

## **STATISTICS WORKSHEET 1**

**1.** Bernoulli Theorem variables take the values 1 and 0.

Ans- a) True

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

Ans- a) Central Limit Theorem

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

Ans- b) Modeling Bounded count data

**4.** Point out the correct statement.

Ans- d) All of the above

**5.** \_\_\_\_\_ random variables are used to model rates.

Ans- c) Poisson

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

Ans- b) False

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

Ans- b) Hypothesis

**8.** Normalized data are centered at \_\_\_\_\_ and have units equal to standard deviations of the original data.

Ans- a) 0

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

Ans- c) Outliers can conform to the regression relationship

**10.** What do you understand by the term normal distribution?

Ans- Normal distribution is also called Gaussian distribution or normal curve.

In this case most of values in the data set fall in the middle of the range and the rest are deviating off symmetrically towards both extreme sides. The ideal condition for the normal distribution is when the mean of values are zero and standard deviation is +1 or -1.

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

Ans- In case the number of missing values or data is less,we can drop the row.Otherwise the nan values can be replaced with the mean,median or mode according to the size of the data set and type of data(categorical and numerical.According to my recommendation Mean imputation technique is common if the datas are missing at random or else Logistic Regression or KNN imputation is helpful.

**12.**What is A/B testing?

Ans- A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.It is a hypothetical testing methodology for making decisions that estimate population parameters based on sample statistics.

**13.** Is mean imputation of missing data acceptable practice?

Ans- If the datas are missing at random Mean imputation is acceptable as it preserves the mean of observed data.Otherwise Mean Imputation is considered unacceptable as it decreases the variance while increasing the bias. Mean imputation ignores feature correlation.

**14.** What is linear regression in statistics?

Ans- Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x).The simplest form of the regression equation with one dependent and one independent variable is defined by the formula  $y = c + b \cdot x$ , where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.

**15.** What are the various branches of statistics?

Ans- Following are the two types branches of statistics:

- 1) Descriptive statistics deals with the presentation and collection of data.  
This is usually the first part of a statistical analysis.

- 2) Inferential statistics, as the name suggests, involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics.