STATISTICS WORKSHEET

I. Answers of (1 -09)

Q1) A

Q2) B

Q3) C

Q4) C

Q5) C

Q6) A

Q7) A

Q8) A

Q9) A

II. Answers of (10 -15)

Q10) **NORMAL DISTRIBUTION:-** The Normal Distribution is an probability distribution which is symmetric about its mean. Graphically It is in the form of " Bell Shape Curve", with two parameters of Mean is' 0' and Standard Deviation is '1'. It generally performs in large samples test i.e Z-test. Here its sample size n>30. Normal distributions are symmetrical about its mean. In Normal Distribution 68% will lies in between 1 Standard Deviation (S.D), 95% will lie in between 2 standard Deviation(S.D), 99.7% will lies between 3 Standard Deviation (S.D).

Q11) **HANDLING MISSING DATA:-** First of all in a given data set will check is there any missing data is there if it is there then will perform some task they are ;

a) **Pair wise detection**: If there is an missing data in given data set then existing values are used in statically testing. Hence parawise observes all information.

b) **Mean substitution:** In this mean value of a variable is used in place of missing data value for that same variable.

c) **List or case deletion**: The most common approach to the missing data omit the missing cases where the data is not present and analyze the remaining data. This approach is known as List or case detection.

Q12) **WHAT IS TESTING A/B:-** The A/B testing is basically known as 'Stastical Hypothesis Testing' or in other words it is also known as 'Stastical Inference'. It is analytical methods for making decisions. We can make decision by taking our assumption true i.e (Hypothesis) testing.

Q13) **IS MEAN IMPUTATION OF MISSING DATA ACCEPTABLE:-** 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. That's a good thing. Since most research studies are interested in the relationship among variables, mean imputation is not a good solution.

Q14) **LINEAR REGRESSION:-** In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables. The case of one explanatory variable is called simple linear regression; for more than one, the process is called multiple linear regression.

Q15) **BRANCHES OF STATSTICS:-** It plays main role in field of research, it collects data from various places and represents in form of graphical representation. It mainly have two branches namely, they are, i) Descriptive Statistics , ii) Inferential Statistics. Descriptive Statistics consists of (Mean, Median, Mode) & Inferential Statistics consists of (Hypothesis testing, CHI-Square testing, F-test, T-test).