Exploring differences using t-tests

The t-test is one kind of inferential statistic used to figure out if there are statistically significant differences between two sets of scores. For example, say you wanted to know if girls did better than boys (to a statistically significant degree) on your work sample post-test. I'm not saying you have to do this on your work sample post-test – but you could using a t-test!

There are two different kinds of t-tests used when in different situations. The first, called an independent samples t-test, is used when the two sets of scores are independent of one another. In the boys and girls instance above, the two groups (boys vs. girls) are independent of one another. The second, called non-independent samples t-test, is used when the two sets of scores are related. For example, if you wanted to know if kids did statistically significantly better on your post-test than on your pre-test, the non-independent sample t-test is for you!

What follows are examples of how to calculate each kind of statistic and room for you to practice doing your own. As with previous activity sheets, do your best and give it a whirl. The idea is to try to work through the logic of the activity! Before you begin, I suggest you review the section on t-tests in unit 8 beginning on pg. 457.

Independent Samples t-test

Here's the formula for independent samples t-test:

$$SS = \sum x^2 - \frac{(\sum x)^2}{n}$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{SS_1 + SS_2}{n_1 + n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

SS = sum of squares... and here's how to calculate it

Given this formula, say you wanted to figure out if there is a statistically significant difference on post-test scores between experimental and control classrooms. The experimental classroom learned to read using the Super Inventory and Comprehension Kit (SICK) and the other used regular old teaching methods. The following post-test data was recorded.

Student #	SICK class	Control class
1	82%	60%
2	70%	72%
3	98%	78%
4	90%	74%
5	84%	58%
6	64%	72%
7	68%	68%
8	88%	94%
9	84%	84%
10	86%	70%
11	80%	72%
12	78%	60%
13	78%	72%
14	94%	70%
15	98%	82%

Using the data and formula from the previous page, watch how I calculate the t for these two sets of scores.

First I make a chart like this:

Student #	SICK class (x_1)	x_1^2	Control class (x_2)	x_2^2
1	82	6724	60	3600
2	70	4900	72	5184
3	98	9604	78	6084
4	90	8100	74	5476
5	84	7056	58	3364
6	64	4096	72	5184
7	68	4624	68	4626
8	88	7744	94	8836
9	84	7056	84	7056
10	86	7396	70	4900
11	80	6400	72	5184
12	78	6084	60	3600
13	78	6084	72	5184
14	94	8836	70	4900
15	98	9604	82	6724

Then I sum: $\Sigma x_1 = 1242$ $\quad \Sigma x_1^2 = 104308$ $\quad \Sigma x_2 = 1086$ $\quad \Sigma x_2^2 = 79902$

$$\bar{x}_1 = \frac{1242}{15} = 82.8 \qquad \bar{x}_2 = \frac{1086}{15} = 72.4$$

↳ this is read as "x bar" ... which means the average of x_1

Finally, I plug all the numbers into the formula and crank it out:

$$SS_1 = 104308 - \frac{(1242)^2}{15} = 1470.4$$

$$SS_2 = 79902 - \frac{(1086)^2}{15} = 1275.6$$

$$t = \frac{82.8 - 72.4}{\sqrt{\left(\frac{1470.4 + 1275.6}{15 + 15 - 2}\right)\left(\frac{1}{15} + \frac{1}{15}\right)}}$$

$$t = \frac{10.4}{\sqrt{\left(\frac{2746}{28}\right)(.133)}} \qquad t = \frac{10.4}{3.616} \qquad \boxed{t = 2.88}$$

So… once I have the value for t (t=2.88), I can go to the probability table showing the significance of t for varying degrees of freedom (download t probability table in unit 8). The formula for coming up with degrees of freedom for independent samples t-test is $n_1 + n_2 - 2$. In this case that's 15 + 15 – 2 or 28. So I go down the degrees of freedom (df) column on the left side of the table until I get to 28 and then start moving across through the figures on the right. The numbers in the table are t-values that indicate, for example, for 28 df and a probability of .20 (the first column), your value of t has to be greater than 1.313 for there to be an 80% confidence in the difference between the scores. See how I got 80%? A .20 probability really means that 80% of the time, the difference would occur.

So… for our example, 2.88 is larger than the numbers in all but the last column. This means that the differences on post-test reading scores are statistically significant to the .01 level ($p<.01$) – in favor of the SICK reading method. In other words, there's less than 1 chance in 100 that these differences in post-test scores happened by chance. We should feel pretty secure, then, in the effectiveness of the SICK method – right!?

Or should we… remember, all inferential statistics are based on several assumptions about a normal distribution of scores and a fairly large sample size. If you were really hard-nosed in your approach, you should plot the two sets of post-test scores on a frequency distribution chart and check to see that each set is relatively normally distributed. As a rule of thumb, you need about 30 scores in a distribution to have confidence in your t-test scores (as dependent on sample size). So… it is fair to have some concern about the reliability of the results of this t-test.

So that's how to do an independent samples t-test! Now you try it!

Your turn:

Say you wanted to know if girls did better than boys on a post-test of mathematical computations. The independent samples t-test will tell you. Here's the data – figure it out! Remember to report the t-value and the p-value. Good luck!

