



大学生创新创业训练计划项目

结题报告

项目级别： 国家级 省级 校级

项目编号： X202210359482

项目名称： 新型冠状病毒肺炎传染模型

项目负责人： 朱思雯

项目组成员： 王悦妍、邓颖、姚佳婧、罗天曦

起止年月： 2022年4月-2023年4月

指导教师： 张神星

所在学院： 数学学院

合肥工业大学

一、基本情况

项目级别	<input type="checkbox"/> 国家级 <input type="checkbox"/> 省级 <input checked="" type="checkbox"/> 校级			项目编号	X202210359482
项目类别	<input checked="" type="checkbox"/> 创新训练 <input type="checkbox"/> 创业训练 <input type="checkbox"/> 创业实践			成果形式	软件程序
项目名称	新型冠状病毒肺炎传染模型				
负责人	朱思雯	学院	数学学院	联系电话	13696576396
指导教师	张神星	学院	数学学院	职 称	副教授
起止时间	2022 年 4 月至 2023 年 4 月			资助经费	1000 元
参加学生信息(包括负责人)					
姓名	学号	学院	专业班级	项目分工	
朱思雯	2021211683	数学学院	统计学 01 班	主持组织项目, 联系指导老师	
王悦妍	2021211660	数学学院	统计学 01 班	整合信息, 构建模型	
邓颖	2021211663	数学学院	统计学 01 班	整合信息, 构建模型	
姚佳婧	2021211677	数学学院	统计学 01 班	查询资料, 查找处理数据	
罗天曦	2021211672	数学学院	信息与计算科学 01 班	查询资料, 工作记录	
一、项目执行情况					
1、本项目的选题背景、目的与意义					
<p>新型冠状病毒肺炎疫情自 2019 年末起, 在全球多个国家和地区爆发并迅速蔓延, 对人类社会产生了严重影响。新冠肺炎疫情的不断反复, 严重影响了人们正常的工作和生活。由于新冠病毒具有传染性强、传播速度快、流行范围广、变异种类多、难以预测和防范等特点, 研究它在变化过程中的关键因变得很有意义。这样的研究将可帮助人们回答在面对类似的流行病时, 如何在尚在局部传播时就检测、预测、控制它, 如何在短时间内确定病原体的种属, 以及如何制定相应的防疫政策。因此我们提出了本项目。</p> <p>本项目将通过对已有的数据进行推演, 模拟出病毒流行的过程、预测地区感染人数的变化趋势, 以及评估防疫措施的有效性, 来预先了解某地区新冠肺炎疫情的发展情况和发展趋势, 为相关部门的决策提供参考, 为实现精准防控和病毒救治争取宝贵的时间, 更加有效地防控疫情, 控制感染人数, 为我国抗击新型冠状病毒肺炎注入冷静和信心, 也为其他传染病模型的建立提供参考, 对其他传染病进行理论和数据研究分析, 使该模型对大部分的传染病具有普遍适用性。</p>					
2、项目的创新点与特色					
<p>新型冠状病毒是 RNA 病毒, 由于 RNA 是单链遗传物质, 因此变异几率比较高, 目前已有的新型冠状病毒肺炎传染模型主要是早期的变异毒株, 如阿尔法、贝塔、德尔塔等, 我们的项目结合了新冠病毒新变异株奥密克戎, 对模型进行了调整, 使整个新型冠状病毒肺炎传染模型更加的完善, 以得到更准确的测试结果。</p>					

