



2020 年江苏省研究生数学建模科研创新实践大赛

题目 数据驱动疫情特征描述、防疫效果度量 及重启策略建模

摘 要

本题要求通过各国疫情数据，建立合适的模型，从中提取出疫情传播信息，评估各国疫情防控的效果，并通过预测疫情走势对“重启”提出建议。

具体任务如下：

- (1) 建立中位数模型估算病毒的平均潜伏期，在此基础上引入弹性变量，增强变量间的相关性。根据各国的数据求得平均潜伏期长度为 5.489 天。
- (2) 在 (1) 的基础上，建立 SEIR 模型，通过易感、潜伏、感染、移出这四种状态转化比例之间的关系，求得各国的病毒基本传染数 R_0 在 3.93-7.43 之间。将 R_0 作为衡量病毒传染能力的指标，通过纵向比较各国各时期的 R_0 ，认为在人为管控力度逐渐加强的情况下，病毒的传染能力随时间逐渐减弱。
- (3) 首先以中国武汉为例，研究了控制传染源，切断传播途径，保护易感人群这三个措施对疫情防控成果的影响。再通过调节参数 K （感染者平均接触人数）和 R_0 ，建立回溯传输模型对各国疫情发展趋势进行演绎。最后通过演绎趋势图和实际趋势图的比较，结合实际情况对各国防疫成果进行客观评价。本模型得出的结果认为中国、韩国、意大利、法国的防疫成果较好；澳大利亚的防疫成果一般；美国、印度的防疫成果较差。
- (4) “重启”条件为：在传染源与易感人群的接触强度和接触时间都增大的情况下，也不会发生二次爆发。因此本题需要找出人群中病毒携带者的比例降到极低的时间节点。通过逻辑回归模型对中国疫情发展进行拟合预测，建议中国在疫情拐点到来后的第 7 天开始“分地区分批”复工复产，并且“重启”期间，政府的管控系数维持在 0.6，即仍需保持很高的警惕性。逻辑回归模型预测中国在 2 月 13 日迎来疫情拐点，建议从 2 月 20 日起，低传染风险地区开始分批复工，重要民生工业首批复工，一个潜伏周期后，恢复交通，放开人员流动，再经历一个潜伏周期，分批恢复服务业。

关键字： 中位数法 基本传染数 SEIR 模型 回溯传输模型 逻辑回归

目录

一、问题重述.....	3
1.1 问题背景.....	3
1.2 问题的提出.....	3
二、问题分析.....	3
三、符号说明.....	4
四、问题一.....	5
4.1 平均潜伏期计算.....	5
4.1.1 模型假设.....	5
4.1.2 建模分析.....	5
4.1.3 模型建立与求解.....	6
4.1.4 模型结论.....	8
4.2 病毒传染强度随时间的变化.....	8
4.2.1 模型假设.....	8
4.2.2 建模分析.....	9
4.2.3 模型建立与求解.....	9
4.2.4 模型应用.....	10
4.2.5 模型结论.....	14
五、问题二.....	14
5.1 模型假设.....	14
5.2 建模分析.....	15
5.3 模型建立与求解.....	15
5.4 模型结论.....	20
六、问题三.....	21
6.1 模型假设.....	21
6.2 建模分析.....	21
6.3 模型建立与求解.....	22
6.4 模型结论.....	23
七、模型总结与评价.....	24
参考文献.....	25

一、问题重述

1.1 问题背景

当下，COVID-19 疫情正肆虐全球，这是一种新病毒疫情，人们对于这场疫情有太多的认知缺失，导致防疫工作非常被动。比如该病毒的潜伏期多长、潜伏期的传染性多大，还比如无症状感染者的传染性多强等。另一方面，疫情的控制关乎千家万户，各国的国情和民情不同，采取控制疫情的手段和程度也不同，导致防疫效果千差万别。此外，COVID-19 疫情对经济的冲击前所未有，世界各国（地区）都面临着防疫和开放经济活动的两难选择，经过几个月的防疫，各国逐步开始“重启”，客观衡量各国“重启”时机选择的合理性将有利于决策者科学地制定“重启”策略。

1.2 问题的提出

现在关于 COVID-19 的潜伏期是通过对感染者的跟踪和回溯得到的大致时间，有较大的不确定性。潜伏期的人（包括无症状感染者）具有传染性，且潜伏期是随机的，因人而异。建立估计潜伏期的分布和传染强度随时间变化的模型，并估计 COVID-19 的平均潜伏时间，将有利于政策制定者更合理的安排防疫举措。

传染病的防治原则：控制传染源；切断传播途径；保护易感人群。各国（地区）的检测程度有较大的差异，而戴口罩和保持社交距离的防疫举措在世界各地执行的情况也不尽相同，造成了各国疫情的传播趋势千差万别。根据疫情传播趋势数据建立数学模型度量各国防疫效果，利用所建立的模型计算并对比美国、意大利、法国、澳大利亚、韩国、印度等国家的防疫效果，分析造成防疫效果差异的原因。

经过几个月的防疫，各国逐步开始“重启”，这会带来传染源与易感人群的接触强度和接触时间的增大，容易产生二次大规模传播，所以确保不出现疫情的二次爆发是“重启”的必要条件。建立“重启”时机选择策略的数据分析模型，分析第一问中某个国家的“重启”时机选择策略，并给出“重启”建议。

二、问题分析

关于问题一：

本题共两小问，需要分别估计 COVID-19 的平均潜伏时间和病毒传染强度随时间的变化趋势。

（1）本题要求根据各国新冠肺炎的实时数据，建立估算病毒平均潜伏期分布的模型。传染病潜伏期长度计算的方法主要有两种：几何均数法和中位数法。从官方新闻中得知，此次疫情的潜伏期以天为单位，是离散型计量数据，且潜伏期少有成指数模式增长的情况发生^[1]。因此，本题采用中位数法。

从 WHO 网站获得意大利、法国、澳大利亚、韩国、印度和中国从首次出现确诊病例至今的总确诊人数和单日新增确诊人数的数据。将总确诊数的 16%,50%,84% 作为中位数法选取数据的标志比例，将得到的首部，中部，尾部的总确诊人数代入到模型中，估算出该国家的病毒潜伏期。

(2) 从官方信息中可以得知, 携带新冠病毒的潜伏者同样具有感染健康人群的能力。因此, 选用 SEIR 模型对人群结构进行建模, 将总样本分为易感者, 潜伏者, 感染者, 移出者四个部分^[5]。病毒感染强度, 政府管控力度和民众自我防范意识等因素都会影响四类人群之间的转化率。本题通过计算各国不同时期四类人群转化率的大小估算病毒的传染强度随时间变化的规律, 并使用基本传染数 R_0 作为传染强度大小的指标。

关于问题二:

几乎所有新病毒造成的传染病都会在疫源地存在遗漏病例的情况, 随着时间的发展, 病例的确诊方式会得到优化, 而且监管力度也会得到加强。根据传染病动力学原理, 传染病的防治原则为: 控制传染源; 切断传播途径; 保护易感人群。控制传染源首先要确定传染源, 这可以快速掌握传染链。实现切断传播途径和保护易感人群的简便方法就是戴口罩和保持社交距离, 这可以有效降低病毒携带者的接触人数。

本题收集了美国、意大利、法国、澳大利亚、韩国、印度等国家的确诊数据集, 建立模型通过调节参数 K 和 R_0 回溯疫情的理想趋势, 并与上述国家实际的防疫效果进行对比, 并分析造成防疫效果差异的原因。

关于问题三:

本题需要预测复工时间, 这要求在传染源与易感人群的接触强度和接触时间的增大条件下, 不产生二次大规模传播。想要达成这个目标需要可流动人群中, 携带病毒人群数量所占的比例足够小。因为模型假设了感染者都会被确诊并有效隔离, 所以本模型将复工时间点设置在疫情到达拐点的一个星期后。疫情到达拐点, 说明病毒传播已经被有效控制。经过上面的计算, 经过 7 天后, 原本的潜伏者绝大部分也转化为感染者, 人群中可以自由活动的病毒携带者比例降到极低。通过中国的实际确诊人数数据, 使用逻辑斯蒂模型对疫情发展趋势进行预测, 找出疫情拐点, 确定复工日期。

三、符号说明

符号	含义
α	中位数组的下限
N	样本例数
I	组距
$\sum f\alpha$	中位数在前一组的累计频数
fmd	中位数组的例数
m_1	产生总确诊人数 16% 的确诊人数的时间
m_0	产生总确诊人数 50% 的确诊人数的时间
m_2	产生总确诊人数 84% 的确诊人数的时间
X	由共同暴露日期至 m_0 为的时间
\underline{j}	取样组数
\overline{X}	平均潜伏期
S	易感状态人群
E	潜伏状态人群

