

Multiple Response Permutation Procedures

Objective: Calculate if there is a significant difference between groups in a multivariate space

Useful for multivariate data that does not meet the assumptions of MANOVA (e.g. Normality and Equal Variances for each variable)

MRPP make **NO Assumptions** therefore any numeric data can be used

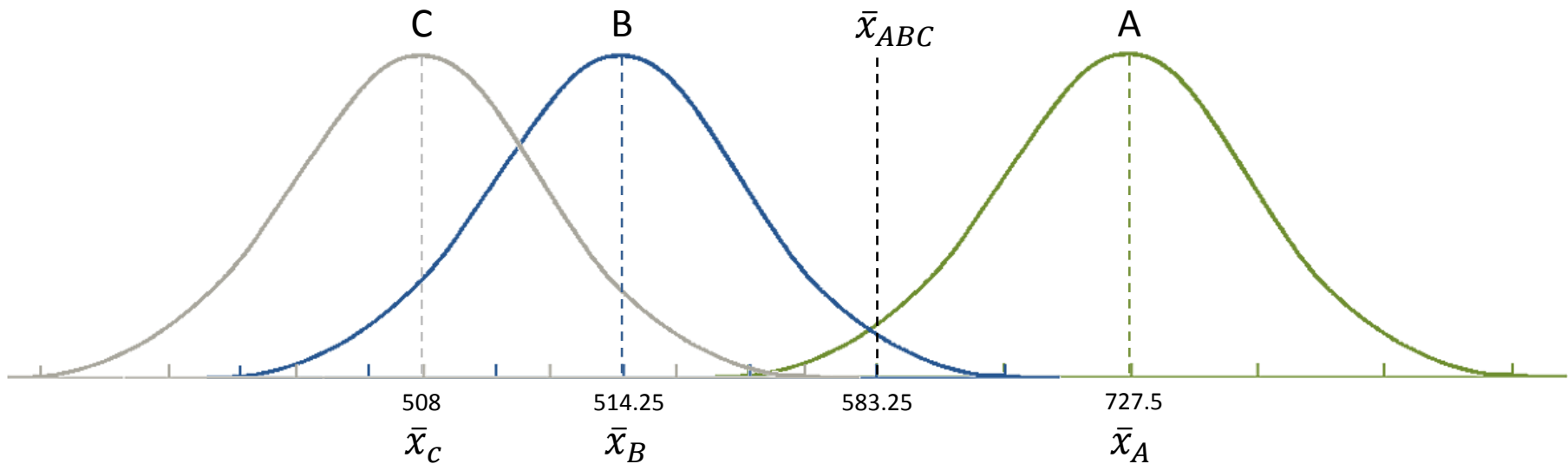
However the assumptions of independence (spatial & temporal) and design considerations (randomization, sufficient replicates, no pseudoreplication) should still be upheld – *good statistical practice!*

MRPP work with absolute differences (we call them distances) where smaller values indicate similarity

Makes the calculations equivalent to sum-of-squares (used in ANOVA)

