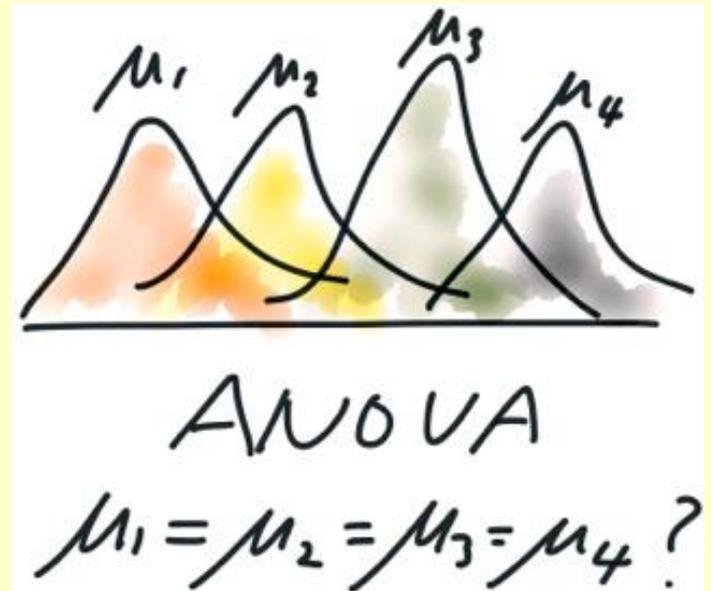


ANOVA approach

➤ Advantages:

- Ideal for evaluating hypotheses
- Ideal to quantify effect size (e.g., differences between groups)
- Address multiple factors at once
- Investigates interaction terms



➤ Disadvantages:

- Requires careful sampling design with replication
- Relies on assumptions of normality and equal variances

ANOVA Designs

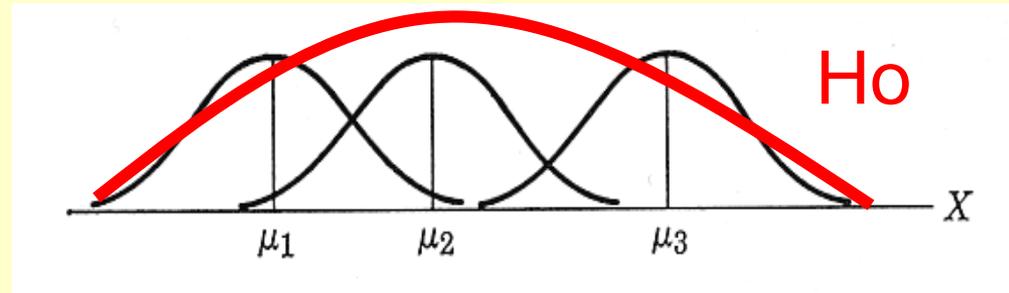
➤ Definitions following [Searle et al. 1992](#)

- **Factor:** A classification that assigns each observation to one level of the classification. In PC-ORD, the factors are chosen from individual columns in the second matrix.
- **Level:** Individual classes of a classification. For example, factor "sex" has 2 levels: "male" and "female." In PC-ORD different levels are assigned integer numerical values. The actual value chosen to represent the levels is unimportant.
- **Cell:** A subset of data occurring at the intersection of one level of every factor being considered. Every data point in a design belongs to one and only one cell of the design.
- **Balanced Design:** Every cell with equal sample size.

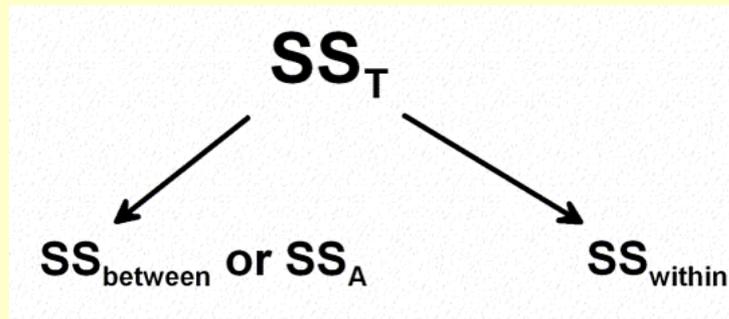
ANOVA – One-Way

➤ Conceptual Approach:

- Consider one Factor (X)



- Calculate amount of variance explained by this factor



- F test quantifies ratio of variance between / within:

$$F = \frac{\text{between-group variability}}{\text{within-group variability}}$$

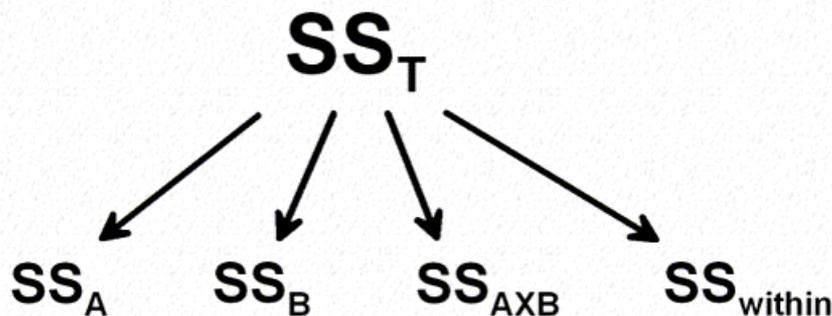
- Pseudo-f statistic:

$$F = \frac{\text{explained variance}}{\text{unexplained variance}}$$

ANOVA – Two-Way

➤ Conceptual Approach:

- Define two or more Factors
- Calculate means for the different factor combinations
- Calculate amount of variance explained by each factor, and by their interaction



		Factor B				A Marginals
		b_1	b_2	b_k	b_q	
Factor A	a_1	X_{i11}	X_{i12}	X_{i1k}	X_{i1q}	$\bar{X}_{.1}$
		X_{n11}	X_{n12}	X_{n1k}	X_{n1q}	
		\bar{X}_{11}	\bar{X}_{12}	\bar{X}_{1k}	\bar{X}_{1q}	
	a_2	X_{i21}	X_{i22}	X_{i2k}	X_{i2q}	$\bar{X}_{.2}$
		X_{n21}	X_{n22}	X_{n2k}	X_{n2q}	
		\bar{X}_{21}	\bar{X}_{22}	\bar{X}_{2k}	\bar{X}_{2q}	
	a_j	X_{ij1}	X_{ij2}	X_{ijk}	X_{ijq}	$\bar{X}_{.j}$
		X_{nj1}	X_{nj2}	X_{njk}	X_{njq}	
		\bar{X}_{j1}	\bar{X}_{j2}	\bar{X}_{jk}	\bar{X}_{jq}	
	a_p	X_{ip1}	X_{ip2}	X_{ipk}	X_{ipq}	$\bar{X}_{.p}$
		X_{np1}	X_{np2}	X_{npk}	X_{npq}	
		\bar{X}_{p1}	\bar{X}_{p2}	\bar{X}_{pk}	\bar{X}_{pq}	
B Marginals		$\bar{X}_{.1}$	$\bar{X}_{.2}$	$\bar{X}_{.k}$	$\bar{X}_{.q}$	$\bar{X}_{..}$
Grand Mean						

Multivariate ANOVAs

➤ *Objectives:*

Discuss general approach of multivariate experiments

Go over the settings / results of multivariate ANOVAs

Go over the settings / results of blocked MRPP

Experiments vs Exploration

➤ Traditionally, ecologists face a dichotomy between experimental and exploratory scientific approaches: pattern description versus experimental manipulation

