STATISTICS WOORKSHEET

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

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

Ans= A 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 that data far from the mean.in graph form normal distribution will apper as a bell curve.

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

Ans= use deletion method to eliminate missing data.also we use regression analysis to eliminate data.for imputation complete case analysis = we directly removes the rows that have missing data .

Arbitratry value imputation.

Frequent category imputation.

12. What is A/B testing?
Ans = A/B testing is basic randomized control experiment. It is a way to compare the two version of a variable to find out which performs better in a controlled environment.

For example:- lets say you own a company and want to increase the sales of your product. Here, either you can use random experiment or you can apply scientific and statistical method.

13. Is mean imputation of missing data acceptable practice?

Ans = The process of replacing null values in a data collection with the data's mean is known as mean imputation.

mean imputation is typically considered terrible practice since it ignores feature correlation. consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score. if we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.

second, mean imputation decreases the variance of our data while increasing bias. as a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

14. What is linear regression in statistics?

Ans = Linear regression attemps to model the relationship between two variables by fitting a linear equation to observed data.a linear regression line has an equation of the form y=a+bx where x is the explanatory variables and y is the dependent variables.

15. What are the various branches of statistics?
Ans = There are three real branches of statistics

    a.   Data collection

STATISTICS WOORKSHEET

b. Descriptive statistics
c. Inferential statistics