

PERMANOVA

Eduard Szöcs

Introduction

Data

Maths

Assumptions

Summary

How to

PERMANOVA

Eduard Szöcs

Institute for Environmental Sciences - University Koblenz-Landau

January 20, 2016

PERMANOVA

Eduard Szöcs

Introduction

Data

Maths

Assumptions

Summary

How to

- ▶ Assumptions of MANOVA:
 - ▶ Independence of the sample units
 - ▶ Multivariate normality
 - ▶ Homogeneity of variance–covariance matrices
- ▶ Euclidean distance useful?
- ▶ Generally **not** met for ecological data sets!
- ▶ Need a robust method to handle complex data sets.

PERMANOVA

Eduard Szöcs

Introduction

Data

Maths

Assumptions

Summary

How to

- ▶ Assumptions of MANOVA:
 - ▶ Independence of the sample units
 - ▶ Multivariate normality
 - ▶ Homogeneity of variance–covariance matrices
- ▶ Euclidean distance useful?
- ▶ Generally **not** met for ecological data sets!
- ▶ Need a robust method to handle complex data sets.

Permutational Multivariate Analysis of Variance Using Distance Matrices

PERMANOVA

Eduard Szöcs

Austral Ecology (2001) **26**, 32–46

A new method for non-parametric multivariate analysis of variance

MARTI J. ANDERSON

*Centre for Research on Ecological Impacts of Coastal Cities, Marine Ecology L
University of Sydney, New South Wales 2006, Australia*



Abstract Hypothesis-testing methods for multivariate data are needed to make rigorous tests about the effects of factors and their interactions in experiments. Analysis of variance

- ▶ Very influential paper in community ecology
 - ▶ Google Scholar: >4500 citations
- ▶ Non-parametric approach combined with ecological distance measures!

Permutational Multivariate Analysis of Variance Using Distance Matrices

PERMANOVA

Eduard Szöcs

Austral Ecology (2001) **26**, 32–46

A new method for non-parametric multivariate analysis of variance

MARTI J. ANDERSON

*Centre for Research on Ecological Impacts of Coastal Cities, Marine Ecology L
University of Sydney, New South Wales 2006, Australia*



Abstract Hypothesis-testing methods for multivariate data are needed to make rigorous tests about the effects of factors and their interactions in experiments. Analysis of variance

- ▶ Very influential paper in community ecology
 - ▶ Google Scholar: >4500 citations
- ▶ Non-parametric approach combined with ecological distance measures!

Permutational Multivariate Analysis of Variance Using Distance Matrices

PERMANOVA

Eduard Szöcs

Austral Ecology (2001) **26**, 32–46

A new method for non-parametric multivariate analysis of variance

MARTI J. ANDERSON

*Centre for Research on Ecological Impacts of Coastal Cities, Marine Ecology L
University of Sydney, New South Wales 2006, Australia*



Abstract Hypothesis-testing methods for multivariate data are needed to make rigorous tests about the effects of factors and their interactions in experiments. Analysis of variance

- ▶ Very influential paper in community ecology
 - ▶ Google Scholar: >4500 citations
- ▶ Non-parametric approach combined with ecological distance measures!

Data set used in this lecture

PERMANOVA

Eduard Szöcs

Introduction

Data

Maths

Assumptions

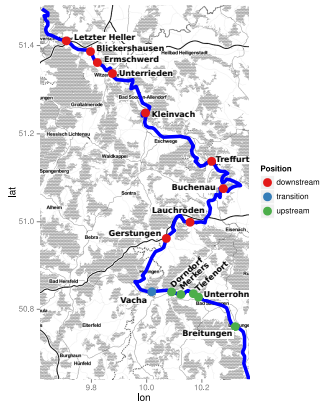
Summary

How to

Macroinvertebrate data from the River Werra¹

Aim : Effect of anthropogenic salinisation on macroinvertebrate communities

- ▶ upstream - downstream design
- ▶ salt brine discharge around Vacha
- ▶ Do the communities differ between up- and downstream?
- ▶ Not the original data (proprietary).



¹Bäthe, Jürgen, and Eckhard Coring. Biological Effects of Anthropogenic Salt-load on the Aquatic Fauna: A Synthesis of 17 Years of Biological Survey on the Rivers Werra and Weser. Limnologia - Ecology and Management of Inland Waters 41(2): 125-133.

