FLIP ROBO

STATISTICS WORKSHEET-1

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. c) Poisson

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

Answer 10-> The most important and commonly encountered probability distribution is known as Normal Distribution. The Normal Distribution is considered the “bell curve”, as many distributions are bell-shaped. It is also the distribution off of which our ideas of standard deviation are based.

Answer 11-> There are 7 ways to handle missing data :-

1. Deleting Rows with missing values
2. Impute missing values for continuous variable
3. Impute missing values for categorical variable
4. Other Imputation Methods
5. Using Algorithms that support missing values
6. Prediction of missing values
7. Imputation using Deep Learning Library — Datawig

Answer 12-> A/B testing is a way to compare two versions of a single variable , typically by testing a subject’s response to variant A against variant B , and determining which of the two variants is more effective.

Answer 13->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 and also 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.

Answer 14-> To predict a dependent variable value (y) based on a given independent variable (x) we perform Linear regression. This regression technique finds out a linear relationship between x (input) and y(output). The relationship between two variables by fitting a linear equation to observed data on which only a variable is considered to be an explanatory variable and the rest is the dependent one.

Answer 15-> The various branches of static are :-

a)Descriptive Analytics -> The interpretation of historical data to better understand changes that have occurred in a business is called Descriptive Analytics.

b)Predictive Analytics ->The advanced analytics that makes predictions about the future outcomes by using historical data combined with statistical modeling, machine learning, and data mining.

c)Prescriptive Analytics-> The data analytics that uses algorithms and analysis of raw data so that the output will be better and effective decisions for a long and also for short span of time .