Statistics

A Note By
Shah Md. Arshad Rahman Ziban

Descriptive Statistics

01 Mean

The average of all data points. It is calculated by dividing the sum of all values by the number of values.

$$\text{Mean} = \frac{\Sigma x_i}{n}$$

Suppose we have the dataset: [10,20,20,30,40]

$$\text{Mean} = \frac{10 + 20 + 20 + 30 + 40}{5} = \frac{120}{5} = 24$$

02 Median

The middle value in a sorted dataset. If there's an even number of values, it's the average of the two middle values.

Suppose dataset: [10,20,20,30,40]

The middle value is 20 (the third value).

Suppose dataset: [10,20,20,30,40,50]

Since the dataset has 6 numbers (even), the middle two values are the 3rd and 4th values:20 and 30.

The average of these two middle values:

$$\text{Median} = \frac{20 + 30}{2} = 25$$

03 Mode

The most frequently occurring value(s) in the dataset.

Suppose we have the dataset: [10,20,20,30,40]

The most frequent value is 20.

04 Variance

A measure of how spread out the data points are from the mean. It is calculated as the average of the squared differences from the mean.

Suppose we have the dataset: [2,4,6] and Mean = 4.

$$\text{Variance}(\sigma) = \frac{\Sigma(x_i - \mu)^2}{n}$$

$$\text{Variance}(\sigma) = \frac{(2 - 4)^2 + (4-4)^2 + (6 - 4)^2}{3} = \frac{4 + 0 + 4}{3} = 2.67$$

(05) Standard Deviation

The standard deviation is the square root of variance, showing how much data deviates from the mean.

$$\text{Standard Deviation} = \sqrt{2.67} \approx 1.63$$

Descriptive Statistics

(01) Hypothesis Testing

Hypothesis testing is like checking if a guess about a group is correct.

Example: A company says their product helps students score 20% better on tests.

Null Hypothesis (Ho): The product doesn't help (no effect).
Alternative Hypothesis (Ha): The product helps improve test scores by 20%.

Use a method (like a t-test) to see if there's enough proof to believe the company's claim or not.

In simple terms: You're testing if the company's claim is true by looking at the test results.

(02) Confidence Intervals

A confidence interval gives a range of values where the true number is likely to be.

Example: A survey says the average height of a population is 5.5 feet, with a confidence interval of 5.4 to 5.6 feet at 95% confidence.

What it means: We're 95% sure the true average height is between 5.4 and 5.6 feet.

(03) T-tests

A t-test compares the averages of two groups to see if they are different.

Example: You want to see if students using a new study method score better than those who don't.

1. Compare the average test scores of both groups using a t-test.
2. The result will tell you if the new method works.

Probability

01 Probability Distributions

A probability distribution shows how likely different outcomes are.

Example: In a bag, there are 4 red balls and 6 blue balls. If you randomly pick one ball, the probability of:

$$\text{Picking a red ball is } \frac{4}{10} = 0.4 \text{ or } 40\%$$

$$\text{Picking a red ball is } \frac{6}{10} = 0.6 \text{ or } 60\%$$

This is a simple discrete probability distribution, as each outcome (red or blue) has a specific probability.

02 Bayes' Theorem

Helps update probabilities when you get new information.

Example:

1% of people have a disease: This means (Disease) = 0.01.

The test is 99% accurate:

- If you have the disease, the test correctly says "Positive" 99% of the time: (Positive Test | Disease) = 0.99 P (Positive Test | Disease) = 0.99.

- If you don't have the disease, the test falsely says "Positive" 1% of the time: (Positive Test | No Disease) = 0.01 P (Positive Test | No Disease) = 0.01.