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.
   1. True
   2. False

**ANSWER: a) True**

1. 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?
   1. Central Limit Theorem
   2. Central Mean Theorem
   3. Centroid Limit Theorem
   4. All of the mentioned

**ANSWER: a) Central Limit Theorem**

1. Which of the following is incorrect with respect to use of Poisson distribution?
   1. Modeling event/time data
   2. Modeling bounded count data
   3. Modeling contingency tables
   4. All of the mentioned

**ANSWER: b) Modeling bounded count data**

1. Point out the correct statement.
   1. The exponent of a normally distributed random variables follows what is called the log- normal distribution
   2. Sums of normally distributed random variables are again normally distributed even if the variables are dependent
   3. The square of a standard normal random variable follows what is called chi-squared distribution
   4. All of the mentioned

**ANSWER: d) All of the mentioned**

1. random variables are used to model rates.
   1. Empirical
   2. Binomial
   3. Poisson
   4. All of the mentioned

**ANSWER: c) Poisson**

1. 10. Usually replacing the standard error by its estimated value does change the CLT.
   1. True
   2. False

**ANSWER: b) False**

1. 1. Which of the following testing is concerned with making decisions using data?
   1. Probability
   2. Hypothesis
   3. Causal
   4. None of the mentioned

**ANSWER: b) Hypothesis**

1. 4. Normalized data are centered at and have units equal to standard deviations of the original data.
   1. 0
   2. 5
   3. 1
   4. 10

**ANSWER: a) 0**

1. Which of the following statement is incorrect with respect to outliers?
   1. Outliers can have varying degrees of influence
   2. Outliers can be the result of spurious or real processes
   3. Outliers cannot conform to the regression relationship
   4. None of the mentioned

**ANSWER: d) None of the mentioned**

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

1. **What do you understand by the term Normal Distribution?**

**ANS:**

Basically it is a probability distribution in which it is symmetric about the mean.

In this distribution it’s show data closest to the mean which are more frequently occurs than the data which are far from the mean.

If we will look in graph from of it, the output will be look like bell curve.

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

**ANS:**

Random forest is a non-parametric imputation method applicable to various variable types that works well with both data missing at random and not missing at random. Random forest uses multiple [decision trees](https://blogs.oracle.com/ai-and-datascience/post/an-introduction-to-building-a-classification-model-using-random-forests-in-python) to estimate missing values and outputs (out of bag) imputation error estimates.

1. **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.

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

**ANS:**

Mean imputation is typically considered terrible practice since it ignores feature correlation.

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

**ANS:**

Linear Regression establishes a relationship between dependent variable (Y) and one or more independent variables (X) using a best fit straight line

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

**ANS:**

The two main branches of statistics are descriptive statistics and inferential statistics. Both of these are employed in scientific analysis of data and both are equally important for the student of statistics.

