## What is t-test?

The t test (also called Student’s T Test) compares two [averages](http://www.statisticshowto.com/average/)([means](http://www.statisticshowto.com/mean/)) and tells you if they are different from each other. The t test also tells you how [significant](http://www.statisticshowto.com/what-is-statistical-significance/)the differences are; In other words it lets you know if those differences could have happened by chance.

What is t-score?

The [t score](http://www.statisticshowto.com/t-score-formula/) is a ratio between the **difference between two groups and the difference within the groups**. The larger the t score, the more difference there is between groups. The smaller the t score, the more similarity there is between groups. A t score of 3 means that the groups are three times as different *from* each other as they are within each other. When you run a t test, the bigger the t-value, the more likely it is that the results are repeatable.

* A large t-score tells you that the groups are different.
* A small t-score tells you that the groups are similar.

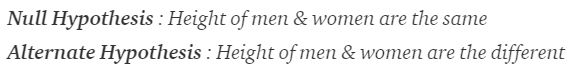
How to perform a 2 sample t-test?

Lets us say we have to test whether the height of men in the population is different from height of women in general. So we take a sample from the population and use the t-test to see if the result is significant.

**Steps:-**

1. **Determine a null and alternate hypothesis.**In general, the null hypothesis will state that the two populations being tested have no statistically significant difference. The alternate hypothesis will state that there is one present. In this example we can say that:

Image for post



**2. Collect sample data**Next step is to collect data for each population group. In our example we will collect 2 sets of data, one with the height of women and one with the height of men. The sample size should ideally be the same but it can be different. Lets say that the sample sizes are nx and ny.

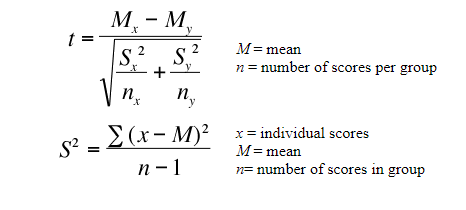
**3. Determine a confidence interval and degrees of freedom**This is what we call alpha (α). The typical value of α is 0.05. This means that there is 95% confidence that the conclusion of this test will be valid. The degree of freedom can be calculated by the the following formula:

Image for post

Image for post

**4. Calculate the t-statistic**t-statistic can be calculated using the below formula:

Image for post



where,Mx and My are the mean values of the two samples of male and female.  
Nx and Ny are the sample space of the two samples  
S is the standard deviation

**5. Calculate the critical t-value from the t distribution**To calculate the critical t-value, we need 2 things, the chosen value of alpha and the degrees of freedom. The formula of critical t-value is complex but it is fixed for a fixed pair of degree of freedom and value of alpha. We therefore use a table to calculate the critical t-value:

## One-sample t-test

It is used to compare a sample mean with a known population mean or some other meaningful, fixed value.

## Independent samples t-test

It is used to compare two means from independent groups

# **One Sample T Test: (Manually)**

We will perform 1 Sample T Test in following steps

* State Hypothesis:
* Compute T Test Statistics
* Compute Critical value using T table (For Two Tailed Test We need to find the critical cut-off value)
* Evaluate null hypothesis

n **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/).

Types of Tests.

There are two main types: one-way and two-way. Two-way tests can be with or without replication.

* One-way ANOVA between groups: used when you want to test **two groups** to see if there’s a difference between them.
* Two way ANOVA without replication: used when you have **one group** and you’re**double-testing**that same group. For example, you’re testing one set of individuals before and after they take a medication to see if it works or not.
* Two way ANOVA with replication: **Two groups**, and the members of those groups are **doing more than one thing**. For example, two groups of patients from different hospitals trying two different therapies.

[Back to Top](https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/anova/#top)

