# Statistics worksheet 1

1. a) True

2. a) Central Limit Theorem

3. b) Modeling bounded count 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

10. Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are frequent in occurance than data far from the mean.

11. Following are the methods to handle the missing values:

- Delete the missing record value
- Create a separate model to handle the missing value
- Through Statistical methods (mean, median, mode)

12. A/B 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 drive business metrics. It eliminates all the guesswork out of website optimization and enables experience optimizers to make data-backed decisions. In A/B testing, A refers to "control" or the original testing variable. Whereas B refers to 'variation' or a new version of the original testing variable.

14. Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. For example, a modeler might want to relate the weights of individuals to their heights using a linear regression model. A linear regression line has an equation of the form $Y = a + bX$, where X is the explanatory variable and Y is the dependent variable. The slope of the line is b, and a is the intercept (the value of y when $x = 0$).