Two sample tests (part I):

We now look at two samples at a time, instead of one.

For example, we have men vs. women, or blood pressure in a control group vs. a medicated group.

In all cases, we're looking at two different groups, and we want to compare these groups to see if they are equal (actually not equal) for some measure.

Do men and women have the same level of some hormone (say insulin)?

Is the blood pressure the same in medicated and control groups?

(Note that both of the above are asking the question in terms of H_0).

There are actually a large number of different two sample tests. For now we'll stick with those that actually measure some quantity (later we may deal with two sample tests that deal with things like proportions).

Let's first introduce the classic *t*-test. *This test assumes that the variances are equal*. Your text starts with this test.

We have two samples, say heights of men and women.

We want to find out if there is a significant difference in the (true) average height of men and women.

The example is a bit silly, since we know there is a difference.

We now set up our hypotheses:

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$

Pick α as usual.

Then calculate t^*.

But now we have two $\bar{y}$'s to contend with. We also have two values for s.

So how do we combine it all?

It turns out that the following equation will give us t^*, which will have a *t*-distribution:

$$t^* = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} \quad where \quad s_p^2 = \frac{SS_1 + SS_2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Note that the last quantity given above is simply a weighted average of the two variances.

Also note that if $n_1 = n_2$, then $s_p^2 = \frac{s_1^2 + s_2^2}{2}$.

This is also called the "equal variance" t-test.

Some comments:

1) Note that the numerator is really $\bar{y}_1 - \mu_1 - (\bar{y}_2 - \mu_2)$, but since our H_0 states that $\mu_1 = \mu_2$, we can get rid of the μ's in the numerator.

2) The denominator is actually just an average of the two standard errors for each of the samples.

3) As given, t^* does have a t distribution.

Once we get our t^*, we compare this as usual to the value in the t-tables:

First we need to get our degrees of freedom - this is easy:

$d.f. = \nu = n_1 + n_2 - 2$

Then reject our H_0 if $|_t*| \geq t_{table}$

otherwise fail to reject.

Let's do an example (exercise 8.1, p.):

We want to know if there is a difference in serum cholesterol levels for male and female turtles:

First set up our hypotheses:

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$

Let's pick $\alpha = 0.05$

Here are our data and some preliminary summary statistics:

								$\bar{y}$	s
males:	220.1	218.6	229.6	228.8	222.0	224.1	226.5	224.2	4.255
females:	223.4	221.5	230.2	224.3	223.8	230.8		225.7	3.867

First we need to get s_p^2:

$$s_p^2 = \frac{6(4.255)^2 + 5(3.867)^2}{11} = 16.67$$

now we can do:

$$t^* = \frac{224.2 - 225.7}{\sqrt{\frac{16.67}{7} + \frac{16.67}{6}}} = -0.6268$$

And without even looking, we know that that's not significant. But let's look up our t_{table} anyway:

$|t^*| = 0.6268 \geq t_{table} = t_{0.05,11} = 2.201$, so we fail to reject.

If you want to do this with R, you can do the following (there are actually several different ways of doing this):

1) Read in the data:

```
males <- scan()
220.1  218.6  229.6  228.8  222.0  224.1  226.5
```
(*hit return twice*)

do the same for females

2) Do the t-test:

```
t.test(males,females,var.equal = TRUE)
```

Don't forget to include the `var.equal = TRUE` bit, or R will do a unequal variance t-test by default (also known as Welch's test).

The output should be fairly self explanatory.

As mentioned, this test assumes that $\sigma_1^2 = \sigma_2^2$ (i.e., that the variances are equal)

If this assumption is not true, you can get yourself in trouble using this test.

The assumption is not so important if $n_1 \approx n_2$ (the test will still do okay).

As a general rule, this test is only used if you're really sure about the equal variance assumption. It's a bit surprising that your text introduces two sample tests with this one.

So if you can't assume equal variances (which you usually shouldn't), what do you do?

Use a version of the t-test that does not assume equal variances. This is also known as Welch's test.

It follows the same basic outline as above, but now we calculate t^* a bit differently:

$$t^* = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Comments:

1) If $n_1 = n_2$ and/or $s_1^2 = s_2^2$, then the two formulas are actually the same

2) This only has an approximate t distribution, but it's a good approximation.

3) It does *much* better if the variances are really unequal.

You then compare the $|t^*|$ to t_{table} just as before. However, there is one major headache here.

Your degrees of freedom become much more difficult to calculate:

$$\nu = \frac{(SE_1^2 + SE_2^2)^2}{\frac{SE_1^4}{n_1 - 1} + \frac{SE_2^4}{n_2 - 1}}$$

This can be a real pain (obviously).

Let's use the same example as above and do Welch's test:

We'll use the same hypothesis and α.

So let's calculate t^*:

$$t^* = \frac{224.2 - 225.7}{\sqrt{\frac{18.10}{7} + \frac{14.95}{6}}} = -0.6318$$

But now we need to calculate our degrees of freedom before we can make our comparison:

Note that $SE_1 = 0.6078$ and $SE_2 = 0.6444$

So we plug all this into our equation:

$$\nu = \frac{(0.6078^2 + 0.6444^2)^2}{\frac{0.6078^4}{6} + \frac{0.6444^4}{5}} = 10.942$$

note: always (always!) round down to the nearest integer.

So we make our comparison:

$|t^*| = 0.6318 \geq t_{\text{table}} = t_{0.05,10} = 2.228$, so we fail to reject.

So we come to the same conclusion as last time.

If you want to use R, the procedure is the same as above, except now you can just do:

```
t.test(males,females)
```

Leave off the `var.equal = TRUE` statement - as mentioned, R defaults to Welch's test.

Essentially, you should always use this version of the *t*-test:

If the assumption of equal variances is true, the other version will be a bit more powerful.

If you're pretty sure the variances are equal, then, yes, go ahead and use the other test. It is better (more powerful).

If the assumption of equal variances is not true (and if n_1 and n_2 are very unequal), the other test can be atrocious. Really awful.

Welch's test will do much better.

Even if the variances are actually equal, this test is only a little bet worse than the other test (only has a little less power).

Concluding comments:

You can test to see if H_0: $\sigma_1 = \sigma_2$. There are two problems with this:

1) You are trying to prove the null hypothesis (which you can't do)

2) The tests are not very powerful.

But if you're interested, see 8.5 in your text.

One sided tests are done as described previously. Just make sure that if H_1: $\mu_1 > \mu_2$, that you verify that your data agree with this (i.e, make sure $\bar{y}_1 > \bar{y}_2$).

If you want to use R for a one sided test, you could just do:

```
t.test(males,females, alternative = "less")
```

Of course you could use `"greater"` as well.

Remember that you need to decide ahead of time if you should/can do a one sided test.

There are other versions of the *t* test (some use trimmed means, for instance), but these are by far the most used.

There are several assumptions of both of these *t*-tests:

1) Data are random	Without random data nothing is valid.
2) Data in each sample are independent	The data in the two samples should be independent. This means that being in one sample shouldn't influence being in the other sample in any way. *We can violate this assumption by doing paired testing* (if we're careful).
3) Data in each sample are normally distributed.	This applies to *each* sample. In other words, you need to verify the normal distribution assumption for each sample. There is a way to combine samples and do just one verification, but this involves residuals. We may discuss this at a later date. *If we violate the normal distribution assumption, other tests will probably do better (e.g, the Mann-Whitney U test)*.