➤ **For example:** PCA / NMDS vs PO / MRPP

➤ **What is an Experiment?**

Change one or more variables in a consistent way

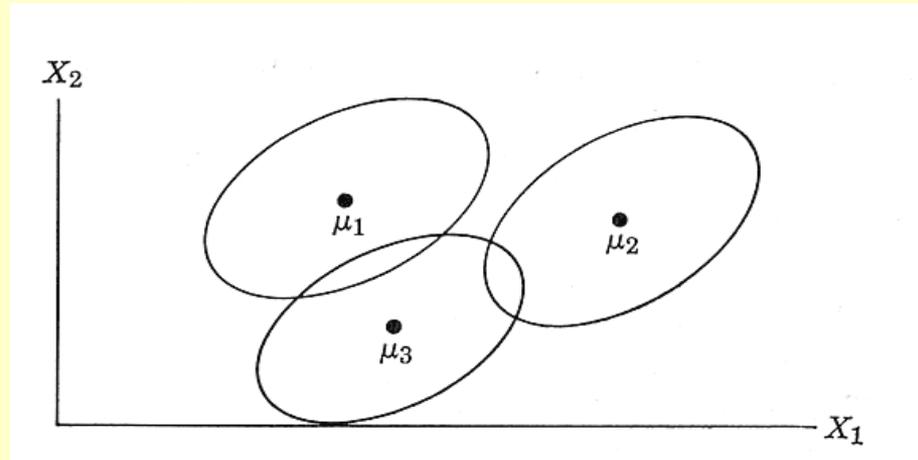
Contrast multiple treatments (& interactions)

Contrast response against controls (unmanipulated)

➤ **Take Home:** Investigator controls the allocation of samples to treatments (replicates)

Multivariate ANOVA

Traditional MANOVA has two properties that render it inappropriate for analysis of ecological community data (spp):



- Relationships among data points are assumed to be Euclidean; yet this distance measure performs poorly with community data (McCune & Grace 2002)
- Calculation of p values assumes multivariate normality, though this assumption is infrequently reasonable for community data (McCune & Grace 2002)

Multivariate ANOVA

- Multivariate ANOVA (MANOVA): An analysis of variance where the response consists of two or more potentially interrelated variables. In contrast, an univariate analysis of variance has only one response variable.
- PerMANOVA performs distance-based multivariate analysis of variance, also known as nonparametric MANOVA or npMANOVA.
- Hypotheses are evaluated with permutation tests, rather than by reference to an assumed (normal) distribution.

● Group 1

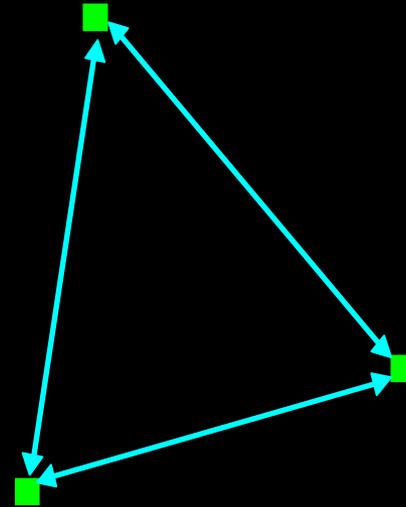
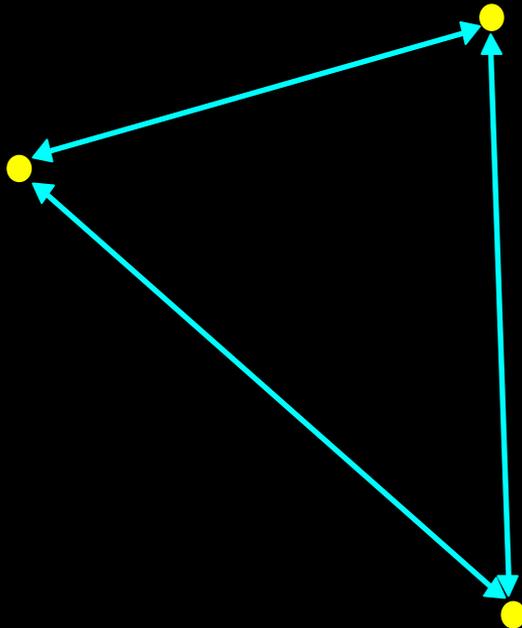
■ Group 2



