STATISTICS WORKSHEET-1 (Answers)

Name: Prashant Pathak (Internship 33) Worksheet_set_1(10 to 15) [FLIP ROBO]

Question 10. What do you understand by term the Normal Distribution?

Answer 10. The normal distribution is the most widely known and use of all distributions. Because the normal distribution approximates many natural phenomena so well, it has developed into a standard of reference for many probability problems.

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

Answer 11. Missing data can be dealt with in a variety of ways. I believe the most common reaction is to ignore it. Choosing to make no decision, on the other hand, indicates that your statistical program will make the decision for you. Another common strategy among those who pay attention is imputation and I recommend mean imputation for handling missing data.

Question 12. What is A/B testing?

Answer 12. A/B testing (also known as bucket testing or split-run testing) is a user experience research methodology. A/B tests consist of a randomized experiment with two variants, A and B. It includes application of statistical hypothesis testing or "two-sample hypothesis testing" as used in the field of statistics.

Question 13. Is mean imputation of missing data acceptable practice?

Answer13. Mean imputation is typically considered terrible practice since it ignores feature correlation. Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.

Second, mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

Question 14. What is linear regression in statistics?

Answer 14. In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables.

Question 15. What are the various branches of statistics?

Answer 15. There are three real branches of statistics: data collection, descriptive statistics and inferential statistics.