Statistics

1) A) True
2) A) Central limit theorem
3) A) modelling event/ time data
4) D) All of the mentioned
5) C) Poisson
6) B) False
7) B) Hypothesis
8) A) 0
9) C) Outliers cannot conform to the regression relationship

Worksheet 2

10) A normal distribution is one in which the values are evenly distributed both above and below the mean. A population has a precisely normal distribution if the mean, mode, and median are all equal. For the population of 3,4,6,5,6,6,7, the mean, mode, and median are all 6.

11) There were three methods to handle missing data, one is Mean or Median Imputation. When data is missing at random, we can use list-wise or pair-wise deletion of the missing observations. Multivariate Imputation by Chained Equations (MICE) assumes that the missing data are Missing at Random & Random Forest.
    Also few methods are deletion and regression analysis to systematically remove data. Data imputation technique perform better.

12) A/B testing is a method of comparing two versions of a webpage or app against each other to determine which one performs better. It enables data-driven decisions and ensuring positive impacts.

13) Replacing null values in a data collection with the data's mean is known as mean imputation. compare different statistical methods, the number of imputations should be even larger than the percentage of missing observations, usually between 100 and 1000, in order to control the Monte Carlo error.

14) Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values.

15) It is divided into two major divisions: descriptive and inferential. Descriptive statistics describe what is going on in a population or data set. Inferential statistics allows to take findings from a sample group and generalize them to a larger population.