Statistics Worksheet -1

1. A
2. A
3. B
4. D
5. C
6. B
7. B
8. A
9. 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 and the standard deviation is 1. It has zero skew and a kurtosis of 3. Normal distributions are symmetrical, but not all symmetrical distributions are normal.In reality, most pricing distributions are not perfectly normal.

11. How do you handle missing data? What imputation techniques do you recommend?

ANSWER: Types of missing Data:

Missing Completely At Random (MCAR): When missing values are randomly distributed across all observations, then we consider the

data to be missing completely at random.

Missing At Random (MAR): The key difference between MCAR and MAR is that under MAR the data is not missing randomly across all observations, but is missing randomly only within sub-samples of data.

Not Missing At Random (NMAR): When the missing data has a structure to it, we cannot treat it as missing at random.

Imputation techniques:

Mean or Median Imputation

Multivariate Imputation by Chained Equations

Random Forest

12. What is A/B testing?

ANSWER: A/B testing (also known as split testing or bucket testing) is a method of comparing two versions of a webpage or app against each other to determine 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.

13. Is mean imputation of missing data acceptable practice?

ANSWER: Its a non standard, it uses Random Forest. It is used to predict the missing data. It also can be used for both i.e continous as well as categorical data and so it makes advantageous over other imputations.

14. What is linear regression in statistics?

ANSWER: Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things: 1) does a set of predictor variables do a good job in predicting an outcome (dependent) variable. 2) Which variables in particular are significant predictors of the outcome variable, and in what way do they–indicated by the magnitude and sign of the beta estimates–impact the outcome variable? These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. 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.

15. What are the various branches of statistics?

ANSWER: Descriptive Statistics

Descriptive statistics deals with the presentation and collection of data. This is usually the first part of a statistical analysis. It is usually not as simple as it sounds, and the statistician needs to be aware of designing experiments, choosing the right focus group and avoid biases that are so easy to creep into the experiment.

Inferential Statistics

Inferential statistics, as the name suggests, involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics. In the end, it is the inferences that make studies important and this aspect is dealt with in inferential statistics.