I	感染状态人群
R	移出状态人群
$S(t)$	t 时刻的易感人群数
$E(t)$	t 时刻的潜伏人群数
$I(t)$	t 时刻的感染人群数
$R(t)$	t 时刻的移出人群数
N	群体总人数
β	单位时间易感人群与感染个体接触并被传染的概率
k	感染者每天平均接触的人数
b	与感染者接触时被传染的概率
γ_1	潜伏个体在单位时间内转化为感染个体的概率
γ_2	感染个体在单位时间内转化为移出状态的概率
R_0	基本传染系数
C	恢复期
M	总确诊人数，包括潜伏和感染并发的人
m	潜伏期的病人数
n	处于潜伏期仍具备感染能力的人数
N	总人数
t	从感染到确诊的平均时间
ω	是权值向量
b	偏置
P	事件发生的概率
x	输入自变量
Y	输出因变量
$P(t)$	总人数函数
r	增长阻力

四、问题一

4.1 平均潜伏期计算

4.1.1 模型假设

- (1) 病毒的潜伏期短
- (2) 病毒源头为一个共同来源
- (3) 流行只有一个高峰
- (4) 病例集中在最长与最短的潜伏期内

4.1.2 建模分析

本题要求根据各国新冠肺炎的实时数据，建立估算病毒平均潜伏期分布的模型。传染病潜伏期长度计算的方法主要有两种：几何均数法和中位数法。从官方新闻中得知，此次疫情的潜伏期以天为单位，是离散型计量数据，且潜伏期少有成指数模式增长的情况发生。^[1]因此，本题采用中位数法。

从 WHO 网站获得意大利、法国、澳大利亚、韩国、印度和中国从首次出现确诊病例至今的总确诊人数和单日新增确诊人数的数据。将总确诊数的 16% ,50% ,84%作为中位数法选取数据的标志比例^[2]，将得到的首部，中部，尾部的总确诊人数代入到模型中，估算出该国家的病毒潜伏期。

4.1.3 模型建立与求解

(1) 中位数法估算病毒的潜伏期模型^{[2][3]}

$$m_d = \alpha + \frac{\frac{N}{2} - \sum f\alpha}{fmd} \cdot i \quad (4-1)$$

其中， α 为中位数组的下限， N 为样本例数， i 为组距， $\sum f\alpha$ 为中位数在上一组的累计频数， fmd 为中位数组的例数。

(2) 平均潜伏期估算:

$$\overline{X} = \frac{(m_2 - m_0) \times (m_0 - m_1)}{(m_2 - m_0) + (m_0 - m_1)} \quad (4-2)$$

m_1, m_0, m_2 为产生总确诊人数 16%，50%，84%的确诊人数的时间， \overline{X} 为平均潜伏期时间长度。

(3) 平均潜伏期估算公式优化

由于影响病例增长的因素错综复杂，增强变量之间的相关性，对模型引入松弛变量，使该模型的适应性更强。

优化后的公式:

$$\overline{X} = \frac{(m_2 - m_0) \times (m_0 - m_1)}{(m_2 - m_0) + (m_0 - m_1)} - j \quad (4-3)$$

其中 j 为取样组数。
 由于病毒的传播尚未结束，因此选取从出现确诊病例的 1 月 24 日为开始日期，以 7 月 13 日作为截止日期，以 3 天为一个步长，将澳大利亚、意大利、法国、印度、韩国、中国等国家的每日新增确诊人数及总确诊人数，代入公式 (4-3) 估算病毒的潜伏期，结果如下表:

表 4-1 澳大利亚潜伏期估算结果

百分比	日期	总人数	$m_2 - m_0$	$m_0 - m_1$	潜伏期
100%	7 月 13 日	10251			
16%	3 月 24 日	1640			
50%	4 月 2 日	5125	31	9	4.975
84%	7 月 7 日	8610			

表 4-2 意大利潜伏期估算结果

百分比	日期	总人数	$m_2 - m_0$	$m_0 - m_1$	潜伏期
100%	7 月 13 日	243061	27	17	7.43
16%	3 月 19 日	38890			
50%	4 月 5 日	121530			
84%	5 月 2 日	204171			

表 4-3 法国潜伏期估算结果

百分比	日期	总人数	$m_2 - m_0$	$m_0 - m_1$	潜伏期
100%	7 月 13 日	161275	30	12	5.57
16%	3 月 27 日	25804			
50%	4 月 8 日	80637			
84%	5 月 8 日	135471			

表 4-4 印度潜伏期估算结果

百分比	日期	总人数	$m_2 - m_0$	$m_0 - m_1$	潜伏期
100%	7 月 13 日	936181	15	30	7
16%	5 月 26 日	149788			
50%	6 月 25 日	468090			
84%	7 月 10 日	786392			

表 4-5 韩国潜伏期估算结果

百分比	日期	总人数	$m_2 - m_0$	$m_0 - m_1$	潜伏期
100%	7 月 13 日	13479	79	9	5.03
16%	2 月 28 日	2156			
50%	3 月 8 日	6739			
84%	5 月 26 日	11322			

表 4-6 中国潜伏期估算结果

百分比	日期	总人数	$m_2 - m_0$	$m_0 - m_1$	潜伏期
100%	7 月 13 日	85568			
16%	2 月 2 日	13690	7	9	3.93
50%	2 月 11 日	42784			
84%	2 月 18 日	71877			

4.1.4 模型结论

综上，本题结论如下图 4-1 所示：

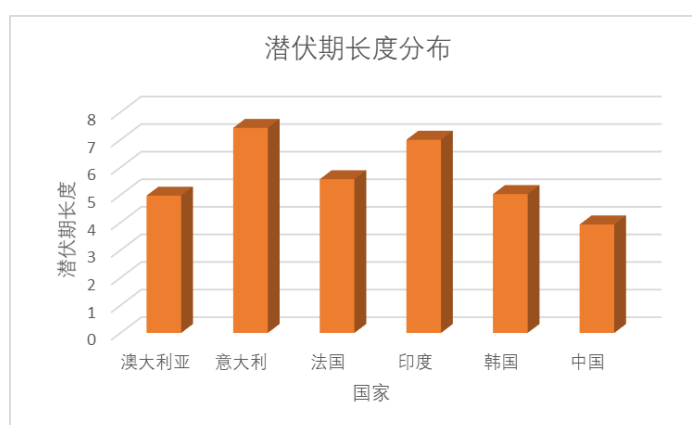


图 4-1 潜伏期长度分布

表 4-7 各国病毒潜伏期估算结果

澳大利亚	意大利	法国	印度	韩国	中国	平均潜伏期
4.975	7.43	5.57	7	5.03	3.93	5.489

利用中位数法，估算出病毒的平均潜伏期为 5.489 天。

4.2 病毒传染强度随时间的变化

4.2.1 模型假设

SEIR 模型中，将研究对象分为 S、E、I、R 四种类型。

- (1) S 为易感状态人群，模型中假定只有患病群体直接接触到的人才处于易感状态。^[4]
- (2) E 为处于潜伏期的人群，表示已经被感染但没有表现出感染症状的群体，且具有传染能力。
- (3) I 为感染状态人群，具有传染能力。
- (4) R 为移出状态人群，包括病愈、死亡、有效隔离的人，不具有病毒传染能力且移出后不会再次被感染。^{[5][6]}

四种类型人群的转化方向如图 4-2 所示：

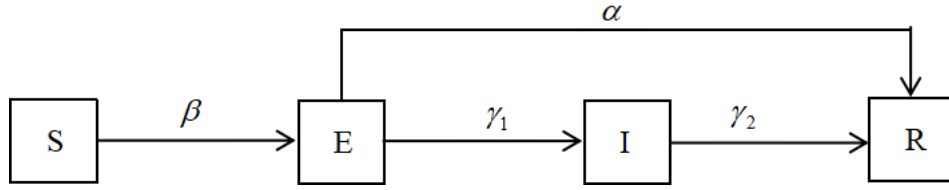


图 4-2 SEIR 传染病动力学模型的人群转化

4.2.2 建模分析

从官方信息中可以得知，携带新冠病毒的潜伏者同样具有感染健康人群的能力。因此，选用 SEIR 模型对人群结构进行建模，将总样本分为易感者，潜伏着，感染者，移出者四个部分。病毒感染强度，政府管控力度和民众自我防范意识等因素都会影响四类人群之间的转化率。本题通过计算各国不同时期四类人群转化率的大小估算病毒的传染强度随时间变化的规律，并使用基本传染数 R_0 作为衡量传染强度大小的指标。

