STATISTICS – WORKSHEET 1 (SATENDER SINGH) DS2302

1)Bernoulli random variables take (only) the values 1 and 0.
a) True b) False
Answer - a

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

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

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

5) _____ random variables are used to model rates.
a) Empirical b) Binomial c) Poisson d) All of the mentioned
Answer - c

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

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

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

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

10) What do you understand by the term Normal Distribution?
Answer: A normal distribution is the proper term for a probability bell curve. In a normal distribution, the mean is zero & the standard deviation is continuous data which follow bell like shape distribution. If there is skewed data then it should be converted into normal distribution or bell shaped distribution.

11) How do you handle missing data? What imputation techniques do you recommend?
Answer: We can handle missing data in many ways like: Mean Substitution, Maximum Likelihood, Median Substitution, Zero Imputation, Mode Imputation. I believe the most common reaction is to ignore it. Choosing to make no decision indicates that your statistical programme will make the decision for you. I will recommend Constant imputation method in dealing with missing data. Constant imputation methods impute a constant value in the replacement of missing data in an observation. There are variations of this technique and some ways for data scientists to make this more effective.

12) What is A/B testing?
Answer: A/B testing is an experiment on two variants to see which performs better on a given metrics. Typically, two consumer groups are exposed to two different versions of the same thing to see if there is a significant difference in metrics like sessions, click-through rate or conversions.

13) Is mean imputation of missing data acceptable practice?
Answer: The process of replacing null values in a data collection with the data's mean is known as mean imputation. Mean imputation is considered terrible practice since it ignores feature correlation. 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?
Answer: Linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables or dependent and independent variables. The case of one explanatory variable is called simple linear and for more than one it is called multiple linear regression.

15) What are the various branches of statistics?
Answer: There are two type of statistics
Descriptive: Which we can describe or know maximum about that and able to explain because we know very much about them. Here data is small eg.- class room
Inferential: Here data is vast, for that we pick some percentage of data and make the conclusion for whole data. We cannot able describe it of our own because here data is very vast we pick some random sample and judge for whole data. We make predictions eg.-stock market