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

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

10. What do you understand by the term Normal Distribution?
11. How do you handle missing data? What imputation techniques do you recommend?
12. What is A/B testing?
13. Is mean imputation of missing data acceptable practice?
14. What is linear regression in statistics?
15. What are the various branches of statistics?

10and 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 is also known as Gaussian distribution . It's a probability distribution which is symmetric about the mean , which shows that data near the mean are more frequent in occurrence than data far from the mean. A normal distribution is the proper term for probability bell curve. In a normal distribution mean is 0 and SD is 1. It has 0 skewness and kurtosis of 3.

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

Ans Missing data csn skew anything for data scientists from economic analysis to clinical trials. Missing data reduces the statistical power of the analysis which can distort the validity of results. When dealing with missing data we can use 2 primary methods;

a- Imputation method
b- Removal

There are following ways to handle missing values in dataset-

a-deleting rows with missing values

b-impute missimg values for continuous variable

c- Impute missing values for categorical variable
d- Use of algorithms
e- Prediction of missing values

The recommended techniques are –

A mean substitution

B iterative model based imputation

C Random imputation of missing data

D- multiple imputation of incomplete multivariate data.

12. What is A/B testing?

A/B testing is consists of a randomised experiment with 2 variants A and B. it includes application of statistical hypothesis testing or 2 sample hypothesis testing. A/B testing is a way to compare two versions of single variable , typically by testing a subjects response to variant A against variant B and determing which of the two are more effective.

13. Is mean imputation of missing data acceptable practice?

No not highly as it has 2 problems-

A- Mean imputation does not preserve the relationships among variables.
B- Mean imputation leads to an underestimate of Standard Errors.

14. Whatislinearregressioninstatistics?

In stastics , linear regression is a linear approach to modelling relationship between a scalar response and one or more explanatory variables.

15. What are the various branches of statistics?

There are 2 main branches of statistics –

1- Descriptive Statistics
2- Inferential statistics