Two sample tests in **R** are identical to one sample ones save for an addition of a second variable. The functions prop.test(), t.test() and wilcox.test() are all available.

A few twists exists. You may need to specify if the variances are assumed to be equal, and you might have matched or paired data.

For a test of proportion we have

> prop.test(c(y1,y2),c(n1.n2), alt="two.sided")

will do a two-tailed significance test for $p_1 = p_2$ where $\hat{p}_i = y_i/n_i$.

For a two-sample test of mean (or median) we can use any of

```
> t.test(x,y,alt="...")  # non-equal variances
> t.test(x,y,var.eq=T, alt="...") # assume equal variances
> wilcox.test(x,y,alt="...")  # wilcox assume x_ij from f(x - mu_i)
```

For reference, if \mathbf{x} and \mathbf{y} are two samples we have

For this project, I would like you to write up your observations as you go along. You can do this with notepad, word, in the terminal ... At the end of the lab, you should turn something in for me to look at. I'm happy to answer questions, but I'll let this be as self-paced as possible. Feel free to work in groups if desired.

These come from the 4/9/2002 New York Times Science Times.

Example Mammograms:

The HIP study begun in New York in the 1960's randomly assigned 60,000 women to two groups – one that received mammograms, and one that didn't. The women were then observed for the next 18 years. Of the 30,000 who had mammograms 153 died, of the 30000 who did not, 196 died.

Let p_1 be the probability of dying of breast cancer if you had a mammogram and p_2 be the probability of dying if you had no mammogram. A statistical test could be $H_0: p_1 = p_2$ against $H_A: p_1 < p_2$.

Carry out the statistical test using the function prop.test.

What is the p-value? Would you reject the null hypothesis at the 0.05 level?

Based on this study (and others) the mammogram became a common tool in preventing breast cancer.

There were possible problems though. In the course of the 18 years some women were dropped from the study -800 who received mammograms and 300 who did not because these women already had cancer. Critics wonder why the imbalance.

Suppose the probability of being dropped is p_i with p_1 the probability a women who had a mammogram was dropped. Based on the data test the hypothesis $H_0: p_1 = p_2$ against the alternative $H_A: p_1 > p_2$. Do you reject or accept the null hypothesis?

Critics argue that if only 10% of the 500 additional women from the mammogram group were dropped then the number of deaths would be similar. Suppose, if fact the true percent was 5%. To be specific, suppose that of the 1100 people dropped from the study, 5% died from breast

cancer evenly from each group. Repeat the hypothesis test. Do you now accept or reject the null hypothesis? What is the *p*-value?

Would you estimate the 5% from the original proportions or some other?

Can you explain why it is likely that many more women were dropped from the mammogram group than from the control group for having breast cancer?

Example Dr. Peter Gotzsce:

An interesting article about this doctor also appeared. He has earned a reputation for critically examining past statistical studies and puncturing the results. The article lists work he has done on the placebo effect, mammograms, AZT treatments and bias in double-blind trials.

The study of AIDS drugs (AZT) was based on the amount of the drug taken. In particular the recommended dose of 1200 mgs was tested against the amount 400 mgs. The latter having fewer side effects and being cheaper.

Suppose a test to decide is done as follows, a numeric measurement of a patients health is made. The differences are recorded after a treatment.

Let 20 AIDS patients be randomly assigned to 2 equal-sized groups. Group 1 received a does of 1200 milligrams, Group 2 400. The data on differences is as follows

Group 1: 2 1 5 1 5 7 -1 8 7 3 Group 2: 7 0 8 1 10 12 2 9 5 2

We want to do a *t*-test of the data. This test requires normal data. Do a side-by-side boxplot of the data to investigate normality. (What does a side-by-side boxplot actually show?) If the data seem normal, the *t*-test is also better if the variances are equal. Does your side-by-side boxplot confirm this? If yes to all the above, then the *t*-test can be done with a command like

```
t.test(x,y,var.equal=T,alt="...")
```

Do a test to see if μ_1 is more than μ_2 . What are the null- and alternative hypotheses. What is the *p*-value. Do you accept or reject at the 0.10 level?

Suppose GlaxoSmithKline disputes the findings, saying it is better to measure lifespans for some measurable quantity. You then run a test for this with 22 people, 12 randomly chosen for a 1200 mg group (group 1) and 12 for a 400 mg group (group 2). The data is now given by

Group 1: 10 25 42 153 472 59 25 201 36 110 Group 2: 159 200 65 147 104 63 265 94 286 97

Again, we want to test the difference. Confirm that the data do not appear to be from a normal distribution. This rules out a t-test. Lets try the a Wilcoxen rank-sum test. Details will come in class about what the test does, but it is a test of the hypotheses

 H_0 : same center H_A : different centers

The data should have the same shape distribution, but does not need to be normal. Verify this is the case for the data we have.

To perform the test is easy. It is done like

wilcox.test(x,y,alt="...")

Perform the test, do we reject at the 0.05 level?

Example Brachytherapy:

In treating Breast Cancer new treatments are being developed which recognize the fact the "When we look at the medical literature we find that the vast majority of recurrences are right around the original tumor site" (Dr. Robert Kuske). As such they are developing local treatments, one such is brachytherapy which inserts a radioactive seed inside the tumor cavity.

One study of 260 patients treated this way reports that only 1 patient had a relapse in 11 years. Whereas the average relapse rate in similar Stage 1 and 2 patients is 10-12 percent in 10 years.

Test the hypothesis that this treatment is better than what is used for the average treatment. What are you hypotheses? What is your control? What kind of study would this be (hint section 1 in chapter 12.)

Example Telling when it really matters:

The thrust of several articles is the question of whether new tests for cancer are too sensitive (mammography, PSA), in that people are being diagnosed and subsequently treated for cancers that might never have caused a problem. (The side effects can be quite severe.) What is needed are methods to test if cancers are truly threatening.

Currently one test involves measuring a tumor's thickness to predict whether it will spread. A new attempt to look at genetic markers based on a new tool called a DNA chip is supposed to improve matters.

Suppose the following data was recorded. A group of 50 people were randomly assigned to two groups. The first group was tested based on thickness and 20 of 25 of the diagnoses proved to be correct. The second group was tested with the DNA chip. It had 23 out of 25 accurately predicted. Is there evidence at the 0.05 level to support the prediction rate is better for the new test? If not, how large must $n = n_1 + n_2$ be so that a test of size n can reject H_0 at the 0.05 level if the sample percentages were the same?

Example Do experts always agree?:

Experts may not always agree. For example, a group of 8 leading pathologists were given identical slides for 37 patients. From these they were to identify the risk of cancer. In the real study the doctor agreed on only 11 of the 37 patients.

To study this numerically, lets suppose that two experts were given 10 identical slides, and they were asked to rate them on a scale of 1 to 100 as to the chance of the slide sample being cancerous. They gave the following ratings:

Expert 1: 59 61 76 86 87 92 73 85 63 86 Expert 2: 58 71 78 87 97 90 80 89 68 89 Do a side by side boxplot. Do they seem to agree? Do a *t*-test if appropriate. Do you conclude that the two seem to be the same? Should this really be tested as a "paired" or "matched" sample? If so, do a t-test using this. Do you get the same results? Explain.