Ashutosh Mishra
Thursday, July 21, 2022

STATISTICS WORKSHEET-1 (ANSWERS)

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?

Normal distribution is basically a perfect condition where the Mean=Median=Mode=0 and standard deviation = (+1 & -1) and the confidence level would be 95%.

For instance – The ratings for a movie performance would definitely be differ with respect to each spectators. i.e. It's not possible that everybody will give the rating 5 or so, but normal distribution is a perfect situation ergo in this case everybody would be giving the same ratings, which is nearly impossible in real life.

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

See, if missing values are present in a data frame then we can remove it or replace it. Like if we have a very large data set to operate let's suppose we have roughly 500 rows and out of that 30 rows are showing null values then we may delete it as this would not affect the data at all as 470 rows are enough to do EDA and also we can replace it if the rows are less.

I will recommend to go for Simple Imputer.

12. What is A/B testing?

An AB test is an example of statistical hypothesis testing, a process whereby a hypothesis is made about the relationship between two data sets and those data sets are then compared against each other to determine if there is a statistically significant relationship or not.

13. Is mean imputation of missing data acceptable practice?

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.

14. What is linear regression in statistics?

Linear Regression is basically a best fitted straight line which contains most points on to it. In other words, linear regression is a model which shows relationship b/w the explanatory (independent) variables and target (dependent) variable.

15. What are the various branches of statistics?