

病毒潜伏期影响因素分析

——基于对数正态分布

王文庆

青海师范大学数学与统计学院, 青海 西宁

收稿日期: 2024年6月1日; 录用日期: 2024年6月21日; 发布日期: 2024年6月30日

摘要

鉴于传染病、疫情等突发性疾病的严峻性, 其出现无疑会对人们的生产生活造成极其严重的冲击。若控制不力, 将引发医疗资源紧张、经济质量下滑等一系列关乎生存与发展的问题。因此, 对发病源的规律性进行深入探究显得尤为重要。通过掌握疾病的发展规律, 医者能够及时制定相应医疗方案, 实施有效干预, 进而遏制疾病的进一步扩散, 控制医疗成本。自2019年COVID-19疫情爆发以来, 该病毒已对全球的经济、卫生、医疗等多个领域造成了巨大冲击, 同时也对人类的生存安全提出了严峻挑战。本文以国家卫生健康委员会公布的各省市感染人数的真实数据为基础, 经过严格的信息完整度筛查, 依据病源仅暴露一次、暴露源清晰、暴露信息记录完整的标准, 筛选出78位患者的病历数据作为研究样本。通过深入分析, 探索了新型冠状病毒的潜伏期天数, 结果表明, 该病毒在人体内的平均潜伏期约为16天。此外, 研究还发现, 潜伏期长短与患者的性别、年龄无显著关联, 但与已感染群体的接触方式和病毒的地域浓度密切相关。

关键词

病毒潜伏期, 描述统计, 对数正态分布, 方差分析

Analysis of Factors Affecting Incubation

Period of Virus

—Based on Log-Normal

Wenqing Wang

School of Mathematics and Statistics, Qinghai Normal University, Xining Qinghai

Received: Jun. 1st, 2024; accepted: Jun. 21st, 2024; published: Jun. 30th, 2024

Abstract

Given the severity of infectious diseases, pandemics, and other sudden illnesses, their emergence

文章引用: 王文庆. 病毒潜伏期影响因素分析[J]. 统计学与应用, 2024, 13(3): 943-953.

DOI: 10.12677/sa.2024.133096

undoubtedly poses a grave impact on people's lives and production. If not controlled effectively, it will lead to a series of issues concerning survival and development, such as the strain on medical resources and a decline in economic quality. Therefore, it is crucial to conduct in-depth research on the regularity of the source of the disease. By grasping the development pattern of the disease, doctors can promptly formulate corresponding medical treatment plans, implement effective interventions, and thereby curb the further spread of the disease and control medical costs. Since the outbreak of the COVID-19 pandemic in 2019, the virus has had a tremendous impact on various fields such as global economy, health, and medical care, while also posing severe challenges to human survival and safety. Based on the real data of the number of infections in various provinces and cities announced by the National Health Commission, this article has strictly screened for information completeness and selected 78 patients' medical record data as research samples according to the criteria of only one exposure to the source of the disease, clear exposure source, and complete exposure information records. Through in-depth analysis, the latent period of the novel coronavirus was explored, and the results showed that the average latent period of the virus in the human body is approximately 16 days. Additionally, the study found that the length of the latent period was not significantly associated with the patient's gender or age, but was closely related to the mode of contact with infected individuals and the regional concentration of the virus.

Keywords

Incubation Period of Virus, Descriptive Statistics, Log-Normal, ANOVA

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在二零一九年的十二月份，中国湖北省武汉市由于不明原因的肺炎疫情而按下了暂停键，全市所有人员居家静默，非必要不得流动，务工人员不得返乡等规定随之到来，全国上下所有成员都为之担忧，随着专家组的不断介入与调查，COVID-19 的真实面貌也被人们所认识，它是一种由新型冠状病毒感染导致的肺部炎症，此次疫情的病毒毒株是人类历史上从未出现过的，因此对它的病理研究是十分必要的。

病毒感染人体的时候并不会马上就使机体免疫做出相应的免疫应答，换句话说它并不会立刻让受感染的机体察觉到它的存在，由于病毒进入人体的时候其信息载体是核糖核酸，病发前需要一定的时间在宿主体内进行转录、繁殖，当宿主机体病毒载量达到一定限度时其免疫系统才会发觉病原体的存在，继而引起免疫系统的一系列反应，与此同时为了保证免疫系统的正常运行，机体感觉器官会发生异常，也就是说此时病理状态被揭露出来了。此次疫病原作为病毒家族的一员也具有相同的暴露历程，医学上将这种机体内有病原体进入，但机体并不会立刻暴露病理状态的时间段叫做病原体的潜伏期，由于此次疫情病毒具有自然疫源性并且具有与一些流感病毒的相似身体反应特征，以及由于该病毒复制快、合胞体嗜性广、在宿主体内存活能力强等原因，使其传播速度极快、变异能力极强，从而导致人们无法及时对传染源进行识别切断，也因此此次疫情对国家多个层面都受到了巨大的冲击，遇上了前所未有的挑战。因此对该病毒的病理研究在帮助人们把握该病毒特性，减少相关损失上是格外重要的，也是必要的。

