Statistics Notes

Complete Guide with Examples & Visualizations

Notes by lord_ritik07

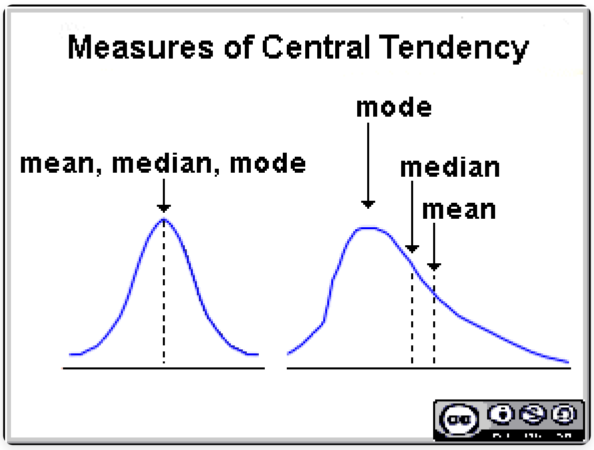
Table of Contents

1. Measures of Central Tendency

> **Central Tendency** is a statistical measure that identifies a single value as representative of an entire dataset. It indicates where the center of a distribution lies.

Types of Central Tendency Measures

- **Arithmetic Mean (AM)** - The average of all values
- **Median** - The middle value when data is ordered
- **Mode** - The most frequently occurring value
- **Geometric Mean (GM)** - The nth root of the product of values
- **Harmonic Mean (HM)** - The reciprocal of the arithmetic mean of reciprocals



Visual comparison of mean, median, and mode in different distribution shapes

Arithmetic Mean ($\bar{X}$)

Ungrouped Data: $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$
Grouped Data: $\bar{X} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{j=1}^{k} f_j}$
Weighted Mean: $\bar{X}_w = \frac{\sum w_i x_i}{\sum w_i}$

Examples

Easy Example Easy

Problem: Find the mean of the following test scores: 85, 90, 78, 92, 88

1. Sum all values: 85 + 90 + 78 + 92 + 88 = 433
2. Divide by number of values: 433 ÷ 5 = 86.6
3. **Answer:** Mean = 86.6

Medium Example Medium

Problem: Calculate the weighted mean for a student with the following grades:

Subject	Grade	Credit Hours

Math	85	4
Science	92	3
History	78	2

1. Multiply each grade by its weight: (85×4) + (92×3) + (78×2) = 340 + 276 + 156 = 772
2. Sum the weights: 4 + 3 + 2 = 9
3. Divide: 772 ÷ 9 = 85.78
4. **Answer:** Weighted mean = 85.78

Hard Example Hard

Problem: Calculate the mean for the following grouped data:

Class Interval	Frequency
0-10	5
10-20	8
20-30	12
30-40	7
40-50	3

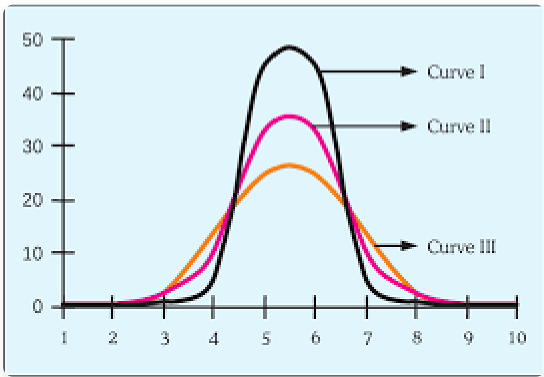
1. Find midpoints of each class: 5, 15, 25, 35, 45
2. Multiply midpoints by frequencies: (5×5) + (15×8) + (25×12) + (35×7) + (45×3) = 25 + 120 + 300 + 245 + 135 = 825
3. Sum frequencies: 5 + 8 + 12 + 7 + 3 = 35
4. Divide: 825 ÷ 35 = 23.57
5. **Answer:** Mean = 23.57

Note: Median is often a better indicator than mean when:

- The dataset contains significant outliers
- The distribution is highly skewed
- You need a measure that's less affected by extreme values

2. Measures of Dispersion

Dispersion measures how spread out the values in a dataset are. While central tendency tells us about the center, dispersion tells us about the spread.



