

国内外疫情发展态势对比分析

刘崑

慕慧君

肖雨婷

王洁

卢晋萍

2020/6/21

摘要

目前, COVID-19 虽然已经成功在中国得到控制, 但仍在全球范围内传播。根据世界卫生组织报告的数据显示, 不同地区和国家 COVID-19 呈现出不同的发展态势。研究疫情发展规律, 掌握疫情的发展态势, 有利于政府安抚民心以及科学有效地制定防控政策。本文对全球疫情的发展态势进行了可视化分析, 然后选取 2019 年 GDP 前十的代表性国家对其 COVID-19 的不同发展趋势进行了分析, 为科学有效地控制疫情提供相关建议。首先利用可视化分析方法对 2019-12-31 日至 2020-05-28 全球各国 COVID-19 每日新增病例进行分析, 接着将可视化分析、聚类分析、Logistic 增长曲线非线性回归这三种方法应用于分析 2019 年 GDP 前十的国家 COVID-19 累计确诊病例、每日新增确诊病例、累计死亡病例这 3 项数据, 并计算每日确诊增长率、死亡率这 2 项数值辅助开展分析。经分析, 在 GDP 前十的国家中, 中国疫情是其中最先也是最短时间将疫情控制住的国家, 疫情防范效果最为突出; 德法意这三个国家疫情形势相似, 和紧邻中国的日本, 这 4 个国家疫情基本得到控制。而美国、英国、加拿大、印度和巴西这五个国家目前疫情均不明朗。可见, 中国疫情防控政策值得其他国家采纳, 如德意法日四个国家疫情均得到控制; 而美国、英国政府不干预致使疫情形势不明朗, 给世界带来重大的健康和经济损失。

关键词: COVID-19; 可视化分析; 聚类分析; Logistic 增长曲线

1 引言

1.1 背景与目的

2020 年初, 不明原因的病毒性肺炎在中国武汉市爆发, 随后, 其病原体被确定为新型冠状病毒, 并在全球多个国家相继出现。根据世界卫生组织报告, 1 月 11 日, 西太平洋地区出现 41 例确诊病例, 1 月 13 日, 东南亚出现 1 例确诊病例, 1 月 20 日, 美洲出现 1 例确诊病例, 1 月 24 日, 欧洲出现 3 例确诊病例, 1 月 29 日, 地中海东部地区出现 4 例确诊病例, 2 月 25 日, 非洲出现 1 例确诊病例。截止 5 月 12 日 6:42 CEST (Central Europe Summer Time, 西太平洋地区新冠肺炎累积确诊病例已达到 161872 例, 东南亚地区累积确诊 107354 例, 美洲累积确诊 1743717 例, 欧洲累积确诊 1755790 例, 地中海东部地区累积确诊 281744 例, 非洲累积确诊 46829 例。

从以上数据可以看出, 同一病毒性肺炎在不同地区和国家呈现出不同的发展态势。研究疫情发展规律, 掌握疫情的发展态势, 有利于政府安抚民心以及科学有效地制定防控政策。本文对全球疫情的发展

态势进行了可视化分析，然后选取一些代表性国家对新冠肺炎的不同发展趋势类型进行了分析，并对各个国家肺炎发展趋势进行了拟合和预测，为科学有效地控制疫情提供相关建议。

1.2 文献综述

随着新型冠状病毒肺炎的全球蔓延，关于新冠肺炎疫情的相关研究也日渐增多。其中，对新冠肺炎疫情发展态势的研究早在疫情爆发的初期就已引起学者们的关注。

杨安吉，颜忠诚运用 logistics 模型对新冠肺炎的发展趋势进行了研究和解释，结果表明，logistic 增长曲线的拟合值与实际观测数据差异较小，中国内地 COVID-19 疫情的累积确诊患者病例数、累积治愈患者例数和累计死亡患者例数变化趋势基本符合 logistic 增长曲线。可以看出 logistic 模型对疫情的发展趋势具有较好的拟合效果；易大莉等人使用聚类分析方法对新冠病毒肺炎疫情的发展进行区域差异分析；武文涛等人运用 SIR 模型，通过计算和预测广东省新型冠状病毒肺炎的基本再生数，研究广东省疫情的流行趋势为其防控和相关政策支持提供依据；杨宇琦等人使用官方数据，计算出此次疫情确诊病例环比增长率、治愈率和死亡率，绘制变化曲线图分析疫情发展态势，并建立 SIR 传染病模型进行趋势预测；张荣强等人采用 Excel、Origin 软件开展 H1N1 和 COVID-19 的描述性流行病学分析，并分析二者的变化规律，探讨和比较 2019 冠状病毒病（COVID-19）和甲型流感（2009-H1N1）疫情初期的变化趋势及流行特征；方法上，《Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world》一文将 logistic 增长模型、广义 logistic 增长模型、广义增长模型和广义 Richards 模型与 1 月 19 日至 3 月 10 日全国、29 个省、4 个重灾区 and 整个欧洲的报告感染病例数进行了校正。不同的模型提供了各个场景预测的上下限。

以上文献多是关于局部地区的新冠肺炎疫情发展趋势的分析，方法上多用到 SIR 模型以及 logistic 增长模型，而关于全球性的疫情态势分析极少。本文着眼全球疫情角度，对国内外疫情发展态势进行分析，从多角度对全球疫情发展态势进行刻画，由于分析的国家地区较多，使用 SIR 模型需要考虑的额外因素较多，因此本文在进行趋势拟合和预测时采用了 logistic 增长模型进行分析，并根据所得的结论为抵制新冠病毒肺炎疫情提供相关建议

