

Lecture 6

(Additional information on t-tests and hypothesis testing)

Lecture 5 focused on perhaps the most common of the t-tests, the two sample assuming equal variance. There are other versions as well; Excel lists two others, the two sample assuming unequal variance and the paired t-test. We will end with some comments about rejecting the null hypothesis.

Choosing between the t-test options

As the names imply each of the three forms of the t-test deal with different types of data sets. The simplest distinction is between the equal and unequal variance tests. Both require that the data be at least interval in nature, come from a normally distributed population, and be independent of each other – that is, collected from different subjects.

The F-test for variance.

To determine if the population variances of two groups are statistically equal – in order to correctly choose the equal variance version of the t-test – we use the F statistic, which is calculated by dividing one variance by the other variance. If the outcome is less than 1.0, the rejection region is in the left tail; if the value is greater than 1.0, the rejection region is in the right tail. In either case, Excel provides the information we need.

To perform a hypothesis test for variance equality we use Excel's *F-Test Two-Sample for Variances* found in the Data Analysis section under the Data tab. The test set-up is very similar to that of the t-test, entering data ranges, checking Labels box if they are included in the data ranges, and identifying the start of the output range. The only unique element in this test is the identification of our alpha level.

Since we are testing for equality of variances, we have a two sample test and the rejection region is again in both tails. This means that our rejection region in each tail is 0.25. The F-test identifies the p-value for the tail the result is in, but does not give us a one and two tail value, only the one tail value. So, compare the calculated p-value against .025 to make the rejection decision. If the p-value is greater than this, we fail to reject the null; if smaller, we reject the null of equal variances.

Excel Example. To test for equality between the male and female salaries in the population, we set up the following hypothesis test.

Research question: Are the male and female population variances for salary equal?

Step 1: H_0 : Male salary variance = Female salary variance

H_a : Male salary variance \neq Female salary variance

Step 2: Reject H_0 if p-value is less than Alpha = 0.025 for one tail.

Step 3: Selected test is the F-test for variance

Step 4: Conduct the test

The screenshot shows the Microsoft Excel interface with the 'Data' tab selected. The 'Data Tools' group is active, displaying options like 'Sort', 'Filter', 'Text to Columns', and 'Forecast'. Below the ribbon, the 'F-Test Two-Sample for Variances' tool has been used to generate the following output table:

	Male	Female						
	24	22		F-Test Two-Sample for Variances				
	24	22				Male	Female	
	25	23		Mean		52	38	
	27	23		Variance		316	334.6667	
	28	23		Observations		25	25	
	28	23		df		24	24	
	40	24		F		0.944223		
	43	24		P(F<=f) one-tail		0.444679		
	47	24		F Critical one-tail		0.440669		
	47	24						
	49	24						
	56	24						
	58	34						

Step 5: Conclusion and interpretation. The test resulted in an F-value less than 1.0, so the statistic is in the left tail. Had we put Females as the first variable we would have gotten a right tail F-value greater than 1.0. This has no bearing on the decision. The F value is larger than the critical F (which is the value for a 1-tail probability of 0.25 – as that was entered for the alpha value).

So, since our p-value (.44 rounded) is $> .025$ and/or our F (0.94 rounded) is greater than our F Critical, we fail to reject the null hypothesis of no differences in variance. The correct t-test would be the two-sample T-test assuming equal variances.

Other T-tests.

We mentioned that Excel has three versions of the t-test. The equal and unequal variance versions are set up in the same way and produce very similar output tables. The only difference

is that the equal variance version provides an estimate of the common variation called *pooled variance* while this row is missing in the unequal variance version.

A third form of the t-test is the *T-Test: Paired Two Sample for Means*. A key requirement for the other versions of the t-test is that the data are independent – that means the data are collected on different groups. In the paired t-test, we generally collect two measures on each subject. An example of paired data would be a pre- and post-test given to students in a statistics class. Another example, using our class case study would be comparing the salary and midpoint for each employee – both are measured in dollars and taken from each person. An example of NON-paired data, would be the grades of males and females at the end of a statistics class. The paired t-test is set up in the same way as the other two versions. It provides the correlation (a measure of how closely one variable changes when another does – to be covered later in the class) coefficient as part of its output.

An Excel Trick. You may have noticed that all of the Excel t-tests are for two samples, yet at times we might want to perform a one-sample test, for example quality control might want to test a sample against a quality standard to see if things have changed or not. Excel does not expressly allow this. BUT, we can do a one-sample test using Excel.

