Mithilesh singh
Basic Statistics Test
Time duration 2 hours

BASIC Statistics Test Questions for Data Analyst

Choose the correct answer: -

Q1. You are a data analyst conducting a study on color blindness and you question 550 people. 110 of them have brown eyes and 54% of them have blue eyes. What percentage of the people you questioned has blue or brown eyes?

(A) 74%
(B) 64%
(C) 84%
(D) 54%

Q2. In which situation is a bar graph preferred over a pie chart?

(A)When there are some large categories in the data.
(B)When the number of categories in the data is low.
(C)When one of the categories in the data is really large.
(D) When the number of categories in the data is high.

Q3. **What is Bessel's correction** ?

(A)A process of hypothesis testing
(B)A process to split sample from a population.
(C) Factor that is used to estimate a populations' standard deviation from its sample
(D) Variables to be used for the correlation coefficient .

Q4. You have taken test of 10 students. Their scores were: 5, 7, 2, 1, 3, 4, 8, 8, 6, 6. What is the interquartile range (IQR)?
(A)4
(B)5.5
(C)5
(D)8

Q5 If the mean and the standard deviation of a continuous random variable that is normally distributed are 20 and 5, respectively, find an interval that contains around 68% of the distribution.
A. (18,24)
B. (15,25)
C. (12,25)
D. (10,30)

Q6 The mean and the SD of a set of variables(normally distributed) are 50 and 5, respectively, find an interval that contains around 82% of the distribution.
A. (60,55)
B. (45,60)
C. (45,55)
D. (50,60)

Q7 What are population and sample in Inferential Statistics, and how are they different?

Statement A. We calculate the statistics using the sample.
Statement B. Using these sample statistics, we make conclusions about the population
Statement C. Margin of error depends on Population size , sample size has no impact on Margin of error..

(A)All statement True.
(B)statement A false.
(c)statement B false
(D)statement C false

Q.8 What is the meaning of the five-number summary in Statistics?

A. Mean,Median,Mode,SD,variance
B. Mean, Medican,Mode,Q1,Q3
C. Low extreme,Q1,Median,Q3,High extreme.
D. None of the above

Q.9 . The heights of adult women are normally distributed with a mean of 62.5 inches and standard deviation of 2.5 inches. Use the Empirical Rule to determine between what two heights 68% of adult women will fail.

A. (52.5,72.5)
B. (55,70)
C. (57.5,67.5)
D. (60,65)

Q 10 The formula to calculate standardized normal random variable is
(A)x - μ/σ
(B)x + μ/σ
(C)x - σ/μ
(D)x + σ/μ

Q11 Consider the probability distribution as standard normal, if the value of μ is 75, the value of x is 120 with the unknown standard deviation of distribution then the value of z-statistic

(A)will be one
(B)will be zero
(C)will be negative
(D)will be positive

Q12 If the value of x is less than μ of standard normal probability distribution then the

(A)z-statistic is negative
(B)z-statistic is positive
(C)f(x) will be even number
(D)f(x) will be prime number

Q 13 The standard normal probability distribution has mean equal to 40, whereas the value of random variable x is 80 and the z-statistic is equal to 1.8 then the standard deviation of standard normal probability distribution is

(A)120
(B)80
(C)40
(D)20

Q 14 In a symmetrical distribution, the third quartile and first quartile of data in distribution must

(A)be at equal distance
(B)not be at equal distance
(C)positive value concentration
(D)negative value concentration

Q 15 A machine produces electrical components.
99.7% of the components have lengths between 1.176 cm and 1.224 cm.
Assuming this data is normally distributed, what are the mean and standard deviation?

(A)Mean = 1.210 cm
S.D. = 0.008 cm
(B)Mean = 1.200 cm
S.D. = 0.004 cm
(C)Mean = 1.190 cm
S.D. = 0.008 cm
(D)Mean = 1.200 cm
S.D. = 0.008 cm

Q 16 In a factory, the weight of the concrete poured into a mold by a machine follows a normal distribution with a mean of 1150 pounds and a standard deviation of 22 pounds. Approximately 95% of molds filled by this machine will hold weights in what interval?

(A)1084 to 1216 pounds
(B)1106 to 1150 pounds
(C)1106 to 1194 pounds
(D)1128 to 1172 pounds

Q17 A 120 ml can of soda has a mean volume of 120 ml, with a standard deviation of 25 ml. How common are cans with less than 115 ml of soda? Calculate the probability.

