

# 对新冠肺炎的 简要分析



2020/4/12

——数据分析大报告

包含肺炎患者去世对新增感染人群影响显著问题、针对全球不同地区（部分主要疫情国家）的痊愈人数的单因素方差分析

小组成员：宋卡妮、李金哲、赵启航

# 目录

引言 .....	2
一、准备阶段 .....	2
1.1 选题原因 .....	2
1.2 数据来源 .....	3
二、源数据基本分析 .....	3
2.1 数据集频数特点: .....	3
2.1.1 不同时期, 各个国家(地区)新增感染人数占全球总感染数的比重 .....	3
2.1.1 不同时期, 各个国家(地区)新增死亡人数占全球总感染数的比重 .....	4
2.2 世界主要地区基本统计(12/31~4/12) .....	5
2.2.1 主要疫情地区从12月31号至4月12号每日新增确诊人数 .....	6
2.2.2 主要疫情地区从12月31号至4月12号每日新增死亡人数 .....	6
2.2.3 问题分析: .....	7
2.3 国际不同地区医疗情况数据特征 .....	8
2.3.1 确诊与死亡人数数据特征(e.g. 华盛顿) .....	8
2.3.2 入院人数数据特征(全球主要疫情国家) .....	8
2.4 缺失值、异常值的分析与处理 .....	9
2.4.1 异常值筛选及处理 .....	9
2.4.2 缺省值分析及处理 .....	11
三、问题探究一: 医疗设备数目是否对死亡人数产生显著影响 .....	12
3.1 分析方法: 单因素方差分析法 .....	12
3.1.1 单因素方差分析步骤: .....	12
3.1.2 单因素方差分析数据: .....	12
3.2 单因素分析结果 .....	12
3.3 方差齐性检验 .....	13
3.4 多重比较检验 .....	13
3.5 趋势检验 .....	13
3.6 分析方法: 多因素方差分析法 .....	14
3.6.1 多因素方差分析步骤: .....	14
3.6.2 多因素方差分析数据: .....	14
3.7 多因素分析结果: .....	15
四、问题探究二: 新增死亡人数对新增感染人数是否产生影响 .....	15
4.1 问题分析 .....	15
4.2 问题提出 .....	16
4.3 方差齐性检验 .....	16

4.4 非参数检验 .....17

4.5 趋势检验.....17

4.6 进一步分析（英国） .....18

五、问题探究三：对医疗需求的趋势分析 .....19

5.1 时间序列分析简述 .....19

5.2 分析步骤.....20

5.2.1 专家建模器.....23

5.2.2 指数平滑.....23

5.3 分析结果及结论.....24

六、报告总结： .....25

七、附录.....25

7.1 新增感染比例图生成代码 .....26

7.2 新增死亡比例图生成代码.....27

引言

为更好地防控新型冠状病毒(2019-nCoV)传染病,保障医疗机构医务人员及其所需资源应是重中之重。本文出发点主要针对全球范围内主要疫情地区的医疗情况简要分析，以期政府、组织有区别的资源倾斜，同时收治新冠肺炎患者的医疗机构内逝世人数与新增人数间是否有一定的联系；给疫情期间医务人员提供可能性的想法，合理有效利用防护用品的同时也涵盖不同程度的患者收治。

一、准备阶段

1.1 选题原因

2019 年底新型冠状病毒肺炎（COVID-19，简称“新冠肺炎”）疫情暴发并呈全球扩散态势，在此背景下，世界卫生组织（WHO）于 2020 年 3 月 11 日宣布将其上调至“大流行病”级别。新冠肺炎疫情的暴发和传播，既是全球化的体现，也是对全球化的考验。全球卫生外交从国际合作的角度看待和解决卫生问题，将人的安全与健康放在首位。

基于以上大背景，我们小组选择针对全球主要疫情地区的医疗需求及患者收治数据进行更进一步的分析。

## 1.2 数据来源

数据集一：Hospitalization

来源：世界银行数据库

其中包含截止到 2020 年 4 月 12 日的全球大部分国家以及主要地区的医疗资源缺少数。含有每天需要的病床数、所需 ICU 数、所需呼吸机数、入院人数、确诊人数、死亡人数。

数据集二：[http://fms.news.cn/swf/2020\\_sjxw/2\\_1\\_xgyq/index.html?v=](http://fms.news.cn/swf/2020_sjxw/2_1_xgyq/index.html?v=0.42056890891670506)

0.42056890891670506

来源：新华网

包含从 2019/12/31 到 2020/4/12 世界国家（除中国）每日确诊人数，每日死亡人数，累计确诊人数，累计死亡人数。

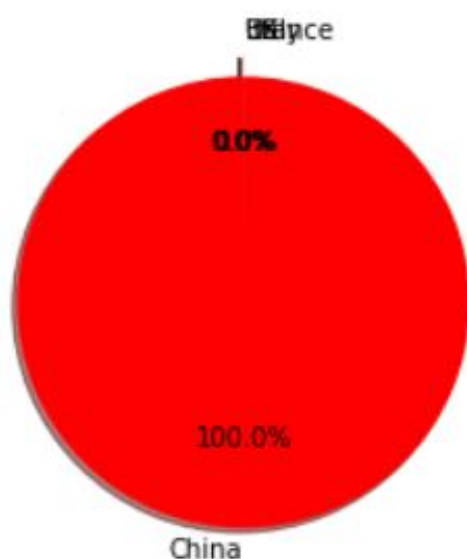
## 二、源数据基本分析

### 2.1 数据集频数特点：

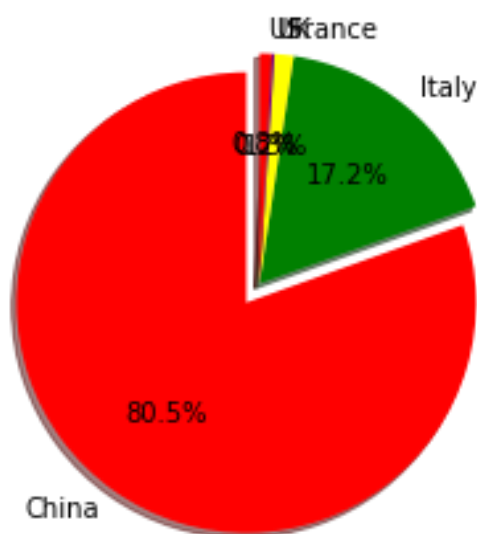
#### 2.1.1 不同时期，各个国家（地区）新增感染人数占全球总感染数的比重

