Example

Confidence Interval on the Mean Variance Unknown, Small Sample

Suppose the football coaches at a state university think that the distribution of the length of field goal attempts made by college football kickers is approximately normal. If the population of field goal attempts is normally distributed or at least approximately normally distributed, then it is appropriate to use a small sample to estimate the population mean with a confidence interval. Assuming the coaches measured 18 randomly selected attempts that yielded a mean of 38 yards and a variance of 64 yards squared, what is the 99% confidence interval for the mean based on this sample of 18 observations?

$$\overline{x} \pm t_{\frac{\alpha}{2}(n-1)} \bullet s_{\overline{x}}$$

$$\overline{x} \pm t_{\frac{.01}{2}(18-1)} \bullet \frac{s}{\sqrt{n}}$$

$$38 \pm 2.898 \frac{\sqrt{64}}{\sqrt{18}}$$

$$38 \pm 2.898(1.885618)$$

$$38 \pm 5.4645$$

$$(32.54, 43.46)$$

The average length of college field goals is estimated to be between 32.54 and 43.46 yards, with 99% confidence. Although it is not certain that the mean length of the college field goals is between 32.54 and 43.46 yards, based on the data these are the most plausible and reasonable values for the parameter of interest, the mean length of college field goals.

Make sure to understand the notation $t_{\frac{.01}{2}(18-1)}$. This notation indicates the value in the t-distribution with 17 degrees of freedom which has 0.005 area under the t(17) curve to the right of the value. To find the value in the t-table locate the row for 17df and locate the column labeled as 0.005. Note that $P(t > t_{\frac{.01}{2}(18-1)}) = 0.005$. The numbers in the subscripts for the t notation indicate the area under the curve to the right and the degrees of freedom.

Example

Hypothesis Test on the Mean
Variance Unknown, Small Sample

Recall the prior example about the field goal attempts that estimated that average length of the attempts with a confidence interval based on 18 observations. Suppose that before the coaches sampled, they believed that the length of the average field goal attempt was 41 yards. Did the sample information that the coaches gathered ($\overline{x} = 38$ and $s^2 = 64$) give them reason to think the average length is significantly different from 41 yards? (Use $\alpha = 0.01$). To address this question the appropriate set of hypotheses are:

$$H_o : \mu = 41 yds.$$

$$H_a : \mu \neq 41 yds.$$

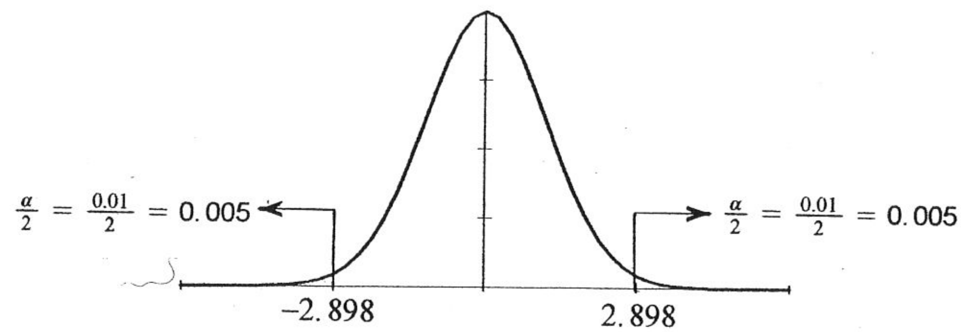
Test statistic for this situation is:

$$t = \frac{\overline{x} - \mu_o}{s_{\overline{x}}} = \frac{\overline{x} - \mu_o}{\frac{s}{\sqrt{n}}} = \frac{38 - 41}{\frac{\sqrt{64}}{\sqrt{18}}} = -1.59.$$

The above test statistic has a t(17) distribution if the mean field goal attempt length is 41 yards. Remember, to say that the distribution for the test statistic is known under a true null implies that the set of possible values and their likelihoods are known for the test statistic if the value stated for the parameter in the null hypothesis is really correct. If the observed test statistic value is an unlikely value in this known set then doubt is cast on the validity of the null hypothesis. The rejection region provides a technique to determine whether the observed test statistic is a likely or an unlikely value in that set. The P-value provides the level of likelihood for the observed test statistic value in that set. The next parts in this example describe these two processes, rejection region and p-value, in detail.