4.2.3 模型建立与求解

记 $S(t)$ ， $E(t)$ ， $I(t)$ ， $R(t)$ 分别为 t 时刻的易感人群数，潜伏人群数，感染人群数，移出人群数。 N 为群体总人数，有

$$S(t) + E(t) + I(t) + R(t) \equiv N(t) \quad (4-4)$$

假设一个易感状态的人在单位时间 τ 里与感染个体接触并被传染的概率为 β ，由于易感人群的比例为 S/N ，时刻 t 网络中总共有 $I(t)$ 个感染个体，所以易感个体的数目按照如下变化率减少^[7]

$$\frac{dS}{dt} = -\frac{\beta SI}{N} \quad (4-5)$$

假设感染者每天平均接触的人数为 k ，接触时传染的概率为 b ，有

$$\beta = kb \quad (4-6)$$

相应地，潜伏个体的数目按照如下变化率增加，并且整体以单位时间概率 γ_1 转化为感染个体^[7]

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \gamma_1 E \quad (4-7)$$

感染个体数目由潜伏群体提供，同时以单位时间概率 γ_2 转化为移出状态^[8]

$$\frac{dI}{dt} = \gamma_1 E - \gamma_2 I \quad (4-8)$$

相应的，感染群体以概率 γ_2 由向移除个体转化

$$\frac{dR}{dt} = \gamma_2 I \quad (4-9)$$

综上，可得：

$$\frac{dI}{dt} = \gamma_1 \beta \frac{IS}{N} - \gamma_2 I \quad (4-10)$$

$$I(t) = e^{(\gamma_1 \beta \frac{IS}{N} - \gamma_2) t} \quad (4-11)$$

$$I(t) = e^{(\gamma_1 k b \frac{IS}{N} - \gamma_2) t} \quad (4-12)$$

基本传染系数^{[7][8]}

$$R_0 = \frac{\beta}{\gamma_2} = \frac{kb}{1/C} \quad (4-13)$$

4.2.4 模型应用

以武汉为例：

由于中国人口总数庞大，湖北地区与湖北以外地区疫情差距极大，为减少结果偏差，中国地区以武汉为研究对象。本次疫情的传染周期一般为 7 天^[9]，故本题中将 7 天作为一组求 b 、 R_0 的值，按照疫情爆发前期（1 月 25 日至 1 月 30 日）、中期（2 月 25 日至 3 月 3 日）、后期（3 月 8 日至 3 月 14 日）选取武汉市卫生健康委员会公示的数据。2020 年 1 月 8 日的确诊病例为 1 人，为第 0 天，武汉人口为 1300 万，死亡率为 3%，COVID-19 的恢复期 C 为 14 天，总确诊人数函数为 $I(t)$ ，有

$$I(0)=1$$

在疫情早期，由于防控意识淡薄，本题中假定感染者每天平均接触人数为 $k=5$ ，系统中的全部人口都处于易感状态，

武汉市疫情前期（1 月 25 日至 1 月 30 日）总确诊人数如表 4-8 所示：

表 4-8 武汉市疫情前期的总确诊人数

日期	总确诊人数/人	t/天
1 月 25 日	618	47
1 月 26 日	698	48
1 月 27 日	1590	49
1 月 28 日	1905	50
1 月 29 日	2261	51
1 月 30 日	2639	52
1 月 31 日	3215	53

将以上数据带入微分方程组，利用 matlab 程序对 SEIR 模型进行求解，可得：

$$b=0.0444(0.04378,0.04501)$$

$$R_0=3.108$$

武汉市疫情中期（2 月 25 日至 3 月 3 日）总确诊人数如表 4-9 所示：

表 4-9 武汉市疫情中期的总确诊人数

日期	总确诊人数/人	t/天
2 月 25 日	47441	78
2 月 26 日	47824	79
2 月 27 日	48137	80
2 月 28 日	48557	81
2 月 29 日	49112	82
3 月 1 日	49315	83
3 月 2 日	49426	84
3 月 3 日	49540	85

将以上数据带入微分方程组，利用 matlab 程序对 SEIR 模型进行求解，可得：

$$b=0.04045(0.03984,0.04106)$$

$$R_0=2.8315$$

武汉市疫情后期（3 月 8 日至 3 月 14 日）总确诊人数如表 4-10 所示：

表 4-10 武汉市疫情晚期的总确诊人数

日期	总确诊人数/人	t/天
3 月 8 日	49948	90
3 月 9 日	49965	91
3 月 10 日	49978	92
3 月 11 日	49986	93
3 月 12 日	49991	94
3 月 13 日	49995	95
3 月 14 日	49999	96

将以上数据带入微分方程组，利用 matlab 程序对 SEIR 模型进行求解，可得：

$$b=0.03738(0.03689, 0.03787)$$

$$R_0=2.6166$$

利用以上模型，计算了韩国、美国、意大利、澳大利亚、法国、印度在不同时期平均接触人数和基本传染系数，结果如下：

表 4-11 韩国在疫情不同时期的总确诊人数

早期			中期			晚期		
日期	总人数	t	日期	总人数	t	日期	总人数	t
2/17	30	28	3/15	8162	55	4/17	10635	88
2/20	104	31	3/18	8320	58	4/20	10674	91
2/23	602	34	3/21	8897	61	4/23	10702	92
2/26	1261	37	3/24	9037	64	4/26	10728	93
2/29	3150	40	3/27	9332	67	4/29	10761	94
3/3	4812	43	3/30	9661	70	5/2	10780	95
3/6	6284	46	4/2	9976	73	5/5	10804	96
b			b			b		
0.05277			0.04053			0.03257		
(0.05205, 0.05349)			(0.03865, 0.04242)			(0.03152, 0.03361)		
R ₀			R ₀			R ₀		
3.6939			2.8371			2.2799		

表 4-12 美国在疫情不同时期的总确诊人数

早期			中期			晚期		
日期	总人数	t	日期	总人数	t	日期	总人数	t
3/9	213	49	4/14	632781	85	6/10	1951096	137
3/12	987	52	4/20	723605	88	6/13	2010391	140
3/15	1678	55	4/23	800926	91	6/16	2079592	143
3/18	3536	58	4/26	899281	94	6/19	2149166	146
3/21	15219	61	4/29	983457	97	6/22	2241178	149
3/24	42164	64	5/2	1067127	100	6/25	2329463	152
3/27	68334	67	5/5	1154985	103	6/28	3452048	155
b			b			b		
0.04732			0.042			0.03358		
(0.04666, 0.04798)			(0.04095, 0.04306)			(0.03293, 0.03423)		
R ₀			R ₀			R ₀		
5.29984			4.704			3.76096		

表 4-13 意大利在疫情不同时期的总确诊人数

早期			中期			晚期		
日期	总人数	t	日期	总人数	t	日期	总人数	t
2/23	76	20	4/17	168941	74	6/10	235561	128
2/26	322	23	4/20	178932	77	6/13	236305	131
2/29	888	26	4/23	187327	80	6/16	237290	134
3/3	2036	29	4/26	195351	83	6/19	238159	137
3/6	3858	32	4/29	201505	86	6/22	238499	140
3/9	7374	35	5/2	207428	89	6/25	239410	143
3/12	12462	38	5/5	211938	92	6/28	240136	146
b			b			b		
0.06422			0.04179			0.03173		
(0.06358, 0.06485)			(0.04026, 0.04331)			(0.03097, 0.03244)		
R ₀			R ₀			R ₀		
4.5395			2.9253			2.2211		

表 4-14 澳大利亚在疫情不同时期的总确诊人数

早期			中期			晚期		
日期	总人数	t	日期	总人数	t	日期	总人数	t
3/6	57	20	4/14	6366	59	6/13	7290	119
3/9	77	23	4/17	6468	62	6/16	7335	122
3/12	122	26	4/20	6612	65	6/25	7521	125
3/15	249	29	4/23	6654	68	6/28	7641	128
3/18	414	32	4/26	6703	71	7/1	7834	131
3/21	873	35	4/29	6738	74	7/4	8255	134
3/24	1709	38	5/2	6767	77	7/7	8586	137
b			0.03826			0.02794		
(0.05291, 0.05365)			(0.03644, 0.04008)			(0.02742, 0.02845)		
R ₀			2.6782			1.9558		

