Paired t-test

Comparing two populations

When comparing two populations, we could be interested in:

- Comparing the *mean* of some property of each population, or
- Comparing the *proportion* of some property in each population.

For example, you may be comparing:

- the running time of software on two different hardware setups;
- the proportion of users who click an ad on one site versus another;
- the effectiveness of a medication versus a placebo;
- the difference in performance of two classes on the same test.

Even if you took two samples from the *same* population, you will almost certainly get different means/proportions. Thus, the relevant question to address is: *is the difference between the two outcomes statistically significant?*

Paired versus unpaired

Consider the following two scenarios:

Scenario 1. Fifteen people were given alcohol, and fifteen people were not. The reaction speed of each group was tested and compared.

Scenario 2. Fifteen people had their reaction speed tested before being given alcohol, and then tested again after being given alcohol.

The difference between the two is:

- In scenario 1, the two samples are entirely independent of one another. The two samples are *unpaired*.
- In scenario 2, the two samples are dependent on one another. Each (before, after) reading is a pair, so the two samples are *paired*.

A paired sample is more informative, so applying the right test can ensure we get the most out of the data.

Paired t-test

The paired t-test is actually just an application of the one-sample t-test. The procedure works by computing the difference for each pair, and then a one-sample t-test is performed on the resulting differences.

A stipulation: it is assumed that the differences follow a normal distribution or that the sample size is large.

Example

Suppose that 5 students were given a diagnostic test before completing a short online course. After completing the module, they then attempted the diagnostic test again. The scores for each student are shown in the table below.

Student	1	2	3	4	5
Score before course	18	23	21	23	15
Score after course	22	22	23	27	20
Difference	$+4$	-1	$+2$	$+4$	$+5$

To a 5% level of significance, do the students perform better after taking the course?

Reminder: df for a one-sample t-test is $n - 1$, where n is the sample size.

Unpaired t-test

Unpaired t-test

The unpaired t-test is more elaborate, because it needs to account for the potential difference in variance for the two populations.
To test the following hypotheses:

$$H_0\colon \mu_1 - \mu_2 = 0 \text{ versus } H_1\colon \mu_1 - \mu_2 \neq 0,$$

we need to compute the test statistic

$$t = \frac{\overline{x} - \overline{y}}{\mathrm{SE}},$$

where $\overline{x}$ and $\overline{y}$ denote the sample means for each population, and SE is the standard error.

A stipulation: it is assumed that each population is normally distributed or the sample sizes are large.

Important! The standard error and degrees of freedom for the unpaired t-test are different compared to the other t-tests we have seen.

Standard error

For two samples of size n_1 and n_2, with sample standard deviations s_1 and s_2, the standard error is given by

$$\mathsf{SE} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

This is the most general form of the unpaired t-test. There are other unpaired t-tests considered in the literature, but they often have unreasonable assumptions such as assuming equal variances.

Degrees of freedom and p-values

The p-value depends on the appropriate choice of t-distribution. For an unpaired t-test, the following degrees of freedom formula is used:

$$\mathrm{df} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}.$$

In STM4PSD, you will *not* need to calculate this yourself.

Then, the p-value is calculated depending on the form of the hypotheses.

With $T \sim t_{\mathrm{df}}$:

- If H_1 is of the form $\mu_1 - \mu_2 \neq 0$, calculate $2P(T \geqslant |t|)$.
- If H_1 is of the form $\mu_1 - \mu_2 > 0$, calculate $P(T \geqslant t)$.
- If H_1 is of the form $\mu_1 - \mu_2 < 0$, calculate $P(T \leqslant t)$.

Example

The shelf life of two competing products (product A and product B) is being compared. To test this, 39 instances of product A and 35 instances of product B were randomly selected for testing. The sample mean for product A was 77 days, and the sample mean for product B was 72 days. The sample standard deviation for products A and B was 11.9 and and 4.2, respectively. To a 5% level of significance, is there a difference in shelf life for the two products?

Confidence intervals

Confidence intervals are calculated by using a t-distribution with the appropriate degrees of freedom.

A 95% confidence interval for the previous example:

Interpretation: the confidence interval is the estimate of the *difference* between the two means.