STATISTICS WORKSHEET-1, Submitted by Faheem Ahmad

# 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

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

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

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

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

Answer: c

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

Answer: b

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

1. 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

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: 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?

Answer:  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.

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

Answer: To handle missing data there are different methods based on the data we are dealing with, which are as follows:-

For categorical data, we can use MODE.

For numerical data, we can use MEAN or MEDIAN.

Likewise, KNN imputation- k neighbors are chosen based on some distance measure and their average is used as an imputation estimate. The method requires the selection of the number of nearest neighbors, and a distance metric. KNN can predict both discrete attributes (the most frequent value among the k nearest neighbors) and continuous attributes (the mean among the k nearest neighbors).

There are various modules in Python from where we can do the imputation such as Simple Imputer, KNN imputer and Iterative imputer.

1. What is A/B testing?

Answer: A/B testing is a user experience research methodology. A/B tests consist of a randomized experiment with two variants, A and B. It includes application of statistical hypothesis testing or "two-sample hypothesis testing" as used in the field of statistics.

1. Is mean imputation of missing data acceptable practice?

Answer: No, not always. It depends on the use-case.

1. What is linear regression in statistics?

Answer: Linear regression consists of finding the best-fitting straight line through the points. The best-fitting line is called a regression line.

1. What are the various branches of statistics?

Answer: The main branches of statistics are descriptive statistics and inferential statistics.