表 4-15 法国在疫情不同时期的总确诊人数

早期			中期			晚期		
日期	总人数	t	日期	总人数	t	日期	总人数	t
2/26	12	30	4/26	122875	90	6/16	152819	141
2/29	57	33	4/29	125464	93	6/19	153557	144
3/3	191	36	5/2	128722	96	6/22	154567	147
3/6	534	39	5/5	130242	99	6/25	155087	150
3/9	1116	42	5/8	135980	102	6/28	156156	153
3/12	2269	45	5/11	137073	105	7/1	157194	156
3/15	4469	48	5/14	138609	108	7/4	158734	159
b			0.03695			0.02982		
(0.04836, 0.04963)			(0.0358, 0.0381)			(0.02924, 0.0304)		
R ₀			2.5865			2.0874		

表 4-16 印度在疫情不同时期的总确诊人数

早期			中期			晚期		
日期	总人数	t	日期	总人数	t	日期	总人数	t
3/24	434	33	5/2	37336	72	6/13	308993	114
3/27	724	36	5/5	46496	75	6/16	343091	117
3/30	1071	39	5/8	56342	78	6/19	380532	120
4/2	1636	42	5/11	67152	81	6/22	425282	123
4/5	3374	45	5/14	78003	84	6/25	473105	126
4/8	5194	48	5/17	90927	87	6/28	528859	129
4/11	7447	51	5/20	106750	90	7/1	585493	132
b			0.0405			0.03479		
(0.0491, 0.0499)			(0.03971, 0.04129)			(0.03425, 0.03532)		
R ₀			4.536			3.89648		

4.2.5 模型结论

各国疫情早期、中期、后期基本传染数的如表 4-17 所示：

表 4-17 不同时期各国家的病毒传染强度

国家\疫情传染强度	早期	中期	晚期
中国（武汉）	3.108	2.8315	2.6166
韩国	3.6939	2.8371	2.2799
美国	5.29984	4.704	3.76096
意大利	4.5395	2.9253	2.2211
澳大利亚	3.7296	2.6782	1.9588
法国	3.43	2.5865	2.0874
印度	5.544	4.536	3.89648

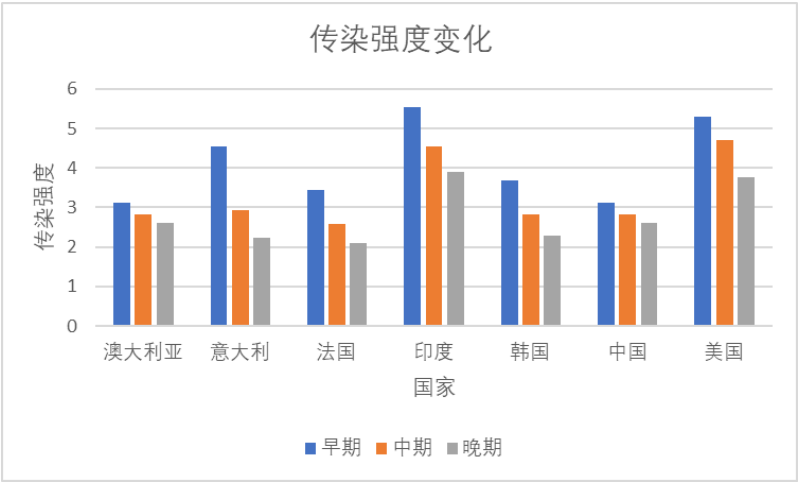


图 4-3 病毒传染强度随时间变化

R_0 值越小说明疫情控制得越好，当 R_0 递减甚至小于 1 后，病毒的传播就会逐渐消失。如图 4-3 所示， R_0 的大小在逐渐变小。说明从疫情爆发至今，各国政府的管控力度不断增大，人民的防范意识不断增强，病毒的传染强度在人为控制下逐渐减弱了。

中国在疫情初期，就对武汉封城，并在全国停工停产，很好地阻断了病毒的传播，所以中国的 R_0 在各国中处于一个很低的水平。意大利早期的 R_0 很大，但在政府决定封城后，中期的 R_0 迅速变小，疫情发展逐渐向好。而美国，在早期疫情爆发后，仍没有有效的防疫举措，可以看到虽然传染强度在减弱，但防疫形式依然严峻。

各国的情况都说明，控制传染源，切断传播途径，保护易感人群可以有效降低病毒传染强度。

五、问题二

5.1 模型假设

- (1) 假设潜伏人群可以自由移动，且随机分布在总人群中；
- (2) 假设感染病毒的人群最终都会被确诊。

5.2 建模分析

几乎所有新病毒造成的传染病都会存在疫源地存在遗漏病例的情况，随着时间的发展，病例的确诊方式会得到优化，而且监管力度也会得到加强。^{[10][11]}根据传染病动力学原理，传染病的防治原则：控制传染源；切断传播途径；保护易感人群。控制传染源首先要确定传染源，这可以快速掌握传染链。实现切断传播途径和保护易感人群的简便方法就是戴口罩和保持社交距离，这可以有效降低病毒携带者的接触人数^{[10][12]}。

本题收集了美国、意大利、法国、澳大利亚、韩国、印度等国家的确诊数据集，建立模型通过调节参数 K 和 R_0 回溯疫情的理想趋势，并与上述国家实际的防疫效果进行对比，并分析造成防疫效果差异的原因。

5.3 模型建立与求解

回溯传播模型：

假设确诊总人数为 M ，包括潜伏和感染并发的人，其中潜伏病人数为 m ，潜伏人群占总确诊人数的比例为

$$P=m/M \tag{5-1}$$

其中 P 由平均潜伏期内感染比例求得^[13]

$$P=n/N \cdot T \tag{5-2}$$

其中 n 为潜伏期仍具备感染能力（即未被有效隔离）的人数， N 为总人数， T 为从感染到确诊的平均时间

$$T=\text{平均潜伏期}+\text{发病到确诊的平均时间} \tag{5-3}$$

由式（5-1）（5-2）（5-3）可以推出感染的人数 M 。

以中国武汉为例，利用回溯传播模型模拟四类人群数(易感人群 S 、潜伏人群 E 、感染人群 I 、移出人群 R)随时间的变化情况。

（1）此时武汉已经封城，但前七天对城内居民的行动未过多限制。此时 $N=1100$ 万， $C=14$ ， $K=5$ ， $R_0=3.108$ 。

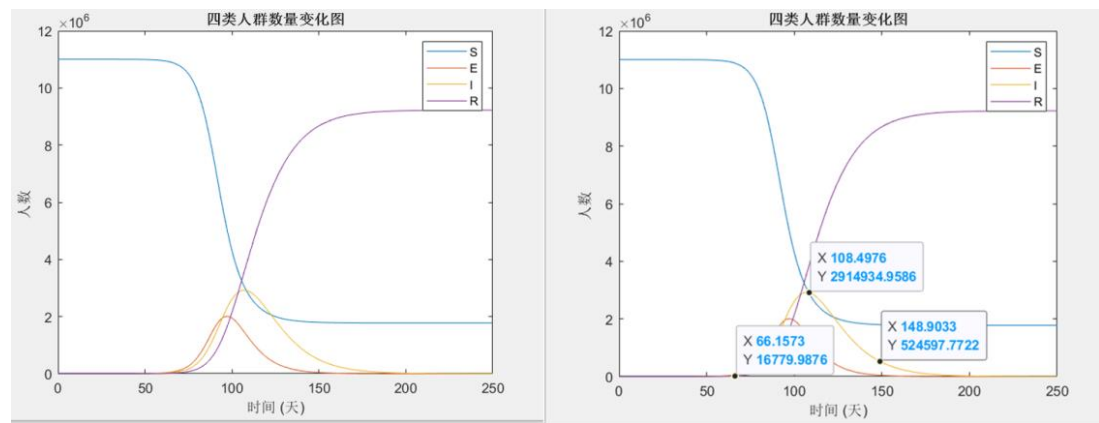


图 5-1 未采取措施的四类人群变化趋势模拟结果

由模拟结果可知，感染人群的曲线在第 66 天左右开始递增，表明中国武汉的疫情在 2 月 6 日左右开始集中爆发，感染人群的曲线在第 108 天达到极大值，此时疫情的传播到达高峰（3 月末，四月初），潜伏期人群数在第 150 天降为 0，表明疫情的传播接近尾声（5 月上旬）。

(2) 根据上一问可知，采取一定的管控手段和人们自发的避免感染，会减小病毒的基本传染数。此时 $N=1100$ 万， $C=14$ ， $K=5$ ， $R_0=2.6166$ 。