3、团队成员分工和合作情况（团队成员分工和合作情况，成员在项目研究中取得的成绩、发挥的作用以及在能力培养和素质提高方面的体验和收获。2000 字左右,附件另附）

本项目团队由王悦妍、邓颖、朱思雯、姚佳婧、罗天曦组成。自项目确定立项以来，为如期完成项目研究目标，团队进行了如下分工：

- (1) 项目初期，队长朱思雯负责协调团队内部工作，统筹推进项目进程。
 - 邓颖、朱思雯共同进行 Python 的基础学习，包括基础语法、数据组合类型等知识，为日后使用该编程语言建立一个具有一定适用性的新型冠状病毒肺炎传染模型打好基础。
 - 王悦妍、姚佳婧开始学习常见的传染病模型和参数估计方法，了解建立模型的过程，通过对不同模型比较，选定较好的模型去建立新型冠状病毒肺炎传染模型。
 - 罗天曦查阅和整理相关资料。
- (2) 项目中期，在学习一定的基础知识后，本队开始着手构建新型冠状病毒肺炎传染模型，持续跟进相关技术的学习与编程实现，详细阅读相关文献。
 - 王悦妍、邓颖了解和学习 SEIR 模型微分方程的建立及其转化迭代形式，并将其转化为数据算法，学习了 Python 语言的基础知识和可视化的相关内容，了解传染病模型的相关 Python 代码，利用 Python 语言编写出可运行的模型代码、对收集的实际数据进行可视化，同时对比该模型与实际情况以便对模型进行进一步的调整。
 - 朱思雯、姚佳婧搜寻相关数据，包括特定省市等地区一段时间内累计确诊、当前患病、累计治愈和死亡、累计易感的人数数据，并进行初步整合和处理。
 - 罗天曦继续查找与本项目有关的基础理论、书籍和文献。
- (3) 项目后期，本团队积极与指导老师沟通，推进模型建立过程。
 - 邓颖、王悦妍多次对得到的参数数据做曲线拟合，利用得出来的参数与时间的函数代入构建的模型中，来预测患病人数，并将预测值和实际数据进行对比，并对结果进行分析，不断调整预测，使预测更加接近于真实值，得出最终的新型冠状病毒肺炎传染模型。
 - 朱思雯、姚佳婧对所收集的数据进行处理，计算出各项比率，包括发病率、治愈率、暴露率等，最后进行整合处理，同时撰写相关报告。
 - 罗天曦协助工作并搜集整理相关论文，进行记录，并负责工作记录本的撰写。

合作情况：本项目团队成员在项目过程中经历了组队选题、项目申报书撰写、项目立项、基础设计方案形成、早期数据收集、中期报告撰写、数据计算及分析、参数拟合、构建新型冠状病毒肺炎传染模型、结题报告撰写等多个阶段的一系列任务。本团队进行了模块化实践实施计划，合理分工完成相应任务，有效地推进了项目进程。团队成员在项目过程中根据各自优势进行了精细化的分工，以期将团队效益及团队竞争力最大化。团队成员在日常学习过程中积累相关经验，积极参与创新创业课题，遇到问题相互沟通，共同克服难题，最终获得圆满成功。

本团队成员各司其职，共同合作，每个人都参与进来，发挥巨大作用，缺一不可。团队成员同属于同一个专业，在一起的时间长，有利于研究项目的讨论与进行，且团队知识能力互补。这次项目使我们认识到要加强学习，才能使我们建立创新理念，接受新思维、新举措，不断创新，跳出思维定势，拓展思维视角，改变自己传统的一些思维习惯，用新的创新思维方法创造性的开展工作、建立模型。认识到加强学习，是我们进一步提高工作能力的需要。只有通过加强学习，才能了解和掌握先进的理念和方法，取他人之长补己之短。我们从被动学习向主动学习转变，克服学习上的懒惰性，尽可能地获得各方面的知识和信息，以适应需要。我们积极主动地学习新知识、掌握新技能、学习建立模型的各种方法并成功构建了新型冠状病毒肺炎传染模型，不断提高

自身的综合能力和素质，更好地完成各项工作。时间虽短，但我们学到了很多，收获颇丰。

我们从 2022 年 4 月开始申请大学生创新创业训练计划项目，指导老师是张神星教授。到现在已经过去了一年多的时间了，我们从最开始的迷茫不知所措到现在成功做出模型。我们感觉其中发生了许多事情，自己也收获了许多。回想过去参加研究项目的过程，从开始的寻找课题到申请立项撰写项目申请书，到查阅相关参考文献，确定实验项目、实施方案和寻找创新点；并制定详细的研究方案和步骤；对项目进行相关调查和研究，到最后确定项目的可行性。一步步走来，这其中的辛苦和辛酸只有经历过的人才懂，其中的经验和成长也只有经历过的人才分享和拥有。这是一次难得的经历，一次让我们得到锻炼和成长。其实这段时间对这个项目的付出和努力，让我们觉得收获最大，体会最多的应该是团队合作方面。一个人不管怎样都是需要团队协作才能把事情做得漂亮，也许一个人也能完成，但是可能要花上十倍百倍的时间，而且完成的可能也没有那么漂亮。“众人拾柴火焰高”，团队合作中需要我们成员间的不断磨合，学会倾听大家的意见和分享你的看法，做到尊重每一个组员，开心地交流与合作。所以很感谢这个项目让我们懂得了许多在平常学习中学不到的知识与能力。

4、项目成果（论文发表、竞赛获奖、专利/著作权、注册公司、参加会议、实物制作等情况）

无

二、研究总结报告（4000 字左右）

1、预定计划执行情况、项目研究和实践情况（含进度安排、完成内容、关键技术及效果等）

预定计划执行情况：

(1) 预期成果

建立完整的新型冠状病毒肺炎传染模型，通过对一般情况下受感染人数的变化规律的分析，在该模型的基础上对当前以及未来新冠疫情的发展趋势进行一定的预测，从而为各种防疫措施提供有效参考，找到恰当的防范方式来为我国的抗击疫情做出一定贡献。

