FLIP ROBO WORKSHEET

STATISTICS WORKSHEET-1

Q 1. Bernoulli random variables take (only) the values 1 and 0.
Ans 1. a) True

Q 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?
Ans 2. a) Central Limit Theorem

Q 3. Which of the following is incorrect with respect to use of Poisson distribution?
Ans 3. b) Modeling bounded count data

Q 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 4. d) All of the mentioned.

Q 5. _____ random variables are used to model rates.
Ans 5. c) Poisson.

Q 6. Usually replacing the standard error by its estimated value does change the CLT.
Ans 6. b) False.

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

Ans 7. b) Hypothesis

Q 8. Normalized data are centered at _____ and have units equal to standard deviations of the original data.

Ans 8. a) 0

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

Ans 9. c) Outliers cannot conform to the regression relationship

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

Ans 10. The normal distribution is a continuous probability distribution that is symmetrical around its mean, most of the observations cluster around the central peak, and the probabilities for values further away from the mean taper off equally in both directions.

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

Ans 11. Following methods are used for handling missing data

- Deleting Rows with missing values.
- Impute missing values for continuous variable.

- Impute missing values for categorical variable.
- Other Imputation Methods.
- Using Algorithms that support missing values.
- Prediction of missing values
- Imputation using Deep Learning Library

  Imputation Techniques used to handle missing data.

1. Complete Case Analysis(CCA)
2. Arbitrary Value Imputation
3. Frequent Category Imputation

Q 12. What is A/B testing?

Ans 12. A/B testing is an analytical method for making decisions that estimates parameters based on sample statistics.

Q 13. Is mean imputation of missing data acceptable practice?

Ans 13. Bad practice in general. If just estimating means: mean imputation preserves the mean of the observed data. Leads to an underestimate of the standard deviation.

Q 14. What is linear regression in statistics?

Ans 14. Linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables.

Q 15. What are the various branches of statistics?

Ans 15. There are three real branches of statistics: data collection, descriptive statistics and inferential statistics