

数字化转型背景下参数区间估计的教学设计

张爱丽

陆军防化学院, 北京 102205

DOI: 10.61369/ETR.2025360024

摘要 : 文章根据置信区间的定义, 讨论置信区间的含义和分位点的选取。用 Matlab 仿真模拟某传染病潜伏期数据的分布情况, 然后根据区间估计给出传染病隔离时间的合理预测。

关键词 : 数字化转型; 学科交叉性; 问题式; 区间估计; 对数正态分布

Teaching Design of Parameter Interval Estimation in the Context of Digital Transformation

Zhang Aili

Army Chemical Defense College, Beijing 102205

Abstract : The article discusses the meaning of confidence intervals and the selection of percentiles based on the definition of confidence intervals. It simulates the distribution of the incubation period data for a certain infectious disease using Matlab, and then provides a reasonable prediction of the isolation period for the infectious disease based on interval estimation.

Keywords : digital transformation; interdisciplinary nature; problem-based; interval estimation; log-normal distribution

引言

随着信息技术的飞速发展, 全球教育领域正面临着从传统教育模式向数字化教育模式转型。这场转型不仅仅是技术的更新迭代, 更是教育理念、教学模式、管理方式的一场革命。探索适应时代要求的课程教学改革路径, 形成数据驱动、人机结合、跨界开放的新型教育生态, 实现大学数学课程“知识立智、应用为主、思政立德”的融合势在必行。本文以参数的区间估计为例, 探索数字背景下的教学设计模式, 主要分为以下四步。

确定核心知识, 形成本质问题, 转化为驱动性问题。通过实践性任务与学生已具备的知识、技能之间的矛盾激发学生主动探究、深入思考的意愿;^[1-3] 同时增强学生的学习体验, 用问题链将知识由点到线, 线到面连接起来, 从整体和全局把握重难点。

依托案例, 引导学生对实际问题进行合理分析和预测, 提出解决问题的意见, 使得学生有获得感。

多学科融合形成合力。一方面, 跨学科研究能拓展理论视野, 为课堂注入活力; 另一方面对学习提出了更高的要求, 为我们从理论到实践架起桥梁。

关注数学素养的形成。以严密的逻辑推理为基础, 培养学生以理性思维解决问题的素养; 与科技前沿相结合, 培养学生良好的科学态度和创新精神, 合理地提出新思想、新理念、新方法的素养; 以数学模型为依托, 培养学生对现实世界中现象和过程进行合理的简化和量化的素养。以数学文化为引领, 带领学生体验数学的和谐、简洁、对称之美。^[4]

本节课设计: 从点估计出发通过问题式揭示区间估计的合理性; 小样本情况下用分布及区间估计的含义是经过科学家长期对真理追求的结果, 以此激发学生对科学的热情; 从样本数据出发和计算机相结合体现利用学科交叉性解决问题的方法。

一、案例 – 区间估计

(一) 问题引入

例1 用螺旋测微器测量某一铜环的厚度7次, 测量数据如下:

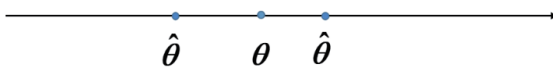
n	1	2	3	4	5	6	7
X/mm	9.515	9.514	9.518	9.516	9.515	9.513	9.517

估计该铜环的厚度。

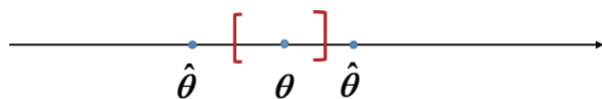
分析: 用参数的点估计, 样本均值代替真实值。铜环的平均厚度 X 的算术平均值估计

$$\bar{X} = \frac{1}{7} \sum_{i=1}^7 H_i = 9.515(mm)$$

由于测量结果具有随机性, 做一组试验得到一个平均值, 如果再做一组试验就会得到另外一个平均值。^[5]



估计值无法反映与真实值的误差范围，使用起来可靠度无法确定。希望能找到一个范围即包含真实值又能反映与真实值的误差。



由测量结果的不确定性，不能保证寻找的区间一定包含真实值，希望寻找的这个区间包含真实值得可靠度比较大，比如95%。一般我们认为多次测量结果 X 满足正态分布。假设 $X \sim N(\mu, \sigma^2)$ 。我们需要估计的就是 μ 。

$$\text{即 } P(\mu < \mu < \bar{\mu}) = 95\%$$

已知 $\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$ ，但是这里 σ 是未知的，而样本标准差 S 是可以计算的，同时在小样本情况下，用 t 分布更准确。

$$\frac{\bar{X} - \mu}{S / \sqrt{n}} \sim t(n-1)$$

$$\text{则 } P\left(-t_{\alpha/2}(n-1) < \frac{\bar{X} - \mu}{S / \sqrt{n}} < t_{\alpha/2}(n-1)\right) = 95\%$$

$$\text{经整得: } P\left(\bar{X} - t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}}\right) = 95\%$$