First impression of the data

PERMANOVA

Eduard Szöcs

Introduction

Data

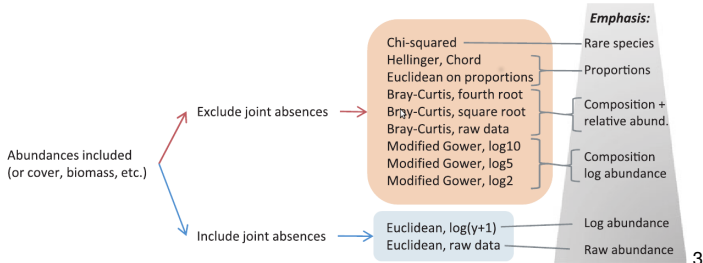
Maths

Assumptions

Summary

How to

► NMDS Bray-Curtis-Distance and $x^{0.25}$ transformation.



³Anderson MJ, Crist TO, Chase JM, Vellend M, Inouye BD, Freestone AL, et al. Navigating the multiple meanings of beta diversity: a roadmap for the practicing ecologist. Ecology Letters. 2011;14(1):19–28.

First impression of the data - NMDS

PERMANOVA

Eduard Szöcs

Introduction

Data

Maths

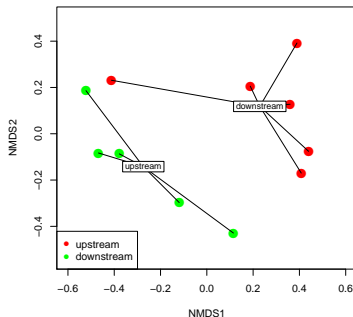
Assumptions

Summary

How to

- ▶ NMDS Bray-Curtis-Distance and $x^{0.25}$ transformation.

- ▶ upstream and downstream sites clearly separate in NMDS.
- ▶ Spread looks similar.
- ▶ Indication of a difference between upstream and downstream.



Recap: ANOVA

PERMANOVA

Eduard Szöcs

Introduction

Data

Maths

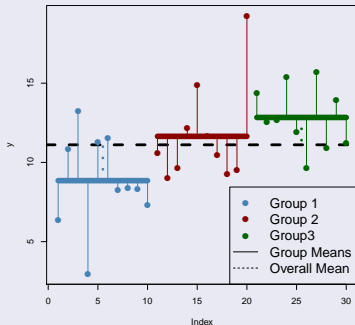
Assumptions

Summary

How to

Question : How is univariate ANOVA calculated?

From univariate...



$$SS_{total} = SS_{residual} + SS_{group}$$

$$F - ratio = \frac{SS_{group}}{SS_{residual}} \cdot \frac{df_{residual}}{df_{group}}$$

Recap: ANOVA

PERMANOVA

Eduard Szöcs

Introduction

Data

Maths

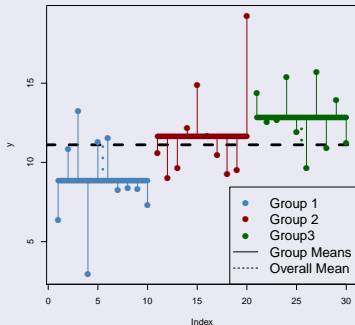
Assumptions

Summary

How to

Question : How is univariate ANOVA calculated?

From univariate...



$$SS_{total} = SS_{residual} + SS_{group}$$

$$F - ratio = \frac{SS_{group}}{SS_{residual}} \cdot \frac{df_{residual}}{df_{group}}$$

Recap: ANOVA

PERMANOVA

Eduard Szöcs

Introduction

Data

Maths

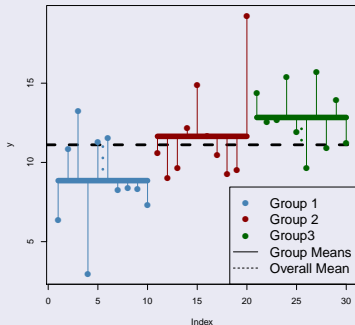
Assumptions

Summary

How to

Question : How is univariate ANOVA calculated?

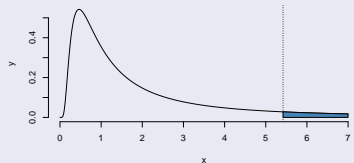
From univariate...



$$SS_{total} = SS_{residual} + SS_{group}$$

$$F - ratio = \frac{SS_{group}}{SS_{residual}} \cdot \frac{df_{residual}}{df_{group}}$$

F-distribution



Distance-based MANOVA

PERMANOVA

Eduard Szöcs

Distance-based MANOVA is analogous!