1.3 写作框架及主要内容

- 1、基于全球疫情态势的描述性统计分析。这部分内容从不同角度，选取不同指标对全球疫情的发展态势进行了刻画和分析。
- 2、聚类分析。对 2019 年 GDP 排名前 10 的国家的疫情发展进行聚类，分析区域差异。
- 3、趋势拟合及预测。对 2019 年 GDP 排名前 10 的国家进行新冠肺炎累积确诊病例的拟合和预测。
- 4、总结。总结前文的分析结果，并结合各个国家的实际情形和相关政策，为抵抗疫情提供合理建议。

2 描述性统计分析

2.1 全球每日新增病例

对各国每日新增病例进行比较分析，取 2019-12-31 日至 2020-05-28 的各国 COVID-19 每日新增病例的数据[数据来源](#)，对比各国 COVID-19 发展的速度和各国的对抗效力。由于各个国家病例数相差较

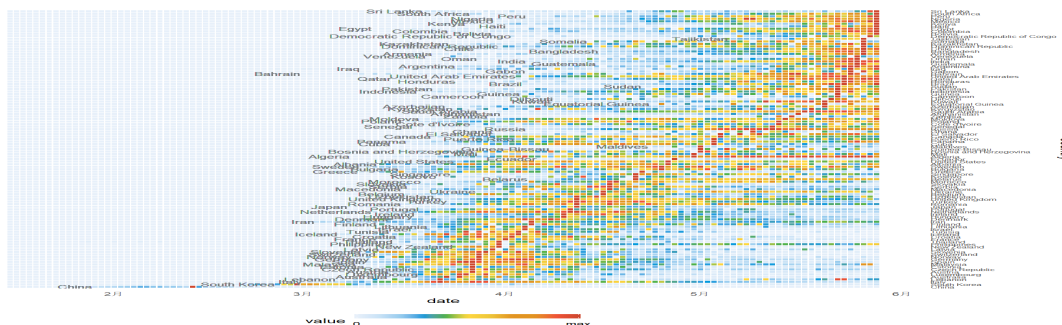


图 1: 全球归一化的每日新增病例

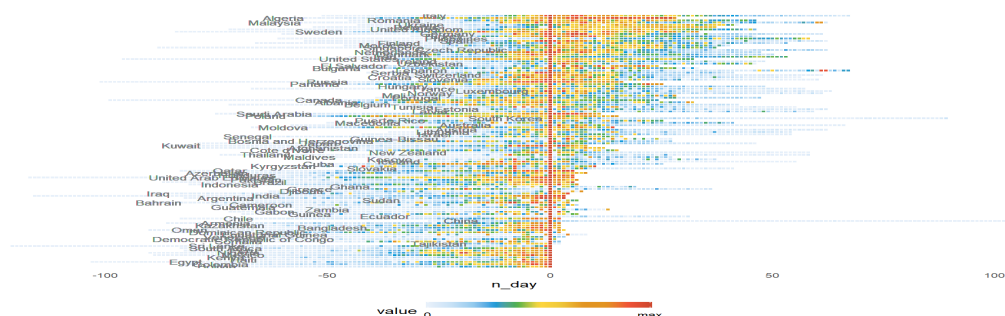


图 2: 各国对抗 COVID-19 的效力比较

大，所以这里对每个国家/地区的新增病例的进行归一化（(1) 式），范围从 0 到 1。公式如下：

$$f(x) = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

现实中的数据并不像数学中的函数那样光滑，每日新增病例具有一定的波动性，单日新增最大值出现并不意味着拐点出现，因此，当持续一段时间的总体趋势在下降，或者连续一段时间下降，才能说一个地区疫情的拐点确实到来了。

动态图 (动态图见 HTML 版) 这种生动的可视化聚焦于病毒从中国开爆发，随后病毒在世界其他地区扩散的时间顺序。为了增强视觉效果，根据每个国家/地区达到每日病例高峰（深红色框）的日期，将前 90 个国家/地区放在两个半对角线上。

为了进行更详细的分析，创建了图 1，它与动态图图的动画相同，但是国家/地区从下到上排序。从图 1 中可以看到，1 月的下旬 COVID-19 开始在中国爆发，到三月左右国内疫情基本得到控制，每日新增数变得很少；紧随中国之后的是韩国，3 月初出现单日新增最高值（首现拐点），持续了几天数量较大的单日新增，之后单日新增数量保持在较小值的一定范围波动。

图 2 以每日最多新增病例为中心，该图显示了前一阶段和后续阶段的时间和强度。图中可以看到，中国的单日新增最大值基本就是拐点，之后每日新增连续下降；意大利、德国、西班牙、法国等出现单日新增最大之后，还持续了较长一段时间的增减反复，才迎来了拐点；而英国达到每日新增病例最大值后的近一个月还未迎来拐点。美国在出现单日新增最大之后，每日新增还在较大值波动，也就是说美国的新冠累计确诊还在较大的增长，拐点还未显现。

