Answers for Statistics Worksheet

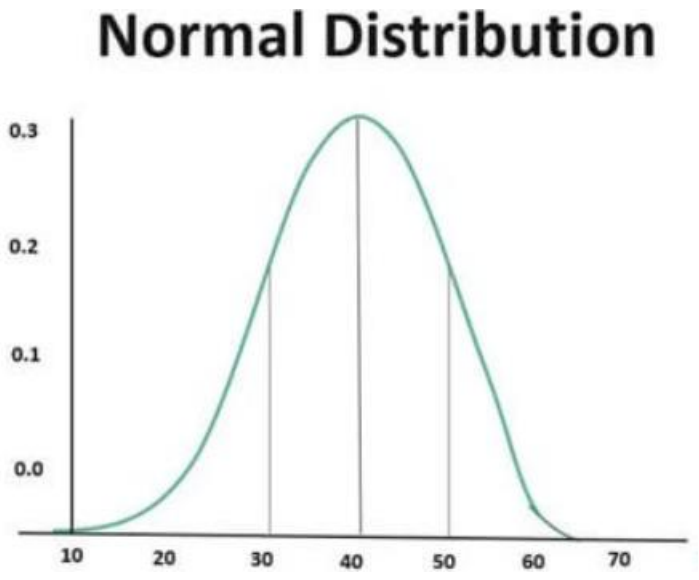
1. A) True
2. A) Central Limit Theorem
3. B) Modeling Bounded count data
4. D) All of the mentioned
5. C) Poisson
6. B) False
7. B) Hypothesis
8. A) 0
9. C) Outliers cannot conform to the regression relationship

10. What do you understand by the term Normal Distribution?

    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 more frequent in occurrence that data far from the mean.

    It is symmetric about the mean and indicates that values near the mean occur more frequently than the values that are farther away from the mean.

    The normal distribution appears as a 'bell curve' when graphed.



11. How do you handle missing data? What imputation techniques do you recommend?

    When dealing with missing data, data scientists can use two primary methods to solve the error: Imputation or data removal.
    The imputation method substitutes reasonable guesses for missing data. It's most useful when the percentage of missing data is low.
    The simplest imputation method is replacing missing values with the mean or median values of the dataset at large, or some similar summary statistic. This has the advantage of being the simplest possible approach, and one that doesn't introduce any undue bias into the dataset.

12. What is A/B testing?

    A/B testing compares two version of an app or webpage to identify the better performer. It's a method that helps you make decisions bases on real data rather than just guessing. It compares options to learn what customers prefer. You can test website/app layouts, email subject lines, product designs, CTA button text, colors etc.

    A/B testing also called split testing or bucket testing compares the performance of two versions of content to see which one appeals more to visitors/viewers. It tests a control

    a. Version against a variant
    b. Version to measure which one is most successful based on your key metrics.

    A/B testing is one of the components of the overarching process of Conversion Rate Optimization (CRO), using which you can gather both qualitative and quantitative user insights. You can further use this collected data to understand user insights. You can further use this collected data to understand user behaviour, engagement rate, pain points, and even satisfaction with website features, including new features, revamped page section, etc. if you're not A/B testing your website, you're website, you're surely losing out on a lot of potential business revenue.

13. Is mean imputation of missing data acceptable practice?

The process of replacing null values in a data collection with the data's mean is known as mean imputation.
In a single imputation method the missing data are filled by some means and the resulting completed data set is used for inference. 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.
Mean imputation is typically considered terrible practice since it ignores feature correlation. Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.
Second, mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

14. What is linear regression in statistics?

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.
Also called simple regression or ordinary least squares(OLS), linear regression establishes the linear relationship between two variables. Linear regression is graphically depicted using a straight line of best fit with the slope defining how the change in one variable impacts a change in other.

15. What are the various branches of statistics?

The two main branches of statistics are Descriptive Statistics and Inferential Statistics.
Both of these are employed in scientific analysis of data and both are equally important for the student of statistics.
In Descriptive Statistics, the Data or Collection Data are described in a summarized way, whereas in Inferential Statistics, we make use of it in order to explain the descriptive kind.
Both of them are used on a large scale.