



# Тестирование гипотез на основе многомерных данных: PERMANOVA

Анализ и визуализация многомерных данных с  
использованием R

Вадим Хайтов, Марина Варфоломеева

- Еще раз про множественные сравнения
- Пример: Поведение песчанок в тесте открытое поле
- PERMANOVA
- Условия применимости PERMANOVA
- Более подробная интерпретация результатов perMANOVA
- Более сложные дизайны в PERMANOVA

### Вы сможете

- Разобраться с основными дизайнами тестирования гипотез в рамках дисперсионного анализа
- Применить функции, реализованные в R, для тестирования гипотез с помощью метода PERMANOVA

**Еще раз про множественные  
сравнения**

## В чем опасность множественных сравнений?

$\alpha$  — это **вероятность совершить ошибку первого рода при тестировании гипотезы** (= вероятность отвергнуть истинную нулевую гипотезу, = вероятность найти различия там, где их нет).

Обычно принимается, что  $H_0$  отвергают на уровне значимости  $\alpha = 0.05$ .

Когда у нас два средних — все просто, сравнение всего одно. Естественно, **вероятность совершить ошибку I рода для группы сравнений  $\alpha_{familywise}$  равна уровню значимости для единственного сравнения  $\alpha_{per\ comparison}$** .

$$\alpha_{familywise} = \alpha_{per\ comparison}$$

Но если сравнений много, то растет вероятность совершить хотябы одну ошибку I рода (найти различия там, где их нет).

## Если сравнений много...

Например, если мы хотим попарно сравнить три значения, нам понадобится сделать 3 сравнения.

Пусть мы решили, что в каждом из сравнений будем использовать уровень значимости  $\alpha_{per\ comparison} = 0.05$ .

Тогда в каждом из сравнений вероятность совершить ошибку первого рода будет  $\alpha_{per\ comparison} = 0.05$ .

Следовательно, вероятность сделать хотябы одну ошибку I рода в группе из 3 сравнений будет  $\alpha_{familywise} = 0.05 + 0.05 + 0.05 = 0.05 \cdot 3 = 0.15$

**Чем больше сравнений, тем больше вероятность обнаружить различия там, где их на самом деле нет.**

**Вероятность совершить хотябы одну ошибку I рода для всей группы из N сравнений  $\alpha_{familywise}$  будет в  $N$  раз больше, чем принятая для одного сравнения  $\alpha_{per\ comparison}$**

$$\alpha_{familywise} = \alpha_{per\ comparison} \cdot N_{comparisons}$$

В таблице даны значения  $\alpha_{familywise}$  для разного числа сравнений, если  $\alpha_{per\ comparison} = 0.05$ :

Число средних	Число сравнений	$\alpha_{familywise}$
2	1	0.05
3	3	0.15
4	6	0.30
5	10	0.50

---

## Для решения проблемы есть два подхода

1. Взять более жесткий порог уровня значимости. (Например, ввести поправку Бонферрони, поправку Хольма-Бонферрони, воспользоваться процедурой Бенджамини-Хохберга).
2. Изменить схему тестирования гипотезы — тестировать не три независимых гипотезы, а одну сложную (так это, например, происходит в ANOVA).

## Поправка Бонферрони

$\alpha_{per\ comparison} = \alpha_{familywise}/N$ , где  $N$  - количество попарных сравнений

Это жесткий способ, т.к с возрастанием числа сравнений приходится резко снижать уровень значимости, и значит, мощность теста тоже снижается. В результате, мы теперь больше рискуем не найти различий там, где они на самом деле есть.