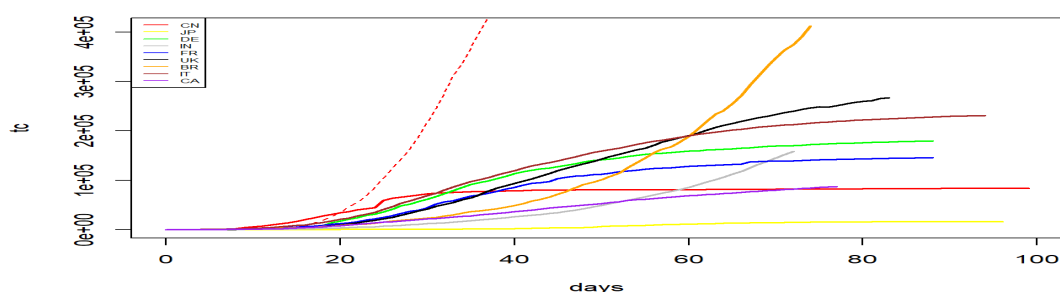


图 3: 2019 年 GDP 前十国家累计确诊数比较

2.2 2019 年 GDP 前十各国疫情态势对比

此次选取了 2019 年 GDP 前 10 的国家进行疫情累计确诊的对比分析, GDP 前十的国家中大多数都是在国际上有一定影响力的发达国家, 且国家综合实力较强, 或者是发展中的人口大国。如果大国在抗疫方面做的很好, 就可以帮助一些国家实力较弱的国家进行抗疫; 如果大国自己的疫情都不能控制, 那么我们极有可能会较长时间的跟病毒做斗争。

为了直观的比较各国疫情时间发展, 横轴取各国第 100 例病人确诊的时间作为起始点, 因为中国跨度较大, 且后期疫情数据变化微小, 发展时间最多取 100 天, 由于美国累计确诊人数较多, 纵轴数值较大, 导致其他国家的累计确诊人数的曲线轨迹不明显; 因此取这十个国家中累计确诊人数第二多的巴西作为纵轴取值标准。

图 3 中红色虚线代表美国, 可以看到在确诊第 100 例新冠肺炎之后的 35 天左右, 美国累计确诊人数已经超过了 35 万人; 同样的时间跨度, 中国抗疫已经接近后期, 剩余的 8 个国家累计确诊还处在不同程度的增长阶段。

在拐点方面, 可以看到疫情较为严重的欧洲: 意大利、德国和法国均在出现累计确诊第 100 例病人 40 多天后出现了拐点; 至 5 月 28 日, 英国距离出现第 100 例 COVID-19 累计确诊病例过了 80 天, 图中英国显示累计确诊病例增速放缓, 如果未来一段时间能够加强疫情防控, 相信拐点很快将会到来; 同时, 巴西、印度的疫情还在进一步恶化, 特别是巴西最近累计确诊人数急速上升, 不到 20 天已经超越了意大利, 累计确诊病例增速加快。日本的在近 5 月 28 日几天增长较为平缓, 只要后续时间疫情突然反跳的情况, 按照目前的趋势日本的 COVID-19 基本得到控制。

由于各国确诊人数相差较大, 比较各国累计死亡人数意义不大。这里将采用 COVID-19 各国累计死亡人数占累计确诊人数的比例来作为死亡率, 公式为:

$$dr = \frac{TotalDeaths}{TotalCases} * 100 \quad (2)$$

图 4 展示了 2019 年 GDP 前十的国家出现第一例 COVID-19h 患者死亡后至 2020 年 5 月 28 日的死亡率的变化, 横轴为各国出现第一例患者死亡后的时间跨度 (死亡率发展时间根据日本的数据截止到 105 天), 纵轴为各国死亡率 (%). 从图中可以看出, 在出现 COVID-19 患者死亡初期, 由于累计确诊人数少, 死亡率会有较大的波动, 随着累计确诊人数增多 (分母增多), 死亡率呈一定的趋势波动。累计确诊人数遥遥领先的美国, 最近一个多月死亡率却在 6% 上下波动; 反而是累计确诊人数不到 15 万人的法国, 死亡率快速增长, 目前已经超过了 19%; 虽然日本近日的死亡率呈一定的上升趋势, 但能够加强疫情防控, 参考中国的死亡率, 这种趋势不会持续上升。除此之外, 意大利的死亡率近来在 14%

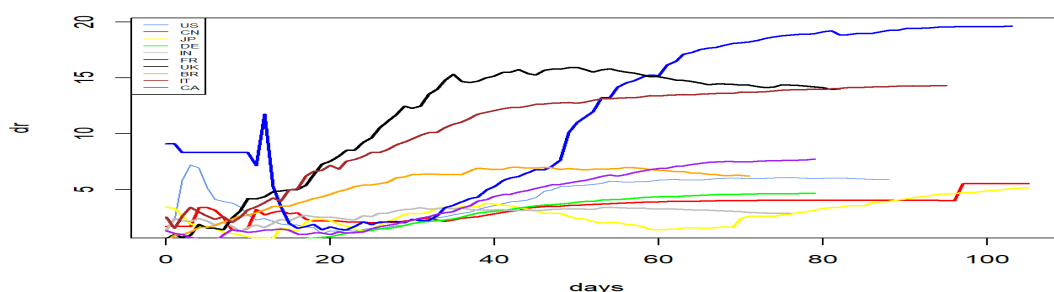


图 4: 2019 年 GDP 前十国家 COVID-19 死亡率比较

波动；英国的死亡率在升到超过 15% 后，缓降至 14% 波动。这些线图中最令人惊讶的是德国，德国的累计确诊人数虽然不是最多，却是高于法国的，从图 4 中也能看出；而德国的死亡率至今从未超过 5%，且最近增速平缓，德国能做到这一点除了健全医疗体系、充足医疗储备之外，也和德国强大的科研能力以及特殊的社会、家庭结构紧密相关。[这篇报道](#)很好的解释了德国能够做到低死亡率、高治愈率的原因。