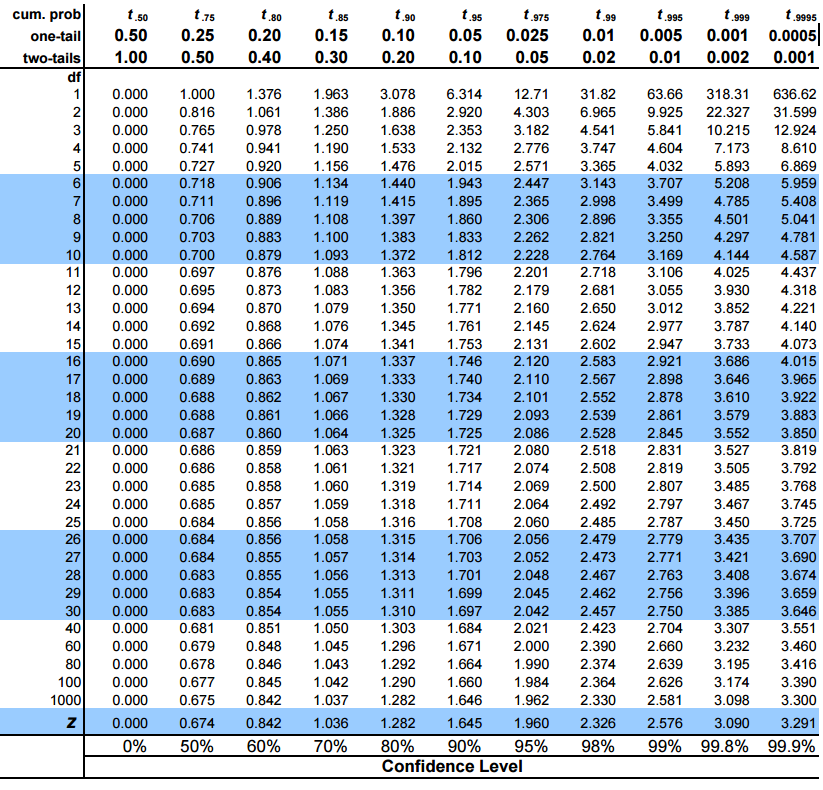
One Way ANOVA

A one way ANOVA is used to compare two means from two independent (unrelated) groups using the [F-distribution](https://www.statisticshowto.com/probability-and-statistics/f-statistic-value-test/#Fdist). The [null hypothesis](https://www.statisticshowto.com/probability-and-statistics/null-hypothesis/) for the test is that the two [means](https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/mean-median-mode/#mean)are equal. Therefore, a [significant](https://www.statisticshowto.com/what-is-statistical-significance/)result means that the two means are unequal.

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.

Image for post



In python, rather than looking up in the table we will use a function from the sciPy package. (I promise u, its the only time we will use it!)

**6. Compare the critical t-values with the calculated t statistic**If the calculated t-statistic is greater than the critical t-value, the test concludes that there is a statistically significant difference between the two populations. Therefore, you reject the null hypothesis that there is no statistically significant difference between the two populations.

In any other case, there is no statistically significant difference between the two populations. The test fails to reject the null hypothesis and we accept the alternate hypothesis which says that the height of men and women are statistically different.

What is an Anova test used for?

The one-way **analysis** of variance (**ANOVA**) is **used to** determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups

# How to Calculate Confidence Intervals in Python

A **confidence interval for a mean**is a range of values that is likely to contain a population mean with a certain level of confidence.

It is calculated as:

**Confidence Interval = x  +/-  t\*(s/√n)**

where:

* **x:**sample mean
* **t:**t-value that corresponds to the confidence level
* **s:**sample standard deviation
* **n:**sample size

This tutorial explains how to calculate confidence intervals in Python.

### **Confidence Intervals Using the t Distribution**

If we’re working with a small sample (n <30), we can use the [t.interval() function](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.t.html) from the scipy.stats library to calculate a confidence interval for a population mean.

The following example shows how to calculate a confidence interval for the true population mean height (in inches) of a certain species of plant, using a sample of 15 plants:

**import numpy as np**

**import scipy.stats as st**

**#define sample data**

**data = [12, 12, 13, 13, 15, 16, 17, 22, 23, 25, 26, 27, 28, 28, 29]**

**#create 95% confidence interval for population mean weight**

**st.t.interval(alpha=0.95, df=len(data)-1, loc=np.mean(data), scale=st.sem(data))**

**(16.758, 24.042)**

The 95% confidence interval for the true population mean height is **(16.758, 24.042)**.

You’ll notice that the larger the confidence level, the wider the confidence interval. For example, here’s how to calculate a 99% C.I. for the exact same data:

**#create 99% confidence interval for same sample**

**st.t.interval(alpha=0.99, df=len(data)-1, loc=np.mean(data), scale=st.sem(data))**

**(15.348, 25.455)**

**Project Report:**

Please add Results & Conclusion section in your report along with an abstract section. These sections will complete your report. Your report will now contain abstract, introduction, literature review, data collection, methodology & experiment and results & conclusion section. You are advised to use overleaf for write-up. Please refer to project guidelines for formatting instructions and further guidance.

**Presentation Slides:**

Your presentation should at least cover following content  
- Title (Project title and Group members) 1 min

- Introduction (What is project, what issues does it solve, what was the motivation behind project etc) 2-3 min  
- Methodology (Elaborate the techniques that you used) 4-6 min  
- Data Description (size of data, how was it collected, some insights of the data etc) 2-3 min  
- Experiments and Results (what experiments were performed, what were the results) 3-5 min  
- Conclusion 2-4 min