迄今为止对于病毒潜伏期的研究已经有很多医务工作者，流行病学专家在重多领域根据病患临床表现对其进行了相应的探索式研究，早期在新型冠状病毒的研究中郭德银等人[1]就如何应对突发性疫情做了相关分析，并对早期突发性疫病防控提供了宝贵的建议，还有很多医学工作者根据患者的临床表现对 COVID-19 病毒引起的疾病进行了诊断性分析，并对此病毒引起的疾病的治疗赵芳等人[2]，邓江霞等人

[3]; 沈燕等人[4]提供了宝贵的意见, 也有学者应用数学的方法, 赵艳婷[5]将微分方程对新冠病毒的动力型传播机制进行了研究, 并利用 RIOTS 求解了分数阶的最优控制问题并对 COVID-19 关于时效性影响的疫苗接种建立了 FO-SEIR 模型, 实现了疫苗接种的控制最优化。还有学者将病毒研究与机器学习理论集合起来, 比如任建强等[6]在预测新型冠状病毒肺炎疫情未来发展趋势的研究中提出了三步预测模型, 将神经网络、随机森林、长短期记忆网络和序列到序列的机器学习算法引入到新冠肺炎传染病疫情预测中, 预测了未来的确诊人数和实际感染规模。后继由甘雨等人[7]提出了基于长短期记忆网络的病毒传染率预测方法, 该方法被证明优于传统 SEIR 模型, 此方法的提出与完善为 COVID-19 疫情的发展趋势上的预测手段提供了一种有效且可行的实用新方法, 后继还有刘小慧等人[8]基于武汉疫情新增感染人数数据的分析, 给出了相关潜伏期和疑似期的分布函数的相关估计, 为新冠肺炎疫情的相关研究提供了宝贵的参考价值。此外对于病毒潜伏期的研究还有徐听怡和顾蓓菁[9]使用了两参数 BS 疲劳寿命分布进行了点估计, 得到了刻度参数和形状参数的近似值, 叶莹等人[10]用秩和检验比较了在不同的暴露方式下新型冠状病毒病毒潜伏期的差异性, 得出了家庭暴露的传播性更强的结果, 对于 COVID-19 病毒潜伏期的研究有周瑜和郑庭庭[11], 邱明悦等人[12]; 张美玲等[13]学者对潜伏期的研究给人们认识病毒提供了理论的支持和实践的指导, 还有很多学者从数据本身的角度对其特征进行了刻画, 赵宜宾等人[14]以天津市新型冠状病毒肺炎病例为样本, 用极大似然方法提取了数字特称拟合模型, 得到了潜伏期均值特征。于洋[15]在寿命检验中使用了对数正态, 于晓红和张来斌[16]使用了威布尔分布, 发现这两个分布都能很好地拟合寿命参数, 对数正态分布是指对数的频数分布呈正态, 对数正态分布的一些重要性质在实际中有着重要的应用价值, 尤其在工程, 医学, 生物学等领域对对数正态分布的有着广泛的应用, 它能够很好地刻画数据分布的特征, 并以此来揭示事物数据背后隐藏的信息。为了帮助人们更好的认识疫情状况, 安抚民众恐慌情绪, 给疫情防控人员更多的可能在最大程度上去更好的控制泛滥的病情, 本文着眼于防疫期间卫生健康组织委员会发布的疫情数据对新型冠状病毒的潜伏期的分布进行了相关研究。

2. 数据概述

2.1. 数据来源

本次研究所用到的数据是来自新冠肺炎疫情爆发初期卫生健康委员会在官网发布的新增感染人员的数据公示, 从多个数据中剔除了患病期间数据记录不完整, 溯源不清楚, 发病日期记录模糊, 数据信息不精确的一些有缺陷的数据, 并删去了数据准确度 < 5 的数据之后, 最终得到了 78 条完整数据信息, 其中提取了感染人员的年龄特征, 性别特征, 感染方式, 感染地点等重要信息, 研究了潜伏期的长短是否与这些因素具有潜在的联系。

