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
b) False
Answer: A) True

2. Which of the following theorem states that the distribution of averages of id 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: c) Poisson

6. 10. Usually replacing the standard error by its estimated value does change the CLT.
a) True
b) False
Answer: A)True

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
Answer: B) Hypothesis

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
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

WORKSHEET

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 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 bell curve.

11. How do you handle missing data? What imputation techniques do you recommend?
Answer: The concept of missing data is implied in the name: it's data that is not captured for a variable for the observation in question.
When dealing with missing data, data scientists can use two primary methods to solve the error:
Before deciding which approach to employ, data scientists must understand why the data is missing.

1) Missing at Random (MAR)
2) Missing Completely at Random (MCAR)
3) Missing Not at Random (MNAR)

These are examples of single imputation methods:

1) Mean, Median and Mode
2) Time-Series Specific Methods
3) Last Observation Carried Forward (LOCF) and Next Observation Carried Backward (NOCB)
4) Linear Interpolati
5) Seasonal Adjustment with Linear Interpolation.

12. What is A/B testing?
Answer: A/B testing is basically statistical hypothesis testing, or, in other words, statistical inference. It is an analytical method for making decision that estimates population parameters based on sample statistics.

13. Is mean imputation of missing data acceptable practice?
Answer: True, imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased.

14. What is linear regression in statistics?
Answer: Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable. A regression model that estimates the relationship between one independent variable and one dependent variable using a straight line.

15. What are the various branches of statistics?
Answer: There are three real branches of statistics:

1) Data Collection
2) Descriptive Statistics
3) Inferential Statistics