A 具有满意的一致性，或其不一致程度是可以接受的。同时对每一个成对比较矩阵 A 计算最大特征根及对应特征向量，利用一致性指标、随机一致性指标和一致性比率做一致性检验。检验 CR ，当 $CR < 0.1$ 时通过，特征向量（归一化后）即为权重向量；若不通过，需重新构造判断矩阵。

在这里对判断矩阵和权重向量进行一致性检验的结果为： $CI=0.06$ ， $CR=0.054$ ，通过一致性检验。

第五步：风险评分计算。基于原始数据矩阵 X ，得到风险评分矩阵 $Score$ 如下：

累计病例 新增病例 人口密度 人均GDP

城市1

城市2

...

城市200

$X_{1,1}$

$X_{2,1}$

...

$X_{200,1}$

$X_{1,2}$

$X_{2,2}$

...

$X_{200,2}$

$X_{1,3}$

$X_{2,3}$

...

$X_{200,3}$

$X_{1,4}$

$X_{2,4}$

...

$X_{200,4}$

然后计算 $Score \times W$ ，得到各地区综合风险评分，见附录。

第六步：对风险评分分类

根据所得到的 200 个风险评分，得出箱式分布图及百分位数表。

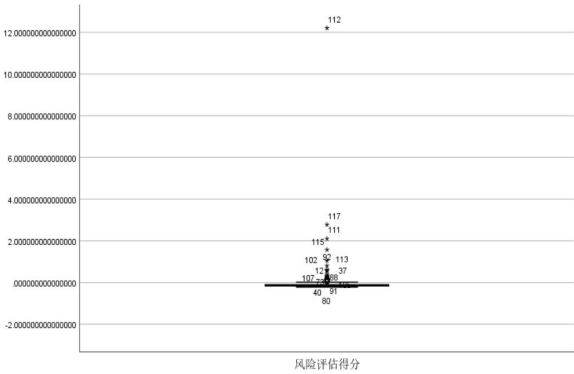


图 5.箱式分布图

从箱式图的分布可以看出，绝大部分城市的疫情风险评估得分较低，极少数风险评估得分较高

		百分位数						
		5	10	25	百分位数 50	75	90	95
加权平均（定义1）	风险评估得分	-.202816549	-.186761540	-.173224688	-.149196227	-.091194064	.1448409168	.3965028086
图基枢纽	风险评估得分			-.173198979	-.149196227	-.091217886		

由百分位数表可得，取风险评估得分小于百分位数 50%即-0.149196227 的城市为疫情情况轻微，大于百分位数 90%即 0.1448409168 的城市为疫情情况严重，在二者之间的为疫情风险等级中等。

下图是选 10 所代表城市的疫情情况展示，其中颜色由绿到红代表疫情风险等级由低到高，所有的 200 个城市疫情情况见附件中。

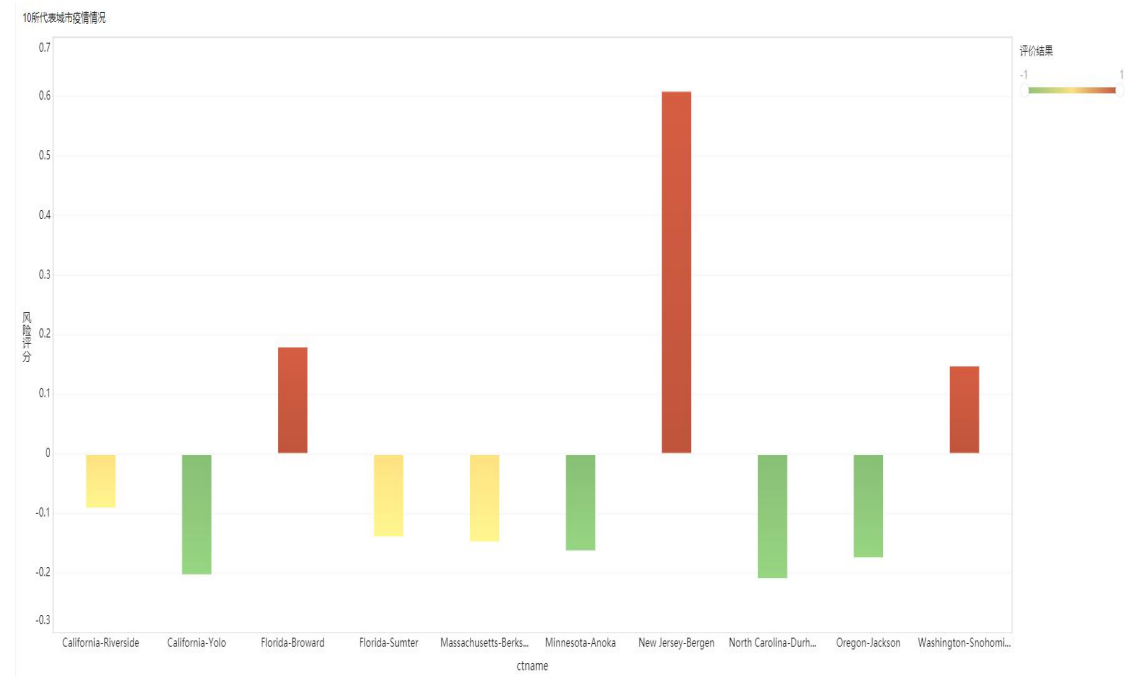


图 6.各城市疫情情况结果

3.2 问题二：疫情发展与城市情况数据相关性研究

3.2.1 问题二分析

截至 2021 年 6 月，全球累计超 1.7 亿人确诊 COVID-19，累计死亡人数高达 370 万，疫情的暴发对人类社会产生了不可估量的影响。由于新冠病毒传播的特性，各国主要推行戴口罩、勤洗手、居家隔离、出行限制、聚集限制等防疫措施，对防控疫情的恶化起到了重要的作用。为了针对疫情发展做好应对措施，对不同城市的城市“属性”与疫情发展情况进行相关性分析，从而根据相关程度较高的城市“属性”来建立评估模型至关重要。

针对本题，我们从经济、政治、文化、社会环境四个层次对城市的“属性”进行了详细地分析，并根据不同层次的城市数据及有关文献给出了分析结果。

3.2.2 数据预处理

首先我们要根据已有数据对疫情发展情况这项指标进行量化，显然累计确诊人数能够较好的表明疫情目前的发展状况，如果考虑过去几天的累计确诊人数便能大致判断出疫情的发展状况即未来走向。

在此我们对数据进行了预处理，采用五项移动平均法的方法构建了一个新的序列，并对新的序列关于时间进行了平均，具体处理步骤如下：

(1) 设每所城市已知数据中最后一天的累计确诊人数为 x_{last} ，则根据五项移动平均法利用后十天的数据可分别得到一组新的序列，生成新序列公式如下：

$$x_{last} = \frac{x_{last} + x_{last-1} + \dots + x_{last-4}}{5}$$

所生成“平均”的累计确诊人数为 $(x'_{last-5}, x'_{last-4}, x'_{last-3}, x'_{last-2}, x'_{last-1}, x'_{last})$ ，为了充分利用已知数据我们采用了以下方法对新生成的序列计算修正后的疫情发展情况指标，即疫情确诊人数新增率 k ：

$$k = \frac{k_1 + k_2 + \dots + k_5}{5}$$

其中 k_1, k_2, \dots, k_5 分别为：

$$k_1 = \frac{x'_{last} - x'_{last-1}}{1} = x'_{last} - x'_{last-1}$$

$$k_2 = \frac{x'_{last-1} - x'_{last-2}}{1} = x'_{last-1} - x'_{last-2}$$

$$\dots$$

$$k_5 = \frac{x'_{last-4} - x'_{last-5}}{1} = x'_{last-4} - x'_{last-5}$$

(2) 从第一问可知采用层次分析法所计算出的疫情风险评估得分主要取决于专家进行打分的评估矩阵，因此具有较强的主观性，也符合我们对疫情风险的直观感受。而在本题中我们所处理后的“修正后的疫情发展情况指标”，在某种程度上也可以认为能够评估疫情的发展情况，因此我们对这两个指标分别进行了统计折线图的绘制，结果如下：