Consider Univariate ANOVA

Used when you have 3 or more samples



$$H_0: \mu_A = \mu_B = \mu_C$$

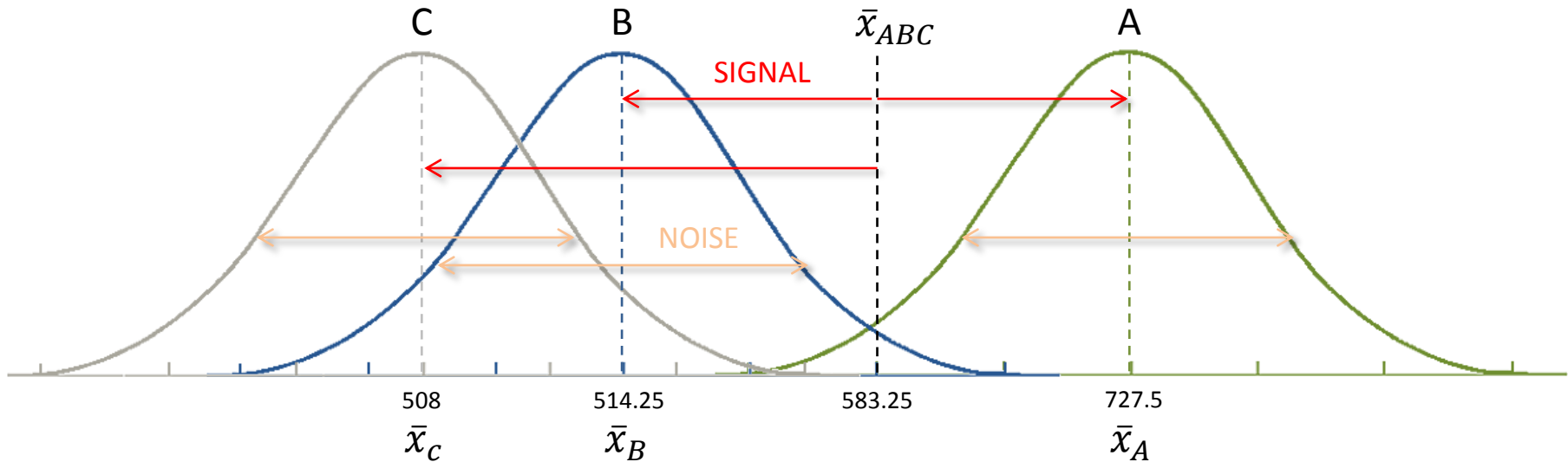
$$H_a: \mu_A \neq \mu_B \neq \mu_C$$

The alternative could be true because all the means are different or just one of them is different than the others

If we reject the null hypothesis we need to perform some further analysis to draw conclusions about which population means differ from the others and by how much

Consider Univariate ANOVA

Used when you have 3 or more samples



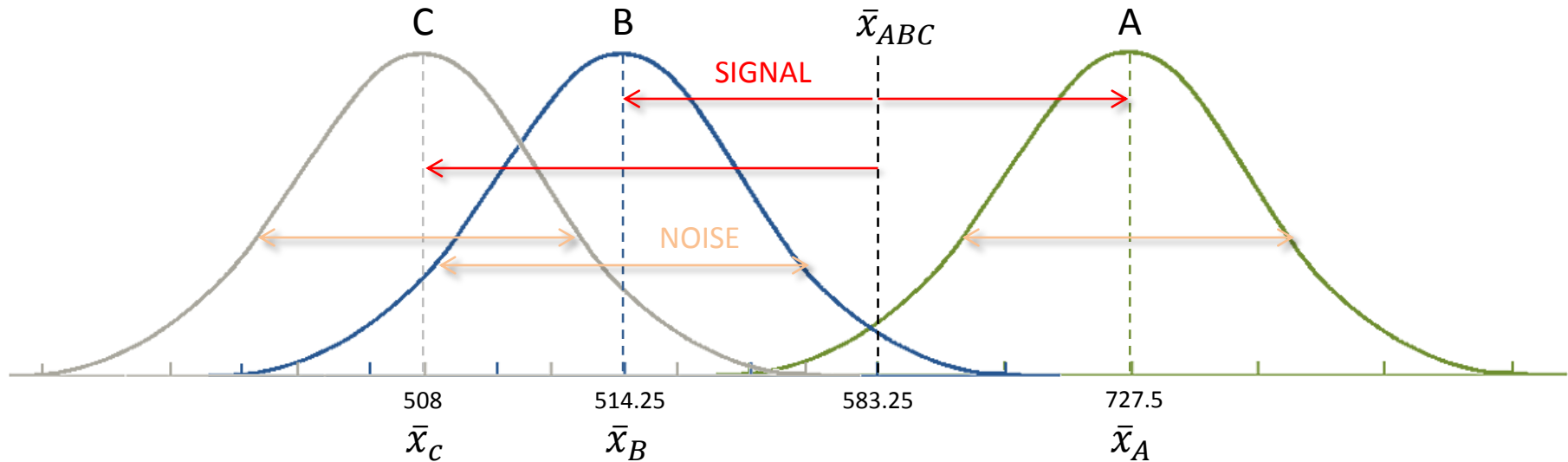
$$F = \frac{\text{signal}}{\text{noise}} \quad F = \frac{\text{variance between}}{\text{variance within}}$$

$$\text{variance between} = \frac{\sum_i^n (\bar{x}_i - \bar{x}_{ALL})^2}{n - 1} * r \quad \text{variance within} = \frac{\sum_i^n \text{variance}_i}{n}$$