(二) 置信区间的定义

定义1 设总体 $X \sim F(x, \theta)$ ， θ 是待估计参数，若对给定的 $\alpha(0 < \alpha < 1)$ 存在两个统计量： $\underline{\theta} = \underline{\theta}(X_1, X_2, \dots, X_n)$ ， $\bar{\theta} = \bar{\theta}(X_1, X_2, \dots, X_n)$ ，使得

$$P(\underline{\theta} < \theta < \bar{\theta}) = 1 - \alpha, \quad \theta \in \Theta$$

则称随机区间 $(\underline{\theta}, \bar{\theta})$ 为 θ 的置信度为 $1 - \alpha$ 的置信区间。 $\underline{\theta}, \bar{\theta}$ 分别称为置信下限和置信上限， $1 - \alpha$ 称为置信度或置信水平。

【问题1】估计值与真值的误差？

$$P(\underline{\theta} < \theta < \bar{\theta}) = 1 - \alpha, \quad \theta \in \Theta$$

置信区间 $(\underline{\theta}, \bar{\theta})$ 长度反映误差范围即精度。

【问题2】估计的可靠性？

置信区间包含真值的可能性即置信度反映了参数估计的可靠度。

置信度与精度之间的关系：置信度与精度成反比关系，即精度越高，置信度越小。^[6]

奈曼原则：先保证给定的置信度，再去寻找有优良精度的区间估计。

通常情况下，我们选择 $\alpha = 0.05, 0.1$ 等。

例1 中 μ 的置信度为 $1 - \alpha$ 的置信区间为

$$\left(\bar{X} - t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}}\right)$$

当 $n = 7$ ，取 $\alpha = 0.05$ 查表得 $t_{\alpha/2} = 2.3646$ 。则 μ 的置信度为95%的置信区间为 $(\bar{X} - 0.002, \bar{X} + 0.002)$ 。

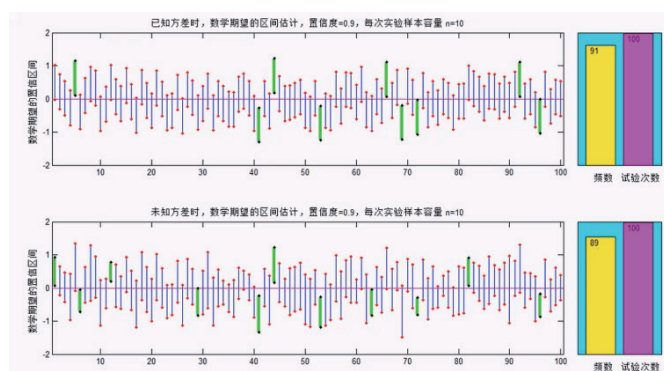
【问题3】下面两种结果的表达，那个是正确的？

1. 区间 $(\bar{X} - 0.002, \bar{X} + 0.002)$ ，包含未知参数 μ 的概率是95%。

2. 未知参数 μ 落在区间 $(\bar{X} - 0.002, \bar{X} + 0.002)$ 的概率是95%。

分析：都不正确，区间 $(\bar{X} - 0.002, \bar{X} + 0.002)$ 和 μ 是确定的值，要么区间一定包含 μ ，要么该区间一定不包含 μ ，不存在概率关系。

置信区间的含义：置信区间是随着样本不同而不同的一个随机变量，而参数的真值是个确定的值^[15]。已知标准正态分布方差已知时，置信度为90%的置信区间，假如做100次实验，约90次包含真值，约10次不包含真值。下图所示是用计算机仿真模拟出方差未知和方差已知时的情况。其中绿色是不包含真值的情况^[7-9]。



(三) 分位点的选取

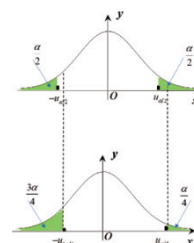
【问题4】上例中分位点为什么选取 $t_{\alpha/2}$ ？

根据奈曼原则：先保证给定的置信度，再去寻找有优良精度的区间估计。那么我们选取的置信区间

$$\left(\bar{X} - t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}}\right)$$

区间长度为 $2 \frac{S}{\sqrt{n}} t_{\alpha/2}$ ，此时长度最短，即精度最高。

以正态分布为例：



取 $\alpha = 0.05$ ， $u_{\alpha/2} - (-u_{\alpha/2}) = 2 \times 1.96 = 3.92$ ，

$$u_{\alpha/4} - (-u_{3\alpha/4}) = 2.24 + 1.78 = 4.02$$

(四) 区间估计的应用

某传染病传播速度快，感染范围广，防控难度大，因此为了能够在最短时间内阻隔传播，斩断病毒传播链条^[11-14]，各地纷纷出台防控通知，在高风险地区旅行后隔离14天，那么为什么我们把隔离的时间定为14天呢？