(A)2.5%
(B)0.15%
(C)2.35%
(D)2.25%

Q 18 **What is the relationship between mean and median in a normal distribution?**

(A) In a normal distribution, the mean is equal to the median
(B) the mean is greater than median
(C) the mean is less than median

Q19 At a local high school, GPA's are normally distributed with a mean of 2.9 and standard deviation of 0.6. What is the GPA of the highest 2.5% of the students?
(A)4.1
(B)4.1 or higher
(C)4.7
(D)4.5 or higher

Q 20 State the case where the median is a better measure when compared to the mean.

Statement(A):-Median is preferred In the case where there are a lot of outliers that can positively or negatively skew data.

(A) statement A is True

(B) (B)statement A is False

Q.21 Variable A is normally distributed with $\mu = 12.00$ and $\sigma = 3.11$. What is the probability that a randomly selected case will have a score of less than 15?

(A)0.72
(B)0.29
(C)0.87
(D)0.12

Q.22 what is central limit theorem ?

Statement A . normal distribution is arrived at when the sample size varies without having an effect on the shape of the population distribution

Statement B The mean of samples will follow a normal distribution normal distribution is arrived at when the sample size varies without having an effect on the shape of the population distribution .

(A)Both statements A and B are True.
(B)Statement A is True , but Statement B is False.
(C)Both statements are False.

Q.23 Students pass a test if they score 50% or more.
The marks of a large number of students were sampled and the mean and standard deviation were calculated as 42% and 8% respectively.
Assuming this data is normally distributed, what percentage of students pass the test?

(A)5
(B)16
(C)24
(D)32

Q.24The mean June midday temperature in Delhi is 36°C and the standard deviation is 3°C
Assuming this data is normally distributed, how many days in June would you expect the midday temperature to be between 39°C and 42°C?
(A)3
(B)4
(C)7
(D)14

Q.25 In a population, if μ=125 and σ=25, how common are values in the 100 – 175 range?

(A)68%
(B)84%
(C)81.5%
(D)95%

Q26 Common visual technique used for univariate analysis is….
(A)Histogram.
(B)Scatter Plot.
(C) Pair Plot.

Q27 Common visual technique used for bi-variate analysis is….
(A)Histogram.
(B)Scatter Plot.
(C) Pair Plot.

Q28. A professor uses the following formula to grade a statistics exam:

y-hat = 0.5 + 0.53x. After obtaining the results the professor realizes that the grades are very low, so he might have been too strict. He decides to level up all results by one point. What will be the new grading equation?

(A)y-hat = 1.5 + 1.53x
(B)y-hat = 1.5 + 0.53x
(C)y-hat = 0.5 + 0.53x
(D)y-hat + 1 = 0.5 + 0.53x

Q29.What can you conclude about a Pearson's r that is bigger than 1?

(A)The correlation is very high.

(B)This is impossible. Correlations are always between -1 and 1.

(C)There is a non-linear relationship between X and Y.

(D)This is impossible. Correlations are always between 0 and 1.

Q.30 The Bayes rule can be used best in……

(A)Solving queries

(B)Dimensionality Reduction.

(c)Decreasing complexity

(D)Answering Probabilistic query

Q31.Why do you use squared residuals when computing the regression line?

(A)To balance the results (because you take the squared root later on).

(B)Because the residuals equal each other out (i.e. their sum equals zero).

(C)To make the differences between the predicted values and the real data-points even clearer.

(D)Because you have an X-value and a Y-value for every data point.

Q32. You roll a dice five times. The outcomes are: 6 6 6 6 6. Then you repeat this and you find: 1 4 3 5 2. Which of the following outcomes is most likely?

(A)The first outcome is more likely.
(B)The second outcome is more likely.
(C)Both outcomes are equally likely

Q33. You want to know how many hours of sleep new parents lose after they had their first baby. You know that the population mean equals 2.3 hours. Because you can't investigate the whole population, you take a sample of 100 people. You find an average sleep loss of 2.1 hours. What is, based on this sample, the point estimate of your population mean?

(A)0.2
(B)2.3
(C)5.4
(D)2.1

Q34.Which assumptions must need to be satisfied for the construction of a confidence interval for a mean?