2.2. 数据的描述统计

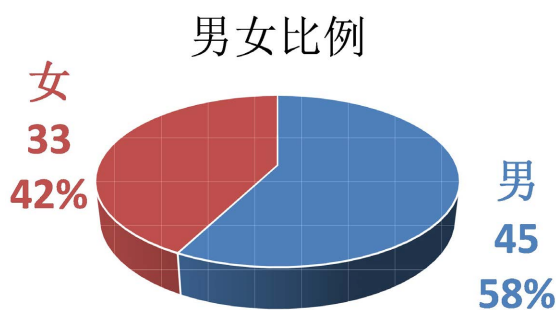


Figure 1. Gender distribution
图 1. 性别分布

表格

本次研究所用到的数据共有 78 条。其中在性别特征上有男性感染者 45 人，女性患者 33 人。男性感染者占比 58%，女性感染者占比 42%，从初步的统计数据来看有男性患者高于女性患者的趋势，见图 1。

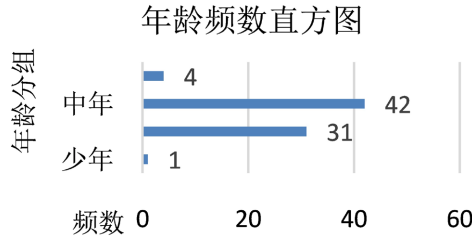


Figure 2. Age distribution
图 2. 年龄分布

在年龄特征上在所有感染者中，年龄最小感染者为 9 岁，年龄最大感染者为 69 岁，感染者的平均感染年龄为 43.2 岁，从初步统计的数据来看感染对象以青壮年群体为主，其次是中年，老年，幼儿居后的特征，见图 2。

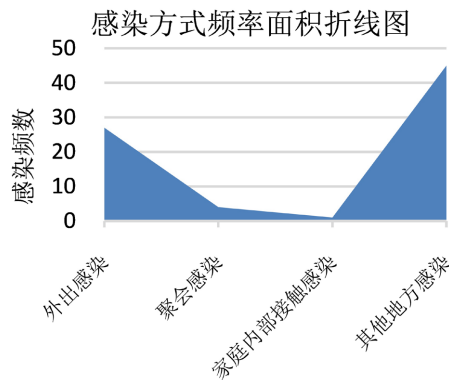


Figure 3. Transmission mode distribution
图 3. 感染方式分布

记录了四种感染方式，即① 外出感染、② 聚会感染、③ 家庭内部接触感染、④ 其他方式感染；感染的地点记录有① 武汉感染、② 湖北非武汉感染、③ 其他地区感染的三种感染地点，根据离武汉的远近程度，可以解释为湖北武汉感染为病毒浓度最高区，湖北非武汉为病毒浓度中高区，其他地区为病毒浓度最低区，见图 3。

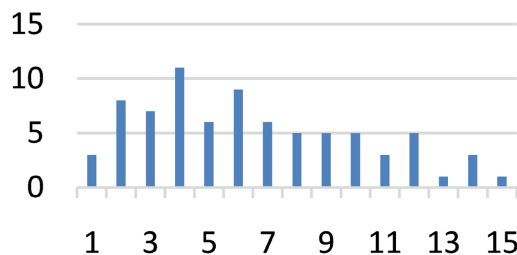


Figure 4. Incubation period distribution
图 4. 潜伏期分布

潜伏期数据初步统计显示，COVID-19 感染者的数据中有最短潜伏期天数为 0 天，也就是说从感染到发病是一个很快的过程，最长潜伏期天数为 16 天，平均潜伏期天数为 5.54 天的特征。也就是说大部分患者潜伏期天数在 5、6 天左右的波动是最小的，见图 4。

对选定样本数据做了初步统计分析，如下表 1 所示，其中中位数，众数，均值这三个指标表示集中趋势的度量，指标值分别为 5，3，5.54，潜伏期的偏度值为 0.59，这些数据显示潜伏期分布形状呈尖峰态左偏，数据中有部分潜伏期数据相较于集中值偏离较大，因此数据呈现出右侧拖尾的现象。因此初步设想新型冠状病毒的潜伏期天数不服从正态分布，而是服从于有偏的正态分布。

Table 1. System resulting data of standard experiment

表 1. 潜伏期特征

病毒潜伏期数字特征统计分析								
样本个数	中位数	众数	均值	方差	最大值	最小值	偏度	峰度
78	5	3	5.54	13.78	16	0	0.59	-0.36

为了检验样本是否服从对数正态分布，对潜伏期样本做了非正态分布的 Shapiro-Wilk 检验，于是得到检验值 w 为 0.94951，得到 p 值为 0.3747，他们均大于置信水平 0.05，也就是说潜伏期的分布函数和对数正态分布的分布函数是没有差异的，因此选用对数正态分布来描述潜伏期天数的分布，见图 5。