A large F-value indicates a significant difference

Consider Univariate ANOVA

Used when you have 3 or more samples



$$\text{variance between} = \frac{\sum_i^{A,B,C} (\bar{x}_i - \bar{x}_{ABC})^2}{3 - 1} * 4 = \frac{(727.5 - 583.25)^2 + (514.25 - 583.25)^2 + (508 - 583.25)^2}{2} * 4$$
$$= \mathbf{62463.25}$$

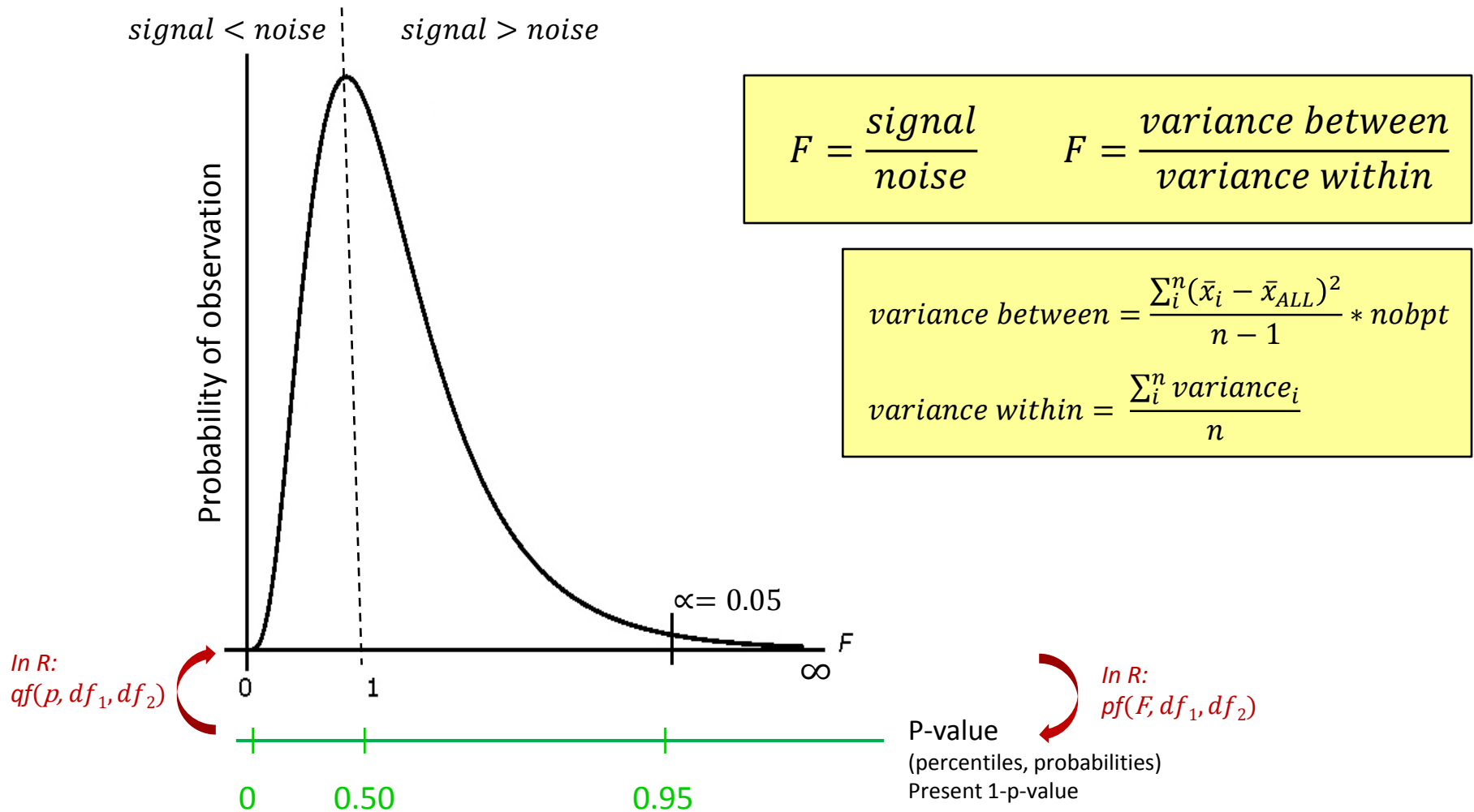
$$\text{variance within} = \frac{\text{var}_A + \text{var}_B + \text{var}_C}{3} = \frac{891.6667 + 819.3333 + 305.5833}{3}$$
$$= \mathbf{672.1943}$$

$$F = \frac{\text{variance between}}{\text{variance within}} = \frac{62463.25}{672.1943} = \mathbf{92.92439}$$

