

Statistics Interview Questions

Guidelines

- No Plagiarism will be entertained in the subjective questions
 - The main aim of all the questions where the answer is numeric value is not the answer, but the approach you followed to get that answer so make sure to explain every step in a clean way.
 - Make sure visuals are easily interpretable and less cluttered so follow the rules of data visualization.
 - Give real-world examples as much as you can in the answers. Good Luck!
-

Q-1 Explain Correlation? With real-world examples

Q-2 Briefly Explain why correlation does not imply causation. With real-world examples

Q-3 What is Covariance. And what is the main difference between correlation & Covariance?

Q-4 What are the 2 main Fundamental plots to catch Correlation?

Q-5 What is a z-score? EXPLAIN intuitively.

Q-6 Explain Normal distribution with its properties. With real-world examples

Q-7 How is a Normal distribution different from a Standard Normal Distribution? Is there a way to convert it into one another? If Yes, Explain briefly that method.

Q-8 What are Outliers? Explain with a mathematical term.

Q-9 What is a left-skewed and right-skewed distribution. Explain Briefly

Q-10 What is a Symmetric distribution? Give all examples of that distribution with some visuals.

Q-11 What is a relationship between mean, the median in Normal Distribution?

Q-12 What do these terms Bell Curve and Gaussian Distribution mean?

Q-13 How is a box-plot different from a scatterplot?

Q-14 What is the difference between Descriptive and Inferential statistics?

Q-15 What is a population and sample in inferential statistics. Explain briefly with good examples.

Q-16 Let A and B be events on the same sample space, with $P(A) = 0.3$ and $P(B) = 0.8$. Can these two events be disjoint?

Q-17 Can we use a z-table for a skewed distribution? Explain

Q-18 Which statistic is most sensitive to outlier – Mean, Median or mode?

Q-19 What is a percentile?

Q-20 Explain the effect of Shifting linear Transformation on mean, median, S.D. & IQR of data?

Q-21 Explain the effect of Scaling linear Transformation on mean, median, S.D. & IQR of data?

Q-22 What's Density Curve. Give a diagram also.

Q-23 Briefly explain the 'Empirical Rule' in statistics. Explaining with a diagram is mandatory, preferably with real-world data examples.

Q-24 What is the mean and S.D. of a Standard normal distribution.

Q-25 Lifespan of African Lizards is normally distributed with mean = 3.1yrs and S.D. = 0.6yrs . Find the Prob. Of Lizard living between 2.5 yrs to 4.3 yrs. Explain visually with Normal distribution.

Q-26 Consider a Normal distribution of student's heights with mean = 15cm and S.D. = 20cm. Rohan's height = 161.4cm.

What Prob. Of students have height less than Rohan's? Explain visually with Normal distribution

Q-27 Consider a Normal distribution of student's heights with mean = 15cm and S.D. = 20cm. Rohan's height = 161.4cm.

What Prob. Of students have height more than Rohan's? Explain visually with Normal distribution

Q-28 What is the Prob of a student having the same height? Give an approximate value (Hint: No calc. is needed here)

Q-29 A set of laptop prices are normally dist. Mean = \$750 and S.D. = \$60. What is the prob? Of lap prices between \$624 and \$768. Explain visually with Normal distribution.

Q-30 Consider Normal dist. Of pulse rates. Mean = 80, S.D. = 9. Additional testing is to be done for the top 40%. Find the minimum pulse rate among those top 40%? Explain visually with Normal distribution

Q-31 Normal dist of avg. wait time for various roads in Lucknow. Mean = 185 secs and S.D. = 11 secs. Anand wants to take the road that is in the bottom 10% of Lucknow. What is the max wait time in that bottom 10%? Explain visually with Normal distribution

Q-32 State the formula for the correlation coefficient

Q-33 Explain each part of the formula you mentioned above, preferably with an example.

Q-34 Ram has 2 children and 1 of them is a girl. What is the probability that the other child is also a girl?

You can assume that there is an equal number of males and females in the world. Define sample spaces properly.

Q-35 A fair six-sided die is rolled twice. What is the probability of getting 2 on the first roll and not getting 4 on the second roll?

Q-36

$$P(A \cup B \cup C) = P(A \cap C^c) + P(C) + P(B \cap A^c \cap C^c)$$

Is it True or False? Explain your answer.

Q-37 Consider a 4-sided die and roll it twice. What is the probability that the number on the first roll is strictly higher than the number on the second roll?

Define sample space properly.

Q-38 Which of the following can't be a Probability event:

- 0.1
- -0.04
- 0.5
- 0.99999
- 1.0001
- 1/89
- 0

Explain why's with your answer.

