

## 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
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
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
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
5. \_\_\_\_\_ random variables are used to model rates.
  - a) Empirical
  - b) Binomial
  - c) Poisson
  - d) All of the mentioned
6. 10. Usually replacing the standard error by its estimated value does change the CLT.
  - a) True
  - b) False
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
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
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

**Q10 and Q15 are subjective answer type questions, Answer them in your own words briefly.**

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

In simple words, we can say that when data lies like bell curve on graph it is Normal distribution. In normal distribution, Mean, Median and Mode of data are same, most of data are near the Mean and Normal distribution curve is symmetric where half size data are above mean and half size data are below mean. If we talk about standard deviation, 65 to 70% data are lying in standard deviation range of  $\pm 1$  to the mean and cover 99.5% data in range of standard deviation  $\pm 3$ .

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

It is varied by every data. We cannot say that this method is best for every data. There are many type of missing data like categorical, numerical. We need to study it that in real sense what should be in place of missing data. Mostly used method is mean, median and mode method. Apart from it KNN and random forest are also used to handle missing values. If we want to handle missing categorical data then we can go with categorical imputer where we can handle data with arbitrary value of Max time repeated value.

12. What is A/B testing?

It is such an interesting technique which is used to understand impact of new feature. In this technique, data are divided into 2 groups. One group contains original data and another group contains changes or upgradation in it. Focus is to know that which dataset is more useful or demandable. In this technique, we use Hypothesis also. Data are divided into Part A and Part B, so it is called A/B testing.

13. Is mean imputation of missing data acceptable practice?

Yes. In most of cases we can apply this method but we cannot say that it is a compatible method for every data set. It totally depends upon dataset and how much data are missing.

14. What is linear regression in statistics?

It is a technique, which is used when we have continuous data and we want to find relation between two variables. This technique is very useful to make predictive model. I can give example of salary, experience and age. As per past data we can build model using linear regression to predict model where we can input experience and age to find tentative salary.

## ANSWER OF MCQ'S

- ANS-1 ----- (A) True
- ANS-2 ----- (A) Central limit Theorem
- ANS-3 ----- (B) Modeling bounded count data
- ANS-4 ----- (D) All of Above
- ANS-5 ----- (C) poisson
- ANS-6 ----- (B) False
- ANS-7 ----- (B) Hypothesis
- ANS-8 ----- (A) 0
- ANS-9 ----- (C) Outlier cannot confirm to the regression relationship



are the various branches of statistics?

15. What

---