

STATISTICS WORK SHEET

1. A
 2. A
 3. B
 4. D
 5. C
 6. B
 7. B
 8. A
 9. C
10. A normal distribution is a symmetrical representation of any data plotted in a form of a bellshape. The distribution is usually symmetrical on both sides of the mean, with the right side of the mean being a mirror image of the left side.
11. Missing data could be handled in three (3) main ways.
- a. Mean or Median Imputation: - This method uses the mean or the median of the non-missing observation. This can be useful in cases where the number of missing observations is low
 - b. Multivariate Imputation by Chained Equation (MICE): - MICE assumes that the missing data are Missing at Random (MAR). It imputes data on a variable-by-variable basis by specifying an imputation model per variable
 - c. Random Forest: - Random Forest is a non-parametric imputation method applicable to various variable types that works well with both data missing at random and not missing at random. Random forest uses multiple decision trees to estimate missing values and outputs OOB (out of bag) imputation error estimates.
- I will recommend the Random Forest imputation method because of its ability to be used for both data missing at random or not missing at random.
12. A/B testing is a way to compare two versions of a single variable, typically by testing a subject's response to variant A against variant B, and determining which of the two variants is more effective
13. The variance of a mean-imputed variable is always biased downward from the variance of the un-imputed variable. This bias affects standard errors, confidence intervals, and other inferential statistics.
14. Linear regression strives to show the relationship between two variables by applying a linear equation to observed data.
15. The major branches of statistics are Descriptive and Inferential Statistics