

Solved Statistics Worksheet Set 3

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

10. What Is Bayes' Theorem?

Ans-

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.

Consider that A and B are any two events from a sample space S where $P(B) \neq 0$. Using our understanding of conditional probability, we have:

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

Similarly, $P(B|A) = P(A \cap B) / P(A)$

It follows that $P(A \cap B) = P(A|B) * P(B) = P(B|A) * P(A)$

Thus, $P(A|B) = P(B|A) * P(A) / P(B)$

Here, $P(A)$ and $P(B)$ are probabilities of observing A and B independently of each other. That's why we can say that they are marginal probabilities. $P(B|A)$ and $P(A|B)$ are conditional probabilities.

11. What is z-score?

Ans-

A z-score (also called a *standard score*) gives one an idea of how far from the mean a data point is. It's a measure of how many standard deviations below or above the population mean a raw score is.

A z-score can be placed on a normal distribution curve. Z-scores range from -3 standard deviations (which would fall to the far left of the normal distribution curve) up to +3 standard deviations (which would fall to the far right of the normal distribution curve). In order to use a z-score, you need to know the mean μ and also the population standard deviation σ .

The basic z score formula for a sample is:

$$z = (x - \mu) / \sigma$$

12. What is t-test?

Ans-

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

The t-test is a test used for hypothesis testing in statistics and uses the t-statistic, the t-distribution values, and the degrees of freedom to determine statistical significance.

13. What is percentile?

Ans-

A percentile (or a centile) 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 (or score) below which 20% of the observations may be found.

The term percentile and the related term percentile rank are often used in the reporting of scores from norm-referenced tests. The 25th percentile is also known as the first quartile (Q1), the 50th percentile as the median or second quartile (Q2), and the 75th percentile as the third quartile (Q3). In general, percentiles and quartiles are specific types of quantiles.

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. The ANOVA test is used to determine the influence that independent variables have on the dependent variable in a regression study. The Formula for ANOVA is:

$$F \text{ (annova coefficient)} = \frac{\text{MST (Mean sum of squares due to treatment)}}{\text{MSE (Mean sum of squares due to error)}}$$

15. How can ANOVA help?

Ans-

The type of ANOVA test used depends on a number of factors. It is applied when data needs to be experimental. Analysis of variance is employed if there is no access to statistical software resulting in computing ANOVA by hand. It is simple to use and best suited for small samples. With many experimental designs, the sample sizes have to be the same for the various factor level combinations. 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.
