

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?
b) Total Variation = Residual Variation + Regression Variation
2. Collection of exchangeable binary outcomes for the same covariate data are called outcomes.
c) binomial
3. How many outcomes are possible with Bernoulli trial?
a) 2
4. If H_0 is true and we reject it is called
a) Type-I error
5. Level of significance is also called:
c) Level of confidence
6. The chance of rejecting a true hypothesis decreases when sample size is:
b) Increase
7. Which of the following testing is concerned with making decisions using data?
b) Hypothesis
8. What is the purpose of multiple testing in statistical inference?
d) All of the mentioned
9. Normalized data are centred at and have units equal to standard deviations of the original data
a) 0

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

10. What Is Bayes' Theorem?

Bayes' Theorem, named after 18th-century British mathematician Thomas Bayes, is a mathematical formula for determining [conditional probability](#). Conditional probability is the likelihood of an outcome occurring, based on a previous outcome having occurred in similar circumstances. Bayes' theorem provides a way to revise existing predictions or theories (update probabilities) given new or additional evidence.

In finance, Bayes' Theorem can be used to rate the [risk](#) of lending money to potential borrowers. The theorem is also called Bayes' Rule or Bayes' Law and is the foundation of the field of Bayesian statistics.

- Bayes' Theorem allows you to update the predicted probabilities of an event by incorporating new information.
- Bayes' Theorem was named after 18th-century mathematician Thomas Bayes.
- It is often employed in finance in calculating or updating risk evaluation.
- The theorem has become a useful element in the implementation of machine learning.
- The theorem was unused for two centuries because of the high volume of calculation capacity required to execute its transactions.

11. What is z-score?

The [Z-score](#), or standard score, is the number of standard deviations a given data point lies above or below the mean. The mean is the average of all values in a group, added together, and then divided by the total number of items in the group.

To calculate the Z-score, subtract the mean from each of the individual data points and divide the result by the standard deviation. Results of zero show the point and the mean equal. A result of one indicates the point is one standard deviation above the mean and when data points are below the mean, the Z-score is negative.

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.

13. What is percentile?

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.

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.

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.