FLIP ROBO

STATISTICS WORKSHEET-3

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

1. Which of the following is the correct formula for total variation?
   a) Total Variation = Residual Variation – Regression Variation
   b) Total Variation = Residual Variation + Regression Variation
   c) Total Variation = Residual Variation * Regression Variation
   d) All of the mentioned
   Answer- B

2. Collection of exchangeable binary outcomes for the same covariate data are called_____outcomes. a) random
   b) direct
   c) binomial
   d) none of the mentioned
   Answer- C

3. How many outcomes are possible with Bernoulli trial?
   a) 2
   b) 3
   c) 4
   d) None of the mentioned
   Answer- A
   d) None of the mentioned

4. If Ho is true and we reject it is called
   a) Type-I error
   b) Type-II error
   c) Standard error
   d) Sampling error

   Answer-A

5. Level of significance is also called:
   a) Power of the test
   b) Size of the test
   c) Level of confidence
   d) Confidence coefficient
   Answer- B

6. The chance of rejecting a true hypothesis decreases when sample size is:
   a) Decrease
   b) Increase
   c) Both of them
   d) None

   Answer- B

7. Which of the following testing is concerned with making decisions using data?
   a) Probability
   b) Hypothesis
   c) Causal

d) None of the mentioned

8. What is the purpose of multiple testing in statistical inference?
   a) Minimize errors
   b) Minimize false positives
   c) Minimize false negatives
   d) All of the mentioned
   Answer- D

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

   a) 0

   b) 5

   c) 1

   d) 10

   Answer- A

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What Is Bayes' Theorem?
Answer-
Bayes theorem is one of the most popular machine learning concepts that helps to calculate the probability of occurring one event with uncertain knowledge while other one has already occurred.
Formula of Bayes theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Here,
• 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 Here, both events A and B are independent events which means probability of outcome of both events does not depends one another.

11. What is z-score?
Answer-
A Z-score is a numerical measurement that describes a value's relationship to the mean of a group of values. 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.
Example, a Z-score below 1.8 suggests a company might be headed for bankruptcy, while a score closer to 3 suggests a company is in solid financial positioning. **Formula of Z-score**:

$$z = \frac{(x - \mu)}{\sigma}$$

Here:
• μ is the mean of the population,
• σ is the standard deviation of the population.
• x is 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. If we want to compare more than two groups, or if we want to do multiple pairwise comparisons, use an ANOVA test or a post-hoc test as t-test can only be used when comparing the means of two groups.

The t-test is a parametric test of difference, it assumes our data:

1. are independent
2. are (approximately) normally distributed.
3. have a similar amount of variance within each group being compared.

13. What is percentile?

In statistics, percentiles are used to understand and interpret data. They indicate the values below which a certain percentage of the data in a data set is found. Percentiles are frequently used to understand test scores and biometric measurements.

The nth percentile of a set of data is the value at which n percent of the data is below it.

In everyday life, percentiles are used to understand values such as test scores, health indicators, and other measurements. For example, an 18-year-old male who is six and a half feet tall is in the 99th percentile for his height. This means that of all the 18-year-old males, 99 percent have a height that is equal to or less than six and a half feet. An 18-year-old male who is only five and a half feet tall, on the other hand, is in the 16th percentile for his height, meaning only 16 percent of males his age is the same height or shorter.

Percentiles for the values in a given data set can be calculated using the formula:

$n = (P/100) \times N$

Here,

• N = number of values in the data set,
• P = percentile, and
• n = ordinal rank of a given value (with the values in the data set sorted from smallest to largest).

14. What is ANOVA?

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. The Formula for ANOVA is:

$F = MST/MSE$

Here:

• F=ANOVA coefficient
• MST=Mean sum of squares due to treatment
• MSE=Mean sum of squares due to error

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.

A one-way ANOVA evaluates the impact of a sole factor on a sole response variable. It determines whether all the samples are the same. The one-way ANOVA is used to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups.

A two-way ANOVA is an extension of the one-way ANOVA. With a one-way, you have one independent variable affecting a dependent variable. With a two-way ANOVA, there are two independents.