Paired tests

So far we have dealt with two sample tests where we assume that the two samples are independent of each other.

If we're careful, we can actually violate this assumption *AND* get more power at the same time!

So let's introduce *paired* tests.

In paired tests, you attempt to pair each observation in one sample with each observation in the other sample.

For example, you want to test the effect of a medicine that increases heart rates in mice.

You take 10 mice and give them the medicine, and put them in a cage.

You take another 10 mice, give them a placebo, and put them in another cage.

What is wrong with the above *design*?

Suppose something happens to one of the cages? Something that's not obvious (e.g., it gets colder, or someone forgets to feed the mice).

That could undermine your experiment. The results you observe might be due to some *other* environmental factor, not the medicine.

(For example: extreme cold might increase the heart rate in mice without medicine due to shivering).

Instead, it might be better to use 10 cages. Each cage gets a medicated mouse and a control mouse.

That way, if a cage is exposed to unusual conditions, it affects both mice equally.

This is called a *paired* design.

In a paired design, you deliberately match up each observation in one sample with each observation in the other sample.

In our little example, each medicated mouse is matched up with a control mouse kept under the *same* conditions.

Other examples of paired designs might include:

A before and after study (a classic example of a paired design): the same subject is measured before some kind of treatment, and again after the treatment. Obviously, each observation "before treatment" is matched with each observation "after treatment".

Determining if there is a difference in length between the index finger and the ring finger using 20 people: the index finger and ring finger of the same person are measured. Again, each observation of index finger is matched with each observation of ring finger.

By correctly pairing things, you can reduce the variation between subjects (e.g., between people). This means a paired test (done correctly) can give you more power.

So now that we know the basics of a paired test, how do you actually do a paired test?

The paired t-test:

You set up your hypothesis as usual:

$$H_0 : \mu_1 = \mu_2 \quad \text{or} \quad H_0 : \mu_d = 0, \quad \text{where} \quad \mu_d = \mu_1 - \mu_2$$

Then pick H_1 and α as appropriate.

Now calculate t^*. We have yet another formula for t^*:

$$t^* = t_s = \frac{d - 0}{SE_{\bar{d}}} \quad \text{where} \quad SE_{\bar{d}} = \frac{s_d}{\sqrt{n_d}}$$

The d refers to *differences* between the observation.

Basically, you calculate all the differences between your observations, and then perform a one sample t-test on the differences.

Finally, we compare $|t^*|$ with t_{table} as usual, using $d.f. = n_d - 1$.

(Note the similarity to a one sample t-test; it's not a coincidence!)

So here's an example:

We want to evaluate the effects of atenolol (a beta blocker) on blood pressure. To simplify things, let's just worry about systolic pressure. We measure 6 patients before and after administration of atenolol and get the following (simulated) results for systolic pressure (in mm Hg):

	Before	After	Difference
	183	152	31
	205	182	23
	233	206	27
	189	165	24
	159	134	25
	212	188	24
$\bar{y}$	196.8	171.2	25.7
s	25.6	26.1	2.9

Notice how much more variability there is in the *Before* and *After* pressures. This is a classic example of a paired design, so we calculate the differences and notice that the variability has dropped considerably. So let's set up our hypotheses:

$$\begin{array}{lll} H_0 : \mu_1 = \mu_2 & \text{or we could do:} & H_0 : \mu_d = 0 \\ H_1 : \mu_1 \neq \mu_2 & \text{or:} & H_1 : \mu_d \neq 0 \end{array}$$

Decide on α. Let's use $\alpha = 0.05$.

Now we calculate t^*:

$$s_{\bar{d}} = \frac{2.9}{\sqrt{6}} = 1.1839$$

so we get:

$$t^* = \frac{25.7}{1.1839} = 21.4$$

So we look up $t_{\text{table}} = t_{0.05,5} = 2.571$.

And since $|t^*| = 21.4 \geq t_{0.05,5} = 2.571$, we *reject* H_0 and conclude that atenolol does lower the systolic pressure in our patients.

Let's make some concluding remarks on paired tests.

Figuring out if something is paired is not always easy.

If you're the one *designing* the experiment, then it should be pretty obvious - you're the one that's doing the pairing.

If you're looking at someone else's data and trying to figure out if the data are paired (which is what we're doing in this class), it's sometimes a bit trickier:

Is there any reason the observation in the first sample is next to the observation in the second sample (i.e., is there any reason the observations are in the same row)?

If the answer is yes, then you are probably dealing with paired data.

See some of the examples of paired data above.

Finally note the obvious: if $n_1 \neq n_2$ then obviously the data are not paired.

A paired test deliberately violates the assumption of independence between samples.

Done correctly, this actually gives you more power.

If you have paired data, a regular un-paired two sample test isn't valid anymore (it can give you much lower power).

What about the assumptions of the paired t-test?

As usual, we need to assume the data are random.

We also need to assume that the differences have a normal distribution: $d \sim N$.

Notice that we don't care about the distribution of sample 1 or sample 2, just the distribution of the differences.

As usual, if n_d gets to be reasonably large we can still use the paired t-test.

If n_d is small and the differences are not normal, there is a paired test similar to the Mann-Whitney U-test:

This is called the Wilcoxon rank sum text, but we don't have the time to study it. If you're interested, Wikipedia has an explanation.

Also note that the sign test (which we discussed briefly) might work, although it doesn't have a lot of power.