# Q&A - Statistics

## What is Statistics? Explain its main types.

Statistics is the science of collecting, analyzing, and interpreting data.  
  
Types:  
- Descriptive: Summarizes data (mean, graphs).  
- Inferential: Makes predictions or conclusions about a population from a sample.

## Define population and sample with examples.

Population: The entire group of interest. Example: All students in a university.  
Sample: A subset of the population. Example: 100 students selected from the university.

## What is the difference between descriptive and inferential statistics?

Descriptive: Summarizes data (tables, charts, averages).  
- Inferential: Uses data from samples to make conclusions or predictions about populations.

## Explain data types (qualitative vs quantitative, discrete vs continuous).

Qualitative: Categorical (e.g., gender, colors).  
- Quantitative: Numerical (e.g., height, marks).  
 • Discrete: Countable (e.g., number of cars).  
 • Continuous: Measurable (e.g., weight).

## What is a variable in statistics? Give examples.

A variable is any characteristic that can take different values.  
Examples: Age, income, test scores.

## Define mean, median, and mode. How are they different?

Mean: Average of values.  
- Median: Middle value in ordered data.  
- Mode: Most frequent value.  
They differ in how they represent central tendency.

## How do you calculate the range of a dataset?

Range = Maximum value – Minimum value.  
Example: For {2, 5, 9}, Range = 9 – 2 = 7.

## What is the standard deviation, and why is it important?

Standard deviation measures the spread of data from the mean.  
It shows how consistent or varied the data is.

## Explain variance and how it relates to standard deviation.

Variance is the average of squared deviations from the mean.  
Standard deviation is the square root of variance.

## What is a frequency distribution? Give an example.

A table showing how often values occur.  
Example: Test scores grouped into ranges with counts.

## Explain the concept of normal distribution and its characteristics.

Normal distribution is a symmetric, bell-shaped curve.  
Characteristics: Mean=Median=Mode, 68-95-99.7% rule applies.

## What is skewness, and how does it affect data interpretation?

Skewness measures asymmetry of data.  
- Positive skew: Tail on right.  
- Negative skew: Tail on left.

## What is kurtosis, and what does it tell us about a dataset?

Kurtosis measures the 'tailedness' of data distribution.  
High kurtosis = more extreme values; low kurtosis = fewer extremes.

## Differentiate between probability and statistics.

Probability: Predicts likelihood of events.  
- Statistics: Analyzes collected data to make conclusions.

## What is a z-score, and how is it calculated?

Z-score = (Value – Mean) / Standard Deviation.  
It measures how many standard deviations a value is from the mean.

## Explain the difference between population standard deviation and sample standard deviation.

Population SD: Uses entire population data.  
- Sample SD: Uses sample data with (n-1) denominator for unbiased estimate.

## What is the Central Limit Theorem, and why is it important?

It states that the sampling distribution of the sample mean approaches a normal distribution as sample size increases, regardless of population shape.  
It allows valid inference using normal probability.

## 18.What is correlation? Differentiate between positive and negative correlation.

Correlation measures strength of relationship between variables.  
- Positive: Variables move in same direction.  
- Negative: Variables move in opposite directions.

## Explain the difference between correlation and causation.

Correlation shows association between variables.  
Causation means one variable directly affects the other.

## What is regression analysis, and when is it used?

Regression analysis models relationships between variables to predict outcomes.  
Used in forecasting (e.g., predicting sales from advertising).

## Explain hypothesis testing and its steps.

Hypothesis testing evaluates claims using data.  
Steps: State hypotheses → Choose test → Set significance level → Analyze data → Accept/Reject null.

## What is a null hypothesis and an alternative hypothesis?

Null (H0): Assumes no effect or difference.  
- Alternative (H1): Assumes there is an effect or difference.

## Explain p-value in hypothesis testing.

P-value is the probability of observing the data if H0 is true.  
Low p-value (<0.05) → reject H0.

## What is the difference between Type I and Type II errors?

Type I: Rejecting true H0 (false positive).  
- Type II: Failing to reject false H0 (false negative).

## What is a confidence interval, and how is it interpreted?

A confidence interval gives a range of values that likely contain the population parameter.  
Example: 95% CI means we are 95% confident the true value lies within the range.

## Explain t-test and when to use it.

T-test compares means between two groups.  
Used when population SD is unknown and sample size is small.

## Explain chi-square test and its applications.

Chi-square test checks relationship between categorical variables.  
Applications: Goodness of fit, independence tests.

## What is ANOVA, and when is it used?

ANOVA (Analysis of Variance) compares means of 3 or more groups.  
Used to test if group differences are statistically significant.

## How do you handle missing data in statistics?

Methods: Remove missing values, replace with mean/median, or use advanced imputation techniques.

## What is sampling bias, and how can it be reduced?

Sampling bias occurs when a sample is not representative of the population.  
Reduce by using random sampling and ensuring diverse selection.