(2) 实际成果

建立了完整的新型冠状病毒肺炎传染模型，该模型可通过对不同病原体传染情况实际数据进行拟合，以得到相对应的参数，对感染者人数进行预测。通过调查和收集上海市已有的新冠感染者人数，对现有报道的上海市感染新冠的人数（包括无症状感染者）进行拟合估计，得到相应的参数。将参数代入已构建的新型冠状病毒肺炎模型中，通过模型进行运算，找出新冠感染人数变化的规律，预测新型冠状病毒肺炎的发展，并通过调整参数，观察预测曲线趋势，以此来比较各种防疫措施的效果，以有效减少可能感染的人数。

(3) 进度安排

总体上首先了解学习了各类传染病模型，并学习借鉴这些模型的优点，分析这些模型需要改进的方面，提出模型的基本设想和需注意的方面。学习利用 Python 语言，基于已有的传染病模型，建立一个具有一定适用性的数学模型新型冠状病毒肺炎传染模型。

2022 年 4 月到 2022 年 6 月，团队成员查找学习与传染病模型相关的基础理论、文献及书籍，了解相关的一系列传染病模型，并对当前疫情的形势有初步了解；对 Python 的基础语法进行学习，为日后使用该编程语言建立新型冠状病毒肺炎传染模型打好基础。详细阅读相关书籍及电子文献，根据团队成员专业方向分工学习并掌握相关专业知识。

2022年7月到2022年8月，团队持续跟进相关模型的学习和编程实现，对传染模型的相关信息进行整合处理，了解相关模型的方程的确立，并深入学习研究了SEIR模型的微分方程的建立及其迭代形式，大致将SEIR模型的微分方程转化为差分方程，以便利用Python建立模型；进一步学习Python语言的基础知识，同时学习了解可视化的相关内容，另外还了解学习传染病模型的相关Python代码，并对其中的算法、结构有了初步了解。

2022年9月到2022年11月，团队收集并筛选统计可用的疫情数据建立新冠肺炎传染病模型，利用Python语言编写出可运行的模型代码、对实际数据进行画图，同时对已有参数做估计和曲线拟合以方便后面进一步调整模型。

2022年12月到2023年3月，我们积极与指导老师沟通，推进模型建立过程。成员对所收集的数据进行处理，计算出各项比率，包括发病率、治愈率、暴露率等，最后进行整合处理，多次通过已有数据提取出所需数据并用不同方法对已有参数做估计和曲线拟合，将得出来的参数代入构建的模型中，来预测患病人数，并将预测值和实际数据进行比较，得出最终的新型冠状病毒肺炎传染病模型。团队首先将参数视为常数利用最小二乘法对于新型冠状病毒肺炎传染病模型中参数作估计，拟合效果不尽人意，于是重新整理数据计算出每天的所需参数，将参数视为关于时间的函数，观察数据的分布，选择适当的函数对参数进行拟合，进而得到预测参数，将参数代入已构建的新型冠状病毒肺炎模型中，通过模型进行运算，通过图像观察新冠肺炎感染者人数的变化趋势。同时着手准备相关报告撰写，查阅相关行业及技术文献，综合并提炼核心内容形成报告文稿。

项目研究：

经典SEIR模型常适用于具有潜伏期的传染病传播研究，通过把人群划分为易感人群(S)、潜伏人群(E)、确诊人群(I)、移出人群(R)来构建传播微分方程，并假设整体人数不变，不考虑出生和自然死亡，迁入迁出来求解微分方程中的传染率参数，再代入传播微分方程以此刻画传染病的传播过程。其中：易感人群是指未感染传染病且没有抗体的健康人群；潜伏人群是指已经被传染病感染但未有发病症状且不具有传染能力的人群；确诊人群是指出现传染病发病症状并具有传染能力的人群；移出人群则是指治愈和因感染病毒死亡的患者人群。根据新冠病毒传播机制，建立模型，该模型的将人划分为易感者、无症状感染者（暴露者）、感染者、移出者。

模型假设：

易感者与患病者有效接触即变为暴露者，暴露者经过潜伏期后成为患病者，患病者可被治愈成为移出者，移出者终身免疫不再感染。

以一天作为模型的最小时间单元。总人数为 N ，不考虑人口的出生与死亡，迁入与迁出，此总人数不变。将 t 时刻各类人群占总人数的比率分别记为 $s(t)$ 、 $e(t)$ 、 $i(t)$ 、 $r(t)$ ，各类人群的数量为 $S(t)$ 、 $E(t)$ 、 $I(t)$ 、 $R(t)$ 。