3. 论文相关理论概念界定

3.1. 对数正态分布

1. 正态分布：正态分布也称常态分布，又称高斯分布，是连续型随机变量中重要的一种分布，统计学中的很多分布都是正态分布的衍生，并且正态分布也是很多分布的极限分布。正态分布是德国数学家 C.F.Gauss 在误差描述上的重大发现。正态分布的曲线高峰位于中央位置，两侧逐渐下降并且完全对称。正态分布的概率密度函数中记 μ 为总体均数， σ 为总体标准差， μ 和 σ 是正态分布的两个参数，若随机变量 x 服从正态分布，记为 $X \sim N(\mu, \sigma^2)$ 。不同的 μ 值和不同的 σ 值对应不同的正态分布曲线，如下图 5 所示。

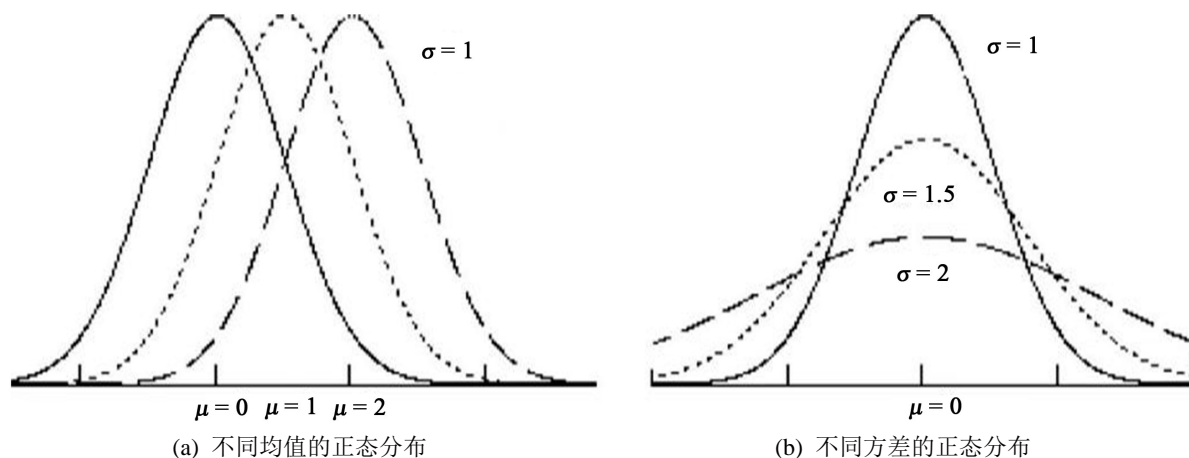


Figure 5. Latent period distribution

图 5. 潜伏期分布

若随机变量 X 服从一个数学期望为 μ ，方差为 σ^2 的正态分布，则数学期望决定了其分布位置，标准差决定了其分布幅度，特别的，当数学期望为 0，方差为 1 时的正态分布叫做标准正态分布。正态分布的概率密度函数定义为：

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

正态分布的性质：① 概率密度曲线在均值处达到最大，并且曲线左右对称。② 标准差决定了概率密度曲线的“陡峭”或“扁平”程度，标准差越大，曲线越扁平；标准差越小，曲线越陡峭。这是因为，标准差越小，意味着大多数变量距离均值的距离越短，也就是说大多数变量都紧密地聚集在均值周围，图形所能覆盖的变量值就少些，图形上呈现瘦高型。相反，标准差越大，数据跨度就越大，分散程度越大，所覆盖的变量就越多，图形呈现“矮胖型”③ 正态随机变量在特定区间上的取值概率由正态曲线下的面积给出，而且其曲线下的总面积等于 1。

2. 对数正态分布：对数正态分布是正态分布的一种衍生，设 X 是取值为正数的连续随机变量，如果 X 取对数后服从正态分布，即 $\ln X \sim N(\mu, \sigma^2)$ ， X 的概率密度函数定义为：

$$f(x, \mu, \sigma) = \begin{cases} \frac{1}{x\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(\ln x - \mu)^2\right], & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (2)$$

则称这个随机变量 X 是服从对数正态分布，记为 $\ln X \sim N(\mu, \sigma^2)$ ，对数正态分布中的期望和方差可以通过极大似然估计来求得。若总体 X 服从对对数正态分布，即：

$$\ln X \sim N(\mu, \sigma^2) (-\infty < \mu < +\infty, \sigma > 0)$$

$x_1, x_2, x_3, \dots, x_n$ 是来自总体 X 的样本，令 $Y = \ln x \sim N(\mu, \sigma^2)$ ，则 $X = e^x$ 的分布函数为：

$$F(x) = P\{X \leq x\} = P\{e^y \leq x\} = \begin{cases} 0, & x \leq 0 \\ P\{Y \leq \ln x\}, & x > 0 \end{cases} = \begin{cases} 0, & x \leq 0 \\ \Phi\left(\frac{\ln x - \mu}{\sigma}\right), & x > 0 \end{cases} \quad (3)$$

