

# 98 Statistics Interview Questions to Hire Top Talent

## Questions

1. What's a p-value? Imagine you're explaining it to a friend who doesn't know stats.
2. Tell me about a time you saw a graph or chart in the news that seemed misleading. What made it seem that way?
3. What's the difference between mean, median, and mode? Can you give me an example where one is better to use than the others?
4. Explain what standard deviation is in simple terms.
5. What is the Central Limit Theorem? Why is it important?
6. How would you explain the difference between correlation and causation to someone?
7. What are Type I and Type II errors in hypothesis testing?
8. Describe a situation where you might use a t-test.
9. What does it mean for data to be normally distributed?
10. What's the difference between descriptive and inferential statistics?
11. If you have a dataset with missing values, what are some ways you could handle them?
12. What is a confidence interval? How do you interpret it?
13. What are some common statistical software packages or programming languages used in data analysis?
14. How would you design a simple survey to collect data on customer satisfaction?
15. What is regression analysis? Give an example of when it might be used.
16. What are outliers? How can they affect your analysis?
17. Explain the concept of variance.
18. Why is sampling important in statistics?
19. What are some different types of data (e.g., nominal, ordinal, interval, ratio)?
20. How do you know if a statistical test is appropriate for a given situation?
21. Describe a time you used data to solve a problem.
22. What are some common data visualization techniques?
23. Explain the meaning of degrees of freedom.
24. What does 'average' mean, and how do you find it?
25. Imagine you have a bag of candies, some red and some blue. How do you figure out which color you're more likely to pick without looking?
26. What's a 'graph', and what can it show you?
27. If someone says something is 'random', what does that mean?
28. Can you explain the difference between a bar chart and a pie chart?
29. What does it mean if something is 'likely' to happen?
30. If you roll a dice, are you more likely to get a 6 or another number?
31. What does it mean to 'predict' something?
32. How would you describe data in simple terms?
33. What is a survey, and what is it used for?
34. If you see a trend, what does it tell you?
35. Explain what a percentage means.
36. Have you ever used a spreadsheet and can you explain how?
37. If I give you a set of numbers, how would you arrange them to make sense of them?
38. Can you explain the term 'more than' and 'less than'?
39. What does it mean for something to be 'most common'?
40. What does 'probability' mean in a simple way?
41. Explain what a line graph shows.
42. How would you show the number of boys and girls in your class, in a simple graph?
43. If you were counting the number of cars that pass by, how would you note it down so you can show your friend later?
44. Explain the difference between qualitative and quantitative data.
45. If you have a lot of data, what is one way to summarize it?
46. What is a 'sample', and why do we use it?
47. What does 'correlation' mean?
48. How would you describe what 'statistics' is, to someone who has never heard of it?
49. Why is it important to collect data carefully?
50. Explain the difference between type I and type II errors in hypothesis testing. Can you provide an example of when each type of error might occur in a real-world scenario?
51. Describe what multicollinearity is and how it can affect a regression model. How can you detect and address multicollinearity?
52. What are the assumptions of linear regression, and how can you check if these assumptions are met in a given dataset?
53. Explain the concept of p-value and its role in hypothesis testing. What are some common misconceptions about p-values?
54. Differentiate between frequentist and Bayesian approaches to statistics. What are the advantages and disadvantages of each approach?
55. Describe the Central Limit Theorem and its significance in statistics. How is it used in statistical inference?
56. What is the difference between correlation and causation? How can you determine if a relationship between two variables is causal?
57. Explain the concept of confidence intervals. How does the confidence level affect the width of the interval?
58. What are the different types of sampling methods? Provide examples of when each method might be appropriate.
59. Describe the concept of statistical power and its relationship to sample size and effect size. How can you increase the power of a statistical test?
60. Explain what regularization is in the context of machine learning. Describe L1 and L2 regularization and their differences.
61. How would you handle missing data in a statistical analysis? Discuss different imputation methods and their potential biases.
62. What is the curse of dimensionality, and how does it affect statistical modeling? What techniques can be used to mitigate it?
63. Describe the difference between parametric and non-parametric statistical tests. When would you use each type of test?
64. Explain the concept of A/B testing and its application in business decision-making. What are some potential pitfalls to avoid when conducting A/B tests?
65. What are some common data visualization techniques, and how can they be used to explore and communicate statistical findings effectively?
66. Describe the concept of time series analysis. What are some common time series models, and when are they appropriate to use?
67. Explain the concept of bootstrapping and its applications in statistical inference. How does it differ from traditional resampling methods?
68. What is survival analysis, and what types of questions can it answer? Describe some common survival analysis techniques.
69. Explain what is the role of statistics in experimentation and design? What are the key components of a well-designed experiment?
70. How would you explain p-values to a non-technical stakeholder, and what are some common misconceptions about them?
71. Describe a situation where you used Bayesian inference. What were the benefits and drawbacks compared to a frequentist approach?
72. How do you handle multicollinearity in a regression model, and what impact does it have on the model's interpretability?
73. Explain the difference between Type I and Type II errors. How do you balance the risk of each in a real-world scenario?
74. Describe your experience with time series analysis. What models have you used, and how did you evaluate their performance?
75. How do you approach outlier detection and treatment in a dataset, and what considerations guide your decisions?
76. What is survival analysis, and in what types of situations would you apply it?
77. Explain the concept of the Central Limit Theorem and its importance in statistical inference. Can you give me a real-world example?
78. How would you design an A/B test to compare two different website designs, and what metrics would you track?
79. Describe your experience with different types of statistical distributions (e.g., normal, Poisson, binomial). When would you use each?
80. How do you assess the goodness-of-fit of a statistical model, and what methods do you use to improve it?
81. What are the key assumptions of linear regression, and how do you check if they are met?
82. Explain the concept of regularization in machine learning. How do L1 and L2 regularization differ, and when would you use each?
83. Describe a situation where you had to deal with missing data. What imputation techniques did you use, and why?
84. How do you approach feature selection in a high-dimensional dataset, and what criteria do you use to evaluate the selected features?
85. Explain the concept of statistical power and its importance in hypothesis testing. How do you calculate the required sample size for a study?
86. How do you validate a statistical model to ensure it generalizes well to new data?
87. Describe your experience with causal inference. What methods have you used to estimate causal effects, and what are the limitations of each?
88. How do you handle imbalanced datasets in classification problems, and what metrics do you use to evaluate model performance?
89. Explain the difference between parametric and non-parametric statistical methods. When would you use each?
90. How would you explain hierarchical modeling to someone without a statistics background, and what are its advantages?
91. Describe a time when you had to communicate complex statistical results to a non-technical audience. What strategies did you use?
92. What are some ethical considerations when working with statistical data, and how do you ensure responsible data analysis?
93. Explain what a statistical interaction is, and how you would test for one in a regression model.
94. Describe a challenging statistical problem you faced and how you overcame it.
95. How do you stay up-to-date with the latest developments in the field of statistics?
96. What is the difference between fixed and random effects in a statistical model, and when is each appropriate?
97. Explain the concept of propensity score matching and how it can be used to reduce bias in observational studies.
98. How would you design a statistical model to predict customer churn for a subscription-based service?