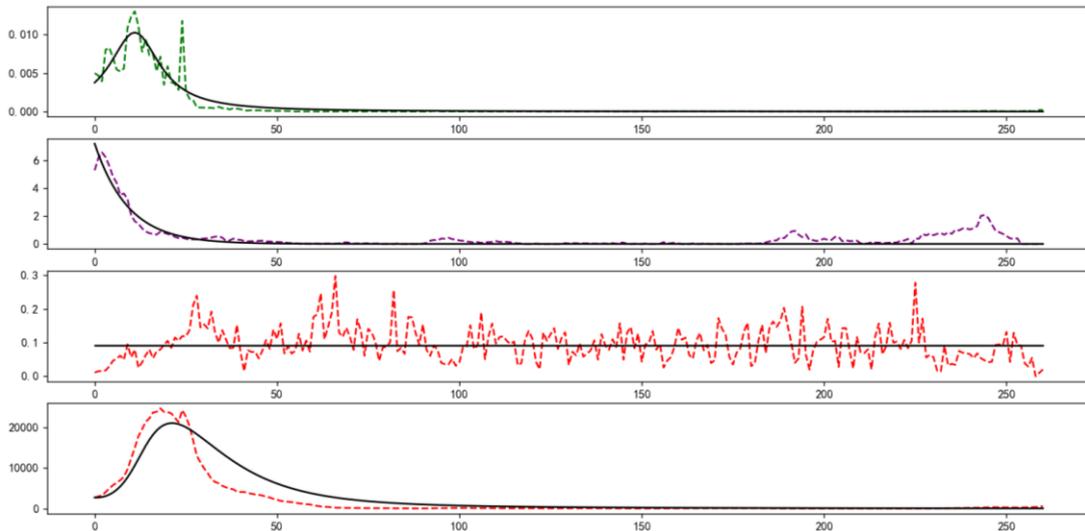
- 初始时刻 $t=0$ 时，各类人数量所占初始比率为 s_0 、 e_0 、 i_0 、 r_0 。
- 日暴露数 λ ，即每个患病者每天有效接触的易感者的平均人数。
- 日发病率 δ ，即每天发病成为患病者的暴露者占暴露者总数的比率。
- 日治愈率 μ ，即每天被治愈的患病者人数占病人总数的比率。
- 平均治愈天数 $1/\mu$ ，又称平均传染期，即从患病到治愈的天数。
- 传染期接触数 $\sigma = \lambda/\mu$ ，即每个患病者在整个传染期 $1/\mu$ 天内，有效接触的易感者人数。

根据模型假设，每个病人每天可使 $\lambda \cdot s(t)$ 个易感者变为暴露者，且患病者人数为 $N \cdot i(t)$ ，所以每天有 $\lambda \cdot s(t) \cdot N \cdot i(t)$ 个易感者变为暴露者。每天的暴露者 $N \cdot e(t)$ 中，又有 $\delta \cdot N \cdot e(t)$ 发病成

为患病者。每天的患病者 $N \cdot i(t)$ 中，又有 $\mu \cdot N \cdot i(t)$ 被治愈成为移出者。

完成内容：

我们通过对已知数据的分析、计算和比较来确定模型中的参数，建立了新冠肺炎传染病模型。具体而言，我们利用上海地区 2022 年 4 月 4 日（第 0 天）至 2022 年 5 月 10 日（第 39 天）的数据来进行参数估计。通过对已知数据计算出每天实际参数值，然后利用曲线拟合参数值关于时间的函数，以求得参数值的估计，再将其所对应的函数代入模型中进行预测。最后利用 Python 语言编写出可运行的模型代码，对实际数据和预测数据进行画图。



上图中第一个图是发病率，其中虚线为实际数据，实线为拟合的曲线；第二个图为暴露数，其中虚线为实际数据，实线为拟合的曲线；治愈率使用均值进行估计，第三个图是确诊患者的人数，其中虚线为实际数据，实线为拟合的曲线。