3 聚类分析

3.1 聚类分析简介

统计学中，聚类分析是将随机现象进行归类的统计学方法，已广泛应用于医学科学研究之中。聚类分析也成为群分析、点群分析，是研究分类的一种多元统计方法。

由于研究的样品或指标（变量）之间存在着程度不同的相似性（亲疏关系），故可根据一批样本的多个观测指标，具体找出一些能够度量样品或指标之间相似程度的统计量，以这些统计量为划分类型的依据，把一些相似程度较大的样品进行（或指标）聚合为一类，把另外一些彼此之间相似程度较大的样品（或指标）又聚合为另一类，关系密切的聚合到一个小的分类单位，关系疏远的聚合到一个大的分类单位，直到把所有的样品（或指标）聚合完毕，这就是聚类的基本思想。

3.2 实证分析

3.2.1 数据来源

数据选取的是 2019 年 GDP 前 10 的国家于 3 月 3 日至 5 月 25 日每日新增病例的数据。数据来源于美国约翰斯·霍普金斯大学发布的疫情数据。

3.2.2 数据处理

原始数据中存在许多异常值，例如负值或者 0 值，又或者突然出现极大值，首先查阅相关疫情报道对这些数据进行修正。对每个国家/地区的新病例的进行归一化（(1) 式），范围从 0 到 1。

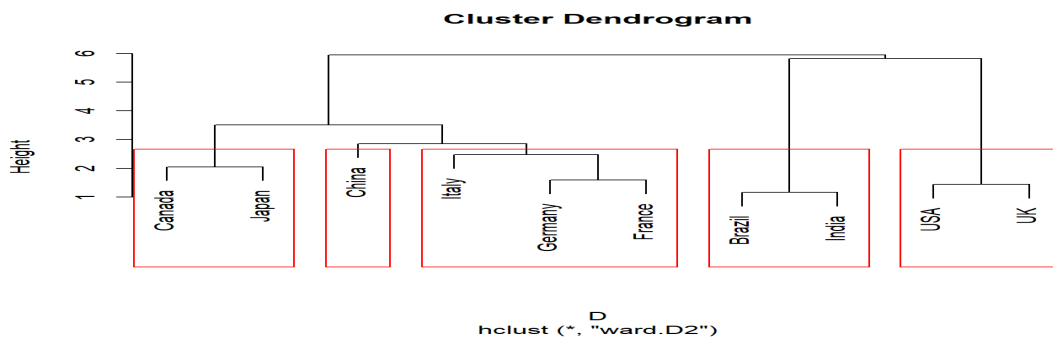


图 5: 2019 年 GDP 前 10 国家聚类结果

3.2.3 方法选择

由图 5 聚类分析结果可知，我们把 10 个国家分为了 5 类，第一类是加拿大和日本，第二类是中国；第三类是意大利，德国和法国这 3 个国家；第四类划分的国家有：巴西和印度；最后一类国家分别是：美国 and 英国。每一类国家在疫情发展的大致趋势上具有相同的走向。接下来图 6 是对每一类国家疫情发展的趋势做进一步的分析：用处理过的数据拟合其发展曲线。

首先，是加拿大和日本这两个国家，这两个国家是疫情防控措施做的较好的国家，日本自疫情在国内发展的初期就采取了严格的防控措施，使得国内疫情每日新增人数最多未超过千人；加拿大同样也在抗疫期间逐步升级其防控力度，严格把关边境口岸，禁止任何出现症状的人员入境等。结果，虽比邻美国，国内疫情发展并没有到像美国一样很严重的境地，但目前加拿大的疫情形势仍然严峻。每日新增确诊人数约千人，预计得到较好的控制还需要一段时间。

第二类是中国，从 3 月份以来国内疫情基本得到了控制，所以从图中可以看出近 3 个月来每日新增确诊人数最多是出现在 3 月份初期，之后是逐步下降的趋势，过程中出现一次波动，可能对应于国内出现无症状患者的情况，但这一情况很快也得到了控制。现在，国内的每日新增病例基本维持在 2 位数以内。

第三类国家有意大利，德国，法国这 3 个国家，这些国家都是欧洲国家，其中，以意大利为代表，从 3 月份以来，疫情迅速爆发，呈现出指数增长的趋势，在当时，成为全球疫情的一个重灾区，后来逐步采取封城、强制佩戴口罩等措施进行防控，现在，这些国家的疫情基本得到了控制，国家开始逐步解封，恢复部分正常的生活节奏。

第四类国家巴西和印度，这两个国家的疫情情况仍然处于一种迅速爆发的状态，近期，巴西每日新增确诊人数近 2 万人，成为除美国之外全球第二个疫情重灾区；印度的疫情也同样严重，这可能和印度人口众多，但相关软件、硬件设施较差有关；从图上也可以看出这两个国家每日确诊人数正在快速增加。