Introduction

Data

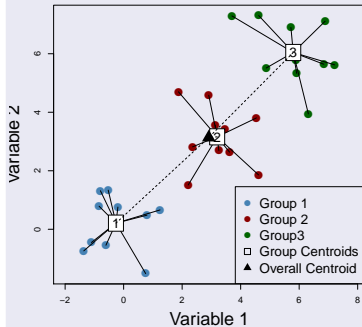
Maths

Assumptions

Summary

How to

... to multivariate ANOVA



- ▶ Partitioning into variance components:

$$SS_{total} = SS_{group} + SS_{residual}$$

- ▶ **centroids**
- ▶ p-value by **permutations**

Variance Partitioning

PERMANOVA

Eduard Szöcs

Introduction

Data

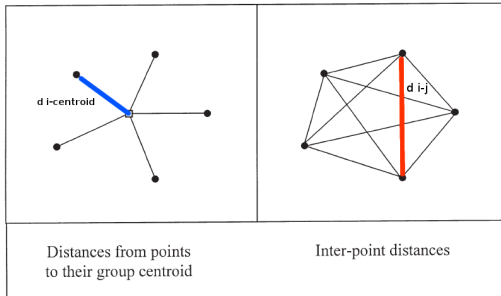
Maths

Assumptions

Summary

How to

- ▶ We can use any **Distance Matrix** to partition the variance.
- ▶ Sum of squared distances from individual points to their centroid is equal to the sum of squared interpoint distances divided by the number of points.
- ▶ $\sum d_{i-\text{centroid}}^2 = \frac{1}{N} \sum d_{i-j}^2$



Variance Partitioning

PERMANOVA

Eduard Szöcs

Introduction

Data

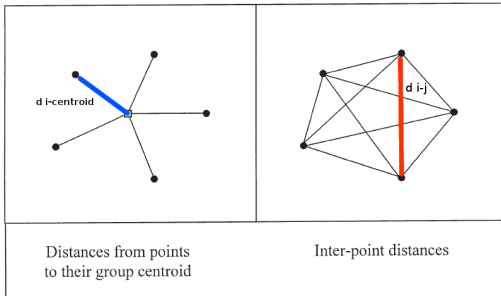
Maths

Assumptions

Summary

How to

- ▶ We can use any **Distance Matrix** to partition the variance.
- ▶ Sum of squared distances from individual points to their centroid is equal to the sum of squared interpoint distances divided by the number of points.
- ▶ $\sum d_{i-\text{centroid}}^2 = \frac{1}{N} \sum d_{i-j}^2$



Variance Partitioning

PERMANOVA

Eduard Szöcs

Introduction

Data

Maths

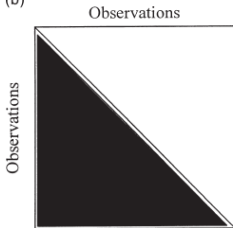
Assumptions

Summary

How to

Like in univariate ANOVA variance can be partitioned:

(b)



$$SS_{total} = \frac{1}{N} \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij}^2$$

N = total number of observations

$$SS_{residual} = \frac{1}{n} \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij}^2 \epsilon_{ij}$$

n = number of observations per group

$$\epsilon_{ij} = \begin{cases} 1, & \text{if observations } i \text{ and } j \text{ are in the same group} \\ 0, & \text{otherwise} \end{cases}$$

$$SS_{group} = SS_{total} - SS_{residual}$$

$$(Pseudo-)F = \frac{SS_{group}}{SS_{residual}} \frac{N-a}{a-1}$$

a = no. groups

p-value is assessed via permutations.