X 的概率密度为：

$$f(x) = F'(x) = \begin{cases} 0, & x \leq 0 \\ \Phi'\left(\frac{\ln x - \mu}{\sigma}\right) \frac{1}{\sigma x}, & x > 0 \end{cases} = \begin{cases} 0, & x \leq 0 \\ \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2}, & x > 0 \end{cases} \quad (4)$$

因此样本的似然函数为：

$$\begin{aligned} \ln L(\mu, \sigma^2) &= \prod_{i=1}^n f(x_i) = (2\pi\sigma^2)^{-\frac{n}{2}} \left(\prod_{i=1}^n x_i^{-1}\right) e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (\ln x_i - \mu)^2} \quad (x_i > 0, i = 1, 2, \dots, n), \\ \ln L(\mu, \sigma^2) &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 + \sum_{i=1}^n \ln \frac{1}{x_i} - \frac{1}{2\sigma^2} \sum_{i=1}^n (\ln x_i - \mu)^2 \end{aligned} \quad (5)$$

分别对 $\ln L$ 关于 μ 和 σ^2 分别求偏导，并令其等于 0，于是有：

$$\begin{cases} \frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (\ln x_i - \mu) = 0 \\ \frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^2} \sum_{i=1}^n (\ln x_i - \mu)^2 = 0 \end{cases} \quad (6)$$

$$\text{求解得到: } \hat{\mu} = \frac{1}{n} \sum_{i=1}^n \ln x_i, \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left(\ln x_i - \frac{1}{n} \sum_{i=1}^n \ln x_i \right)^2$$

由于 $X = e^Y, Y \sim N(\mu, \sigma^2)$, 于是有:

$$\begin{aligned} EX &= Ee^Y = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^y e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\frac{1}{2\sigma^2}(y^2 - 2\mu y + \mu^2 - 2\sigma^2 y)} dy \\ &= e^{\frac{\mu + \frac{1}{2}\sigma^2}{\sigma}} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left[\frac{y - (\mu + \sigma^2)}{\sigma}\right]^2} dy \\ &= e^{\frac{\mu + \frac{1}{2}\sigma^2}{\sigma}} \end{aligned} \quad (7)$$

根据最大似然估计的不变性原理, EX 的最大似然估计为 $\widehat{EX} = e^{\frac{\hat{\mu} + \frac{1}{2}\hat{\sigma}^2}{\sigma}}$,

$$\text{其中 } \hat{\mu} = \frac{1}{n} \sum_{i=1}^n \ln X_i, \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left(\ln X_i - \frac{1}{n} \sum_{i=1}^n \ln X_i \right)^2.$$

3.2. 平均超出量函数

为了更直观的描述潜伏期的规律, 本文引入平均超出量函数进行统计描述。平均超出量函数在统计推断中是用最大值的模型来研究的, 在每一组观测数据中用最大值来建立模型, 但是有一个弊端就是在每一组观测数据中只取数据的最大值, 这对于其他数据来说会造成部分数据的浪费, 因此就要用超出某一个大一点的数的所有的数据来进行建模。设 X_1, X_2, \dots, X_n 是独立同分布的, 其分布函数是 F , 对某个固定的值 $u < x^+ = \sup\{x: 0 < F(x) < 1\}$, 称 u 为这组值的阈值, 如果 $X_i > u$, 则称 X_i 为这组数据的超阈值, 称 $X_i - u$ 为超出量, 于是有:

$$F_u(x) = \Pr(X - u \leq x | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)}, x \geq 0 \quad (8)$$

其中称 $F_u(x)$ 为随机变量 X 的超过阈值 u 的超出量分布函数, 简称超出量分布。

称:

$$F_{[u]}(x) = \Pr(X \leq x | X > u) = \frac{F(x) - F(u)}{1 - F(u)}, x \geq 0 \quad (9)$$

称为随机变量 X 的超阈值的分布函数。

$$e(u) = E(X - u | X > u) \quad (10)$$

称为 X 的平均超出量函数。

3.3. 基于对数正态分布的病毒潜伏期特征分析

首先对筛选病患数据进行整理, 进而对潜伏期特征进行了初步特征分析。

从分位数图可以知道数据的分布情况, 从图 6 来看并不服从正态分布, 因此进一步进行 S-W 检验, 得到检验 P 值为 0.0037, 小于 0.05 水平, 因此验证了潜伏期天数不服从正态分布, 考虑到数据总量大于 30 ($78 > 30$), 并且从整体来看, 潜伏期天数的分位数图和 QQ 图存在明显的分布不均匀, 所以在 P 值远

