1.08 Tests for two groups: Two dependent means

In this video we'll discuss how to compare two dependent groups on a quantitative variable, using a **paired samples t-test**. We'll also see how to calculate the corresponding confidence interval.

We use a paired samples t-test or confidence interval if we have a **quantitative response variable** and a **binary independent variable** that distinguishes two paired samples. An example of a research question is: Does the mean score on a happiness scale become lower after people have children versus the mean score before they had children?

Assumptions

The samples are **dependent** or **paired,** which means the samples consist of repeated measurements obtained from the same participants, or different measurements of paired participants. In both cases the difference scores of the paired observations are used in the paired t-test.
These difference scores are assumed to be distributed normally. The paired t-test is robust against violation of this assumption for large samples due to the Central Limit Theorem. It's also robust for small samples when using a two-sided test. Normality is important only if the number of difference scores is small. In this course we will consider a sample of fewer than thirty pairs small. Remember, these rule of thumb are somewhat arbitrary!

Statistical hypotheses

The statistical hypotheses are expressed in terms of the mean difference score μ_d in the population. If there is no difference in the population, we expect μ_d to equal zero. This is the null hypothesis. Possible alternative hypotheses are that mean population difference score is unequal to zero, or that it is greater than zero or smaller than zero. The interpretation depends on how you calculate the difference scores, subtracting group 2 from group 1, or the other way around.

Test statistic

The test statistic t equals the mean difference score minus the expected value under the null hypothesis, which is zero, divided by the standard error, which equals the standard deviation of the difference scores, divided by the square root of the sample size - the number of paired observations.

Test statistic distribution and p value

The test statistic follows a Student t-distribution, with degrees of freedom equal to the sample size minus one. With the test statistic and degrees of freedom we can calculate or look up the accompanying one-sided or two-sided p-value and compare it to the predetermined significance level. Based on this outcome we reject or fail to reject the null hypothesis.

Example

Suppose I want to test whether a raw meat diet is healthier for cats than regular canned food. Instead of comparing two separate groups of cats that are exposed to different diets I can expose one group of cats to both diets. I randomly assign half of the sample to a raw meat diet for two months and the other half to a canned food diet. Cat health is measured on a scale between zero and ten by a veterinarian based on lab results. Now I switch all the cats to the opposite diet. After two months I measure their health again. I now have two health measurements for each cat, one after eating raw and one after eating canned food for two months. I expect the raw diet to result in better health scores compared to the canned food.

First I have to transform the raw scores into difference scores and check whether the scores are distributed normally, by looking at the histograms of the difference scores. The distribution looks normal, but even if it hadn't, the sample size is large enough to ensure a robust one-sided test. The null hypothesis states that the mean difference score is zero. My one-side alternative hypothesis is that the difference will be larger than zero, since I subtracted the canned food health scores from the raw meat health scores. I'll set the significance level to 0.05.

The test statistic value is -0.12 divided by 1.91 divided by the square root of 77. This equals -0.55. Unexpectedly the value is negative and falls in the left tail. The degrees of freedom are 76. If I use a table to look up the p-value I find that it lies somewhere between 0.50 and 0.90. If I calculate the p-value with statistical software, I find a value of 0.71.
This is much larger than the significance level of 0.05, so I *can't* reject the null hypothesis. I *can't* conclude that the mean health score in the population is higher for cats fed on raw meat, compared with cats fed on canned food.

Confidence interval

We can calculate a confidence interval for the difference score using this formula: the mean difference score plus and minus t times the standard error. Plus and minus t equals the t values associated with the required confidence level and the degrees of freedom equaling the sample size minus one. With 76 degrees of freedom and a confidence level of 95%, the values are minus and plus 2.00, based on a table look-up, rounding down to the nearest degrees of freedom of 60. The standard error is calculated the same way as before. Also, remember that we need to meet the same assumptions required for a two-sided t-test.

The confidence interval for our example data is -0.12 plus and minus 2.00 times 0.22. This results in a confidence interval that ranges from -0.55 to 0.31. Since the value zero - no difference in the means - lies inside the interval, zero is considered a plausible value. This means that if we had performed a two-sided test it would have also been non-significant.