STAT 305: Chapter 6

Introduction to formal statistical inference

Amin Shirazi

ashirazist.github.io/stat305.github.io

Chapter 6.1: Large-sample confidence intervals for a mean

Large Sample Confidence Interval

Formal statistical inference uses probability theory to quantify the reliability of data-based conclusions. We want information on a population.e.g

> μ
> - true mean fill weight of food jams
> - true mean strength of metal bars
>
> λ
> - true mean of the number of accidents on a highway in Iowa

We can then use:

to estimate the parameters of population:

1. Point estimates:
   e.g **sample mean** $\overline{X}$ of the strength of metal bars is 4.83.

   > We would then say that $\overline{X}$ is an estimate for true (population) mean μ.

Large Sample Confidence Interval

1. Interval estimates:
   μ is likely to be inside an interval. (e.g $\mu \in (2.84, 5.35)$)

Then we can say **we are confident that the true mean of the strength of metal bars** (μ) **is somewhere in the** $(2.84, 5.35)$

But the question is *how confident*?

Large Sample Confidence Interval

Many important engineering applications of statistics fit the following mold. Values for parameters of a data-generating process are unknown. Based on data, the goal is

we just focus on the population mean (μ)

> 1.identify an inteval of values likely to contain an *unknown parameter*
>
> 2.qualify "how likely" the interval is to cover the correct value of the unknown parameter.

Confidence Interval

Definition and the use

Confidence Interval

Definition: confidence interval for a *parameter* (or function of one or more parameters) is a *data-based interval* of numbers thought likely to contain the parameter (or function of one or more parameters) possessing a stated probability-based confidence or reliability.

> A confidence interval is a realization of a **random interval**, an interval on the real line with a random variable at one or both of the endpoints.

Example:[Instrumental drift]

Let Z be a measure of instrumental drift of a random voltmeter that comes out of a certain factory. Say $Z \sim N(0,1)$. Define a random interval:

$$(Z-2, Z+2)$$

What is the probability that -1 is inside the interval?

$$\begin{aligned} P(-1 \text{ is in } (Z-2, Z+2)) &= P(Z-2 < -1 < Z+2) \\ &= P(Z-1 < -1 < Z+3) \\ &= P(-1 < -Z < 3) \\ &= P(-3 < Z < 1) \\ &= \Phi(1) - \Phi(-3) \\ &= 0.84. \end{aligned}$$

digression :

a standard Normal is

$(-3, 3)$

$P(Z > 4) \approx 0$

$P(Z < 4) \approx 1$