В таблице даны значения  $\alpha_{per\ comparison}$  после поправки Бонферрони, сохраняющие  $\alpha_{familywise} = 0.05$ :

Число средних	Число сравнений	$\alpha_{per\ comparison}$
2	1	0.050
3	3	0.017
4	6	0.008
5	10	0.005
6	15	0.003
7	21	0.002
8	28	0.002
9	36	0.001

---

8/55



# Классический дисперсионный анализ



Рональд Эйлмер Фишер

Пусть имеется три градации фактора  $A_{1...3}$

- Почему появляется *межгрупповая* изменчивость, то есть разные средние значения для групп по фактору  $A$ ?
- Почему  $\bar{Y}_{gen}$  появляется *внутригрупповая* изменчивость, то есть разные значения  $y$  в группах?

**A1**

$y_{1,1}$

$y_{2,1}$

$y_{3,1}$

$\bar{Y}_{A1}$

**A2**

$y_{1,2}$

$y_{2,2}$

$y_{3,2}$

$\bar{Y}_{A2}$

**A3**

$y_{1,3}$

$y_{2,3}$

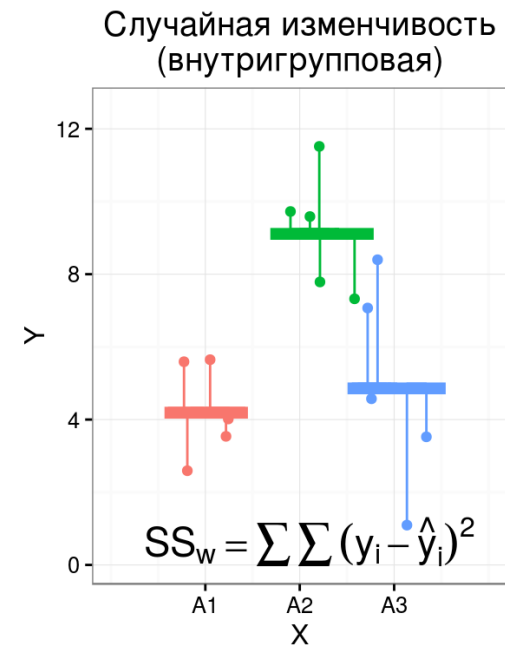
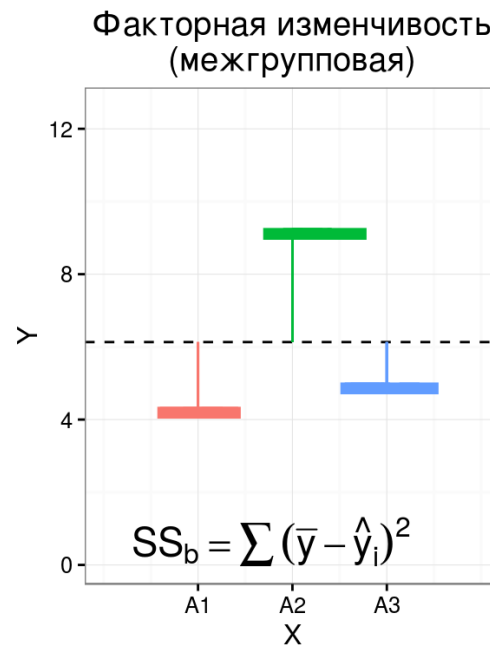
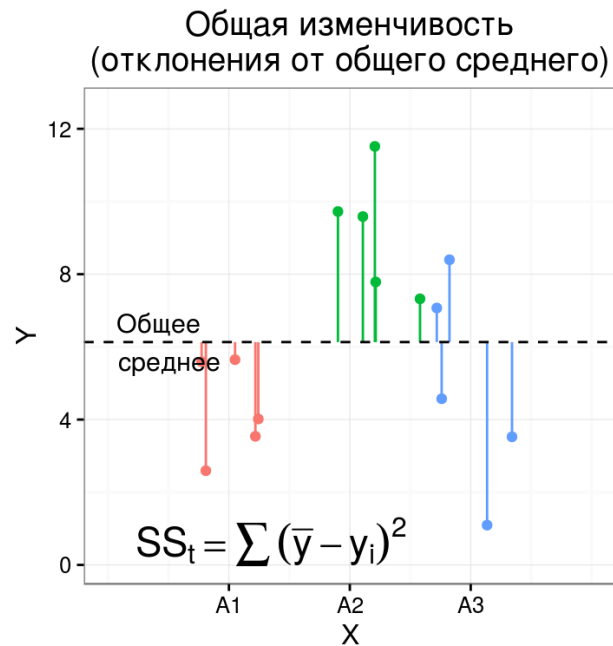
$y_{3,3}$

