Hypothesis Test Notes

Test Statistics

The ability to "test" what someone says about a population is largely dependent on being about to tell if the random sample value was significantly different than the population value. In other words does the random sample value significantly disagree with what the person said about the population?

That is a problem. The answer to this is it sometimes impossible to tell with your own eyes. Let us suppose that the population percentage is 0.25 (25%) and the sample value is 0.22 (22%). Is the sample significantly different? We don't know. Sometimes 3% is a significant difference and sometimes 3% is not significant.

So how do we tell if our sample data significantly disagrees with a population value?

The answer to this is we need to measure the difference in a very special way. Knowing how many miles different or percentage points different is not going to help us. We need to know how many "standard errors" different they are.

This is called a <u>Test Statistic</u>.

<u>Example 1</u>

Let's look at the percentage problem where the population percentage is 0.25 (25%) and the sample value is 0.22 (22%). We know the sample value is % lower, but we do <u>not</u> know if that is significant. Another important bit of information is the sample size. In this case it was 1000.

To find this out calculate the test statistic.

Formulas for Test Statistics follow a general pattern. Remember a test statistic counts how many standard errors that the sample value is above or below the population value. So the formula looks like the following:

$$\frac{(\text{Sample Value} - \text{Population Value})}{\text{Standard Error}}$$

Remember in the last unit we saw that statisticians often used formulas to approximate the standard error.

Here is the formula for the test statistic when comparing a sample percentage $(\hat{p})$ and population percentages *(π).* Recall that the number of standard deviations is often represented with at Z-score or T-score, so it is not surprising that the test statistic is often a T or Z. Also remember that z-scores and t-scores are often rounded to the hundredths place.

$$Z = \frac{(\hat{p} - \pi)}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

Let's plug in our numbers. Remember that the sample percentage $\hat{p} = 0.22$ and population percentages *π = 0.25* and the sample size *n = 1000.*

$$z = \frac{(\hat{p} - p)}{\sqrt{\frac{p(1-p)}{n}}} = \frac{(0.22 - 0.25)}{\sqrt{\frac{0.25(1-0.25)}{1000}}} = \frac{(0.22 - 0.25)}{\sqrt{\frac{0.25(0.75)}{1000}}} = \frac{-0.03}{0.013693} \approx -2.19$$

Our z-score test statistic was about −2.19. So our sample percentage (22%) was 2.19 standard errors below the population percentage 25%.

So is the sample proportion 0.22 significant different than the population proportion (0.25) in the null hypothesis? Here is the significance rule for judging test statistics.

Significance Rule:

If the test statistic falls in one of the tails determined by a critical value, then the sample data significantly disagrees with population parameter in the null hypothesis.

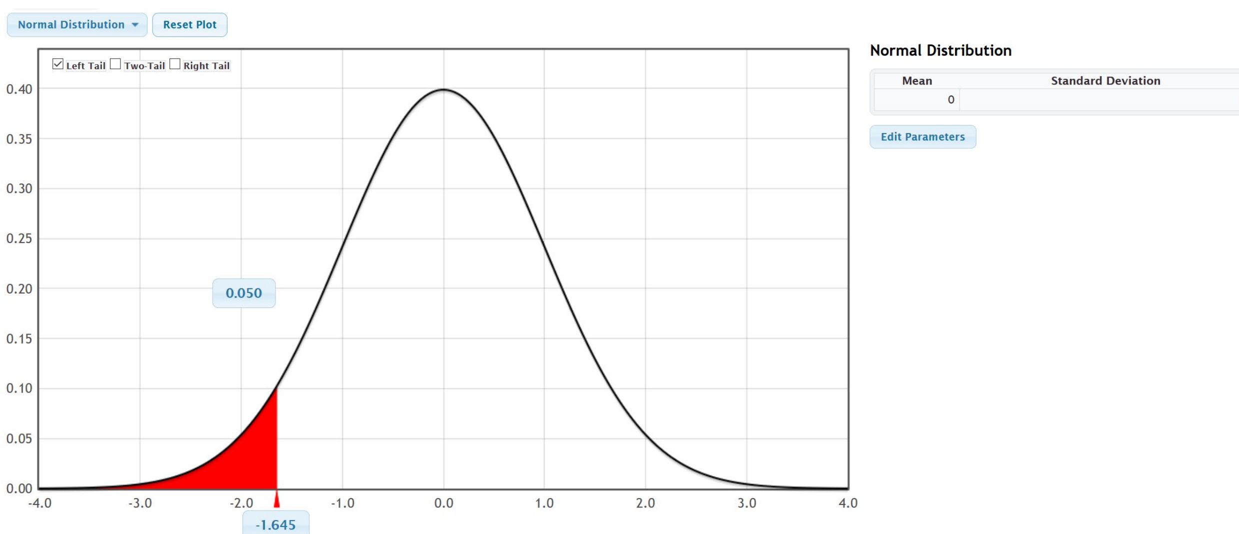
If the test statistic does NOT fall in one of the tails determined by a critical value, then the sample data does NOT significantly disagree with the population parameter in the null hypothesis.

To determine significance, we need to draw a picture and calculate the critical value. Critical values are based on the significance levels (opposite of confidence levels).

Confidence Levels (1 – α)	Significance Levels (Alpha Levels "α")
0.90	0.10
0.95	0.05
0.99	0.10

So let's suppose we are dealing with a left tailed test with a 5% significance level (α = 0.05). The significance level is the probability in the tail or tails. In this case we will have 0.05 in the left tail. The Z-score corresponding to 5% in the left tail will be our critical value.

