

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

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

c) binomial

3. How many outcomes are possible with Bernoulli trial?
- a) 2
 - b) 3
 - c) 4
 - d) None of the mentioned

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

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

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

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

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

d) All of the mentioned



WORKSHEET

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

a) 0

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?

10. Bayes Theorem is very helpful to estimate the precision of values. To know how, let's start this from the very beginning. Bayes Theorem is a statement and theorem given by an 18th-century mathematician from Britain. The formula he deduced is effective and being used in conditional probability. Now, what exactly is conditional probability?

It is a term used for the likelihood of deriving an answer to a question or the probable outcome on the basis of its previous outcomes. A single statement is a method or process of cross-checking or revising the existing predictions to eliminate the chances and possibilities of making mistakes. It is how we state the Bayes Theorem in Machine Learning.

11. The Z-score, or standard score, is a fractional representation of standard deviations from the mean value. Accordingly, z-scores often have a distribution with no average and standard deviation of 1. Formally, the z-score is defined as:

Calculating the Standard Score (Z-Score)

$$\text{Standard Score, } z = \frac{X - \mu}{\sigma}$$

TERMS:

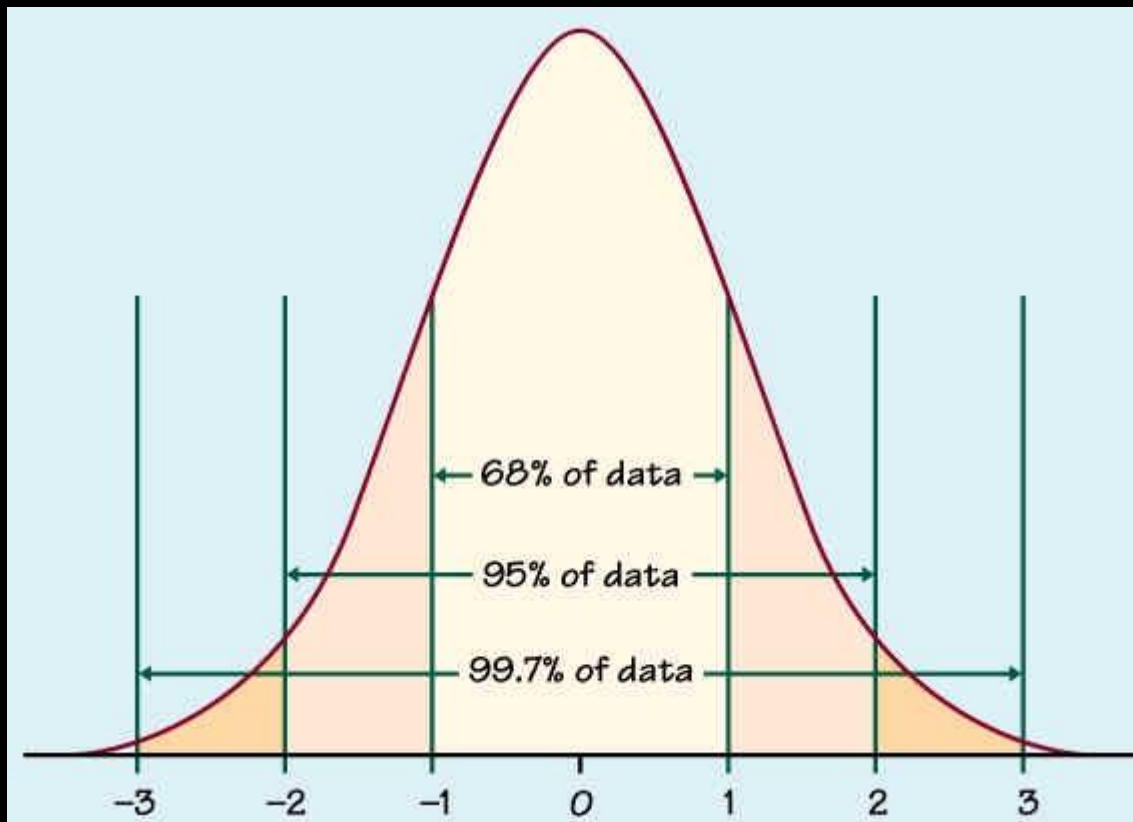
μ = mean (pronounced 'mu')

X = score

σ = standard deviation (pronounced 'sigma')

One cannot calculate the z-score without first knowing the mean and standard deviation of the complete population.

As described above, the z-score works by taking a sample score and subtracting the mean score, before then dividing by the standard deviation of the total population. It can be easy to think of this using an example. Imagine having a score of 70 out of 100, where the mean score is 60, and the standard deviation is 15. Using the z-score function, one can calculate the z-score of .6667 from the data. By comparing the z-score the standard normal distribution table, one can determine that the probability of a score being greater than .67 is .2514 or 25.14%. In short, the score is better than roughly 75% of the rest of the scores.



12. The t-test compares the means (averages) of two populations to determine how different they are from each other. The test generates a T-score and P-value, which quantify exactly how different each population is and the likelihood that this difference can be explained by chance or sampling error.

There are three variations of the t-test used for different scenarios:

- Independent Samples: compares the averages for two groups.
- One Sample: tests the averages of a single group against a known average.
- Paired Sample: compares averages from the same group at different points of time.

The null hypothesis for the Independent and One Sample tests is that the means of both groups are identical. For the Paired Sample, the null hypothesis is that the pairs of differences between both tests are equal.

Regardless of the technique used, the test will output a t-score. This score is simply the ratio between the mean difference across two groups, as well as the difference within the groups. The bigger the t score, the larger the difference between samples, which also means the test results are more likely reproducible. A smaller score means more similarity between groups.

13. Percentiles

Percentiles are used in statistics data to provide you a number that expresses the value that a given percent of the values are lower than.

Example: We have an array of the ages of all the people that working in same office
ages = [25,31,43,48,50,41,39,60,52,32,27,46,47,55]

What is 75. percentile? The answer is 48, meaning that 75% of the people are 48 or younger.

14. Analysis of variance, or ANOVA, is **a statistical method that separates observed variance data into different components to use for additional tests**. A one-way ANOVA is used for three or more groups of data, to gain information about the relationship between the dependent and independent variables.

15. It provides the overall test of equality of group means. It can control the overall type I error rate (i.e. false positive finding) It is a parametric test so it is more powerful, if normality assumptions hold true.



FLIP ROBO