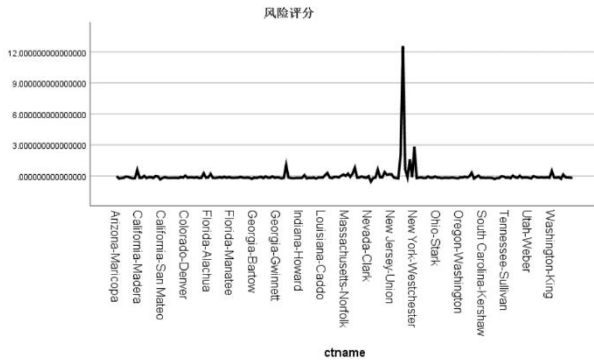


图 7.风险评分

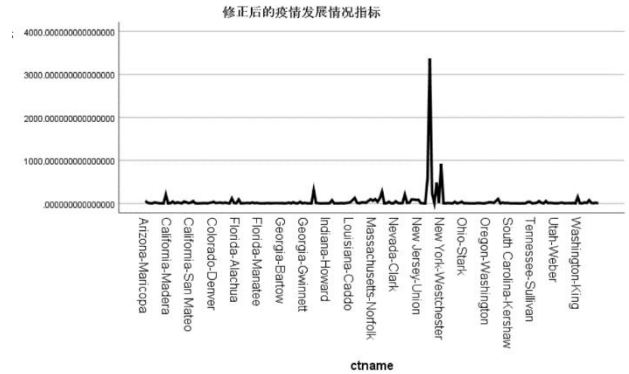


图 8.修正后的疫情发展情况指标

从图中可以看出，不同城市之间相对风险评分和修正后的疫情发展情况指标十分类似，这也与我们先前的主观感觉不谋而合，即疫情发展情况指标与风险评分有着重要的联系，故在本文中我们采用了修正后的疫情发展情况指标进行分析。

3.2.3 经济层面

(1) 人均 GDP

人均 GDP 对不同城市的分布情况显然位置,且也并不满足较好的线性相关关系,因此我们使用了 Spearman 相关分析方法,该方法是通过利用两变量的秩次大小做线性相关分析,并且对原始变量的分布不做要求,适用范围较广。

将各个城市的人均 GDP 和修正后的疫情发展情况指标进行 Spearman 分析结果如下:

相关性				
			pergdp	修正后的疫情发展情况指标
斯皮尔曼 Rho	pergdp	相关系数	1.000	.538**
		Sig. (双尾)	.	.000
		N	199	199
	修正后的疫情发展情况指标	相关系数	.538**	1.000
		Sig. (双尾)	.000	.
		N	199	200

** . 在 0.01 级别 (双尾), 相关性显著。

从上表中可以看出修正后的疫情发展情况指标与 pergdp 的斯皮尔曼相关系数为 0.538, 显著性水平为 0.000 (近似) 通过了显著性检验, 因此可以认为修正后的疫情发展情况指标和人均 GDP 之间的相关性显著, 即人均 GDP 可以作为一项影响疫情发展情况的的城市数据。

(2) 城市医疗水平

权威医学杂志《柳叶刀》在 2018 年发布了《2016 年全球疾病、伤害和风险因素负担研究报告》(Global Burden of Diseases, Injuries, and Risk Factors Study 2016, 下称 GBD), 对 195 个国家和地区在 1990 到 2016 年间个人医疗服务的获取情况和质量进行了评价。

这份报告将在有效护理情况下不应发生死亡的 32 种疾病的死亡率作为评估指标, 依此计算出世界各国的医疗可及性和质量指数 (Healthcare Access and Quality Index, 以下简称 HAQ 指数)。各国指数分布在 0 到 100 之间, 分值越高, 说明该国个人获得的医疗服务机会和质量越高。

通过查阅数据, 全球较高及较低的医疗水平的部分排名如下, 该数据来自于报告《全球 195 个国家和地区的医疗质量与可及性排名》[3]。

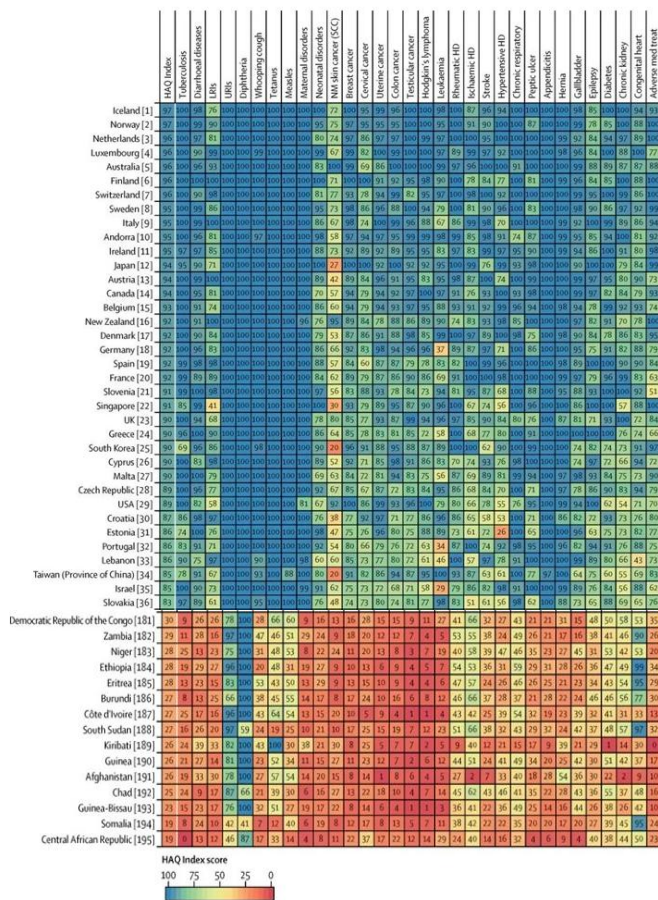


图 9.全球部分国家医疗水平示意图

3. 2. 4 政治层面

(1) 社会体制

在疫情面前，社会体制对疫情发展的影响非常显著，中国特色社会主义制度所具有的显著优势，是抵御风险挑战、提高国家治理效能的根本保证。疫情发生后，全国上下紧急行动，依托强大综合国力，开展全方位的人力组织战、物资保障战、科技突击战、资源运动战，全力支援湖北省和武汉市抗击疫情，在最短时间集中最大力量阻断疫情传播。“中方行动速度之快、规模之大，世所罕见，展现出中国速度、中国规模、中国效率”

开展新中国成立以来规模最大的医疗支援行动。调动全国医疗资源和力量，全力支持湖北省和武汉市医疗救治。自 1 月 24 日除夕至 3 月 8 日，全国共调集 346 支国家医疗队、4.26 万名医务人员、900 多名公共卫生人员驰援湖北。19 个省份以对口支援、以省包市的方式支援湖北省除武汉市以外 16 个地市，各省在发生疫情、防控救治任务十分繁重的情况下，集中优质医疗资源支援湖北省和武汉市。人民解放军派出 4000 多名医务人员支援湖北，承担火神山医院等 3 家医疗机构的医疗救治任务，空军出动运输机紧急运送医疗物资。各医疗队从接受

