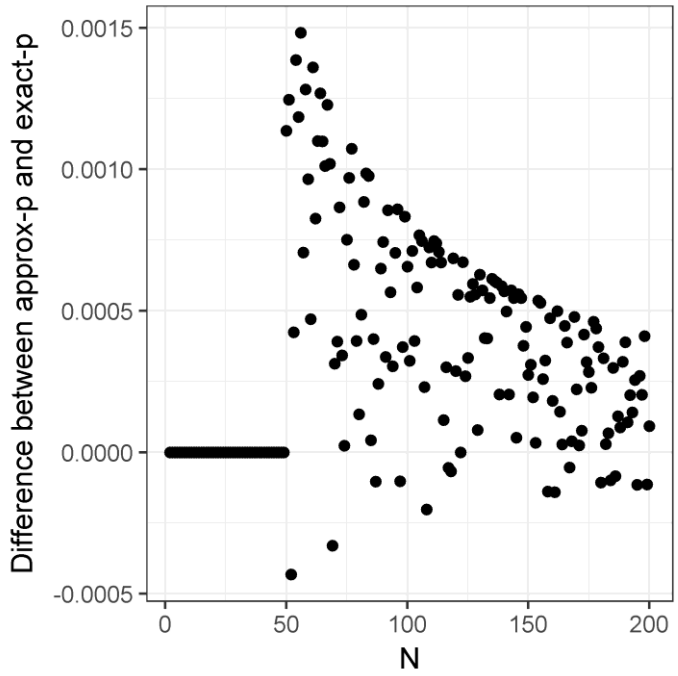
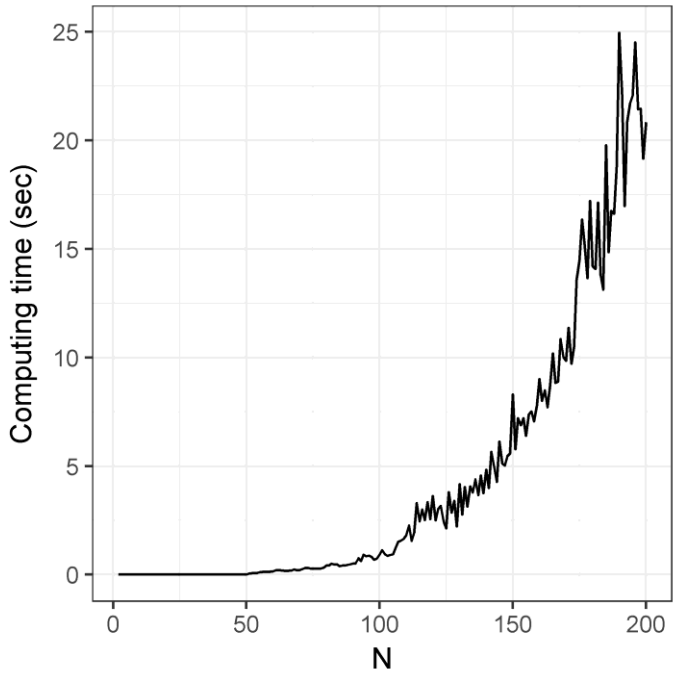
RenR 480/711

Non-parametric statistics

Permutational Tests

Wilcoxon test – approximated versus exact p-values





→ No differences for less than 50 samples (exact method used)
→ Neglectable differences for more than 50 samples
→ Computing time increases exponentially

Non-parametric tests

Assumptions:

Non-parametric tests do not assume:

- Normality
- Equal variances
- No large outliers

Important is, however: (as for all statistical tests)

- Independence of samples
- Replication
- Randomization

Permutational tests

Example

ID	FARM	VARIETY	YIELD	Distance within	Distance between
1	Farm1	A	720	20	203
2	Farm1	A	740	30	
3	Farm1	A	690	50	
5	Farm1	B	515	35	199
6	Farm1	B	480	30	
7	Farm1	B	545	85	
9	Farm1	C	540	38	4
10	Farm1	C	502	30	
11	Farm1	C	510	8	

$$SS\ Distance_{within} = \frac{1}{n}\sum_{i}^{n\ Groups}\sum_{i}^{n}(Distance_i^2)$$

$$SS\ Distance_{within} = \frac{1}{3}(20^2+30^2+50^2)+\frac{1}{3}(35^2+30^2+85^2)+\frac{1}{3}(38^2+30^2+8^2)$$

$$SS\ Distance_{within} = 4186$$

$$SS\ Distance_{between} = \sum_{i}^{n}(Distance_i^2)$$

$$SS\ Distance_{between} = 203^2+199^2+4^2$$

$$SS\ Distance_{between} = 81094$$

$$D = \frac{signal}{noise} = \frac{\frac{SS\ distance_{between}}{n\ of\ groups-1}}{\frac{SS\ distance_{within}}{n\ of\ observations - n\ of\ groups}}$$

$$D = \frac{\frac{81094}{3-1}}{\frac{4186}{9-3}} = 58.369$$

Permutational tests

Permutation #1
Shuffling of the labels in column VARIETY

ID	FARM	VARIETY	YIELD	Distance within	Distance between
1	Farm1	A	720		
2	Farm1	C	740		
3	Farm1	C	690		
5	Farm1	B	515		
6	Farm1	C	480		
7	Farm1	B	545		
9	Farm1	B	540		
10	Farm1	A	502		
11	Farm1	A	510		

