

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

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

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

Ans. c) binomial

3. How many outcomes are possible with Bernoulli trial?

- a) 2
- b) 3
- c) 4
- d) None of the mentioned

Ans. a) 2

4. If H_0 is true and we reject it is called

- a) Type-I error
- b) Type-II error

- c) Standard error
- d) Sampling error

Ans. a) Type-I error

5. Level of significance is also called:

- a) Power of the test
- b) Size of the test
- c) Level of confidence
- d) Confidence coefficient

Ans. b) Size of the test

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

- a) Decrease
- b) Increase
- c) Both of them
- d) None

Ans. b) Increase

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

- a) Probability
- b) Hypothesis
- c) Causal
- d) None of the mentioned

Ans. b) Hypothesis

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

Ans. d) All of the mentioned

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

Ans. a) 0

Q10 and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What Is Bayes' Theorem?

Ans. Bayes' Theorem governs the likelihood that one event is based on the occurrence of some other events. It depends upon the concept of conditional probability. This theorem gives us the probability of some events depending upon some conditions related to the event.

Formula of Bayes' Theorem

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

where ,

probability of occurrence of event A

probability of occurrence of event B

probability of occurrence of event A given B

probability of occurrence of event B given A

Bayes' Theorem Statement

Let $B_1, B_2, B_3, \dots, B_n$ is a series of n non-empty sets with sample space as X , where all the sets have a non-zero probability of occurrence

and they form a partition of X. Let E be an event associated with X, then by Bayes' Theorem we get:

$$p(B_i|X) = \frac{p(B_i) * p(X|B_i)}{\sum_{j=1}^n p(B_j) * p(X|B_j)}$$

For all $i = 1, 2, 3, \dots, n$

11. What is z-score?

Ans. A z-score is the distance between a data point and the mean using standard deviations. Z-score can be positive or negative. The sign tells whether the observation is above or below the mean. The negative z-score denotes the observation is below the mean and positive z-score denotes the observation is above the mean. The zero z-score means the observation is equal to mean. Z-score is also known as standard score.

Calculation of z-scores:

The z-score is calculated by taking deviation of data point from mean divided by standard deviation.

Formula for z-score

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

Where, x = data point of interest

μ = mean of population

σ = standard deviation of population

In case if we don't know the population values we can use the sample values i.e., sample mean and sample standard deviation.

Specifically, the z-score follow the standard normal distribution with mean 0 and SD 1. However, skewed data will produce z-score which are similarly skewed.

12. What is t-test?

Ans. A statistical test used to compare the means of two groups of data.

A t-test is an inferential statistic used to determine if there is a significant difference between means of two groups and how they are related. T-tests are used when the datasets are normally distributed and have unknown variances.

For example, a dataset recorded from tossing a coin 100 times.

A t-test compares the average values of two datasets and determines if they came from the same population.

For example, a sample of students from class A and sample of students from class B would not likely have the same mean and standard deviation.

Formula:

$$t = \frac{\text{mean1} - \text{mean2}}{\frac{s(\text{diff})}{\sqrt{n}}}$$

Where,

Mean1 = average value of first sample set

Mean2 = average value of second sample set

S(diff) = SD of the differences of paired data values

n = sample size (number of paired differences)

13. What is percentile?

Ans. A percentile is a measure used in statistics indicating the value below which a given percentage of observations in a group of observations fall.

For example, the 20th percentile is the value below which 20% of the observations may be found.

The percentiles are also known centiles.

The 25th percentile is also known as 1st quartile.

The 50th percentile is known as 2nd quartile or median.

The 75th percentile is known as 3rd quartile.

The term percentile is often used in the reporting of scores from non-referenced tests.

For example, if a score is at the 85th percentile it is equal to the value below which the 85% of the observations may be found.

Formula:

$$p = \left(\frac{n}{N} \right) * 100$$

Where,

n = number of values below X

N = total count of population

14. What is ANOVA?

Ans. 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 ANOVA test allows a comparison of more than two groups at the same time to determine whether a relationship exists between them. The result of the ANOVA formula, the F statistic (also called the F-ratio), allows for the analysis of multiple groups of data to determine the variability between samples and within samples.

If no real difference exists between the tested groups, which is called the null hypothesis, the result of the ANOVA's F-ratio statistic will be close to 1. The distribution of all possible values of the F statistic is the F-distribution. This is actually a group of distribution functions, with two characteristic numbers, called the numerator degrees of freedom and the denominator degrees of freedom.

Formula:

$$F = \frac{MST}{MSE}$$

Where, F is the coefficient of ANOVA

MST is the mean sum of squares due to treatments

MSE is the mean sum of squares due to error

15. How can ANOVA help?

Ans. 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.

There are two main types of ANOVA: one-way (or unidirectional) and two-way. There also variations of ANOVA. For example, MANOVA (multivariate ANOVA) differs from ANOVA as the former tests for multiple dependent variables simultaneously while the latter assesses only one dependent variable at a time. One-way or two-way refers to the number of independent variables in your analysis of variance test. 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. For example, a two-way ANOVA allows a company to compare worker productivity based on two independent variables, such as salary and skill set. It is utilized to observe the interaction between the two factors and tests the effect of two factors at the same time.