指令到组建 2 小时内完成，24 小时内抵达，并自带 7 天防护物资，抵达后迅速开展救治。在全国紧急调配全自动测温仪、负压救护车、呼吸机、心电监护仪等重点医疗物资支援湖北省和武汉市（表 1）。从全国调集 4 万名建设者和几千台机械设备，仅用 10 天建成有 1000 张病床的火神山医院，仅用 12 天建成有 1600 张病床的雷神山医院。短短 10 多天建成 16 座方舱医院，共有 1.4 万余张床位。加强临床血液供应，10 个省份无偿支援湖北省红细胞 4.5 万单位，血小板 1762 个治疗量，新鲜冰冻血浆 137 万毫升（不含恢复期血浆）。大规模、强有力的医疗支援行动，有力保障了湖北省和武汉市救治，极大缓解了重灾区医疗资源严重不足的压力。相关数据如下表所示[4]。

图 10.疫情发生以来调往湖北省医疗物资情况

序号	类别	品种	单位	数量
1	医疗设备	全自动测温仪	台	20033
2		负压救护车	辆	1065
3		呼吸机	台	17655
4		心电监护仪	台	15746
5	消杀用品	84 消毒液	吨	1874
6	防护用品	免洗手液	万瓶	71.4
7		医用手套	万副	198.7
8		防护服	万套	773
9		医用 N95 口罩	万只	498
10	防控药品	医用非 N95 口罩	万只	2720
11		磷酸氯喹	万片/粒	40
12		阿比多尔	万片/粒	360

图 11.重点医疗物资生产情况

	类别	指标		日产量较疫情 初期(1月底) 增长倍数	
		产品	日产能		
1	防护用品	医用防护服(万套)	189	80	90.6 倍
2	消杀用品	免洗手消毒液(吨)	409	308	2.6 倍
		84 消毒液(万箱)	36.6	11.7	1.6 倍
3	医疗设备	全自动红外测温仪(万台)	1.07	0.34	23.3 倍
4	检测用品	病毒检测试剂(万人份)	1020	760	58 倍

与之形成鲜明对比的是美国，但此次美国抗疫不力，新冠肺炎确诊人数和死亡人数暴增，疫情尚未得到有效防控，仍在持续蔓延。美国是一个总统共和制的联邦制国家，实行三权分立相互制衡、两党制的政党制度。美国的两党争霸现象自 19 世纪延续至今，党争已经激化到接近引发宪政危机的程度，这不仅在严重撕裂美国政治与社会，还深刻影响到国家治理的效能。美国的 50 个州全部因新冠肺炎疫情进入“重大灾难”状态，新冠病毒正以不同的方式影响着每一个美国人，而美国却无法整合所有公共卫生资源并尽可能地动员全体社会资源进行有效应对。由此可见不同的意识形态对疫情的发展与控制具有十分重要的影响。

(2) 政府政策

政府政策的制定对疫情的发展与控制起着至关重要的作用，以武汉的疫情防控为例，在疫情开始蔓延的第一时间国家就紧急制定了相关政策，采取阻断病毒传播的关键一招，坚决果断关闭离汉离鄂通道，武汉保卫战、湖北保卫战全面打响。[4]左图展示了疫情的不同时间段武汉市政府采取的相关政策。中共中央成立应对疫情工作领导小组，并向湖北等疫情严重地区派出中央指导组。国务院先后建立联防联控机制、复

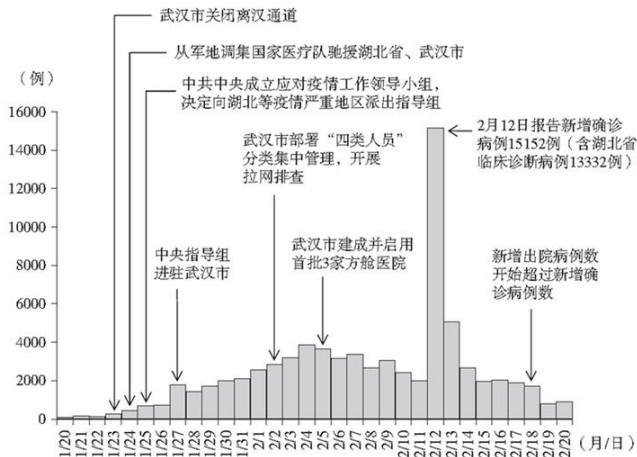


图 12.疫情不同时间段武汉采取的有关政策

工复产推进工作机制。全国集中资源和力量驰援湖北省和武汉市。各地启动重大突发公共卫生事件应急响应。最全面最严格最彻底的全国疫情防控正式展开，疫情蔓延势头初步遏制。

而反观美国，没有采取严格的疫情防控政策，既没有限制人员流通，也没有强制要求戴口罩，因此疫情一直没法得到有效控制。由此可见，政府所制定得有关政策是影响疫情发展与控制的关键因素。

3.2.5 文化层面

(1) 受教育水平

受教育水平我们采用高中及以上学历的占比来衡量，占比越高说明受教育水平高，反之则受教育水平低，下图为不同收入水平国家的受教育水平占比[5]。

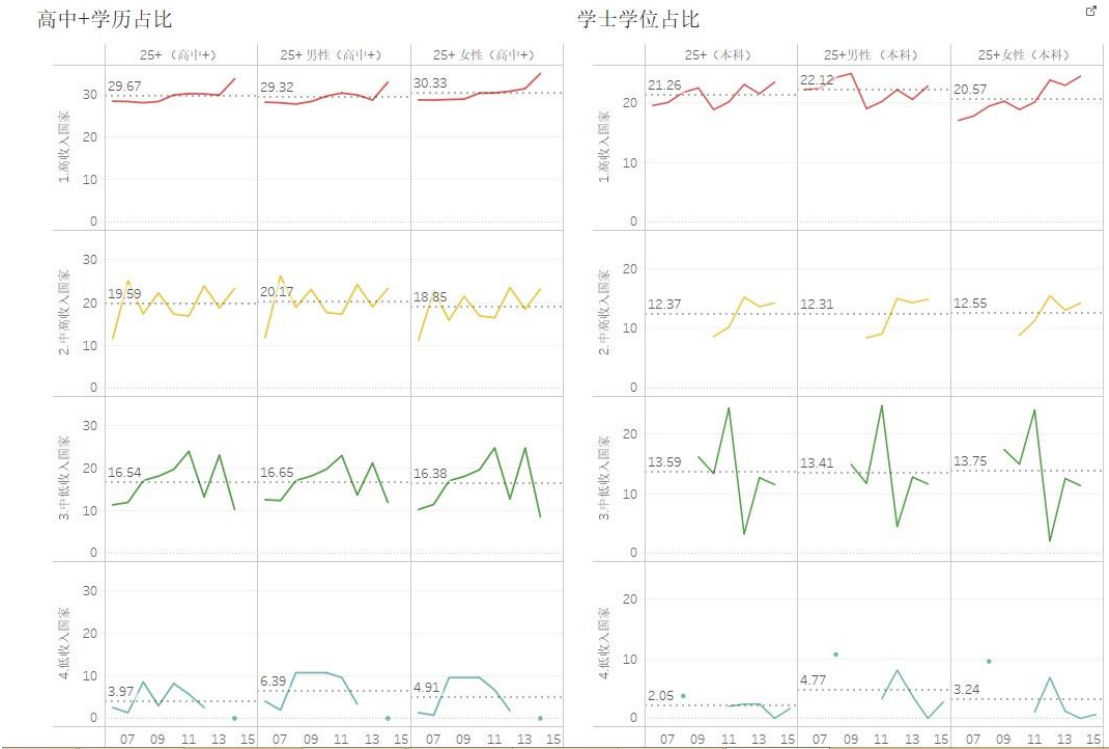


图 13.不同收入国家的受教育水平程度

与世界疫情发展情况相对比可见相对来说受教育水平较高的地区疫情控制得较好。受教育水平高，则意味着能够掌握更多医疗卫生方面的知识，能够更加频繁的接触到有关疫情的相关信息，同时也能对政府提出的抗疫政策做出更加积极的响应。所以受教育水平与疫情的发展情况有着一定的相关性。

