

Lesson 13

Hypothesis Testing with the t-test Statistic

Outline

Unknown Population Values

The t-distribution

-t-table

Confidence Intervals

Unknown Population Values

When we are testing a hypothesis we usually don't know parameters from the population. That is, we don't know the mean and standard deviation of an entire population most of the time. So, the t-test is exactly like the z-test computationally, but instead of using the standard deviation from the population we use the standard deviation from the sample.

The formula is: $t = \frac{\bar{X} - \mu}{s_{\bar{x}}}$, where $s_{\bar{x}} = \frac{s}{\sqrt{n}}$

The standard deviation from the sample (S), when used to estimate a population in this way, is computed differently than the standard deviation from the population. Recall that the sample standard deviation is "S" and is computed with n-1 in the denominator (see prior lesson). Most of the time you will be given this value, but in the homework packet there are problems where you must compute it yourself.

The t-distribution

There are several conceptual differences when the statistic uses the standard deviation from the sample instead of the population. When we use the sample to estimate the population it will be much smaller than the population. Because of this fact the distribution will not be as regular or "normal" in shape. It will tend to be flatter and more spread out than population distribution, and so are not as "normal" in shape as a larger set of values would yield. In fact, the t-distribution is a family of distributions (like the z-distribution), that vary as a function of sample size. The larger the sample size the more normal in shape the distribution will be. Thus, the critical value that cuts off 5% of the distribution will be different than on the z-score. Since the distribution is more spread out, a higher value on the scale will be needed to cut off just 5% of the distribution.

The practical results of doing a t-test is that 1) there is a difference in the formula notation, and 2) the critical values will vary depending on the size of the sample we are using. Thus, all the steps you have already learned stay the same, but when you see that the problem gives the standard deviation from the sample (S) instead of the population (σ), you write the formula with "t" instead of "z", and you use a different table to find the critical value.

The t-table

Critical values for the t-test will vary depending on the sample size we are using, and as usual whether it is one-tail or two-tail, and due to the alpha level. These critical values

are in the Appendices in the back of your book. See page A27 in your text. Notice that we have one and two-tail columns at the top and degrees of freedom (df) down the side. Degrees of freedom are a way of accounting for the sample size. For this test $df = n - 1$. Cross index the correct column with the degrees of freedom you compute. Note that this is a table of critical values rather than a table of areas like the z-table.

Also note, that as n approaches infinity, the t-distribution approaches the z-distribution. If you look at the bottom row (at the infinity symbol) you will see all the critical values for the z-test we learned on the last exam.

Confidence Intervals

If we reject the null with our hypothesis test, we can compute a confidence interval. Confidence intervals are a way to estimate the parameters of the unknown population. Since our decision to reject the null means that there are two populations instead of just the one we know about, confidence intervals give us an idea about the mean of the new unknown population.

See the Confidence Interval demonstration on the web page or click here <http://faculty.uncfsu.edu/dwallace/sci.html> for the rest of the lesson.