Student #	**Boys Score (x_1)**	x_1^2	**Girls Score (x_2)**	x_2^2
1	22		26	
2	26		30	
3	22		22	
4	18		26	
5	18		26	
6	22		24	
7	22		30	
8	22		20	
9	28		14	
10	18		30	

Put your answers here:

t = _____

p< _____

Non-independent samples t-test

Now let's take a look at the kind of t-test to use when the two sets of scores are related. The easiest way to think about this is the case of pre and post-test data. Say you wanted to know if the same group of students did statistically significantly better (or worse, for that matter) on the post-test than on the pre-test. In this case, the two sets of scores are from the same kids – so the non-independent samples t-test is the inferential statistic to use.

Here the formula for non-independent samples t-test:

$$t = \frac{\bar{D}}{\sqrt{\dfrac{\Sigma D^2 - \dfrac{(\Sigma D)^2}{n}}{n(n-1)}}}$$

D = difference between scores

$\bar{D}$ = average distance between scores!

Given this formula, and the pre/post-test scenario described above, look how I calculate the t statistic for these two groups of scores. Keep in mind that student #1 on the pre-test has to be the same student as #1 on the post-test – in other words, they have to match.

Student #	Pre-test Score	Post-test Score
1	62	88
2	34	60
3	56	60
4	60	92
5	80	100
6	68	78
7	48	72
8	50	76
9	48	84
10	22	58

The first thing I do is to make a chart that looks like this:

Student #	Pre-test Score	Post-test Score	D	D^2
1	62	70	8	64
2	34	50	16	256
3	56	60	4	16
4	70	70	0	0
5	80	82	2	4
6	68	78	10	100
7	48	64	16	256
8	50	56	6	36
9	48	40	-8 – subtract!	64
10	22	30	8	16
Sum totals...			$\Sigma D = 62$	$\Sigma D^2 = 812$

$\bar{D} = 6.2$

Go to the next page to see how I plugged the numbers into the formula and cranked it out!

Here's the calculations:

$$t = \frac{6.2}{\sqrt{\dfrac{812 - \dfrac{(62)^2}{10}}{10(9)}}} \qquad t = \frac{6.2}{\sqrt{\dfrac{812 - 384.4}{90}}} \qquad t = \frac{6.2}{2.18}$$

$$\boxed{t = 2.84}$$

In the end, I got a t-value of 2.84. The degrees of freedom (df) formula for non-independent samples is n-1, where n is the number of pairs of scores. So in this example df = 10-1 or 9. So if I go to the values of t chart on page 561 of the book and look across the df column of 9 I see that 2.84 is just barely bigger than the value of t for a probability of .02 but less than the value of t for a probability of .01. So I can say that, for these two sets of scores my t=2.84 which registers a p<.02. This is a little atypical as we commonly use either .05, .01, or .001 as the "markers" in education research but… we're able to be a little more specific than the value of .05.

Your turn:

Say you had 10 goldfish in your pond and you fed them new Super Miracle Andesine and Caratine Krackers (SMACK) for a number of weeks. At the end of that time, you want to know if they grew a statistically significantly amount – how's that for a variation on the pre/post-test idea! Here's the data, calculate a non-independent samples t-test to see if there's a difference.

Fish #	Pre length	Post length	D	D^2
1	8	9		
2	8	10		
3	6	9		
4	10	11		
5	12	12		
6	4	8		
7	5	9		
8	2	9		
9	10	11		
10	9	12		

Put your answers here:

t = _____

p< _____

Questions about t-tests and the calculations above:

1. Give an example of an instance in your life as a teacher when using t-tests might be an interesting or important thing to do. Be sure to indicate whether your example calls for independent samples t-test or non-independent samples t-test.

2. Looking at the chart for distribution of t (probability table for t), why do the values of t get smaller as you move down each column in the chart? For example, looking at the first column where the probability is for .20, why do the values for t get smaller as you go down the column?

3. Similarly, why do the values for t get larger as you move across the rows? For example, looking at the row where df equal 20, why do the t values get larger and larger as you move across the row?

4. One of the assumptions when using t-tests is that the number of scores in each group is equal. This is easy to understand when using non-independent samples t-test (pre/post tests – as the two groups are really the same kids at different time points) but what do you think researchers do in independent samples t-tests when the two groups aren't of equal size?

5. What other kind of assumptions must be examined when employing t-tests (as well as most inferential tests)? Go back and look in the inferential statistics chapter if necessary.

6. In the last example, where you fed your goldfish SMACK and measured their growth, there's one huge threat to internal validity. Think about it for a minute before you look at the chart you downloaded from unit 8 on threats to internal and external validity to consider different options. Really, you should have compared two independent groups of fish – the treatment group and a control group – using an independent samples t-test! But that's not the threat I'm talking about – can you think of it? (think, think… maturation… got it?)

7. The last issue I want to mention is the fact that t-tests call for the two groups of numbers to be the same kind of numbers – on the same scale, for example. In our examples, test scores were either percentages, raw test scores, or, in the case of the fish, some units of length. But… what do you do when you want to compare kids scores on the math test vs. kids scores on the reading test? No problem if the scores are on the same scale – like percentage correct or something. If the scores are on different scales – say raw scores out of a possible 20 on the math test and raw scores out of 1000 on the reading test, then you can either convert the scores to percentages or convert everything to a standard score like z-scores. This allows you to compare apples and oranges… at least compare them according to their central tendency. Crazy stuff here… but very useful! All statistics are based on these ideas of normal distribution and standard error. Anyway… good work if you've made it this far!