100

Statistics Interview Questions 📚

*Seen in **data scientist** and **data analyst** interviews at FAANGs, startups and consulting firms*

Descriptive Statistics

1. Define mean, median, and mode. How do they differ?
2. What is the difference between population variance and sample variance?
3. How do outliers impact mean and median?
4. Explain the concept of a quartile and interquartile range.
5. What is skewness and kurtosis?
6. How can you measure the spread of data?
7. What is a boxplot and what information does it convey?
8. Describe a situation where the median is more appropriate than the mean.
9. What is a Z-score?
10. Explain the 68-95-99.7 rule.
11. How do you handle missing data?
12. How would you treat outliers in your dataset?
13. What is a percentile?
14. Describe the difference between covariance and correlation.
15. How do you normalize data?
16. What are the benefits of using standard deviation over range?
17. Define the five-number summary in statistics.
18. How would you compare the distributions of two different datasets?
19. Describe the relationship between variance and standard deviation.
20. How is a mode useful when analyzing categorical data?

Probability

1. Define probability.
2. Explain the difference between joint and conditional probability.
3. How is the law of total probability used?
4. Describe Bayes' theorem and its significance.
5. Differentiate between independent and mutually exclusive events.
6. Explain the binomial distribution and when it is used.
7. What is the central limit theorem?
8. How would you use the Poisson distribution in real-world applications?
9. Define expectation and variance for a random variable.
10. What is a probability mass function (PMF)?
11. Explain cumulative distribution function (CDF).
12. Describe the properties of a normal distribution.
13. How is a standard normal distribution different from a normal distribution?
14. What is the law of large numbers?
15. Explain the concept of marginal probability.
16. Describe how the exponential distribution is related to the Poisson distribution.
17. What's the difference between a discrete and a continuous probability distribution?
18. Explain the geometric distribution.
19. When would you use the uniform distribution?
20. How do you compute the expected value of a discrete random variable?

Statistical Testing

1. Define hypothesis testing.
2. What is a null hypothesis and an alternative hypothesis?
3. Explain Type I and Type II errors.
4. What is a p-value?
5. How do you set the significance level in hypothesis testing?
6. What are the assumptions for a t-test?
7. Differentiate between one-sample, two-sample, and paired t-tests.
8. Define a confidence interval.
9. How is the chi-squared test used?
10. What is an ANOVA test?
11. How do power and sample size relate in hypothesis testing?
12. Define the standard error.
13. What is a false discovery rate?
14. How do you test the normality of data?
15. Describe bootstrapping and its advantages.
16. Explain the concept of multicollinearity.
17. When would you use a non-parametric test?
18. How do you correct for multiple comparisons?
19. Explain the difference between one-tailed and two-tailed tests.
20. How do you interpret the results of a linear regression in terms of hypothesis testing?

Regression Analysis

1. What is linear regression?
2. How do you interpret the coefficients of a linear regression model?
3. What assumptions are made in linear regression?
4. Explain the difference between simple and multiple linear regression.
5. What is R-squared and how is it used?
6. How do you check for multicollinearity in a dataset?
7. What is heteroskedasticity and how can it be detected?
8. Explain the concept of regularization in regression.
9. Differentiate between L1 and L2 regularization.
10. How do you handle categorical variables in regression analysis?
11. What is logistic regression and how does it differ from linear regression?
12. How do you evaluate the performance of a regression model?
13. What is the residual plot and what patterns in this plot might indicate a problem with the model?
14. Describe interaction terms in regression and when they might be used.
15. What is the purpose of using polynomial regression?
16. How do you handle missing values in regression analysis?
17. What are the consequences of overfitting in regression?
18. Describe stepwise regression.
19. Explain the concept of a confounding variable.
20. How can you test the linearity assumption in regression?

Experimental Design

1. Define random sampling.
2. What is stratified sampling and when is it used?
3. Differentiate between a sample and a population.
4. Describe the concept of sampling bias.
5. What is cluster sampling?
6. How would you set up an A/B test for a website redesign?
7. Explain the concept of a control group.
8. How do you determine the sample size needed for an experiment?
9. What is the difference between cross-sectional and longitudinal studies?
10. How can you ensure the results of an experiment are statistically significant?
11. Explain the placebo effect.
12. What is a confounding variable in experimental design?
13. Describe the difference between internal and external validity.
14. What is a quasi-experiment?
15. How do you randomize subjects in a clinical trial?
16. What precautions should be taken when conducting multiple A/B tests simultaneously?
17. Explain the concept of experimental power.
18. What is the purpose of blinding in an experiment?
19. Describe the potential pitfalls of convenience sampling.
20. How can you control for confounders in an experimental design?

Join DataInterview.com

Access courses, mock interviewing and private community with coaches and candidates like you

Created by interviewers at top companies like Google and Met