Inference from Small Samples

The t-distribution is used to make inference about a population mean μ if

1. The population from which the sample is drawn is (approximately) normally distributed.
2. The sample size is small (i.e., $n < 40$).
3. The population standard deviation σ is not known.

The degrees of freedom are $df = n - 1$.

A $(1-\boldsymbol{\alpha})100\%$ **Small-Sample Confidence Interval for** $\boldsymbol{\mu}$ is of the form

$$\overline{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}.$$

Small-Sample Test about a Population Mean

Null Hypothesis: $H_0\colon \mu = \mu_0$

Test Statistic: $t = \dfrac{\overline{x} - \mu_0}{s/\sqrt{n}}$

Alternative Hypothesis	Rejection region at the significance level α
$H_a\colon \mu > \mu_0$	$t > t_\alpha$ (upper-tailed test)
$H_a\colon \mu < \mu_0$	$t < -t_\alpha$ (lower-tailed test)
$H_a\colon \mu \neq \mu_0$	$t > t_{\alpha/2}$ or $t < -t_{\alpha/2}$ (two-tailed test)

The t-distribution is used to make inference about $\mu_1 - \mu_2$ if

1. The two populations from which the samples are drawn are (approximately) normally distributed.
2. The samples are small (i.e., $n_1 < 40$ and $n_2 < 40$) and independent.
3. The standard deviation σ_1 and σ_2 of the two populations are unknown but are equal.

The degrees of freedom are $df = n_1 + n_2 - 2$.

A $(\mathbf{1}-\boldsymbol{\alpha})\mathbf{100\%}$ **Small-Sample Confidence Interval for** $\boldsymbol{\mu_1} - \boldsymbol{\mu_2}$ is of the form

$$(\overline{x}_1 - \overline{x}_2) \pm t_{\alpha/2}\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

where $s^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$ is the pooled variance for the two samples.

Small-Sample Test about a Difference Between two Population Means

Null Hypothesis: $H_0\colon \mu_1 - \mu_2 = D_0$

Test Statistic: $t = \dfrac{\overline{x}_1 - \overline{x}_2 - D_0}{\sqrt{s^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$, where s^2 is the pooled variance.

Alternative Hypothesis	Rejection region at the significance level α
$H_a\colon \mu_1 - \mu_2 > D_0$	$t > t_\alpha$ (upper-tailed test)
$H_a\colon \mu_1 - \mu_2 < D_0$	$t < -t_\alpha$ (lower-tailed test)
$H_a\colon \mu_1 - \mu_2 \neq D_0$	$t > t_{\alpha/2}$ or $t < -t_{\alpha/2}$ (two-tailed test)

If the two populations do not have equal standard deviations, we do not use the pooled variance. Instead we use the following formulas for standard error and degrees of freedom.

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad \text{and} \quad df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$