Please submit your answer on March 26th.only

BASIC Statistics test   **No. of questions 40   Time Duration:- 90 mins**

Write the answer sheet as with correct answer only:-
For example-
1-B
2-D
…..

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? [Your answer should consist of just the number, no additional characters – so if you think the answer is 41% enter the number 41]

(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. Ten students completed an exam. Their scores were: 5, 7, 2, 1, 3, 4, 8, 8, 6, 6. What is the Median?

(A)4
(B)5.5
(C)5
(D)6.5

Q4. Ten students completed an exam. 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 50 and 5, respectively, find an interval that contains 68% of the distribution.
A. (60,55)
B. (55,45)
C. (45,55)
D. (50,60)

Q6. If the mean and the standard deviation of a continuous random variable that is normally distributed are 28 and 3 respectively, find an interval that contains approximate 82% of the distribution.

A. (25,34)
B. (25,31)
C. (20,35)
D. (19,37)

Q.7 A competency test has scores with a mean of 80 and a standard deviation of 10. A histogram of the data shows that the distribution is normal. Use the Empirical Rule to find the percentage of scores less than 80.

A. 99.7%
B. 95%
C. 68%
D. 50%

Q 8. The formula to calculate the StandardScaler

(A)x - μ / σ

(B)x + μ / σ

(C)x - σ / μ

(D)x + σ / μ

Q9 Assume 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. Consider the probability distribution as standard normal

(A)will be one

(B)will be zero

(C)will be negative

(D)will be positive

10. Let's Assume you are in a senior position of a company making a decision to Acquire company A or B.

company A which has $ 21 Million Value,
 company B which has $ 15 Million Value.
Deal price is same say $ 15 Million for company A & B.
However, the problem is that if you acquire company A there is 90% chance that the Government will challenge the acquisition (some Taxes irregularity as challenged) and a 60% chance the government will win.
If Government win a value of A drops to $ 14 Million because of Legal Fees and Pending Taxes.
Even If government losses, the value drops to $ 19 Million because of Legal fees.
Acquiring which company is the best deal? What is the effective Value?

(A)Company A because effective value is 16-17 Million USD
(B)Company B
(c)Company A because effective value is 17-17.5 Million USD
(D)Company A because effective value is 15.5-16.0 Million USD

Q11.

In a school there are 1000 students, out of which 430 are girls, it is known that out of 430 girls 10% girls are studying in class 12 .

A student chosen randomly from the school, what is the probability that chosen one is a student of class 12?.

It is given that the chosen student is a girl.

WRITE the correct percentage: -

Q12 Which of the following statement is true, for a Standard normal Probability Distribution. If the value of x is less than μ.

(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 Calculate the standard deviation of standard normal probability distribution where mean equal to 40, whereas the value of random variable x is 80 and the z-statistic is equal to 1.8 , the SD will be -

(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

Q15 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%

Q16 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 approximately 2.5% of the students?
(A)6.1 or higher
(B)4.1 or higher
(C)4.7
(D)4.5 or higher

Q 17 The mean life of a tire is 30,000 km. The standard deviation is 2000 km. Then, 84% of all tires will have a life between _____ km and _____ km.

(A)30,000 km and 32,000 km.
(B)24,000 km and 34,000 km.
(C)26,000 km and 34,000 km.
(D)27,000 km and 31,000 km.

Q.18 Variable A is normally distributed with μ = 12.00 and σ = 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.19 The 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

Q20 Common visual technique used for univariate analysis is....
(A)Histogram.
(B)Scatter Plot.
(C) Pair Plot.

Q21 Common visual technique used for bi-variate analysis is....
(A)Histogram.
(B)Scatter Plot.
(C) Pair Plot.

Q22. 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

Q23. How can you interpret about a Pearson's r that is more than 1?

(A)The correlation is very high.

(B) Correlations are always between -1 and 1.

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

(D) Correlations are always between 0 and 1.

Q24. 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

Q25. You are doing a survey that how many hours of sleep new parents lose after they had their first baby. You know that the population mean equals 2.1 hours. Because you can't investigate the whole population, you take a sample of 200 people. You find an average sleep loss of 1.9 hours. What is, based on this sample, the point estimate of your population mean?

(A)1.9
(B)0.2
(C)4.0
(D)2.1

Q26. To calculate the confidence interval for a mean Which assumptions must need to be satisfied?

(A)The confidence level must be 90%.

(B)The sample must be random.

(C)The confidence Level must be 99%.

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

Q27. 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.

Q28. 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.

Q29. If we do not reject the null hypothesis, we conclude that?
A) there is not enough statistical evidence to infer that the alternative hypothesis is true
B there is 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.

Q30.Suppose that we reject a null hypothesis at the 5% level of significance. For which of the following level of significance do we MUST reject the null hypothesis?

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

31. When both variables increase or decrease concurrently and at a constant rate,

(A)a positive linear relationship exists.
(B)a negative linear relationship exists.
(C)No relationship exists.
(D)Both positive and Negative relationship exists.

32. In a monotonic relationship, the variables tend to move in the same relative direction, but not necessarily at a constant rate.

(A)True
(B)False

33. These are the values that occur most frequently.
Player #1 #2 #3 #4 #5 #6 #7 #8 #9 #10
Home Runs 8 8 11 12 15 15 17 19 19 27

The following dataset has
(A)Mode values are: 8, 15, 19.
(B)Mode is 19
(c)No mode in this dataset.

34. In hypothesis testing
Statement I
If Ha has the form $\mu<\mu0$ the test is called a left-tailed test.
Statement II
If Ha has the form $\mu>\mu0$ the test is called a right-tailed test
(A)Statement I is True
(B)Statement II is True
(c) Both statements are True
(D)Both statements are False

(35)The width of a confidence interval increases when the margin of error increases, which happens when the:

- (I)Significance level increases;

- (II)Sample size decreases;

- (III)Sample variance increases.

  (A)statement I is true

  (B)Statement II is true

  (C)Statement III is true.

  (D)All statements are True

(36)The Bayes rule can be used best in……
(A)Solving queries
(B)Dimensionality Reduction.
(c)Decreasing complexity
(D)Answering Probabilistic query

(37)   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

(38)statement I :- A skewed distribution is a distribution where the values in the dataset are not normalized and the distribution curve is inclined towards one side.

Statement II:-A uniform distribution on the other hand is a symmetric distribution where the probability of occurrence of each point is same for a given range of values in the dataset.

   (A) Statement I is True.
   (B) Statement II is true
   (C) Both are true
   (D) Both are false

(39) What is multicollinearity

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.

(40) Statement –I :- False Negatives (FN): We predicted no, but they actually do have the disease.

Statement –II:- False Positives (FP): We predicted yes, but they don't actually have the disease.

(A)Statement I is true.

(B)Statement II is true

(C)Both statement true

(D)Both statement False.