Q-39 Urvin randomly picks 4 cards from a deck of 52-cards and places them back into the deck (Any set of 4 cards is equally likely). Then, Amit randomly chooses 8 cards out of the same deck (Any set of 8 cards is equally likely). Assume that the choice of 4 cards by Urvin and the choice of 8 cards by Amit are independent. What is the probability that all 4 cards chosen by Urvin are in the set of 8 cards chosen by Amit? Explain

Context: A player is randomly dealt a sequence of 13 cards from a deck of 52-cards. All sequences of 13 cards are equally likely. In an equivalent model, the cards are chosen and dealt one at a time. When choosing a card, the dealer is equally likely to pick any of the cards that remain in the deck.

Q- 40 If you dealt 13 cards, what is the probability that the 13th card is a King?

Q-41 A fair six-sided die is rolled 6 times. What is the probability of getting all outcomes as unique?

Q-42 A group of 60 students is randomly split into 3 classes of equal size. All partitions are equally likely. A and B are two students belonging to that group. What is the probability that A and B will end up in the same class?

Q-43 We have two coins, A and B. For each toss of coin A, the probability of getting head is $1/2$ and for each toss of coin B, the probability of getting heads is $1/3$. All tosses of the same coin are independent. We select a coin at random. and toss it till we get a head. The probability of selecting coin A is $\frac{1}{4}$ and coin B is $3/4$. What is the expected number of tosses to get the first heads?

Q-44 How would u calculate range & interquartile range.

Q-45 Differentiate between quantitative and qualitative data.

Q-46 Explain with examples -- Categorical ordinal and nominal variables

Q-47 Explain with examples – Numerical Continuous and Discrete data.

Q-48 What is the five-number summary and how is it useful?

Q-49 Suppose a life insurance company sells a \$240,000 one-year term life insurance policy to a 25-year-old female for \$210. The probability that the female survives the year is .999592. Find the expected value of this policy for the insurance company.

Q-50 When can a probability event be independent of itself?

Q-51 Name and explain any 3 outlier detection techniques.

Q-52 Name and explain any 3 outlier handling techniques

Q-53 What would be your approach if I ask you the mean length of all road trips in India? (With a confidence of 95%)

Q-54 Can you relate Standard deviation with Standard error. Explain

Q-55 What are the various effects/trade-offs of changing the width of Confidence Intervals?

Q-56 On increasing Margin of error, what happens to C.I.?

Q-57 What is the difference between 95% confidence and 99% confidence. Can you give a scenario for both?

Q-58 Explain the Degree of freedom

Q-59 What do you think will happen if DF (deg. Of freedom) is large?

Q-60 What is Statistical power?

Q-61 What are a z-test and t-test? Explain briefly

Q-62 When to use which distribution/test? (z & t-test)

Q-63 What are null and alternate hypotheses?

Q-64 Give examples for the above question for one-sided and 2-sided tests.

Q-65 What is the p-value? Explain briefly with example and visual

Q-66 How would you conduct a hypothesis test? Briefly tell every step.

Q-67 Calculate the p-value using any of your favourite tool (python, excel etc) and describe it here.

Q-68 What do we mean by – making a decision based on comparing p-value with significance level?

Q-69 If the null hypothesis consists of a constant value only, what type of test is it you think? Explain in your best way.

Q-70 Do you think the critical values of 95% one tail and 90% two tail are same. If yes, why?

Q-71 Cross-fertilizing a red and a white flower produces red flowers 25% of the time. Now we cross-fertilize five pairs of red and white flowers and produce

five offspring. What is the probability that there are no red flower plants in the five offspring?

Q-72 A roulette wheel has 38 slots – 18 red, 18 black, and 2 green. You play five games and always bet on red slots. How many games can you expect to win?

Q-73 A roulette wheel has 38 slots, 18 are red, 18 are black, and 2 are green. You play five games and always bet on red. What is the probability that you win all the 5 games?

Q-74 Some test scores follow a normal distribution with a mean of 18 and a standard deviation of 6. What proportion of test takers have scored between 18 and 24?

Q-75 A jar contains 4 marbles. 3 Red & 1 white. Two marbles are drawn with replacement after each draw. What is the probability that the same colour marble is drawn twice?

Q-76 Which of the following events is most likely? Give why

- A) At least one 6, when 6 dice are rolled
- B) At least 2 sixes when 12 dice are rolled
- C) At least 3 sixes when 18 dice are rolled
- D) All the above have same probability

Q-77 What is the main difference between Overfitting and underfitting?

Q-78 What is selection bias in statistics?