Different distributions with the same mean but different dispersion

Range

Range = Maximum value - Minimum value

Variance

Population Variance: $\sigma^2 = \frac{\sum(x_i - \mu)^2}{N}$

Sample Variance: $s^2 = \frac{\sum(x_i-\bar{x})^2}{n-1}$

Standard Deviation

Population SD: $\sigma = \sqrt{\frac{\sum(x_i-\mu)^2}{N}}$

Sample SD: $s = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}$

Examples

Easy Example Easy

Problem: Find the range of: 12, 15, 18, 22, 25, 30

1. Identify maximum value: 30
2. Identify minimum value: 12
3. Calculate range: 30 - 12 = 18
4. **Answer:** Range = 18

Medium Example Medium

Problem: Calculate variance and standard deviation for: 5, 7, 9, 11, 13

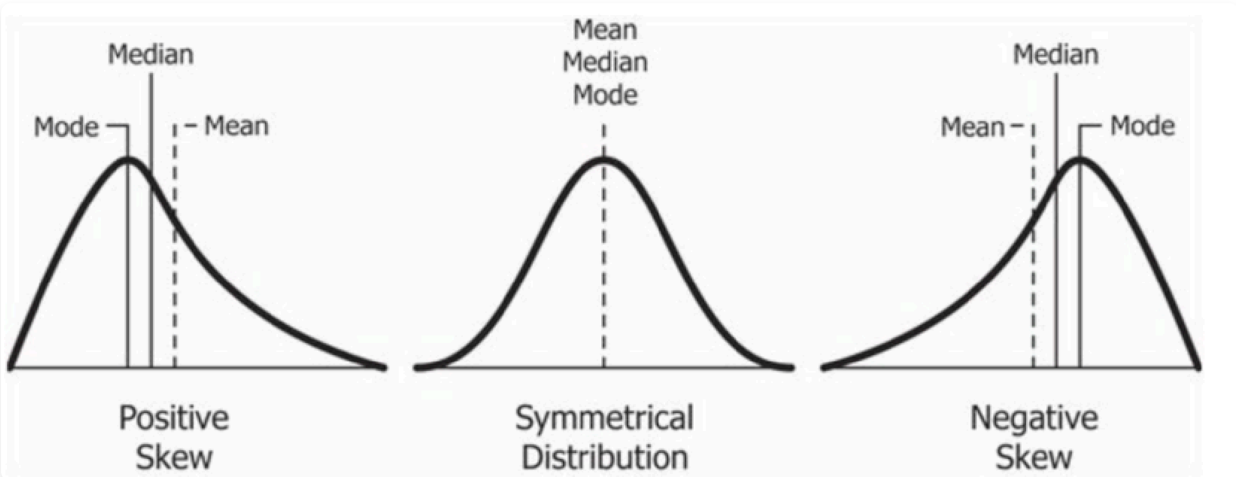
1. Calculate mean: (5+7+9+11+13)/5 = 45/5 = 9
2. Find squared differences: $(5-9)^2=16$, $(7-9)^2=4$, $(9-9)^2=0$, $(11-9)^2=4$, $(13-9)^2=16$
3. Sum squared differences: 16+4+0+4+16 = 40
4. Variance (population): 40/5 = 8
5. Standard deviation: √8 ≈ 2.83
6. **Answer:** Variance = 8, SD ≈ 2.83

Key Points:

- Range is simple but sensitive to outliers
- Variance measures average squared deviation from mean
- Standard deviation is in the same units as the original data
- For samples, we use n-1 in the denominator (Bessel's correction)

3. Measures of Skewness

Skewness measures the asymmetry of a probability distribution. A symmetric distribution has skewness of zero.



Visual representation of positive skew (right-tailed), negative skew (left-tailed), and symmetric distributions

Pearson's Coefficient of Skewness

$$Sk = \frac{3(\text{Mean} - \text{Median})}{\text{Standard Deviation}}$$

Examples

Easy Example Easy

Problem: For a dataset with mean=50, median=45, and SD=10, calculate skewness

1. Apply formula: Sk = 3(50-45)/10
2. Calculate: 3×5/10 = 15/10 = 1.5
3. **Answer:** Skewness = 1.5 (positively skewed)

Medium Example Medium

Problem: Interpret the following skewness values: -0.8, 0, 2.1

1. Skewness = -0.8: Moderately negatively skewed (left-tailed)
2. Skewness = 0: Symmetrical distribution
3. Skewness = 2.1: Highly positively skewed (right-tailed)

Interpretation Guide:

- Skewness between -0.5 and 0.5: Approximately symmetric
- Skewness between -1 and -0.5 or 0.5 and 1: Moderately skewed
- Skewness less than -1 or greater than 1: Highly skewed

4. Probability Theory

Probability quantifies the likelihood of events occurring, ranging from 0 (impossible) to 1 (certain).

