M378K: October 24th, 2025.

The Statistical Set-Up.

Population. e.g., all the people in this class

Sample. e.g., a committee of 4 students

We assume that it's a **REPRESENTATIVE SAMPLE.**

Use the same word for the results of measuring (or polling) from the population w/ an unknown but common dist'n.

Def'n. A random sample of size n from a distribution D is a random vector

$$(Y_1, Y_2, \dots, Y_n)$$

such that:

1. $Y_1, \dots, Y_n$ are independent
2. every Y_i has the distribution D.

Example. Consider 10 measurement $Y_1, Y_2, \dots, Y_{10}$.
Care was taken so that they're independent.

It's a standard model to assume that

Y_i are normally distributed w/ an unknown mean μ.

Scenario #1. We know the standard deviation 0.1.

Then, $Y_i \sim N(\mu, \sigma = 0.1)$, $i = 1..10$

Scenario #2. We don't know the standard deviation σ.

Then, $Y_i \sim N(\mu, \sigma)$, $i = 1..10$

Def'n. Any function of the random sample is called a STATISTIC.

A POINT ESTIMATOR is any function (rule, procedure) of the random sample $(Y_1, \dots, Y_n)$ which included only known constants (w/ the purpose of estimating a model parameter).

An interval estimator is a pair of point estimators.

ESTIMATORS MUSTN'T CONTAIN THE UNKNOWN PARAMETER WE'RE TRYING TO ESTIMATE.

e.g.,

- $\dfrac{\#\text{ of people in the "commitee" who like ice cream}}{4}$,

  i.e., a sample proportion;

- in the normal example, we look @ the sample mean

$$\boxed{\bar{Y} = \frac{Y_1 + Y_2 + \dots + Y_n}{n}}$$

M378K Introduction to Mathematical Statistics

Problem Set #13

Order Statistics.

Problem 13.1. *An insurance company is handling claims from two categories of drivers: the good drivers and the bad drivers. The waiting time for the first claim from a* **good** *driver is modeled by an exponential random variable T_g with mean 6 (in years). The waiting time for the first claim from a* **bad** *driver is modeled by an exponential random variable T_b with mean 3 (in years). We assume that the random variables T_g and T_b are independent.*

What is the distribution of the waiting time T until the first claim occurs (regardless of the type of driver this claim was filed by)?

$\rightarrow$: $T = \min(T_g, T_b)$

$S_T = [0, \infty)$

$t > 0$:
$$\begin{aligned} F_T(t) &= \mathbb{P}[T \leq t] \\ &= \mathbb{P}[\min(T_b, T_g) \leq t] \\ &= 1 - \mathbb{P}[\min(T_b, T_g) > t] \\ &= 1 - \mathbb{P}[T_b > t, T_g > t] \quad \text{(independent)} \\ &= 1 - \mathbb{P}[T_b > t] \cdot \mathbb{P}[T_g > t] \\ &= 1 - e^{-\frac{t}{\tau_b}} \cdot e^{-\frac{t}{\tau_g}} \\ &= 1 - e^{-t\left(\frac{1}{\tau_b} + \frac{1}{\tau_g}\right)} \end{aligned}$$

$$T \sim E(\tau) \quad \text{w/} \quad \tau = \frac{1}{\frac{1}{\tau_g} + \frac{1}{\tau_b}} = \frac{1}{\frac{1}{6} + \frac{1}{3}} \Rightarrow$$

$$\boxed{\tau = 2}$$

$\square$

Definition 13.1. *Let $Y_1, \ldots, Y_n$ be a* **random sample**. *The random sample ordered in an increasing order is called an* order statistic *and denoted by*

$$Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}.$$

Question Write $Y_{(1)}$ as a function of $Y_1, Y_2, \ldots, Y_n$.

$$\boxed{Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)}$$

Question Write $Y_{(n)}$ as a function of $Y_1, Y_2, \ldots, Y_n$.

$$\boxed{Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)}$$

Problem 13.2. *What is the distribution function of the random variable $Y_{(n)}$?*

: for $y \in \mathbb{R}$: $F_{Y_{(n)}}(y) = \mathbb{P}[Y_{(n)} \leq y] = \mathbb{P}[\max(Y_1, \ldots, Y_n) \leq y]$

$= \mathbb{P}[Y_1 \leq y, Y_2 \leq y, \ldots, Y_n \leq y]$ (independence)

$= \mathbb{P}[Y_1 \leq y] \cdot \ldots \mathbb{P}[Y_n \leq y]$ (identically dist'd)

$= (\mathbb{P}[Y_1 \leq y])^n = (F_Y(y))^n$

Problem 13.3. *Assume that the random sample comes from a density f_Y. Is the r.v. $Y_{(n)}$ continuous? If so, what is its density $g_{(n)}$?*