● Group 1

■ Group 2

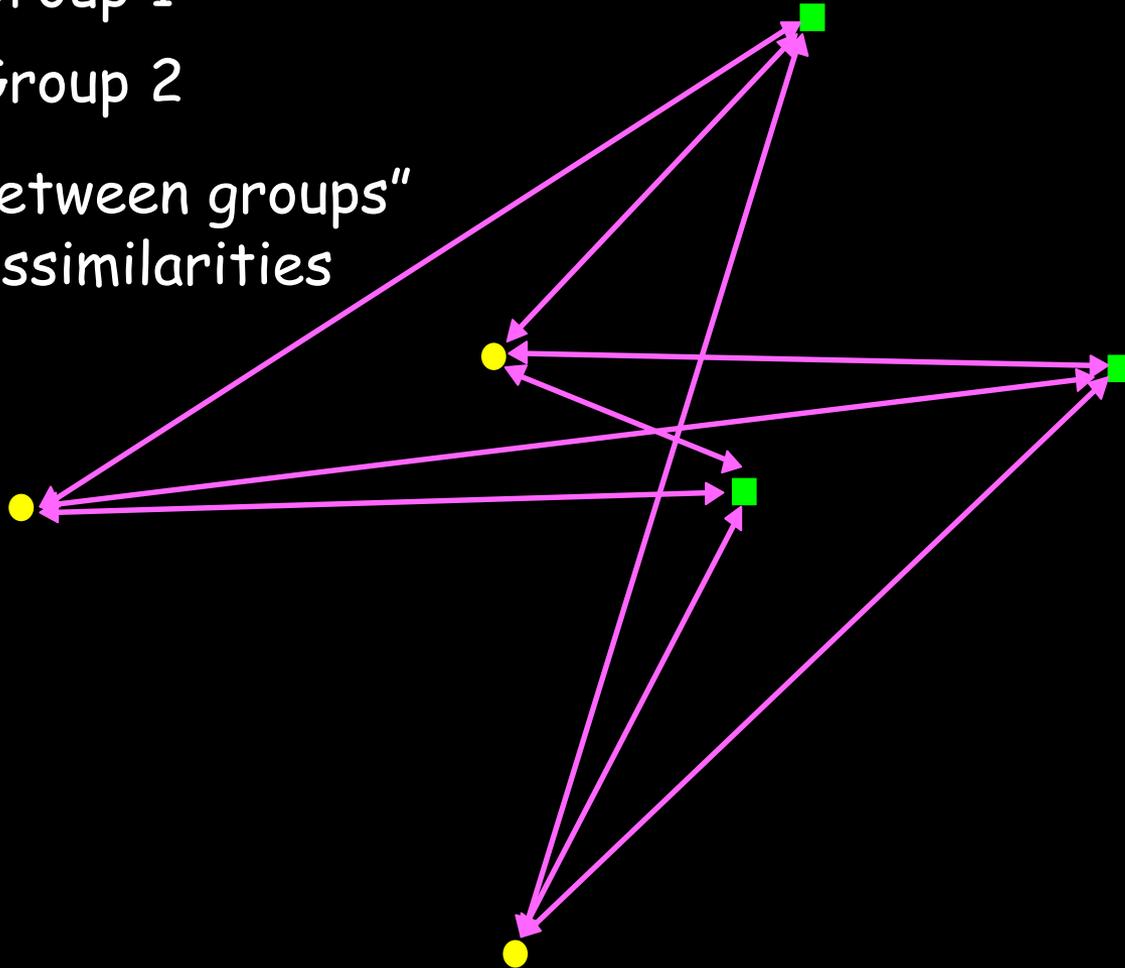
↔ "Within groups"
dissimilarities



● Group 1

■ Group 2

↔ "Between groups"
dissimilarities

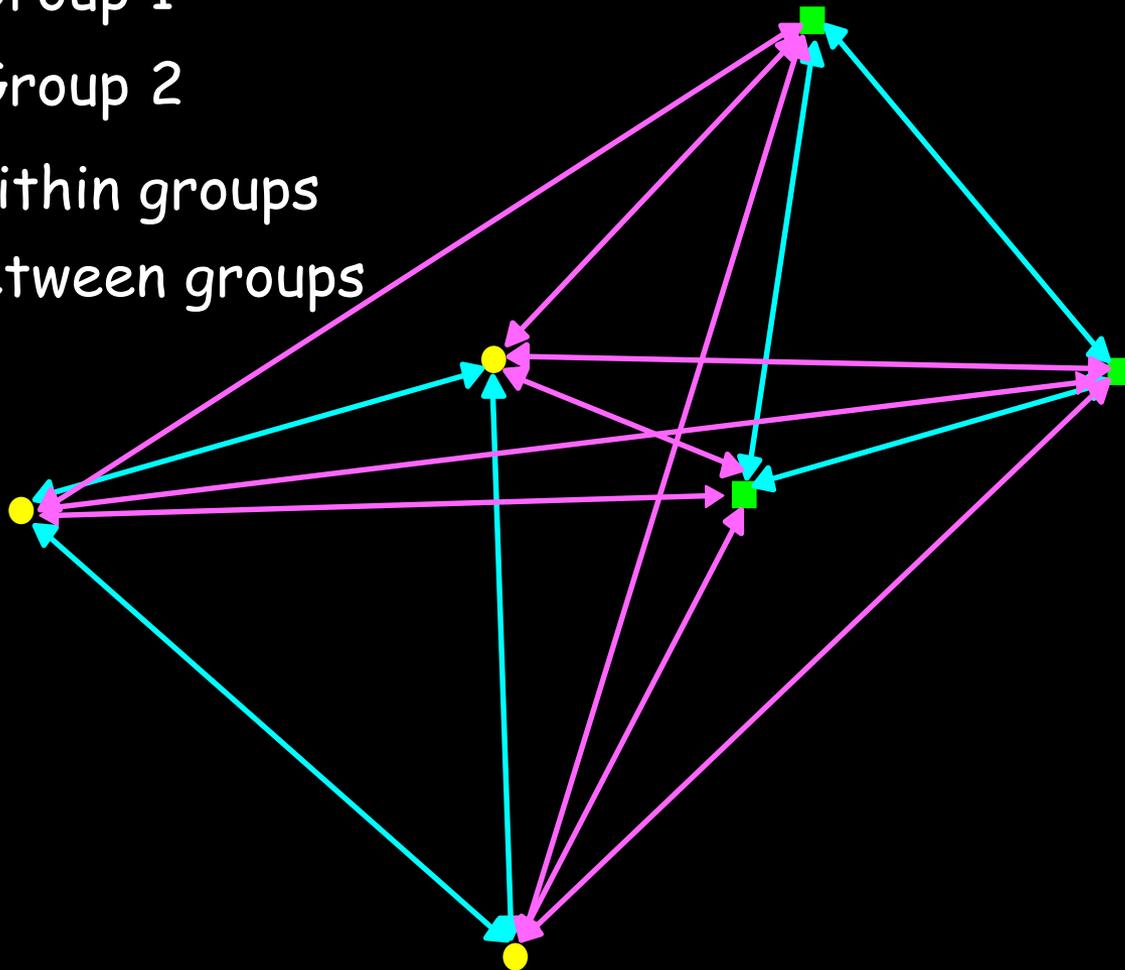


● Group 1

■ Group 2

↔ Within groups

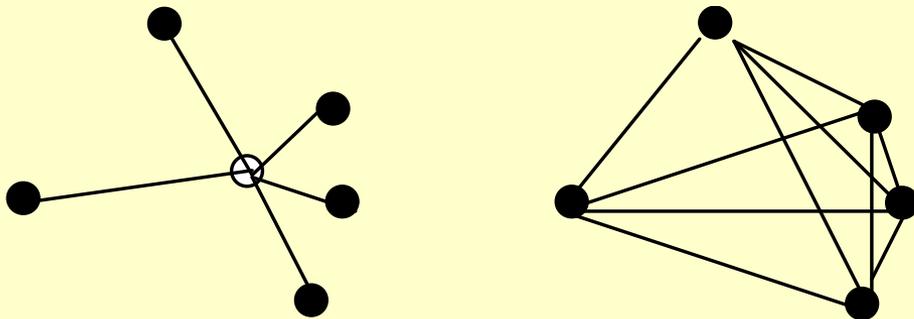
↔ Between groups



Multivariate ANOVA

Key to development of PerMANOVA was **Anderson's (2001)** recognition that sums of squares could be calculated directly using distances among data points, rather than the distances from the data points to the mean. She explains the problem:

"In the case of an analysis based on Euclidean distances, the average for each variable across the observations within a group constitutes the measure of central location for the group in Euclidean space, called a centroid. For many distance measures, however, the calculation of a central location may be problematic"



Sums of distances from points to centroid (left) calculated from average squared interpoint distance (right).

PerMANOVA – How it Works

The total sum of squares of a distance matrix \mathbf{D} with N rows and N columns is:

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

The residual (within-group) sum of squares for a one-way classification is:

$$SS_R = \frac{1}{n} \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij}^2 \varepsilon_{ij}$$

where n is the number of observations per group,
 N is the number of sample units,
 $\varepsilon_{ij} = 1$ if i and j in same group; $\varepsilon_{ij} = 0$ if in different groups.

PerMANOVA – How it Works

Thus, the sum of squares between groups is:

$$SS_A = SS_T - SS_R$$

This allows user to calculate a pseudo- F -ratio:

$$F = \frac{SS_A / (a - 1)}{SS_R / (N - a)}$$

Explained Variance

Unexplained Variance

where a is group number and N is sample number.

Note: If distance matrix contains Euclidean distances, then pseudo- F = parametric univariate F ratio.

PerMANOVA – Limitations

➤ Limitations and non-limitations

- Maximum number of levels is 500 for each grouping variable.
- Maximum number of factors is 2.
- All cells must be filled (no empty cells)
- Design must be balanced, with equal number of observations in each cell.
- Each cell must have replication (more than one observation, $n > 1$)

Groups	Tools	Window
	Cluster Analysis	
	Two-way Cluster Analysis	
	MRPP	
	Blocked MRPP (MRBP)	
	TWINSPAN	
	Indicator Species Analysis	
	Mantel Test	
	PerMANOVA	

Island / Habitat	A	B
1	n	n
2	n	n

PerMANOVA - Output

- PerMANOVA writes a result file. No graphics produced.
- To explore the relationships among the data points and groups, use ordination to provide a visual summary.
- **NOTE:** Univariate vs. multivariate ANOVA:

You can use perMANOVA in PC-ORD to perform either univariate or multivariate permutation-based ANOVA.

