WORKSHEET

STATISTICS WORKSHEET-1

Q1. (A) True

Q2. (A) Central Limit Theorem

Q3. (B) Modeling bounded count data

Q4. (B)

Q5. (C) Poisson

Q6. (B) False

Q7. (B) Hypothesis

Q8. (A) 0

Q9. (C) Outliers cannot conform to the regression relationship

10. 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 than data far from the mean. In graph form, normal distribution will appear as a bell curve.

Q11. a. Use deletion methods to eliminate missing data. The deletion methods only work for certain datasets where participants have missing fields.

b. Use regression analysis to systematically eliminate data.

c. Data scientists can use data imputation techniques.

And imputation techniques are below-:

1. Complete Case Analysis (CCA).

2. Arbitrary Value Imputation.

3. Frequent Category Imputation.

Q12. A/B testing is a user experience research methodology. A/B tests consist of a randomized experiment with two variants, A and B. It includes application of statistical hypothesis testing or "two-sample hypothesis testing" as used in the field of statistics.

Q13. 1. Bad practice in general.

2. If just estimating means: mean imputation preserves the mean of the observed data.

3. Leads to an underestimate of the standard deviation.

4. Distorts relationships between variables by “pulling”. Estimates of the correlation toward zero.

Q14. 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).

Q15. There are three real branches of statistics:

1. Data collection
2. Descriptive statistics
3. Inferential statistics