Statistics Worksheet:

Question 1 to 9:

1. A) True
2. A) Central Limit Theorem
3. C) Modeling Bounded Count Data
4. D) All the mentioned
5. C) Poisson
6. B) False
7. B) Hypothesis
8. A) 0
9. C) Outliers cannot conform to the regression relationship

Question 10 to 15:

10. The normal distribution is a probability distribution that is symmetric about the mean, indicating that data near the mean occur more frequently than data that is far from the mean. In graphical form, the normal distribution will be displayed as a bell-shaped curve. It is a symmetrical distribution in which most observations are clustered around the central peak, and the probability of values far from the average decreases equally in both directions.

11. Missing data could skew result of analysis which in return will lead to produce biased estimate and invalid result. However, it can be dealt by using two primary methods, namely imputation and removal of data.

    There are majorly two types of methods to delete data while handling the missing data namely, listwise and dropping variables. Also, pairwise is also been used to delete data.

    When some data is missing in a data set it make sense to delete data. Instead of deletion imputation of value of missing data is also a solution. It can deliver reasonably reliable results. There are many imputation methods such as; Mean, Median, & Mode, Time series specific methods, Last Observation Carried Forward (LOCF) & Next Observation Carried Backward (NOCB), Linear Interpolation, Seasonal Adjustment with Linear Interpolation. Apart from this multiple imputation & KNN is also used to treat missing data value.

    I would recommend multiple imputation to handle missing data values.

12. An AB test is statistical hypothesis testing, a process whereby a hypothesis is made about the relationship between two data sets and those data sets are then compared against each other to determine if there is a statistically significant relationship or not.

13. Mean imputation does not preserve the relationships among variables and leads to an underestimate of standard errors. So believe it is not acceptable practice.

14. Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things: does a set of predictor variables do a good job in predicting an outcome dependent variable? 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.

15. The two main branches of statistics are descriptive statistics and inferential statistics. Both of these are employed in scientific analysis of data.