3.2.6 社会环境层面

(1) 月份平均温度

查阅有关文献[6]和资料我们发现，虽然新冠病毒某些特性以及人们免疫系统和行为方式等可能会随季节和气温变化，从而对疫情发展产生一定影响，但以来升温或者降温来控制疫情的观点缺乏科研证据支持。

从数据分析的角度，我们同样对月份平均温度和修正后的疫情发展情况指标

进行了相关性分析，得出结果如下图：

相关性

		修正后的疫情 发展情况指标	Meantem202 0
斯皮尔曼 Rho	修正后的疫情发展情况指 标	相关系数	1.000
		Sig. (双尾)	.996
		N	200
	Meantem2020	相关系数	.000
		Sig. (双尾)	.996
		N	200

因此从上表中可以看出，月份平均温度与疫情未来的发展情况相关性并不强，因此不能作为预测或评判疫情发展情况的一个主要指标。

(2) 城市所处纬度

城市所处纬度即地理位置，与该城市的气候有着较强的相关性，但相较于疫情发展情况，由上述月份平均温度的讨论可知，并不能对疫情产生着较大的直接影响。下图为采用 Spearman 相关性检验对其进行的分析结果：

相关性

		修正后的疫情 发展情况指标	Latitude
斯皮尔曼 Rho	修正后的疫情发展情况指 标	相关系数	1.000
		Sig. (双尾)	.967
		N	200
	Latitude	相关系数	-.003
		Sig. (双尾)	.967
		N	200

从上表中我们可以看出，纬度和修正后的疫情发展情况指标并没有显著的相关性，故不能将其作为评判疫情发展情况的一个主要指标。

(3) 人口密度

截止至目前，世界上许多最致命、最触目惊心的疫情都爆发在城市，尤其是人口稠密的一些地区，这些城市也成为了新冠病毒最易传播、疫情最严峻的地区。这是因为新冠本身就属于一种流行病，想要遏制其传染就需要保持社交距离，但城市的聚集性要求人们群体性的聚集并共享空间，因此也就造成了疫情的扩散。

下图为利用数据对各个城市的人口密度和修正后的疫情发展情况指标进行 Spearman 相关性检验的结果：

相关性

		修正后的疫情 发展情况指标	density
斯皮尔曼 Rho	修正后的疫情发展情况指标	相关系数	1.000
		Sig. (双尾)	.000
		N	200
	density	相关系数	.724**
		Sig. (双尾)	.000
		N	200

** 在 0.01 级别（双尾），相关性显著。

从图中我们可以看出，修正后的疫情发展情况指标与人口密度的斯皮尔曼相关系数为 0.724，显著性水平为 0.000（近似），通过了显著性检验，因此可以认为人口密度与疫情的发展情况具有显著的相关性，即可以将其视为疫情发展的主要指标。

（4）老人百分比

人口老龄化是全球很多国家都面临的一个问题，老年人越来越多，而由于身体免疫力较低、获取疫情信息的能力不强等原因，老年人也成为了最容易感染新冠肺炎病毒的人群。同样，我们将各个城市的老人百分比即老年人占城市总人口的百分比与修正后的疫情发展情况指标进行了相关性分析，所得结果如下图所示：

相关性

		修正后的疫情 发展情况指标	prop.elder
斯皮尔曼 Rho	修正后的疫情发展情况指标	相关系数	1.000
		Sig. (双尾)	.000
		N	200
	prop.elder	相关系数	-.271**
		Sig. (双尾)	.000
		N	200

** 在 0.01 级别（双尾），相关性显著。

从上表中我们可以看出，老人百分比与修正后的疫情发展情况有着一定的相关性，但与人口密度等因素相比影响不大，因此可以认为老人百分比与疫情发展情况有着一定的关系，但将其作为预测评判疫情未来发展情况的主要指标所起效果并不明显。

3.3 问题三：温州疫情预测模型

3.3.1 问题三分析

去年武汉的疫情爆发牵动着全国上下人们的心弦，但与此同时千里之外的温州却成为了湖北以外疫情最严重的城市，其疫情严重情况不但远超北京、上海，甚至超过了紧邻湖北、人口众多的重庆。因此对于温州疫情情况的研究，构建温州疫情预测模型对我国甚至全球打赢疫情防控阻击战有着重要意义。

从本文第二问的分析与解答中我们知道了疫情的发展情况不仅仅与该城市的人口密度、人均 GDP、医疗水平有着联系，更与该城市所指定的防控疫情政策有着重要关系，甚至是能够决定疫情未来的走向。因此针对本题，我们不仅仅利用了时间序列 ARIMA 模型进行分析，也采用了 SEIR 传染病动力学模型对温州市的新冠疫情情况进行预测和评估，并将二者预测结果与温州市的实际疫情情况作对比，分析出各自模型的优缺点，给出了疫情防控的一些有效控制建议。

3.3.2 时间序列 ARIMA 模型的求解与分析

(1) ARIMA 模型的建立[7]

时间序列的模型建立与普通回归问题不同之处在于时间序列和时间有关，且随着上升或下降的趋势，出现了季节性趋势的形式，比如一些季节性传染病的情况就需要利用这类模型进行求解。ARIMA 是一种根据时间序列集合预测未来发展情况或者进行表现分析的模型，以下为 ARIMA 模型的建模步骤：

① 时间序列预处理

计算序列的自相关系数和偏相关系数，评估序列的平稳性和纯随机性，如果不满足条件，再通过差分法来把原来的累计确诊人数序列平稳化，后续数据处理的正确性。以下为我们对该累计确诊人数序列处理后的可视化结果：

