# 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) 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) Central limit theorem

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

a) Modelling event/time data b) Modelling bounded count data

c) Modelling contingency tables d) All of the mentioned

Ans) Modelling 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

5. ______ random variables are used to model rates.

a) Empirical b) Binomial

c) Poisson d) All of the mentioned

Ans) Poisson Random Variable

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

a) True b) False

Ans) 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

Ans) 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

Ans) 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

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

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

Normal distribution, 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 than data far from the mean. In graph form, normal distribution will appear as a bell curve.

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

Usual technique for missing data handling is data deletion-two common ones include Listwise Deletion and Pairwise Deletion. Listwise Deletion means deleting any participants or data entries with missing values. This method is particularly advantageous to samples where there is a large volume of data because values can be deleted without significantly distorting readings. Pairwise deletion is the process of eliminating information when a particular data point, vital for testing, is missing. Pairwise deletion saves more data compared to likewise deletion because the former only deletes entries where variables were necessary for testing, while the latter deletes entire entries if any data is missing, regardless of its importance.

General data imputation techniques to handle missing data: Average imputation and common-point imputation. Average imputation uses the average value of the responses from other data entries to fill out missing values. Common-point imputation, on the other hand utilise the middle point or the most commonly chosen value. For example, on a five-point scale, the substitute value will be 3.

12. What is A/B testing?

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.

13. Is mean imputation of missing data acceptable practice?

Imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased. That's a good thing. But any statistic that uses the imputed data will have a standard error that's too low. Ultimately, because your standard errors are too low, so are your p-values. Now you're making Type I errors without realizing it.That's not good. Anyway imputation of missing data is acceptable.

14. What is linear regression in statistics?

In statistics, linear regression is an approach for modelling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variable) denoted X. The case of one explanatory variable is called simple linear regression.

15. What are the various branches of statistics

The two main branches of statistics are descriptive statistics and inferential statistics. Both of these are employed in scientific analysis.

## Descriptive Statistics

Descriptive statistics deals with the presentation and collection of data. This is usually the first part of a statistical analysis. It is usually not as simple as it sounds, and the statistician needs to be aware of designing experiments, choosing the right focus group and avoid biases that are so easy to creep into the experiment.

## Inferential Statistics

Inferential statistics, as the name suggests, involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics. In the end, it is the inferences that make studies important and this aspect is dealt with in inferential statistics.