第五类国家是美国和英国，美国一直是全球疫情最为严重的国家，近期仍然未得到很好的控制，确诊人数达到近 170 万人，其原因一部分源于政府在疫情爆发初期的不作为，防控措施不到位；另一方面也源自于美国公民的不在意导致疫情蔓延到如此严重的局面。英国感染人数 25 万左右，死亡人数超过 3.6 万人，成为了欧洲死亡人数最多的国家，并且英国的疫情似乎没有减弱的趋势。主要原因还要归于政府的防疫策略上，英国放弃了追踪接触者的战略，而是将检测重点放在新冠肺炎重症患者身上。导致错失尽早切断疫情传播途径的良机。最终，英国和美国还需要一段时间才能宣告抗疫的胜利。

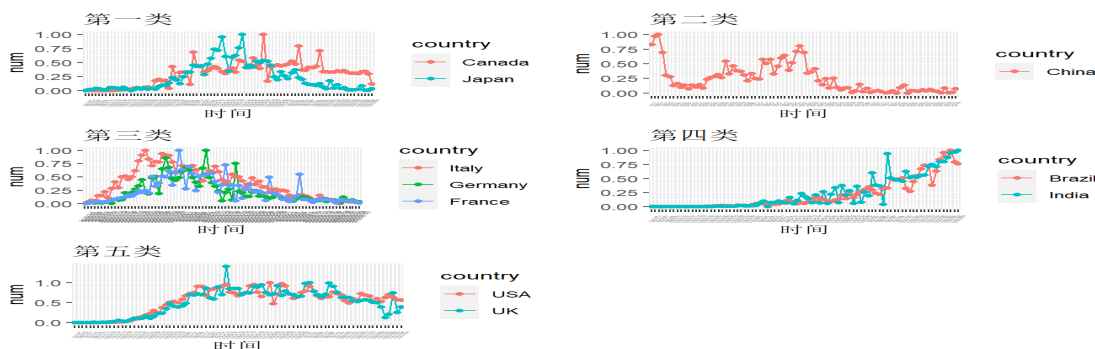


图 6: 各类拟合曲线图

4 趋势拟合及预测

本章节通过拟合患者比例和增长率的变化来分析疫情趋势，将复杂的外界影响因素视为患者数量增长的环境压力，以拟合生态增长曲线的方式解读 2019-nCoV 病毒的感染患者例数变化，分析国内外 COVID-19 疫情现状并预估发展趋势。

4.1 COVID-19 预测预报非线性模型及其机理简介

在时间序列里，有些变量的增长量最初比较小，随时间的增加逐渐增长而达到一个快速增长时期，而后增长速度趋缓，最终达到稳定的总增长量，这一过程若用曲线来表示，则是一种拉长的 S 形曲线。这种 S 形曲线因变量的增长特性的不同而呈现出多样性变化。通常传染病的累计病例遵循 S 形曲线增长。Logistic 函数是一种常见的 S 形函数，它是皮埃尔·弗朗索瓦·韦吕勒在 1844 或 1845 年在研究它与人口增长的关系时命名的。该模型广泛应用于生物繁殖和生长过程、人口增长过程模拟。在分析 COVID-19 疫情变化时，可以将患者看作在限定环境中数量增长的物种，实际中病毒的传播也存在人为管控或自然阻力，因此 2019-nCoV 病毒的感染者例数不会无限制的增长，在一定空间范围内应符合 Logistic 增长曲线。

Logistic 方程, 即常微分方程如下:

$$\frac{dP}{dt} = rP * (1 - \frac{P}{K}) \quad (3)$$

将上述方程求解，可以得到 Logistic 函数:

$$P(t) = \frac{K P_0 e^{rt}}{K + P_0 (e^{rt} - 1)} \quad (4)$$

其中 P_0 为初始值, r 为衡量曲线变化快慢, t 为时间。 $\frac{dP}{dt}$ 为种群增长率 (单位时间个体数量的改变), P 为种群的大小 (个体的数量), K 为可能出现的最大种群数 (上渐近线) 或承载力。

4.2 数据说明与方法

4.2.1 数据说明

本章节选用 2019 年 GDP 前十的国家：美国、中国、日本、德国、印度、法国、英国、巴西、意大利和加拿大。对 GDP-TOP10 的国家自 2019 年 12 月 31 日至 2020 年 5 月 24 日 COVID-19 累计确诊病例进行拟合和预测，以及对 GDP-TOP10 的国家最终流行规模的预测值进行比较。同时，计算 GDP-TOP10 的国家自 2019 年 12 月 31 日至 2020 年 5 月 24 日每日确诊增长率数值辅助开展分析，预估 GDP-TOP10 的国家发展趋势。

4.2.2 方法说明

(1) **曲线拟合方法** 曲线拟合 (curve fitting) 是指选择适当的曲线类型来拟合观测数据，并用拟合的曲线方程分析变量之间的关系。本章节将每日累计确诊病例数据制成散点图，采用单个变化量对时间的响应进行非线性回归拟合，取得拟合方程的系数，并绘制回归曲线。

