Statistics Worksheet-1

Q1. a)
Q2. a)
Q3. b)
Q4. d)
Q5. c)
Q6. b)
Q7. b)
Q8. a)
Q9. c)

Q10. Normal distribution is a probability function tha describes how the values of a variable are distributed. It arranges the data set in such a manner that most values cluster in the middle of the range and the rest of it tapers off symmetrically towards either extreme.

Q11. When the percentage of missing values to the record is small we can ignore them. However, if the percentage is high the most commonly used way is to substitute a value (eg. mean, median,etc.)
Deletion method can be used to eliminate the missing data. It can only work for datasets where participants have missing fields. Regression method can be used to predict null values if the rest of the values are well connected.

Q12. A/B testing is a method of comparing two different versions of a webpage or an app against each other to find out which one performs better. A/B testing is essentially an experiment where two or more variants of a page are shown to users at random, and statistical analysis is used to determine which variation performs better for a given conversion goal.

Q13. Mean imputation of a missing data is not considered as a good practice in general. It preserves the mean of the observed data which leads to underestimate the standard deviation.

Q14. Linear regression is a commonly used type of predictive analysis which examines if a set of predictor variables do a good job in predicting an outcome (dependent) variable; and which variables are particular are significant predictors of the outcome variable and in what way do they impact the outcome variable.
The simplest form of the regression equation with one dependent and one independent variable is defined by the formula y = c + b*x, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.

Q15. There are two branches of statistics:

1. Descriptive statistics: It involves collection and presentation of data. Different areas of study require different kinds of analysis using descriptive statistics.
2. Inferential statistics: it involves drawing the right inferences (conclusions) from the statistical analysis that has been performed using the right descriptive statistics.