We can look this up using StatKey. Go to www.lock5stat.com and click on "StatKey". Under theoretical distributions, click on "normal". Click on "left tail" and put in 0.05 in the upper probability box. The number below will be our critical value.



So our left tail starts at −1.645. Where does our test statistic fall? Our test statistic of Z = −2.19 falls in the left tail. So the sample percentage of 22% is significantly lower than the population percentage of 25%. The sample data significantly disagrees with the null hypothesis.

<u>Example 2</u>

What do we do if we want to know if a sample mean is significantly different from a population mean? Sometimes 13 pounds is a lot and sometimes 13 pounds is very little. An article in a health magazine claims that the mean average weight of all men is about 175 pounds. A random sample of 60 men found that the sample mean was 188 pounds with a standard deviation of 96 pounds. So the sample mean 188 is 13 pounds heavier than the population mean of 175. Is that significant?

The answer again is we don't know. We would need calculate a test statistic to see if 13 pounds is a lot in this situation. Here is the formula for calculating test statistics to compare sample and population means. Notice it follows the same general pattern and seeks to count how many standard errors the sample value is above or below the population value. Notice we label the test statistic as a T-score since again it is the number of standard deviations (errors) and the T-score is more accurate than the Z-scores for small quantitative data sets.

$$T = \frac{(\text{Sample Value } - \text{ Population Value})}{\text{Standard Error}} = \frac{(\bar{x} - \mu)}{\left(\frac{s}{\sqrt{n}}\right)}$$

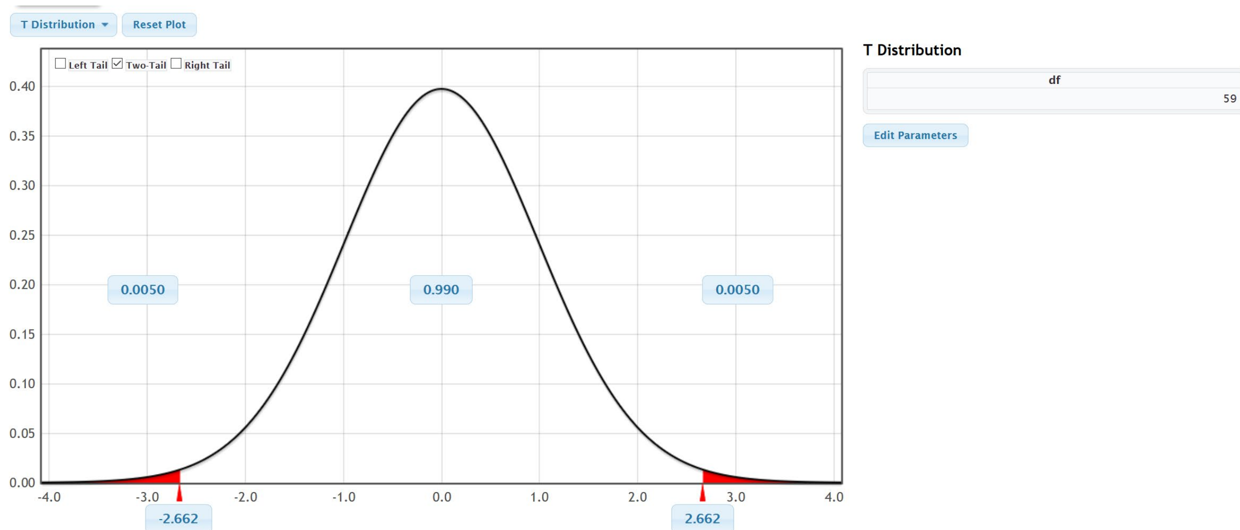
Now use the formula to calculate the test statistic. Remember the sample mean x-bar is 188, the population mean mu is 175, the standard deviation s is 96 and sample size n is 60.

$$T = \frac{(\bar{x} - \mu)}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{(188 - 175)}{\left(\frac{96}{\sqrt{60}}\right)} = \frac{+13}{12.39355} \approx +1.0489 \approx +1.05$$

Notice our sample mean of 188 pounds is not a significant disagreement with population mean of 175 pounds. Our sample mean of 188 pounds is only 1.05 standard errors above the population mean of 175.

Is it significant? Again, we need to see if the test statistic falls in a tail determined by the critical value. Suppose we are using a 1% significance level and a two-tailed test. We can look up critical value T-score using StatKey. T-scores use degrees of freedom (n – 1 = 60 – 1 = 59).

Go to www.lock5stat.com and click on “StatKey”. Under theoretical distributions, click on “t”. Click on “two tail”. Since it is two-tailed we will need to put half of our significance level in each tail. So put in 0.005 in either of the upper probability boxes in either tail. The other tail will adapt. The numbers below will be our critical values.



So the tails start at ±2.662 and the T-test statistic of +1.05 does NOT fall in either tail. So the sample mean of 188 pounds is NOT significantly higher than the population mean of 175 pounds. The sample data does NOT significantly disagree with the null hypothesis.

Important Notes:

- *It is important to understand how test statistics work and what the formulas mean. It is not important to calculate these by hand. Statistics programs like Statcato and StatCrunch can calculate the test statistic and the critical value in very quickly with much better accuracy. It is important that you can explain the meaning of the test statistic and what it tells us about significance.*

- **Significance Rule:**

  If the test statistic falls in one of the tails determined by a critical value, then the sample data significantly disagrees with population parameter in the null hypothesis.

  If the test statistic does NOT fall in one of the tails determined by a critical value, then the sample data does NOT significantly disagree with the population parameter in the null hypothesis.