Difference of means test ($t$-test)

The significance of differences between a sample mean, and a (perhaps hypothetical) “true” mean, or between two sample means, can be assessed using the $t$-statistic calculated as part of the $t$-test. The $t$-statistic may be thought of as a scaled difference between the two means, where the absolute difference between means is rescaled using an estimate of the variability of the means. The reference distribution for the $t$-statistic is the $t$-distribution, shape of which varies slightly as a function of sample size for $n < 30$, and strongly resembles the normal distribution in its shape.

The one-sample $t$-statistic is

$$t = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

where $\bar{X}$ is the sample mean, $\mu$ is the true or hypothetized mean, $s$ is the sample standard deviation, and $n$ is the sample size. The specific $t$-distribution that serves as the reference distribution for the $t$-statistic depends on the “degrees of freedom” ($df$) of the test statistic. For this one-sample test, $df = n - 1$.

The two-sample $t$-statistic is

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}_1-\bar{X}_2}}$$

Where $\bar{X}_1$ and $\bar{X}_2$ are the means of the two samples, and $\sigma_{\bar{X}_1-\bar{X}_2}$ is a measure of the variability of the differences between the sample means. When the population variances are assumed to be equal, a pooled variance estimate is calculated as the weighted average (by sample size) of the two sample variances

$$\sigma_{pooled}^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

and then

$$\sigma_{\bar{X}_1-\bar{X}_2} = (\sigma_{pooled}^2)^{0.5}((1/n_1) + (1/n_2))^{0.5}.$$ 

The degrees of freedom that define the specific $t$-distribution for this straightforward
case is given by
\[ df = n_1 + n_2 - 2. \]

If the population variances are not assumed to be equal, the separate sample variances are used as an estimate of \( \sigma_{\bar{X}_1 - \bar{X}_2} \):
\[ \sigma_{\bar{X}_1 - \bar{X}_2} = ((s_1^2 / n_1) + (s_2^2 / n_2))^{0.5}. \]

In this more complicated case, the degrees of freedom is given by
\[ df = \frac{(s_1^2 / n_1) + (s_2^2 / n_2)}{\left[ \frac{s_1^2 / n_1}{n_1 - 1} + \frac{s_2^2 / n_2}{n_2 - 1} \right]}. \]