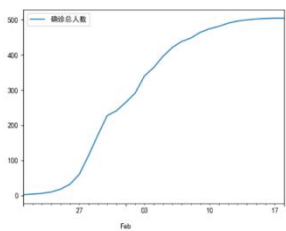


图 14. 未经过差分的时序图

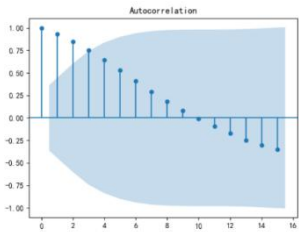


图 15 未经过差分的自相关图

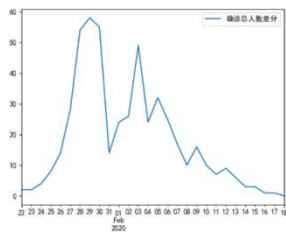


图 16 经过差分的时序图

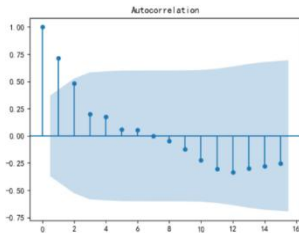


图 17. 经过差分的自相关图

从以上可视化结果中可以看出，数据经过差分处理后，能够满足平稳性检测和白噪声监测即纯随机性检测。

②模型识别

模型识别即从已知的模型中选择一个与给出的时间序列过程相吻合的模型。模型识别的方法很多，例如 Box-Jenkins 模型识别方法等。

③模型定阶

在确定了模型的类型之后，还需要知道模型的阶数，可使用 BIC 准则法进行定阶，在本文中我们便采用了 BIC 准则法对该疫情预测模型进行定阶。

④参数估计

对模型的参数进行估计的方法通常有相关矩估计法、最小二乘估计以及极大似然估计等。

(2) 模型结果的分析与讨论

得到 ARIMA 时间序列模型之后，我们将其预测的 20 天疫情情况，即累计确诊人数预测结果列出了如下表格：

2.19	2.20	2.21	2.22	2.23	2.24	2.25	2.26	2.27	2.28
516	533	551	568	586	603	621	638	656	673
2.29	2.30	2.31	3.1	3.2	3.3	3.4	3.5	3.6	3.7
691	708	726	744	761	779	796	814	831	849

其中第一行第三行为预测日期，第二行第四行为 ARIMA 预测的累计确诊人数。再将其预测的结果绘制成折线统计图，与温州的实际疫情情况进行对比。

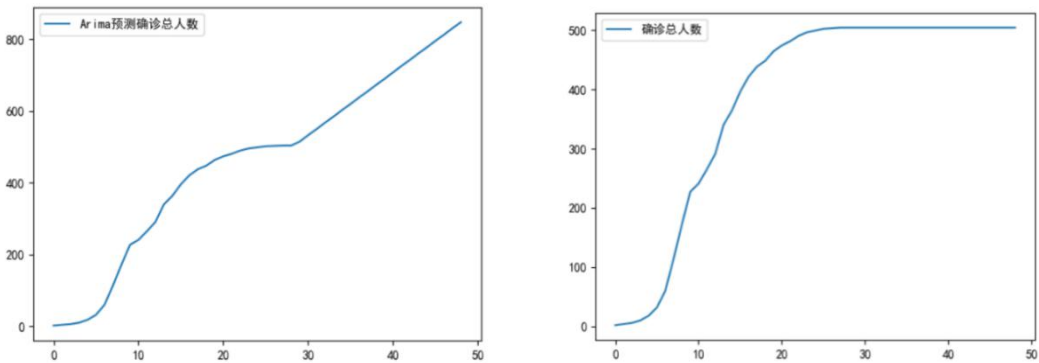


图 18.ARIMA 模型预测结果对比图

从上图可以看出，Arima 预测模型的短期预测结果相对准确，但长期预测结果与温州实际疫情累计确诊人数并不符合。这是由于 Arima 模型只考虑了内生变量，没有考虑其他外生变量的影响，比如政府在某些时间点实施相应的防控政策会导致疫情发展情况出现转折点等情况。而且从 Arima 模型的本质上讲其只能捕捉线性关系，而疫情情况随时间的变化趋势并不简单的只是线性关系，因此通过 Arima 模型的预测结果在短期内可以为疫情防控措施的制定提供一些思路，但长期的预测还需要有具体的实际数据来支撑。

3.3.3 SEIR 传染病动力学模型的求解与分析

SEIR 模型，就是把一个封闭区域中的人群分为四类：S（susceptible/易感者）、E（exposed/潜伏者）、I（infectious/感染者）、R（recovered/康复者），通过各环节的转化率、感染者的治愈率等对传染病的传播时间和传播规模进行预测。其中 SEIR 模型的传染病动力学微分方程为：

$$\begin{aligned} S_i &= S_{i-1} - \frac{\gamma\beta_1 I_{i-1} S_{i-1}}{N} - \frac{\gamma_1\beta_2 E_{i-1} S_{i-1}}{N} \\ E_i &= E_{i-1} + \frac{\gamma_1\beta_1 I_{i-1} S_{i-1}}{N} - \alpha E_{i-1} + \frac{\gamma_1\beta_2 E_{i-1} S_{i-1}}{N} \\ I &= I_{i-1} + \alpha E_{i-1} - \gamma I_{i-1} \\ R_i &= R_{i-1} + \gamma I_{i-1} \end{aligned}$$

上述方程中的 S 是易感染者例数，由于新型冠状病毒所有人群普遍易感染，所以这里 S 就是总例数； $\alpha \approx 0.3$ 是潜伏者转化为感染者的概率，E 是潜伏者例数，I 是感染者例数，R 表示移除例数， β_1 、 β_2 分别表示感染者和潜伏者传染给易感人群的概率，依据国家卫健委公布的数据计算得到 $\beta_1=0.053$ ， $\beta_2 \approx 0.1$ 、 r_1 、 r_2 是感染者和潜伏者接触易感染者例数，在没有管控之前 $r=r_2=12$ ，由于政府实施较为严格的管控，进行居家隔离，因此这里 $r=r_2=3$ ，即最多只能接触 3 例； γ 表示康复的概率为 0.6，这里不是严格按照恢复率来计算的，而是综合考量了不同人群的特性以及人口密度来计算的。

（1）SEIR 模型的建立[8]

我们通过查询国家卫健委官方网站和湖北卫健委官方网站公布的温州市新冠肺炎感染例数等数据对感染者接触易感者的人数、传染概率、潜伏者接触易感者的人数等参数进行了计算，得到了以下参数数据表格：

感染者接触易感者的人数 r	感染概率 B	潜伏者接触易感者的人数 r_2	潜伏者传染正常人的概率 B_2	潜伏者转化为感染者概率 a	康复概率 y
12	0.053	12	0.1	0.3	0.6

因此最终 SEIR 模型各指标的递推表达式为

$$S[i+1] = S[i] - r * B * \frac{S[i]}{N} - r_2 * B_2 * S[i] * \frac{E[i]}{N}$$

$$E[i+1] = E[i] + r * B * S[i] * \frac{I[i]}{N} - a * E[i] + r_2 * B_2 * S[i] * \frac{E[i]}{N}$$

$$I[i+1] = I[i] + a * E[i]$$

$$R[i+1] = R[i] + y * I[i]$$

(2) 模型结果的分析与讨论

得到 SEIR 传染病动力学模型之后，我们将其预测的 20 天疫情情况，即累计确诊人数预测结果列出了如下表格：

2.19	2.20	2.21	2.22	2.23	2.24	2.25	2.26	2.27	2.28
492	493	495	496	498	499	501	502	504	505
2.29	2.30	2.31	3.1	3.2	3.3	3.4	3.5	3.6	3.7
507	509	510	512	513	515	515	515	515	515

