Name: Anita Vinod Khamkar

Batch No.1844

Institude: DataTrained

WORKSHEET 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

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

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

(c) Modeling bounded count data

4. Point out the correct statement.

(d)All the mentioned

5. _____ random variables are used to model rates.

c) Poisson

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

b) False

7. 1. Which of the following testing is concerned with making decisions using data?

b) Hypothesis

8. 4. Normalized data are centered at_____and have units equal to standard deviations of the original data.

a) 0

9. Which of the following statement is incorrect with respect to outliers?

c) Outliers cannot conform to the regression relationship

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

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

Ans: Normal distribution , also know 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?

Ans: 1. Deleting Rows with missing values.

2. Impute missing values for continues variable

3. Impute missing values for categorical variable

4. Other Imputation Methods

5. Using Algorithms that support missing values

6. Prediction of missing values

7. Imputation using Deep Learning Library--Datawig

12. What is A/B testing?

Ans: A/B testing is a basic randomized control experiment. It is a way to compare the two version of a variable to find out which perform better in a controlled envioronment.

13. Is mean imputation of missing data acceptable practice?

Ans: Mean imputation reduces the variance of the imputed variables. Mean imputation shrinks standard errors, which invalidates most hypothesis tests and the calculation of confidence interval. Mean imputation does not prevent relationships between variables such as correlations.

14. What is linear regression in statistics?

Ans: Linear regression is an attempt to model the relationship between two variable by fitting a linear equation to observed data, where one variable is considered to be an explanatory variable and the other as a dependent variable.

15. What are the various branches of statistics?

Ans: Statistics is a study of presentation, analysis, collection, interpretation and organization of data

There are **two main branches** of statistics
- Inferential Statistic.
- Descriptive Statistic.

Inferential Statistics:
Inferential statistics used to make inference and describe about the population. These stats are more useful when its not easy or possible to examine each member of the population.

Descriptive Statistics:
Descriptive statistics are use to get a brief summary of data. You can have the summary of data in numerical or graphical form.