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

Ans: a) 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

Ans: a) 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

Ans: 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

Ans: d) All of the mentioned

5. _____ random variables are used to model rates.

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

Ans: c) Poisson

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

a) True

b) False

Ans: b) False

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

a) Probability

b) Hypothesis

c) Causal

d) None of the mentioned

Ans: b) Hypothesis

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

a) 0

b) 5

c) 1

d) 10

Ans: 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

Ans: c) Outliers cannot conform to the regression relationship

WORKSHEET

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

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

Ans : A normal distribution is the proper term for a probability bell curve. In a normal distribution the mean is zero and the standard deviation is 1. It has zero skew and a kurtosis of 3. Normal distributions are symmetrical, but not all symmetrical distributions are normal. In reality, most pricing distributions are not perfectly normal.

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

Ans: There are several techniques for handling missing data. Some of the most common imputation techniques include:

1. Mean imputation: Replace missing values with the mean of the non-missing values.

2. Median imputation: Replace missing values with the median of the non-missing values.

3. Mode imputation: Replace missing values with the mode of the non-missing values.

4. Class-based imputation: Replace missing values with the mean, median, or mode of the non-missing values in the same class.

5. Model-based imputation: Use a model to predict missing values based on other features.

12. What is A/B testing?

Ans: A/B testing (also known as bucket testing, split-run testing, or split testing) is a user experience research methodology. A/B tests consist of a randomized experiment that usually involves two variants (A and B), although the concept can be also extended to multiple variants of the same variable. It includes application of statistical hypothesis testing or "two-sample hypothesis testing" as used in the field of statistics1.

In simpler terms, A/B testing is a way to compare multiple versions of a single webpage or app against each other to determine which one performs better. It is a method to measure the effectiveness of different versions of a webpage or app in terms of user engagement and conversion rates.

13. Is mean imputation of missing data acceptable practice?

Ans: It is acceptable when the missing value proportion is not large enough.

But, when the missing values are large enough and you impute them with the

mean, the standard errors will be lesser than what they actually would have been.

Small standard errors can lead to small p-values and this can create problems for us,

because some variables will start appearing significant, which are ideally not significant.

14. What is linear regression in statistics?

Ans: Linear regression is a statistical method that is used to establish a relationship between two variables. It is used to predict the value of one variable based on the value of another variable. In simple linear regression, there are two variables: the independent variable (x) and the dependent variable (y). The formula for simple linear regression is:

y = B0 + B1x

where y is the predicted value of the dependent variable (y) for any given value of the independent variable (x). B0 is the intercept, which is the predicted value of y when x is 0. B1 is the regression coefficient, which tells us how much we expect y to change as x changes.

15. What are the various branches of statistics?

Ans: There are two main branches of statistics: Descriptive statistics and Inferential statistics. Descriptive statistics deals with the collection, analysis, interpretation, and presentation of data. It is used to describe and summarize data using measures such as mean, median, mode, variance, standard deviation, etc. Inferential statistics deals with making predictions or inferences about a population based on a sample of data. It is used to test hypotheses and make predictions about the population using statistical models1.

Other branches of statistics include Biostatistics, Business statistics, Econometrics, Engineering statistics, Social statistics, etc. These branches are applied in different fields to solve problems and make decisions1.