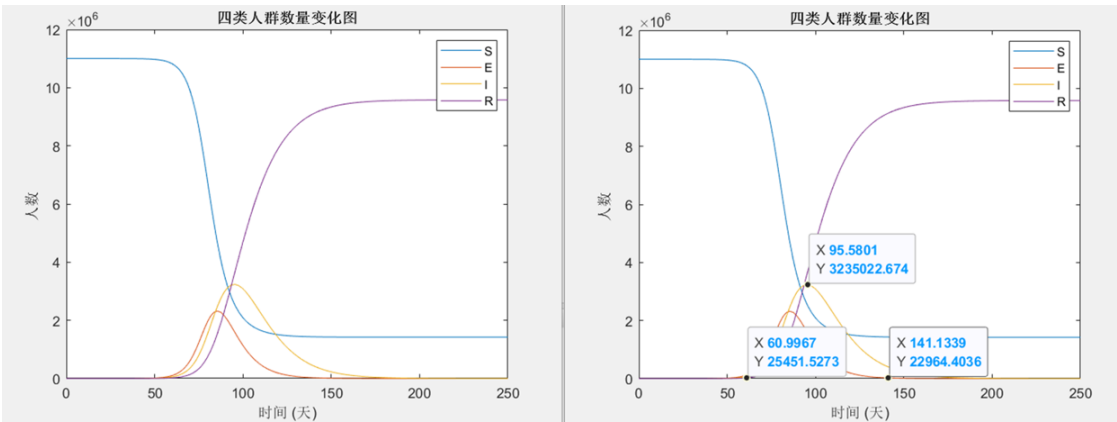


图 5-2 采取一定管控手段后四类人群变化趋势模拟结果

疫情在第 60 天左右开始集中爆发（1 月末），第 95 天左右达到高峰（3 月中旬），第 140 天后接近尾声（4 月下旬）。这也说明早期武汉政府披露的感染人数可能比实际少得多，这使得初期的基本传染数较大。（检测跟不上，医疗设备缺乏，无法满足检测和治疗的需求）

(3) 很快，政府对武汉的交通进行了限制，并关闭了各种不必要的公共场所。此时保持传染基本数不变，感染者每天平均接触到的人数 k 减少。此时 $N=1100$ 万， $C=14$ ， $K=1$ ， $R_0=3.108$ 。

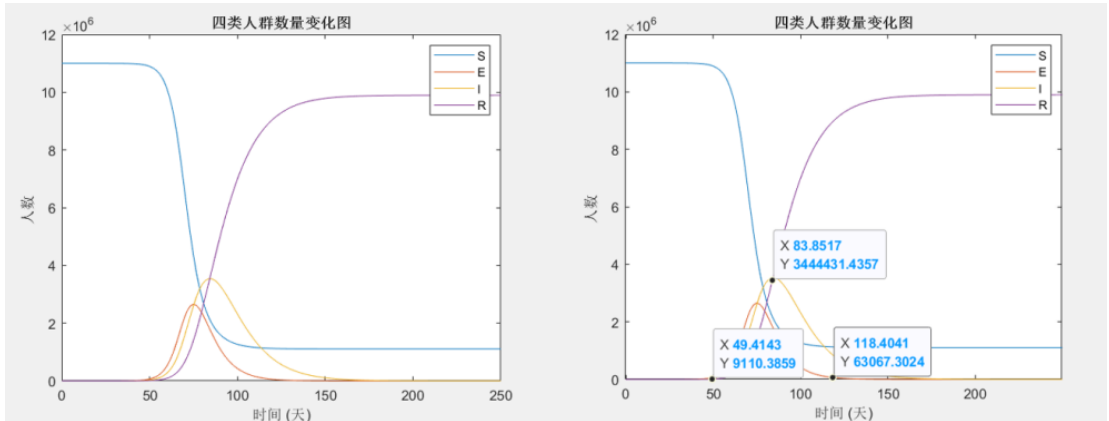


图 5-3 交通管制后四类人群变化趋势模拟结果

疫情在第 50 天左右开始集中爆发（1 月 25 日左右），第 84 天左右达到高峰（3 月上旬），第 120 天后接近尾声（4 月中旬）。该结果表明，管控措施对疫情防治作用很明显。

(4) 从上一问可知，随着时间推移，人们的防范意识和政府管控加强，基本传染系数 R_0 会减少。此时 $N=1100$ 万， $C=14$ ， $K=1$ ， $R_0=2.6166$ 。

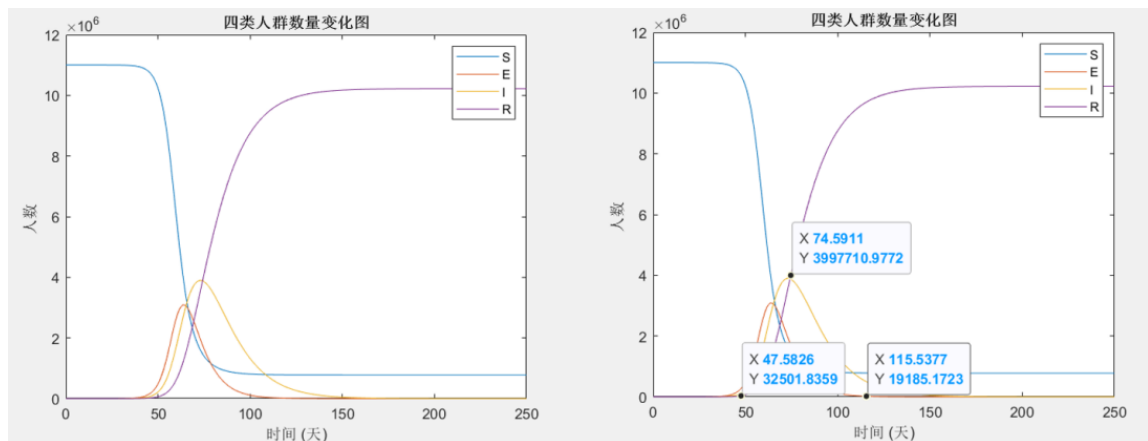


图 5-4 政府强化管控后四类人群变化趋势模拟结果

疫情在第 48 天左右开始集中爆发（1 月 23 日左右），在第 75 天左右达到高峰（2 月末），在第 125 天后接近尾声（4 月上旬）。与武汉实际情况比较相符，说明政府当前的管控力度非常强。

前面已经验证了中国疫情防控取得了很大成效，接下来将使用回溯传播模型预测各国的理想防疫趋势，并和实际趋势进行比较来评判各国的防疫成果。

中国武汉、美国、韩国、印度、意大利、法国、澳大利亚的四类人群数(易感人群 S、潜伏人群 E、感染人群 I、移出人群 R)随时间的变化情况，并与实际的数据相比较。

使用下表数据进行回溯模型建模：

表 5-1 各国回溯模型数据

国家\参数	N	K	R_0
中国（武汉）	11000000	1	2.1166
美国	330000000	8	3.76096
韩国	52230000	5	2.2799
印度	1354051854	8	3.89648
意大利	60484000	3	2.2211
法国	65233271	5	2.2874
澳大利亚	25440000	5	1.9558

