

## 7 SIGNIFICANCE TESTS

### 7.02 Test about proportion

Suppose you're interested in the question how many Americans have scuba-diving experience. Suppose you also have good reasons to believe that less than 3 percent of all Americans have ever explored marine life with an oxygen cylinder on their backs. This means that your alternative hypothesis is that  $\pi$  is smaller than 0.03. Your null hypothesis is that  $\pi$  equals 0.03. In this video I'll explain how you should conduct a significance test if you're interested in a proportion.

What we do when we conduct a significance test is this: we assume that the population value we're interested in has a certain value and assess if it is likely that the sample we have collected actually comes from a population with this assumed parameter value. Because we look at a sample, we focus on the sampling distribution. We can determine, for example, what the sampling distribution of the sample proportion looks like, given the assumed population parameter value of 0.03. This is what we do when we conduct a test: we assess how many standard deviations (and because we're dealing with the sampling distribution we talk about standard errors) the observed sample proportion is removed from the population proportion according to the null hypothesis. This number of standard errors is what we refer to as the **test statistic**.

Suppose we have drawn a sample of 1000 Americans and that the proportion of respondents that has scuba-diving experience equals 0.02. What we're going to do is *this*. What you see here is the sampling distribution of the sample proportion *assuming* that the null hypothesis is true and the population proportion thus equals 0.03. How likely is a sample proportion of 0.02 if the population proportion is indeed 0.03? To answer that question we compute the test statistic, or the number of standard errors the sample statistic is removed from the assumed population parameter.

The number of standard errors from the mean is represented by a z-score. We can compute how many z-scores the sample statistic is removed from the population parameter by means of *this* formula: the z-score equals the population proportion assumed under the null hypothesis, subtracted from the sample proportion, divided by the standard error assumed under the null hypothesis. The standard error under the null hypothesis equals the square root of the null hypothesis proportion multiplied with 1 minus that value, divided by the sample size  $n$ .

Let's first compute the null hypothesis standard error. That's the square root of 0.03 multiplied with 0.97 divided by 1000. That's about 0.005. So, our test statistic is 0.02 minus 0.03 divided by about 0.005. That equals -1.85. This means that our sample proportion falls 1.85 standard errors below the population proportion when the null hypothesis is true. *This* is what that looks like in a graph. Is there enough proof to reject our null hypothesis? Well, on the basis of this information, we can look up the probability that our test statistic takes a value like the observed test statistic or even lower. Just look at the z-table. This probability equals 0.0322. This is what we call the **p-value**. This p-value shows us that finding a sample proportion of 0.02 if the population proportion is actually 0.03 is unlikely. But is it unlikely enough to reject the null hypothesis?

Well, that depends on the **significance level** we choose. *Before* we conduct our test, we decide how small the p-value needs to be to actually reject the null hypothesis. The most commonly used significance level is 0.05. That means that if the p-value is equal to or smaller than 0.05, we say that our sample provides enough evidence to reject the null hypothesis. Our p-value of 0.0322 is smaller than the significance level of 0.05. So, if we set our significance level at 0.05, we reject our null hypothesis. This is also represented by what we call the **rejection region**. It is displayed *here*. The critical z-value that forms the border of the rejection region is -1.64. You can look that up in the z-

table. It is the z-score that corresponds to a left-tail probability of 0.05. Our test statistic of -1.85 falls within the rejection region. We thus reject our null hypothesis and conclude that the proportion of Americans with scuba-diving experience is lower than 0.03. We say that our result is **statistically significant**.

In this example our test was based on the alternative hypothesis that  $\pi$  is *smaller* than 0.03. Therefore, we only focused on one side of the sampling distribution: the left one. We performed what we call a **one-tailed test**. What if our alternative hypothesis would be:  $\pi$  is *not equal to* 0.03? In that case we wouldn't focus on just the left side of the distribution, but on both sides of the distribution. In that case, we don't perform a one-tailed, but a **two-tailed test**. If we stick to our significance level of 0.05, this means that the left side corresponds to a cumulative probability of 0.025 and the right side too. The critical values corresponding to this rejection area are -1.96 and 1.96. You can see that in the z-table here. Now, our test statistic of -1.85 doesn't fall within the rejection area anymore. This means that we cannot reject our null hypothesis that  $\pi$  is equal to 0.03. This implies that choosing a one- or two-tailed test can make a huge difference for your conclusions. In practice two-tailed tests are used much more often. My advice: only use the one-sided alternative if you have very good theoretical reasons to do so.

Now, what happens if we change the significance level? We can, for instance, set the significance level at 0.01. This means that we reject the null hypothesis if the p-value is less than 0.01. If we do a one-tailed test, a significance level of 0.01 corresponds to a critical value of -2.33. In our example, we don't reject the null hypothesis. After all, our test statistic does not fall in the rejection region. We know that because: (1) our p-value is larger than 0.01; and (2) our test statistic (which is -1.85) is less extreme than the critical value (which is -2.33). As you can see, just like whether you choose a one- or a two-tailed test, your chosen significance level can strongly affect your results. Important to remember: most significance tests are two-tailed, and based on a significance level of 0.05.