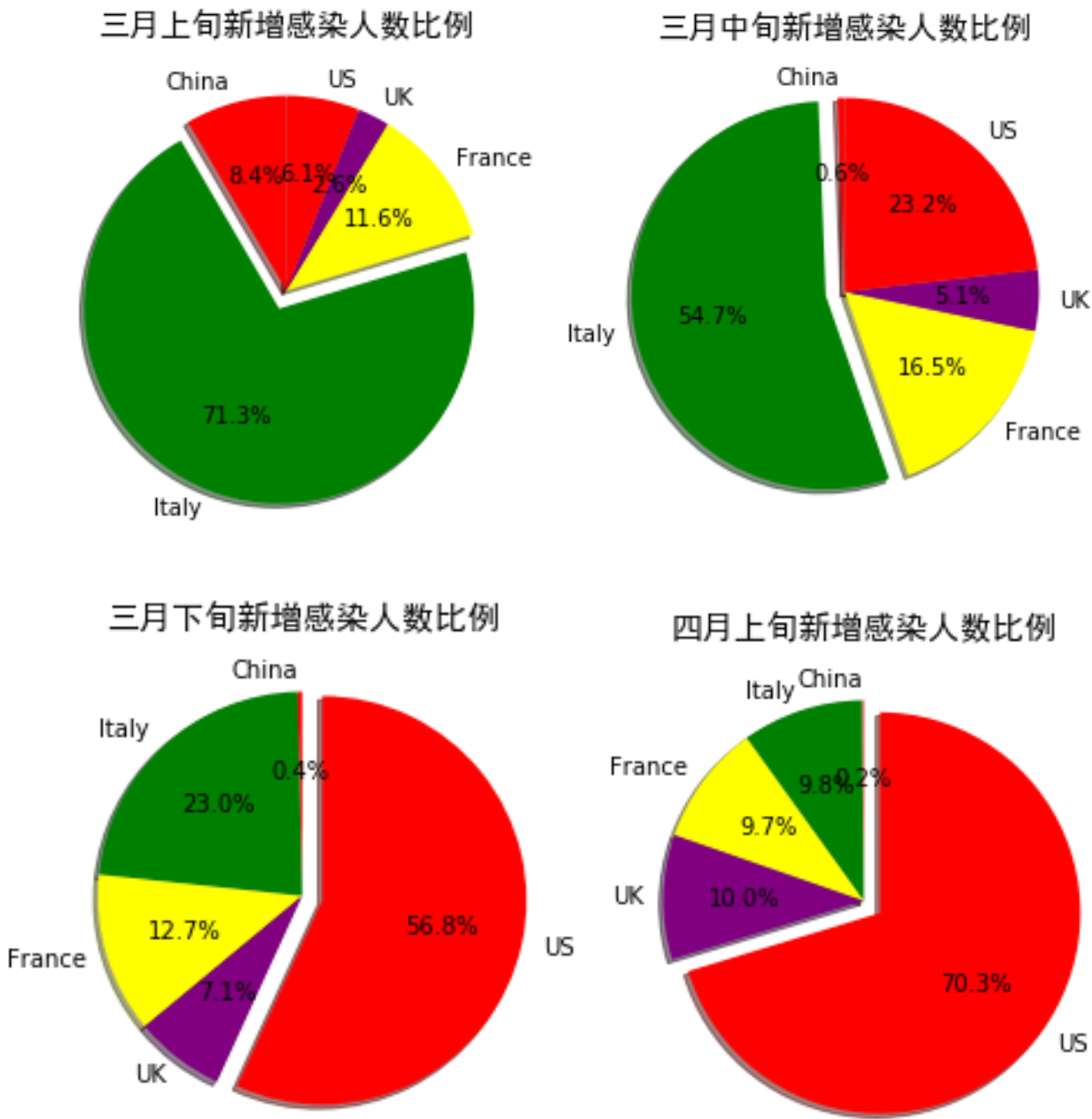
【生成图例的代码详见附录】

二月中旬新增感染人数比例

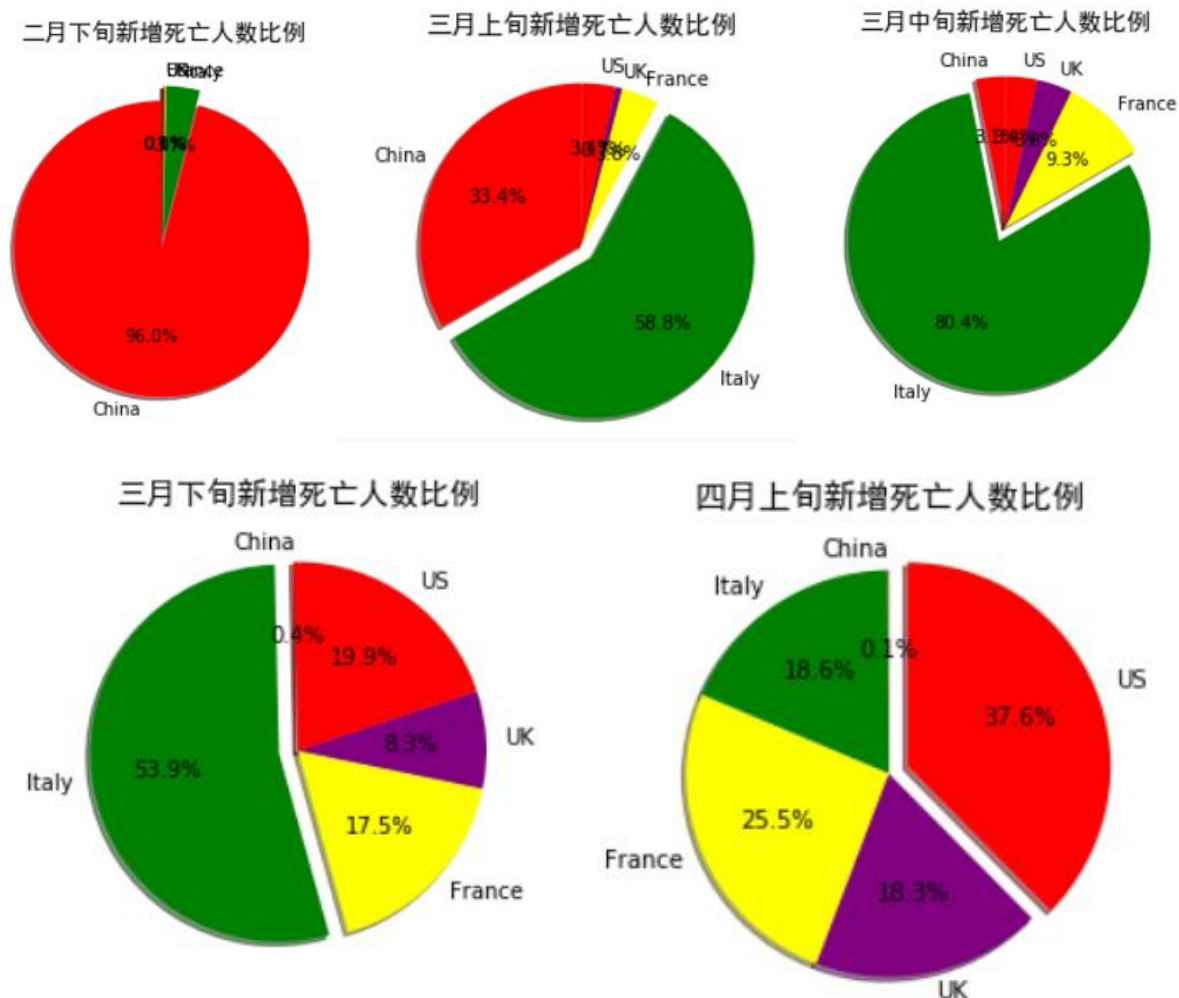


二月下旬新增感染人数比例





2.1.1 不同时期，各个国家（地区）新增死亡人数占全球总感染数的比重



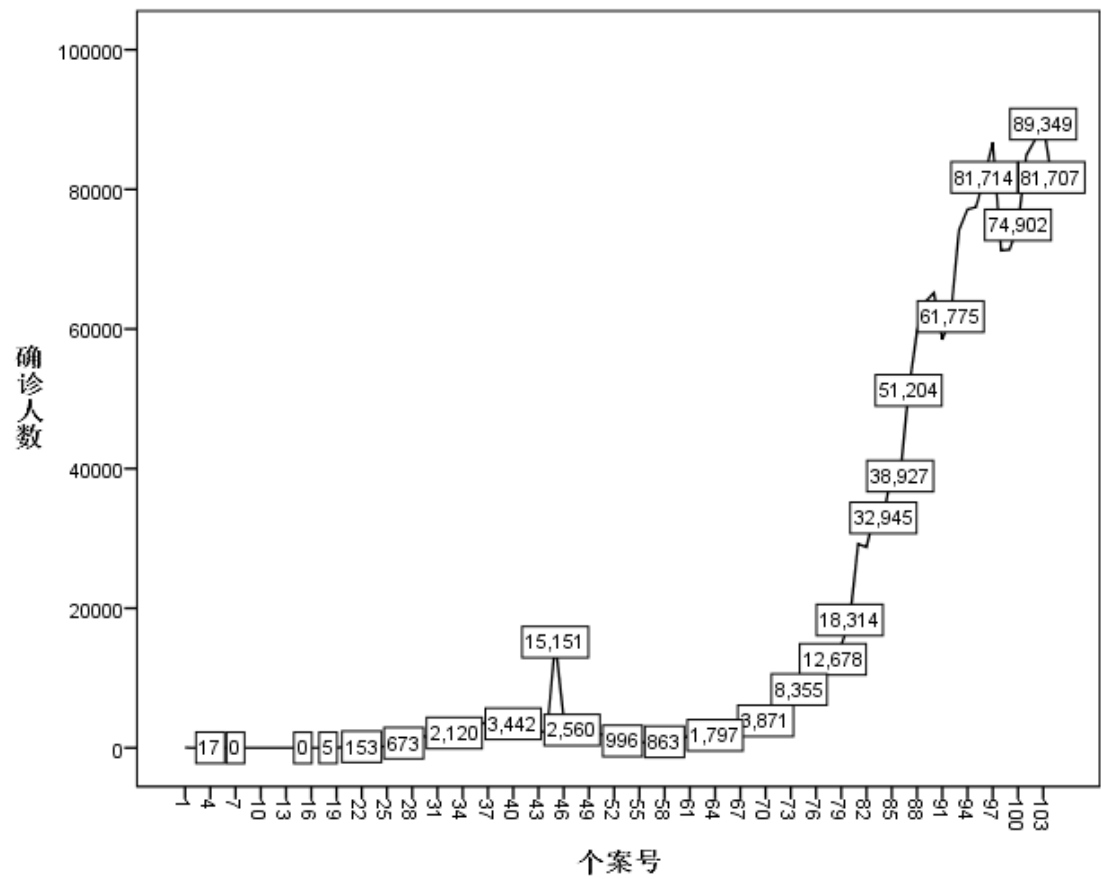
### 统计结果分析:

以二月中旬为界限，在此之前仅中国发现新冠病例，在当年二月下半个月，在欧洲出现病例，且以意大利为欧洲主要疫情来源国，截止二月底欧洲发病人数占全球总人数的 17.3%。3 月起意大利、法国、英国等欧洲国家疫情爆发，新增发病数为 71.3%、11.6%、2.6%，且此比率除意大利外在当月仍在上升，说明英国、法国等疫情严重恶化。另外，美国也在此时疫情爆发。至 3 月底，主要疫情国已由中国绝对性的转移为美国、意大利，而意大利新增感染数比率持续下降，一是可能由于意大利防控初见成效，二是由于美国疫情严重恶化，感染人数急剧上升。

而 3 月上下的新增死亡数比率和新感染人数变化基本一致。其中，中国的死亡人数比率下降趋势比新增感染更明显。除此之外，美国的死亡人数上升不如新增确诊数上升的迅速。

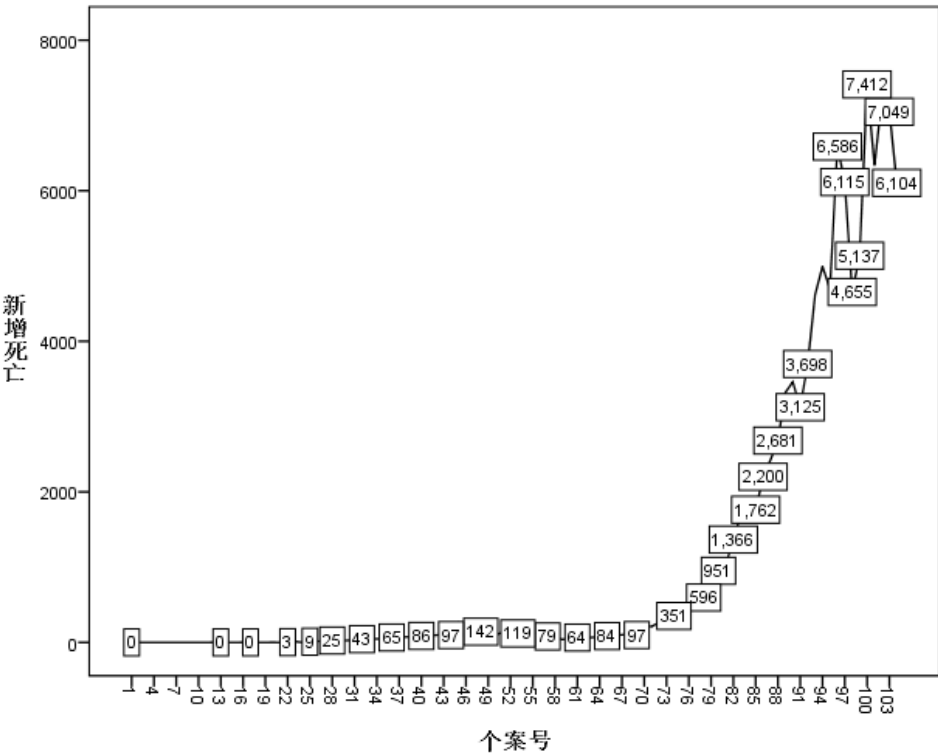
## 2.2 世界主要地区基本统计 (12/31~4/12)

