STATISTICS WORKSHEET-1

Answer of Q1 – **(A)** True

Answer of Q2 – **(A)** Central Limit Theorem

Answer of Q3 – **(B)** Modeling bounded count data

Answer of Q4 – **(D)** All of the mentioned

Answer of Q5 – **(C)** Poisson

Answer of Q6 – **(B)** False

Answer of Q7 – **(B)** Hypothesis

Answer of Q8 – (A) 0

Answer of Q9 – **(C)** Outliers cannot conform to the regression relationship.

Answer of Q10 –

The normal distribution is a continuous probability distribution that is symmetrical on both sides of the mean, so the right side of the center is a mirror image of the left side. The area under the normal distribution curve represents probability and the total area under the curve sums to one.
Most of the continuous data values in a normal distribution tend to cluster around the mean, and the further a value is from the mean, the less likely it is to occur. The tails are asymptotic, which means that they approach but never quite meet the horizon (i.e. x-axis).
For a perfectly normal distribution the mean, median and mode will be the same value, visually represented by the peak of the curve.

Answer of Q11 –

One way of handling missing values is the deletion of the rows or columns having null values. If any columns/rows have more than half of the values as null then you can drop the entire column/row. We can also use some advanced imputation techniques to solve missing value.

Form my personal experience: - KNN Imputer.

Answer of Q12 –

A/B testing, also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics. Essentially, A/B testing eliminates all the guesswork out of website optimization and enables experience optimizers to make data-backed decisions. In A/B testing, A refers to 'control' or the original testing variable. Whereas B refers to 'variation' or a new version of the original testing variable.
The version that moves your business metric(s) in the positive direction is known as the 'winner.' Implementing the changes of this winning variation on your tested page(s) / element(s) can help optimize your website and increase business ROI.

Answer of Q13 –

Mean imputation: So simple. And yet so dangerous.
Perhaps that's a bit dramatic, but means imputation (also called mean substitution) really ought to be a last resort. 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. This method can lead into severely biased estimates even if data are MCAR.

Since most research studies are interested in the relationship among variables, mean imputation is not a good solution.

Answer of Q14 –

In statistics, linear regre**s**sion is a linear approach for modeling 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. This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable.

Linear Regression is one of the most fundamental and widely known Machine Learning Algorithms which people start with. Building blocks of a Linear Regression Model are:

Discreet/continuous independent variables
A best-fit regression line

Continuous dependent variable i.e. A Linear Regression model predicts the dependent variable using a regression line based on the independent variables. The equation of the Linear Regression is: $Y=a+b*X + e$

Where, a is the intercept, b is the slope of the line, and e is the error term. The equation above is used to predict the value of the target variable based on the given predictor variable(s).

Answer of Q15 –

The two main branches of statistics are:

(1) Descriptive Statistics

(2) Inferential Statistics

(1) Descriptive Statistics – Through graphs or tables, or numerical calculations, descriptive statistics uses the data to provide descriptions of the population.

(2) Inferential Statistics – Based on the data sample taken from the population, inferential statistics makes the predictions and inferences.