Statistics Fundamentals

Lecture: Population vs. Sample

The Core of Data Collection

In statistics, we aim to understand the world through data. But we can't always study everyone or everything. This is where the crucial concepts of **Population** and **Sample** come into play. Let's dive in!

The Population

The **population** is the ***entire*** group that you want to study and draw conclusions about. It contains every member of the group of interest.

Examples:

- All students enrolled at a specific university.
- All smartphones manufactured by a company in a year.
- Every voter in a country.

The Challenge: Studying an entire population is often impossible or impractical due to time, cost, and logistics.

The Sample

A **sample** is a specific, smaller group that you will collect data from. It is a ***subset*** of the population.

Examples:

- 1,000 students selected from that university.
- 500 smartphones tested from the production line.
- 1,500 voters contacted for a national poll.

The Goal: The sample should be **representative** of the population, so we can use its data to make inferences about the whole group.

A Simple Analogy: Tasting Soup

Imagine you've made a large pot of soup. You want to know if it's seasoned correctly.

The **entire pot of soup** is the **Population**.

You don't need to eat the whole pot! You take **one spoonful** to taste. This is your **Sample**.

If the spoonful tastes good, you infer the whole pot is good. For this to work, you must stir the soup first to make your sample representative!

Describing Our Data

Once we have data, either from a population (a census) or a sample (a survey), we need ways to summarize it. This is where **descriptive statistics** come in. We'll look at two key types: measures of central tendency and measures of variability.

Measures of Central Tendency: Finding the "Center"

These measures describe a typical or central value in a dataset.

Mean (Average)

The sum of all values divided by the count of values. It's the most common measure of center, but it can be sensitive to outliers (very high or low values).

Population Mean (μ)

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

Here, N is the total number of items in the population.

Sample Mean ($\bar{x}$)

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Here, n is the number of items in the sample.

Measures of Variability: Describing the "Spread"

These measures tell us how spread out or dispersed the data points are.

Variance

The average of the squared differences from the Mean. A larger variance means the data is more spread out.

Population Variance (σ^2)

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

The average squared distance from the population mean μ.

Sample Variance (s^2)

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

We divide by $n-1$ (Bessel's correction) to get a better, unbiased estimate of the population variance.

Key Differences at a Glance

Aspect	Population
Definition	The entire group of individuals.
Characteristic	A numerical value describing the population is a **Parame**
Goal	The group we want to make conclusions about.
Data Collection	Called a **census**.

Check Your Understanding

1. A biologist wants to study the average wingspan of all monarch butterflies in North America. They capture and

measure 200 butterflies. What is the population?

- ○ The 200 captured butterflies.
- ○ All monarch butterflies in North America.
- ○ The average wingspan of the 200 butterflies.

2. To gauge student satisfaction, a university emails a survey to 500 randomly selected students. What is the sample?

- ○ All students at the university.
- ○ The students who respond to the survey.
- ○ The 500 students who received the survey.

3. A city's census finds the average household income for all residents is $65,000. This value ($65,000) is a...

- ○ Parameter.
- ○ Statistic.
- ○ Sample.

4. A researcher calculates the average height from a sample of 100 oak trees to be 75 feet. This value (75 feet) is represented by which symbol?

- ○ μ (mu)
- ○ $\bar{x}$ (x-bar)
- ○ σ^2 (sigma-squared)

Check Answers