Different distances within and between

$$SS\ Distance_{within} = \frac{1}{n}\sum_i^{n\ Groups}\sum_i^n(Distance_i^2)$$

$$SS\ Distance_{within} = \frac{1}{3}(20^2+30^2+50^2)+\frac{1}{3}(35^2+30^2+85^2)+\frac{1}{3}(38^2+30^2+8^2)$$

$$SS\ Distance_{within} = 4186$$

$$SS\ Distance_{between} = \sum_i^n(Distance_i^2)$$

$$SS\ Distance_{between} = 203^2+199^2+4^2$$

$$SS\ Distance_{between} = 81094$$

Different D

$$D = \frac{signal}{noise} = \frac{\frac{SS\ distance_{between}}{n\ of\ groups-1}}{\frac{SS\ distance_{within}}{n\ of\ observations - n\ of\ groups}}$$

$$D = \frac{\frac{81094}{3-1}}{\frac{4186}{9-3}} = 58.369$$

Permutational tests

Test statistic based on distances

Calculation steps:

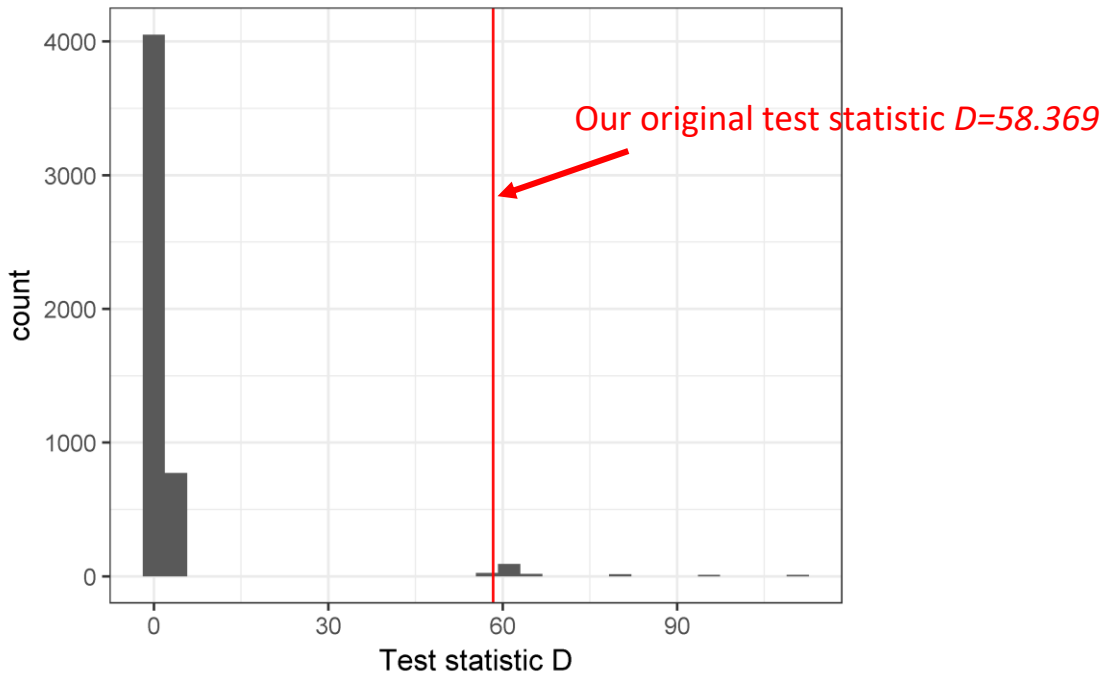
1. Calculate distances between and within groups
2. Calculate sum of squares based on distances
3. Calculate a test statistic based on your data:

$$\text{Delta} = \frac{signal}{noise} = \frac{distance_{between}}{distance_{within}}$$

4. Shuffle (permutate) the variable names and calculate the test statistic
5. Build a distribution by repeating 5000 times (your choice)
6. Compare your test statistic to this distribution

Permutational tests

Repeat 5000 times to build a distribution



How many times did we get a value of D larger than our test statistic?
Count = 174 times

What is the probability of getting a value of D larger than our test statistic?

174 out of 5000 times
= 174/5000
= 0.0348 → p-value

The shape of the distribution does not matter!