If you have only one column in your main matrix (one variable), then the analysis is univariate.

If you perform the univariate analysis with Euclidean distance, the result will match an ANOVA analysis.

PerMANOVA - Example

- Indian Ocean Seabirds: Equal Replication (15 samples)
- Group Definitions – Relate back to Hypothesis

3 Watermasses:

- 1) Tropical:
SST > 20 deg. C.
- 2) Subtropical:
20 < SST < 18
- 3) Transition:
18 > SST

Main - WORK.WK1					Second - WORK2.WK1		
15					15	plots	
42					2	variabl	
	Q	Q	Q	Q		C	C
	SPPT	YNAL	LISH	LTJA		waterma	product
plot1	0	0	0	0	plot1	1	1
plot2	0	0	12.5	0	plot2	1	1
plot3	28.21	0	0	0	plot3	1	1
plot4	23.98	2.04	3.06	1.02	plot4	2	2
plot5	15.63	3.23	2.16	0	plot5	3	3
plot6	2.03	0.08	0	0	plot6	3	2
plot7	1.64	0	0	0	plot7	3	3
plot8	0.19	0	0	0	plot8	3	3
plot9	2.29	0.08	0	0	plot9	3	2
plot10	0.67	69.02	0	0	plot10	2	3
plot11	0	6.1	0	0	plot11	2	2
plot12	0	0	0	0	plot12	2	2
plot13	0	0	0	1.25	plot13	2	3
plot14	0	0	3.37	2.25	plot14	1	1
plot15	0	0	1.57	0.2	plot15	1	1

PerMANOVA - Example

- One-Way Test:
One Factor
(watermass)

- Distance
Metric Used
(Relative
Sorensen)

Distance Measure

- Sorensen (Bray-Curtis)
- Relative Sorensen
- Jaccard
- Euclidean (Pythagorean)
- Relative Euclidean
- Correlation
- Chi-squared
- Squared Euclidean

- Randomizations
(n = 999)
- Pair-wise Comparisons
(Post-hoc Tests)

PerMANOVA Setup

Design

- One way
- Two-way factorial
- One fixed factor and one level nested
- Two levels nested
- Randomized complete blocks

Replicates within [watermass] as grouping variable.

Factor One

- watermass
- productivity

Factor Two

- watermass
- productivity

Randomization test

Make pairwise comparisons

Write F statistics for each permutation (bulky output)

Count unique values of F (recommended for small data sets)

OK Cancel

PerMANOVA - Example

- Pseudo-F
- Df: 14 (12, 2)

IndianOceanBirds_Groups

Groups were defined by values of: watermass

Main matrix has: 15 plots by 42 species

Distance measure = Relative Sorensen

Evaluation of differences in species between groups.

Design: One-way

Randomization test of significance of pseudo F values

Number of randomizations: 4999

Random number seed: 4553 selected by time.

<u>Source</u>	<u>d.f.</u>	SS	MS	F	p *
<u>watermas</u>	2	2.1219	1.0609	3.9684	0.001000
Residual	12	3.2082	0.26735		
Total	14	5.3301			

- Variance explained

Variance components estimated for random effects model (Model II)

Ignore variance components if you consider the factor to have fixed effects.

COMPONENTS OF VARIANCE

Source	Variance	% of variation
<u>watermas</u>	0.15872	37.252
Residual	0.26735	62.748
Total	0.42607	100.000

PerMANOVA - Example

- P values calculated with permutations

Statistics from randomizations

```
-----  
F from randomized groups  
-----  
Number  
Source      F      Mean      Maximum      S.Dev      observed F      p *  
-----  
watermass  3.96840  1.02572  5.45999  0.00563  4  0.001000
```

* proportion of randomized trials with indicator value equal to or exceeding the observed indicator value.
 $p = (1 + \text{number of runs } \geq \text{observed}) / (1 + \text{number of randomized runs})$

- Pair-wise comparisons

```
PAIRWISE COMPARISONS for factor watermas  
Note: p values are not corrected for multiple comparisons.  
-----  
Level vs. Level      t      p  
-----  
1 vs. 2      1.1737  0.188400  
1 vs. 3      2.3813  0.007800  
2 vs. 3      2.4737  0.007000  
-----  
***** PerMANOVA finished *****
```

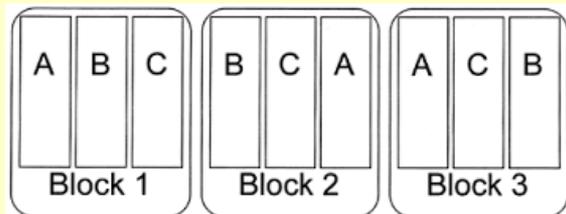
Expanded ANOVA Designs

- Two-way
(interaction)

- Blocking

- Nestedness

- Randomized
Complete Block



PerMANOVA Setup

Design

- One way
- Two-way factorial
- One fixed factor and one level nested
- Two levels nested
- Randomized complete blocks

Replicates within [watermass] as grouping variable.

Factor One

- watermass
- productivity

Factor Two

- watermass
- productivity

Distance Measure

- Sorensen (Bray-Curtis)
- Relative Sorensen
- Jaccard
- Euclidean (Pythagorean)
- Relative Euclidean
- Correlation
- Chi-squared
- Squared Euclidean

Randomization test

Make pairwise comparisons

Write F statistics for each permutation (bulky output)

Count unique values of F (recommended for small data sets)

OK Cancel Help

Nestedness

Nested layouts needed when constraints prevent the crossing every level of one factor with every level of another factor. Thus, fewer than all levels of one factor occur within each level of the other factor.

If Factor B is nested within Factor A, then some level of Factor B can only occur within some other level of Factor A and there can be no interaction of Factor A and Factor B.

For example:

We are studying demographic rates in different countries and continents. Yet, countries are only represented within their continents. Thus, countries are nested within continents.

Blocking

Blocking to "remove" the effect of nuisance factors.

For block designs, one factor or variable is of primary interest. However, there are other uncontrolled (nuisance) factors.

Blocking can be used to reduce or eliminate the contribution to experimental error contributed by these nuisance factors.

Create homogeneous blocks in which the nuisance factors are held constant, and the factor of interest is allowed to vary.

Within blocks, it is possible to assess the effect of the factor of interest, without worrying about variations due to changes in the block factors, which are accounted for in the analysis.

Blocking

A nuisance factor can be used as a blocking factor if every level of the primary factor **occurs the same number of times with each level of the nuisance factor**. The analysis will focus on varying the primary factor within each experiment block.