假如有这样一组100名患者潜伏期数据：

9.2, 3.9, 2, 24, 3.6, 6.2, 11, 3.8, 7.3, 8.8, 6.4, 4.8, 5.9, 9.6, 4.3, 7.1, 5.6, 4.6,
4.4, 7.4, 3.2, 13.4, 3.1, 4.3, 3.6, 2, 3.4, 5.4, 7.6, 2.7, 1.7, 3, 4.9, 6.9, 2.7, 2.7, 2, 5, 7.6, 6.7, 17.1, 8.6, 6.5, 2.7, 2.9, 3.9, 5.8, 7.7, 1.3, 3.2, 3.5, 6.3, 2.9, 17,
4.6, 5.6, 8.8, 5.6, 2.2, 13.7, 5.6, 4.5, 4.3, 9.5, 1.8, 10.7, 4.3, 16.6, 2.1, 5.1, 7.5, 2.6, 6.5, 8.6, 7.7, 4.3, 3.7, 6.4, 6.7, 14.8, 5.6, 4.3, 1.6, 3, 4.8, 11.7, 4.3, 16.5, 13.3, 4, 2.9, 4.9, 10.1, 6.4, 8.1, 7.1, 2.6, 3.3, 4.2, 7.6

1. 用 MATLAB 仿真模拟: 结论是该组数据符合对数正态分布。

2. 对数正态分布潜伏期的区间估计:

由于正态分布 μ 的置信区间为:

$$\left(\bar{X} - t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}} \right)$$

对数正态分布 μ 的置信区间为:

$$\left(\frac{1}{n} \sum_{i=1}^n \ln X_i - t_{\alpha/2}(n-1) \frac{W}{\sqrt{n}}, \frac{1}{n} \sum_{i=1}^n \ln X_i + t_{\alpha/2}(n-1) \frac{W}{\sqrt{n}} \right)$$

这里 $W^2 = \frac{1}{n-1} \sum_{i=1}^n \left(\ln X_i - \frac{1}{n} \sum_{i=1}^n \ln X_i \right)^2$, 当 $\alpha = 0.05$, $n = 100$

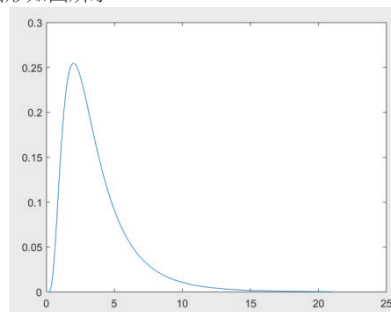
时, $W^2 = 0.35$

$$\frac{1}{n} \sum_{i=1}^n \ln X_i - t_{\alpha/2}(n-1) \frac{W}{\sqrt{n}} = 1.65 - 1.65 \times \frac{0.59}{10} = 1.54$$

$$\frac{1}{n} \sum_{i=1}^n \ln X_i + t_{\alpha/2}(n-1) \frac{W}{\sqrt{n}} = 1.76$$

所以潜伏期 μ 的 95% 的置信区间为: (4.7, 5.8)。

分布图形如图所示



$$P(x \leq \ln 14) = P\left(\frac{x - \mu}{\sigma} \leq \frac{\ln 14 - 0.9}{0.62}\right) = P\left(\frac{x - \mu}{\sigma} \leq 2.80\right) \approx \Phi(2.8) = 0.9974$$

这说明隔离 14 天, 若有 10000 名患者, 会有约 26 人潜伏期超过 14 天, 这是小概率事件, 因此我们认为隔离 14 天是合理的。^[10]

参考文献

- [1] Qun li, M. Med., Xuhua Guan. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. [J]. the new England journal of medicine. 2020.03.26.
- [2] 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. 2011.10.
- [3] 盛骤, 谢式千, 潘承毅. 《概率论与数理统计》[M]. 高等教育出版社, 第 4 版. 2016.6.
- [4] 郭高, 肖燕婷, 陈娟娟. 《概率论与数理统计应用案例分析》[M]. 科学出版社, 第 1 版. 2020.4.
- [5] [美] David Salsburg 著, 刘清山译. 《女士品茶》[M]. 江西人民出版社, 第 1 版. 2016.8.
- [6] 国防科技大学《工程应用数学基础》, 军队职业教育网.
- [7] 北京电子科技大学《概率论与数理统计》, 中国大学生慕课网.
- [8] 王凤肆, 青天福. 《概率论与数理统计课程教学执行计划》[M]. 上海交通大学出版社, 2015 年 9 月.
- [9] 胡健颖, 孙山泽. 抽样调查的理论、方法和应用 [M]. 北京: 北京大学出版社, 2000.
- [10] 陈爱萍, 黄蕾全. 问题式学习的内涵、特征与策略 [J]. 教育科学研究. 2008(1):38-42.
- [11] 刘雄伟, 朱健民, 李建平. 问题式课堂教学设计案例分析——以方向导数与梯度为例 [J]. 2013.10. 高等教育研究学报 (36):67-70.
- [12] [美] 理查德·E 梅耶著, 盛群力, 丁旭, 钟丽佳译. 应用学习科学——心理学大师给教师的建议 [M]. 北京: 中国轻工业出版社, 2020.
- [13] 郑太年. 学习科学与教学变革 [M]. 上海: 上海教育出版社, 2019.
- [14] [美] 康斯坦丝·瑞德著, 姚慕生, 陈克艰译. 奈曼——来自生活的统计学家 [M]. 上海: 上海科学技术出版社, 2001.
- [15] 何晓群. 多元统计分析 [M]. 北京, 中国人民大学出版社, 2017 年 8 月.