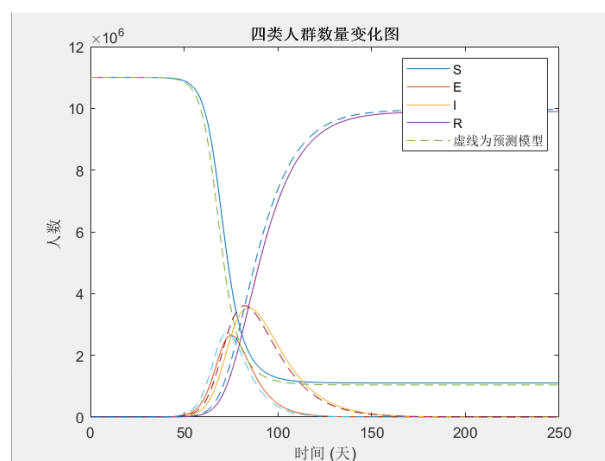


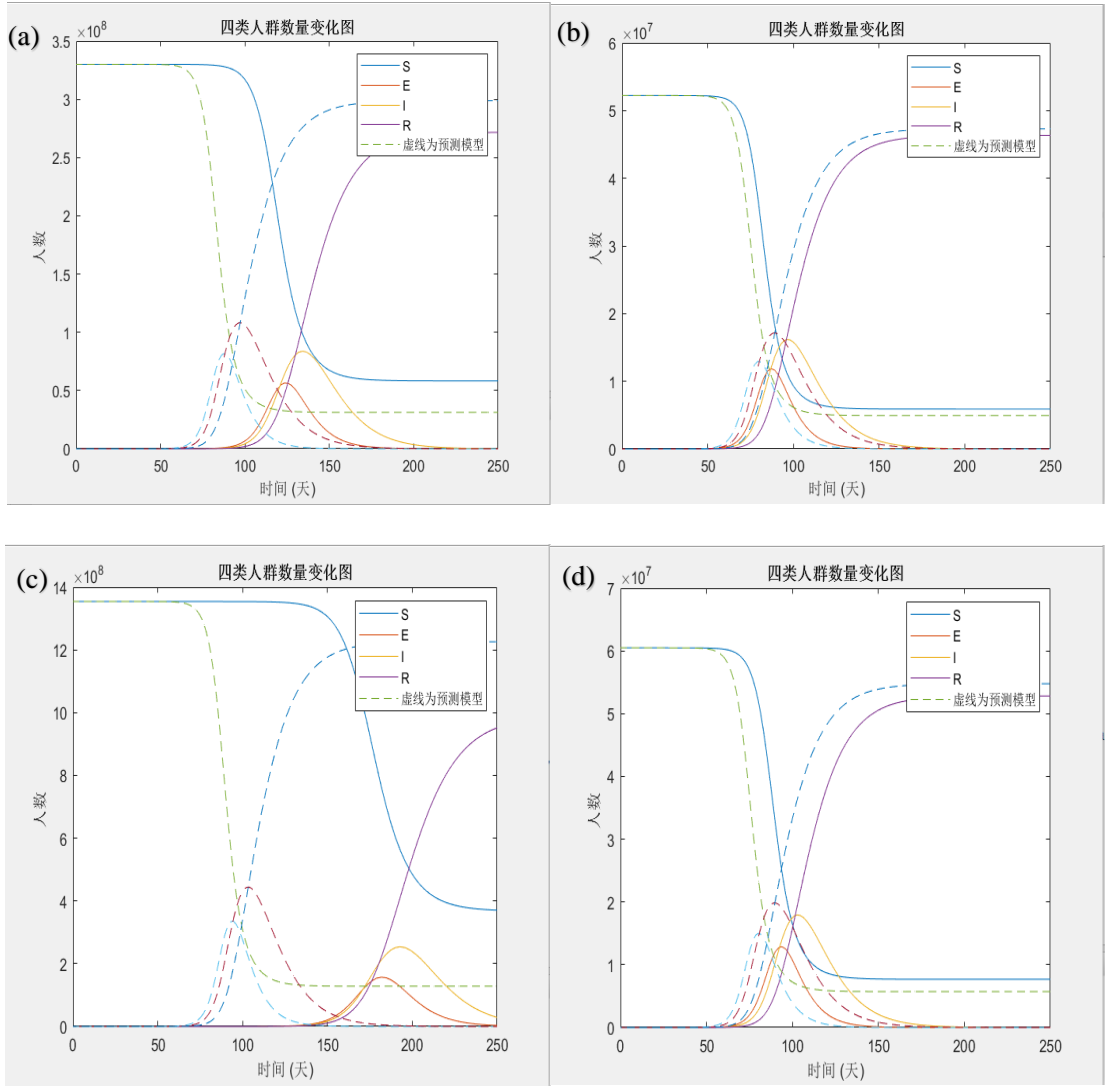
图 5-5 中国（武汉）的模拟结果与真实数据对比

从控制感染源的角度来讲，中国很早成功检测并向世界公布了新冠病毒的基因序列，并大批量生产检验试纸。在疫情重点区域进行了大规模核酸检测，提出了“应收尽收，不漏一人”的口号^[2]。中国强大的科研能力和政府执行力使国家很快就控制住了传染源。

从切断传播途径的角度来看，在疫情初期，党中央就迅速下令，对武汉进行封城，全国范围内停运市内交通，并关闭人员密集型服务场所^[2]。全国人民万众一心，自发的在家中隔离，减少外出活动。这些措施都极大地阻断了病毒从疫源地向全国范围内扩散，也极大地降低了潜伏者感染健康者的可能性。

从保护易感人群的角度来说，中国的方舱医院对感染者进行了有效隔离，阻断了二次感染，极大地降低了感染总人数，使疫情快速到达拐点，加快了疫情结束的脚步。

但上图中，理想回溯的图像和实际图像并未完全重合。这是因为在疫情初期，人们对病毒的意识不足，政府的检测能力较弱，并且存在大量的潜伏期病人仍在自由活动，因而在这一时期实际的感染人数应多于确诊人数。这使得总确诊人数略高于理想模型，疫情高峰、结束期的到来也相应延后。



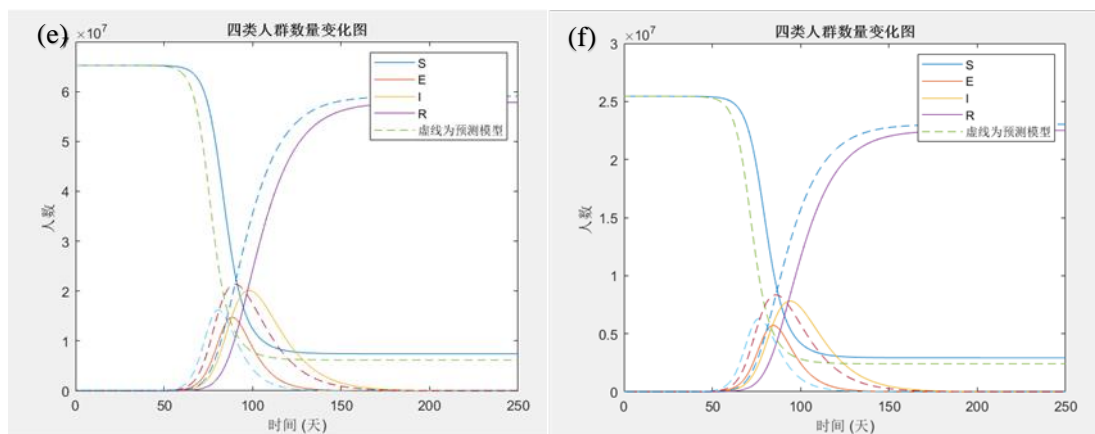


图 5-6 各国回溯模型与实际模型对比 (a)美国 (b) 韩国
(c) 印度 (d) 意大利 (e) 法国 (f) 澳大利亚

(a) 美国

控制感染源：美国每天可以确诊数万的感染病例，说明其具有强大的检验能力。

切断感染源：未停工停产，民众不戴口罩且可以参加聚集性活动，

保护易感人群：未建立方舱医院，无法有效隔离易感人群。

从图 (a) 中可以看出，美国的疫情高峰和结束日期相较于预测模型都大大推迟了。这说明美国政府并没有有效地切断传染源，保护易感人群，实际的感染人数仍呈井喷式爆发，实际感染人数远超每日新增的确诊人数。实际移出群体数量也大量减少，说明由于防疫失利，造成了大量的民众因感染肺炎而死亡。

(b) 韩国

控制感染源：韩国政府的病毒检测能力很强。

切断传染源：未停工停产，但民众普遍戴口罩。

保护易感人群：因为较好的医疗基础，韩国医院对已经确诊的人进行了较为有效的隔离。

从图 (b) 可以看出，韩国政府的防控工作做的较好，民众普遍戴口罩也极大地减弱了病毒的传染，但未停工停产还是造成了部分地区性传染事件的发生。

(c) 印度

控制感染源：印度人口众多，数量庞大的贫困人口得不到及时的检测。

切断传染源：占人口绝大多数的贫困地区是医疗和防控的真空区，且人口密度极大，印度政府无法真正地切断传染源。

保护易感人群：印度基础设施和医疗储备很差，完全做不到有效隔离。

从图 (c) 中可以看出，印度的检测速度远远落后于病毒的传染速度。印度政府没有办法控制疫情的发展。

(d) 意大利

控制感染源：意大利政府在接受了中国政府的大量援助后，获得了大量的检验试纸，检验能力较强。

切断感染源：停工停产，禁止了民众的非必须外出活动，要求全民戴口罩。

保护易感人群：学习了中国的方舱医院，有效的阻断了病毒的二次传染。

从图（d）可以看出，意大利整体防控做得较好，但疫情早期并未采取有效的措施，造成了大面积感染和医疗击穿的情况，且意大利老龄化严重,有一定数量的人群未确诊就因为新冠病毒去世。