其中第一行第三行为预测日期，第二行第四行为 ARIMA 预测的累计确诊人数。再讲其预测的结果绘制成折线统计图，与温州的实际疫情情况进行对比。

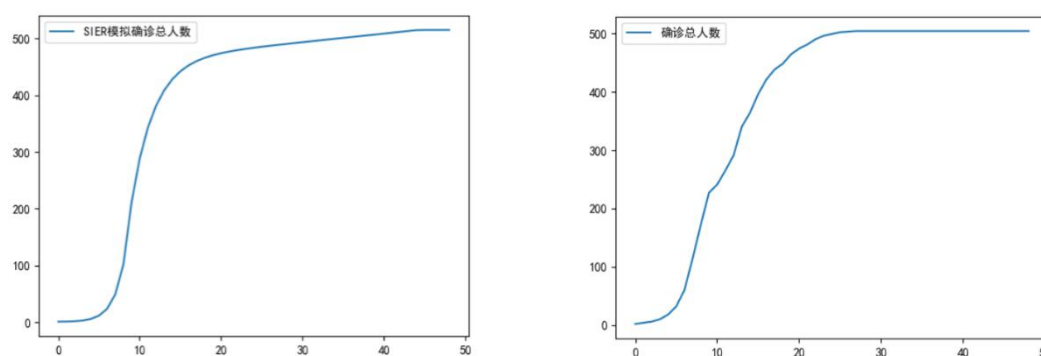


图 19. SEIR 模型预测结果对比图

从上图中可以看出 SEIR 模型的预测结果与温州实际总确诊人数十分相似，这是由于我们所采用的 SEIR 模型并不是简单的考虑了传统的传染病动力学因素，也考虑了一些国家强力干预和春节后的回程高峰消除等因素对温州疫情发展情况的影响，因此其预测的疫情发展情况考虑较为全面，预测效果较好。

4. 模型评价和改进

4.1 新冠疫情等级评估模型

在该模型中，我们采用了单指标风险评价法，系统聚类法以及层次分析法。以上方法的分类方法都具有一定的合理性，但是这些方法有各自的优缺点。

对于单指标风险评价法，优势是计算简单、容易操作但此方法主观性太强且没有考虑社会经济因素的影响，没有综合考虑影响疫情发展情况的各个因素。系统聚类方法将风险向量视为4维空间的一个点，然后基于各个地区特征的相似性将距离临近的点聚为一类，方法简单，易于操作，但该方法同等对待各个指标的重要性，结果的可解释性较弱。层次分析法综合多个指标的信息，可以对各个地区的风险进行量化评分，因此也可以对各个地区风险进行排序，便于领导层决策管理；其缺点是其判断矩阵较难获得，需要查询相关文献或请有关专家进行评判。

综上所述，针对问题一，层次分析法效果良好，能够较好的满足各城市地区疫情情况分类的要求。

4.2 Spearman 相关性检验

常用的 Pearson 相关检验适用于两变量均符合正态分布时的情况，因此在第二问中的疫情发展情况与城市数据指标的相关性探究中该方法并不适用。我们采用了另外一种相关性分析方法 Spearman 检验。该方法能够利用单调方程评价两个统计变量的相关性，且适用范围较广，能够较好的满足探城市疫情发展情况与已有数据的相关性检验。针对其他无法用数据衡量的城市指标，我们通过查询相关文献和资料也对其进行了一一分析和整理，并给出了相关性结论，较好的完成了任务二的要求。

4.3 温州疫情预测模型

4.3.1 ARIMA 时间序列分析模型

ARIMA 时间序列分析模型建立过程较为容易，只需要内生变量而不需要借助其他外生变量，在此处只需要借助原有的总确诊人数就可以进行时间序列分析预测但是要求时序数据是稳定的 (stationary)，或者是通过差分化(differencing)后是稳定的，并且本质上只能捕捉线性关系，而不能捕捉非线性关系，因为他只会捕捉到动态的信息，对于其它的一些客观信息譬如 GDP，人口密度，政策等信息是无法捕捉到的，因此会比较局限。

4.3.2 SEIR 模型

整体而言是非常适合用来模拟传染病类型的预测和分析，此处是新冠的预测和分析，并且综合考量了城市的政策情况，感染者接触情况，传染情况，疾病的致死率发病率，恢复率等等因素，根据之前的疫情数据，进行数值模拟来确定一系列相关的参数。

但 SEIR 只是传染病传播的基本模型，还有很多没有考虑的因素，如人口的出生与死亡，迁入和迁出等。还有可以考虑的更细致的因素，如人群流动速度、易感人群的年龄分布、不同人群对疾病的易感性，患病者的症状轻重，人口密度，

医疗卫生程度，检验检疫手段，政府重视程度，人群心理因素等。这些因素都对暴露数、发病率、治愈率、传染期长度有着直接或间接的影响。但是 SEIR 整体而言算是一个比较优秀的模型了

SEIR 和 ARIMA 比较而言，这里更适合用 SEIR，因为 ARIMA 时间序列分析在这里很难只根据总确诊人数提取出有效信息，譬如发病率，致死率，人口密度，总人口数，经济状况，病人接触情况，因此很难预测出准确和有效的模型，但是 SEIR 模型这里的一系列参数，即潜伏期，传染率，病人接触情况，恢复率等可以根据官方的数据进行数值模拟，进行很好的估算，同时如果有政策改动，譬如临时封城等，SEIR 模型里面也会有较为直观的体现。因此这里更适合于用 SEIR 传染病动力学模型进行预测和评估。

4.4 模型扩展方向

（1）疫情各个城市的分类模型在本文中主要利用了四项城市指标，但也可以通过进一步考虑城市政府的政策影响力等来优化改分类评估模型。

（2）对于 ARIMA 模型可以考虑将一系列环境影响因素标量化，譬如政策改进 0-10 为封管程度，病人接触情况，传播情况等等也可以标量化，这样结果就不会那么单一仅仅根据总确诊人数来确定预测疫情的未来发展趋势；对于 SIER 模型可以考虑构建再多加一些权重因素，譬如人群流动速度、易感人群的年龄分布、不同人群对疾病的易感性，患病者的症状轻重，等等，如果再细致一点将这些特征引入，SIER 模型应该会表现更好。

5. 参考文献

- [1] 中华人民共和国中央人民政府, 图表:国务院联防联控机制印发《关于科学防治精准施策分区分级做好新冠肺炎疫情防控工作的指导意见》
http://www.gov.cn/xinwen/2020-02/18/content_5480834.htm, 2021. 6. 4
- [2] 李辉智, 易大莉, 李高明. 重庆市新型冠状病毒肺炎疫情风险分区分级评估方法比较研究[J]. 重庆医科大学学报, 2020, 45(07):870-875.
- [3] GBD 2016 Healthcare Access and Quality Collaborators. Measuring performance on the Healthcare Access and Quality Index for 195 countries and territories and selected subnational locations: asystematic analysis from the Global Burden of Disease Study 2016. Lancet. 2018 Jun 2.
- [4] 中华人民共和国中央人民政府, 抗击新冠肺炎疫情的中国行动
http://www.gov.cn/zhengce/2020-06/07/content_5517737.htm, 2021. 6. 5
- [5] 特布鲁, 世界主要国家的教育水平,
<https://zhuanlan.zhihu.com/p/28398447>, 2021. 6. 5
- [6] 张斌, 气温升高了, 新冠病毒会被遏制吗? 看看专家怎么说 | 特别关注,
https://www.sohu.com/a/391877717_100012122, 2021. 6. 5
- [7] 杨真真, 谢艳秋, 靳旭东, 庄桂敏. 基于 ARIMA 时间序列模型的传染病发展趋势预测——以 COVID-19 为例[J]. 中国科技信息, 2021(Z1):70-72.
- [8] 曹盛力, 冯沛华, 时朋朋. 修正 SEIR 传染病动力学模型应用于湖北省 2019 冠状病毒病 (COVID-19) 疫情预测和评估[J]. 浙江大学学报(医学版), 2020, 49(02):178-184.

