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
Answer – 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
b) Central Mean Theorem
c) Centroid Limit Theorem
d) All of the mentioned
Answer – a) Central Limit Theorem
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
Answer -b) Modeling bounded count data
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
Answer-d) All of the mentioned
5. _____ random variables are used to model rates.
a) Empirical
b) Binomial
c) Poisson
d) All of the mentioned
Answer-b)Poission
6. Usually replacing the standard error by its estimated value does change the CLT.
a) True
b) False
Answer – b)False
7. Which of the following testing is concerned with making decisions using data?
a) Probability
b) Hypothesis
c) Causal
d) None of the mentioned
Answer-b)Hypothesis
8.Normalized data are centered at_____and have units equal to standard deviations of the original data.
a) 0
b) 5
c) 1
d) 10
Answer-a)0
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
Answer-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?

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.

In graphical form, the normal distribution appears as a "bell curve".

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

Answer - To tidy up your data, your options usually include accepting, removing, or recreating the missing data. You should consider how to deal with each case of missing data based on your assessment of why the data are missing.
->Are these data missing for random or non-random reasons?
->Are the data missing because they represent zero or null values?
->Was the question or measure poorly designed?
Imputation techniques to recommend :-
->simpleImputer
-> Hot deck imputation
-> Cold deck imputation

12. What is A/B testing?

Answer- A/B testing is a common and powerful marketing technique. Before sending out a marketing message, a marketer would send "test" versions to a portion of audience members to see which performs better.
Using an A/B test gives you an idea of what delights your customers the most. For example, if you're using a promoted tweet campaign to support a new product launch, you might test different versions of that tweet using automated software, so you can tell how to market to your audience.
After seeing which tweet had the most clicks, you can then use *that* version in marketing messages, as well as a template for the duration of the campaign. That way, engagement is more of a guarantee.

13. Is mean imputation of missing data acceptable practice?

Answer- You can fill in missing values with the mean of the variable over the time period of observation. Pros: Easy to compute and understand. Decent option if you know your variables to be distributed normally. Cons: You if your data has a trend your added values may make your charting look odd. Also, this is not acceptable if your variables have an odd distribution that makes the mean value meaningless.

14.What is linear regression in statistics?

Answer- Linear Regression is a statistical technique used to model the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables). When there is one

independent variable (iv), the procedure is known as simple Linear Regression; when there are more iv's, statisticians refer to it as multiple Linear Regression.

15.What are the various branches of statistics?

Answer- There are two main branches of statistics: Descriptive Statistics and Inferential Statistics.

Descriptive Statistics deals with the collection of data, its presentation in various forms, such as tables, graphs and diagrams and finding averages and other measures which would describe the data.

Inferential Statistics is used to interpret data more abstractly.