（e）法国

控制感染源：作为老牌资本主义强国，法国政府的病毒检验能力很好。

切断传染源：未停工停产，但鼓励民众戴口罩。

保护易感人群：对感染的人群进行了有效隔离。

从图（e）可以看出，法国政府由于自身实力雄厚，对疫情的控制也很好。但疫情早期，仍然对申根国开放边境，输入了流动的病毒携带者，影响了疫情的发展进程。

（f）澳大利亚

控制感染源：澳大利亚政府病毒检测能力较好。

切断感染源：未停工停产，民众戴口罩的意识一般，但地广人稀，有自身得天独厚的优势。

保护易感人群：对感染的人群进行了有效隔离。

从图（f）可以看出，澳大利亚政府的疫情防控工作做得很一般，但是得益于自身较好的基础医疗积累和地广人稀的天然优势，新冠肺炎并没有爆炸式传播。

5.4 模型结论

各国防疫成果评价如图 5-7 所示，按 1-5 评级，1 为最差，5 为最好。

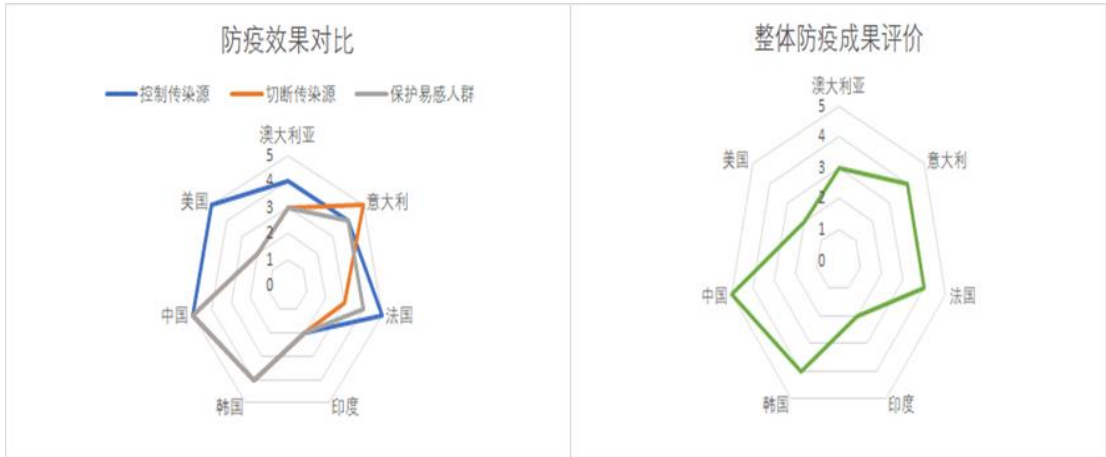


图 5-7 各国防疫成果对比图

从上图可以知道，控制传染源，切断传播途径，保护易感人群是控制疫情的有效手段。检测能力受国力影响很大，老牌强国的检测能力都不错，而印度受制于庞大的贫困人口，心有余而力不足。由于经济原因，绝大多数国家无法像中国，意大利一样，大面积停工停产，但减少不必要聚集活动和戴口罩是切断传播途径

的有效方法。保护易感人群可以极大地减缓感染人数的增加，在医院隔离资源饱和的情况下，方舱医院是有效隔离的最佳选择。疫情防控不仅仅是政府工作，民众配合也十分重要，中国民众自发居家隔离，美国民众不戴口罩且外出聚集，这都极大的影响了疫情发展。各国的实际情况同样说明越及时采取有效措施，防控效果越好。

表 5-2 各国疫情趋势发展

国家\疫情趋势	集中爆发	疫情高峰	预计疫情尾声
美国	3 月上旬	7 月中旬	10 月上旬
中国	1 月下旬	2 月末	4 月中旬
韩国	2 月下旬	5 月上旬	6 月下旬
意大利	3 月上旬	5 月下旬	8 月中旬
法国	3 月中旬	6 月上旬	8 月中旬
澳大利亚	3 月中旬	6 月中旬	8 月中旬
印度	4 月上旬	7 月中旬	12 月上旬

表 5-3 对比了以上 7 个国家从控制传染源、切断传播途径、保护易感人群三方面的防疫效果。

表 5-3 各国防疫效果对比

国家	控制传染源	切断传播途径	保护易感人群	防疫效果
中国	好	好	好	好
美国	好	较差	较差	较差
澳大利亚	较好	一般	一般	一般
法国	好	一般	较好	较好
意大利	较好	好	较好	较好
印度	较差	较差	较差	较差
韩国	好	较好	较好	较好

六、问题三

6.1 模型假设

- (1) 假设本模型使用的是只存在一次疫情高峰的数据。
- (2) 本模型认定出现拐点后，疫情得到控制。
- (3) 所有感染人群都能确诊并有效隔离。
- (4) 不考虑境外输入情况。

6.2 建模分析

本题需要预测复工时间，这要求在传染源与易感人群的接触强度和接触时间的增大条件下，不产生二次大规模传播。想要达成这个目标需要可流动人群中，携带病毒人群数量所占的比例足够小。因为模型假设了感染者都会被确诊并有效隔离，所以本模型将复工时间点设置在疫情到达拐点的一个星期后。疫情到达拐点，说明病毒传播已经被有效控制^[13]。经过上面的计算，经过 7 天后，原本的潜伏者绝大部分也转化为感染者，人群中病毒携带者的比例降到极低。

本题通过中国的实际确诊人数数据，使用逻辑斯蒂模型对疫情发展趋势进行预测，找出疫情拐点，确定复工日期。

6.3 模型建立与求解

建立逻辑斯蒂回归模型

$$P(Y=1|x) = \frac{\exp(\omega \cdot x + b)}{1 + \exp(\omega \cdot x + b)} \quad (6-1)$$

$$P(Y=0|x) = \frac{1}{1 + \exp(\omega \cdot x + b)} \quad (6-2)$$

$x \in R^n$ 是输入, P 是输出, $\omega \in R^n$ 和 $b \in R^1$ 是参数, ω 是权值向量, b 是偏置, $\omega \cdot x$ 是 ω 和 x 的内积。逻辑斯蒂回归比较两个条件概率值的大小, 将 x 分到概率值较大的那一类^[14]。

如果事件发生的概率是 P , 那么该事件的几率是 $\frac{P}{1-P}$

该事件的对数几率是

$$\text{logit}(P) = \log \frac{P}{1-P} \quad (6-3)$$

对逻辑回归来说

$$\log \frac{P(Y=1|x)}{1-P(Y=0|x)} = \omega \cdot x \quad (6-4)$$

就是说, $Y=1$ 的对数几率是由输入 x 的线性函数表示的概率, 即逻辑斯蒂模型^{[14][15][16]}。

对本题建立逻辑斯蒂模型

$$P(t) = \frac{P_0 e^{rt}}{1 - P_0(e^{rt} - 1)} \quad (6-5)$$

$P(t)$ 为总人数函数, r 为增长阻力 (政府管理、民众意识)。

运用 python 对逻辑斯蒂模型进行训练, 得出如图 6-1 的所示的逻辑拟合图像:

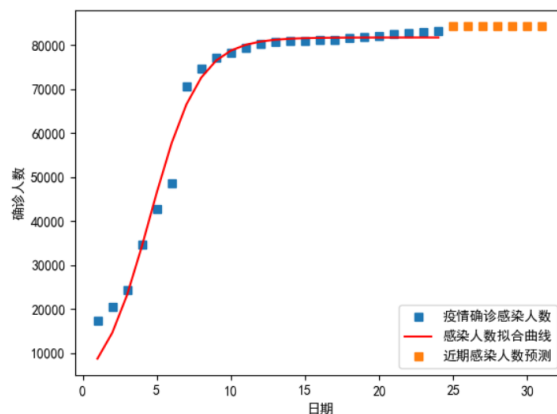


图 6-1 逻辑拟合曲线

本组训练数据使用 2 月 3 日作为日期坐标轴的零点。本模型要求政府的管控力度系数达到 0.6，即全程保持高度警惕。从图像中可以发现，在首日的 10 天后（即 2 月 13 日），疫情增速明显放缓，到达拐点。根据模型的设定，7 天后（即 2 月 20 日）为“重启”日期。本逻辑拟合曲线对复工之后的感染人数进行了预测，20 号之后感染人数只有零星增长，不具备二次爆发的条件。

