

T-Test

Data Science | CCDATSCL

The **t-test** tells you how significant the differences between group means are.

It lets you know if those differences in means could have **happened by chance**.

The **t-test** is usually used when data sets follow a normal distribution but you don't know the population variance.

T-Score

The **t-score** is a ratio between the difference between two groups and the difference within the groups.

Larger t-scores = more difference between groups.

Smaller t-scores = more similarity between groups.

t-scores and p-values

Every **t-score** has a **p-value** to go with it.

A **p-value** from a t-test is the probability that the results from your sample data **occurred by chance**.

Low p-values indicate your data **did not occur by chance**.

For example, a **p-value** of .01 means that there is only a **1% probability** that the results from an experiment **happened by chance**.

Types of t-test

- An **independent samples t-test**
 - compares the means for two groups.

$$t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

where,

'x' bar is the mean of the sample,

μ is the assumed mean,

σ is the standard deviation

and n is the number of observations

- A **paired sample t-test**
 - compares means from the same group at different times (say, one year apart).

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

where

- t = Student's t-test
- \bar{x}_1 = mean of first group
- \bar{x}_2 = mean of second group
- s_1 = standard deviation of group 1
- s_2 = standard deviation of group 1
- n_1 = number of observations in group 1
- n_2 = number of observations in group 2

- A **one sample t-test**

- tests the mean of a single group against a known mean.

$$t = \frac{\sum(x_1 - x_2)}{\frac{s}{\sqrt{n}}}$$

where

t = Student's t-test

$x_1 - x_2$ = Difference mean of the pairs

s = standard deviation

n = sample size

Paired sample t-test example

Calculate a t-test for the following data of the number of times people prefer coffee or tea in five time intervals.

Coffee	Tea
4	3
5	8
7	6
6	4
9	7

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

$\bar{x}_1 = 31/5 = 6.2$
 $\bar{x}_2 = 28/5 = 5.6$
 $\sum(x_1 - \bar{x}_1)^2 = 14.8$
 $\sum(x_2 - \bar{x}_2)^2 = 17.2$
 $s_1 = 14.8/4 = 3.7$
 $s_2 = 17.2/4 = 4.3$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

Applying the known values in the t-test formula, we get
 $t = \frac{6.2 - 5.6}{\sqrt{\left(\frac{3.7^2}{5} + \frac{4.3^2}{5}\right)}}$
 $= \frac{0.6}{\sqrt{1.6}} = 0.6/1.26 = 0.47$
 $t = 0.47$

Independent sample t-test example

A company wants to improve its sales. The previous sales data indicated that the average sale of 25 salesmen was **\$50 per transaction**.

After training, the recent data showed an average sale of **\$80 per transaction**. If the **standard deviation is \$15**, find the t-score.

Has the training provided improved the sales?

Solution:

H_0 accepted hypothesis: the population mean = the claimed value $\Rightarrow \mu = \mu_0$

H_0 alternate hypothesis: the population mean not equal to the claimed value $\Rightarrow \mu \neq \mu_0$

$$\text{t-test formula for independent test is } t = \frac{m - \mu}{\frac{s}{\sqrt{n}}}$$

Degree of freedom	0.50	0.25	0.20	0.15	0.10	0.05
24	0.000	0.685	0.857	1.059	1.318	1.711
25	0.000	0.684	0.856	1.058	1.316	1.708
26	0.000	0.684	0.856	1.058	1.315	1.706

Mean sale = 80, $\mu = 50$, $s = 15$ and $n = 25$

substituting the values, we get $t = (80-50)/(15/\sqrt{25})$

$$t = (30 \times 5)/10 = 10$$

looking at the t-table we find $10 > 1.711$. (i.e. CV for $\alpha = 0.05$). \therefore the accepted hypothesis is not true. Thus we conclude that the training boosted the sales.