

STATISTICS WORKSHEET-3

1. b) Total Variation = Residual Variation + Regression Variation
2. c) binomial
3. a) 2
4. a) Type-I error
5. b) Size of the test
6. b) Increase
7. b) Hypothesis
8. d) All of the mentioned
9. a) 0
10. 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.

Formula



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

A, B = events

$P(A|B)$ = probability of A given B is true

$P(B|A)$ = probability of B given A is true

$P(A), P(B)$ = the independent probabilities of A and B

STATISTICS WORKSHEET-3

14. In statistics, the z-scores or standard score is the number of standard deviations by which the value of a raw score (i.e., an observed value or data point) is above or below the mean value of what is being observed or measured. Raw scores above the mean have positive standard scores, while those below the mean have negative standard scores.

It is calculated by subtracting the population mean from an individual raw score and then dividing the difference by the population standard deviation. This process of converting a raw score into a standard score is called standardising or normalising (however, "normalising" can refer to many types of ratios; see normalisation for more).

Standard scores are most commonly called z-scores

Formula

$$Z = \frac{x - \mu}{\sigma}$$

Z = standard score

x = observed value

μ = mean of the sample

σ = standard deviation of the sample

STATISTICS WORKSHEET-3

15. 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 formula for computing the t-value and degrees of freedom for a paired t-test is:

$$T = \frac{mean1 - mean2}{\frac{s(diff)}{\sqrt{(n)}}}$$

where:

mean1 and mean2=The average values of each of the sample sets

s(diff)=The standard deviation of the differences of the paired data values

n=The sample size (the number of paired differences)

n-1=The degrees of freedom

16. 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.

Formula:

STATISTICS WORKSHEET-3

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

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

n = Total number of observations

17. Analysis of Variance (ANOVA) is a statistical formula used to compare variances across the means (or average) of different groups. A range of scenarios use it to determine if there is any difference between the means of different groups.

18. 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.