STATISTICS ASSIGNMENT 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

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 Ho 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

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What Is Bayes' Theorem?

Answer:

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.

- 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.

Applications of Bayes' Theorem are widespread and not limited to the financial realm. For example, Bayes' theorem can be used to determine the accuracy of medical test results by taking into consideration how likely any given person is to have a disease and the general accuracy of the test. Bayes' theorem relies on incorporating prior probability distributions in order to generate posterior probabilities.

Prior probability, in Bayesian statistical inference, is the probability of an event occurring before new data is collected. In other words, it represents the best rational assessment of the probability of a particular outcome based on current knowledge before an experiment is performed.

Posterior probability is the revised probability of an event occurring after taking into consideration the new information. Posterior probability is calculated by updating the prior probability using Bayes' theorem. In statistical terms, the posterior probability is the probability of event occurring given that event B has occurred.

11. What is z-score?

Answer:

A Z-score is a numerical measurement that describes a value's relationship to the mean of a group of values. Z-score is measured in terms of standard deviations from the mean. If a Z-score is 0, it indicates that the data point's score is identical to the mean score. A Z-score of 1.0 would indicate a value that is one standard deviation from the mean. Z-scores may be positive or negative, with a positive value indicating the score is above the mean and a negative score indicating it is below the mean.

- A Z-Score is a statistical measurement of a score's relationship to the mean in a group of scores.
- A Z-score can reveal to a trader if a value is typical for a specified data set or if it is atypical.
- In general, a Z-score below 1.8 suggests a company might be headed for bankruptcy, while a score closer to 3 suggests a company is in solid financial positioning.

Z-scores reveal to statisticians and traders whether a score is typical for a specified data set or if it is atypical. Z-scores also make it possible for analysts to adapt scores from various data sets to make scores that can be compared to one another more accurately.

12. What is t-test?

Answer:

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, like the data set recorded from flipping a coin 100 times.

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.

- A t-test is an inferential statistic used to determine if there is a statistically significant difference between the means of two variables.
- The t-test is a test used for hypothesis testing in statistics.
- Calculating a t-test requires three fundamental data values including the difference between the mean values from each data set, the standard deviation of each group, and the number of data values.
- T-tests can be dependent or independent.

A t-test compares the average values of two data sets and determines if they came from the same population. In the above examples, a sample of students from class A and a sample of students from class B would not likely have the same mean and standard deviation. Similarly, samples taken from the placebo-fed control group and those taken from the drug prescribed group should have a slightly different mean and standard deviation.

Mathematically, the t-test takes a sample from each of the two sets and establishes the problem statement. It assumes a null hypothesis that the two means are equal.

Using the formulas, values are calculated and compared against the standard values. The assumed null hypothesis is accepted or rejected accordingly. If the null hypothesis qualifies to be rejected, it indicates that data readings are strong and are probably not due to chance.

The t-test is just one of many tests used for this purpose. Statisticians use additional tests other than the t-test to examine more variables and larger sample sizes. For a large sample size, statisticians use a z-test. Other testing options include the chi-square test and the f-test.

13. What is percentile?

Answer:

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. For example, if a score is at the 86th percentile, where 86 is the percentile rank, it is equal to the value below which 86% of the observations may be found. In contrast, if it is in the 86th percentile, the score is at or below the value of which 86% of the observations may be found. *Every score is in the 100th percentile*.
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?

Answer:

A common approach to figure out a reliable treatment method would be to analyse the days it took the patients to be cured. We can use a statistical technique which can compare these three treatment samples and depict how different these samples are from one another. Such a technique, which compares the samples on the basis of their means, is called ANOVA.

Analysis of variance (ANOVA) is a statistical technique that is used to check if the means of two or more groups are significantly different from each other. ANOVA checks the impact of one or more factors by comparing the means of different samples. We can use ANOVA to prove/disprove if all the medication treatments were equally effective or not.

15. How can ANOVA help?

Answer:

ANOVA helps you compare how different groups are different from each other and allows you to see if any two groups are statistically similar.

Because it can be a complex procedure, it's not often used in journalism (unless you're one of those fancy data-driven journalists) but it is frequently used in academic research. For example, let's say you're studying how different brands of salad dressing affect the taste of salad (the dependent variable).

You would have your different brands as independent variables maybe Caesar dressing, Italian dressing, Blue Cheese dressing, and Thousand Island dressing. You could poll your participants on their preferred salad flavour before and after trying all four dressings.

The ANOVA method would help you evaluate whether or not there was a statistically significant difference between the four dressings 'effect on participants' preferred flavour profiles. Then you might be able to say something like "participants preferred salad flavours were most highly influenced by Blue Cheese dressing."