One-way ANOVA in R:

```
anova(lm(YIELD~VARIETY))
```

F-Distribution (family of distributions- shape is dependent on degrees of freedom)



The larger the F-value the further into the tail – AND the smaller the probability that the calculated F-value was found by chance, MEANING there is a high probability that something is causing a significant difference between the groups

The math behind MRPP

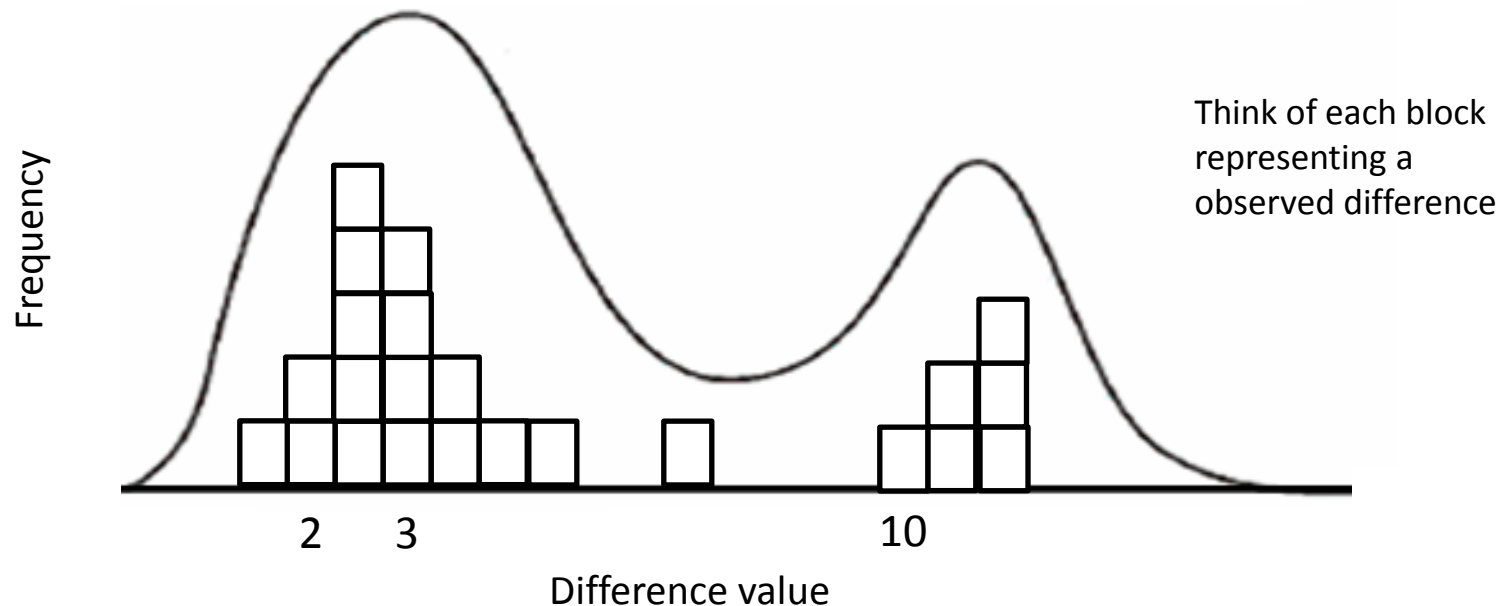
$$D = \frac{\text{signal}}{\text{noise}} \quad D = \frac{\text{distance between groups}}{\text{distance within groups}}$$

MRPP calculates distances between all observations within each group and generates a weighted average of distances (weighted by the number of observations within each group).

MRPP generates noise by randomly shuffling the class variables within the dataset. After shuffling, the weighted average of distances within the random groups are recalculated.