The rejection region identifies likely versus unlikely test statistic values in the following way. The significance level of a hypothesis test, α, identifies the level of probability assigned to the unlikely test statistic values. Alpha is the probability of rejecting a true null hypothesis; it is the chance of seeing an extreme or unlikely test statistic value even if the null hypothesis is true. The probability, α, is set prior to the hypothesis test. In this example it is stated to be 0.01. The consequences of rejecting a true null hypothesis should be considered when the significance level is chosen. In this two-tail example based on a significance level of 0.01 the rejection region would be below –2.898 and above 2.898. The values that exceed 2.898 in magnitude in the t(17) distribution have only 1% chance of occurring. Those values have been identified as the unlikely values in the known distribution of the test statistic assuming that the null hypothesis is true.

To locate the critical or cut off value of 2.898 for the rejection region look in the t-table on the 17 df row and in the column labeled with the right tail area, which in this case is half of alpha or 0.005. On the intersection of that row and column in the t-table the value 2.898 is listed. The following graph relates the critical values of -2.898 and 2.898 to the tail areas that comprise α.



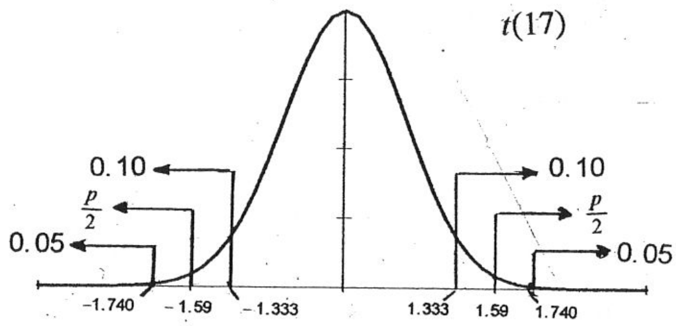
Rejection Region in t(17), two-tail test, $\alpha = 0.01$.

Since the *t*-test statistic is -1.59, which does not exceed 2.898 in magnitude so therefore does not fall in the rejection region, we do not reject the null hypothesis. Conclude, the data do not indicate that the mean field goal attempt is significantly different from 41 yards at the 0.01 level.

The null hypothesis in this example cannot be rejected at the 1% significance level. What is the lowest significance level for which this null hypothesis could be rejected? What is the likelihood of observing a test statistic as extreme as the one generated by the data if the null hypothesis is true? What error rate will have to be tolerated if the conclusion is to reject the null hypothesis? These questions all answered, the p-value of a hypothesis test.

The p-value of a hypothesis test provides information about whether or not the null hypothesis should be rejected. The smaller the p-value the greater the evidence against the null hypothesis. If the data provide an extreme test statistic value that results in a small p-value then the observed data are not very likely if the parameter value stated in the null hypothesis is correct. If data are observed that are highly unlikely when $\mu = 41yds.$ then it is concluded that the mean length is probably not 41yds.

The p-value of a hypothesis test is also called the observed significance level or OSL for the hypothesis test. The p-value or OSL represents the lowest value of alpha, the significance level, for which the null hypothesis could be rejected. The p-value tells the probability of observing a test statistic at least as extreme as the one observed from the data if the null hypothesis is true. The p-value is the tail area out past the observed test statistic in the known distribution under a true null hypothesis. Literally, the *p*-value is the tail area that is cut off by the test statistic value. For example, consider the test statistic $= -1.59$ in the prior example. The *p*-value associated with -1.59 in a two-tail test would be the area below -1.59 in the t(17) and the area above 1.59 in the same t-distribution. The p-value based on the text t-tables can not be stated specifically, but will be determined to be between two values. The graph indicates how the p-value is bounded by two numbers.



P-value,2-tail test, t(17), $0.05 < \frac{p}{2} < 0.10 \Rightarrow 0.10 < p < 0.20$.

The p-value can be approximated by recognizing that 1.59 is between 1.333 and 1.740 in the distribution, so the p-value will be between the areas cut off by those numbers. The value 1.333 has 0.10 probability to the right and 1.740 has 0.05 probability to the right, so half of the p-value (since it's only one of the two tails involved) is between 0.05 and 0.10. The p-value for the hypothesis test is between 10 and 20%.

What does this *p*-value mean? In a hypothesis test the *p*-value represents the error rate that must be tolerated if the null hypothesis is rejected. In this case there is 10%-20% error rate associated with the decision that the null hypothesis is false. More than 10% error is usually too much. With such a large p-value the null hypothesis would usually not be rejected. Conclude, that the data do not indicate that the mean field goal length differs significantly from 41 yards.