Variance Partitioning

PERMANOVA

Eduard Szöcs

Introduction

Data

Maths

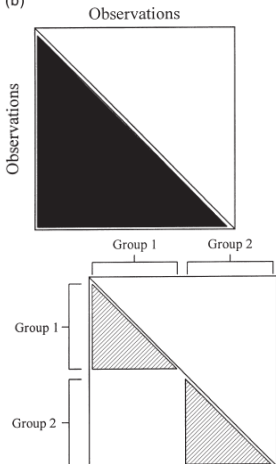
Assumptions

Summary

How to

Like in univariate ANOVA variance can be partitioned:

(b)



$$SS_{total} = \frac{1}{N} \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij}^2$$

N = total number of observations

$$SS_{residual} = \frac{1}{n} \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij}^2 \epsilon_{ij}$$

n = number of observations per group

$$\epsilon_{ij} = \begin{cases} 1, & \text{if observations } i \text{ and } j \text{ are in the same group} \\ 0, & \text{otherwise} \end{cases}$$

$$SS_{group} = SS_{total} - SS_{residual}$$

$$(Pseudo-)F = \frac{SS_{group}}{SS_{residual}} \frac{N-a}{a-1}$$

a = no. groups

p-value is assessed via permutations.

Variance Partitioning

PERMANOVA

Eduard Szöcs

Introduction

Data

Maths

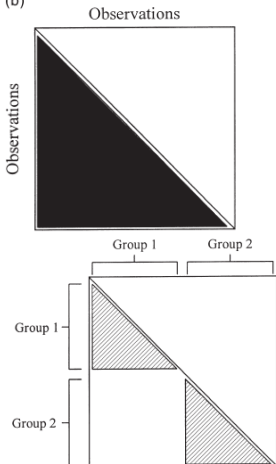
Assumptions

Summary

How to

Like in univariate ANOVA variance can be partitioned:

(b)



$$SS_{total} = \frac{1}{N} \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij}^2$$

N = total number of observations

$$SS_{residual} = \frac{1}{n} \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij}^2 \epsilon_{ij}$$

n = number of observations per group

$$\epsilon_{ij} = \begin{cases} 1, & \text{if observations } i \text{ and } j \text{ are in the same group} \\ 0, & \text{otherwise} \end{cases}$$

$$SS_{group} = SS_{total} - SS_{residual}$$

$$(Pseudo-)F = \frac{SS_{group}}{SS_{residual}} \frac{N-a}{a-1}$$

a = no. groups

p-value is assessed via permutations.

Variance Partitioning

PERMANOVA

Eduard Szöcs

Introduction

Data

Maths

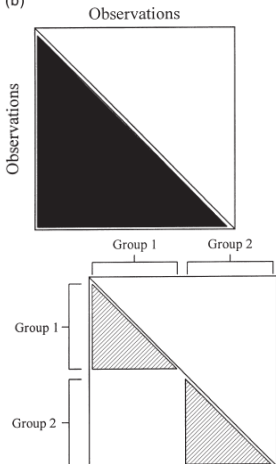
Assumptions

Summary

How to

Like in univariate ANOVA variance can be partitioned:

(b)



$$SS_{total} = \frac{1}{N} \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij}^2$$

N = total number of observations

$$SS_{residual} = \frac{1}{n} \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij}^2 \epsilon_{ij}$$

n = number of observations per group

$$\epsilon_{ij} = \begin{cases} 1, & \text{if observations } i \text{ and } j \text{ are in the same group} \\ 0, & \text{otherwise} \end{cases}$$

$$SS_{group} = SS_{total} - SS_{residual}$$

$$(Pseudo-)F = \frac{SS_{group}}{SS_{residual}} \frac{N-a}{a-1}$$

a = no. groups

p-value is assessed via permutations.

p-values using permutations

PERMANOVA

Eduard Szöcs

Introduction

Data

Maths

Assumptions

Summary

How to

- ▶ Cannot use Fisher's F-ratio

- ▶ normal distribution?
- ▶ euclidean distance?

- ▶ Instead use **permutations**

- ▶ shuffle data randomly
- ▶ compute F-Ratio (F_{perm})
- ▶ repeat many times

- ▶ compare **F of randomized data** with **original F**.

- ▶
$$p = \frac{\text{No. of } F_{perm} \geq F}{\text{No. of permutations} + 1}$$

p-values using permutations

PERMANOVA

Eduard Szöcs

Introduction

Data

Maths

Assumptions

Summary

How to

▶ Cannot use Fisher's F-ratio

- ▶ normal distribution?
- ▶ euclidean distance?