This is equivalent to “noise”

Reshuffling (permutation procedure) is repeated until you get a distribution of average distances



The math behind MRPP

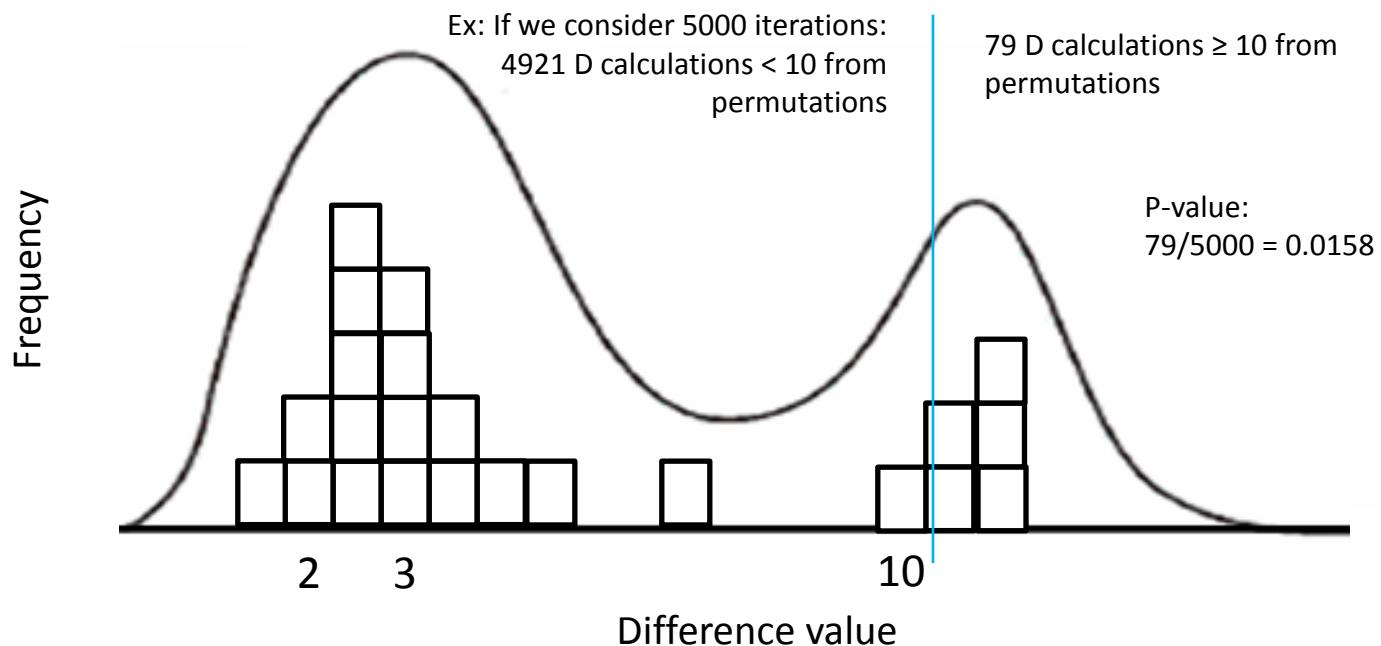
$$D = \frac{\text{signal}}{\text{noise}} \quad D = \frac{\text{distance between groups}}{\text{distance within groups}}$$

Since we are using permutations (iteratively reshuffling data) to generate the distribution of D from our raw data, the shape of the D distribution is dependent on your data

Now the probability of randomly getting a smaller distance than the average distances for the true groups can be calculated

This is the p-value

For permutation tests we can compare D to an expected distribution of D the same way we do when we calculate an F-value



permMANOVA in R

MRPP can be calculated for individual factors in R (we do this in Lab 6.1)

BUT, we can run one or multiple factors (and multiple response variables) at once using Permutational Multivariate Analysis of Variance

Matrix of response variables
These MUST be numeric

permMANOVA in R:

```
adonis(ResponseMatrix, EquationOfPredictors, distance=method) (vegan package)
```

Equation of Predictors (like ANOVA):

Variable1

include single predictor

Variable1+Variable2

include multiple predictors without interaction

Variable1*Variable2

include multiple predictors with interaction

Distance Method to use for calculations:

"euclidian" "manhattan" "bray"

the ones we all ready know (Lab 5)

"gower" "altGower" "canberra" "kulczynski"

"morisita" "horn" "binomial" "cao"

} alternative options – look at `help(adonis)` for more details

permMANOVA interactions

The more predictor variables you include in your analysis the more complicated the results

