STATISTICS WORKSHEET-3 Answers

Ans.1 :- Total Variation = Residual Variation + Regression Variation

Ans.2 :- binomial

Ans.3 :- 2

Ans.4 :- Type-I error

Ans.5 :- Size of the test

Ans.6 :- Increase

Ans.7 :- Hypothesis

Ans.8 :- All of the mentioned

Ans.9 :- 0

Ans.10 :-

Bayes' Theorem:-

- Bayes' Theorem, named after 18th-century British mathematician Thomas Bayes, is a mathematical formula used to determine the conditional probability of events.

- Essentially, the Bayes' theorem describes the probability of an event based on prior knowledge of the conditions that might be relevant to the event.

- The theorem is named after English statistician, Thomas Bayes, who discovered the formula in 1763.

- It is considered the foundation of the special statistical inference approach called the Bayes' inference.

- Besides statistics, the Bayes' theorem is also used in various disciplines, with medicine and pharmacology as the most notable examples.

- In addition, the theorem is commonly employed in different fields of finance.

- Some of the applications include but are not limited to, modelling the risk of lending money to borrowers or forecasting the probability of the success of an investment.

- Formula for Bayes' Theorem:-

  P(A | B) = P(B | A) * P(A) / P(B)

- P(A | B) – the probability of event A occurring, given event B has occurred

- P(B | A) – the probability of event B occurring, given event A has occurred

- P(A) – the probability of event A

- P(B) – the probability of event B

Ans.11 :-

Z-Score:-

- A Z-score is a numerical measurement that describes a value's relationship to the mean of a group of values.

- A Z-score (also called a *standard score*) gives you an idea of how far from the mean a data point is.

- Z-score is measured in terms of standard deviations from the mean.

- If a Z-score is 0, it indicates that the data point's score is identical to the mean score.

- A Z-score of 1.0 would indicate a value that is one standard deviation from the mean.

- Z-scores may be positive or negative, with a positive value indicating the score is above the mean and a negative score indicating it is below the mean.

- A Z-Score is a statistical measurement of a score's relationship to the mean in a group of scores.

Ans.12 :-

t-test:-

- A t-test is an inferential statistic used to determine if there is a significant difference between the means of two groups and how they are related.

- T-tests are used when the data sets follow a normal distribution and have unknown variances, like the data set recorded from flipping a coin 100 times.

- The t-test is a test used for hypothesis testing in statistics and uses the t-statistic, the t-distribution values, and the degrees of freedom to determine statistical significance.

- It is often used in hypothesis testing to determine whether a process or treatment actually has an effect on the population of interest, or whether two groups are different from one another.

- It lets you know if those differences in means could have happened by chance.

- T-tests can be dependent or independent.

Ans.13 :-

Percentile:-

- Percentile is defined as the value below which a given percentage falls under.
- For example, in a group of 20 children, Ben is the 4th tallest and 80% of the children are shorter than you.
- Hence, it means that Ben is at the 80th percentile.
- It is most commonly used in competitive exams such as SAT, LSAT, etc.
- The percentile formula is used when we need to compare the exact values or numbers over the other numbers from the given data i.e. the accuracy of the number.
- Hence, the percentile formula is:

Percentile = (n/N) × 100

Where,

- n = ordinal rank of the given value or value below the number
- N = number of values in the data set
- P = percentile

Ans.14 :-

Percentile:-

- Analysis of variance (ANOVA) is an analysis tool used in statistics that splits an observed aggregate variability found inside a data set into two parts: systematic factors and random factors.

- The systematic factors have a statistical influence on the given data set, while the random factors do not.

- Analysts use the ANOVA test to determine the influence that independent variables have on the dependent variable in a regression study.

- ANOVA is also called the Fisher analysis of variance, and it is the extension of the t- and z-tests.

- Hence, the ANOVA formula is:

  F = MST / MSE

  Where,

- F=ANOVA coefficient

- MST=Mean sum of squares due to treatment

- MSE=Mean sum of squares due to error

Ans.15 :-

How can ANOVA help?:-

- ANOVA is helpful for testing three or more variables.
- It is similar to multiple two-sample t-tests. However, it results in fewer type I errors and is appropriate for a range of issues.
- ANOVA groups differences by comparing the means of each group and includes spreading out the variance into diverse sources.
- It is employed with subjects, test groups, between groups and within groups.
- The one-way ANOVA can help you know whether or not there are significant differences between the means of your independent variables
- such as the first example: age, sex, income
- You may want to test multiple independent variables
- (such as Location, employment status or education).
- When you understand how each independent variable's mean is different from the others, you can begin to understand which of them has a connection to your dependent variable (landing page clicks), and begin to learn what is driving that behaviour.