

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) 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 : a) Power 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: A) Minimize errors

9) Normalized data are centred at _____ and have units equal to standard deviations of the original data

- e) 0
- f) 5
- g) 1
- h) 10

ANS : a) 0

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

10) What Is Bayes' Theorem?

ANS: In statistics and probability theory, the Bayes' theorem (also known as the Bayes' rule) is a mathematical formula used to determine the conditional probability of events. Essentially, the Bayes' theorem describes the probability of an event based on prior knowledge of the conditions that might be relevant to the event.

The theorem is named after English statistician, Thomas Bayes, who discovered the formula in 1763. It is considered the foundation of the special statistical inference approach called the Bayes' inference.

Formula for Bayes' Theorem

The Bayes' theorem is expressed in the following formula:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where:

- $P(A|B)$ – the probability of event A occurring, given event B has occurred
- $P(B|A)$ – the probability of event B occurring, given event A has occurred
- $P(A)$ – the probability of event A
- $P(B)$ – the probability of event B

Note that events A and B are independent events (i.e., the probability of the outcome of event A does not depend on the probability of the outcome of event B).

A special case of the Bayes' theorem is when event A is a binary variable. In such a case, the theorem is expressed in the following way:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A^-)P(A^-) + P(B|A^+)P(A^+)}$$

Where:

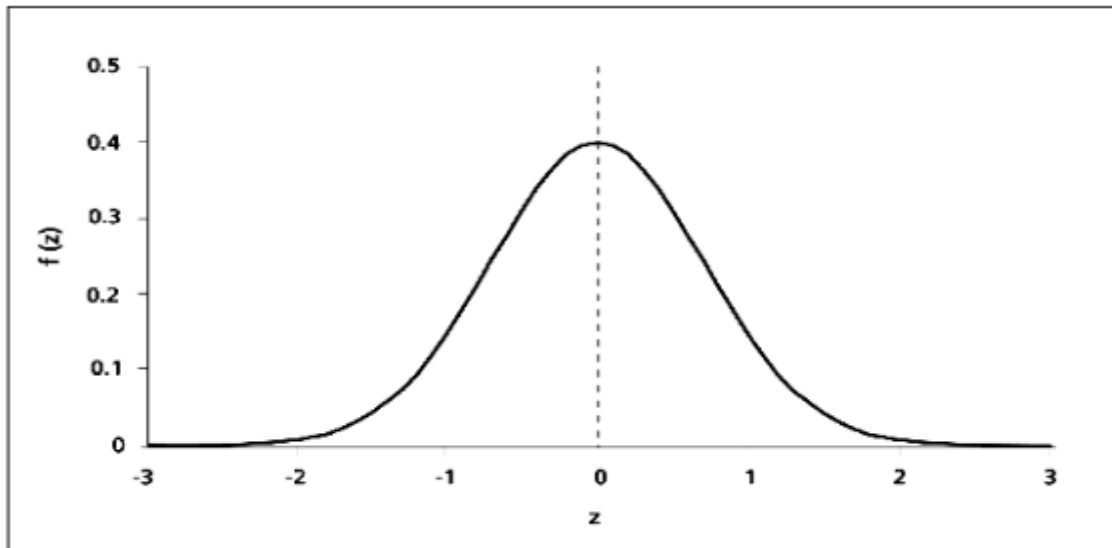
- $P(B|A^-)$ – the probability of event B occurring given that event A^- has occurred
- $P(B|A^+)$ – the probability of event B occurring given that event A^+ has occurred

In the special case above, events A^- and A^+ are mutually exclusive outcomes of event A.

11).What is z-score?

A z-score describes the position of a raw score in terms of its distance from the mean, when measured in standard deviation units. The z-score is positive if the value lies above the mean, and negative if it lies below the mean.

It is also known as a standard score, because it allows comparison of scores on different kinds of variables by standardizing the distribution. A standard normal distribution (SND) is a normally shaped distribution with a mean of 0 and a standard deviation (SD)



The formula for calculating a z-score is $z = (x - \mu) / \sigma$, where x is the raw score, μ is the population mean, and σ is the population standard deviation.

As the formula shows, the z-score is simply the raw score minus the population mean, divided by the population standard deviation.

12)What is t-test?

A t-test is **a statistical test that compares the means of two samples**. It is used in hypothesis testing, with a null hypothesis that the difference in group means is zero and an alternate hypothesis that the difference in group means is different from zero.

For comparing the mean of a population \bar{x} from n samples, with a specified theoretical mean μ , we use a one-sample t-test.

t-Test Formula



$$t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

where,

'x' bar is the mean of the sample,

μ is the assumed mean,

σ is the standard deviation

and n is the number of observations

where σ/\sqrt{n} is the standard error

13)What is percentile?

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.

You can calculate percentiles in statistics using the following formula:

$$P_x = \frac{x(n + 1)}{100}$$

P_x = The value at which x percentage of data lie below that value

n = Total number of observations

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.

The Formula for ANOVA is:

$$F = MST/MSE$$

where: F=ANOVA coefficient

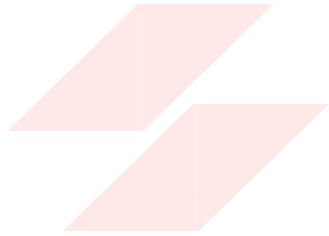
MST=Mean sum of squares due to treatment

MSE=Mean sum of squares due to error

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.

The one-way ANOVA can help you know whether or not there are significant differences between the means of your independent variables (such as the first example: age, sex, income). When you understand how each independent variable's mean is different from the others, you can begin to understand which of them has a connection to your dependent variable (landing page clicks), and begin to learn what is driving that behavior.



FLIP ROBO