6. 附录

6.1 新冠疫情评分及等级（层次分析法）

地区	风险评分	评价结果
Arizona-Maricopa	-0.0142558	中等
Arizona-Pima	-0.242526644	轻微
Arizona-Pinal	-0.198235698	轻微
Arkansas-Jefferson	-0.187016326	轻微
California-Alameda	-0.053271894	中等
California-Contra Costa	-0.074058689	中等
California-Fresno	-0.172199376	轻微
California-Humboldt	-0.224110991	轻微
California-Imperial	-0.208455982	轻微
California-Los Angeles	0.589080144	严重
California-Madera	-0.185375844	轻微
California-Marin	-0.184300713	轻微
California-Orange	0.019833442	中等
California-Placer	-0.202237328	轻微
California-Riverside	-0.118324924	中等
California-Sacramento	-0.096701161	中等
California-San Benito	-0.198275999	轻微
California-San Diego	0.001413711	中等
California-San Francisco	-0.032812628	中等
California-San Luis Obispo	-0.329914664	轻微
California-San Mateo	-0.190295703	轻微
California-Santa Clara	-0.102864132	中等
California-Santa Cruz	-0.176672284	轻微
California-Shasta	-0.186657913	轻微
California-Solano	-0.180294378	轻微
California-Sonoma	-0.182148882	轻微
California-Ventura	-0.168281614	轻微
California-Yolo	-0.202915047	轻微
Chongqing	-0.088820436	中等
Colorado-Arapahoe	-0.155201107	轻微
Colorado-Denver	0.035721556	中等
Colorado-Douglas	-0.168777085	轻微
Colorado-Eagle	-0.132091129	中等
Colorado-El Paso	-0.126990576	中等
Colorado-Gunnison	-0.172023125	轻微
Colorado-Jefferson	-0.107374475	中等
Colorado-Larimer	-0.204081105	轻微
Colorado-Summit	-0.183564495	轻微
Connecticut-Fairfield	0.264577734	严重
Connecticut-Litchfield	-0.16367813	轻微
Florida-Alachua	-0.137251499	中等
Florida-Broward	0.214437608	严重
Florida-Charlotte	-0.177297009	轻微
Florida-Clay	-0.193222341	轻微
Florida-Collier	-0.161437908	轻微
Florida-Duval	-0.099053969	中等
Florida-Escambia	-0.177725958	轻微
Florida-Hillsborough	-0.061904872	中等
Florida-Lake	-0.17400545	轻微
Florida-Lee	-0.098530218	中等
Florida-Manatee	-0.165997492	轻微
Florida-Nassau	-0.18312866	轻微
Florida-Okaloosa	-0.158866673	轻微
Florida-Pasco	-0.163845456	轻微
Florida-Pinellas	-0.084722761	中等
Florida-Santa Rosa	-0.15702843	轻微
Florida-Sarasota	-0.176454891	轻微
Florida-Seminole	-0.142476874	中等
Florida-Sumter	-0.163940763	轻微
Florida-Volusia	-0.280327849	轻微
Georgia-Bartow	-0.138234552	中等
Georgia-Charlton	-0.193828147	轻微
Georgia-Cherokee	-0.12875632	中等
Georgia-Cobb	-0.066605724	中等
Georgia-Coweta	-0.207588177	轻微
Georgia-DeKalb	-0.039924827	中等
Georgia-Fayette	-0.158605726	轻微
Georgia-Floyd	-0.172744381	轻微
Georgia-Fulton	-0.034184061	中等
Georgia-Gordon	-0.172801167	轻微
Georgia-Gwinnett	-0.122873759	中等
Georgia-Lee	-0.156078353	轻微
Georgia-Lowndes	-0.229189211	轻微
Georgia-Polk	-0.174997543	轻微
Illinois-Cook	1.061550792	严重
Illinois-Kane	-0.143449204	中等
Illinois-McHenry	-0.195903811	轻微
Indiana-Adams	-0.213163632	轻微
Indiana-Boone	-0.196491075	轻微

Indiana-Hendricks	-0.173552578	轻微
Indiana-Howard	-0.199424606	轻微
Indiana-Johnson	-0.174659542	轻微
Indiana-Marion	0.089374799	中等
Indiana-Noble	-0.232680152	轻微
Iowa-Johnson	-0.168634662	轻微
Iowa-Pottawattamie	-0.198801992	轻微
Kansas-Johnson	-0.142423523	中等
Kentucky-Harrison	-0.250209287	轻微
Kentucky-Jefferson	-0.147742118	中等
Louisiana-Ascension	-0.139622708	中等
Louisiana-Caddo	-0.151470814	轻微
Louisiana-Jefferson	0.08012936	中等
Louisiana-Orleans	0.296409609	严重
Maryland-Baltimore-city	-0.142843275	中等
Maryland-Harford	-0.206794931	轻微
Maryland-Montgomery	-0.08062983	中等
Maryland-Prince George's	-0.084873629	中等
Massachusetts-Berkshire	-0.140821236	中等
Massachusetts-Essex	0.022049946	中等
Massachusetts-Middlesex	0.147870593	严重
Massachusetts-Norfolk	0.005793176	中等
Massachusetts-Suffolk	0.229027305	严重
Massachusetts-Worcester	-0.077777582	中等
Michigan-Oakland	0.219563732	严重
Michigan-Wayne	0.802633121	严重
Minnesota-Anoka	-0.167683464	轻微
Minnesota-Ramsey	-0.139422861	中等
Missouri-St. Louis	-0.043160994	中等
Nebraska-Douglas	-0.135027827	中等
Nebraska-Knox	-0.171763865	轻微
Nevada-Clark	-0.02094263	中等
Nevada-Washoe	-0.536624411	轻微
New Hampshire-Grafton	-0.195029579	轻微
New Hampshire-Rockingham	-0.150213048	轻微
New Jersey-Bergen	0.627994247	严重
New Jersey-Burlington	-0.120473632	中等
New Jersey-Camden	-0.128880159	中等
New Jersey-Hudson	0.385073908	严重
New Jersey-Monmouth	0.116987622	中等
New Jersey-Passaic	0.176417835	严重
New Jersey-Union	0.173170129	严重
New Mexico-Bernalillo	-0.141890437	中等
New Mexico-Santa Fe	-0.200388415	轻微
New Mexico-Socorro	-0.223064629	轻微
New York-Nassau	2.124899509	严重
New York-New York	12.55583512	严重
New York-Rockland	0.62139176	严重
New York-Saratoga	-0.131892424	中等
New York-Suffolk	1.594300932	严重
New York-Ulster	-0.109712581	中等
New York-Westchester	2.815403565	严重
North Carolina-Cabarrus	-0.202992673	轻微
North Carolina-Chatham	-0.156871712	轻微
North Carolina-Durham	-0.129743784	中等
North Carolina-Forsyth	-0.202449223	轻微
North Carolina-Johnston	-0.187949268	轻微
North Carolina-Mecklenburg	-0.046087027	中等
North Carolina-Onslow	-0.172563785	轻微
North Carolina-Wake	-0.165071568	轻微
Ohio-Cuyahoga	-0.050064294	中等
Ohio-Stark	-0.152020338	轻微
Oklahoma-Tulsa	-0.156603327	轻微
Oregon-Deschutes	-0.209275296	轻微
Oregon-Douglas	-0.169297699	轻微
Oregon-Jackson	-0.172973579	轻微
Oregon-Klamath	-0.186916672	轻微
Oregon-Marion	-0.155849674	轻微
Oregon-Multnomah	-0.156571305	轻微
Oregon-Polk	-0.193399881	轻微
Oregon-Umatilla	-0.224864519	轻微
Oregon-Washington	-0.124924508	中等
Pennsylvania-Bucks	-0.140966516	中等
Pennsylvania-Delaware	-0.059057094	中等
Pennsylvania-Monroe	-0.140194339	中等
Pennsylvania-Montgomery	-0.041966158	中等
Pennsylvania-Philadelphia	0.315962241	严重
Pennsylvania-Wayne	-0.243144995	轻微
Rhode Island-Providence	-0.092142748	中等
Shanghai	0.042051915	中等
South Carolina-Charleston	-0.182203903	轻微
South Carolina-Kershaw	-0.154257568	轻微
South Carolina-Lancaster	-0.192162587	轻微
South Carolina-Spartanburg	-0.167611942	轻微
South Dakota-Beadle	-0.175556675	轻微
South Dakota-Charles Mix	-0.188961852	轻微
South Dakota-Davison	-0.294327065	轻微
South Dakota-Minnehaha	-0.19266919	轻微
South Dakota-Pennington	-0.206573693	轻微
Tennessee-Davidson	-0.022705442	中等
Tennessee-Shelby	-0.054575196	中等
Tennessee-Sullivan	-0.167528089	轻微
Tennessee-Williamson	-0.133471353	中等
Texas-Collin	-0.215771713	轻微
Texas-Dallas	0.036129332	中等
Texas-Fort Bend	-0.121544414	中等
Texas-Gregg	-0.180550439	轻微
Texas-Harris	0.026912532	中等
Texas-Montgomery	-0.17521687	轻微
Texas-Tarrant	-0.13494275	中等
Utah-Davis	-0.134416677	中等
Utah-Weber	-0.170603617	轻微
Vermont-Bennington	-0.22745104	轻微
Virginia-Arlington	-0.006958029	中等
Virginia-Fairfax	-0.102710514	中等
Virginia-Henrico	-0.15154166	轻微
Virginia-Loudoun	-0.129791603	中等
Washington-Clark	-0.134803817	中等
Washington-Grant	-0.157546541	轻微
Washington-Island	-0.121856456	中等
Washington-Jefferson	-0.163462328	轻微
Washington-King	0.500824587	严重
Washington-Kitsap	-0.143110597	中等
Washington-Kittitas	-0.1644894	轻微
Washington-Pierce	-0.104037217	中等
Washington-Skagit	-0.268020029	轻微
Washington-Snohomish	0.149397799	严重
Washington-Whatcom	-0.131394433	中等
Wenzhou	-0.140217937	中等
Wisconsin-Dane	-0.156544213	轻微
Wisconsin-Pierce	-0.192015487	轻微

