

WORKSHEET 3 - STATISTICS

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

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

3. How many outcomes are possible with Bernoulli trial?

- a) 2
- b) 3
- c) 4
- d) None of the mentioned

ANS: A

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

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: B

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

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

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: D

9. Normalized data are centered at and have units equal to standard deviations of the original data

- a) 0
- b) 5
- c) 1

d) 10

ANS: A

10. What Is Bayes' Theorem?

ANS: Bayes' theorem describes the probability of occurrence of an event related to any condition.

It is also considered for the case of conditional probability

If A and B are two events the formula for Bayes theorem is:

$$P(A|B) = [P(B|A) \cdot P(A)] / P(B)$$

Where P(A) and P(B) are the probabilities of events A and B.

P(A|B) is the probability of event A given B

P(B|A) is the probability of event B given A.

11. What is z-score?

ANS: A z-score (also called a standard score) gives you an idea of how far from the mean a data point is. But more technically it's a measure of how many standard deviations below or above the population mean a raw score is. A z-score can be placed on a normal distribution curve. Z-scores range from **-3** standard deviations (which would fall to the far left of the normal distribution curve) up to **+3** standard deviations (which would fall to the far right of the normal distribution curve). In order to use a z-score, you need to know the mean μ and also the population standard deviation σ . Z-scores are a way to compare results to a "normal" population. Results from tests or surveys have thousands of possible results and units; those results can often seem meaningless.

For example, knowing that someone's weight is 150 pounds might be good information, but if we want to compare it to the "average" person's weight, looking at a vast table of data can be overwhelming (especially if some weights are recorded in kilograms). A z-score can tell us where that person's weight is compared to the average population's mean weight.

z score formula:-

$$z = (x - \mu) / \sigma$$

12. What is t-test?

ANS: A T-test is the final statistical measure for determining differences between two means that may or may not be related. The t test is usually used when data sets follow a normal distribution but we don't know the population variance. A T-test studies a set of data gathered from two similar or different groups to determine the probability of the difference in the result than what is usually obtained. The accuracy of the test depends on various factors, including the distribution patterns used and the variants influencing the collected samples. Depending on the parameters, the test is conducted, and a T-value is obtained as the statistical inference of the probability of the usual

resultant being driven by chance.

For example, if one wishes to figure out if the mean of the length of petals of a flower belonging to two different species is the same, a T-test can be done. The user can select petals randomly from two other species of that flower and come to a standard conclusion. The final T-test interpretation could be obtained in either of the two ways:

A null hypothesis signifies that the difference between the means is zero and where both the means are shown as equal.

An alternate hypothesis implies the difference between the means is different from zero. This hypothesis rejects the null hypothesis, indicating that the data set is quite accurate and not by chance.

This T-test, however, is only valid and should be done when the mean or average of only two categories or groups needs to be compared. As soon as the number of comparisons to be made is more than two, conducting this is not recommended.

13. What is percentile?

ANS: In statistics, 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:

$$Px = x (n + 1) / 100$$

Px = The value at which x percentage

of data lie below that value

n = Total number of observations

14. What is ANOVA?

ANS: ANOVA stands for Analysis of Variance it is a statistical analysis technique in which data sets are compared and measured to determine their significance. or 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.

15. How can ANOVA help?

ANS: The one-way ANOVA can help you know whether or not there are significant differences between the means of your independent variables (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 , and begin to learn what is driving that behaviour.