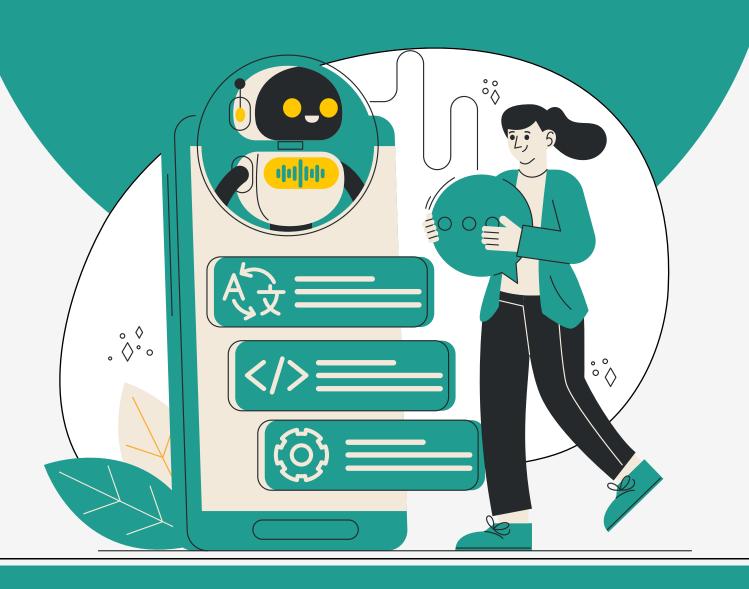# Advanced statistics

Interview Questions

(Practice Project)

Easy

1. What is hypothesis testing, and why is it important in statistics?

Ans: Hypothesis testing is a statistical method used to make inferences about population parameters based on sample data. It helps in determining whether an observed effect is statistically significant or occurred by chance.

2. Explain the difference between null and alternative hypotheses.

Ans: The null hypothesis (H0) represents the status quo or the assumption to be tested. The alternative hypothesis (H1) contradicts the null hypothesis and represents what researchers are trying to find evidence for.

3. What is a Type I error? Can you provide an example?

Ans: A Type I error occurs when the null hypothesis is rejected when it is actually true. For example, in a clinical trial, concluding that a new drug is effective when it has no actual effect.

4. Explain the process of hypothesis testing.

Ans: The process involves formulating null and alternative hypotheses, selecting an appropriate test statistic, determining the significance level, calculating the test statistic and p-value, and making a decision to either reject or fail to reject the null hypothesis based on the p-value.

5. How would you interpret the results of a hypothesis test?

Ans: The interpretation depends on the p-value compared to the chosen significance level. If the p-value is less than or equal to the significance level, the null hypothesis is rejected, and the result is considered statistically significant. Otherwise, the null hypothesis is not rejected.

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

Ans: A p-value is the probability of obtaining test results at least as extreme as the observed results, assuming that the null hypothesis is true. It measures the evidence against the null hypothesis.

7. How do you interpret a p-value?

Ans: A p-value less than the significance level (e.g., 0.05) suggests strong evidence against the null hypothesis, leading to its rejection. A p-value higher than the significance level indicates weak evidence against the null hypothesis, so it is not rejected.

8. What is the significance level (α), and how is it related to the p-value?

Ans: The significance level (α) is a threshold chosen by the researcher, usually 0.05 or 0.01, that determines when to reject the null hypothesis. If the p-value is less than or equal to α, the null hypothesis is rejected.

Medium

9. How would you explain a p-value to someone with no statistical background?

Ans: A p-value is a number that helps us understand whether the results of an experiment or study could have happened by chance. A small p-value means the results are unlikely to be due to random chance, suggesting that there is a real effect or difference.

10. What is a confidence interval?

Ans: A confidence interval (CI) is a range of values, derived from a dataset, that is likely to contain the value of an unknown population parameter. The interval has an associated confidence level that quantifies the level of confidence that the parameter lies within the interval. For example, a 95% confidence interval suggests that if the same population is sampled multiple times, approximately 95% of the intervals will contain the population parameter.

11. How is the confidence interval affected by sample size?

Ans: The width of the confidence interval decreases as the sample size increases. This is because larger samples tend to yield more accurate estimates of the population parameter, reducing the margin of error.

12. What is a significance level in statistical hypothesis testing?

Ans: The significance level, often denoted by α, is the probability of rejecting the null hypothesis when it is actually true. It represents the threshold for determining whether a p-value indicates a statistically significant result.

13. How would you interpret a p-value of 0.03 in the context of a significance level of 0.05?

Ans: If the p-value is 0.03 and the significance level is 0.05, you would reject the null hypothesis because the p-value is less than the significance level. This suggests that there is sufficient evidence to conclude that the observed effect is statistically significant.

14. How does the choice of significance level affect the power of a test?

Ans: The choice of significance level affects the power of a test, which is the probability of correctly rejecting a false null hypothesis. A lower significance level reduces the probability of a Type I error but increases the probability of a Type II error (failing to reject a false null hypothesis), thereby reducing the test's power. Conversely, a higher significance level increases the power but also increases the risk of a Type I error.

15. What is a Z-test and when is it used?

Ans: A Z-test is a statistical test used to determine whether there is a significant difference between sample and population means or between the means of two samples, assuming the data is normally distributed and the population variance is known or the sample size is large (n > 30).

16. Describe the assumptions required for a Z-test to be valid.

Ans: The assumptions for a Z-test to be valid include:

The data is normally distributed, especially important for small sample sizes.
The population variance is known, or the sample size is sufficiently large (n > 30) to justify using the sample standard deviation as an estimate.
The samples are independent of each other.
The data is measured at an interval or ratio scale.

17. What are the different types of T-tests?

Ans: The main types of T-tests are:

One-sample T-test: compares the sample mean to a known value or population mean.
Independent (two-sample) T-test: compares the means of two independent groups.
Paired (dependent) T-test: compares the means of two related groups or the same group at different times.

18. Explain the assumptions necessary for conducting a T-test.

Ans: The assumptions for conducting a T-test are:

The data should be continuous (interval or ratio scale).
The sample data should be drawn from a normally distributed population.
In the case of an independent T-test, the two samples should have equal variances (homogeneity of variance).
The observations should be independent of each other.

19. What is a Chi-Square test and when is it used?

Ans: A Chi-Square test is a statistical test used to determine whether there is a significant association between categorical variables. It is commonly used for testing relationships in contingency tables.

20. What is ANOVA and when is it used?

Ans: ANOVA (Analysis of Variance) is a statistical method used to compare the means of three or more groups to see if at least one of them is significantly different from the others. It is used when comparing multiple groups to understand if there are any statistically significant differences between the means.

21. How do you interpret a significant F-ratio in ANOVA?

Ans: A significant F-ratio indicates that there is a statistically significant difference between the group means. It means that at least one of the groups is different from the others.

22. What is the Central Limit Theorem and why is it important in inferential statistics?

Ans: The Central Limit Theorem states that the sampling distribution of the sample mean will approach a normal distribution as the sample size increases, regardless of the population's distribution. This is important because it allows for the use of normal distribution properties in inferential statistics, even when the population distribution is not normal.