Statistics Worksheet – 3

1. Which of the following is the correct formula for total variation?

Ans : b) Total Variation = Residual Variation + Regression Variation

2. Collection of exchangeable binary outcomes for the same covariate data are called outcomes.

Ans : c) binomial

3. How many outcomes are possible with Bernoulli trial?

Ans : a) 2

4. If Ho is true and we reject it is called

Ans : a) Type-I error

5. Level of significance is also called:

Ans : b) Size of the test

6. The chance of rejecting a true hypothesis decreases when sample size is:

Ans : b) Increase

7. Which of the following testing is concerned with making decisions using data?

Ans : b) Hypothesis

8. What is the purpose of multiple testing in statistical inference?

Ans : d) All of the mentioned

9. Normalized data are centred at and have units equal to standard deviations of the original data

Ans : a) 0

10. What Is Bayes' Theorem?

In statistics and probability theory, the Bayes' theorem 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 Bayes' theorem is expressed in the following formula:

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

Where:

$P(A|B)$ – 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

11. What is z-score?

Z-score measures the distance between a data point and the mean using standard deviations. Z-scores can be positive or negative. The sign tells you whether the observation is above or below the mean. For example, a z-score of +2 indicates that the data point falls two standard deviations above the mean, while a -2 signifies it is two standard deviations below the mean. A z-score of zero equals the mean. Z-score is also called as standard score.

Formula to calculate z-score is :

$z = x-\mu/\sigma$

μ = population mean

σ = population standard deviation

x = raw score

12. What is t-test ?

A t test is a statistical test that is used to compare the means of two groups. 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.

A t-test can only be used when comparing the means of two groups. If we want to compare more than two groups, or if we want to do multiple pairwise comparisons, we can use an ANOVA test or a post-hoc test.

The t-test is a parametric test of difference, meaning that it makes the same assumptions about the data as other parametric tests. The t-test assumes the data:

- are independent
- are (approximately) normally distributed
- have a similar amount of variance within each group being compared

If the data do not fit these assumptions, we can try a nonparametric alternative to the t-test, such as the Wilcoxon Signed-Rank test for data with unequal variances.

13. What is percentile?

In statistics, a percentile is a term that describes how a score compares to other scores from the same set. While there is no universal definition of percentile, it is commonly expressed as the percentage of values in a set of data scores that fall below a given value.

14. What is ANOVA?

An ANOVA test is a type of statistical test used to determine if there is a statistically significant difference between two or more categorical groups by testing for differences of means using variance.

Another Key part of ANOVA is that it splits the independent variable into 2 or more groups. For example, one or more groups might be expected to influences the dependent variable while the other group is used as a control group, and is not expected to influence the dependent variable.

15. How can ANOVA help?

The one-way ANOVA can help to know whether or not there are significant differences between the means of the independent variables (such as the first example: age, sex, income). When we understand how each independent variable's mean is different from the others, you can begin to understand which of them has a connection to the dependent variable (landing page clicks), and begin to learn what is driving that behaviour.