结合 S 值、 R^2 值对拟合程度进行判断，其中 S 值描述了数据值与拟合值的距离，表示评估模型描述响应值的程度，其值越低，模型描述响应的程度越高； R^2 值是度量拟合优度的统计量，介于 0-1 之间，也称拟合的确定系数， R^2 越高可表示拟合优度越高。 S 值和 R^2 值表达式如下所示：

$$S = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5)$$

$$R^2 = 1 - \frac{SSE * (n - k)}{(var(y)) * (n - 1)} \quad (6)$$

(2) Logistic 增长曲线

本章节所采用的 R 包为 `nlrwr` 包中的 `nls()` 函数，相关使用可参考《Nonlinear Regression with R》。函数 `nls()` 自带一组自启动函数，即 “self-starter functions”，其可以自动设置初始值，Logistic 的自启动函数为 `SSlogis(x, Asym, xmid, scal)`，其表达式为：

$$\frac{Asym}{(1 + \exp(\frac{xmid - x}{scal}))} \quad (7)$$

与 Logistic 函数表达式略有不同，此处参数 $Asym$ 代表渐近线的数字参数，其值可被作为最终流行规模的预测值，及 Logistic 函数表达式中的 K 值，参数 $xmid$ 表示曲线拐点处的值，参数 $scal$ 表示 x 轴上的数字比例参数。

(3) 每日确诊增长率数值

本章节所采用的每日确诊增长率数值的表达式为：

$$r(t) := \log\left(\frac{P(t)}{P(t-1)}\right) \quad (8)$$

参考《neralized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world》

Country	Day	Asym	lower_Asym	upper_Asym	xmid	scal	R_2	S
United States	125	1.63E+06	1585238	1669113	93.4	11.3	0.996	1.60E+11
China	146	8.27E+04	82364	83007	41.6	4.66	0.998	3.32E+08
Japan	131	1.66E+04	16442	16707	92.7	7.93	0.999	6.49E+06
Germany	118	1.71E+05	168671	172389	69.2	8.00	0.997	1.92E+09
India	116	2.72E+05	250912	296895	117.0	12.7	0.999	1.35E+08
France	121	1.40E+05	138596	141285	73.6	8.23	0.998	9.41E+08
United Kingdom	114	2.62E+05	256924	268136	83.9	10.8	0.997	2.54E+09
Brazil	89	1.07E+06	939451	1254222	98.9	13.2	0.999	5.90E+08
Italy	115	2.23E+05	220316	226127	64.6	10.2	0.996	3.52E+09
Canada	120	8.68E+04	85062	88657	90.9	11.3	0.998	2.23E+08

图 7: Logistic 增长曲线参数估计

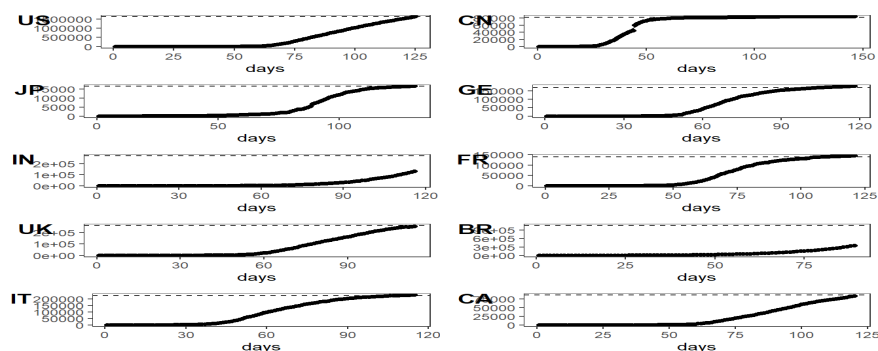


图 8: 累计确诊病例回归拟合

4.3 结果与分析

4.3.1 累计确诊病例回归结果分析

从图 7 中可以看到获得的 Logistic 增长曲线的参数估计, 以及 S 值和 R^2 值。根据 R^2 值一列, 可以看到利用 Logistics 增长曲线对 GDP-TOP10 的国家自 2019 年 12 月 31 日至 2020 年 5 月 24 日累计确诊病例数据进行拟合效果显著, 且在进行回归分析时, GDP-TOP10 的国家 Logistic 增长曲线的估计参数均通过显著性检验。参数 Asym 代表渐近线的数字参数, 通常被用来作为最终流行规模预测值, 从表中可以看到疫情由高到低的顺序为: 美国、巴西、印度、英国、意大利、德国、法国、加拿大、中国、日本。而我国作为第一人口大国, 累计确诊病例远低于其他国家, 可见我国政府的管控措施显著。参数 xmid 表示 x 曲线拐点处的值, 将参数 xmid 值与 Day 值比较, 可见印度和巴西还未达到拐点值, 且印度即将迎来拐点, 其他国家均达到拐点。

4.3.2 累计确诊病例回归拟合分析

图 8 是利用获得 Logistics 增长曲线对 GDP-TOP10 的国家累计确诊病例数据的拟合, 从图中可以看到拟合效果良好。中国、日本、德国、法国和意大利这五个国家, 疫情基本稳定, 总增长量基本平稳, 长时间停留在渐近线附近。美国、英国和加拿大这三个国家, 刚刚抵达渐进线处, 疫情情况还不明朗, 政府应加强控制。印度和巴西这两个国家, 离渐近线较远, 疫情还未达到拐点, 累计确诊病例可能会继续上升。