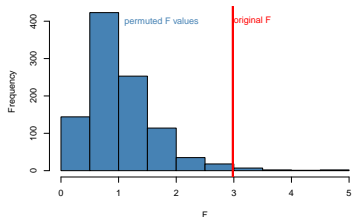
▶ Instead use **permutations**

- ▶ shuffle data randomly
- ▶ compute F-Ratio (F_{perm})
- ▶ repeat many times

▶ compare **F of randomized data** with **original F**.

$$p = \frac{\text{No. of } F_{perm} \geq F}{\text{No. of permutations} + 1}$$

Histogramm of permuted F values



Permutations (II)

PERMANOVA

Eduard Szöcs

Introduction

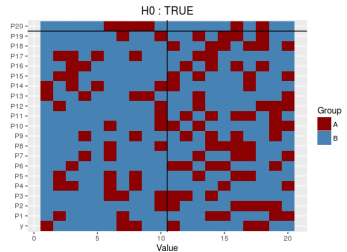
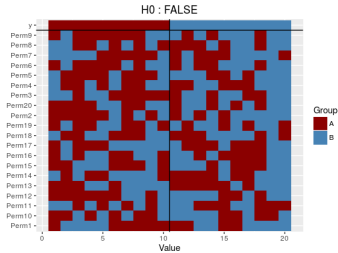
Data

Maths

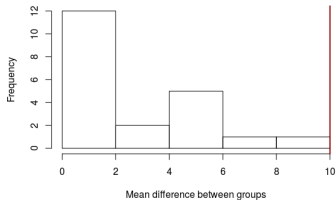
Assumptions

Summary

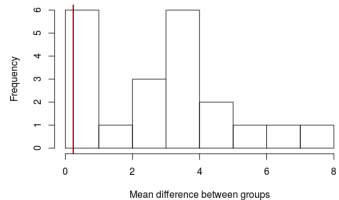
How to



Hist. of differences between groups



Hist. of differences between groups



Assumptions of PERMANOVA

PERMANOVA

Eduard Szöcs

Introduction

Data

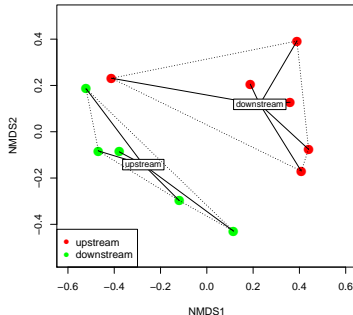
Maths

Assumptions

Summary

How to

- ▶ equal **dispersions**
- ▶ Visual inspection
- ▶ Multivariate analogue to Levene's test available.⁴
- ▶ Multivariate Dispersion
 - ▶ β -diversity
 - ▶ functional diversity
 - ▶ see literature folder



⁴ Anderson, M. J. 2006. Distance Based Tests for Homogeneity of Multivariate Dispersions. Biometrics 62 (1): 245-253.

Assumptions of PERMANOVA

PERMANOVA

Eduard Szöcs

Introduction

Data

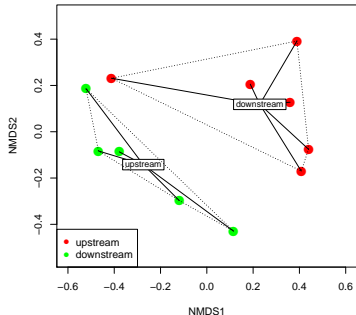
Maths

Assumptions

Summary

How to

- ▶ equal **dispersions**
- ▶ Visual inspection
- ▶ Multivariate analogue to Levene's test available.⁴
- ▶ Multivariate Dispersion
 - ▶ β -diversity
 - ▶ functional diversity
 - ▶ see literature folder



⁴Anderson, M. J. 2006. Distance Based Tests for Homogeneity of Multivariate Dispersions. Biometrics 62 (1): 245-253.

Summary

PERMANOVA

Eduard Szöcs

Introduction

Data

Maths

Assumptions

Summary

How to

PERMANOVA is a

- ▶ flexible (any distance measure) and
 - ▶ easy (analogue to univariate Anova)
- tool for ecologists.

However,

- ▶ non-parametric does no mean assumption free.

Summary

PERMANOVA

Eduard Szöcs

Introduction

Data

Maths

Assumptions

Summary

How to

PERMANOVA is a

- ▶ flexible (any distance measure) and
 - ▶ easy (analogue to univariate Anova)
- tool for ecologists.

However,

- ▶ non-parametric does no mean assumption free.

Lets do it in R!

PERMANOVA

Eduard Szöcs

Introduction

Data

Maths

Assumptions

Summary

How to



*use***R**!