Supplement on the Kruskal-Wallis test

So what do you do if you don’t meet the assumptions of an ANOVA?

{There are other ways of dealing with things like unequal variances and non-normal data, but we won’t learn them here. Most of these require transformations, and transformations come with a whole set of problems that we don’t have time to deal with. See 2.7 [2.7] particularly the section on “non-linear transformations to get an idea of what kind of transformations we mean. The text doesn’t really point out what some of the problems are when using transformations.}

As long as the data are still random, we can employ what’s called the Kruskal-Wallis test. The only assumption we still need to worry about is “randomness”.

But what are we testing now?

- H0: All k population distributions are identical
- H1: At least one of these is different (tends to yield larger/smaller observations than the others).

- 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* ≥ to W table (but see below), then reject.

So, what is W*?

It’s a bit messy, and is given by:

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

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 \( K_1 \) and \( K_2 \) in the Mann-Whitney test, except now we’re using ranks instead of “number smaller in other sample”)

- 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. It’s actually very similar to the “calculator” formula for variance that we didn’t have a chance to talk about (but are in my notes).

- 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 \( n' \) 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:

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

- As mentioned, this is an approximation, and exact tables do exist, but they take up a good portion of a book! Minitab seems to do approximation. (More sophisticated software like SAS has an option for getting the exact values of \( W \)).

- So how does it all work?? An example (let’s use the Sheep example, but this time we’ll do a KW test instead of an ANOVA)

- let’s do
  H0: the diets are all the same.
  H1: at least one of the diets is different.
  \( \alpha = .05 \)

<table>
<thead>
<tr>
<th>Diet 1</th>
<th>Diet 2</th>
<th>Diet 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>value</td>
<td>rank</td>
<td>rank(^2)</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>16</td>
<td>8.5</td>
<td>72.25</td>
</tr>
<tr>
<td>9</td>
<td>3.5</td>
<td>12.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum (( R_i ))</td>
<td>14</td>
<td>88.5</td>
</tr>
<tr>
<td>( n_i )</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>
- 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 \( W_{\text{table}} \), 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.

(Incidentally, the same result we got with ANOVA)

- Okay, let’s summarize:

- When do you use KW? When you don’t meet the assumptions of 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? Well, use a little common sense. Don’t just assume they’re unequal unless you only have two categories (then use a t-test or Mann-Whitney!). This is kind of the opposite to what was said before. If you’re worried about them being seriously unequal, use a KW test, but be aware that to use the KW test for means (or medians), it too, assumes equal variances.

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

  - Why not use it all the time?

    - ANOVA is much more flexible. As mentioned, 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 dig into that 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.

- The easy way to do this (as usual) is to use a computer package; for example, here's the printout from R:

R:

```
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 two-way “non-parametric” tests; if you need to use something like this, you can look up the “Quade” test of the “Friedman” test (the Friedman test is related to the sign test).

  - A “non-parametric” test is a test that doesn't assume a particular distribution for the date. The sign test, the Mann-Whitney test and the Kruskal-Wallis test are three examples that we've talked about. We'll discuss at least one more.