STATISTICS WORKSHEET-1

ANSWER 1 TO 9

Answer1= (A) True

Answer2= (A) Central limit theorem

Answer3= (B) Modeling bounded count data

Answer4= (D) All of the mentioned

Answer5= (C) Poisson

Answer6= (B) False

Answer7= (A) Probability

Answer8= (A) 0

Answer9= (C) Outliers cannot conform to the regression relationship

ANSWER 10 TO 15

Answer 10 = Normal Distribution is a probability distribution that is symmetrical on both sides of the mean , so the right side of the center is the mirror image of the left side. The area under the curve represents probability. It is also called the Bell curve because the graph of its probability density looks like a bell.

Answer 11 = Missing data reduced the statistical power of the analysis , which effects the validity of the result. It appears when no value is available in one or more variable of an individual.

I will recommend Mean or Median imputation technique :-

At random when the data is missing , list-wise or pair-wise can help us in deletion of the missing observations. However, there are multiple reasons why this technique is not the most feasible option:

1: To produce a reliable analysis with a non-missing data this technique may not have that much enough observation.

2: To be a part of the analysis, External factors may require some specific observations.

3: In predictive analytics, missing data can prevent the prediction for those observations which have missing data .

Answer 12 = An A/B test is an example of statistical hypothesis testing, a process in which a hypothesis is made about the relationship between two data sets and those data sets are then compared against each other to determine if there is a statistically significant relationship or not.

Answer 13 = Yes , imputing the mean preserves the mean of the observed data . so at random if the data is completely missing, the mean remain unbiased .

Answer 14 = Linear regression is an attempt to model the relationship between two variables by fitting a linear equation to observe data ,where one variable is considered to be an explanatory variable and the other as a dependent variable.

Answer 15 = The three branches of statistics :-

Data collection – data collection is the process of gathering and measuring information on variables of interest , in an established systematic fashion that enables one to answer stated research question , test hypotheses and evaluate outcomes.

Descriptive statistics – descriptive statistics are used to calculate , describe and summarize data in a logical and efficient way. It reported numerically in the manuscript .

Inferential statistics – inferential statistics allow us to make predictions from a given sample data set. Its aim is to form interpretations and make a broad statement of the population data beyond the immediate data available.