Design 1: Blocked

subplot	plot	region
1	1	1
2	1	1
1	2	1
2	2	1
1	3	1
2	3	1
1	1	2
2	1	2
1	2	2
2	2	2
1	3	2
2	3	2

Design 2: Nested

subplot	plot	region
1	1	1
2	1	1
1	2	1
2	2	1
1	3	1
2	3	1
1	4	2
2	4	2
1	5	2
2	5	2
1	6	2
2	6	2

Examples:

- Seasonal vegetation at multiple sites (samples linked by location)
- Paired comparisons (morning / night samples at same location)

Blocked MRPP (MRBP)

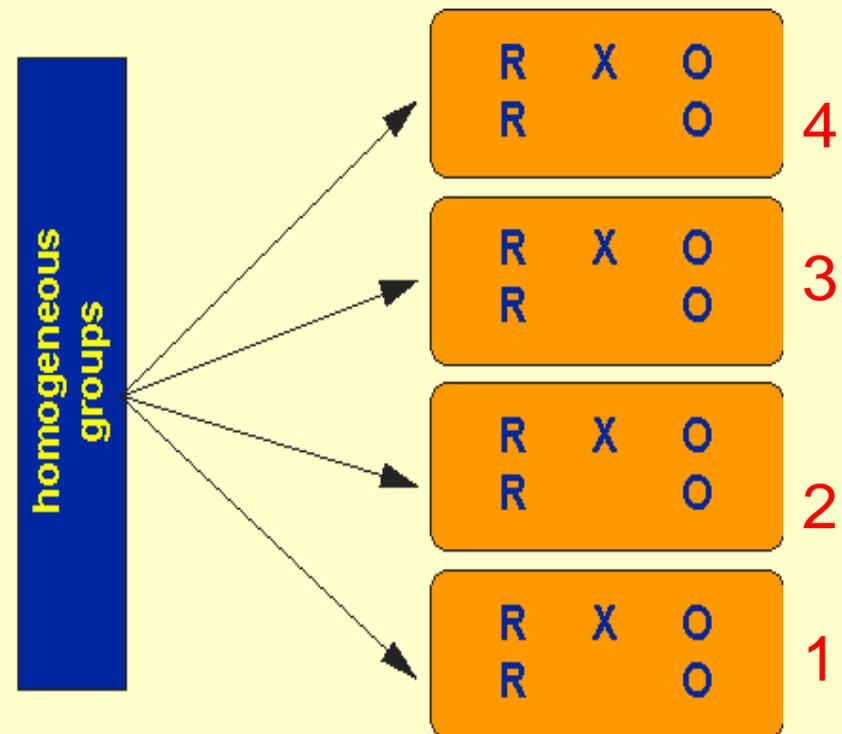
Randomized block or paired-sample data can be analyzed with a variant of MRPP called blocked MRPP (MRBP).

Blocking arranges experimental units into homogeneous groups different from each other (called blocks)

A blocking factor is a source of variability that is not of primary interest to the experimenter, but may affect the result.

Categorical Blocking Factor

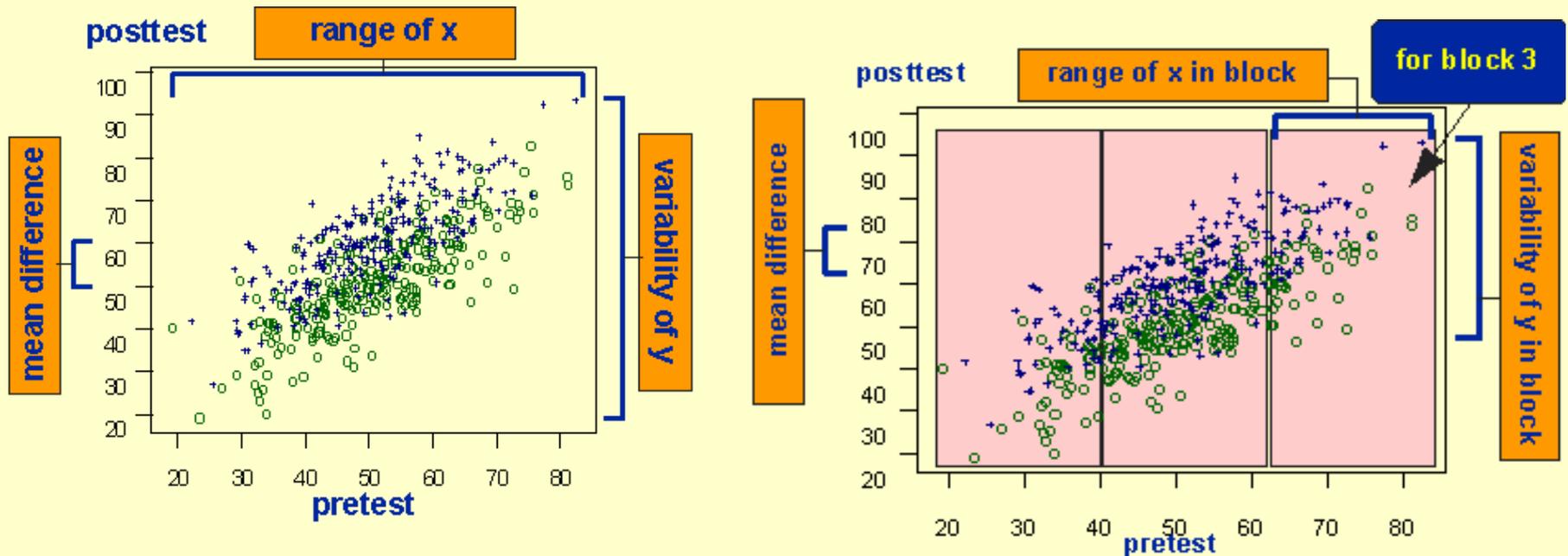
Example: Does software use (R / O) affect stats skills, after controlling for grade level?



Blocked MRPP (MRBP)

Continuous Blocking Factor Example: Does software use (R / O) affect post-test score, after controlling for skill level?

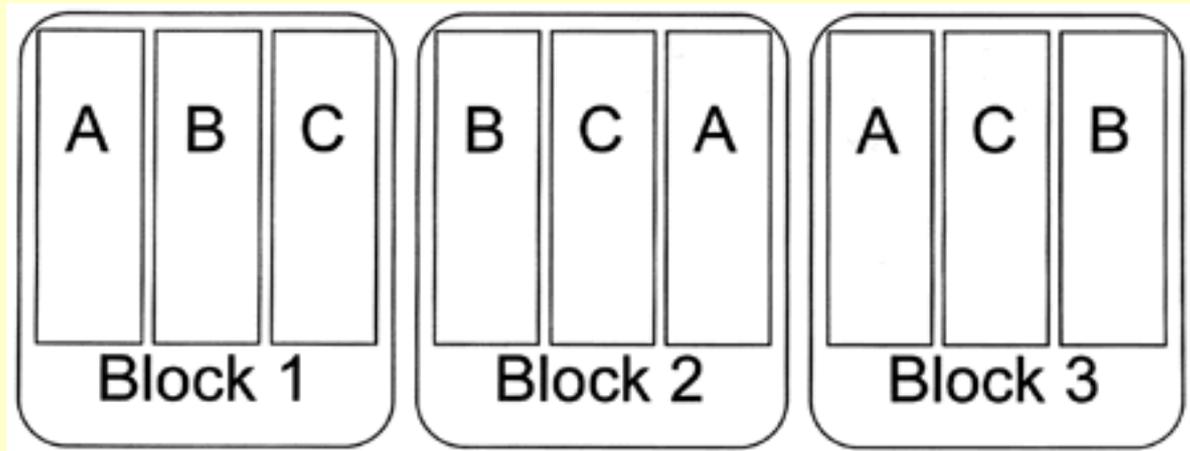
The factors of interest are depicted by crosses / circles.



The general rule is to:

“Block what you can, randomize what you cannot”

Blocking in MRPP



Limitations: this analysis **requires a balanced design**

- One sample unit for each block / treatment combination.
 - Number of treatments must be equal among blocks.
 - Each treatment (A, B, C) must be present in each block.
- (**NOTE:** PC-ORD allows up to 1000 blocks and 100 groups).

