**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

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

a) 0 b) 5 c) 1 d) 10

and have units equal to standard deviations of the original data

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

10. What Is Bayes' Theorem?

Ans. Bayes' theorem named after Reverend [Thomas Bayes](https://en.wikipedia.org/wiki/Thomas_Bayes), describes the [probability](https://en.wikipedia.org/wiki/Probability) of an [event](https://en.wikipedia.org/wiki/Event_(probability_theory)), based on prior knowledge of conditions that might be related to the event.[[1]](https://en.wikipedia.org/wiki/Bayes%27_theorem#cite_note-1) For example, if the risk of developing health problems is known to increase with age, Bayes' theorem allows the risk to an individual of a known age to be assessed more accurately (by conditioning it on their age) than simply assuming that the individual is typical of the population as a whole.

{\displaystyle P(A\mid B)={\frac {P(B\mid A)P(A)}{P(B)}}} Bayes' Theorem is a way of finding a [probability](https://www.mathsisfun.com/data/probability.html) when we know certain other probabilities.

The formula is:

P(A|B) = P(A) P(B|A)/**P(B)**

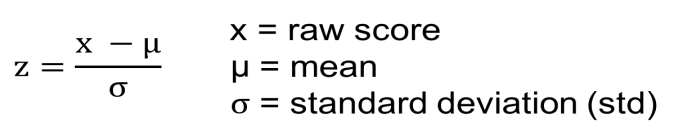
|  |  |  |
| --- | --- | --- |
| Which tells us: |  | how often A happens *given that B happens*, written P(A|B), |
| When we know: |  | how often B happens *given that A happens*, written P(B|A) |
|  |  | and how likely A is on its own, written P(A) |
|  |  | and how likely B is on its own, written P(B) |

11. What is z-score?

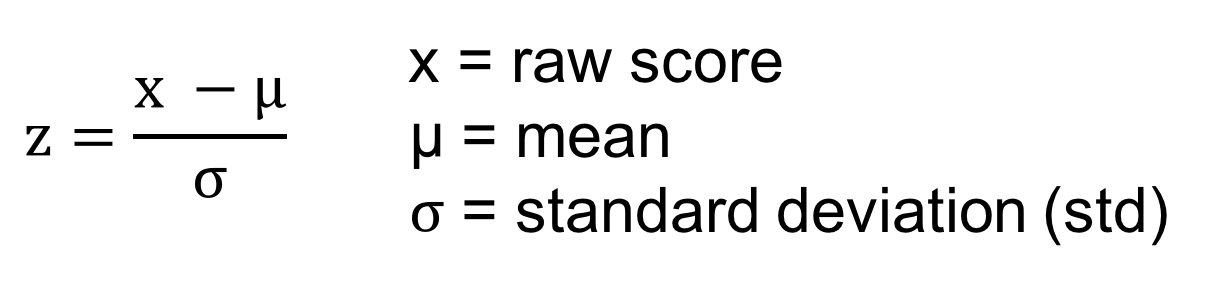
Ans. A Z-score is a numerical measurement that describes a value's relationship to the mean of a group of values. Z-score is measured in terms of [standard deviations](https://www.investopedia.com/terms/s/standarddeviation.asp) from the mean. If a Z-score is 0, it indicates that the data point's score 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.

In finance, Z-scores are measures of an observation's variability and can be used by traders to help determine market volatility. The Z-score is also sometimes known as the [Altman Z-score](https://www.investopedia.com/terms/a/altman.asp).

* 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.



12. What is t-test?

Ans. A t-test is a type of inferential [statistic](https://www.investopedia.com/terms/s/statistics.asp) used to determine if there is a significant difference between the means of two groups, which may be related in certain features. It is mostly used when the data sets, like the data set recorded as the outcome from flipping a coin 100 times, would follow a normal distribution and may have unknown variances. A t-test is used as a hypothesis testing tool, which allows testing of an [assumption](https://www.investopedia.com/ask/answers/073115/what-assumptions-are-made-when-conducting-ttest.asp) applicable to a population.

* A t-test is a type of inferential statistic used to determine if there is a significant difference between the means of two groups, which may be related in certain features.
* The t-test is one of many tests used for the purpose of [hypothesis testing](https://www.investopedia.com/terms/h/hypothesistesting.asp) in statistics.
* Calculating a t-test requires three key data values. They include the difference between the mean values from each data set (called the mean difference), the standard deviation of each group, and the number of data values of each group.
* There are several different types of t-test that can be performed depending on the data and type of analysis required

13. What is percentile?

Ans. Percentile scores for individual test takers represent how an individual test taker’s score compares to the scores of other test takers within a particular comparison group. Percentile scores range from the 1st  through 99th percentile, indicating the percentage of scores in the comparison group which are lower than the test taker’s score. For example, if your scores report says that a test taker with a [CCTST](https://www.insightassessment.com/article/california-critical-thinking-skills-test-cctst-2)Reasoning Skills Overall score of 19 is in the 67th percentile, this means that this test taker has tested better than 66% of the test takers compared to an aggregated sample of test takers like themselves. As the test administrator, you choose the comparison group when you make a test assignment for your employees or students.

14. What is ANOVA?

Ans. 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.

* 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.
* If no true variance exists between the groups, the ANOVA's F-ratio should equal close to 1.

The Formula for ANOVA is:

F= MSE/MST

​ 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?

Ans. An ANOVA test is a way to find out if survey or experiment results are [significant](https://www.statisticshowto.com/what-is-statistical-significance/). In other words, they help you to figure out if you need to [reject the null hypothesis](https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/support-or-reject-null-hypothesis/) or accept the [alternate hypothesis](https://www.statisticshowto.com/what-is-an-alternate-hypothesis/). The ANOVA test allows a comparison of more than two groups at the same time to determine whether a relationship exists between them. The result of the ANOVA formula, the F statistic (also called the F-ratio), allows for the analysis of multiple groups of data to determine the variability between samples and within samples. If no real difference exists between the tested groups, which is called the [null hypothesis](https://www.investopedia.com/terms/n/null_hypothesis.asp), the result of the ANOVA's F-ratio statistic will be close to 1. The distribution of all possible values of the F statistic is the F-distribution. This is actually a group of distribution functions, with two characteristic numbers, called the numerator [degrees of freedom](https://www.investopedia.com/terms/d/degrees-of-freedom.asp) and the denominator degrees of freedom.

Basically, you’re testing groups to see if there’s a difference between them. Examples of when you might want to test different groups:

* A group of psychiatric patients are trying three different therapies: counseling, medication and biofeedback. You want to see if one therapy is better than the others.
* A manufacturer has two different processes to make light bulbs. They want to know if one process is better than the other.
* Students from different colleges take the same exam. You want to see if one college outperforms the other.