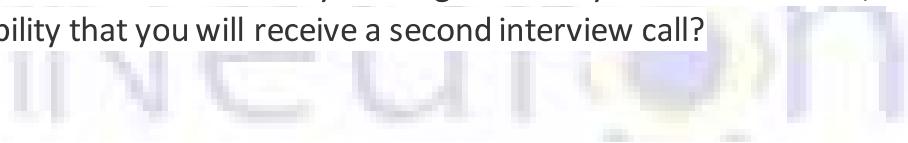
Q-79 What is a Sampling Distribution?

Q-80 What is the Law of Large Numbers?

Q-81 What is the Central Limit Theorem? And what's the use?

Q-82 Briefly explain type-I and type-II Error? Give diagrams also.

Q-83 Suppose you were interviewed for a technical role. 50% of the people who sat for the first interview received the call for second interview. 95% of the people who got a call for second interview felt good about their first interview. 75% of people who did not receive a second call, also felt good about their first interview. If you felt good after your first interview, what is the probability that you will receive a second interview call?



Q-84 A coin of diameter 1-inches is thrown on a table covered with a grid of lines each two inches apart. What is the probability that the coin lands inside a square without touching any of the lines of the grid? You can assume that the person throwing has no skill in throwing the coin and is throwing it randomly.

You can assume that the person throwing has no skill in throwing the coin and is throwing it randomly.

Q-85 Consider the following probability density function: What is the probability for $X \leq 6$ i.e. $P(X \leq 6)$

$$f(x) = \frac{1}{8} e^{-x/8} \text{ for } x \geq 0$$

What is the probability for $X \leq 6$ i.e. $P(X \leq 6)$

Q-86 In a class of 30 students, approximately what is the probability that two of the students have their birthday on the same day (defined by same day and month) (assuming it's not a leap year)?

For example – Students with birthday 3rd Jan 1993 and 3rd Jan 1994 would be a favourable event.

Q-87 Assume you sell sandwiches. 70% people choose Veg, and the rest choose non-veg. What is the probability of selling 2 veg sandwiches to the next 3 customers?

Q-88 Anand is playing a lottery game where he must pick 2 numbers from 0 to 9 followed by an English alphabet (from 26-letters). He may choose the same number both times.

If his ticket matches the 2 numbers and 1 letter drawn in order, he wins the grand prize and receives \$10405. If just his letter matches but one or both of the numbers do not match, he wins \$100. Under any other circumstance, he wins nothing. The game costs him \$5 to play. Suppose he has chosen 04R to play.

What is the expected net profit from playing this ticket?

Context: HIV is still a very scary disease to even get tested for. The US military tests its recruits for HIV when they are recruited. They are tested on three rounds of Elisa(an HIV test) before they are termed to be positive.

The prior probability of anyone having HIV is 0.00148. The true positive rate for Elisa is 93% and the true negative rate is 99%.

Q-89 What is the probability that a recruit has HIV, given he tested positive on first Elisa test? The prior probability of anyone having HIV is 0.00148. The true positive rate for Elisa is 93% and the true negative rate is 99%.

Q-90 What is the probability of having HIV, given he tested positive on Elisa the second time as well.

The prior probability of anyone having HIV is 0.00148. The true positive rate for Elisa is 93% and the true negative rate is 99%.

Q-91 Suppose you're playing a game in which we toss a fair coin multiple times. You have already lost thrice where you guessed heads but a tails appeared. Which of the below statements would be correct in this case?

- A) You should guess heads again since the tails has already occurred thrice and its more likely for heads to occur now
- B) You should say tails because guessing heads is not making you win
- C) You have the same probability of winning in guessing either, hence whatever you guess there is just a 50-50 chance of winning or losing
- D) None of these

Q-92 What is Bayesian?

Q-93 What is Frequentist?

Q-94 What is Likelihood?

Q-95 The inference using the frequentist approach will always yield the same result as the Bayesian approach. **Yes/no with Reason**

Q-96 What is Quantile? Is there a difference between quantile and quartile?

Q-97 What is the difference between skewness and kurtosis? Give answer explaining visually

Q-98 What is Moment?

Q-99 State the binomial probability formula.

Q-100 Hospital records show that 75% of patients suffering from a disease die due to that disease. What is the probability that 4 out of the 6 randomly selected patients recover?

Q-101 Give and explain the types of Statistical Sampling techniques.

Q-102 What is the main difference between Data-Mining and Data Profiling?

Q-103 How would you deal with a multi-source data problems?

Q-104 What are types of Missing values.

Q-105 For each type stated above explain what it is with a real world example.

Q-106 Can a machine learning algorithm be used for imputing missing values. If yes, then explain a few methods/algos.

