One way ANOVA when the data are not normally distributed (The Kruskal-Wallis test).

Suppose you have a one way design, and want to do an ANOVA, but discover that your data are seriously not normal?

Just like with the MWU test as "replacement" for the t-test, there is the Kruskal-Wallis test for a one way ANOVA.

In fact, if you have just two groups, the KW test will give you the identical results to a MWU test (sort of like a ANOVA gives the same result as a equal variance t-test).

Before we look at the KW test, note that there are other things that we can do:

Remember transformations. They can make non-normal data normal.

That's probably the only way to deal with more complicated designs. Things like a two way ANOVA, nested designs, etc., don't have non-parametric equivalents (well, there is one for two way designs... more on that later).

Check out your text in chapter 13 for some ideas.

Let's introduce the KW test:

Just like the MWU test, if make no assumptions other than randomness, we can use the KW test pretty much anytime. It has no other assumptions we need to worry about.

But what are we testing now?

H_0: All k population distributions are identical

H_1: At least one of these is different (tends to yield larger/smaller observations than the others).

Just like last time, though, if we assume distributions are similar except in location, we can use means (or medians) instead.

Then proceed as usual:

- select α

- calculate test statistic (let's call it W^*)

- compare it to a value from KW tables

- if $W^* \geq$ to W_{table} (but see below), then reject.

So, what is W^*?

It's a bit messy, and is given by (note I'm doing things just a bit differently from the text):

$$W^* = \frac{1}{S^2}\left(\frac{\sum_{i=1}^{k} R_i^2}{n_i} - \frac{N(N+1)^2}{4}\right)$$

(Your text (on page 217) uses a slightly different formula).

Yech! But let's not give up yet. It's really not that difficult. Let's figure out what the various parts of this are:

N = total sample size (= n^*, using our ANOVA notation)

R_i = sum of the ranks of the i^{th} sample:

$$R_i = \sum_{j=1}^{n_i} R(X_{ij})$$

This says, take the rank of each of the x's and sum these values for each sample (this is very similar to the way we did things in the MWU test (in this class)).

Notice that the very first thing we'll have to do is to rank our observations from smallest to largest (write in the rank next to our each of our observations).

S^2 = an analogue of the variance (notice this is capitalized). It's given as follows:

$$S^2 = \frac{1}{N-1}\left(\sum_{\text{all ranks}} R(X_{ij})^2 - N\frac{(N+1)^2}{4}\right)$$

Okay, now you're really worried. But let's stick with it for just a little longer.

This says, Sum up all the square of each of the ranks, then subtract the other quantity (the one involving N). That's not too bad.

Now we know how to calculate W^* (it'll be much more obvious after an example). But what about W?

- This is a bit of a problem. Remember how you needed to have n and n' (n_1 and n_2 in our class) in the Mann-Whitney U-test?

  - now we have 2, 3, 4 ... or more samples, so we really can't list the probabilities for everything.

- Let's take another quick look at W^*. It turns out if there are no ties, we can re-write it as:

$$W^* = \frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(N+1)$$

This is identical to the formula in your text for "no ties".

(If there are no ties, then use this formula - it's a lot easier)

The stuff in the sum symbol is kind of analogous to what we "observed" -we're using the actual ranks and getting a kind of sum (squared and adjusted by n_i, but still a sum).

The 3(N+1) term is kind of what we would expect for the sums if the ranks were basically equal in each sample.

The term in front of the sum symbol is another type of "expected" value. {Note: the sum of a bunch of numbers going from 1...n = $n(n+1)/2$, so you might recognize that both 3(N+1) and the quantity out front are a little similar)}.

So we have an observed quantity (-) an expected quantity, where the observed is also divided by an expected quantity. Sound just a little familiar?

As it turns out W^* has (approximately) the Chi-square distribution.

So what do you look up for W? A Chi-square value with k-1 degrees of freedom.

Your text takes a different approach, and for small sample sizes and less than four groups gives you the exact values in table B.13. We'll just stick with the Chi-square approach since it's much easier.

As mentioned, this is an approximation, and exact tables do exist, but they take up a good portion of a book, and your text only does this for small n and k up to 4.

So how does it all work?? An example (let's use an example similar to the pig diet example, but this time we're looking at sheep (it's the same example the 214 text uses to introduce one way ANOVA):

Let's do H_0: the diets are all the same.
H_1: at least one of the diets is different.
$\alpha = .05$

	Diet 1			Diet 2			Diet 3		
	value	rank	rank^2	value	rank	rank^2	value	rank	rank^2
	8	2	4	9	3.5	12.25	15	7	49
	16	8.5	72.25	16	8.5	72.25	10	5	25
	9	3.5	12.25	21	12	144	17	10	100
				11	6	36	6	1	1
				18	11	121			
Sum (R_i)		14	88.5		41	385		23	175
n_i		3			5			4	

Notice that $N = n^* = 12$; also notice that the highest rank is 12, so everything's fine so far (if one or more values are tied, you need to use the average rank).

Now we calculate S^2:

$$S^2 = \frac{1}{11}\left(649 - 12\frac{13^2}{4}\right) = 12.909$$

Now we "merely" plug all this into W^* to get:

$$W^* = \frac{1}{12.909}\left(\frac{14^2}{3} + \frac{41^2}{5} + \frac{23^2}{4} - 12\frac{(13^2)}{4}\right) = 2.075$$

Let's get W_{table}, using our Chi-square table and 2 *d.f.*:

$W_{\text{table}} = 5.99$

And since $W^* < W_{\text{table}}$, we "fail to reject" and conclude we have no evidence to show the diets are different.

Let's summarize:

When do you use KW? When you don't meet the assumptions of (one way) ANOVA:

As usual, with a larger sample size, ANOVA will start to do better, and you don't need to worry about the normal assumption as much.

What about equal variances? If you're worried about unequal variances, don't test for equal means. Use the general form of the KW test.

How about power? Pretty good; even when the data are normal it doesn't do too badly.

Why not use it all the time?

As you probably know by now, ANOVA is much more flexible. The number of designs available for ANOVA is almost endless.

And even though the power isn't bad, it isn't the best test to use if data are normal.

Multiple comparisons are available for the KW test, but we just don't have time to discuss this as well.

We didn't do too much with the theory here. The basic idea is very similar to that of the Mann-Whitney U test, the only odd thing is that we can get away with using Chi-square tables instead of exact tables.

Doing Kruskal-Wallis in R:

The basic setup is almost identical to a one way ANOVA. Here's the above diet example in R:

```
gain <- scan()
8 16 9 9 16 21 11 18 15 10 17 6
```

```
diet <- scan()
1 1 1 2 2 2 2 2 3 3 3 3

diet <- factor(diet)

kruskal.test(gain ~ diet)


        Kruskal-Wallis rank sum test

data:  gain by diet
Kruskal-Wallis chi-squared = 2.0748, df = 2, p-value = 0.3544
```

(R calls the test statistics “Kruskal-Wallis chi-squared”)

Finally, what about two way or more complicated designs?

The Kruskal-Wallis test is designed for one-way type analyses.

There are other non-parametric tests for more complicated designs; if you need to use something like this, you can look up the “Quade” test or the “Friedman” test (the Friedman test is related to the sign test).

The Friedman test works for two way designs, but isn't very powerful.

The Quade test works for blocked designs, and is worth looking at if you have serious problems with normality and are using a blocked design.