Blocked MRPP (MRBP)

Given b blocks and g groups (treatments), the MRPP statistic is modified to:

$$\delta = \sum_{i=1}^g C_i x_i$$

where $\Delta(x_{ij}, x_{ik})$ is the distance between points x_{ij} and x_{ik} in the p dimensional space.

$$\delta = \left[g \binom{b}{2} \right]^{-1} \sum_{i=1}^g \sum_{j < k} \Delta(x_{ij}, x_{ik})$$

Note that for paired-sample data, b is the number of linked observations (blocks) and g is the number of treatments.

Delta is average distance between blocks within treatments.

Blocked MRPP (MRBP)

Approach:

Null hypothesis assigns equal probabilities to each of the $M = (g!)^b$ possible allocations of the g p -dimensional measurements to g treatments within each of b blocks.

In other words, the observed values are randomly reassigned to different treatments in each block. They remain in their block.

Interpretation:

Like MRPP, small δ imply that, given p dimensional space, samples are more similar within than across treatments.

The added features of MRBP are that:

- distances summed with respect to the blocks
- the user has the option of aligning blocks so that all treatments in a given block have a median of zero.

MRBP - Example

- Indian Ocean Seabirds: Equal Replication (15 samples)
- Group Definitions – Watermasses
- Data from multiple years: Sampled all watermasses

3 Watermasses:

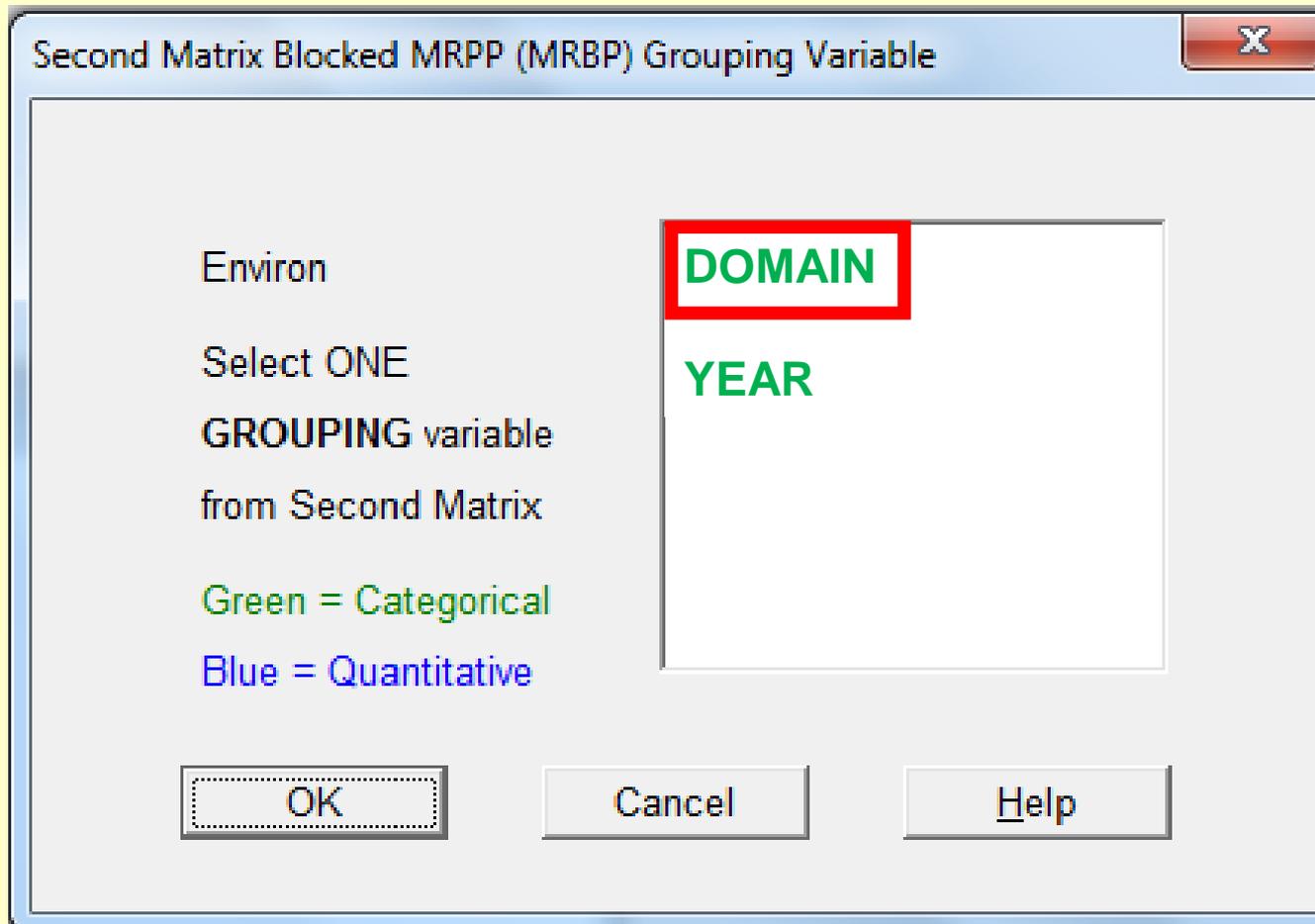
- 1) Tropical:
SST > 20 deg. C.
- 2) Subtropical:
20 < SST < 18
- 3) Transition:
18 > SST

Main - WORK.WK1					Second - WORK2.WK1		
15					15	plots	
42					2	variabl	
	Q	Q	Q	Q		C	C
	SPPT	YNAL	LISH	LTJA		waterma	product
plot1	0	0	0	0	plot1	1	1
plot2	0	0	12.5	0	plot2	1	1
plot3	28.21	0	0	0	plot3	1	1
plot4	23.98	2.04	3.06	1.02	plot4	2	2
plot5	15.63	3.23	2.16	0	plot5	3	3
plot6	2.03	0.08	0	0	plot6	3	2
plot7	1.64	0	0	0	plot7	3	3
plot8	0.19	0	0	0	plot8	3	3
plot9	2.29	0.08	0	0	plot9	3	2
plot10	0.67	69.02	0	0	plot10	2	3
plot11	0	6.1	0	0	plot11	2	2
plot12	0	0	0	0	plot12	2	2
plot13	0	0	0	1.25	plot13	2	3
plot14	0	0	3.37	2.25	plot14	1	1
plot15	0	0	1.57	0.2	plot15	1	1