Q-107 What is KPI. Give a real world example

Q-108 What is Design of experiment? (w.r.t Statistics)

Q-109 What is the 80/20 rule of Businesses?

Q-110 What is Hot-Deck and Cold-Deck Imputation?

Q-111 What is the probability of throwing two fair dice when the sum is 5 and 8?

Q-112 State the case where median is better than mean for imputing missing data? Give a real world example.

Q-113 What is six-sigma in statistics?

Q-114 Given a left-skewed distribution that has a median of 60, what conclusions can we draw about the mean and the mode of the data?

Q-115 Discuss some scenarios where you will keep the outliers in data?

Q-116 If there is a 30% prob. Of seeing a car in any 20-min interval . State the prob. Of seeing atleast one car in an hour?

Q-117 Are Symmetric distributions always unimodal?

Q-118 What is Bessel's Correction?

Q-119 Differentiate between Univariate, Bivariate and Multivariate analysis?

Q-120 What types of variables are used for Pearson's correlation coefficient?

Q-121 In a scatter diagram, what is the line that is drawn above or below the regression line called?

Q-122 In an observation, there is a high correlation between the time a person studies and the amount of marks he gets. What can be inferred from this?

Q-123 What type of data does not have a log-normal distribution or a Gaussian distribution?

Q-124 What would you do if your data is not normal? Discuss your approach in brief here.

Q-125 Any 5 types of data transformations techniques to turn features to normal features.

Q-126 The students of a particular class were given two tests for evaluation. Twenty-five percent of the class cleared both the tests and forty-five percent of the students were able to clear the first test.

Calculate the percentage of students who passed the second test given that they were also able to pass the first test.

Q-127 While it is said that the probabilities of having a boy or a girl are the same, let's assume that the actual probability of having a boy is slightly higher at 0.51. Suppose a couple plans to have 3 children. What is the probability that exactly 2 of them will be boys?

Q-128 Heights of 10 year-olds, regardless of gender, closely follow a normal distribution with mean 55 inches and standard deviation 6 inches. Which of the following is true?

- A) We would expect more number of 10 year-olds to be shorter than 55 inches than the number of them who are taller than 55 inches
- B) Roughly 95% of 10 year-olds are between 37 and 73 inches tall
- C) A 10-year-old who is 65 inches tall would be considered more unusual than a 10-year-old who is 45 inches tall
- D) None of these

Q-129 About 30% of human twins are identical, and the rest are fraternal. Identical twins are necessarily the same sex, half are males and the other half are females. One-quarter of fraternal twins are both males, one-quarter both female, and one-half are mixed: one male, one female. You have just become a parent of twins and are told they are both girls. Given this information, what is the probability that they are identical?

Q-130 Rob has fever and the doctor suspects it to be typhoid. To be sure, the doctor wants to conduct the test. The test results positive when the patient actually has typhoid 80% of the time. The test gives positive when the patient does not have typhoid 10% of the time. If 1% of the population has typhoid, what is the probability that Rob has typhoid provided he tested positive?

Q-131 Arun is having two coins in his hand. Out of the two coins, one is a real coin and the second one is a faulty one with Tails on both sides. He blindfolds himself to choose a random coin and tosses it in the air. The coin falls down with Tails facing upwards. What is the probability that this tail is shown by the faulty coin?

Q-132 A fly has a life between 4-6 days. What is the probability that the fly will die at exactly 5 days ?

Q-133 What type of graph would you use for trend analysis. Give a visual also.

Q-134 Explain Survivorship bias.

Q-135 What are confounding variables?

Q-136 What is a point estimate (in inferential stats)?

Q-137 Explain the Experiment Design types in statistics. Give examples if you can.

Q-138 What is the Conditional Probability? Explain with example

Q-139 What is Bayes theorem, give formula also, explaining it. Explain with example.

Q-140 Difference between Binomial and Bernoulli distribution.

Q-141 What is a Poisson Distribution?

Q-142 Differentiate between Random Variables & Geometric Random variables.

Q-143 Give the formula for computing confidence interval for mean for a sample.

Q-144 Give the formula for computing confidence interval for proportion for a sample.

Q-145 Discuss the Assumption of Confidence interval in brief.

Q-146 State the assumption of Central Limit theorem.

Q-147 What is bootstrapping ? When is it useful in statistics?

Q-148 What is Simple Linear regression? How is it Different from Multiple Linear regression? Explain using real world examples.

Q-149 What is regression analysis? State the four most fundamental assumptions of regression? (you can tell more if you know)

Q-150 Explain the above assumptions in detail with examples.

Q-151 What is an A/B Test?

Q-152 Explain the steps involved in conducting an A/B Test?