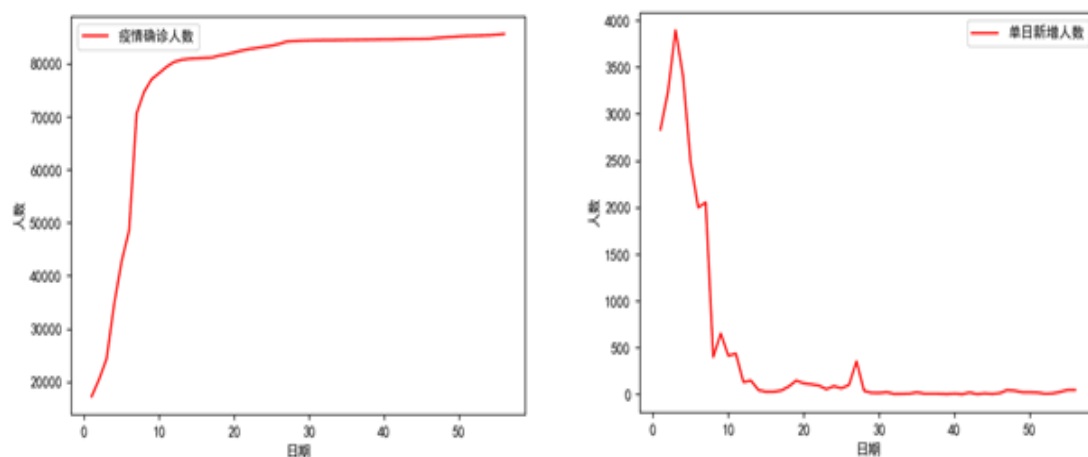


图 6-2 中国实际总确诊人数和单日新增趋势

从上图的中国疫情实际趋势图可以得知，大致在首日的 12 天后（即 2 月 15 日）确诊人数总数不再急速增长且单日新增维持在了一个较低的水平。

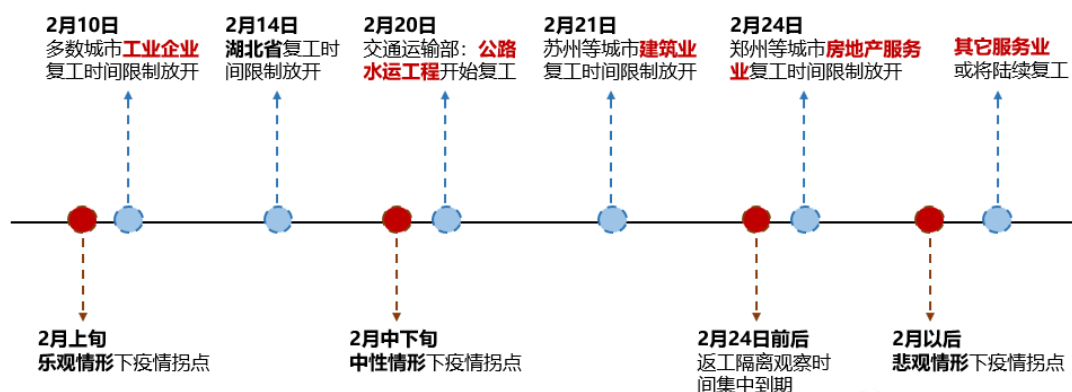


图 6-3 中国实际的“重启”时间轴

中国的“重启”策略是分地区分批重启：传染风险低的地区先“重启”，风险高的地方后“重启”。维持重要民生的工业企业首批重启，保证人民生活质量；在稳步控制住疫情后，恢复交通运输，放开人员流动；在确保不会二次爆发后，服务业复工。

6.4 模型结论

本题的模型由中国的确诊人数数据预测得到。

整个模型的前提是，中国政府的管控力度系数为 0.6，全程对疫情保持高度警惕。

“重启”的条件是在传染源与易感人群的接触强度和接触时间的增大条件下，不产生二次大规模传播。通过模型拟合，给出的“重启”时间节点是到达拐点后的 7 天。模型预测的拐点日期为 2 月 13 日前后，即建议在 2 月 20 日开始“重启”。

“重启”的策略的是分地区分批重启。首先在传染风险低的地区“重启”重要民生工业；一个潜伏周期后“重启”城市间的交通运输，开始人员流动；在经过一个潜伏周期后分批“重启”服务业，逐步复工复产。

对于顺利实现预测的“重启”进程有如下建议：

- (1)保持警惕，尽最大力度切断病毒传播，对感染人员有效隔离。
- (2)民众需要坚持戴口罩，保持有效的社交距离，减少不必要的社交。
- (3)对火车站，机场等人流量大的地方做好人员检测工作，及时阻断病毒携带者的流动。
- (4)民众要保持良好的心态，不造谣，不传谣，共同维护稳定的社会治安。

七. 模型总结与评价

1. 从官方新闻中得知，此次疫情的潜伏期以天为单位，是离散型计量数据，且潜伏期少有成指数模式增长的情况发生。因此，本题采用中位数法求潜伏期天数均值。各国实际的防疫情况不同会导致结果受到人为因素的影响，比如检测能力较弱的国家，推算的潜伏期长度较长。
2. 使用 SEIR 模型，可以充分考虑潜伏群体（已感染未发病）的传染能力。可以通过四种状态转化比例之间的关系，求得传染系数和基本传染数，推演病毒的传染强度随时间的变化。由于实际情况远比想象复杂，所以该模型只能显示一个大致的走向。
3. 回溯传播模型可以通过调节参数，对疫情发展进行很好的模拟。通过与实际趋势的比较，能够客观评价出各国的防疫成果。但具体影响因素是方方面面的，本文只考虑了主要的影响因素。
4. 对逻辑回归模型进行训练，预测了中国未来的疫情走向，估算了比较合理的“重启”条件和时间。但真的要实现“安全”重启，需要全国上下共同努力，不能因为认为疫情得到控制就放松警惕。

参考文献

- [1] 范如国, 王奕博, 罗明, 张应青, 朱超平, 基于 SEIR 的新冠肺炎传播模型及拐点预测分析[J], 电子科技大学学报, 49(03): 369-374, 2020。
- [2] 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), 49(02): 178-184, 2020。
- [3] 王园园, 具有时滞的 SEIR 的肺结核模型研究[D], 西安科技大学, 2012。
- [4] 王茜, 王婷婷, SIRS 传染病模型的稳定性分析[J], 首都师范大学学报: 自然科学版, 37(2): 5-11, 2016。
- [5] 屈小博, 程杰, 新冠肺炎疫情对劳动力市场的影响及政策反应[J], 河北师范大学学报(哲学社会科学版), 43(04): 126-133, 2020。
- [6] 叶振宇, 全球新冠肺炎疫情对我国区域经济的影响与应对[J], 河北师范大学学报(哲学社会科学版), 43(04): 134-140, 2020。
- [7] 张磊, 欧盟应对新冠肺炎疫情机制及其局限[J], 国际论坛, (04): 116-132+159-160, 2020。
- [8] 田斌群, 陈道远, 田卫群, 王行环, 刘同族, 张银高, 郑新民, 李世文, 一个预测新型冠状病毒肺炎流行趋势的简易模型[J/OL], 武汉大学学报(医学版): 1-4, 2020。
- [9] Huang NordenE, Qiao Fangli, A data driven time-dependent transmission rate for tracking an epidemic: a case study of 2019-nCoV[J], Pubmed, 65(6): 1-3, 2020。
- [10] 苏莉莉, 李维德, 杨爱玲, 陈万宝, 一类乙肝病毒传染动力学模型的分析与应用[J], 兰州大学学报(自然科学版), 54(03): 410-416+423, 2018。
- [11] 李建全, 娄洁, 娄梅枝, 离散的 SI 和 SIS 传染病模型的研究[J], 应用数学和力学, (01): 104-110, 2008。
- [12] Delamater Paul L, Street Erica J, Leslie Timothy F, Yang Y Tony, Jacobsen Kathryn H, Complexity of the Basic Reproduction Number(R_0)[J], Pubmed, 25(1): 1-4, 2019。韩晓娜, SARS 流行病学传播动力模型研究[D], 中国人民解放军军事医学科学院, 2006。
- [13] 李建全, 马知恩, 两类带有确定潜伏期的 SEIS 传染病模型的分析[J], 系统科学与数学, (02): 228-236, 2006。
- [14] 曾哲淳, 赵冬, 李岩, 郭强, 石鹏, 李哲, 印惠俊, 李洋, 应用系统动力学模型对 SARS 疫情传播及主要防控措施效果的计算机模拟仿真研究[J], 中华流行病学杂志, (03): 11-15, 2005。
- [15] Chen Y C, Lu P E, Chang C S, et al, A Time-dependent SIR model for COVID-19 with Undetectable Infected Persons[J], 2020。
- [16] 严阅, 陈瑜, 刘可伋, 等, 基于一类时滞动力学系统对新型冠状病毒肺炎疫情的建模和预测[J], 中国科学: 数学, (1): 1-12, 2020。