MRBP SetUp

➤ Select Grouping Variable

- Matrix 2
- Categorical

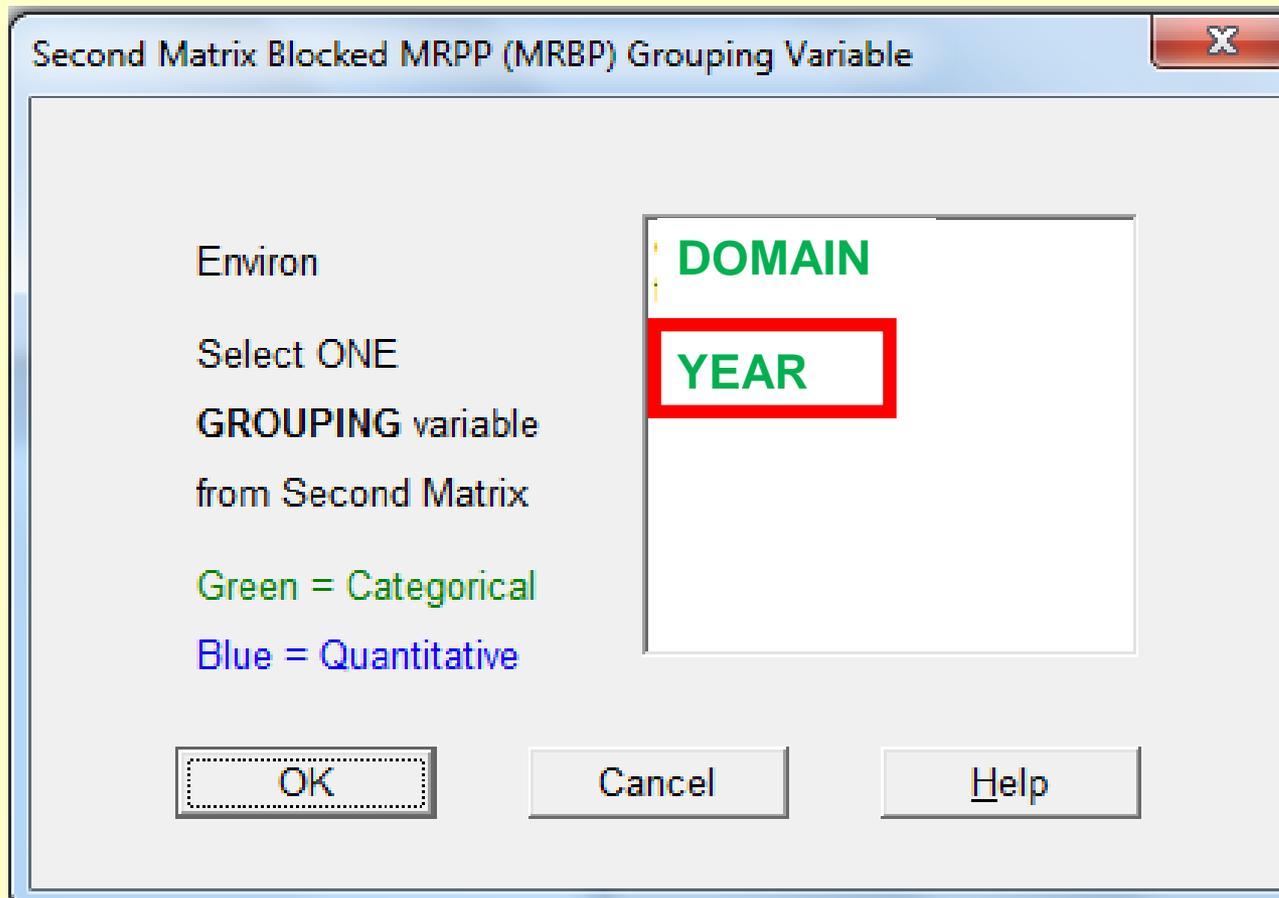


MRBP SetUp

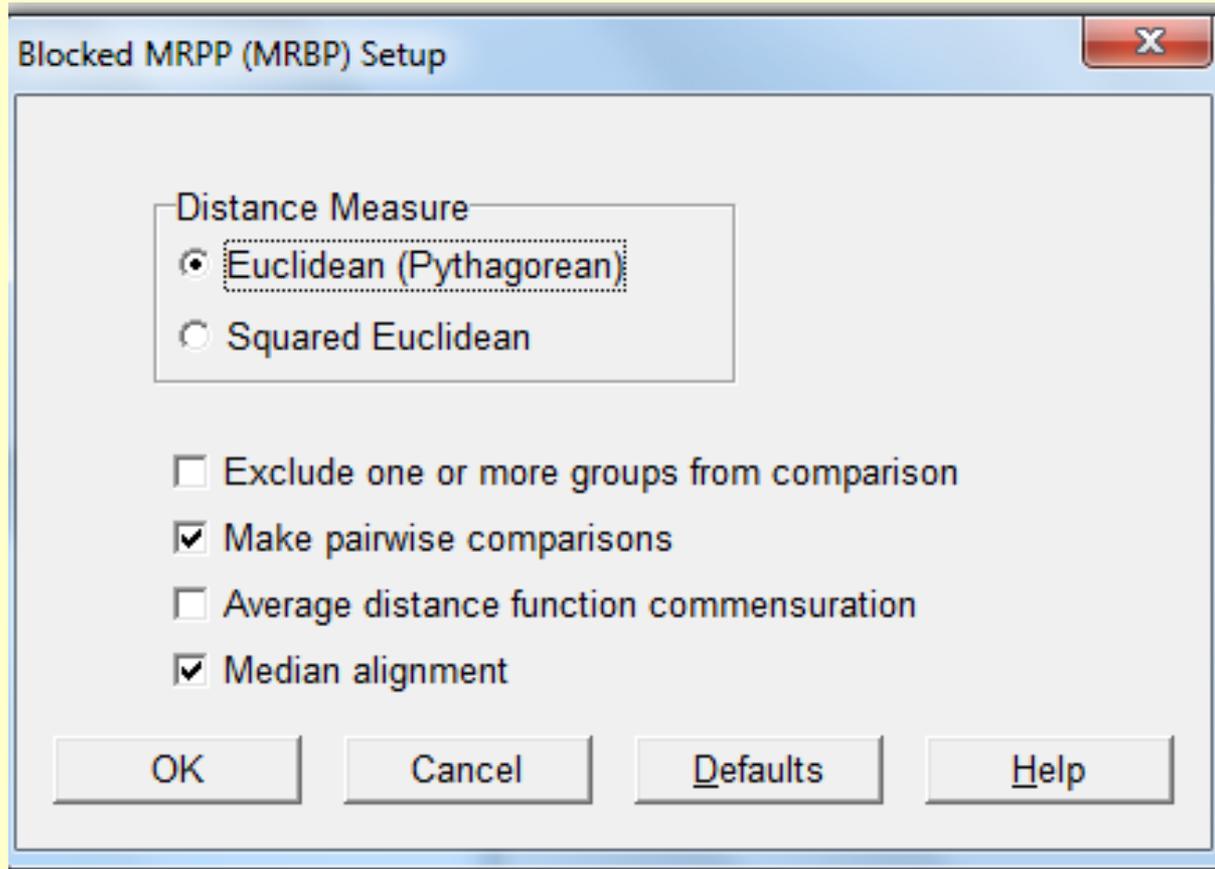
➤ Select Blocking Variable

• Matrix 2

• Categorical



MRBP SetUp



➤ Recommended Settings

- Pairwise comparisons
- Median alignment

Limitation: Euclidean Distance metric

MRBP

Median Alignment Within Blocks

If the median for each variable in each block is subtracted from the raw data for each block, then the medians are said to be aligned to zero for all blocks.

Alignment is desirable in a randomized block design.

But if the problem is conceptualized as paired agreement, say between a model prediction and observed data, then the median alignment is not performed.

MRBP

Median Alignment Within Blocks

Table: Comparison of raw data and data aligned within blocks to zero as input to Blocked MRPP.