关键技术：

在新型冠状病毒传播初期收集到少量数据的基础上，建立模型，预测以后疫情的发展情况，故只用一部分数据进行参数拟合，而用其余数据作检验。

先使用已知数据算出每天所需参数，将得到的真实参数数据画出散点图后，寻找与真实参数数据分布基本相同的函数，然后用这些函数来拟合，再利用最小二乘法得到函数中的参数，用得到的函数来拟合参数随时间的变化，模型中的参数一般分为常量参数和变量参数，常量参数利用平均值确定，变量参数通过曲线拟合的方法确定。在对数据处理中， μ 表示治愈率， δ 表示发病率， λ 表示暴露数。

$$\mu = \frac{\text{新增治愈人数} + \text{新增死亡人数}}{\text{前一天的当前患病人数}}$$

$$\delta = \frac{\text{新增确诊人数}}{\text{前一天的当前无症状人数}}$$

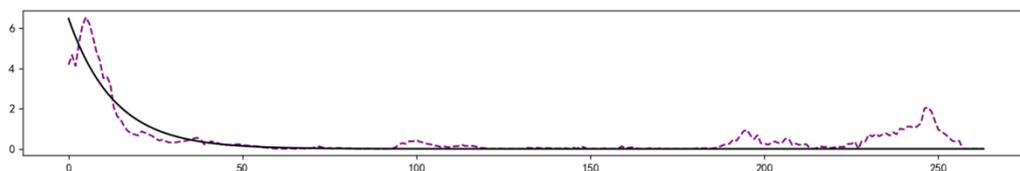
$$\lambda = \frac{\text{新增无症状人数}}{\text{前一天的累计确诊人数} + \text{前一天的累计治愈和死亡人数}}$$

取时间单位为天，得递推关系为：

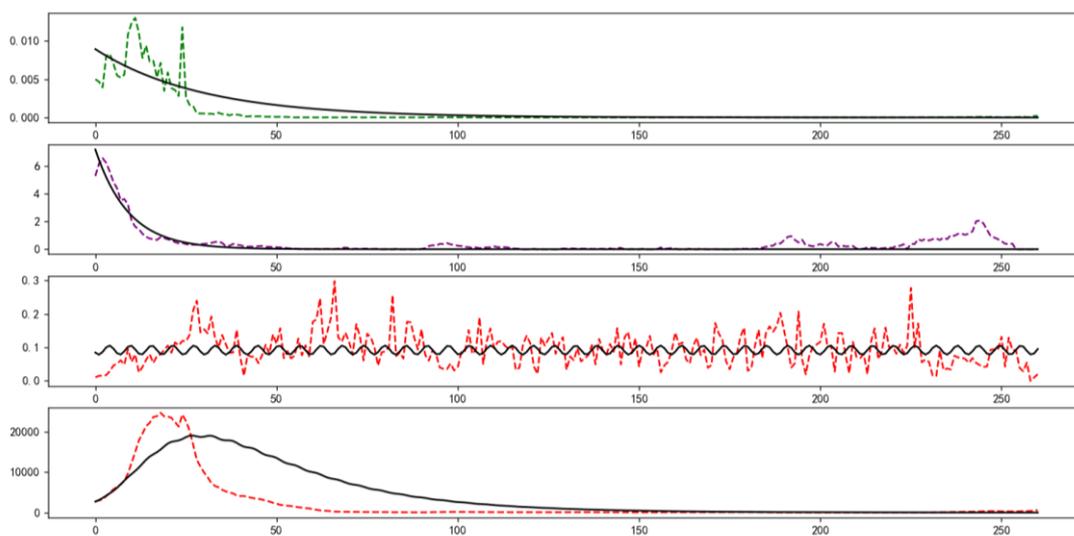
$$\begin{aligned}
 S(t+1) &= S(t) - \lambda(t) S(t)I(t)/N \\
 E(t+1) &= E(t) + \lambda(t)S(t)I(t)/N - \delta(t)E(t) \\
 I(t+1) &= I(t) + \delta(t)E(t) - \mu I(t)
 \end{aligned}$$