远小于 0.05 水平，此时也可以认为数据不存在严重的偏差，所以不用考虑做数据转换。由对数分布均值公式求得 $E(x) = 5.437$ ，于是认为 COVID-19 的潜伏期天数分布大约在六天左右，因此符合早期 COVID-19 大流行的时候结合对疑似感染病患进行为期一周的隔离观察，认为是合理的，用对数正态分布的公式求得的方差为 23.432，所以不同个体间潜伏期的长短是有很大的差异的。求得潜伏期 0.95 分位数 $x_{0.95} = 16$ ，也就是说 95% 的患者潜伏期长达 16 天之久，因此与早期对病患隔离期限长达十四天之久的相应措施是吻合的，见图 7。

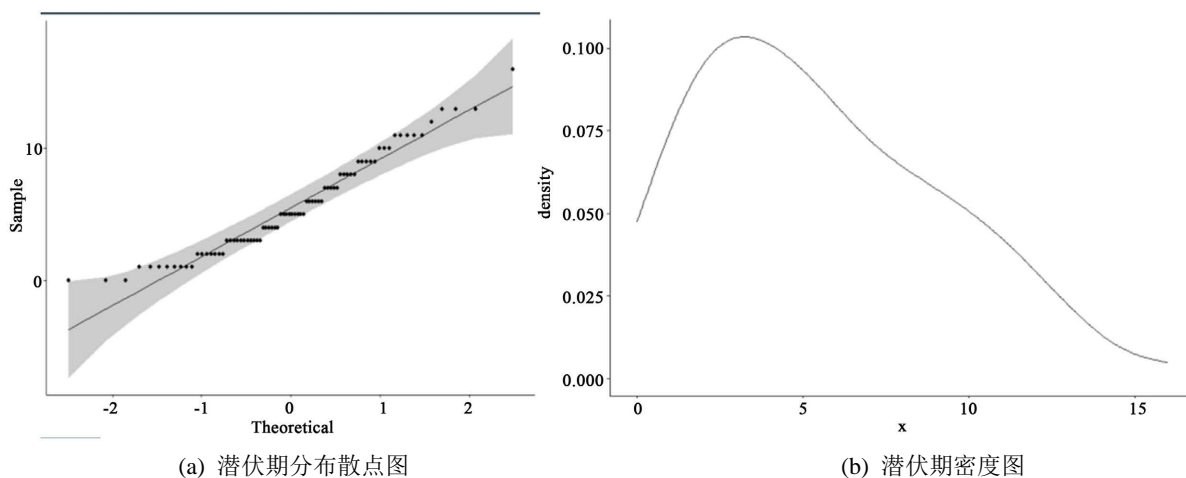


Figure 6. Latent period scatter plot distribution
图 6. 潜伏期散点分布

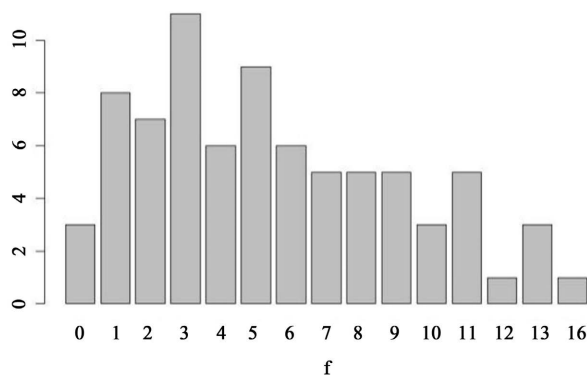


Figure 7. Earliest onset distribution bar chart
图 7. 潜伏期分布条形图

4. 潜伏期影响因素的方差分析

对于新型冠状病毒的潜伏期的影响因素，为了探究本文提取的性别，年龄，感染方式，感染地点等因素是否会对潜伏期长短有显著的影响，因此分别对各类因素进行方差分析。

4.1. 年龄对潜伏期的影响

对所选取的样本数据按照年龄分布将其分成了四个群体，分别为少年群体，青年群体，中年群体，和老年群体这四个群体，其中少年群体年龄分布为七周岁到十五周岁，青年群体为十六周岁到四十周岁，中年群体为四十一周岁到六十五周岁，老年群体为六十六周岁及以上。本文数据来源是最初确诊数据，

可能由于少年和老年群体作为社会重点保护对象的原因，初期感染人数是较少的，因此也就导致了本文筛选的数据中少年群体和老年群体都比较少，不足以体现统计随机性，因此本文着重对青年和中年群体进行关于潜伏期的方差分析。

Table 2. Age analysis table
表 2. 年龄检验表

群体	样本数	最大值	最小值	均值	方差	偏度	峰度	K-S 检验 D 值	K-S 检验 P 值
青年	31	13	0	4.839	14.340	0.696	-0.596	0.77683	0.221
中年	42	16	0	5.571	12.007	0.746	0.463	0.88201	0.226