	Raw Data		Aligned Data	
	Block 1	Block 2	Block 1	Block 2
Group 1	4	9	1.5	1.5
Group 2	2	7	-0.5	-0.5
Group 3	3	8	0.5	0.5
Group 4	1	2	-1.5	-5.5
Median	2.5	7.5	0	0
Observed δ	$5 = (5+5+5+1)/4$		$1 = (0+0+0+4)/4$	
Expected δ	4.375		2.225	
Agreement (A)	0.086		0.556	
p	0.184		0.016	

MRBP

Average Distance Function Commensuration

This option equalizes the contribution of each variable to the distance function. For each variable m sum of deviations (Dev_m) is calculated:

$$Dev_m = \sum_{i=1}^g \sum_{j=1}^d \sum_{k=1}^g \sum_{l=1}^d |x_{mij} - x_{mkl}|^V$$

The exponent V is set to 2 for squared Euclidean distance or 1 for Euclidean distance. Then each element x of the data matrix is divided by the sum of the deviations for the corresponding variable to produce the transformed value y :

$$y_{mij} = x_{mij} / Dev_m$$

MRBP Example – from PC-ORD

SETUP

Community response to forest thinning in blocks:

Analysis of randomized block data with MRBP:

39 variables (species), 5 blocks (sites) , 4 groups (times)

Groups defined by treatments: Treat

Blocks defined by Stands: StanBloc

Input data has: 20 plots by 39 attrib

OPTIONS

Distance measure: Euclidean

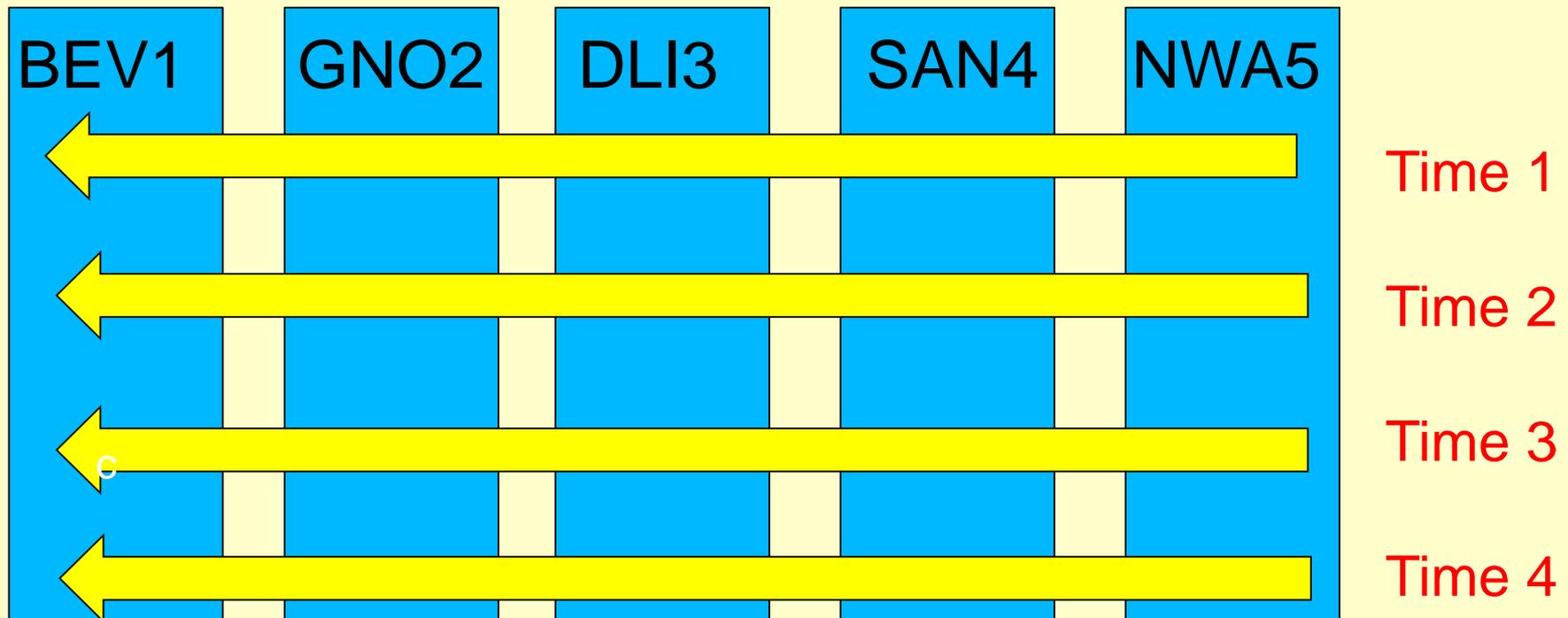
Median alignment performed

No average distance function commensuration

MRBP Example – from PC-ORD

QUESTION: 39 species in 5 forest patches (stanblocks):

Test effect of time (succession), by sampling 4 times
Control the individual differences in the 5 locations



MRBP Example – from PC-ORD

GROUP: 1

Identifier: 1 Size: 5

Members: BEV1-1 GNO2-1 DLI3-1 SAN4-1 NWA5-1

GROUP: 2

Identifier: 2 Size: 5

Members: BEV1-2 GNO2-2 DLI3-2 SAN4-2 NWA5-2

GROUP: 3

Identifier: 3 Size: 5

Members: BEV1-3 GNO2-3 DLI3-3 SAN4-3 NWA5-3

GROUP: 4

Identifier: 4 Size: 5

Members: BEV1-4 GNO2-4 DLI3-4 SAN4-4 NWA5-4

MRBP Example – from PC-ORD

Test statistic: $T = -4.1880177$

Observed delta = 6.9528671

Expected delta = 7.4888375

Variance of delta = 0.16378141E-01

Skewness of delta = -0.48910187

Chance-corrected within-group agreement, $A = 0.07156924$

$A = 1$ - (observed delta/expected delta)

$A_{max} = 1$ when all items are identical within groups (delta=0)

$A = 0$ when heterogeneity within groups equals expectation by chance

$A < 0$ with more heterogeneity within groups than expected by chance

Probability of a smaller or equal delta, $p = 0.00041511$

Multivariate Experiments

- Investigators can use careful sampling design to address gradients / changes in community composition.
- **For example:** outgroup poles / replicate samples

Natural Experiments

- Investigators can take advantage of natural variability to simulate manipulative treatments.
- **For example:** one hillside burned and another one did not

PerMANOVA - Final Thoughts

- Features: Ideal for evaluating specific hypotheses
 - Address multiple factors at once
 - Investigates interaction terms
- Limitations:
 - Requires careful blocked design / replication
 - Blocking allows for controlling additional factors
 - Nesting allows for scenarios without full replication

PerMANOVA – Experimental Design

- You must choose one of the following designs.
- For each design, specify the design variables (factors), by choosing them from a list of variables in second matrix.

One way:

Replicates within [_____] as grouping variable

Two-way factorial:

Replicates within [_____] and [_____] as grouping variables

PerMANOVA – Experimental Design

One fixed factor and one level nested (Mixed Model):

Replicates within [_____] nested within [_____] as grouping variable

Two levels nested (Model II):

Replicates nested within [_____] , nested within [_____]

Randomized complete blocks:

Blocks are [_____] , fixed factor [_____] is grouping variable

PerMANOVA & ANOSIM – References

- MANOVA:

Anderson, M. J. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 26:32-46.

- ANOSIM (perMANOVA with ranked data):

Clarke, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* 18, 117-143.

MRBP – References

Mielke, P. W., Jr., and K.J. Berry. 1982. An extended class of permutation techniques for matched pairs. *Commun. Statist.-Theor. Meth.* 11:1197-1207.

Mielke, P. W. and H. K. Iyer. 1982. Permutation techniques for analyzing multiresponse data from randomized block experiments. *Commun. Statist. A* 11:1427-1437.

Biondini, M.E., C.D. Bonham, and E.F. Redente. 1985. Secondary successional patterns in a sagebrush (*Artemisia tridentata*) community as they relate to soil disturbance and soil biological activity. *Vegetatio* 60: 25-36.

Zimmerman, G.M., H. Goetz, and P. W. Mielke, Jr. 1985. Use of an improved statistical method for group comparisons to study effects of prairie fire. *Ecology* 66: 606-611.