2.2.1 主要疫情地区从 12 月 31 号至 4 月 12 号每日新增确诊人数



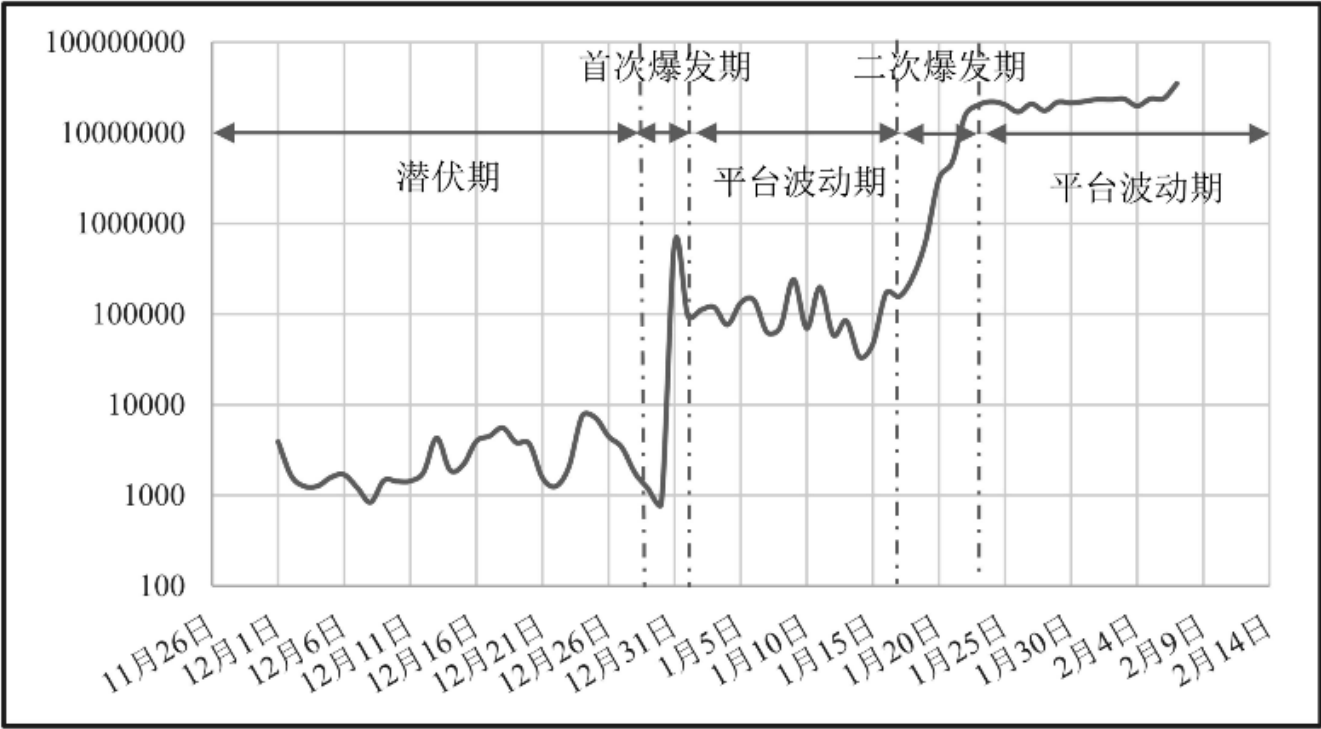
2.2.2 主要疫情地区从 12 月 31 号至 4 月 12 号每日新增死亡人数