由上表 2 可知，对于不同年龄群体的数据特征和各群体的 K-S 检验表明不同年龄群体的潜伏期是服从正态分布的，对青年群体和中年群体的数据做了方差的一致性检验，得到的 P 值为 0.118，数据 P 值大于 0.05，因此，认为方差没有太大的差异。在样本数据服从正态分布并且方差相等的前提下，对各组样本数据关于潜伏期天数做均值的单因素方差分析得到了 P 值为 0.4003，大于 0.05，因此认为各个年龄段的潜伏期是没有显著性差异的，也就是说潜伏期的长短和年龄是没有关系的，这与现实生活中的真实情况是吻合的。

4.2. 性别对潜伏期的影响

将本文所选的数据按照男性，女性进行分组，并分别进行数据特征分析和正态 K-S 检验。

Table 3. Gender validation table
表 3. 性别检验表

组别	样本数	最大值	最小值	均值	方差	偏度	峰度	K-S 检验 D 值	K-S 检验 P 值
男性	33	12	0	4.889	10.328	0.432	-0.855	0.94125	0.2624
女性	45	16	0	6.424	17.564	0.449	-0.674	0.61403	0.772

由上表 3 可知，不同性别的群体数据特征和正态性 K-S 检验结果表明各组数据潜伏期是服从正态分布的，对男性和女性两组的潜伏期做方差的齐性检验得到 P 值为 0.0709，大于 0.05 水平所以认为男性和女性的方差是相等的。在方差相等，并且各组数据样本服从正态分布的前提下对男性和女性群体的潜伏期天数做了单因素方差分析得到 P 值为 0.518，大于 0.05，因此认为性别对潜伏期天数的长短是没有显著性影响的，也就是说新型冠状病毒感染后潜伏期的长短跟性别是没有直接的关系的。

4.3. 感染方式对潜伏期的影响

在对感染方式进行相关探究时，本文筛选出来的数据中对于四种感染方式即外出感染，聚会感染，家庭内部接触感染，以及其他地方感染考虑到感染方式上多数为生活接触和普通接触，于是将病患接触方式归为生活接触和普通接触两类进行讨论。

Table 4. Infection pathway validation table
表 4. 感染途径检验表

感染方式	样本数	最大值	最小值	均值	方差	偏度	峰度	K-S 检验 D 值	K-S 检验 P 值
生活接触	33	16	0	6.727	18.955	0.219	-0.978	0.653	0.084
普通接触	45	11	0	4.667	8.500	0.496	-0.507	0.561	0.247

由上表 4 数据的特征因子值以及 K-S 检验值得到数据是来自服从正态总体的，于是做了方差的一致性检验得到了检验 P 值为 $0.0145 > 0.05$ ，因此认为方差是相同的。在方差相等并且样本数据服从正态分布的前提条件下，对病患不同的感染方式，生活接触和普通接触的数据信息进行了单因素方差分析，得到了检验的 P 值为 0.0145，大于 0.05 水平，因此不拒绝原假设，即 COVID-19 感染之后，潜伏期的长短和感染途径之间的影响是显著的，也就是说新冠病毒的感染与人们的生存环境是有关系的，这也是与现实生活人们预防感染所提出来的要求是基本相符的。

4.4. 病毒浓度对潜伏期的影响

根据本文所筛选的病患信息记录的数据，将地点数据分成了三组来进行研究，分组即在湖北武汉感染，在湖北感染了但是并不是在武汉所感染感染，以及感染了 COVID-19 但是并不是在湖北感染，也就是说在除武汉省以外的地方感染了病毒。

Table 5. Regional concentration validation table
表 5. 地域浓度检验表

感染地点	样本数	最大值	最小值	平均值	方差	偏度	峰度	K-S 检验 D 值	K-S 检验 P 值
武汉	58	16	0	4.500	13.784	0.591	-0.358	0.678	0.251
湖北非武汉	12	16	1	8.167	14.973	0.7026	-0.346	0.890	0.331
其他	8	13	4	9.125	12.210	0.1023	-0.863	0.472	0.069

由上表 5 可知，感染 COVID-19 的地点数据特征以及感染地点的正态性 K-S 检验表明各个数据是服从正态分布的，随后对感染地点不同分类数据进行了方差齐性检验，得到检验 P 值为 $0.0145 > 0.05$ ，因此认为数据的方差是相同的。在数据服从正态分布并且方差相等的前提下再对数据进行了均值单因素方差分析，得到检验 P 值为 0.0145，小于水平 0.05，故不拒绝原假设，认为感染地点对潜伏期的长短的影响是显著的，也就是说 COVID-19 的潜伏期长短与感染地点是有关系的，这符合生活常识，在湖北武汉 COVID-19 感染者最多，湖北非武汉次之，因此认为病毒浓度是依次降低的，这可能与当地毒株浓度以及毒株变异种类有着密切的联系，具体真实情况还有待进一步验证。

