STATISTICS WORKSHEET-1

Submitted by DARSHIK A S

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

[✓ a) True]

b) False

2) **Which of the following theorems 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

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

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]

5) **______ random variables are used to model rates.**

a) Empirical

b) Binomial

[✓ c) Poisson]

d) All of the mentioned

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

a) True

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

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

9) Which of the following statements 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

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

Normal distribution or Gaussian distribution is a continuous probability distribution that is symmetrical on both sides of the mean and its mean, mode and median are all equal. This type of distribution forms a bell curve. The curve represents the probability and the area under the curve is one. The shape of the curve is determined by the mean and the standard deviation.

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

Missing data can be handled through imputation and deletion. We can delete observations with missing values or drop features with a large number of missing values. However, deleting too much data results in loss of information so it may not be the most effective method. So instead of deletion, imputing missing values could yield more reliable results. Some of the methods that can be used for imputation are:

1. Mean or Median or Mode Imputation: It is a commonly used technique where the missing data is imputed with mean or median or mode of the non-missing observations. However, for a large number of missing values this type of imputation can result in loss of variation.
2. Regression imputation: If there is a high correlation between the missing variable and other variables. We can use a simple regression model to predict the missing values using the highly correlated variables.

12) What is A/B testing?

In A/B testing, a hypothesis is made about the relationship between two sets of data and those data sets are compared against each other to identify if there is a statistical significant relationship or not.

13) Is mean imputation of missing data acceptable practice?

Mean imputation enables us to keep the full sample size. However, it is a bad practice in general. It distorts the relationship between variables and leads to underestimate of standard errors.

14) What is linear regression in statistics?

Linear regression is a linear approach to determine the strength and relationship between one dependent variable and a series of other independent variables.

15) What are the various branches of statistics?

1. Descriptive Statistic: In this statistic the data is summarised through the given observations.
2. Inferential Statistic: This type of statistic allows us to use information collected from a sample to make decisions or inferences about a population.