2.2.3 问题分析：

下图为中国防控时期图，作为最早疫情爆发区以中国防控情况作为对照组进行分析



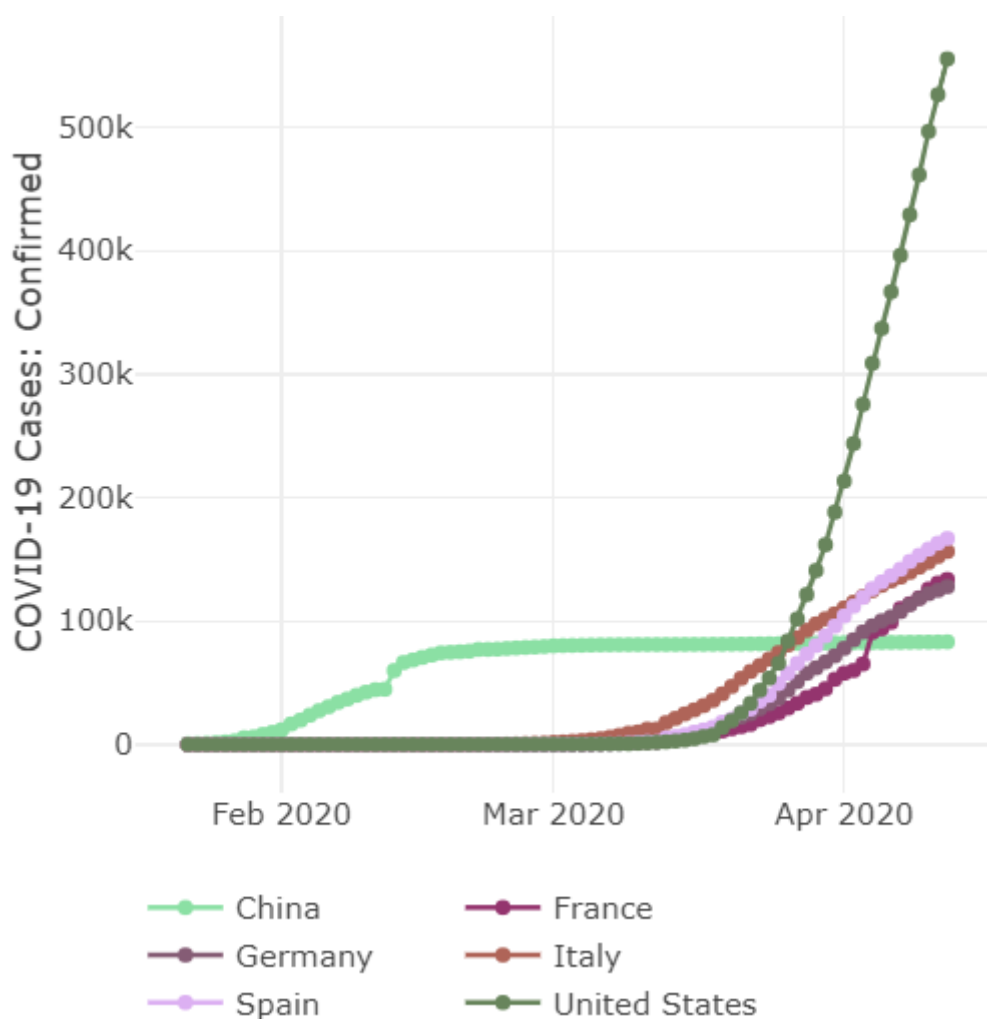


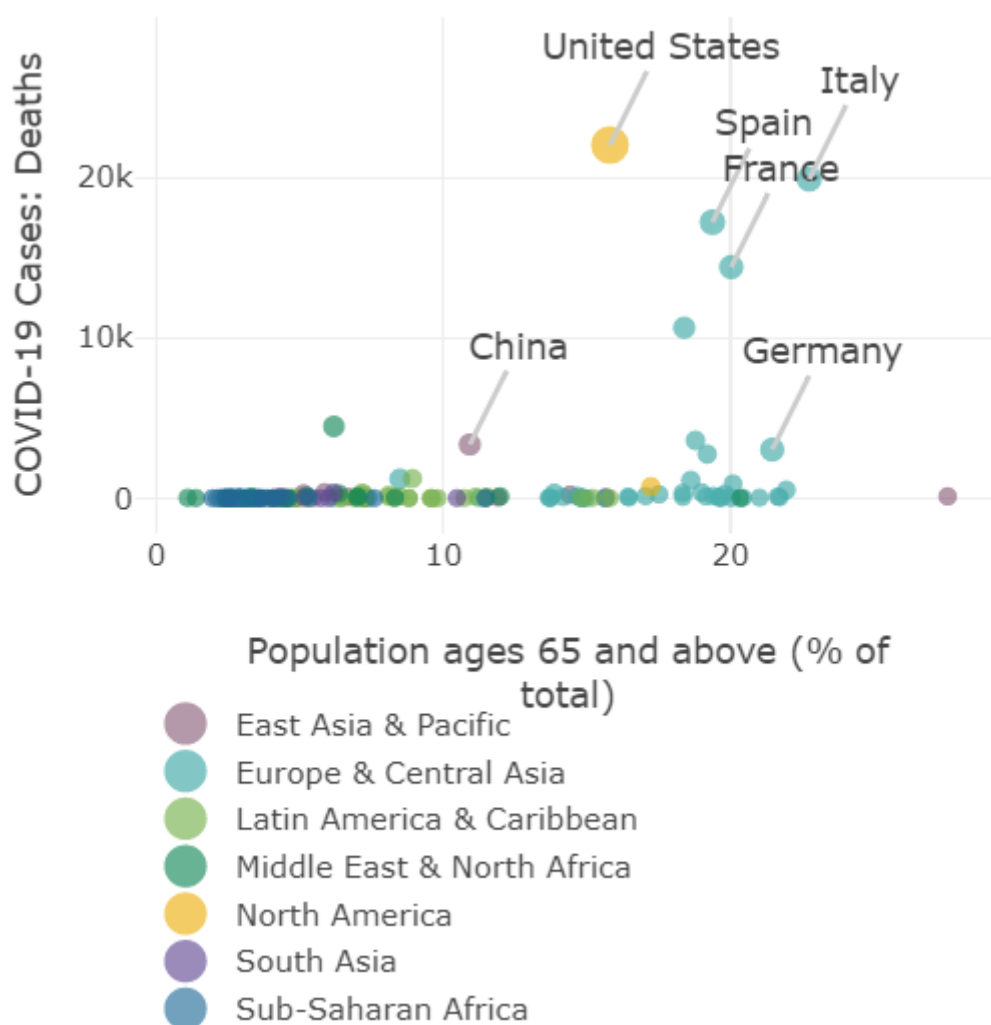
可以看出全球主要疫情地区地区在近日迎来确诊患者数量的爆发，由于急剧上升前新增确诊与新增死亡人数位于极低的水平，而且爆发后至今未显示出一个有效的下跌趋势，所以参照中国疫情状态图，基本判断为处于首次爆发期。

## 2.3 国际不同地区医疗情况数据特征

### 2.3.1 确诊与死亡人数数据特征(e.g. 华盛顿)

### 2.3.2 入院人数数据特征（全球主要疫情国家）

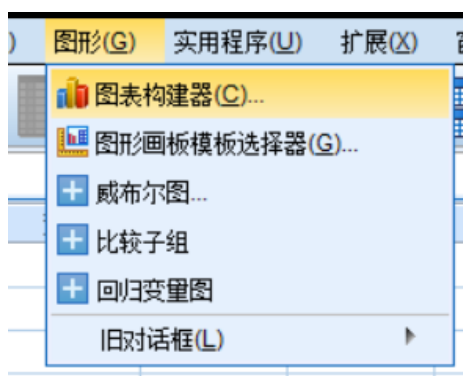


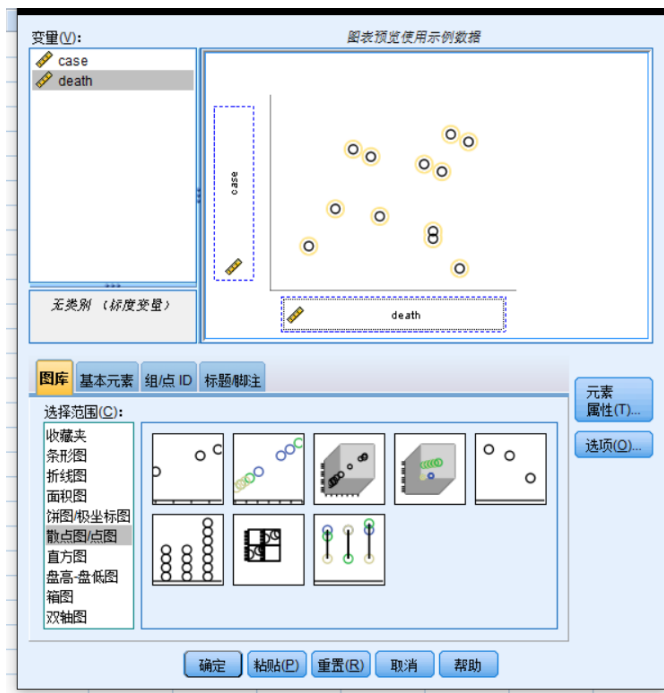


## 2.4 缺失值、异常值的分析与处理

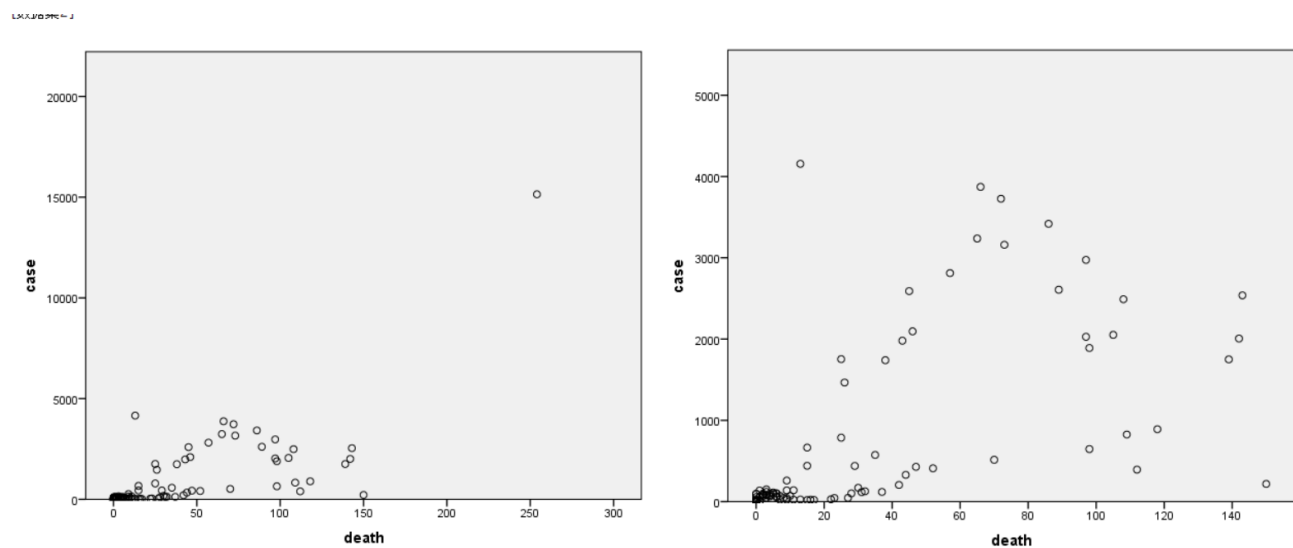
### 2.4.1 异常值筛选及处理

由于中国是最早开展新型冠状病毒肺炎防疫的国家，因此结果可以作为对照样本。根据中国的新增死亡人数和新增确诊人数构建散点图，判断数据是否具有异常点



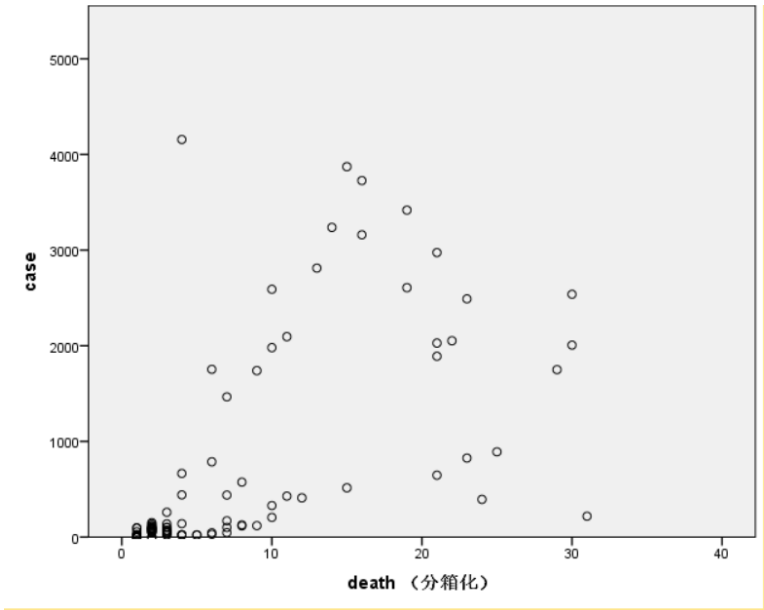


根据下图（左）可以明显看出（254，15141）该点为异常点，可以剔除（右）

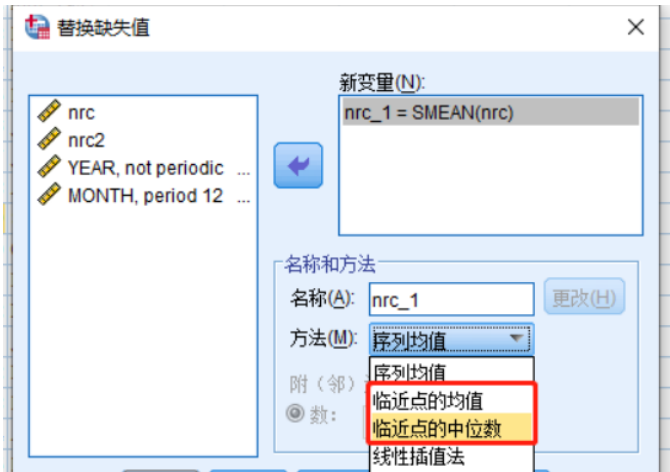


由于数据具有明显的离散化趋势，故将每日新增死亡人数进行可视化分箱操作，操作如下，处理后结果如下





2.4.2 缺省值分析及处理



### 三、问题探究一：医疗设备数目是否对死亡人数产生显著影响

#### 3.1 分析方法：单因素方差分析法

研究一个控制变量的不同水平是否对观测变量产生了显著影响。

##### 3.1.1 单因素方差分析步骤：

1、明确观测变量和控制变量。

2、剖析观测变量的方差。方差分析认为：观测变量值的变动会受控制变量和随机变量两方面的影响。据此，单因素方差分析将观测变量总的离差平方和分解为组间离差平方和和组内离差平方和两部分，用数学形式表述为： $SST=SSA+SSE$ 。

