STATISTICS ANSWERS

1. b) False
2. a) Central Limit Theorem
3. b) Modeling bounded count data
4. c) The square of a standard normal random variable follows what is called chi-squared distribution
5. c) Poisson
6. b) False
7. b) Hypothesis
8. a) 0
9. c) Outliers cannot conform to the regression relationship
10. The normal distribution is also called bell curve. It is a continuous probability distribution that is symmetric around its mean. Majority of the data falls within one standard deviation of the mean, and as you move farther away from the mean, the probability decreases.
11. We can handle missing data with many ways like :
    - Removing records with missing values.
    - Replacing missing values with the mean or median of the observed values.
    - Using statistical models to impute missing values.
    - Creating multiple imputed datasets to account for uncertainty in missing data.
12. It is a statistical method used to compare two versions of a variable to determine which performs better. It is commonly used in marketing and product development to assess the impact of changes (e.g., a new webpage design or a different advertising strategy) on user behavior. This method is also called split testing.
13. While mean imputation is a simple and quick method to handle missing data, but it has many drawbacks which are :
    - It may introduce bias, especially if the missing data are not missing completely at random.
    - It assumes that the missing values have the same mean as the observed values, which might not be the case.
14. It is a statistical method used to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the observed data. The goal is to find the line that best fits the data points and can be used to predict the value of the dependent variable based on the values of the independent variables.
15. Statistics has mainly two branches :
    - Descriptive statistics - Involves the methods of organizing and summarizing data, such as measures of central tendency (mean, median, mode) and measures of dispersion (range, variance, standard deviation).

- Inferential statistics- Involves making inferences or predictions about a population based on a sample of data. This includes hypothesis testing, confidence intervals, and regression analysis, among other techniques.