其中：

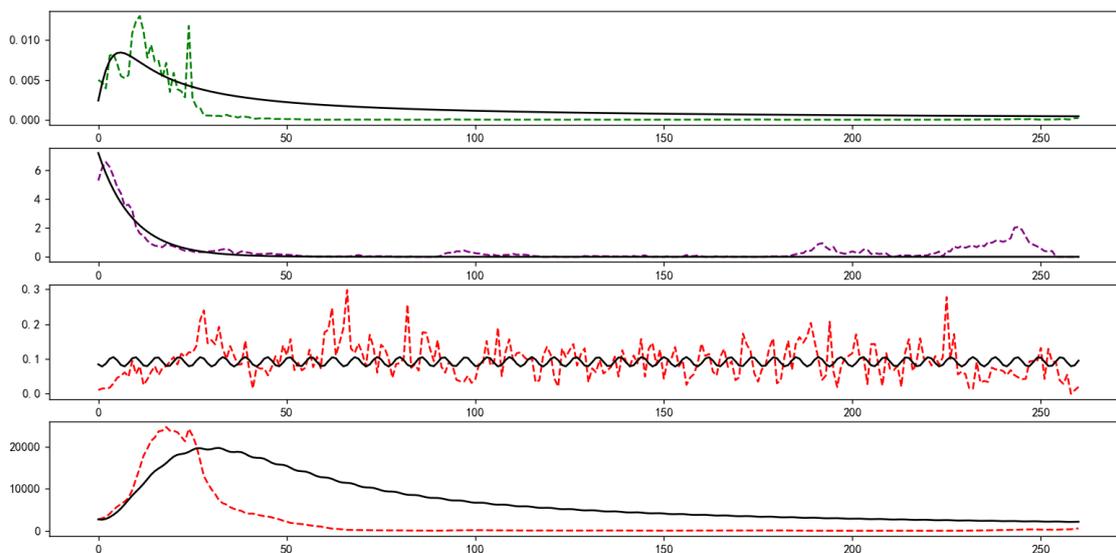
(1) λ 用 $a_1 e^{-a_2 t}$ 拟合，效果如下：



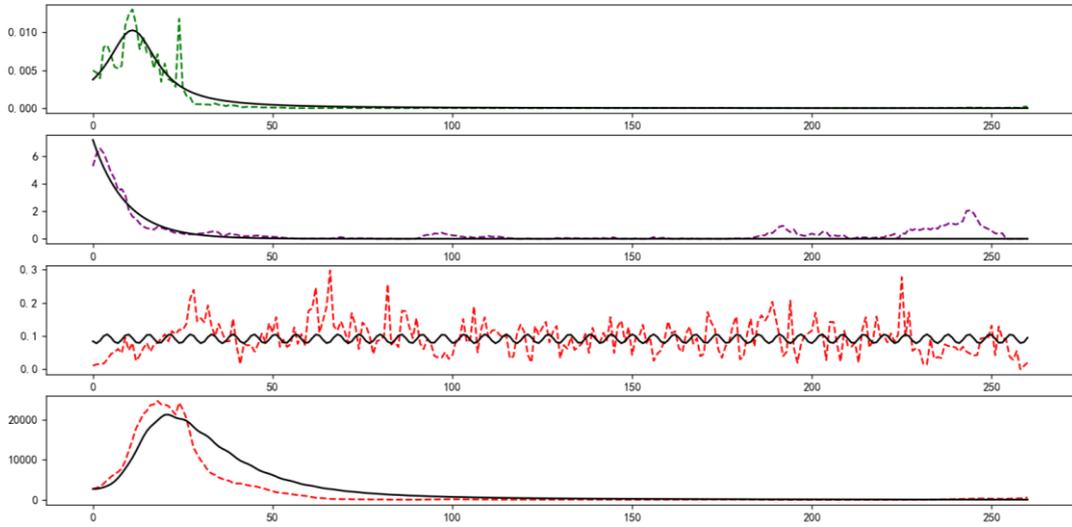
(2) δ 用 $e^{-b_2 t}$ 拟合，效果如下：



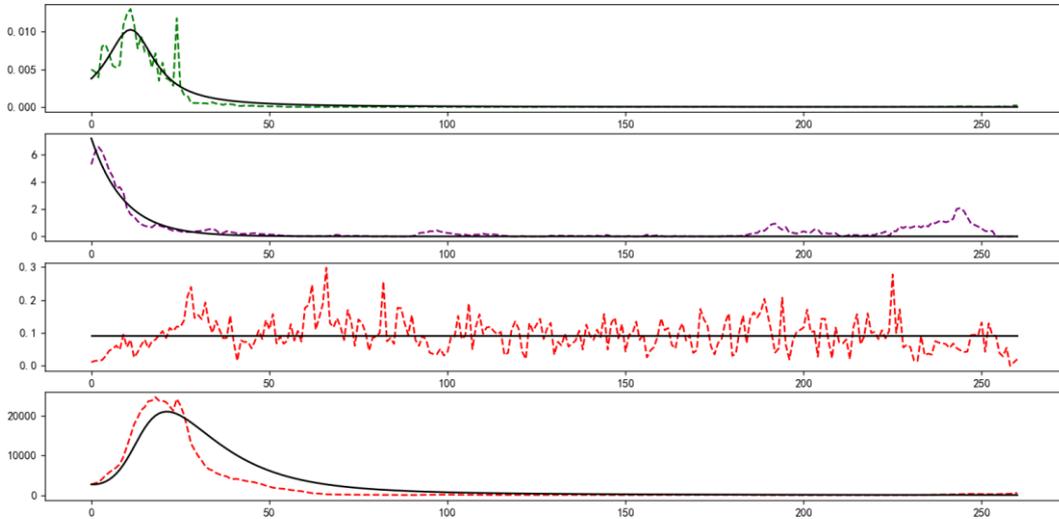
用 $\frac{b_3(t+b_4)}{b_2+b_1(t+b_4)^2}$ 拟合，效果如下：