$\bar{Y}_{A3}$

$\bar{Y}_{gen}$

## Суммарная дисперсия может быть разложена на две составляющие

$$SS_{total} = SS_{between} + SS_{within}$$



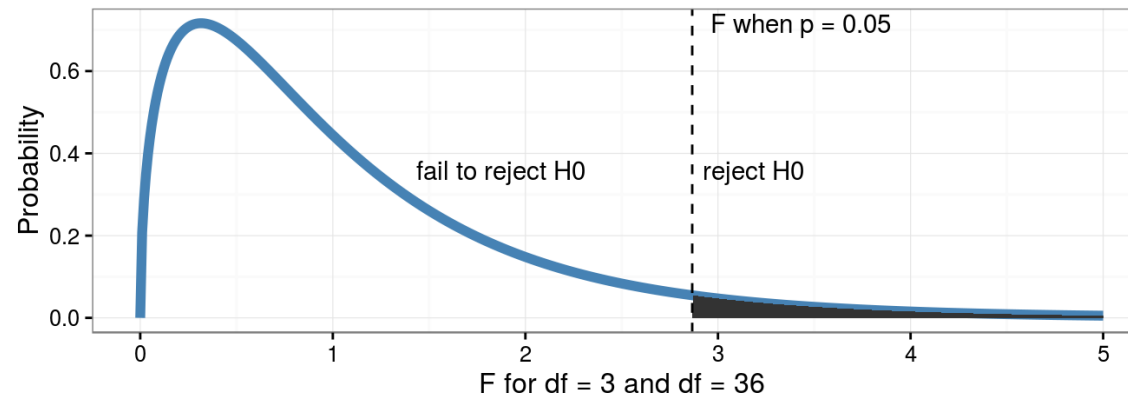
Для тестирования гипотезы о влиянии фактора надо сравнить межгрупповую изменчивость ( $SS_{between}$ ) и внутригрупповую ( $SS_{within}$ ).

Для этого сравнения Фишер предложил статистику

$$F = \frac{SS_{between}/(a-1)}{SS_{within}/(N-a)}$$

### F-распределение

Если межгрупповая изменчивость равна внутригрупповой, то  $F$  принадлежит  $F$ -распределению с двумя параметрами  $df_{between} = a - 1$  и  $df_{within} = N - a$ , где  $a$  — число классов,  $N$  — общее количество объектов в анализе.



## Основные дизайны в ANOVA

1. Однофакторный дизайн (One-way ANOVA)
2. Многофакторный ортогональный дизайн (Multi-way orthogonal ANOVA)
3. Иерархический дизайн (Nested ANOVA)  
etc...

Еще много разных дизайнов

Величину, которую мы изучаем называют зависимой переменной  
Величины, от которых она зависит называют предикторами (независимыми переменными), или факторами

Предикторы бывают дискретными и непрерывными

## Однофакторный дизайн (один дискретный предиктор)

A1	A2	A3
y	y	y
y	y	y
y	y	y

---

выявляется влияние фактора A

## Многофакторный ортогональный дизайн (два и более дискретных предиктора)

<i>factorA</i> <i>factorB</i>	A1	A2	A3
B1	y	y	y
B1	y	y	y
B1	y	y	y
B2	y	y	y
B2	y	y	y
B2	y	y	y
B3	y	y	y
B3	y	y	y
B3	y	y	y

---

Выявляется влияние фактора A, B и A\*B

## Иерархический дизайн

один или более предикторов, иерархически починенные некоторому группирующему фактору, "блоку"

Блок <i>B</i>	A1	A2	A3
B1	y	y	y
B1	y	y	y
B1	y	y	y
B2	y	y	y
B2	y	y	y
B2	y	y	y
B3	y	y	y
B3	y	y	y
B3	y	y	y

---

Выявляется влияние фактора В и фактора А внутри градаций фактора В

## ANOVA разработан для одномерных данных.

Что делать если мы хотим оценивать объект по многим признакам сразу?