The reason is a bit technical, but boils down to the fact that the two-sample unequal variance formula will reduce to the one-sample formula when one of the variables has a variance equal to 0. So using the unequal variance t-test, we enter the variable we are interested – such as salary – as variable one and the hypothesized value we are testing against – such as 45 for our case – as variable two, ensuring that we have the same number of variables in each column.

Here is an example of this outcome.

Research question: Is the female population salary mean = 45?

Step 1: H_0 : Female salary mean = 45

H_a : Female salary mean \neq 45

Step 2: Reject the null hypothesis is less than Alpha = 0.05

Step 3: Selected test is the two sample unequal variance t-test

Step 4: Conduct the test

Original Data Set [Compatibility Mode] - Excel

Formulas Data Review View Tell me what you want to do

Alignment Number Styles

0.0677242368743384

G	H	I	J	K	L	M	N	O	P
		Female	Hypothesis			t-Test: Two-Sample Assuming Unequal Variances			
		22	45						
		22	45				<i>Female</i>	<i>Hypothesis</i>	
		23	45		Mean	38	45		
		23	45		Variance	334.6667	0		
		23	45		Observatic	25	25		
		23	45		Hypothesiz	0			
		24	45		df	24			
		24	45		t Stat	-1.91321			
		24	45		P(T<=t) one-tail	0.033862			
		24	45		t Critical one-tail	1.710882			
		24	45		P(T<=t) two-tail	0.067724			
		24	45		t Critical two-tail	2.063899			
		34	45						
		34	45						
		35	45						
		36	45						

Step 5: Conclusions and Interpretation. Since the two tail p-value is greater than ($>$) .025 and/or the absolute value of the t-statistic is less than the critical two tail t value, we fail to reject the null hypothesis. Our research question answer is that, based upon this sample, the overall female salary average could equal 45.

Miscellaneous Issues on Hypothesis Testing

Errors. Statistical tests are based on probabilities, there is a possibility that we could make the wrong decision in either rejecting or failing to reject the null hypothesis. Rejecting the null hypothesis when it is true is called a Type I error. Accepting (failing to reject) the null when it is false is called a Type II error.

Both errors are minimized somewhat by increasing the sample size we work with. A type I error is generally considered the more severe of the two (imagine saying a new medicine works when it does not), and is managed by the selection of our alpha value – the smaller the alpha, the harder it is to reject the null hypothesis (or, put another way, the more evidence is needed to convince us to reject the null). Managing the Type II error probability is slightly more complicated and is dealt with in more advanced statistics class. Choosing an alpha of .05 for most test situations has been found to provide a good balance between these two errors.

Reason for Rejection. While we are not spending time on the formulas behind our statistical outcomes, there is one general issue with virtually all statistical tests. A larger sample size makes it easier to reject the null hypothesis. What is a non-statistically significant outcome based upon a sample size of 25, could very easily be found significant with a sample size of, for example, 25,000. This is one reason to be cautious of very large sample studies – far from meaning the results are better, it could mean the rejection of the null was due to the sample size and not the variables that were being tested.

The **effect size** measure helps us investigate the *cause* of rejecting the null. The name is somewhat misleading to those just learning about it; it does NOT mean the size of the difference being tested. The significance of that difference is tested with our statistical test. What it does measure is the *effect* the variables had on the rejection (that is, is the outcome practically significant and one we should make decisions using) versus the impact of the sample size on the rejection (meaning the result is not particularly meaningful in the real world).

For the two-sample t-test, either equal or unequal variance, the effect size is measured by Cohen's D. Unfortunately, Excel does not yet provide this calculation automatically, however it is fairly easy to generate.

Cohen's D = (absolute value of the difference between the means)/the standard deviation of both samples combined.

Note: the total standard deviation is not given in the t-test outputs, and is not the same as the square root of the pooled variance estimate. To get this value, use the *fx* function *stdev.s* on the entire data set – both samples at the same time.

Interpreting the effect size outcome is fairly simple. Effect sizes are generally between 0 and 1. A large effect (a value around .8 or larger) means the variables and their interactions caused the rejection of the null, and the result has a lot of practical significance for decision making. A small effect (a value around .2 or less) means the sample size was more responsible for the rejection decision than the variable outcomes. The medium effect (values around .5) are harder to interpret and would suggest additional study (Tanner & Youssef-Morgan, 2013).

References

Lind, D. A., Marchel, W. G., & Wathen, S. A. (2008). *Statistical Techniques in Business & Finance*. (13th Ed.) Boston: McGraw-Hill Irwin.

Tanner, D. E. & Youssef-Morgan, C. M. (2013). *Statistics for Managers*. San Deigeo, CA: Bridgepoint Education.