7 SIGNIFICANCE TESTS

7.04 Step-by-step plan

Compare the following two expectations: (1) you expect that more than half of all certified divers in America have more than 35 hours of diving experience; and (2) the mean number of hours of diving experience of all certified American divers is more than 35 hours. At first sight, these two expectations look very similar. But in the first case you're dealing with a proportion (you're interested in the proportion of certified divers with more than 35 hours of diving experience), and in the second case with a mean (you want to know the mean number of hours). So, when conducting a significance test, you should think very carefully about your approach. In this video I will guide you through a step-by-step plan.

Suppose you've asked a simple random sample of 500 certified divers how many hours of experience they have. Suppose you find that when it comes to your sample of 500, a proportion of 0.57 has more than 35 hours of diving-experience and the mean number of hours of experience is 35.5. The standard deviation is 8. In our sample, the distribution of the variable "hours of diving-experience" is approximately normal.

The first step is: Assess if you're dealing with a proportion or with a mean. We already discussed that. In the first case we're dealing with a proportion, and in the second case with a mean.

Step 2: Formulate your hypotheses. In the case of a proportion, your null hypothesis looks like this: Pi equals Pi-zero. In the case of a mean it is Mu equals Mu-zero. We can have three types of alternative hypotheses: Pi or Mu does not equal Pi- or Mu-zero (here you do a two-sided test); Pi or Mu is larger than Pi- or Mu-zero (here you do a one-sided right-tail test); and Pi or Mu is smaller than Pi- or Mu-zero (here you do a one-sided left-tail test). Our null hypotheses are Pi equals 0.5 and Mu equals 35. The alternative hypotheses are: Pi is larger than 0.5 and Mu is larger than 35. We thus have to conduct right-tailed tests.

Step 3: Check if your assumptions are met. In both cases, randomization is of essential importance. You're data must have been collected by means of a random sample or a randomized experiment. In the case of a proportion, an additional assumption is that the product of your sample size and the population proportion according to your null hypothesis and the product of your sample size and one minus the population proportion according to your null hypothesis must be equal to or larger than 15. If you're dealing with a mean, your population distribution should be approximately normal. However, in practice, this is only of importance if your sample size is small and you do a one-tailed test. Regarding our example, all assumptions are met. We have a simple random sample, our n is large and the sample distribution of the variable "hours of diving experience" is approximately normal.

Step 4: Determine your significance level alpha. The most common significance level is 0.05. Our test will be based on an alpha of 0.05.

Step 5: compute your test statistic. In the case of a proportion this is the relevant formula, and in the case of the mean, this is the formula we use. Note that in the case of a proportion we use the z-distribution and in the case of a mean the t-distribution. In our examples we get values of 0.57 minus 0.5 divided by the square root of 0.5 times 0.5 divided by 500. That equals 3.13. And 35.5 minus 35 divided by 8 divided by the square root of 500. That equals 1.40.

Step 6: Draw the relevant sampling distribution and show the null hypothesis value and the test statistic, supplemented with the rejection region and the corresponding critical value. In the case of our proportion, this is the distribution. The critical value corresponding to a right-tailed test with a significance level of 0.05 can be found in the z-table. It is 1.64, so it looks like this. In the case of our mean, this is the distribution. We look up the critical value in the t-table. That's 1.66. That looks like this.

Step 7: assess if your test statistic is located in the rejection region or not. In our first example this is the case: our test statistic (which is 3.15) is further removed from the mean of the sampling distribution than the critical value (which is 1.64). But in the second example this is not the case. Our test statistic (which is 1.40) lies closer to the mean than our critical value (which is 1.66).

Note that in the case of a proportion you can also look up the p-value in the z-table. The p-value is the probability that our test statistic takes a value like the observed test statistic or even more extreme, given the null hypothesis. Our test statistic is 3.13. The corresponding p-value is 0.0009. This value is much smaller than 0.05, which also means that our test statistic is located in the rejection region. Statistical packages can also give you your exact p-value when you use the t-distribution. It is impossible, however, to look it up in the t-table. The reason is that the t-table is not specific enough. Since the shape of the distribution changes with the degrees of freedom, we would need a separate table for each number of degrees of freedom. That's why for each number of degrees of freedom only the most important t-values are listed, that is: those that correspond to the most commonly used significance levels for one- and two-tailed tests.

The next step (8) is: Decide if the null hypothesis should be rejected. The answer is 'yes' when it comes to our proportion-example and 'no' regarding our mean-example.

Finally step 9: Interpret your findings substantively. We can conclude that more than half of all certified divers in America have more than 35 hours of diving experience. However, we cannot conclude that their mean number of hours of diving experience is more than 35 hours.

Before I conclude, one warning: the decision not to reject your null hypothesis does not imply that you *accept* your null hypothesis. In our second example we did not reject our null hypothesis that the mean number of hours of diving experience equals 35. But that doesn't mean that we can conclude that the mean number of hours of diving experience is exactly 35 hours!