3、通过比较观测变量总离差平方和各部分所占的比例，推断控制变量是否给观测变量带来了显著影响。

##### 3.1.2 单因素方差分析数据：

以 Wisconsin 城市每日呼吸机需求数目与每日死亡人数为例，分析呼吸机数目是否对死亡人数产生了显著影响。

#### 3.2 单因素分析结果

**ANOVA**

死亡数每天

	平方和	自由度	均方	F	显著性
组间	1316.551	9	146.283	19.070	.000
组内	299.166	39	7.671		
总计	1615.717	48			

上表是呼吸机数量对死亡人数的单因素方差分析结果。可以看出：观测变量每日死亡人数的离差平方总和为 **1615.717**。如果仅考虑呼吸机数量单个因素的影响，则每日死亡人数总变差中，不同呼吸机数量可解释的变差为 **1316.551**，抽样误差引起的变差为 **299.166**，组间方差和组内方差分别为 **146.283** 和 **7.671**，两者相除所得的 F 统计量的观测值为 **19.070**，对应的概率 P-值近似为 **0**。

假定显著性水平  $\alpha$  为 **0.05**，由于概率 P-值小于显著性水平  $\alpha$ ，应拒绝原假设，认为不同呼吸机数目对死亡人数产生了显著影响，呼吸机数目不同时死亡人数有显著差异。

3.3 方差齐性检验

死亡数每天

方差齐性检验			
莱文统计	自由度 1	自由度 2	显著性
3.194	9	39	.005

概率 P 值等于 0.005，小于 0.05，拒绝原假设，认为各水平下观测值的方差分布不同，单因素方差分析无意义。下面的多重比较检验使用方差不等的塔姆黑尼方法。

3.4 多重比较检验

事后检验

多重比较

因变量: 死亡数每天

塔姆黑尼

(I) 呼吸机数需求数	(J) 呼吸机数需求数	平均值差值 (I-J)	标准误差	显著性	95% 置信区间	
					下限	上限
10.00	20.00	-.66275	.57071	1.000	-3.0526	1.7271
	30.00	-2.22500 <sup>*</sup>	.36846	.007	-3.9111	-.5389
	40.00	-2.32225	1.05107	.994	-11.9283	7.4838
	60.00	-4.60075	1.38519	.825	-18.9722	9.7707
	80.00	-7.78142	1.73148	.216	-18.6781	3.1153
	90.00	-12.49395	1.96116	.109	-27.8324	2.8445
	100.00	-11.76300	1.77114	.222	-31.3876	7.8616
	110.00	-11.99050	2.01298	.307	-34.8395	10.8585
20.00	120.00	-15.45125	1.00110	.633	-179.5049	148.6024
	10.00	.66275	.57071	1.000	-1.7271	3.0526
	30.00	-1.56225	.49499	.429	-3.9318	.8073
	40.00	-1.55950	1.10182	1.000	-9.9105	6.7915
	60.00	-3.93800	1.42408	.920	-16.7697	8.8937
	80.00	-7.11867	1.76274	.285	-17.6608	3.4234
	90.00	-11.83120	1.98882	.118	-26.5439	2.8815
	100.00	-11.10025	1.80172	.227	-29.2280	7.0275
	110.00	-11.32775	2.03994	.323	-32.7521	10.0966
	120.00	-14.78850	1.05426	.453	-88.5174	58.9404

观察显著性那一列，如果概率 p 值小于 0.05，就拒绝原假设，认为二者有显著性差异，从结果看呼吸机数量不同时每日死亡数绝大部分没有显著性差异，而在我们的认知里呼吸机对治疗新冠肺炎，控制疫情有重要的作用，猜测这个结论与我们的认知不同的原因是呼吸机每日需求量只能代表当日的重症病人，而每日的死亡人数可能是前几天的重症病人，呼吸机需求量并不能与死亡人数成正比。

3.5 趋势检验

ANOVA						
死亡数每天						
		平方和	自由度	均方	F	显著性
组间	(组合)	1316.551	9	146.283	19.070	.000
	线性项 加权	1258.535	1	1258.535	164.066	.000
	偏差	58.016	8	7.252	.945	.491
	二次项 加权	9.534	1	9.534	1.243	.272
	偏差	48.483	7	6.926	.903	.514
组内		299.166	39	7.671		
总计		1615.717	48			

为了验证上述的猜测，我们对死亡数和呼吸机数做了趋势检验，从概率 P 值看出，二次项的概率 P 值大于 0.05，呼吸机数目与死亡数目符合二次项的分布，并不是线性分布，上述猜测正确。

3.6 分析方法：多因素方差分析法

3.6.1 多因素方差分析步骤：

- 1、提出原假设。
- 2、选择检验统计量。
- 3、计算检验统计量的观测值和概率 P 值
- 4、给定显著性水平 $\alpha$ ，并作出决策

3.6.2 多因素方差分析数据：

以 Wisconsi 城市每日呼吸机需求数目、病床需求数、ICU 需求数与每日死亡人数为例，分析三者是否对死亡人数产生了显著影响。





3.7 多因素分析结果：

主体间效应检验

因变量：死亡数每天

源	III 类平方和	自由度	均方	F	显著性
修正模型	1467.663 <sup>a</sup>	26	56.449	8.388	.000
截距	1531.314	1	1531.314	227.544	.000
病床需求数	111.403	7	15.915	2.365	.058
呼吸机数需求数	102.917	7	14.702	2.185	.076
ICU需求数	66.170	9	7.352	1.092	.407
误差	148.055	22	6.730		
总计	3742.732	49			
修正后总计	1615.717	48			

a. R 方 = .908 (调整后 R 方 = .800)

从概率  $p$  值看出三者并没有对死亡人数造成显著差异，与单因素方差分析的结论相同，同样猜测是因为每日死亡人数与若干天前的病床需求数，呼吸机数，ICU 数有关，医疗设备缺乏导致病情恶化，几天后死亡。

四、问题探究二：新增死亡人数对新增感染人数是否产生影响

4.1 问题分析

假设新增感染人数仅与新增死亡人数相关，

由于数据呈现配对关系，且新增死亡人数拆分的数据组别数目为 30。故考虑  $t$  检验和方差检验。

正态性检验 <sup>c,d,e,f,g,h,i,j</sup>						
		柯尔莫戈洛夫-斯米诺夫 <sup>a</sup>			夏皮洛-威尔克	
death (分箱化)		统计	自由度	显著性	统计	自由度
case	1	.289	21	.000	.619	21
	2	.117	19	.200 <sup>*</sup>	.958	19
	3	.241	12	.053	.794	12
	4	.388	7	.002	.587	7
	5	.260	2	.		
	6	.273	4	.	.861	4
	7	.304	5	.148	.753	5
	8	.378	3	.	.766	3
	9	.260	2	.		
	10	.286	4	.	.859	4
	11	.260	2	.		
	15	.260	2	.		
	16	.260	2	.		
	19	.260	2	.		
	21	.252	4	.	.961	4
	23	.260	2	.		
	30	.260	2	.		

\* 临界显著性的下限

经过正态检验分析，可以明显的看出存在少量新增死亡人数组别的概率  $p$  值小于  $0.05$ ，即拒绝接受原假设，因变量与正态分布存在明显差异，与  $t$  检验成立条件‘来自正态分布总体’不符，因此使用方差检验。

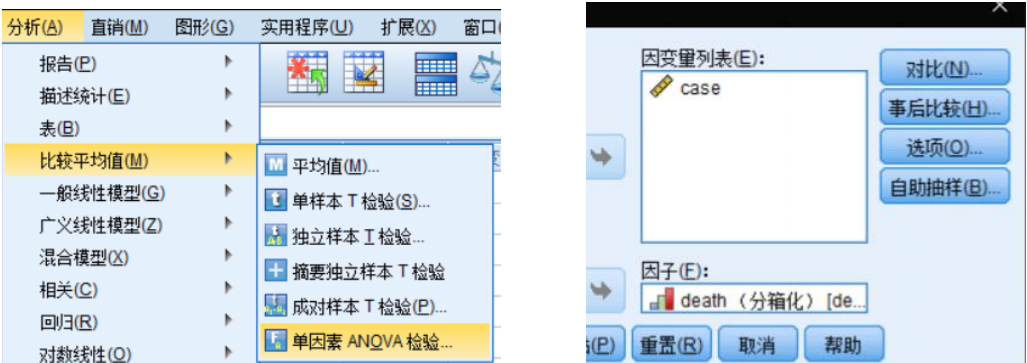
4.2 问题提出

由于方差检验是研究单个因素对观测变量的影响，故在此将问题锁定为不同的新增死亡人数是否对新增感染人数起到显著性的影响。

提出假设

$H_0$ : 新增死亡人数对新增感染人数无影响

$H_1$ : 新增死亡人数对新增感染人数有影响



结果如下

ANOVA					
case					
	平方和	自由度	均方	F	显著性
组间	82336510.14	24	3430687.922	7.728	.000
组内	34624466.27	78	443903.414		
总计	116960976.4	102			

由方差检验的结果，可以明显得到新增感染人数与新增死亡人数的相关性概值十分小，可以约等于  $0$ ；而且远小于接受  $H_0$  的概率  $\alpha=0.05$ 。所以拒接原假  $H_0$ ，接受假  $H_1$ 。即新增死亡人数对新增感染人数有影响