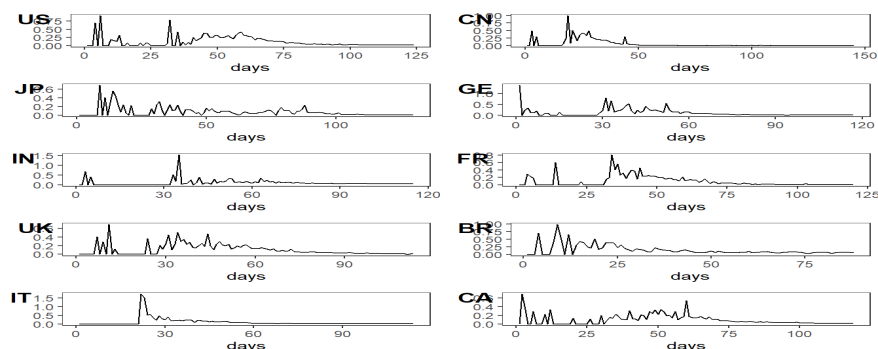


图 9: 每日确诊增长率趋势

4.3.3 每日确诊增长率趋势分析

图 9 是每日确诊增长率折线图, 从图中可以看到中国每日确诊增长率已长时间趋于 0。其次, 日本每日确诊增长率也开始趋于 0, 为 0.0008。印度和巴西这两个国家每日确诊增长率最大, 分别为 0.052 和 0.048, 累计确诊病例仍会继续大幅增加。美国、英国、加拿大这三个国家每日确诊增长率为第二大, 均在 0.013 附近。德国、法国、意大利这三个国家每日确诊增长率接近, 在 0.002 附近。

综上所述, 中国疫情已经得到控制, 疫情防范效果最为突出, 是十个国家最先也是最短时间将疫情控制住的国家, 也体现我们国家领导人坚持“不计成本, 生命至上, 人民至上”的原则。日本疫情也基本得到控制, 是十个国家第二位将疫情控制住的国家。德国、法国、意大利这三个国家疫情形势相似, 疫情也基本得到控制。印度和巴西这两个国家疫情形势较为相似, 是十个国家中疫情还仍未达到拐点的两个国家, 这两个国家累计确诊病例仍会继续大幅增加, 政府应加强管理措施。美国、英国、加拿大这三个国家疫情形势接近, 疫情态势还不太明朗。

5 总结与建议

5.1 总结

(1) “纵观世界风云, 这边风景独好” 通过对 COVID-19 每日新增数据的可视化, 可以直观地对各国疫情发展有全局的把握。中国、韩国等虽然属于最早出现疫情的一批国家, 但是由于国家政府采取强有力的疫情控制政策和大多数民众的配合, 尤其是中国, 使得本国疫情得到很好的控制。从全球归一化的每日新增病例动态图可以看出, 中韩两国在三月左右先后出现拐点, 说明两国疫情从爆发到得到基本控制仅花费一个月左右的时间。

从 GDP 前 10 的国家的疫情累计确诊对比分析可以看出, 在同样的时间跨度, 中国抗疫进入成果保卫战阶段, 目前主要任务是在如何实现全面复工复产的同时防止疫情的反复。日本在疫情大规模爆发后, 开始学习中国科学抗疫, 在近几日增长较为平缓, 按目前趋势线所示疫情也得到基本控制。进一步的, 通过对 Logistic 增长曲线的估计参数 Asym 排序, 得出的结论也与描述统计部分一致, 中国和日本在疫情由高到低的顺序中分别为第九和第十。而且利用 Logistic 增长曲线对 GDP 前 10 的国家累计确诊病例数据的拟合可以看出中国和日本疫情基本得到控制, 总增长量基本平稳, 长时间停留在渐近线附近。并且从每日确诊增长率折线图可以看到中国每日确诊增长率已长时间趋于 0; 日本每日确诊增长率也开始趋于 0, 为 0.0008。

(2) “革命尚未成功, 同志仍需努力” 虽然通过 Logistic 增长曲线对德国、法国和意大利这三个国家累计确诊病例数据的拟合, 拟合结果与中日相似, 总增长量基本平稳, 长时间停留在渐近线附近, 聚类分析的结果也与此基本一致。德法意三个国家每日确诊增长率在 0.002 附近, 并且从 GDP 前 10 的死亡率折线图可以看出, 法国死亡率目前已经超过的 19%, 目前还在快速增长; 意大利的死亡率近来在 14% 波动; 德国得益于健全医疗体系等国情死亡率至今还未超过 5%, 但是整体来看还是需要还处于疫情防控的攻坚阶段。这三个国家均出现疫情缓和的情况, 但仍需谨防后续时间疫情突然出现反跳的情况。

(3) “黎明之前最黑暗” 通过前文的分析, 虽然美国、英国、加拿大、印度和巴西这五个国家目前疫情均不乐观, 但根据聚类分析的结果并结合国家政府对待疫情态度的不同, 抗疫政策的差异以及综合国力的差距, 各国疫情如此严重的成因各不相同。

根据 Logistic 增长曲线拟合的分析, 美国、英国和加拿大这三个国家, 刚刚抵达渐进线处, 且通过对 GDP 前 10 最终流行规模预测值的排序, 美国第一位, 英国第四位, 加拿大第七位。截至 5 月 28 日美国累计确诊人数已经近 170 万人, 虽然美国目前 GDP 排名第一, 国家综合实力也处于领先地位, 但是由于国家领导人对疫情的不重视和不当的抗疫政策, 导致美国即使拥有先进的医疗技术和最好的医疗资源, 却沦为新冠疫情的“重灾区”。3 月 5 日, 英国出现首例死亡病例, 距今已有 74 天, 但累计确诊至今仍在缓慢增长, 拐点也仍未出现, 并且英国的死亡率超过 10%, 截至目前虽有所回落, 但仍在 14% 波动。截至 5 月 28 日, 加拿大累计确诊也达 88989, 较上日新增 899 例, 所以加拿大目前疫情情况还不明朗, 但疫情严重程度较英美两国轻, 所以聚类的结果将其与日本归为同类。

