STATISTICS WORKSHEET1

Q1) A

Q2) A

Q3) B

Q4) D

Q5) C

Q6) B

Q7) B

Q8) A

Q9) C

Q10) What do you understand by the term Normal Distribution?

ANSWER: Normal Distribution is a probability that is symmetric about the mean, which shows the data near to the mean. All normal distribution looks like a symmetric, bell-shaped curve. The graph of the normal distribution depends on two factors, I.e mean and standard deviation.

The mean of the distribution determines the location of the center of the graph and standard deviation determines the height and width of the graph. When the standard deviation is small, the curve is tall and narrow and when the standard deviation is big, the curve is short and wide.

Q11) How do you handle missing data? What imputation techniques do you recommend?

ANSWER: Missing data can be handelled by using deletion methods to eliminate missing data. This method works only when datasets have missing fields ot it can be handelled by regression analysis to systematically eliminate data.

There are many imputation techniques used,but the best technique is

1) By ignoring the records with missing values.

2) By substituting the value such as mean.

3) By predicting missing values.

Q12) What is A/B testing?

ANSWER: A/B testing is a way to compare two versions of single variable, typically by testing a subjects response to variant A against variant B, and determining which of the variants is more effective.

For Example: A/B tests are being used for conducting complex experiments on subjects such as network effects when users are offline, how online services affect user actions, and how users influence one another.

Q13) Is mean imputation of missing data acceptable practice?

ANSWER: Yes, Mean imputation of missing data is acceptable as imputing the mean secures the mean of the observed data. So, if the datas are missing completely at random, the estimate of the mean remains unbiased.

Q14) What is linear regression in statistics?

ANSWER: 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 and for more than one, the process is called multiple linear regression. Linear regression is the most eidely used statistical technique. The result of the linear regression can be used to make prediction about data.

Q)15. What are the various branches of statistics?

ANSWER: There are two branches of statistics:

1)Descriptive Statistics: Descriptive Statistics deals with the presentation and collection of the data.

2)Inferential Statistics: Inferential Statistics involves drawing the right conclusion from the statistical analysis that has been performed using descriptive statistics.

Both of these are used in scientific analysis of data and both are equally important.