4.3 方差齐性检验

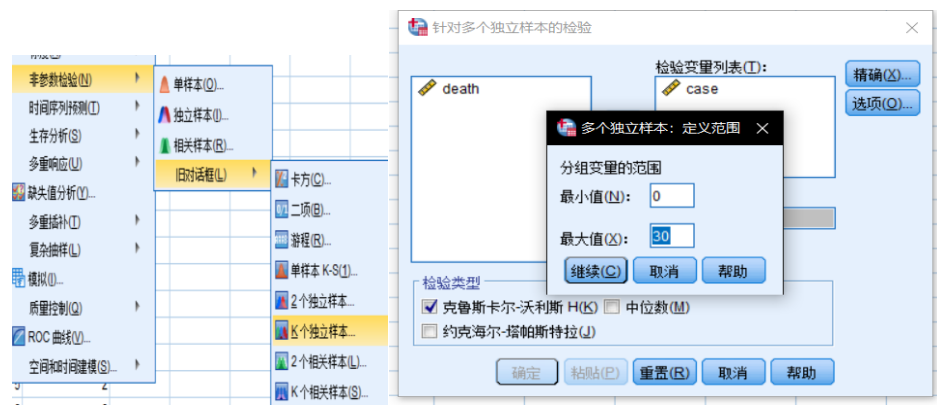
方差齐性检验

case			
莱文统计	自由度 1	自由度 2	显著性
7.818	16	78	.000

概率 P 值等于 0.005，小于 0.05，拒绝原假设，认为因素非等方差。使用非参数检验，进行分析。

4.4 非参数检验

- 提出假设
- H0: 不同新增死亡人数，新增感染人数是相同的
- H1: 不同新增死亡人数，新增感染人数是不同的



结果如下

检验统计 <sup>a,b</sup>	
	case
卡方	75.496
自由度	23
渐近显著性	.000
a. 克鲁斯卡尔-沃利斯检验	
b. 分组变量: death (分箱化)	

由于渐进显著性近似为 0.000，且  $0.000 < 0.01$  所以拒绝原假设 H0，得出结论：H1 不同新增死亡人数，新增感染人数是不同的

4.5 趋势检验

ANOVA						
case						
		平方和	自由度	均方	F	显著性
组间	(组合)	82336510.14	24	3430687.922	7.728	.000
	线性项	45748150.40	1	45748150.40	103.059	.000
	加权偏差	36588359.74	23	1590798.249	3.584	.000
组内		34624466.27	78	443903.414		
总计		116960976.4	102			

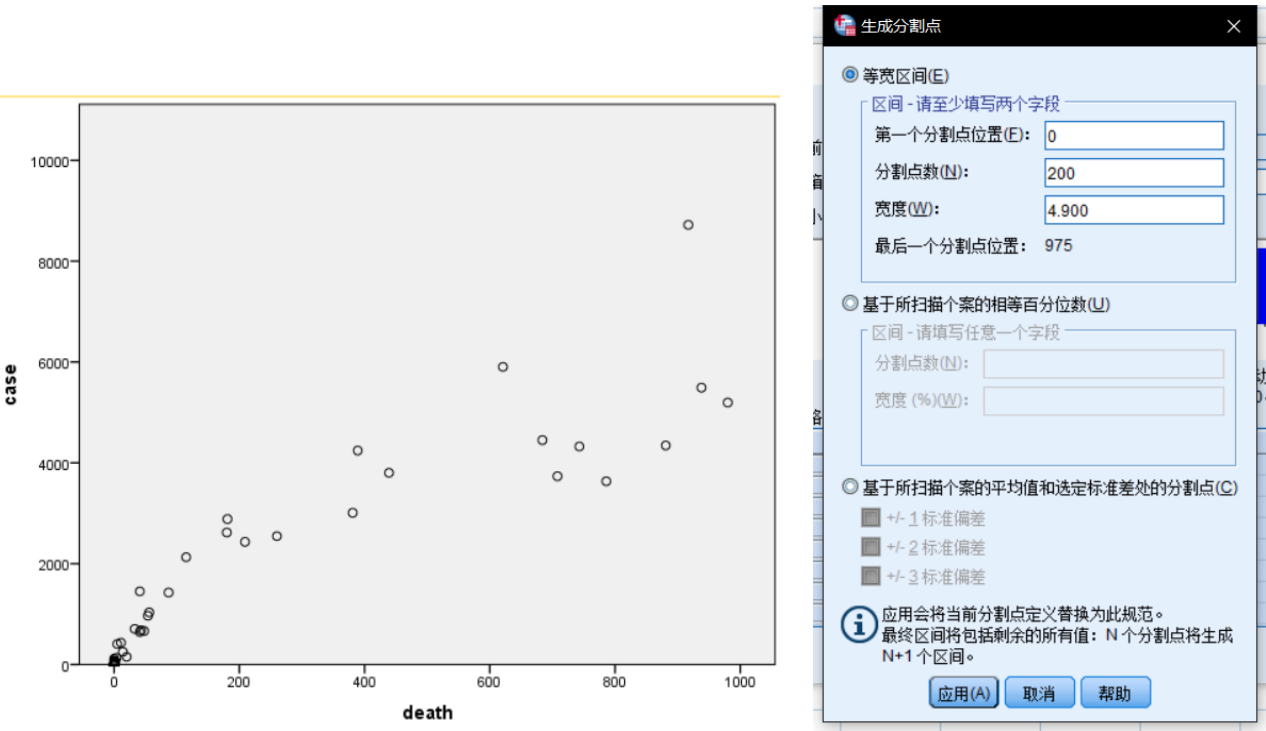
为了验证上述的猜测，我们对新增死亡人数和新增感染人数进行趋势检验，从概率 **P** 值看出，线性项的概率 **P** 值小于 **0.05**，即新增死亡人数与新增感染人数符合线性项的分布，上述猜测正确。

4.6 进一步分析（英国）

由于当前数据为仅为中国的，在其基础上再进行其他国家的数据分析，以保证结论的可靠性；若结论存在差异，则分析原因

【下面以英国进行分析】

由数据散点图可以明显的看出数据具有两极分化的趋势，具有一次函数的趋势。下面进行数据的拆分，为保证数据间隔相同，故将宽度控制在 **5** 左右。



同理进行方差分析，

ANOVA					
case	平方和	自由度	均方	F	显著性
组间	274140474.0	25	10965618.96	1714.106	.000
组内	498987.913	78	6397.281		
总计	274639461.9	103			

可以得到新增感染人数与新增死亡人数的相关性概值  $p$  值约等于 0；而且远小于接受  $H_0$  的概率  $\alpha=0.05$ 。所以拒接原假  $H_0$ ，接受假  $H_1$ 。即新增死亡人数对新增感染人数有影响。

可以证明新增感染人数对新增死亡有显著影响：

## 五、问题探究三：对医疗需求的趋势分析

### 5.1 时间序列分析简述

时间序列是指将同一统计指标的数值按其发生的时间先后顺序排列而成的数列。时间序列分析试图通过研究过去来预测未来。时间序列分析在工程、金融、科技等众多领域有着广泛的应用。通过将时间序列分析与分类模型相结合，更好的应用于数据检测、预测等场景。

其应满足如下条件：

依时间顺序排列起来的一系列观测值（观测值之间不独立）

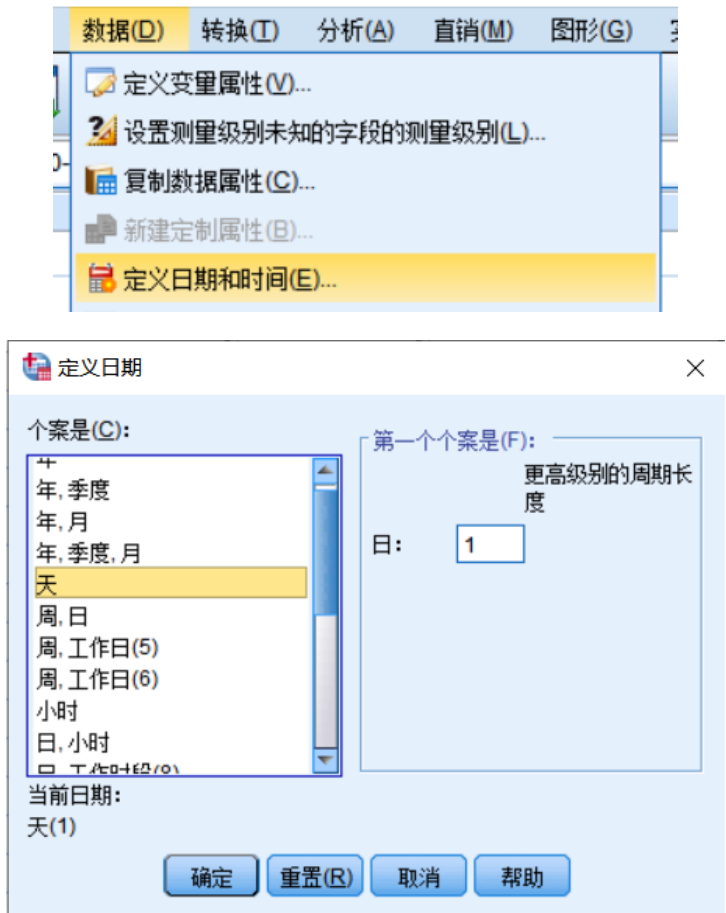
它考虑的不是变量间的因果关系，而是重点考察变量在时间方面的发展变化规律，并为之建立数学模型