If you include more than one predictor variable (treatment) – you should investigate if there is a significant interaction between your treatments

All this means is we want to know if the responses behave differently depending on which combination of the predictors we are considering

E.g. Fertilizer A causes a large effect when it is applied to Soil1, but a small or no effect when applied to Soil2

permMANOVA in R

permMANOVA outputs represent a **HIGH LEVEL** summary

Multiple treatments which include at least 2 factors each

Multiple response variables (think of analyzing the response of multiple species – trying to find a common pattern)

We therefore have to carefully pull apart the analysis results to make interpretations

Simplest case –	All p-values are found to be NOT Significant	Pack up & Go Home You're done!
Moderate case –	Main effect(s) are found to be significant No significant interaction	Further analysis needed
Complex case –	Everything is significant	Complexity of analysis is maximized

permMANOVA in R

```
> adonis(measurements~soils, distance="bray")  
  
Call:  
adonis(formula = measurements ~ soils, distance = "bray")  
  
Permutation: free  
Number of permutations: 999  
  
Terms added sequentially (first to last)  
  
      Df SumsOfSqs MeanSqs F.Model    R2 Pr(>F)  
soils  2    1.4308  0.71540  4.2402 0.33282 0.001 ***  
Residuals 17    2.8682  0.16872      0.66718  
Total    19    4.2990      1.00000  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
> adonis(measurements~fertilizer, distance="bray")
```



```
Call:  
adonis(formula = measurements ~ fertilizer, distance = "bray")  
  
Permutation: free  
Number of permutations: 999  
  
Terms added sequentially (first to last)  
  
      Df SumsOfSqs MeanSqs F.Model    R2 Pr(>F)  
fertilizer 4    1.5239  0.38097  2.0592 0.35447 0.022 *  
Residuals 15    2.7751  0.18501      0.64553  
Total    19    4.2990      1.00000  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



We can read MANOVA outputs like an ANOVA table

Moderate Example:

MANOVA with one predictor variable

OR

If only main predictor variable(s) are found to be significant

No significant interaction

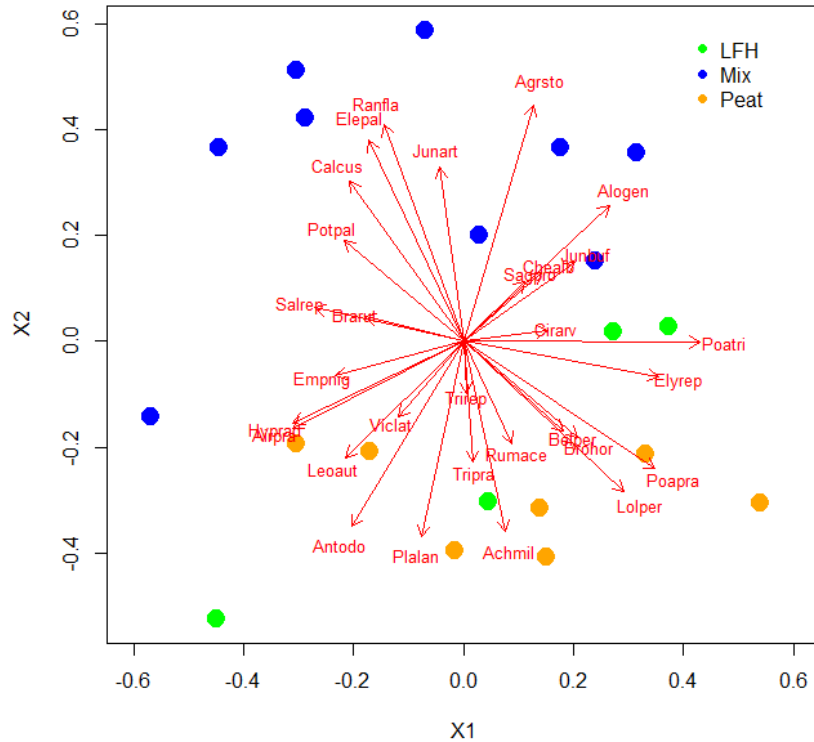
A significant p-value tells us there is a significant difference among groups somewhere

It does **NOT** identify if the trend is true for all response variables OR if a single (or a couple) of response variables are driving the finding of a significant difference

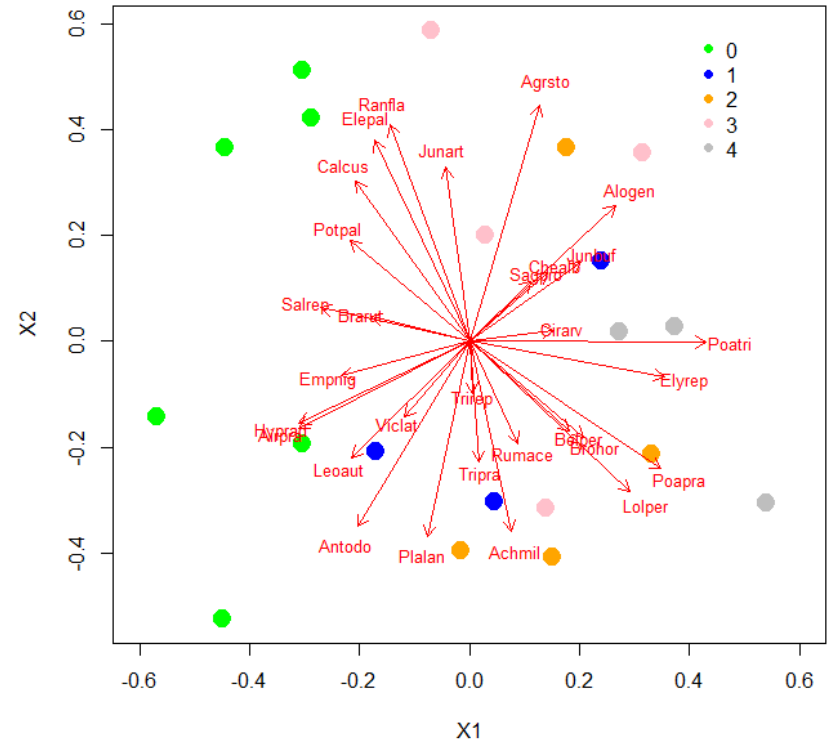
If we find a significant difference in a MAIN effect (single treatment) we can build an NMDS to visualize the differences among species

NMDS to interpret permMANOVA output

Treatment: Soil



Treatment: Fertilizer



We can look at the direction of the species arrows to make inferences as to how which ones are associated with the treatment factors (soil OR fertilizer)

If you want more information on differences for the species with the biggest trends (longest arrows) you can run Permutational ANOVA (univariate) on individual species – Lab 6

permMANOVA in R

