

# STATISTICS WORKSHEET\_1

1. A

2. A

3. D

4. D

5. C

6. A

7. B

8. A

9. C

10. In a normal distribution, data is symmetrically distributed with no [skew](#).

When plotted on a graph, the data follows a bell shape.

11. The common ways of handling missing values are Deletions and Imputations. KNN Imputation I will recommend for missing values.

12. [A/B testing](#), also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics.

13. It is acceptable when the missing value proportion is not large enough. But, when the missing values are large enough and you impute them with the mean, the standard errors will be lesser than what they actually would have been.

14. Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things: (1) does a set of predictor variables do a good job in predicting an outcome (dependent) variable? (2) Which variables in particular are significant predictors of the outcome variable, and in what way do they—indicated by the magnitude and sign of the beta estimates—impact the outcome

variable? These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. The simplest form of the regression equation with one dependent and one independent variable is defined by the formula  $y = c + b \cdot x$ , where  $y$  = estimated dependent variable score,  $c$  = constant,  $b$  = regression coefficient, and  $x$  = score on the independent variable.

15. The two main branches of statistics are [descriptive statistics](#) and [inferential statistics](#).