有足够长的数据序列；数据序列的变动是稳定而规律的

由于疫情的治疗情况由许多因素决定，我们只能用一个  $t$ (时间)来替代所有的影响因素来研究自变量是如何变化的，对于未来一段时间医疗资源需求的短期预测。

## 5.2 分析步骤

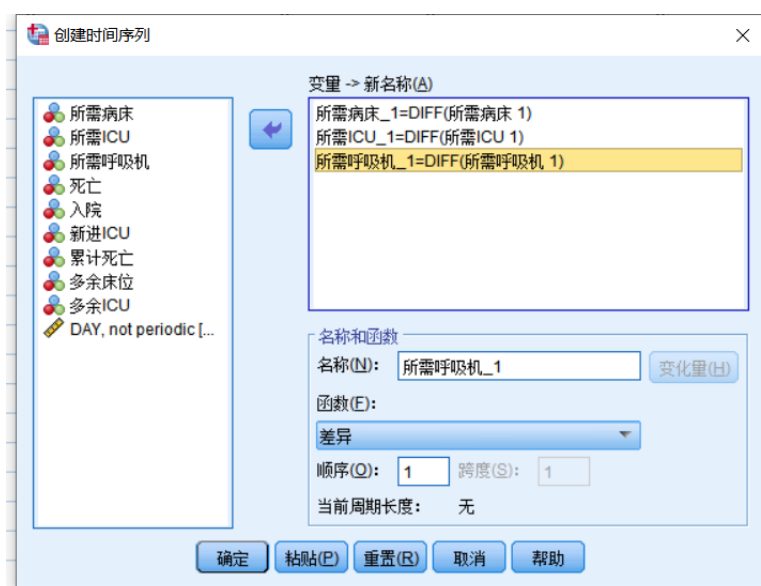
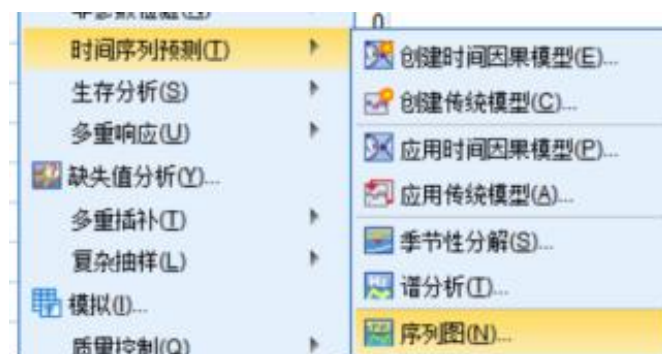
我们以每天定义时间序列（之后用周，日期）



接下来：为了帮我们找到适当的模型，最好先绘制时间序列。时间序列的可视化检查通常可以帮助我们进行选择。另外，我们需要清晰以下几点：

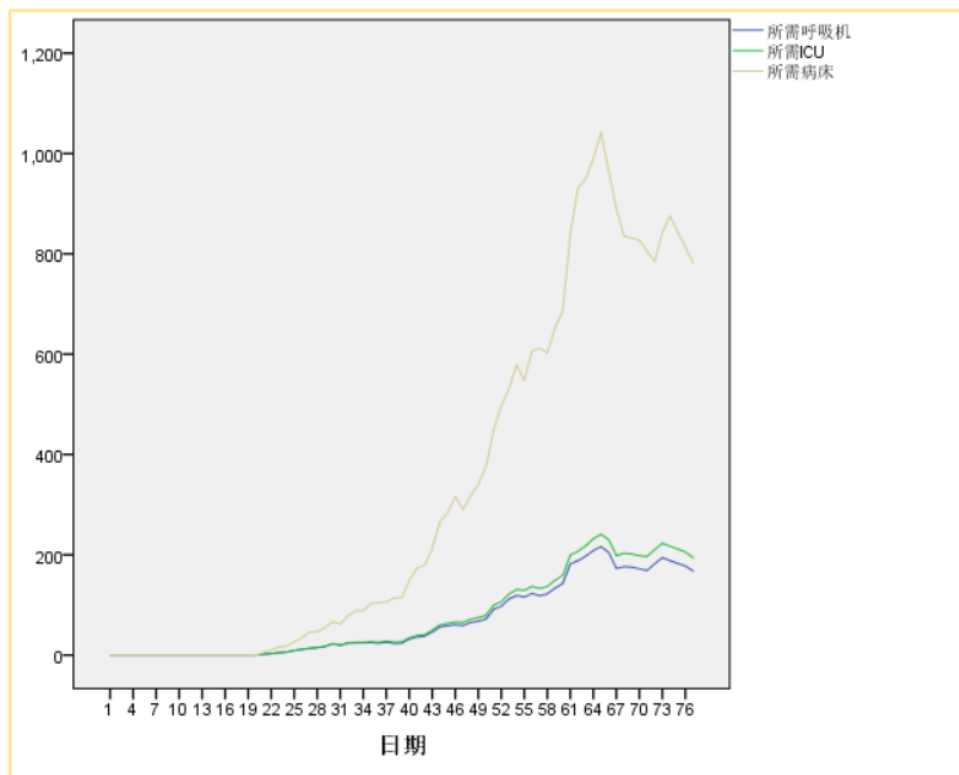
- 此序列是否存在整体趋势？如果是，趋势是显示持续存在还是显示将随时间而消逝？

• 此序列是否显示季节变化？如果是，那么这种季节的波动是随时间而加剧还是持续稳定存在？



由下图我们看到：此序列显示整体上升趋势，即序列值随时间而增加。上升趋势似乎将持续，即为线性趋势。此序列不存在一个明显的季节特征，即时间高点在 64 日（无周期性质）。病床、ICU、呼吸机的需求量变化显示随上升序列而增长，且初步表表象与线性拟合相同。