```
> adonis(measurements~fertilizer*soils, distance="bray")

Call:
adonis(formula = measurements ~ fertilizer * soils, distance = "bray")

Permutation: free
Number of permutations: 999

Terms added sequentially (first to last)

      Df SumsOfSqs MeanSqs F.Model    R2 Pr(>F)
fertilizer  4  1.5239 0.38097  3.0015 0.35447 0.002 **
soils       2  1.1723 0.58614  4.6179 0.27269 0.001 ***
fertilizer:soils 5  0.5874 0.11749  0.9256 0.13664 0.569
Residuals  8  1.0154 0.12693                0.23620
Total     19  4.2990                1.00000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can read MANOVA outputs like an ANOVA table

Complex Example:

MANOVA with more than one predictor variable
Significant interaction



Let's pretend this p-value is less than 0.05

A significant interaction p-value tells us the responses behave differently depending on which combination of the predictors we are considering

It does **NOT** identify if the trend is true for all response variables OR if a single (or a couple) of response variables are driving the finding of a significant difference

If we find a significant difference in a INTERACTION effect a simple NMDS visualization will not be enough

We need to consider the species individually because they are not acting the same

We can do this with Permutational ANOVAs and pairwise comparisons (univariate) – in Lab 6

Permutational ANOVA in R

Permutational ANOVA is simply analyzed in R using the ImPerm package

However Package ImPerm was build under R version 2.15.1 and has never been updated with R

This cause problems when we want to install the package

The latest version of the package 1.2.1 has been uploaded to the class website for you to download (Windows version .zip file and Mac version .tar.gz file). Save this file to a file path on your C: drive

You need to install it using the “Install package(s) from local zip file(s)...” option in the Packages tab on the R Gui

If you are using R Studio – install packages from Package Achieve File dropdown