Statistics worksheet 1_Answeersheet

Q1. Answer: a) True

Q2. Answer: a) Central Limit Theorem

Q3. Answer: b) Modeling bounded count data

Q4. Answer: d) All of the mentioned

Q5. Answer: c) Poisson

Q6. Answer: b) False

Q7. Answer: b) Hypothesis

Q8. Answer: a) 0

Q9. Answer: c) Outliers cannot conform to the regression relationship

Q10. Answer: A normal distribution refers to a probability distribution where the values of a random variable are distributed symmetrically. These values are equally distributed on the left and the right side of the central tendency. Thus, a bell-shaped curve is formed.

Q11. Answer:
Imputation vs Removing Data: Before jumping to the methods of data imputation, we have to understand the reason why data goes missing.

Deletion: Listwise deletion (complete-case analysis) removes all data for an observation that has one or more missing values. Particularly if the missing data is limited to a small number of observations, you may just opt to eliminate those cases from the analysis. However in most cases, it is often disadvantageous to use listwise deletion. This is because the assumptions of MCAR (Missing Completely at Random) are typically rare to support. As a result, listwise deletion methods produce biased parameters and estimates.

Mean, Median and Mode.: Computing the overall mean, median or mode is a very basic imputation method, it is the only tested function that takes no advantage of the time series characteristics or relationship between the variables. It is very fast, but has clear disadvantages. One disadvantage is that mean imputation reduces variance in the dataset.

Q12. Answer: A/B testing is a shorthand for a simple randomized controlled experiment, in which two samples (A and B) of a single vector-variable are compared. These values are similar except for one variation which might affect a user's behavior. A/B tests are widely considered the simplest form of controlled experiment.

Q13. Answer: Mean imputation preserves the mean of the dataset with missing values, This, however, is only appropriate if we assume that our data is normally distributed where it is common to assume that most observations are around the mean anyway. It also is substantially helpful, for small missing data cases.

Q14. Answer: In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables). The case of one explanatory variable is called simple linear regression; for more than one, the process is called multiple linear regression.

Q15. Answer: The two main branches of statistics are:

- Descriptive Statistics
- Inferential Statistics