**Two sample t-test**

The two-sample t-test is used to determine if two population means are equal. A common application is to test if a new process or treatment is superior to a current process or treatment.

There are several variations on this test.

* The data may either be paired or not paired. By paired, we mean that there is a one-to-one correspondence between the values in the two samples. That is, if X1, X2, ..., Xn and Y1, Y2, ... , Yn are the two samples, then Xi corresponds to Yi. For paired samples, the difference Xi - Yi is usually calculated. For unpaired samples, the sample sizes for the two samples may or may not be equal. The formulas for paired data are somewhat simpler than the formulas for unpaired data.
* The variances of the two samples may be assumed to be equal or unequal. Equal variances yield somewhat simpler formulas, although with computers this is no longer a significant issue.
* In some applications, you may want to adopt a new process or treatment only if it exceeds the current treatment by some threshold. In this case, we can state the null hypothesis in the form that the difference between the two populations means is equal to some constant μ1−μ2=d0 where the constant is the desired threshold.

Definition: The two-sample t-test for unpaired data is defined as:

H0: µ1 = µ2

H1: µ1 ≠ µ2

Test Statistic: T=Y1¯−Y2¯ / √ (S12 / N1 + S22 / N2)

where N1 and N2 are the sample sizes, Y1¯ and Y2¯ are the sample means, and S12and S22 are the sample variances.

Where

Sp2=(N1−1)s12+(N2−1)s22 / (N1+N2−2)

|  |  |
| --- | --- |
| Significance Level: | α. |
| Critical Region: | Reject the null hypothesis that the two means are equal if  |T| > t1-α/2,ν  where t1-α/2,ν is the [critical value](https://www.itl.nist.gov/div898/handbook/eda/section3/eda3672.htm) of the [t distribution](https://www.itl.nist.gov/div898/handbook/eda/section3/eda3664.htm) with ν degrees of freedom where  υ=(s21/N1+s22/N2)2(s21/N1)2/(N1−1) +(s22/N2)2/(N2−1)  If equal variances are assumed, then ν = N1 + N2 - 2 |

Example:

The following two-sample t-test was generated for the [“2](https://www.itl.nist.gov/div898/handbook/eda/section3/eda3531.htm) sample” data set. The data set contains miles per gallon for U.S. cars (sample 1) and for Japanese cars (sample 2); the summary statistics for each sample are shown below.

SAMPLE 1:

NUMBER OF OBSERVATIONS = 249

MEAN = 20.14458

STANDARD DEVIATION = 6.41470

STANDARD ERROR OF THE MEAN = 0.40652

SAMPLE 2:

NUMBER OF OBSERVATIONS = 79

MEAN = 30.48101

STANDARD DEVIATION = 6.10771

STANDARD ERROR OF THE MEAN = 0.68717

We are testing the hypothesis that the population means are equal for the two samples. We assume that the variances for the two samples are equal.

H0: μ1 = μ2

Ha: μ1 ≠ μ2

Test statistic: T = -12.62059

Pooled standard deviation: sp = 6.34260

Degrees of freedom: ν = 326

Significance level: α = 0.05

Critical value (upper tail): t1-α/2,ν = 1.9673

Critical region: Reject H0 if |T| > 1.9673

The absolute value of the test statistic for our example, 12.62059, is greater than the critical value of 1.9673, so we reject the null hypothesis and conclude that the two population means are different at the 0.05 significance level.

In general, there are three possible alternative hypotheses and rejection regions for the one-sample t-test:

|  |  |
| --- | --- |
| **Alternative Hypothesis** | **Rejection Region** |
| Ha: μ1 ≠ μ2 | |*T*| > *t*1-α/2,ν |
| Ha: μ1 > μ2 | *T* > *t*1-α,ν |
| Ha: μ1 < μ2 | *T* < *t*α,ν |

For our two-tailed t-test, the critical value is t1-α/2,ν = 1.9673, where α = 0.05 and ν = 326. If we were to perform an upper, one-tailed test, the critical value would be t1-α,ν = 1.6495. The rejection regions for three possible alternative hypotheses using our example data are shown below.

