Independent Two-sample t-test

1. Assumptions

- **Experimental Design:** The sample forms two independent treatment groups.
- **Null Hypothesis:** The population means of the two treatment groups are not significantly different from each other.
- **Population Distribution:** Each treatment group is approximately normally distributed.
- **Sample Size:** Arbitrary for both treatment groups.
- **Variances:** The variances are assumed to be the same for both treatment groups. If the variances are not equal, finding a suitable test of hypothesis is called the Behrens-Fisher Problem. There are several proposed solutions to this problem. We will not discuss the details here.

2. Inputs for the independent two-sample t-test

- Sample sizes of treatment groups: n_1 and n_2
- Sample means of treatment groups: $\bar{x}_1$ and $\bar{x}_1$
- Standard deviations of treatment groups : s_1 and s_2
- Pooled standard deviation: $s_{\text{pooled}} = \sqrt{\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$
- Standard error of $\mu_2 - \mu_1$: $\text{SE}_{\text{diff}} = \sqrt{s^2_{\text{pooled}}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$
- Null hypothesis: $\mu_1 = \mu_2$
- The level of the test: α

3. Five Steps for Performing the Test of Hypothesis

1. State null and alternative hypotheses:

$$H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 \neq \mu_2$$

2. Compute test statistic:

$$t = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)}{\mathrm{SE}_{\mathrm{diff}}},$$

   assuming the null hypothesis.

3. Compute $100(1 - \alpha)\%$ confidence interval I for z.

4. If $t \in I$, accept H_0; if $t \notin I$, reject H_1 and accept H_0.

5. Compute p-value.

4. Discussion

For the one-sample t-test, the sample variance forms an unbiased estimator of the population variance because

$$E(s_x^2) = \sigma_x^2.$$

Similarly, for the independent two-sample t-test assuming equal variances $\sigma_x^1 = \sigma_x^2$, it can be shown that the pooled sample variance forms an unbiased estimator of the sample variance.
It can also be shown that the t statistic

$$t = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)}{\mathrm{SE}_{\mathrm{diff}}}$$

has a t distribution with $n_1 + n_2 - 2$ degrees of freedom, so we can use the t tables to find a confidence interval for the test statistic.

4. A Sample Problem

Mendenhall and Sincich: p. 55: To measure whether a new method for teaching reading works better than the standard method, 22 children that have had difficulties learning how to read are selected. 10 children are are taught reading using the new method; 12 children are taught using the standard method. All children are then given a standardized test. Perform a test of the null hypothesis that there is no difference between the two teaching methods.

Solution: Here are the inputs:

$$\begin{aligned} n_1 &= 12 \qquad n_2 = 10 \qquad \bar{x}_1 = 72.333 \qquad \bar{x}_2 = 76.400 \\ s_1 &= 6.3237 \qquad s_2 = 5.8348 \qquad \alpha = 0.05 \end{aligned}$$

We compute the pooled standard deviation

$$\begin{aligned} s_{\text{pooled}} &= \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \\ &= \sqrt{\frac{(12-1)6.3237^2 + (10-1)5.8348^2}{12+10-2}} \\ &= 6.108 \end{aligned}$$

and the standard error for the difference:

$$\text{SE}_{\text{diff}} = \sqrt{s_{\text{pooled}}^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = \sqrt{6.1085^2\left(\frac{1}{12}+\frac{1}{10}\right)} = 2.616$$

The five steps of the t-test:

1. State the null and alternative hypotheses:

$$H_0 : \mu_1 = \mu_2, \qquad H_1 : \mu_1 \neq \mu_2$$

2. Compute the test statistic:

$$t = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)}{\text{SE}_{\text{diff}}} = \frac{(76.40 - 72.33) - 0}{2.616} = 1.55$$

3. Compute a $100(1-\alpha)\%$ confidence interval I: use the t table with $n_1 + n_2 - 2 = 12 + 10 - 2 = 20$ degrees of freedom to show that $[-2.086, 2.086]$ is a 95% confidence interval for t.

4. Determine whether to accept or reject H_0: $-1.55 \in [-2.086, 2.086]$, so accept H_0.

5. Use SAS or R to determine the p-value: 0.1634.