# Statistics– WORKSHEET 6

## 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?
   1. Total Variation = Residual Variation – Regression Variation
   2. Total Variation = Residual Variation + Regression Variation
   3. Total Variation = Residual Variation \* Regression Variation
   4. All of the mentioned
2. Collection of exchangeable binary outcomes for the same covariate data are called outcomes.
   1. random
   2. direct
   3. binomial
   4. none of the mentioned
3. How many outcomes are possible with bernoulli trial?
   1. 2
   2. 3
   3. 4
   4. None of the mentioned
4. If Ho is true and we reject it, then it is called:
5. Type-I error



1. Type-II error
2. Standard error
3. Sampling error
4. Level of significance is also called:
5. Power of the test
6. Size of the test
7. Level of confidence
8. Confidence coefficient
9. The chance of rejecting a true hypothesis decreases when sample size:
10. Decreases
11. Increases
12. Both of them
13. None of them
14. Which of the following testing is concerned with making decisions using data?
15. Probability
16. Hypothesis
17. Causal
18. None of the mentioned
19. What is the purpose of multiple testing in statistical inference?
20. Minimize errors
21. Minimize false positives
22. Minimize false negatives
23. All of the mentioned
24. Normalized data is centered at and has unit 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.

1. What Is Bayes' Theorem?

In [probability theory](https://en.wikipedia.org/wiki/Probability_theory) and [statistics](https://en.wikipedia.org/wiki/Statistics), Bayes' theorem (alternatively Bayes' law or Bayes' rule), 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. 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 his age) than simply assuming that the individual is typical of the population as a whole.

One of the many applications of Bayes' theorem is [Bayesian inference](https://en.wikipedia.org/wiki/Bayesian_inference), a particular approach to [statistical inference](https://en.wikipedia.org/wiki/Statistical_inference). When applied, the probabilities involved in the theorem may have different [probability interpretations](https://en.wikipedia.org/wiki/Probability_interpretation). With [Bayesian probability](https://en.wikipedia.org/wiki/Bayesian_probability) interpretation, the theorem expresses how a degree of belief, expressed as a probability, should rationally change to account for the availability of related evidence. Bayesian inference is fundamental to [Bayesian statistics](https://en.wikipedia.org/wiki/Bayesian_statistics).

1. 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) of 1

1. What is t-test?

The t test tells you how [significant](https://www.statisticshowto.com/what-is-statistical-significance/)the differences between groups are; In other words it lets you know if those differences (measured in [means](https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/mean-median-mode/)) could have happened by chance.

A very simple example: Let’s say you have a cold and you try a naturopathic remedy. Your cold lasts a couple of days. The next time you have a cold, you buy an over-the-counter pharmaceutical and the cold lasts a week. You survey your friends and they all tell you that their colds were of a shorter duration (an average of 3 days) when they took the homeopathic remedy. What you really want to know is, are these results repeatable? A t test can tell you by comparing the means of the two groups and letting you know the probability of those results happening by chance

**Another example:** Student’s T-tests can be used in real life to compare [averages](https://calculushowto.com/average-value-of-a-function/#def). For example, a drug company may want to test a new cancer drug to find out if it improves life expectancy. In an experiment, there’s always a [control group](https://www.statisticshowto.com/control-group/) (a group who are given a placebo, or “sugar pill”). The control group may show an average life expectancy of +5 years, while the group taking the new drug might have a life expectancy of +6 years. It would seem that the drug might work. But it could be due to a fluke. To test this, researchers would use a Student’s t-test to find out if the results are repeatable for an entire population.

1. What is a percentile?

In [statistics](https://en.wikipedia.org/wiki/Statistics), a percentile (or a centile) is a type of [quantile](https://en.wikipedia.org/wiki/Quantile) which divides the given [probability distribution](https://en.wikipedia.org/wiki/Probability_distribution), or [sample](https://en.wikipedia.org/wiki/Sample_(statistics)), into 100 equal-sized intervals; this allows the data to be analyzed in terms of [percentages](https://en.wikipedia.org/wiki/Percentage). For example, the 20th percentile is the value (or score) below which 20% of the [observations](https://en.wikipedia.org/wiki/Observation_(statistics)) are found, and above which 80% are found.

The term percentile and the related term [percentile rank](https://en.wikipedia.org/wiki/Percentile_rank) are often used in the reporting of scores from [norm-referenced tests](https://en.wikipedia.org/wiki/Norm-referenced_test). 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 (carefully contrast with in the 86th percentile, which means the score is at or below the value below which 86% of the observations may be found—every score is in the 100th percentile).[[dubious](https://en.wikipedia.org/wiki/Wikipedia:Accuracy_dispute#Disputed_statement) – [discuss](https://en.wikipedia.org/wiki/Talk:Percentile#Inconsistency_-_Can_a_score_be_%22in%22_a_percentile,_or_not?)][[citation needed](https://en.wikipedia.org/wiki/Wikipedia:Citation_needed)] The 25th percentile is also known as the first [quartile](https://en.wikipedia.org/wiki/Quartile) (Q1), the 50th percentile as the [median](https://en.wikipedia.org/wiki/Median) or second quartile (Q2), and the 75th percentile as the third quartile (Q3).

1. What is ANOVA?

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/support-or-reject-null-hypothesis/) or accept the [alternate hypothesis](https://www.statisticshowto.com/what-is-an-alternate-hypothesis/).

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.

1. How can ANOVA help?

## Examples of when to use a one way ANOVA

**Situation 1:** You have a group of individuals randomly split into smaller groups and completing different tasks. For example, you might be studying the effects of tea on weight loss and form three groups: green tea, black tea, and no tea.  
**Situation 2:** Similar to situation 1, but in this case the individuals are split into groups based on an attribute they possess. For example, you might be studying leg strength of people according to weight. You could split participants into weight categories (obese, overweight and normal) and measure their leg strength on a weight machine.

### Advantages

1. MANOVA enables you to test multiple dependent variables.
2. MANOVA can protect against [Type I errors.](https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/type-i-error-type-ii-error-decision/)