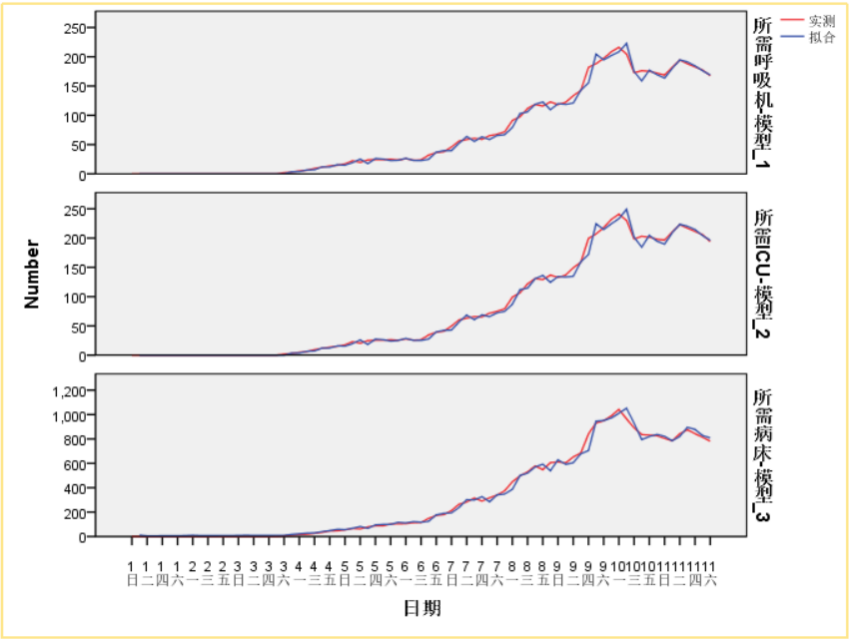
此时，我们对时间序列的特征有了大致的了解，便可以开始尝试构建预测模型。

SPSS 提供了三类预测方法：1-专家建模器，2-指数平滑法，3-ARIMA

以下我们分别用专家建模，指数平滑（简单、Holt）

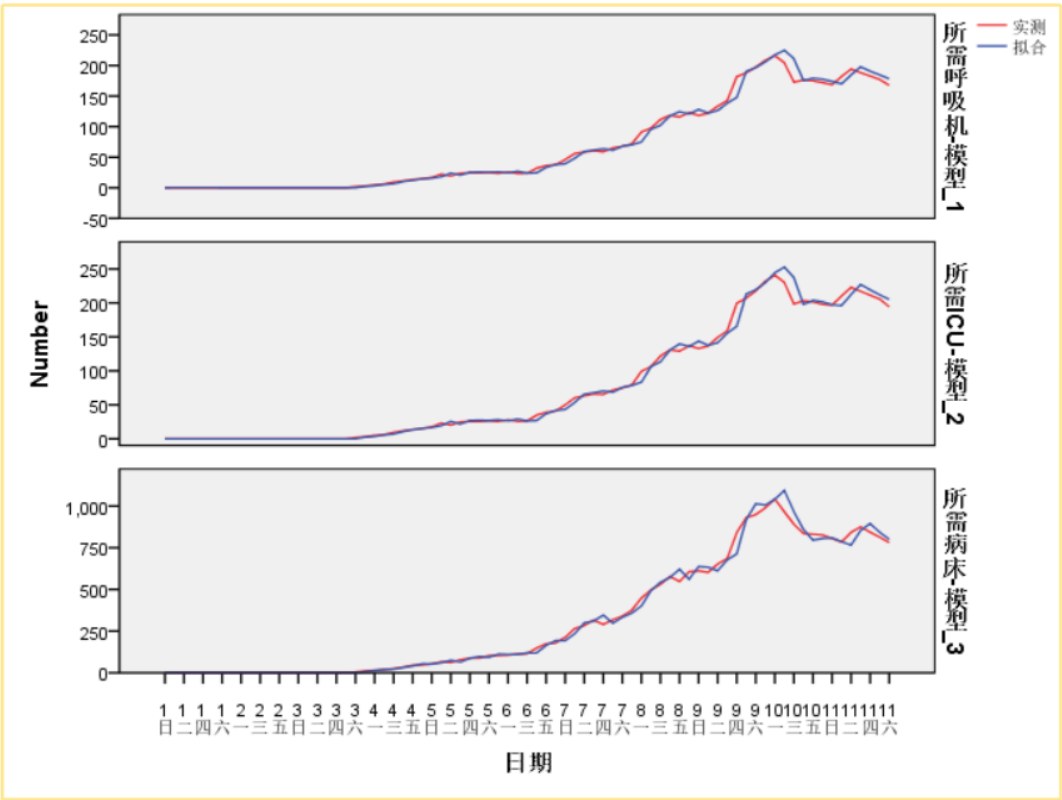


5.2.1 专家建模器

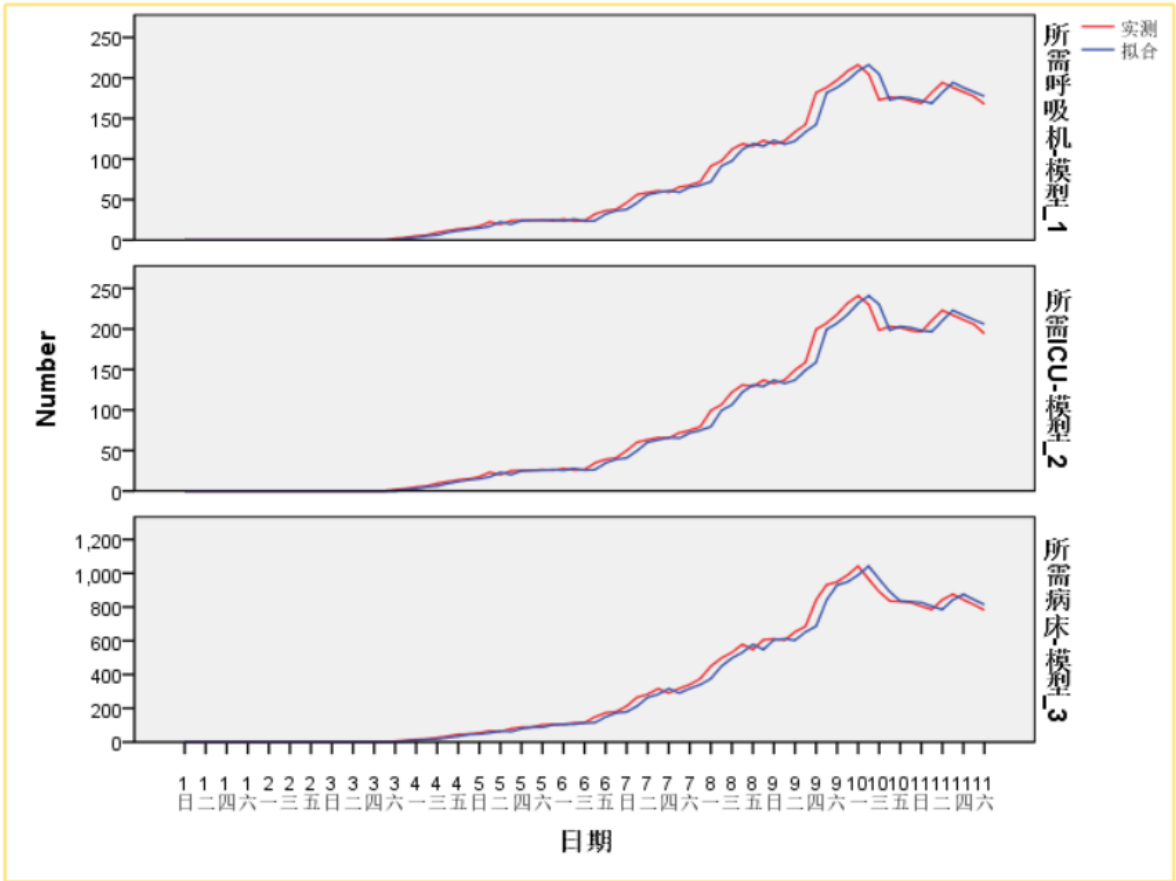


5.2.2 指数平滑

简单模型预测



Holt 线性趋势预测



模型统计						
模型	预测变量数	模型拟合度统计	杨-博克斯 Q(18)			离群值数
		平稳 R 方	统计	DF	显著性	
所需呼吸机-模型_1	0	-.081	57.846	17	.000	0
所需ICU-模型_2	0	-.100	59.554	17	.000	0
所需病床-模型_3	0	-.098	54.387	17	.000	0

5.3 分析结果及结论

简单模型预测（即无趋势也无季节）

首先我们采用最为简单的建模方法，就是简单模型，这里先不讨论模型的检验，只是直观的看一下预测模型的拟合情况，最后我们确定了预测模型后再我们讨论检验和预测值。

Holt 线性指数平滑法

一般选择：针对等级的平滑系数  $\alpha=0.1$ ,针对趋势的平滑系数  $\gamma=0.2$ ;

从图中我们看到，简单模型确实显示了渐进的上升趋势，但拟合程度较专家建模器拟合程度不高并不是我们期望的结果，但是没有季节性变化，也没有周期性改变，证明所需医疗资源直观的讲不会基于现阶段随时间上下波动，基本上与线性预测没有差异。

## 六、报告总结：

这次的实验让我体会到数据的重要性，找到合适的数据是十分困难的，很多数据不满足检验分析方法的前提，没有办法生搬硬套，数据的误差很有可能得出与我们认知相反的结论，要充分分析其中的原因，这与我们平时做实验有很大的不同。同时实验过程中复习了许多课上讲解的内容，很好地巩固了知识，并对这些方法有了新的理解。

这次实验令我们加深了对数据分析课程的理解，不但使我们巩固了老师上课讲的基础知识，更使我们对它们的中心思想、使用方式有了进一步的理解，从而提高了我们对数据处理、数据分析的能力。并且本次实验中也出现了课堂中未曾出现的数据结果，这就又促进我们学习更多的关于数据分析的知识，从而起到提高能力的效果。

## 七、附录

## 7.1 新增感染比例图生成代码

```
import matplotlib.pyplot as plt

from matplotlib.font_manager import FontProperties

font = FontProperties(fname=r"C:\Windows\Fonts\simhei.ttf", size=14)

China_case = [59, 235-59, 9714-235, 40206-9714, 74652-40206, 79929-74652, 80879-79929,
81229-80879, 82241-81229, 82925-82241]

Italy_case = [0, 0, 3, 3-3, 3-3, 1128-3, 9172-1128, 41035-9172, 101739-41035, 143626-101739]

France_case = [0, 0, 6, 11-6, 12-11, 100-12, 1412-100, 10995-1412, 44550-10995, 86334-44550]

United_Kingdom_case = [0, 0, 2, 4-2, 9-4, 23-9, 321-23, 3277-321, 22141-3277, 65077-22141]

United_States_case = [0, 0, 6, 12-6, 15-12, 69-15, 754-69, 14250-754, 164620-14250, 466033-
164620]

label = ['China', 'Italy', 'France', 'UK', 'US']

for i in range(len(China_death)):

    indic = []

    data = []

    data = [China_case[i], Italy_case[i], France_case[i], United_Kingdom_case[i], United_States_case[i]]

    titleList = [

        '一月上旬新增感染人数比例', '一月中旬新增感染人数比例', '一月下旬新增感染人数比例',

        '二月上旬新增感染人数比例', '二月中旬新增感染人数比例', '二月下旬新增感染人数比例',

        '三月上旬新增感染人数比例', '三月中旬新增感染人数比例', '三月下旬新增感染人数比例',

        '四月上旬新增感染人数比例',

    ]

    title = titleList[i]

    #我们将数据最大的突出显示

    for value in data:

        if value == max(data):

            indic.append(0.1)

        else:
```

```

indic.append(0)

plt.pie(
    data,
    labels=label,
    colors=color,
    startangle=90,
    shadow=True,
    explode=tuple(indic),#tuple 方法用于将列表转化为元组
    autopct='%1.1f%%'##是数字 1，不是 l
)

plt.title(title, FontProperties=font)

plt.show()

```

## 7.2 新增死亡比例图生成代码

```

import matplotlib.pyplot as plt

from matplotlib.font_manager import FontProperties

font = FontProperties(fname=r"C:\Windows\Fonts\simhei.ttf", size=14)

China_death = [0, 3, 213-3, 909-213, 2120-909, 2872-2120, 3139-2892, 3253-3139, 3309-3253,
3340-3309]

Italy_death = [0, 0, 0, 0, 0, 29, 464-29, 3407-464, 11591-3407, 18281-11591]

France_death = [0, 0, 0, 0, 1, 2-1, 30-2, 372-30, 3024-372, 12210-3024]

United_Kingdom_death = [0, 0, 0, 0, 0, 0, 5, 144-5, 1408-144, 7978-1408]

United_States_death = [0, 0, 0, 0, 0, 1, 26-1, 150-26, 3170-150, 16690-3170]

label = ['China', 'Italy', 'France', 'UK', 'US']

for i in range(len(China_death)):
    indic = []
    data = []

```

```
data = [China_death[i], Italy_death[i], France_death[i], United_Kingdom_death[i],
United_States_death[i]]

titleList = [
    '一月上旬新增死亡人数比例','一月中旬新增死亡人数比例','一月下旬新增死亡人数比例',
    '二月上旬新增死亡人数比例','二月中旬新增死亡人数比例','二月下旬新增死亡人数比例',
    '三月上旬新增死亡人数比例','三月中旬新增死亡人数比例','三月下旬新增死亡人数比例',
    '四月上旬新增死亡人数比例',
    ]

title = titleList[i]

#我们将数据最大的突出显示

for value in data:
    if value == max(data):
        indic.append(0.1)
    else:
        indic.append(0)

plt.pie(
    data,
    labels=label,
    colors=color,
    startangle=90,
    shadow=True,
    explode=tuple(indic),#tuple 方法用于将列表转化为元组
    autopct='%1.1f%%'##是数字 1，不是 l
)

plt.title(title, FontProperties=font)

plt.show()
```