**1. What is Statistics? Explain its main types.**

Statistics is the science of collecting, analyzing, interpreting, and presenting data.  
**Types:**

* **Descriptive Statistics** → Summarizes data (mean, charts, tables).
* **Inferential Statistics** → Makes predictions or generalizations about a population using sample data.

**2. Define population and sample with examples.**

* **Population**: The entire group you want to study.  
   Example: All students in a university.
* **Sample**: A smaller group taken from the population.  
   Example: 200 students selected for a survey.

**3. What is the difference between descriptive and inferential statistics?**

* **Descriptive**: Describes and summarizes data (e.g., average test score).
* **Inferential**: Makes conclusions or predictions about a population (e.g., predicting election results from a survey).

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

* **Qualitative (categorical)**: Non-numerical, e.g., gender, color.
* **Quantitative (numerical)**: Numbers, e.g., age, height.
  + **Discrete**: Countable values (e.g., number of children).
  + **Continuous**: Any value in a range (e.g., weight, temperature).

**5. What is a variable in statistics? Give examples.**

A variable is a characteristic that can change or vary.  
 Examples: Age, income, blood pressure.

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

* **Mean**: Average value.
* **Median**: Middle value when data is ordered.
* **Mode**: Most frequent value.  
   Example: Data = [2, 3, 3, 4, 5] → Mean = 3.4, Median = 3, Mode = 3.

**7. How do you calculate the range of a dataset?**

**Range = Maximum value – Minimum value**  
 Example: [10, 15, 20] → Range = 20 – 10 = 10.

**8. What is the standard deviation, and why is it important?**

It measures how spread out the data is from the mean.  
 Importance: Tells whether data points are close to average or widely spread.

**9. Explain variance and how it relates to standard deviation.**

Variance measures the average squared differences from the mean.  
 Standard deviation = square root of variance.

**10. What is a frequency distribution? Give an example.**

It shows how often each value occurs in a dataset.  
👉 Example: Test scores → 5 students scored 50, 8 scored 60, 2 scored 70.

**11. Explain the concept of normal distribution and its characteristics.**

Normal distribution is a bell-shaped curve where most values cluster around the mean.  
**Characteristics:**

* Symmetrical.
* Mean = Median = Mode.
* 68% of data lies within 1 SD, 95% within 2 SDs, 99.7% within 3 SDs.

**12. What is skewness, and how does it affect data interpretation?**

Skewness measures the asymmetry of data.

* **Positive skew**: Tail on the right (income distribution).
* **Negative skew**: Tail on the left (exam marks with many high scores).

**13. What is kurtosis, and what does it tell us about a dataset?**

Kurtosis measures the "peakedness" of data distribution.

* **High kurtosis**: Heavy tails, more outliers.
* **Low kurtosis**: Flat distribution.

**14. Differentiate between probability and statistics.**

* **Probability**: Predicts likelihood of events (before experiment).
* **Statistics**: Analyzes outcomes of data (after experiment).

**15. What is a z-score, and how is it calculated?**

Z-score tells how many standard deviations a value is from the mean.  
 Formula: z=X−μσz = \frac{X - \mu}{\sigma}z=σX−μ​

**16. Explain the difference between population standard deviation and sample standard deviation.**

* **Population SD (σ)**: Uses all data.
* **Sample SD (s)**: Uses sample data, divides by (n–1) to avoid bias.

**17. What is the Central Limit Theorem, and why is it important?**

It states that the sampling distribution of the mean approaches normal distribution as sample size increases.  
👉 Importance: Allows us to use normal distribution for hypothesis testing.

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

Correlation measures the relationship between two variables.

* **Positive**: Both increase (height & weight).
* **Negative**: One increases, other decreases (exercise & body fat).

**19. Explain the difference between correlation and causation.**

* **Correlation**: Two variables move together.
* **Causation**: One variable actually causes the other to change.

**20. What is regression analysis, and when is it used?**

Regression finds the relationship between variables to predict outcomes.  
 Example: Predicting sales based on advertising spend.

**21. Explain hypothesis testing and its steps.**

It tests assumptions about a population using sample data.  
**Steps:**

1. State null & alternative hypothesis.
2. Choose significance level (α).
3. Calculate test statistic.
4. Find p-value.
5. Accept or reject null hypothesis.

**22. What is a null hypothesis and an alternative hypothesis?**

* **Null (H₀)**: No effect or no difference.
* **Alternative (H₁)**: There is an effect or difference.

**23. Explain p-value in hypothesis testing.**

The p-value shows the probability of getting results as extreme as the observed ones, assuming H₀ is true.  
 If p < 0.05, reject H₀.

**24. What is the difference between Type I and Type II errors?**

* **Type I Error**: Rejecting a true null hypothesis (false positive).
* **Type II Error**: Failing to reject a false null hypothesis (false negative).

**25. What is a confidence interval, and how is it interpreted?**

It gives a range of values within which the true population parameter is expected to lie.  
 Example: 95% CI = [48, 52] means we are 95% confident the true mean is between 48 and 52.

**26. Explain t-test and when to use it.**

A t-test compares means of two groups.  
Example: Comparing test scores of two classes.

**27. Explain chi-square test and its applications.**

Chi-square test checks the relationship between categorical variables.  
Example: Testing if gender and product preference are related.

**28. What is ANOVA, and when is it used?**

ANOVA (Analysis of Variance) compares means of 3 or more groups.  
 Example: Comparing average marks of students from 3 schools.

**29. How do you handle missing data in statistics?**

* Remove missing rows.
* Replace with mean/median/mode.
* Use advanced methods like regression or imputation.

**30. What is sampling bias, and how can it be reduced?**

Sampling bias occurs when a sample does not represent the population.  
 Example: Surveying only urban people for a national study.  
**Reduce by:** Random sampling, larger samples, avoiding selective groups.