Statistical Analysis – Interview Questions & Answers

Prepared for: Harsha Hiralal Dhunde • Date: August 28, 2025

1. What is Statistical Analysis?

Statistical analysis is the process of collecting, analyzing, interpreting, and presenting data to identify patterns, trends, and relationships. It includes descriptive statistics (mean, median, mode) and inferential statistics (hypothesis testing, regression).

2. Explain Descriptive vs. Inferential Statistics.

Descriptive statistics summarize data (mean, median, standard deviation, histograms). Inferential statistics generalize from a sample to a population (hypothesis testing, confidence intervals, regression).

3. What is the difference between Mean, Median, and Mode?

Mean is the average; Median is the middle value when sorted; Mode is the most frequent value. For skewed data (e.g., income), the median is often more robust than the mean.

4. What is Variance and Standard Deviation?

Variance measures average squared deviation from the mean; Standard Deviation is its square root and is in the same units as the data. Lower SD indicates more consistent data.

5. What is Correlation vs. Causation?

Correlation (−1 to +1) measures linear association between variables. Causation implies one variable directly affects another. Correlation does not imply causation.

6. What is a p-value in hypothesis testing?

The p-value is the probability of observing results at least as extreme as the data, assuming the null hypothesis is true. If $p < 0.05$, the result is typically considered statistically significant.

7. What is the difference between Type I and Type II Errors?

Type I error (false positive): rejecting a true null. Type II error (false negative): failing to reject a false null. Balancing these requires setting appropriate significance and power.

8. What is Regression Analysis?

Regression models relationships between a dependent variable and one or more independent variables. Linear regression predicts continuous outcomes; logistic regression predicts categorical outcomes.

9. Explain Confidence Interval.

A confidence interval provides a range likely to contain the true population parameter. Example: We are 95% confident the average customer age is between 32 and 36 years.

10. What is ANOVA (Analysis of Variance)?

ANOVA tests whether the means of three or more groups are statistically different by comparing between-group and within-group variances.

11. What is the Central Limit Theorem (CLT)?

CLT states that the sampling distribution of the sample mean approaches normality as sample size grows (often $n \geq 30$), regardless of the population distribution.

12. What is Multicollinearity in regression?

When independent variables are highly correlated, coefficient estimates become unstable. It can be detected using Variance Inflation Factor (VIF) and mitigated via feature selection or regularization.

13. What is the difference between Population and Sample?

Population is the entire set of units of interest; Sample is a subset used for analysis when studying the whole population is impractical.

14. What is an Outlier? How do you handle it?

An outlier is an observation far from others. Handle by verifying data quality, transforming (e.g., log), capping via IQR method, or modeling with robust techniques.

15. What statistical methods have you used in projects?

Descriptive statistics for summarization, correlation analysis, hypothesis testing, and regression for prediction. Moving averages and variance analysis were used in dashboards for business insights.

— End of Document —