用 $\frac{b_1}{b_4+b_2(t+b_3)^2+t}$ 拟合，效果如下：

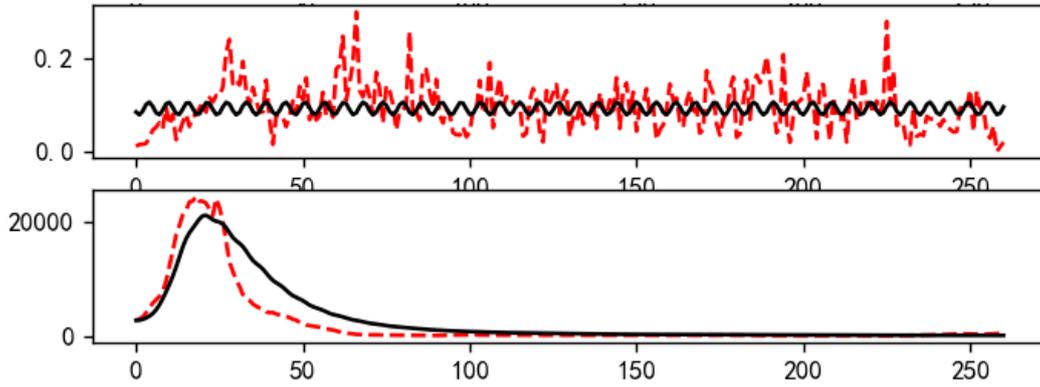


用 $\frac{b_1}{b_4+b_2(t+b_3)^2}$ 拟合，效果如下：

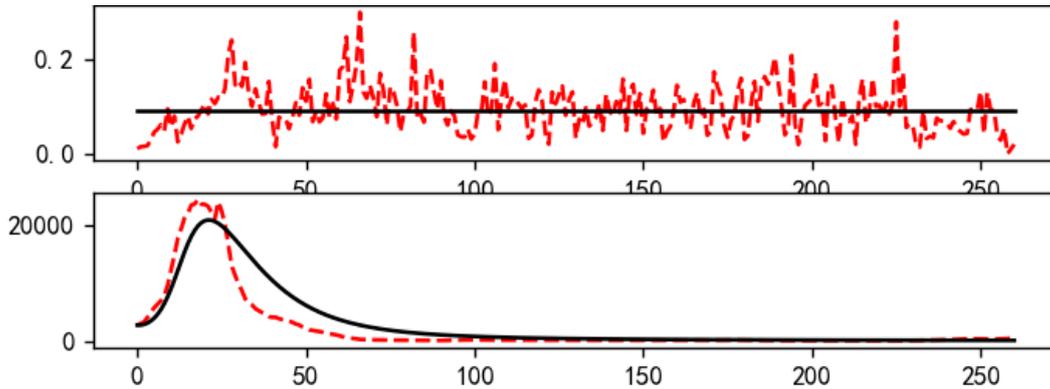


最终取拟合效果较好的 $\frac{b_1}{b_4+b_2(t+b_3)^2}$.

(3) 由于真实数据分布与周期函数有些相近，故希望通过对于三角函数的加减运算以及调整振幅，得到一个与真实数据分布拟合较好的函数，故 μ 先用 $c_1 \sin(c_2 t + c_3) + c_4$ 进行参数拟合，效果如下：



因为真实数据的分布特性，使用均值进行估计，效果如下：



比较两个预测结果，使用 $c_1 \sin(c_2 t + c_3) + c_4$ 进行参数拟合预测效果要更好一些，故最终选定使用 $c_1 \sin(c_2 t + c_3) + c_4$ 对 μ 进行预测。

总结一下： $\lambda(t)$ 的含义是第 t 天每个新冠肺炎感染者传染的人数， $\delta(t)$ 是第 t 天新冠肺炎患者的发病率， μ 是新冠肺炎患者每天治愈和死亡率，它们确定的方法分别是：

$$\lambda(t) = a_1 e^{-a_2 t}, \quad \delta(t) = \frac{b_1}{b_4 + b_2(t + b_3)^2}, \quad \mu = \frac{1}{t + 1} \sum_{i=0}^t \mu_i.$$

