

Homework 9

Put your name and student ID here

2021-04-29

Q1: Let X_1, \dots, X_n be an iid sample of Poisson distribution with parameter $\lambda > 0$. Find an approximate $100(1 - \alpha)\%$ confidence interval for λ .

Q2: Suppose that an event A was observed 36 times out of 120 independent experiments. Use CLT to find an approximate 95% confidence interval for $P(A)$.

Q3: Let X_1, \dots, X_n be an iid sample from a distribution with CDF $F(x)$.

- (a) Show that the empirical CDF $\hat{F}_n(x)$ is an unbiased estimate of $F(x)$ for any fixed $x \in \mathbb{R}$.
- (b) Find the variance of $\hat{F}_n(x)$.
- (c) Now suppose that $F(x) = 1 - \exp(-\lambda x)$ for $x > 0$ and 0 otherwise. Inspecting whether the variance of $\hat{F}_n(x)$ attains the lower bound of Cramer-Rao inequality for estimating $F(x)$ with fixed $x > 0$. (In fact, there exists a better unbiased estimator for $F(x)$ than the empirical CDF for this case.)

Q4: 两个样本比较问题（开放性题目）

两个样本比较是统计中很常见的问题。如，比较男生和女生的某些指标（身高、体重、考试成绩等）；在医学上，为了验证某药物的效果，需要比较实验组（服用药物）和对照组（没有服用药物）的差异性。两样本问题数不胜数。为了方便比较，我们经常做出这样的假设：两个样本所属的总体为独立的正态总体。在课上，我们讨论了两个独立正态总体均值差异和方差差异，给出相应的点估计和区间估计。你的任务如下：

1. 根据感兴趣的应用背景，自选两样本数据。这些数据可以通过发放调查问卷得到，也可以使用已有的数据。如果是通过调查问卷获得，需要说明问卷的内容和调查的对象等信息。如果是使用已有的数据，需要引用数据的来源以及说明该数据的相关背景。
2. 描述性数据分析：通过图表等方式比较两样本数据，比如箱线图、密度直方图/核估计等
3. 异常数据处理（如有）
4. 正态性假设的验证：通过观察密度估计图像，判断是否能够用正态假设进行分析
5. 分别求出每个总体的均值和方差的点估计和 95% 置信区间（需要说明你所用到的假设）
6. 分别求出均值差和方差比的点估计和 95% 置信区间（需要说明你所用到的假设）
7. 基于上述分析得出你的结论：回答这两个样本是否存在差异，并由此能得到什么有价值的信息。比如，你正在比较实验组（服用药物）和对照组（没有服用药物）的差异性，如果发现这两个样本没有差异，则或许可以说明该药物对治疗没有效果。

PS: 如果实在找不到合适的的数据，可以使用 R package `ds1labs` 中的身高数据 “heights” (in inches).