

STATISTICS WORKSHEET-3

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

Answers Are Marked in Yellow Highlights

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
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
3. How many outcomes are possible with Bernoulli trial?
 - a) 2
 - b) 3
 - c) 4
 - d) None of the mentioned
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
5. Level of significance is also called:
 - a) Power of the test
 - b) Size of the test
 - c) Level of confidence
 - d) Confidence coefficient
6. The chance of rejecting a true hypothesis decreases when sample size is:
 - a) Decrease
 - b) Increase
 - c) Both of them
 - d) None
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

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

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

10. What Is Bayes' Theorem?

Answer:

Bayes theorem is a mathematical formula, which is used to determine the conditional probability of the given event. Conditional probability is defined as the likelihood that an event will occur, based on the occurrence of a previous outcome. It is also considered for the case of **conditional probability**. Bayes theorem is also known as the formula for the **probability of “causes”**

Conditional probability: Bayes' Theorem

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

© Byjus.com

Here, P(A) = how likely A happens(Prior knowledge)- The probability of a hypothesis is true before any evidence is present.

P(B) = how likely B happens(Marginalization)- The probability of observing the evidence.

P(A/B) = how likely A happens given that B has happened(Posterior)-The probability of a hypothesis is true given the evidence.

P(B/A) = how likely B happens given that A has happened(Likelihood)- The probability of seeing the evidence if the hypothesis is true.

Example : Assume that the chances of a person having a skin disease are 40%. Assuming that skin creams and drinking enough water reduces the risk of skin disease by 30% and prescription of a certain drug reduces its chance by 20%. At a time, a patient can choose any one of the two options with equal probabilities. It is given that after picking one of the options, the patient selected at random has the skin disease. Find the probability that the patient picked the option of skin creams and drinking enough water using the Bayes theorem.

Solution: Assume E1: The patient uses skin creams and drinks enough water;

E2: The patient uses the drug;

A: The selected patient has the skin disease

$$P(E1) = P(E2) = 1/2$$

Using the probabilities known to us, we have

$$P(A|E1) = 0.4 \times (1-0.3) = 0.28$$

$$P(A|E2) = 0.4 \times (1-0.2) = 0.32$$

Using Bayes Theorem, the probability that the selected patient uses skin creams and drinks enough water is given by,

$$P(E1|A) = \frac{P(A|E1)P(E1)}{P(A|E1)P(E1) + P(A|E2)P(E2)} = \frac{P(A|E1)P(E1)}{P(A|E1)P(E1) + P(A|E2)P(E2)}$$

$$= \frac{(0.28 \times 0.5)}{(0.28 \times 0.5 + 0.32 \times 0.5)}$$

$$= 0.14/(0.14 + 0.16)$$

$$= 0.47$$

Answer: The probability that the patient picked the first option is 0.47

11. What is z-score?

Answer: A z score is a type of statistical measurement that gives an idea of how far a raw score is from the mean of a distribution. A z score is used in a z test for [hypothesis testing](#). It is also used in prediction intervals to determine the probability of a [random variable](#) falling between a range of values.

A z score can be defined as a measure of the number of [standard deviations](#) by which a score is below or above the [mean](#) of a distribution. In other words, it is used to determine the distance of a score from the mean. If the z score is positive it indicates that the score is above the mean. If it is negative then the score will be below the mean. However, if the z score is 0 it denotes that the data point is the same as the mean.

Z Score Formula

To calculate a z score, knowledge of the mean and standard deviation is required. When the population mean and population standard deviation are known then the z score formula is given as follows:

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

Here, μ = population mean

σ = population standard deviation

x = raw score

Z Score Interpretation

If a **z score is 3** it implies that the raw score is 3 standard deviations above the mean.

A **z score of -3** indicates that the raw score is 3 deviations below the mean.

The z score also shows where the raw score will be on a normal distribution curve.

Example 1: Jake scored 70 marks on a test. The mean score of the class was 60 with a standard deviation of 15. Calculate the z score for the marks secured by Jake using the z score formula.

Solution: To find: z score for marks secured by Jake

Given: Marks secured by Jake, $x = 70$

Standard deviation, $\sigma = 15$

Mean marks, $\mu = 60$

Using z score formula,

$$z \text{ score for secured marks} = z = \frac{(70 - 60)}{15}$$

$$= \frac{10}{15}$$

$$= 0.6667$$

Answer: z score for Jake's marks = 0.6667

12. What is t-test?

A t-test is a statistical test that compares the means of two samples. It is used in hypothesis testing, with a null hypothesis that the difference in group means is zero and an alternate hypothesis that the difference in group means is different from zero.

Example: If you want to know whether the mean petal length of iris flowers differs according to their species. You find two different species of irises growing in a garden and measure 25 petals of each species. You can test the difference between these two groups using a t test and null and alternative hypotheses.

- The null hypothesis (H_0) is that the true difference between these group means is zero.
- The alternate hypothesis (H_a) is that the true difference is different from zero.

13. What is percentile?

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?

Analysis of Variance

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 influence 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?

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 influence the dependent variable while the other group is used as a control group, and is not expected to influence the dependent variable.