Basic Probability Rules

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$P(A^c) = 1 - P(A)$$
If A and B are mutually exclusive: $P(A \cup B) = P(A) + P(B)$

Examples

Easy Example Easy

Problem: What's the probability of rolling an even number on a fair die?

1. Even numbers on a die: 2, 4, 6 (3 outcomes)
2. Total possible outcomes: 6
3. Probability = 3/6 = 1/2
4. **Answer:** P(even) = 0.5

Medium Example Medium

Problem: In a class, 30% of students take math, 25% take science, and 10% take both. What's the probability a randomly selected student takes math or science?

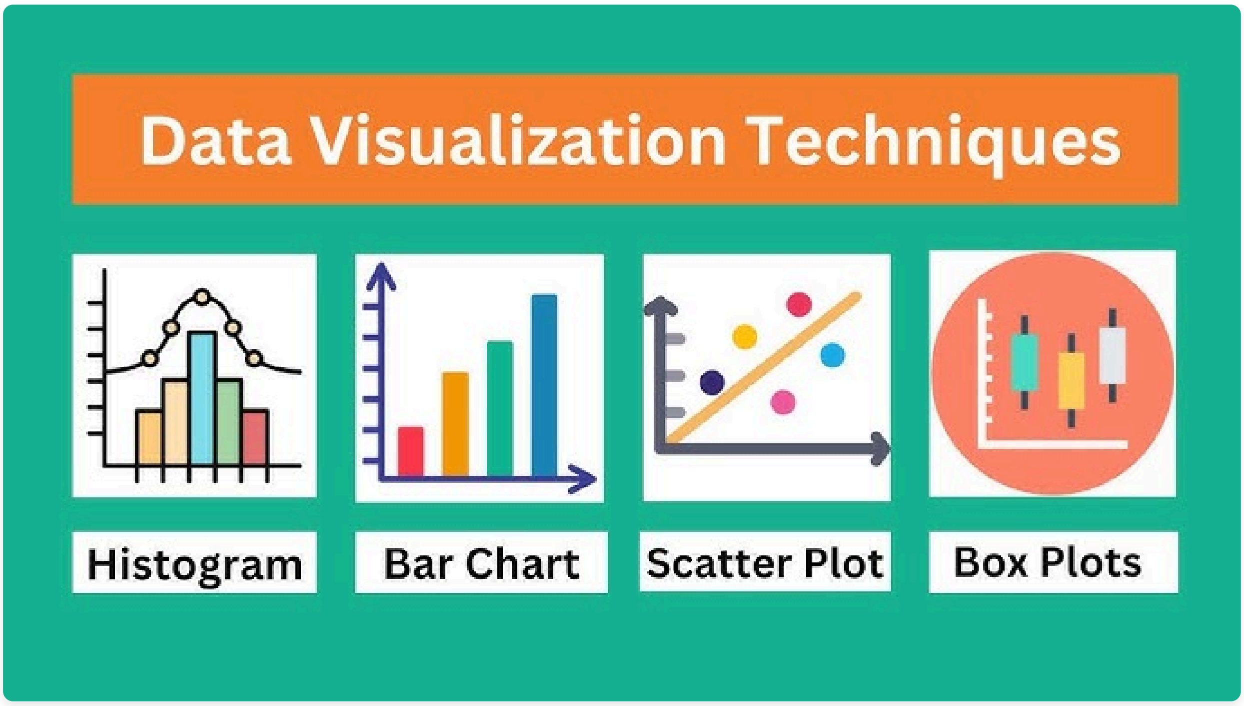
1. P(Math) = 0.30, P(Science) = 0.25, P(Both) = 0.10
2. Apply addition rule: P(Math ∪ Science) = 0.30 + 0.25 - 0.10
3. Calculate: 0.45
4. **Answer:** Probability = 0.45 or 45%

5. Data Representation

> **Data representation** involves visualizing data through charts, graphs, and plots to reveal patterns, trends, and relationships.

Common Data Visualization Methods

- **Histograms** - For frequency distributions
- **Box plots** - For five-number summaries
- **Scatter plots** - For relationships between variables
- **Bar charts** - For categorical data
- **Pie charts** - For proportional data



Examples of histogram, box plot, and scatter plot visualizations

Examples

Easy Example Easy

Problem: Create a simple bar chart for favorite fruits among 20 people: Apples (8), Bananas (5), Oranges (7)

1. Draw horizontal and vertical axes
2. Label categories on horizontal axis: Apples, Bananas, Oranges
3. Scale vertical axis from 0 to 8
4. Draw bars with heights: Apples=8, Bananas=5, Oranges=7