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)

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)

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)

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 : c)

6. 10. Usually replacing the standard error by its estimated value does change the CLT.
   a) True
   b) False

   ANS : a)

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)

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)

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.

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

Ans 1.) Normal distribution, also call Gaussian distribution, the most common distribution for independent, randomly generated variables. This distribution , with its symmetrical 'bell shaped' density curve is of fundamental importance in both statistical model and practice. Its roles include the following:

a.) It is good model for the distribution of measurements that occur in practice in a wide variety of different situations.

b.) It is a "building block" for many other distributions.

c.) Much of large sample statistical inference is based on it, and some procedure requires an assumption that a variable is normally distributed.

This distribution has two parameters, which can conveniently expressed directly as the mean and the standard deviation of the distribution. This distribution is a probability distribution that is symmetrical about mean

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

ANS 2.)

- Use regression analysis to symmetrically eliminate data.
- Use deletion method to eliminate missing data.
- Using algorithm that support missing values.
- Mean or Median imputation.
- Random Forest.

3. What is A/B testing?

ANS 3.)

A/B testing in its simplest sense is an experiment on two variants to see which performs better based on a given metric. It 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.

4. Is mean imputation of missing data acceptable practice?

ANS 4.)

True, 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.

5. What is linear regression in statistics?

ANS 5.)

Simple linear regression uses one independent variable to explain or predict the outcome of the dependent variable, while multiple linear regression uses two or more independent variables to predict the outcome. Regression can help finance and investment professionals as well as professionals in other businesses.

6. What are the various branches of statistics?

ANS 6.)

There are three real branches of statistics: data collection, descriptive statistics and inferential statistics.