

## 1.05 Tests for two groups: Two independent proportions

In this video we'll see how we can compare two independent groups on a categorical, binary variable, using a z-test for two proportions. We'll also see how to calculate the corresponding confidence interval.

We use the z-test or confidence interval for two independent proportions if we have a **binary response variable** and **binary independent variable** that distinguishes two independent groups or samples. Examples of research questions could be: Are men more often smokers than women? Or: Is the proportion of people with epilepsy larger in Europe than in South-America?

### Assumptions

To perform a z-test for two proportions, the samples need to be **independent**; the cases should be assigned to the groups randomly in an experimental design or drawn randomly from the population in a non-experimental design. This is to ensure there's no relation or dependence between cases in the different groups.

We also need to have **sufficient numbers of observations** in each sample: for one-sided tests this means at least ten negative cases and ten positive cases in each sample. For two-sided tests and confidence intervals we need at least five negative and five positive cases in each sample. If this requirement isn't met, we can use **Fisher's exact** z-test instead.

### Statistical hypotheses

The statistical hypotheses are expressed in terms of the difference between the two population proportions. If the populations are the same, this difference will be zero. This is the null hypothesis. Possible alternative hypotheses are that the proportions are unequal and the difference is thereby unequal to zero, or that the difference will be greater than zero or smaller than zero.

### Test statistic

The test statistic  $z$  equals the estimate minus the expected value under the null hypothesis, divided by the standard error under the null hypothesis. The population difference under the null equals zero, so this simplifies to this formula.  $p$ -hat is the pooled proportion assuming the proportions are the same. The best estimate of this 'common' proportion is the pooled proportion, the average proportion correcting for sample size. It equals the

<sup>1</sup>total number of positive cases in both groups divided by the total number of cases.

### Test statistic distribution and p value

The test statistic has a standard normal distribution. One-sided or two-sided p-values are easily determined using software or a table. 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

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. I expect the raw meat group to have less urinary problems. Here are the data. The proportion of cats with urinary problems in the two samples are 0.10 and 0.18.

The number of observations are sufficient for two-sided and one-sided tests. These are imaginary data; we'll assume they're independent. The null hypothesis states that the difference in proportions will be zero, my one-side alternative hypothesis is that the difference will be smaller than zero if I subtract the canned proportion from the raw proportion. Before I calculate the test statistic and determine the p-value, we'll set the significance level to 0.05.

The test statistic value is 0.10 minus 0.18 divided by the square root of (0.14 times one minus 0.14 - where 0.14 equals all 42 cats with urinary problems divided by the total of 298 cats) times one divided by 150 plus one divided by 148. This equals -2.04. The value is negative; we expected a negative value so we determine the p-value by calculating or looking up the area under the curve in the left tail. The value is 0.02. 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 for cats eating raw meat than for cats eating canned food. Of course, to say this is strong evidence of a causal relation between cat diet and urinary problems requires a detailed analysis of the quality of the research design, measurement instruments and sampling methods.

### Confidence interval

We can calculate a confidence interval for the difference in proportions using this formula: the difference in sample proportions plus and minus z times the standard error. Plus and minus z equals the z values associated with the required confidence level, so for example - 1.96 and + 1.96 for a 95%

---

<sup>1</sup> show  $p\text{-hat} = \text{total positive cases} / \text{total sample size} =$

confidence interval. The standard error of the difference score equals the square root of the sum of the group proportion times its complement - the group's variance, divided by its sample size, for each group.

Notice that we were *not* using the pooled proportion, since we're not assuming a null hypothesis value. Also, remember that the same assumptions as for a two-sided null hypothesis test are required.

The confidence interval for our example data is 0.10 minus 0.18 plus and minus 1.96 times the standard error, which equals 0.04. This results in a confidence interval that ranges from -0.161 to -0.004. This corresponds to a significant two-sided test, since the value zero - no difference in the proportions - lies outside the interval, which means it's an implausible value. The interval does lie very close to zero though!

One final way to compare two independent groups on a binary response variable is to express the proportions as a ratio, referred to as the **relative risk**. This is especially useful if the proportions are very small and the relevance of the absolute difference is hard to evaluate.

Suppose we compare the proportion of people who suffer a heart attack in two groups of healthy middle-aged people. In one group that exercises regularly, the proportion is 0.0054. In the other group that doesn't exercise the proportion is 0.0068. Even if it's significant, is the difference in proportion of 0.0014 relevant? If we look at the **relative risk** we see that the group that doesn't exercise is 1.25 times - or 25 percent - more likely to have a heart attack. Confidence intervals can be computed for relative risks, but we won't go into the calculation method here. As you can imagine relative risk is often used in medical sciences and epidemiology where low frequency diseases and symptoms form common topics of research.