印巴属于发展中国家, 医疗水平与欧美发达国家相比有很大差距。截至 5 月 28 日, 巴西、印度的疫情还在进一步恶化, 累计确诊病例增速加快, 并且通过拟合, 印巴两国的拟合曲线离渐近线较远, 疫情还未达到拐点, 累计确诊病例可能会继续上升。截至目前, 印度累计确诊已逾 15 万人。虽然相比美国、英国而言, 印度的疫情并不严重, 而且死亡率也没有意大利等国家高。但是在隐患上来看, 印度最高。之所以这样说是因为, 印度是个拥有 13 亿人口的人口大国, 医疗卫生水平落后, 贫富差距悬殊, 还有之前蝗灾的影响, 印度如今的日检测量以及规模都比较小, 加上检测试剂等物资不足, 这些都使得印度无法开展全国性的疫情排查。巴西总统在疫情期间仍然鼓励民众继续上街, 一方面由于巴西医疗水平较差, 医疗设施落后, 不足以支撑大规模的核酸检测, 另一方面是巴西属于出口型国家, 受到疫情的冲击, 经济发展困难难以支撑国家福利。

5.2 建议

中国目前绝大部分省市已经实现稳定有序的复工复产, 这对全球各国如何度过疫情时期以及实现科学有序的复工复产都具有十分重要的借鉴意义。结合分析结果, 建议部分只对处在疫情不同发展阶段的国家提出是否可进行复工复产的建议, 由于各国国情的不同, 没有具体的复工复产建议。

对于像日本这样疫情已经得到基本控制的国家, 可以在新增确诊实现零增长并度过一段时间的潜伏观察期, 分行业有序科学实现复工复产, 但仍要警惕境外输入以及做好集聚场所, 如学校、看守所、养老院等防控工作, 防止疫情出现再次爆发。对于像德法意此类的国家疫情开始出现缓和的国家, 也不要放松警惕, 继续做好疫情的防控工作, 民众也要积极配合政府的抗疫政策, 配合隔离措施, 既可避免疫情的再次爆发也为后面复工复产的实现奠定良好的基础。而对于英美加以及印巴此类国家, 尤其是美国, 建议执行强有力的疫情防控政策, 在疫情未得到有效控制之前, 不建议实行复工复产。

6 致谢

弹指间一学期即将过去,在此次课程中我们都收获颇多。首先,感谢学院为我们安排这门课程。其次,感谢闫军老师为我们带来了全新的工具,我们从一个小白一步步入门了 git、github 以及 r Markdown 等工具的使用;同时,在学习这门课的过程中,极大的锻炼了我们自主学习和解决问题的能力,这将是我们未来学习生活和职业生涯中及其宝贵的财富。最后,感谢合作伙伴的配合以及同学们的讨论和建议,希望我们在未来都能够正视自己的不足、不断的进取。

贡献声明

- (1) 刘崑 (组长): 负责报告中的第二部分描述性统计分析、第六部分致谢以及报告的整理汇总。
- (2) 慕慧君: 负责报告中的摘要以及第四部分趋势拟合及预测。
- (3) 肖雨婷: 负责报告中的第三部分聚类分析中聚类分析简介以及第五部分总结与建议。
- (4) 王洁: 负责报告中的第一部分引言以及参考文献。
- (5) 卢晋萍: 负责报告中的第三部分聚类分析中的实证分析。

参考文献

- [1] Christian Ritz,Jens Carl Streibig.Nonlinear Regression with R.SpringerLink.2008.Preprint at:<https://link.springer.com/book/10.1007/978-0-387-09616-2>
- [2] Ke Wu,Didier Darcet,Qian Wang,Didier Sornette.Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world.medRxiv.2020.Preprint at: <https://www.medrxiv.org/content/medrxiv/early/2020/03/16/2020.03.11.20034363.full.pdf>
- [3] 杨吉安, 颜忠诚. 基于生态模型的新型冠状病毒肺炎疫情分析 [J/OL]. 首都师范大学学报 (自然科学版):1-12[2020-05-21]
- [4] 易大莉, 李高明, 冷怀明. 新型冠状病毒肺炎疫情发展区域差异的聚类分析 [J/OL]. 重庆医科大学学报:1-6[2020-05-27]
- [5] 武文韬, 柏如海, 李达宁, 冯敖梓, 徐安定, 吕军. 广东省新型冠状病毒肺炎疫情流行趋势的初步预测 [J]. 暨南大学学报 (自然科学与医学版),2020,41(02):181-185.
- [6] 杨雨琦, 孙琦, 王悦欣, 严雪凌, 乐涛. 重庆市新型冠状病毒肺炎 (NCP) 疫情分析与趋势预测 [J]. 重庆师范大学学报 (自然科学版),2020,37(01):135-140.
- [7] 张荣强, 刘辉, 李凤英, 李向文, 刘启玲, 李星慧, 孙娜, 张蓓, 史传道. 全球 2019 冠状病毒病 (COVID-19) 与甲型流感 (2009-H1N1) 疫情初期的流行病学对比分析及中医药防治策略研究 [J/OL]. 陕西中医药大学学报:1-7[2020-05-29].