STATISTICS WORKSHEET ANSWER SHEET

Answer 1 - a) True

Answer 2 - a) Central Limit Theorem

Answer 3 - b) Modeling bounded count data

Answer 4 - d) All of the mentioned

Answer 5 - c) Poisson

Answer 6 - b) False

Answer 7 - b) Hypothesis

Answer 8 - a) 0

Answer 9 - c) Outliers cannot conform to the regression relationship

Answer 10 - A normal distribution is a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest taper off symmetrically toward either extreme. The middle of the range is also known as the mean of the distribution.

Answer 11 - Missing data can be dealt with in a variety of ways.

1. Another common strategy among those who pay attention is imputation.
2. Mean imputation.
3. Substitution.
4. Hot deck imputation.
5. Cold deck imputation.
6. Regression imputation.

Answer 12 - A/B testing is a type of experiment in which you split your web traffic or user base into two groups, and show two different versions of a web page, app, email, and so on, with the goal of comparing the results to find the more successful version.

Answer 13 –

Mean imputation (MI) is one such method in which the mean of the observed values for each variable is computed and the missing values for that variable are imputed by this mean. This method can lead into severely biased estimates even if data are MCAR

Answer 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*x, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.

Answer 15 – There are three real branches of statistics: data collection, descriptive statistics and inferential statistics