STATISTICS WORKSHEET-1

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

1.Bernoulli random variables take (only) the values 1 and 0.

a) True
b) False

Answer: True

2. Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?

a) Central Limit Theorem
b) Central Mean Theorem
c) Centroid Limit Theorem
d) All of the mentioned

Answer: c) Central Limit Theorem

3. Which of the following is incorrect with respect to use of Poisson distribution?

a) Modeling event/time data
b) Modeling bounded count data
c) Modeling contingency tables
d) All of the mentioned

Answer: b) Modeling bounded count data

4. Point out the correct statement.

a) The exponent of a normally distributed random variables follows what is called the log- normal distribution
b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent
c) The square of a standard normal random variable follows what is called chi-squared distribution
d) All of the mentioned

Answer: d) All of the mentioned

5. _____ random variables are used to model rates.
a) Empirical
b) Binomial
c) Poisson
d) All of the mentioned

Answer: c) Poisson.

6. 10. Usually replacing the standard error by its estimated value does change the CLT.

a) True
b) False

Answer: b) False

7. Which of the following testing is concerned with making decisions using data?

a) Probability
b) Hypothesis
c) Causal
d) None of the mentioned

Answer: b) Hypothesis.

8. Normalized data are centered at_____and have units equal to standard deviations of the original data.

a) 0
b) 5
c) 1
d) 10

Answer: a) 0

9. Which of the following statement is incorrect with respect to outliers?

a) Outliers can have varying degrees of influence
b) Outliers can be the result of spurious or real processes
c) Outliers cannot conform to the regression relationship
d) None of the mentioned

Answer: c) Outliers cannot conform to the regression relationship

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What do you understand by the term Normal Distribution?

Answer: Normal Distribution is also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence that data far from the mean.
In graphical form the normal distribution appear as a bell curve.
Height ,birth weight, reading ability or SAT scores are just a few example of such variables .because normal distribution variables are s common.

11. How do you handle missing data? What imputation techniques do you recommend?

Answer: when dealing with missing data, data scientist can use two primary method to solve the error ;imputation or removal of data. the imputation method develop reasonable guesses for missing data. it's most useful when the % of missing data is low. most advanced methodology for performing missing data imputation is multiple imputation. In that imputation we generate missing values from the data set many times.

12. What is A/B testing?

Answer: it is also know as split testing or bucket testing. it is method of comparing two version of Web page or app against each other to determine which one perform better. it is example of statistical hypothesis testing. the statement follows,'' if__, then__, because_".

13. Is mean imputation of missing data acceptable practice?

Answer: The process of replacing null values in a data collection with the data's mean is known as mean imputation.
Mean imputation is typically considered terrible practice since it ignores feature correlation.
Second is , Mean imputation decreases the variance of our data while increasing bias.

14. What is linear regression in statistics?

Answer: In Statistics, Linear regression is a linear approach for modeling the relationship between a scalar response and one or more explanatory variables i.e known as dependent or independent variables.
Ex---if an individual are 70 inches tall, we would predict his weight to be: weight = 80+2 x(70) = 220 lbs.

15. What are the various branches of statistics?

Answer: It has 2 main branches:

1) Descriptive statistics---it deals with methods of collection of data , its presentation and organization in various form,such as distribution table and graph. It shows the data in understandable way.
2) Inferential or predictive ----It is used to make a comparison or predictions about a large group , known as population ,using information gathered about a small part of a population called the sample.