6.2 SEIR 模型 Python 代码

```
T = 50 #天数
N = 9300000 #总人数
E = [0]*T #潜伏者人数
I = [1]*T #感染者
S = [N-I[0]]*T #易感人数
R = [0]*T #康复人数
r = 12 #感染者接触易感者的人数
B = 0.053 #传染概率
r2 = 12 #潜伏者接触易感者的人数
B2 = 0.1 #潜伏者传染正常人的概率
a = 0.3 #潜伏者转化为感染者概率
y = 0.6 #康复概率
for i in range(T-1):
    if i+1>=10: #从第 9 天开始，有关部门采取措施，平均接触人数和平均感染率显著下降一直到 3 月 5 号左
        右接触封禁，即到 46 天解除封禁，
            r=3
            r2=3
            B = 0.001
            B2 = 0.001
    if i+1>=46:
        r= 0
        B= 0
        r2 = 0.000001
        B2 = 0.000001
        a= a* 0.3/2.0
        y = 0.6
    S[i+1] = S[i] - r*B*S[i]*I[i]/N - r2*B2*S[i]*E[i]/N
    E[i+1] = E[i] + r*B*S[i]*I[i]/N - a*E[i] + r2*B2*S[i]*E[i]/N
    I[i+1] = I[i] + a*E[i]-y*I[i]
    R[i+1] = R[i] + y*I[i]
import matplotlib.pyplot as plt
plt.xlabel('Date')
plt.ylabel('Number of latent patients')
#plt.plot(S)
plt.plot(E)
#plt.plot(I)
#plt.text(,0,' \leftarrow 正弦');
#plt.text(20,400,'number of E');
#plt.plot(R)
plt.show()
```

6.3 ARIMA 模型 Python 代码

```
import pandas as pd
import numpy as np
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
from datetime import datetime
from pandas import Series, DataFrame
import warnings
warnings.filterwarnings('ignore', 'statsmodels.tsa.arima_model.ARMA',
                        FutureWarning)
warnings.filterwarnings('ignore', 'statsmodels.tsa.arima_model.ARIMA',
                        FutureWarning)

data = pd.read_excel('testdf.xlsx', index_col= u'日期')
data.index = pd.DatetimeIndex(data.index).to_period()
data1 = pd.read_excel('testdf1.xlsx', index_col = u'日期')
forecastnum = 5
plt.rcParams['font.sans-serif'] = ['SimHei']
plt.rcParams['axes.unicode_minus'] = False
data.plot()
data1.plot()
plt.show()
#以上是画时序图

from statsmodels.graphics.tsaplots import plot_acf
plot_acf(data).show()
#以上是自相关图

from statsmodels.tsa.stattools import adfuller as ADF
#print(u'原始序列的 ADF 检验结果为: ', ADF(data[u'确诊总人数']))
#原始序列的 ADF 检验结果为: (-3.7667177304726867, 0.003266793578451666, 9, 19, {'1%': -3.8326031418574136,
'5%': -3.0312271701414204, '10%': -2.655519584487535}, 114.61375882813807)
#得出的结论是不平稳

D_data = data.diff().dropna()
D_data.columns = [u'确诊总人数差分']
#差分后的结果
D_data.plot()
plt.show()
#时序图
plot_acf(D_data).show()
plt.show()
#自相关图

from statsmodels.graphics.tsaplots import plot_pacf
#偏自相关图
#plot_pacf(D_data).show()
#平稳性检测
#print(u'差分序列的 ADF 检验结果为: ', ADF(D_data[u'确诊总人数差分']))
```

```

#差分序列的 ADF 检验结果为: (-1.2113874732141632, 0.6686780253791897, 9, 18, {'1%': -3.859073285322359, '5%':
-3.0420456927297668, '10%': -2.6609064197530863}, 123.53995327978535)
#from statsmodels.stats.diagnostic import acorr_ljungbox
#print(u'差分序列的白噪声检验结果为: ', acorr_ljungbox(D_data, lags=1))
#差分序列的白噪声检验结果为: (array([15.83641633]), array([6.90605033e-05]))
from statsmodels.tsa.arima_model import ARIMA
#定阶
#一般阶数不超过 length/10
pmax = int(len(data)/30)
#一般阶数不超过 length/10
qmax = int(len(data)/20)
#bic 矩阵
bic_matrix = []
for p in range(pmax+1):
    tmp = []
    for q in range(qmax+1):
        try:
            tmp.append(ARIMA(data, (p, 1, q)).fit().bic)
        except:
            tmp.append(None)
    bic_matrix.append(tmp)
#从中可以找出最小值
bic_matrix = pd.DataFrame(bic_matrix)
#先用 stack 展平, 然后用 idxmin 找出最小值位置。
p,q = bic_matrix.stack().idxmin()
print(u' BIC 最小的 p 值和 q 值为: %s、%s' %(p,q))
#取 BIC 信息量达到最小的模型阶数, 结果 p 为 0, q 为 1, 定阶完成。
#建立 ARIMA(0, 1, 1)模型
model = ARIMA(data, (p, 1, q)).fit(dis=0)
#作为期 20 天的预测, 返回预测结果、标准误差、置信区间。
print(model.forecast(20))

```