5. 结语

本文依据武汉地区 COVID-19 感染患者的初步数据信息，经过筛选，聚焦于性别、年龄、感染方式及感染地点等关键变量，旨在深入探究 COVID-19 感染后潜伏期的长短问题。研究结果显示，潜伏期的长短与感染方式及感染地点存在显著关联，而与性别、年龄等因素并无直接联系，这表明该病毒对宿主呈现出无差别感染的特性。本研究所使用的数据源自武汉疫情初期的记录，鉴于当时湖北省武汉市病毒浓度相对较高，患者数量众多，故在分析潜伏期长短影响因素时，感染地点和感染方式被认定为显著因素。然而，需要指出的是，本文关于病毒潜伏期长短的研究仅限于武汉疫情初期的数据。鉴于 COVID-19 病毒的高度变异性及其不同变体所展现的免疫逃逸性，加之人们疫苗接种率的提升，影响因素正持续变化。因此，对于 COVID-19 感染后潜伏期的影响因素，尚需进一步的研究与探索。

参考文献

- [1] 郭德银, 江佳富, 宋宏彬, 秦天, 李振军, 张定梅, 黄森忠, 舒跃龙, 徐建青, 姜世勃, 郑涛, 田怀玉, 郝荣章, 徐建国. 2020-2021 年度新型冠状病毒肺炎疫情发展趋势分析与应对[J]. 疾病监测, 2020, 35(12): 1068-1072.

-
- [2] 赵芳, 赵静, 张海霞, 战寒秋, 王硕, 董伟杰. 传染病定点收治医院静脉药物配置中心新型冠状病毒肺炎疫情综合应急防控措施[J]. 中国消毒学杂志, 2022, 39(12): 958-961.
- [3] 邓江霞, 江博文, 刘传, 杨柳, 古赛. 新型冠状病毒肺炎疫情期间非高风险地区发热门诊患者的临床特征分析[J]. 现代医药卫生, 2020, 36(24): 3949-3951.
- [4] 沈燕, 高晓东, 胡必杰. 新型冠状病毒肺炎疫情防控期间内镜诊疗相关感染防控建议[J]. 微生物与感染, 2020, 15(6): 408-412.
- [5] 赵艳婷. 新冠病毒传播分数阶动力学研究: 建模预测、数值分析及优化管控[D]: [博士学位论文]. 合肥: 中国科学技术大学, 2022.
- [6] 任建强, 崔亚鹏, 倪顺江. 基于机器学习的新冠肺炎疫情趋势预测方法[J]. 清华大学学报(自然科学版), 2023, 63(6): 1003-1011.
- [7] 甘雨, 吴雨, 王建勇. 新冠肺炎疫情趋势预测模型[J]. 智能系统学报, 2021, 16(3): 528-536.
- [8] 刘小惠, 何阳, 麻先思, 罗良清. 有关新冠肺炎潜伏期和疑似期的统计数据分析: 基于湖北省外 2172 条确诊数据[J]. 应用数学学报, 2020, 43(2): 278-294.
- [9] 徐听怡, 顾蓓青. 基于 BS 疲劳寿命分布的新型冠状病毒肺炎潜伏期的实证研究[J]. 数理统计与管理, 2022, 41(6): 982-988.
- [10] 叶莹, 范威, 王文华, 等. 新型冠状病毒肺炎潜伏期分析[J]. 实用预防医学, 2021, 28(2): 129-131.
- [11] 周瑜, 郑庭庭. 具有潜伏期时滞的登革热病毒传播模型研究[J]. 首都师范大学学报(自然科学版), 2020, 41(4): 7-12.
- [12] 邱明悦, 胡涛, 崔恒建. 双区间删失下新冠病毒肺炎潜伏期分布的参数估计[J]. 应用数学学报, 2020, 43(2): 200-210.
- [13] 张美玲. 基于广义差异比混合治愈模型对新型冠状病毒肺炎潜伏期的统计分析[D]: [硕士学位论文]. 大连: 大连理工大学, 2021.
- [14] 赵宜宾, 张艳芳, 任晴晴. 基于对数正态分布的新型冠状病毒肺炎病例统计特征分析[J]. 工程数学学报, 2022, 39(4): 589-598.
- [15] 于洋. 对数正态分布的几个性质及其参数估计[J]. 廊坊师范学院学报(自然科学版), 2011, 11(5): 8-11.
- [16] 于晓红, 张来斌, 王朝晖, 等. 基于新的威布尔分布参数估计法的设备寿命可靠性分析[J]. 机械强度, 2007, 29(6): 932-936.