

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

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

3. How many outcomes are possible with Bernoulli trial?

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

Ans. a

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

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

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

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

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

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

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

- 10. What Is Bayes' Theorem?
- 11. What is z-score?
- 12. What is t-test?
- 13. What is percentile?
- 14. What is ANOVA?
- 15. How can ANOVA help?

Bayes' Theorem :-

Bayes' Theorem states that the conditional probability of an event, based on the occurrence of another event, is equal to the likelihood of the second event given the first event multiplied by the probability of the first event.

Z-score :-

Z score is also called standard score. This score helps to understand if a data value is greater or smaller than mean and how far away it is from the mean. More

specifically, Z score tells how many standard deviations away a data point is from the mean.

The z-scores can be used to compare data with different measurements and for normalization of data for machine learning algorithms and comparisons.

$$Z \text{ score} = (x - \text{mean}) / \text{std. deviation.}$$

We can calculate z-scores in Python using `scipy.stats.zscore`,.

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.

*one sample t test- The one-sample t-test is a statistical hypothesis test used to determine whether an unknown population mean is different from a specific value.

For the one-sample t-test, we need one continuous variable.

*two sample t test- Two- sample t-test is used when the data of two samples are statistically independent

*paired t test- while the paired t-test is used when data is in the form of matched pairs.25-Ju

Percentile :-

a value on a scale of 100 that indicates the percent of a distribution that is equal to or below it a score in the 95th 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.

Percentiles represent how your child's weight compares to other children of the same height. (i.e. 70th percentile refers to your child being heavier than 70% of all other children at her height).

ANOVA TEST :-

Analysis of variances

ANOVA stands for Analysis of Variance. It's a statistical test that was developed by Ronald Fisher in 1918 and has been in use ever since. Put simply, ANOVA tells you if there are any statistical differences between the means of three or more independent groups.

The t-test is a method that determines whether two populations are statistically different from each other, whereas ANOVA determines whether three or more populations are statistically different from each other.

ANOVA is used as a test of means for two or more populations.

ANOVA used variance based F test to cheque group mean equality

one-way(one factor),two-way(two factor),ANOVA(factor in an independent variable)

ANOVA HYPOTHESIS

null hypothesis H_0 -groups mean are equal(no variances in mean of group)

Alternate hypothesis H_1 -at least one mean of group is different from other group(all mean are not equal)