STA 103 Lecture 1: Statistical Concepts and Notations

Instructor: Wookyeong Song

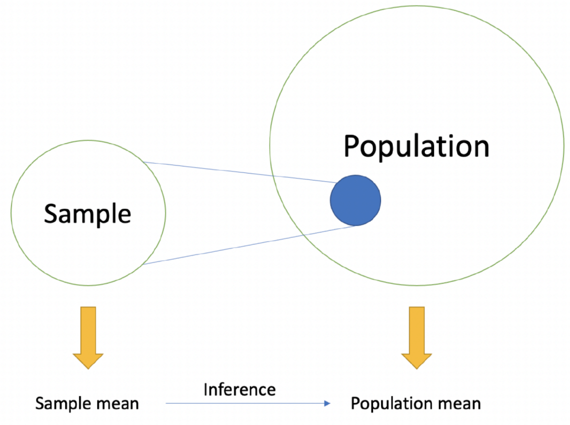
Department of Statistics, University of California, Davis

Aug 4th, 2025

Introduction to Statistics

- **Definition**: Statistics is the science of collecting, organizing, analyzing, and interpreting data to support decision-making.
- **Application**: A business may use customer feedback scores to identify areas for service improvement or optimize pricing strategy using sales data.

Introduction to Statistics



Fundamental Statistical Terminology

- Population: The entire group under study (e.g., all populations of US, all customers of Youtube).
- Sample: A subset selected from the population (e.g., 1000 selected people for presidential election prediction, 5000 selected Youtube customers).
- Parameter: A numeric summary describing a population (e.g., population average daily usage of Youtube). Often denoted as μ (mean), σ^2 (variance).
- Statistic: A numeric summary based on sample data (e.g., average of daily usage of Youtube out of 5000 Youtube customers). Often denoted as $\bar{x}$ (sample mean), s^2 (sample variance).
- (Random) Variable: A numeric characteristic measured on each subject (e.g., daily usage of Youtube, montly credit card spending).

Types of Statistical Analysis

- Descriptive Statistics: Summarize data using graphs, charts, averages, and percentages. (Midterm 1, Midterm 2)
  - Example: A marketing team reports the average spending per customer during a promotional event.

- Inferential Statistics: Make generalizations about a population based on sample data. Estimation, confidence interval, hypothesis testing. (After Midterm 2)
  - A retailer surveys 200 shoppers to estimate the satisfaction level of all store visitors. Find 95% confidence interval of the satisfaction level given 200 samples.
  - For A/B test, is new product B statistically better than old product A? (Youtube 2 minutes 1 ad vs 1 minute 2 ads).

Types of Data

- Qualitative Data: Non-numeric, i.e., labels or name.
  - Example: product category - electronics, groceries.
- Quantitative Data **(Our focus)**: Numeric and measurable.
  - Discrete Variable: Countable (e.g., number of website visits).
  - Continuous Variable: Any value in a range (e.g., sales revenue, heights, weights).

Category	**Symbol**	**Description**
Population	μ, σ, p	Mean, standard deviation, proportion
Sample	$\bar{x}$, s, $\hat{p}$	Mean, standard deviation, proportion

List and Summation

- Data as Lists: In statistics, data is typically represented as a list of values. For example, the daily sales revenue from a store might be written as

$$x_1, x_2, \ldots, x_n, \quad \{x_i\}_{i=1}^{n}$$

  where x_i is the revenue on the i-th day, and n is the total number of days observed.

- If we observe another list, the number of customers visits each day:

$$y_1, y_2, \ldots, y_m, \quad \{y_i\}_{i=1}^{m}.$$

  we label this second list with different notation to avoid confusion.

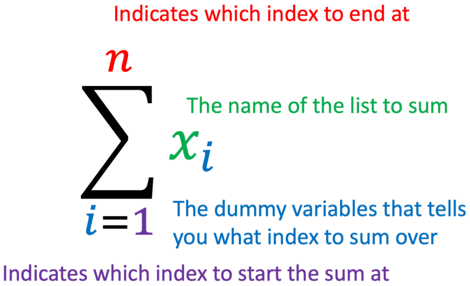
Summation Notation

- **Definition**: The summation operator Σ is shorthand for adding the elements of a list.

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n.$$

- In words, this can be read as "the sum of x_i as i ranged from 1 to n."

- The subscript i is a dummy variable, meaning that you can replace it with whatever letter you want so long as it is not used elsewhere.



Summation Notation

- Note that the following expresses the same thing.

$$\sum_{k=1}^{n} x_k = x_1 + x_2 + \cdots + x_n.$$

- Later in the quarter we will encounter things like

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

- **Example**: Consider the restaurant's daily visit counts over four days: $x_1, x_2, x_3, x_4 = 18, 28, 40, 21$. Find

$$\sum_{i=1}^{2} x_i,\ \sum_{i=1}^{4} x_i,\ \frac{1}{4}\sum_{i=1}^{2}(2x_i + 1),\ \sum_{i=1}^{2} \frac{x_i}{i},\ \sum_{i=1}^{4} x_1,\ \frac{1}{100}\sum_{i=1}^{100} \pi.$$

Properties of Summation

- Let a and b be constants and x_i, y_i be lists of the same length.
  - Constant Factor: $\sum_{i=1}^{n} a x_i = a \sum_{i=1}^{n} x_i$.
  - Additivity: $\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$.
  - Linearity: $\sum_{i=1}^{n} (a x_i + b) = a \sum_{i=1}^{n} x_i + bn$.
- **Warning**: For non-linear functions:

$$\sum_{i=1}^{n} x_i^2 \neq \left(\sum_{i=1}^{n} x_i \right)^2 .$$

Practice Problems

- **Problem 1**: Let daily temperatures in Fahrenheit: $x_1, x_2, \ldots, x_{30}$, and we want to find average in Celsius: $y_i = \frac{5}{9}(x_i - 32)$. Suppose we have $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 72$. Find average of Celsius $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

- Using linear properties:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{5}{9}x_i - \frac{160}{9}\right) = \frac{5}{9}\bar{x} - \frac{160}{9} = \frac{200}{9}.$$

- **Warning**: Avoid recalculating each y_i: you only need the mean of the original data $\bar{x}$.

Practice Problems

- **Problem 2**: Let r_i be the revenue on day i for 10 days, $i = 1, 2, \ldots, 10$, and assume $\sum_{i=1}^{n} r_i = 12,000$ dollars. Now the store offers a 10% discount and expects revenue to drop accordingly, assume no increase in sales. Estimate the new average revenue.

- **Answer**:

$$\bar{r}_{\text{new}} = 0.9 \times \frac{12,000}{10} = 1,080.$$