Statistics – Theory with Real-Life Examples

1. Mean (Arithmetic Average)
Definition:
The mean of a data set is the sum of the observations divided by the number of observations. And, mean is the most commonly used measure of central tendency.
• The mean is usually referred to as average.
• In arithmetic average, we have to divide the sum of the values by the number of values which is another typical value.

Formula:
Population Mean: $\boldsymbol{\mu = (\Sigma x) / N}$
Sample Mean: $\boldsymbol{\bar{x} = (\Sigma x) / n}$
Where:
μ (mu) = population mean
$\bar{x}$ (x-bar) = sample mean
Σx = sum of all values
N = total population size
n = sample size

Real-Life Examples:
Example 1: Basketball Player Performance
Scenario: LeBron James' points per game over 5 games
Game 1: 28 points
Game 2: 35 points
Game 3: 24 points
Game 4: 31 points
Game 5: 42 points
Calculation:
Mean = (28 + 35 + 24 + 31 + 42) / 5 = 160 / 5 = 32 points per game

Example 2: Patient Temperature Monitoring
Scenario: Hospital monitors patient temperature every 4 hours
6 AM: 98.2°F
10 AM: 98.6°F
2 PM: 99.1°F
6 PM: 98.8°F
10 PM: 98.5°F
Calculation:
Mean = 493.2 / 5 = 98.64°F

2. Median

Definition:
The median is the middle value when a dataset is arranged in ascending or descending order. It divides the dataset into two equal halves.
Formula:
For odd number of values (n is odd):
Median = **Value at position (n + 1) / 2**
For even number of values (n is even):
Median = **(Value at position n/2 + Value at position (n/2 + 1)) / 2**

Real-life Examples:
Example 1: Employee Salaries
Scenario: Tech company salaries (including CEO)
5 Junior Developers: 60,000 each
3 Senior Developers: 90,000 each
2 Managers: 120,000 each
1 CEO: 500,000
Sorted: 60k, 60k, 60k, 60k, 60k, 90k, 90k, 90k, 120k, 120k, 500k
Median: 6^{th} value = 90,000

Example 2: Marathon Finish Times
Scenario: Times for 9 runners (in minutes)
185, 192, 198, 205, 210, 215, 225, 240, 310
Median: 5^{th} value = 210 minutes

3. Mode:

Definition:
The mode is the value that appears most frequently in a dataset. A dataset can have no mode, one mode (unimodal), two modes (bimodal), or multiple modes (multimodal).
Formula:
The mode is found by counting frequencies:
Mode = Value with highest frequency

Real-life Examples:
Example 1: Shoe Store Inventory
Scenario: Shoe sizes sold in one day
Size 7: 3 pairs
Size 8: 7 pairs
Size 9: 12 pairs
Size 10: 8 pairs
Size 11: 4 pairs
Size 12: 2 pairs
Mode: Size 9 (12 pairs sold)

Example 2: Customer Service Ratings
Scenario: Restaurant feedback scores (1-5 scale)
Rating 1: 2 customers
Rating 2: 5 customers
Rating 3: 8 customers
Rating 4: 15 customers
Rating 5: 10 customers
Mode: Rating 4

4. Standard Deviation:

Definition:
Standard deviation measures the amount of variation or dispersion of values from the mean. It quantifies how spread out the data values are from the mean.

Formula:
Population Standard Deviation: $\boldsymbol{\sigma = \sqrt{[(\Sigma(x - \mu)^2) / N]}}$
Sample Standard Deviation: $\mathbf{S = \sqrt{[(\Sigma(x - \bar{x})^2) / (n - 1)]}}$
Where:
σ (sigma) = population standard deviation
s = sample standard deviation
x = each value
μ = population mean
$\bar{x}$ = sample mean
N = population size
n = sample size

Note: Standard deviation is expressed in the same units as the original data, making it interpretable. A small standard deviation indicates data points are close to the mean, while a large standard deviation indicates data points are spread out.

