Lecture 19

Non-parametric ANOVA

In the previous lecture, we looked at randomization tests, or permutation tests for comparing two samples. We ca extend the analysis to three or more samples, but now, there are many more possible combinations to consider. We'll look at the logic of the test procedure. To run this test even for this situation should involve computational tools to automate the many calculations that are required.

Let's consider a small sample case where we have 4 observations for three treatments. Since this is equivalent to an ANOVA test, we can use the F-statistic that we use for regular ANOVA. We can already see that even for a small sample, our test statistic calculation is involved. Suppose we come up with a test statistic like 2.51.

Then we need to randomize the data from all three samples into one set, and select 4 observations to be treatment one, and four more to be treatment 2, and the remaining ones to be treatment 3. Then calculate the F statistic on this new version of the data.

We would need to repeat this process for every possible combination of three treatment groups selected randomly. Because there are so many possible situations, it may be a little easier to think about the situation in terms of ordering. If we care about the ordering of groups, there are 6 ways to reorganize the data to obtain the same F-statistic that we started with.

{[1,2,3], [1,3,2], [2,1,3], [2,3,1], [3,1,2], [3,2,1]}

The number here is equal to 3! Because there are three treatments.

To obtain the P-value, we will need to determine how many arrangements of the data produce an F-statistic that is higher than the one we started with. Each possible moving around of observations that produces a higher F-statistic has the same six arrangements as above. This will be the numerator in the probability.

Finally, we would need to know how many possible permutations there are of all the data, whether the F-statistic is higher or not. The formula for this $(n!)^k$ where n is the common sample size, and k is the number of treatments. Thus, here that is $(4!)^3 = 13{,}824$. That will become the denominator of our probability.

The equivalent procedure if you have the computing power is to create all possible permutations of all 12 observations in this case, and then compute the F-statistic on all of those. There will be a lot of duplication (since within the treatments the order doesn't matter and permutations treat the order like it does matter), and the number of computations gets very large very fast, but it will get you to the same outcome. There are some many possible outcomes that sampling procedures are much more efficient.

This procedure is computationally intensive, but does not depend on the observations in each treatment being distributed in any particular way.

We have other methods for computing non-parametric ANOVAs that relax the requirements for normality. One such method is the Kruskal-Wallis test, that like the Wilcoxon test, uses rank to test for equality of means.

Kruskal Wallis

Like the two-sample Wilcoxon test, all the data in this test is pooled into a single sample to be ranked. Let's call N the total number of observations in all the samples. If all the samples have similar values, then the expected value of each sample's rank is $E(R) = \frac{N+1}{2}$ (the middle rank).

This test measures how much the samples deviate from this value. The test statistic K can be expressed as

$$K = \frac{12}{N(N+1)} \sum_{i=1}^{I} J_i \left(\bar{R}_i - \frac{N+1}{2} \right)^2$$

Where J_i are the number of observations in each sample (treatment group).

Calculating this formula using the mean rank of the treatments and the squared expression is more computationally expensive than a simplified expression.

$$K = \frac{12}{N(N+1)} \sum_{i=1}^{I} \frac{R_{i\cdot}^2}{J_i} - 3(N+1)$$

Where $R_{i\cdot}^2$ is the total ranks not the average.

It turns out that when the sample sizes are "large enough" the distribution of K is χ^2. This is true if there are 3 treatments when the number of observations in each sample is greater than 6, or if there are more than three treatments, when the number of observations in each treatment is greater than 5. The χ^2 distribution would use degrees of freedom equal to one less than the number of treatments.

To run the Kruskal-Wallis test, use the function Kruskal.test() with basically the same syntax as the aov() function.

When we run this on the mtcars data using cyl (as a factor) as the grouping variable for mpg, we get the following output.

```
        Kruskal-Wallis rank sum test

data:  mpg by cyl
Kruskal-Wallis chi-squared = 25.746, df = 2, p-value = 2.566e-06
```

This is a similar outcome to what we obtained when we test this same data with the regular ANOVA function. Which we would expect since the data is approximately normal. The biggest differences will arise when the data is not very normal.

When we did traditional ANOVA, we had the option to use pairwise t-tests to see how the data might be grouped, or Tukey's method. The pairwise.wilcox.test() allows us to use the non-parametric equivalent of the t-test to see the groupings. A table below shows the pairwise P-values for differences.

Pairwise comparisons using Wilcoxon rank sum test with continuity correction

data: mtcars$mpg and mtcars$cyl

	4	6
6	0.001	-
8	8.3e-05	0.001

P value adjustment method: BH

Documentation for the pairwise test can provide you with additional options for this test.

Friedman's ANOVA

Friedman's ANOVA is the non-parametric alternative to the Repeated Measures ANOVA. Like the previous test for one-way ANOVA, this test assumes the observations are independent and from a continuous distribution, and the sample sizes are sufficiently large (i.e. bigger than 5 per condition).

Let's consider an example:

A random sample of men undergo an exercise program to see if it lowers their cholesterol levels. Data is collected at month a, month b, and month c, in order. Cholesterol levels are not normally distributed so regular ANOVA is not useful. Our null hypothesis is that the treatment has no effect.

We could add a further factor to our study by looking at both men and women, or by looking at different types of exercise programs.

It is accessed in R using the friedman.test() function. You will need to specify a variable containing the measurements, a group (treatment) variable and the blocking variable. The mechanics of the test follows the model of the Kruskal-Wallis test using ranks and constructing a similar test statistic. We obtain a test statistic by the formula

$$F = \left(\frac{12}{Nk(k+1)}\right)\sum_{i=1}^{I} R_i^2 - 3Nk(k+1)$$

Here N is the number of subjects, k is the number of conditions, R is the sum of the ranks in each condition.

The test statistic obtained is distributed by the χ^2 distribution. As with Kruskal-Wallis, the degrees of freedom is the number of conditions minus one.

It's worth noting that the test will not execute (it will produce an error), if we try to run the test without a complete block design (there must be observations in every combination of the group and blocking variables and a consistent number of observations).

If we find we are able to reject the null hypothesis, then there are several options for post-test analyses to detect differences between pairs of conditions. Earlier, we saw the example of using pairwise

Wilcoxon tests for this purpose. A link in the references below provides some additional options and R packages in which to find them.

We can use fictional data to test the function and look at the output.

```
        Friedman rank sum test

data:  data$score, data$drug and data$person
Friedman chi-squared = 13.56, df = 3, p-value = 0.00357
```

When we come back from break we'll look at computational approaches through various simulations using permutation test sampling method, and bootstrapping.

References:

1. https://faculty.ksu.edu.sa/sites/default/files/probability_and_statistics_for_engineering_and_the_sciences.pdf
2. https://onlinestatbook.com/2/distribution_free_tests/randomization_two_or_more.html
3. https://onlinestatbook.com/2/distribution_free_tests/rank_two_or_more.html
4. https://www.mathsisfun.com/combinatorics/combinations-permutations-calculator.html
5. http://www.sthda.com/english/wiki/kruskal-wallis-test-in-r
6. https://www.statology.org/friedman-test-r/
7. https://www.statisticshowto.com/friedmans-test/
8. https://www.statstest.com/friedman-test/
9. https://psych.unl.edu/psycrs/handcomp/hcfried.PDF
10. https://rcompanion.org/handbook/F_10.html