STATISTICS WORKSHEET

1. A
2. A
3. B
4. D
5. C
6. B
7. B
8. A
9. A

10. Normal distribution is used to analyse data when there is an equally likely chance of being above or below the mean for continuous data whose histogram fits a bell curve. It is the most widely known and used of all distributions.
11. We handle missing data by using the imputation method. I would recommend the multiple imputation technique.
12. A/B testing, also known as split testing or bucket testing, is a methodology for comparing two versions of a web page or application against each other to determine which one performs better.
13. The mean imputation of missing data is not an acceptable practice because it ignores feature correlations and is not suitable for large amounts of data.When there are many missing variables, mean or median results can result in a loss of variation in the data.
14. Linear regression is a regression model that estimates the relationship between one independent variable and one dependent variable using a straight line. It is used to predict the value of a variable based on the value of another variable. The variable I want to predict is called the dependent variable while the variable I am using to predict the other variable's value is called the independent variable. Linear regression is used when the value of y is in decimals.

    It is represented as: $y = a + bx + e$
    y is the dependent value
    a is intercept
    b is coefficient
    x is independent
    e is error
15. The various branches of statistics include
- descriptive statistics
- Inferential statistics