(A)The independent variable must be discrete.

(B)The sample must be random.

(C)The relationship between X and Y must be linear.

(D)The sample mean must be equal to the population mean.

Q35. You draw a sample from the population of Gujrat voters. You do this by randomly selecting 100 voters from each municipality. What kind of sample is this?

(A)Stratified random
(B)Cluster random
(C)Random
(D)Convenience

Q36. A type I error means that:

(A)The null hypothesis is true, and you do not reject the null hypothesis.
(B)The null hypothesis is true, and you reject the null hypothesis.
(C)The null hypothesis is false, and you reject the null hypothesis.
(D)The null hypothesis is false and cannot reject the null hypothesis.

Q37. Relationship between confidence level and confidence interval.

(A)With increased in confidence level the confidence interval range increase.
(B) With increased in confidence level the confidence interval range decrease.
(c) No impact on confidence interval range with respect to confidence level.

Q38. If the null hypothesis is false then which of the following is accepted?
A) Null Hypothesis
B) Positive Hypothesis
C) Negative Hypothesis
D) Alternative Hypothesis.

Q 39 If we do not reject the null hypothesis, we conclude that:

A there is enough statistical evidence to infer that the alternative hypothesis is true
B there is not enough statistical evidence to infer that the alternative hypothesis is true
C there is enough statistical evidence to infer that the null hypothesis is true
D the test is statistically insignificant at whatever level of significance the test was conducted at

Q40 The purpose of hypothesis testing is to:

A test how far the mean of a sample is from zero
B determine whether a statistical result is significant
C determine the appropriate value of the significance level
D derive the standard error of the data

41.The p-value obtained from a classical hypothesis test is:

A the probability that the null hypothesis is true given the data
B the probability that the null hypothesis is false given the data
C the probability of observing the data or more extreme values if the null hypothesis is true
D the probability of observing the data or more extreme values if the alternative hypothesis is true

Q42. Which of the following statement about hypothesis testing is true?

A If the p-value is greater than the significance level, we fail to reject Ho
B A Type II error is rejecting the null when it is actually true
C If the p-value is greater than the significance level, we reject Ho
D None of the above statements are true

Q43.Suppose that we reject a null hypothesis at the 5% level of significance. For which of the following level of significance do we also reject the null hypothesis?

A 6%
B 2.5%
C 4%
D 3%

Q.44 Which of the following p-values will lead us to reject the null hypothesis if the significance level of the test if 5%?

A 0.15
B 0.10
C 0.06
D 0.025

Q. 45 If a hypothesis is rejected at the 0.025 level of significance, it:

A Must Be Rejected At Any Level
B Must Be Rejected At The 0.01 Level
C Must Not Be Rejected At The 0.01 Level
D May Or May Not Be Rejected At The 0.01 Level

Q.46 A feature F1 can take certain value: A, B, C, D, E, & F and represents grade of students from a college. Here feature type is_____

(A)Ordinal

(B)Nominal

(C)Categorical

(D)None of the above.

Q.47. If machine learning model output involves target variable, then that model is called as_____

A. predictive model

B. descriptive model

C. reinforcement learning

D. All of above

Q.48 What do you understand by Hypothesis in the content of Machine Learning?

Statement I : - In machine learning, a hypothesis represents a mathematical function that an algorithm uses to represent the relationship between the target variable and features.

Statement II :- In machine learning, a hypothesis don't represent a mathematical function of any algorithm.

(A)Statement I is true

(B)Statement II is true

Q.49 What is multi-collinearity

Statement I :- A single dependent variable depends on several independent variables is called Multi collinearity.

Statement II:- Independent variables are deduced to possess high correlations with each other is called Multi collinearity.

(A)Statement I is true Statement II is False.

(B)Statement I is False Statement II is True

(c)Both statements are True.

(D)Both Statements are False.

Q.50 What is a Receiver Operating Characteristic (ROC) curve?

(I) A ROC curve is a graph that plots True Positive Rate vs False Positive Rate.

(II)It displays the performance of a classification algorithm at all classification thresholds.

(III)It is dedicated to only few classification algorithms; we can not apply in all classification problem.

(IV)Curves that are pointed towards the left corner of the plot belong to bad classifiers.

(A)Statement iv is false.

(B)statement ii and III are is false

(c) All are True

(D)Statement III and IV are False.