1.07 Tests for two groups: Two dependent proportions

In this video we'll see how we can compare two paired groups on a binary variable, using a z-test for two dependent proportions, called McNemar's test.
We use McNemar z-test if we have a **binary response variable** and a **binary independent variable** that distinguished between two related or paired samples.
The samples are either dependent because they consist of the same cases - measured at two different times, or they consist of matched pairs. These could be naturally occurring pairs, such as twins or couples in a romantic relationship. The pairs can also be *created* based on similarity on relevant background variables such as age, sex and education.

An example of a research question is: Do people smoke less after they've been exposed to an aggressive anti-smoking campaign? Here we measure the same group of people twice. We determine if they smoke before and after the campaign and we see if the proportion has dropped.

Here's another example: Suppose I want to compare the proportion of cats with urinary problems in a sample of cats on a raw meat diet and a sample fed with canned food. Instead of using independent samples I create matched pairs of cats, based on their age, their sex, whether they were sterilized and the number of other pets in the household. Essentially I'm treating the two cats in a pair as if they are interchangeable, except for their diet.

To see how we can test hypotheses about these pairs we need to look at all possible outcomes per pair. Here they are: Both cats have no urinary problems; only the cat with a canned diet has problems; only the cat with a raw diet has problems, or both have problems. If a raw diet is healthier we expect more instances of the second combination, with cats on raw food being healthy and their counterpart on canned food having problems. We also expect *less* instances of combination three, with cats on raw food having urinary problems and their counterpart being healthy. The other combinations don't matter; they don't contribute to the difference between the diets.
To determine the effect of diet in these dependent samples we just have to look for balance or imbalance in the inconsistent combinations. To make it easier we put the combinations in a two-by-two table with healthy and problem cats on a raw diet in the rows and healthy and problem cats on canned food in the columns. In the diagonals we see the

irrelevant, consistent combinations. The off-diagonals represent the interesting, inconsistent combinations. With this two-by-two table in place we can now turn to the test itself.

Assumptions

To perform McNemar's test we need to have a **sufficient number of observations**. For one-sided tests the sum of the inconsistent cases should be at least thirty. Two-sided tests work well, even for small samples. No additional assumptions are required.

Statistical hypotheses

The null hypothesis states that the population proportion in the first group - or at the first measurement - is the same as in the second group - or at the second measurement: H_0: p_1 - p_2 = 0. This corresponds to an equal number of inconsistent cases. Possible alternative hypotheses are that the proportions are unequal, or that the second proportion will be greater or smaller than the first. The interpretation depends on how you define the groups and combinations.

Test statistic

The formula for the test statistic z is very simple: $z = \frac{n_{01} - n_{10}}{\sqrt{n_{01} + n_{10}}}$; it equals the difference between the two off-diagonal elements, the inconsistent cases, divided by the square root of the sum of these elements.

Test statistic distribution and p value

The resulting test statistic has a standard normal distribution. We calculate or look up the accompanying one-sided or two-sided p-value, depending on the alternative hypothesis. If the p-value is smaller or equal to the predetermined significance level we reject the null hypothesis. If it's larger we fail to reject the null hypothesis.

Example

In our example we have enough inconsistent pairs, the sum is 51. The null hypothesis states that the proportion of urinary problems will be the same for raw and can-fed cats. Suppose we believe raw-fed cats will be healthier and have less urinary problems. This is our alternative hypothesis. If we put the raw diet group in the rows, we would expect the top right cell to be larger than the left bottom cell. So we would expect the difference to be positive. We'll set the significance level to 0.05.
When we calculate the actual z-value we find a value of 2.38. As expected, the value is positive and falls in the right tail of the distribution. If we look up the p-value in a table, or calculate it with statistical software,

we find a value of 0.01. This value is smaller than the significance level of 0.05, so we can reject the null hypothesis in favor of the hypothesis that the proportion of cats with urinary problems is lower when we compare them with cats on a canned food diet.

We already came to this conclusion in an earlier video based on the same example with two *in*dependent samples. So why go through the hassle of finding matching pairs or measuring the same cats again after changing their diet, which will take twice as long? Well by treating the pairs as interchangeable or by measuring the same cases twice, we can eliminate a lot of random error in the samples due to individual differences and irrelevant background variables. This results in smaller standard errors, which in turn results in a larger probability to reject the null hypothesis.