2、项目实施的收获和体会，项目工作有哪些不足，有哪些问题尚需深入研究，项目工作中的困难、问题和建议

收获和体会：

我们在刚实施创新训练项目时，每个人都没有接触过这方面的知识，对创新训练项目简直一头雾水，但是这阻止不了我们前进的步伐。前期，我们获取理论知识，即阅读书籍，每个人找到自己负责的模块，到图书馆查找书籍，深钻研，以最快的速度进入角色。有时候，思维可能会出现“停滞不前”的现象，我们就去学习相关专业的理论知识。通过理论学习，我们的水平有了巨大的提高，这也成为我们项目的不竭的动力。我们在项目实施过程中，抓住根本，团队成员各司其职，有效衔接，分工明确，责任到人。虽然我们在合作期间也会有意见不一致的时候，但我们会及时交流，及时调整。

通过这一年的研究，我们深深感受到指导老师和团队合作的重要性和必要性，同时我们也从中获得了锻炼和成长，学习能力的提高扩展了我们的思维能力；共同的学习目标培养了我们的团队精神；学习的毅力铸就了我们执行的坚韧性。非常感谢指导老师，也感谢团队成员不计回报的

付出。

项目工作的不足：

由于我们利用的是较为简单的 SEIR 模型，考虑因素不多，在后期也没有加入更多参数，没有加入对国家防控措施和个人防护措施、人口流动、天气等因素的考虑，所以最终的模型拟合效果与实际情况存在一定差距，对数据的处理比较粗糙。

尚需深入研究的问题：

在疾病的传播过程中，对流行状况影响很大的是再生数，即一个病人在患病期间平均感染的人数，这在流行病传播中是个十分重要的参数。要将一个传染病控制住，实际上就是要通过各种措施使得再生数小于 1。

(1) 在实际应用中不断增加和更新数据进行短期的预测，预测的误差就会减少。

(2) 还可以考虑更多因素包括媒体宣传对心理和行为的影响、公共卫生水平、控制力度、平均气温、地区差异、人口流动、职业分布、年龄分布、医院内外的传播和超级传播者等。考虑到新型冠状病毒与一般传染病不同的是有着相当高的突变率。

工作中的困难、问题和建议：

最初，我们想利用最小二乘法对于模型中参数做拟合，但是拟合出来效果并不好。开始建模之后，我们先从已总结好的数据中提取出所需数据，并用 Python 函数 `xlrd.open_workbook()` 导入数据，但是程序一直显示出错。在我们反复调试并更改 Excel 格式后，成功读取了数据。然后我们用 `append` 函数将所需数据加入列表中，选了 80 天的数据先尝试求解参数。其中求 μ 的方程为：

$$\sum_{t=0}^{79} \left(R(t) - R(0) - \mu \sum_{k=0}^{t-1} I(k) \right)^2 = S, \quad \frac{\partial S}{\partial \mu} \sum_{t=0}^{79} \left(R(t) - R(0) - \mu \sum_{k=1}^{t-1} I(k) \right)^2 = 0.$$

安装求解方程的函数所在的程序解出 μ ，然后求解 δ 和 λ/N

$$\sum_{t=0}^{79} \left(I(t) - (1 - \mu)^{t-1} I(0) - \delta \sum_{k=0}^{t-1} (1 - \mu)^{t-1-k} E(k) \right) = 0,$$

$$\sum_{t=0}^{79} \left(E(t) - (1 - \delta)^{t-1} E(0) - \frac{\lambda}{N} \sum_{k=0}^t (1 - \delta)^{t-1-k} S(k) I(k) \right) = 0.$$

然后导入了数据对预测结果作检验，代入初值求出了预测值，实验结果与实际情况偏差很大。最后同时改变模拟方式才取得了较好的模拟效果。

三、经费开支与报销情况

目前，经费使用了 450 元，主要用于购买相关书籍，搜集资料，已完成报销。购买的书籍有：

- 《流行导论（第二版）》
- 《利用 Python 进行数据分析》
- 《统计学：从数据到结论》
- 《数学之英文写作》
- 《数学模型（第五版）》
- 《数学模型（第五版）习题参考答案》

导师意见：

通过整个研究过程，小组成员的学习和探索能力得到了长足的进步。项目完成达到了预期效果。

签名：

年 月 日

学院评审意见：

单位（盖章）：

负责人签字：

年 月 日

学校专家组评审意见：

专家组组长签字：

年 月 日

学校审批意见：

主管部门（盖章）：

负责人签字：

年 月 日

注：结题报告（一式一份）完成后与《合肥工业大学大学生创新创业训练计划项目工作记录本》、成果证明材料由学院统一存档。