

7 SIGNIFICANCE TESTS

7.05 Significance test and confidence interval

Suppose you've asked a simple random sample of 500 certified scuba-divers how many hours of diving experience they have. The mean diving experience is 36 hours. The standard deviation is 8 hours. The sample distribution of the variable "hours of diving-experience" is approximately normal. Based on this sample information, you want to draw inferences about population parameter μ . This is what we call inferential statistics: based on sample information we draw conclusions about the population from which the sample is drawn. There are two methods of inferential statistics: one, inference about interval estimation by means of confidence intervals; and two, inference about point estimation using significance tests. In this video I will show you that these methods are closely related.

Suppose that you expect that the mean number of diving hours of certified scuba-divers differs from 35 and that you're going to do a significance test. We're interested in a mean, so these are our statistical hypotheses: the null hypothesis is $\mu = 35$; and the alternative hypothesis is: $\mu \neq 35$. Our assumptions are met. Our analysis is based on a simple random sample and we're dealing with a large sample, which is, moreover, also approximately normally distributed.

Our test statistic equals 36 minus 35 divided by 8 divided by the square root of 500. That equals 2.80. The sampling distribution looks like this, and we can find in the t-table that the critical value corresponding to a two-tailed test with a significance level of 0.05 is 1.984. So this is the rejection region. Our test statistic is located within the rejection region, so we reject the null hypothesis and conclude that the mean number of diving hours of certified scuba-divers indeed differs from 35.

Now, what happens if we construct a 95% confidence interval? This is the formula we employ: the sample mean plus and minus the t-score for the 95% confidence level times the standard error which equals the standard deviation divided by the square root of the sample size. The relevant t-score is 1.984, which leads to the following equation: 36 plus and minus 1.984 multiplied with 8 divided by the square root of 500. The lower endpoint then is 35.29. The higher endpoint is 36.71. We can thus be confident that with repeated sampling, this interval would contain the actual population mean 95 percent of the time. The interval gives us a range of plausible values for the population mean.

Just like the significance test, this confidence interval suggests that the population mean differs from 35. In general, the results of a two-tailed significance test are in line with conclusions coming from a confidence interval. More specifically: if the p-value in a two-tailed significance test is equal to or smaller than 0.05, a 95% confidence interval does not contain the null hypothesis value. Similarly, if the p-value in a two-tailed test is larger than 0.05, then the 95% confidence interval *will* contain the null hypothesis value. This makes sense, right? The null hypothesis value is a plausible value, so we shouldn't reject the null hypothesis.

This is represented in *this* figure. You can see that the observed value of 36 is located within the rejection region, and that the confidence interval does not contain the null hypothesis population mean. Now suppose the observed mean is 35.5 instead of 36. In that case our test statistic becomes 1.40. You can see that here. In this case our test statistic does not fall within the rejection area. We therefore do *not* reject the null hypothesis. Similarly, the new confidence interval with the endpoints 34.79 and 36.21 *does* contain the null hypothesis mean of 35. We can be confident that with repeated sampling the interval will contain the actual population mean 95% of the time. This means the null hypothesis value of 35 is a plausible value and that we should not reject the null hypothesis.

This means that, although the approach of constructing a confidence interval and the approach of two-tailed hypothesis testing seem very different, they are mathematically related and are consistent with each other.