Real-Life Examples
Example 1: Investment Risk Assessment
Scenario: Annual returns for two mutual funds over 10 years
Fund A: 8%, 12%, 10%, 9%, 11%, 7%, 13%, 8%, 11%, 9%
Mean return: 9.8%
Standard Deviation: 1.9%

Fund B: -5%, 25%, 3%, 18%, -2%, 22%, 5%, 15%, -1%, 20%
Mean return: 10%
Standard Deviation: 10.8%

Analysis:
Fund B has slightly higher average return
Fund B has much higher SD (10.8% vs 1.9%)
Fund A is more stable/less risky

Example 2: Manufacturing Quality Control
Scenario: Smartphone battery life (hours) from production line
Target: 24 hours
Sample measurements: 23.8, 24.1, 23.9, 24.2, 24.0, 23.7, 24.3, 23.9, 24.1, 24.0
Calculation:
Mean = 24.0 hours
Standard Deviation = 0.18 hours

5. Variance:

Definition:
Variance is the average of the squared differences from the mean. It measures how far a dataset is spread out. Variance is the square of the standard deviation.

Formula:
Population Variance: $\boldsymbol{\sigma^2 = \Sigma(x - \mu)^2 / N}$
Sample Variance: $\boldsymbol{s^2 = \Sigma(x - \bar{x})^2 / (n - 1)}$

Real-Life Examples
Example 1: Weather Temperature Variability
Scenario: Daily temperatures (°C) in two cities over a week

City A (Coastal):
Temperatures: 20, 21, 19, 22, 20, 21, 20
Mean: 20.4°C
Variance: 0.86°C^2

City B (Desert):
Temperatures: 15, 25, 30, 18, 35, 20, 28
Mean: 24.4°C
Variance: 49.1°C^2

Example 2: Sales Performance Consistency
Scenario: Monthly sales (units) for two salespersons
Salesperson A: 100, 95, 105, 98, 102, 96, 104, 99, 101, 100
Mean: 100 units
Variance: 12 units^2
Salesperson B: 80, 120, 70, 130, 90, 110, 75, 125, 85, 115
Mean: 100 units
Variance: 525 units^2

Business Impact:
Same average sales
A is consistent (easier inventory planning)
B is erratic (harder to predict)
A gets "consistency bonus"

6. Quartiles:

Definition:
Quartiles are values that divide a ranked dataset into four equal parts. Each quartile contains 25% of the data points.

Types and Formulas:
Q1 (First Quartile): 25th percentile
Q2 (Second Quartile): 50th percentile (median)
Q3 (Third Quartile): 75th percentile
IQR (Interquartile Range): Q3 – Q1

Position Formula:
Position of Qk = **k(n + 1) / 4**
Where k = 1, 2, or 3 for Q1, Q2, Q3 respectively

Note: Quartiles help understand data distribution and identify outliers. The IQR contains the middle 50% of data and is resistant to outliers.

Example 1: Student Performance Assessment
Scenario: Test scores for 100 students

Q1 = 65 (bottom 25%)
Q2 = 78 (median)
Q3 = 88 (top 25%)
IQR = 23

Educational Decisions:
Students below Q1: Need remedial support
Q1-Q2: Additional practice recommended
Q2-Q3: On track
Above Q3: Advanced placement candidates
IQR = 23: Reasonable spread in abilities

Example 2 : Healthcare Wait Times
Scenario: Emergency room wait times (minutes)
Data from 200 patients
Q1 = 25 minutes
Q2 = 45 minutes
Q3 = 75 minutes
IQR = 50 minutes

Hospital Management Use:
Staffing: Add staff when median wait >45 min
Performance metric: Keep 75% of waits <75 min
Outlier detection: Investigate waits >150 min (Q3 + 1.5×IQR)
Patient communication: “Most patients seen within 45-75 minutes”