This is Professor Rogers. In this presentation, we look at t-tests.

T-tests are used when one variable is numeric and the other one is a categorical variable with two categories. t is the same Student’s t used in determining a confidence interval. With regard to the elements of inference, the magnitude is expressed as an absolute number and is not standardized. Direction is not present in a traditional sense, but we can speak of one group being “larger” or “higher” than another group. Statistical significance of the distance is present.

The slide shown here illustrates our intent. Here we see two groups, one whose mean is x-subscript-1 and the other whose mean is x-subscript-2. We measure the distance between the value of the two groups and try to determine whether the distance is statistically different from zero.

Just like a confidence interval can be built around the sample mean to estimate the population mean, each group mean also has its own confidence interval, and whether these intervals overlap has to be taken into account. Thus, the difference between the group means could be very small and even non-existent when the distance between the confidence intervals is small. . .

. . . or there could be a large difference when the distance between the confidence intervals is large.

The example in this video deals with a commonly discussed issue in criminal justice: Are violent crime rates in the South higher than the rest of the country? We examine the question using state-level data. On the surface, the crime rates look different. The average violent crime rate in the South is 471 crimes per 100,000 people, while in the rest of the country it is 318. The question is whether this difference is statistically significant.

The path for the test is Analyze > Compare Means > Independent Samples T Test. Notice that three different statistics claim the name t-tests. It is important to use the one labeled Independent Samples.

Here is the command dialog box.

The numeric variable is toggled into the top box on the right. The variable I selected for analysis is the log of the violent crime rate. I emphasize here my use of the logged variable. Crime rates are skewed to the right and the t-test assumes that the distribution is normally distributed, so the data is transformed through a logarithmic transformation so that it conforms to the assumptions of the statistic.

The categorical variable goes in the bottom box called Grouping Variable. I have already created a dummy variable for states in the South—it has a value of one if the state is in the South and 0 if it is outside the South.

The variable South has question marks next to it. We need to define the two groups, so we select the Define Groups option to get the dialog box that is placed at the bottom. Group 1 has been designated as those states with a value of 1 on South and Group 2 is states with a value of 0. We are using a dummy variable for the group—the Cut Point option allows us to use a numeric variable and split it into two groups by simply entering the point where we want the division to occur, but we obviously do not need the cut point in this analysis.

Here is what the command dialog box looks like after the groups are defined.

The output gives us two tables. The first table contains the descriptive statistics for the two groups. We see the N, mean, standard deviation, and standard error for each group. Note that the logarithmic transformation for violent crime rate makes the two areas seem much closer together. South has a mean of 6.07 and states outside the South have a mean of 5.70. With the t-test, we do not have direction in the sense that relationships are positive or negative, but we can say that one group is larger or higher than another. In the example, we can say that the South has a higher violent crime rate than the rest of the country. However, this relationship addresses only one element of inference. Could it be that the difference between the two areas simply are an artifact of the shape of the distribution, and once we normalize the distribution the difference disintegrates?

The answer can be found in the second table, which contains the result of the t-test analysis. The table is quite wide.

It can be broken into two natural parts, and each part will be blown up.

The first part of the table addresses whether the variances of the two groups are equal. If the two variances are equal we calculate the t-test one way. If they are not equal, we have to calculate the t-test another way. The test of the assumption of equality is the Levine’s test, which is the first half of the second table.

In the first half of the second table, the column we want to look at is the Significance column. If the value is greater than or equal to .05, we can assume that the variances are equal and our calculation should be based on the first assumption. If the values are less than .05, we can assume that they are different and we should be using the calculations of the second assumption. The value here suggests that we use the assumption of equal variances. A word of warning: This significance value applies only to the Levine’s test. A common mistake is for students to confuse the statistical significance of the Levine’s test with the statistical significance of the t-test.

Once we have made our decision about the equality of variances, we move to the right half of the second table. The results for both the equal-variance and unequal-variance tests are reported. Because we have equal variances, we look at the top line. If we had unequal ones, we would look at the bottom line.

Let’s move from left to right along the line. The value t is the value of t that it takes for the null hypothesis to be true. Here we see the value of t is 3.221. Right away this number gives us a clue about our results. We know from previous videos that t for a 95% confidence interval with an infinite number of degrees of freedom is 1.96. If we have a larger t, then the value of the null hypothesis is outside of the range. 3.221 is larger than 1.96, so we know that the p-value is going to be less than .05 once we see the significance value.

The df column stands for degrees of freedom. With t and the degrees of freedom, we could find the p-value manually in a t-table, but we don’t have to.

The reason we do not have to is the third column, which gives us the statistical significance of the test, one of our three elements of inference. The threshold for statistical significance is .05. Since .002 is less than .05, we can conclude that the difference between the two groups is statistically significance at the .05 level of significance.

The next column gives us the calculation for the difference in means. The difference between the mean of South and the rest of the country is .37334. This difference is the same regardless the assumption about variance. This number is the magnitude of our relationship, but it is a hard number to use because it is not standardized. We have to use our judgment in determining whether this difference should be treated as small or large.

Next we have the standard error of the difference.

Finally, we have the low and high values of the 95% confidence interval. Based on these numbers, we can say the true difference in the population is somewhere between .14040 and .60629.

Thus, here is our hypothesis test. We see that zero falls outside the 95% confidence interval, so we reject the null hypothesis that distance between the two groups is non-existent.

Let’s take a quick look at the underlying calculations. The t-statistic has as its numerator the difference of the means of the two groups, and this difference is standardized by the use of the standard deviations in the denominator.

With unequal variances, the calculation for the denominator changes, but the denominator still is built around the standard deviations.

Let me close with mention of one controversy. Levine’s test is not universally accepted and not included in some statistical software. The issue is that the statistical significance number of the Levine test masks underlying aspects of the distribution of each group that should be studied more closely, much like the way we study skewness and kurtosis. Some statisticians assert that the Levine’s test should not substitute for one’s own judgment. Reliance on your own judgment creates situations where one cannot tell whether to assume equal or unequal variances. When this situation occurs, assume variances are unequal, which shows good faith on the researcher’s part because it is harder to establish significance.

The effect size for a t-test is Cohen’s d. Unfortunately, Cohen’s d is not reported in the output and must be calculated manually. The general formula is the difference of means divided by the pooled standard deviation. The pooled standard deviation is the square root of the sums of the variances of the two groups divided by two.

The interpretation of effect size is built on the absolute value of d, and its thresholds are different for what we are accustomed to seeing in most bivariate statistics. The threshold for a weak relationship begins at .2, a medium effect begins a .5, and a strong effect begins at .8.

In the case of our example regarding the difference in violent crime rates between Southern and non-Southern counties, the effect size is .93. In other words, the difference demonstrates a strong relationship between violent crime rates and region of the country.

T-tests are used when one variable is numeric and the other variable has two categories. The test determines whether the distance between the two means is statistically significant. Cohen’s d provides a standardized effect size. Direction does not exist in a traditional sense, though one group can be said to be “larger” or “higher” than another.