Примеры:

- Сообщество как целое
- Поведение как целое
- Ответы респондентов на взаимосвязанные вопросы в анкетах
- Варианты решений:
- MANOVA (Fisher, 1925, Wilks, 1932)
- distance-based Redundancy Analysis (db-RDA) (Legendre, Anderson, 1999)
- PERMANOVA (Anderson, 2001; McArdle, Anderson, 2001)

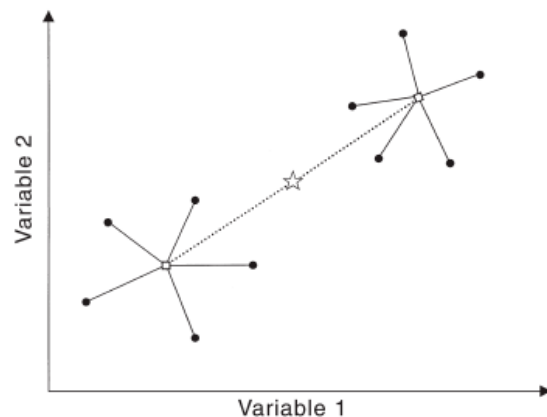


# Многомерный дисперсионный анализ (MANOVA)

Давно разработан параметрический метод MANOVA (Multivariate Analysis Of Variance). Он дает возможность проводить анализ аналогичный ANOVA. В основе MANOVA лежат представление о многомерном нормальном распределении и расстояниях между центроидами.

В MANOVA сравниваются:

- **отклонения точек от групповых центроидов** (аналог  $SS_{within}$ )
- **отклонения групповых центроидов от общего центроида** (аналог  $SS_{between}$ ).



Ограничения MANOVA:

- Нормальность распределения
- Гомогенность дисперсий

# distance-based Redundancy Analysis (dbRDA)

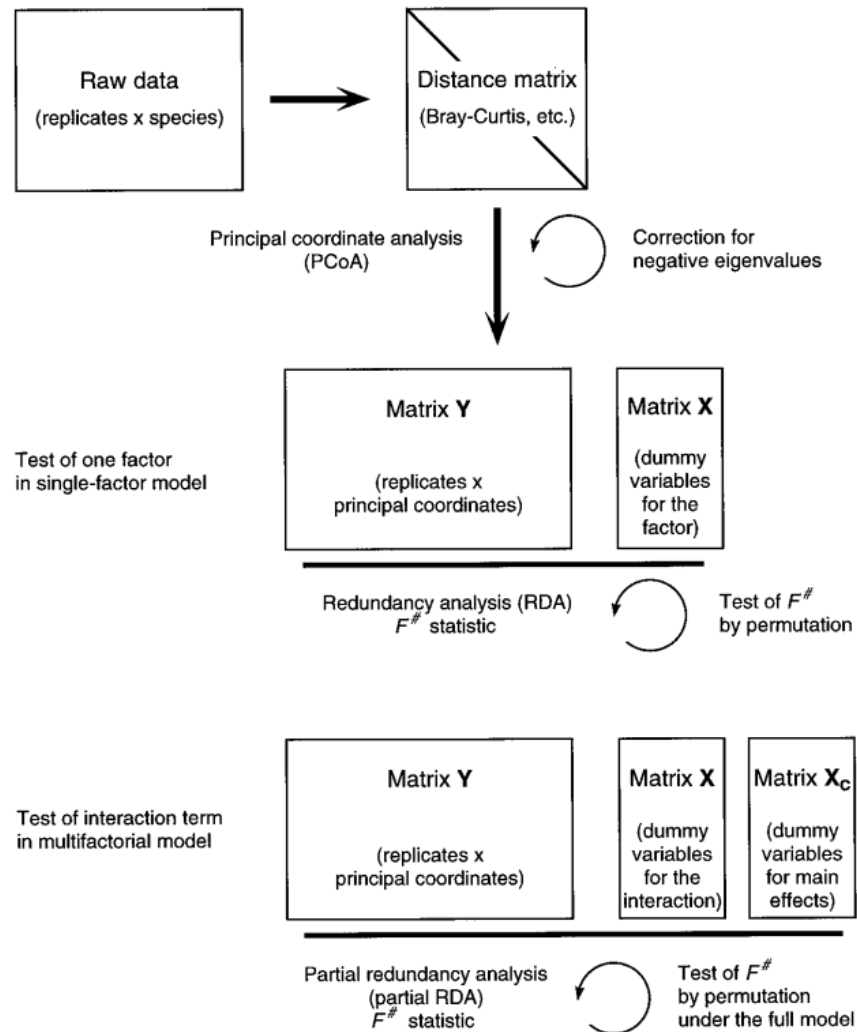


Рис. из Legendre, Anderson, 1999.

Подробнее о db-RDA расскажем в конце курса.

# Permutational Multivariate Analysis of Variance (PERMANOVA)



Марти Джейн Андерсон

*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 Laboratories A11,  
University of Sydney, New South Wales 2006, Australia*

**Abstract** Hypothesis-testing methods for multivariate data are needed to make rigorous probability statements about the effects of factors and their interactions in experiments. Analysis of variance is particularly powerful for the analysis of univariate data. The traditional multivariate analogues, however, are too stringent in their assumptions for most ecological multivariate data sets. Non-parametric methods, based on permutation tests, are preferable.

**Пример: Поведение песчанок в  
тесте открытое поле**

## Пример: Поведение песчанок в тесте открытое поле

**Гипотеза:** Разные виды песчанок демонстрируют различия поведения в тесте "Открытое поле"



(*Pachyuromys duprasi*)

Оценка поведения песчанок трех видов по семи признакам

- Время до выхода в квадрат открытого поля
- Количество актов мочеиспускания
- Количество актов дефекации
- Количество пересеченных квадратов
- Число вертикальных стоек
- Количество актов смещенной активности
- Время проведенное в центре квадрата открытого поля

Виды:

- Карликовая песчанка (*Gerbillus gerbillus*)
- Монгольская песчанка (*Meriones unguiculatus*)
- Жирнохвостая песчанка

## Данные наблюдений

```
pesch <- read.csv("data/pesch.csv", header = TRUE, sep = ";")  
head(pesch)
```

```
##  Gender Species Time_to_entrance Urination Defecation Quadr_Number  
## 1      f    karl              0          1          0           47  
## 2      f    karl             20          0          3          317  
## 3      m    karl            181          0          0          177  
## 4      f    karl              0          0          0           32  
## 5      f    karl            139          0          0          205  
## 6      m    karl              0          0          2           38  
##  Vert_Number Displ_Act Time_in_Centre  
## 1           11         4              0  
## 2           58         3              6  
## 3           55         6              2  
## 4            5         3              0  
## 5           29        10              3  
## 6            9         8              0
```

## Задание

- Какое расстояние можно использовать с этими данными?
- Как можно преобразовать данные?
- Постройте ординацию объектов в осях MDS и раскрасьте точки в соответствии с видами

## Решение

Поскольку измеренные признаки варьируют в разных масштабах целесообразно провести лгарифмирование данных

```
options(digits = 4)
log_pesch <- pesch
log_pesch[, 3:ncol(pesch)] <- log(pesch[, 3:ncol(pesch)] +1)
head(log_pesch)
```

```
## Gender Species Time_to_entrance Urination Defecation Quadr_Number
## 1      f    karl           0.000      0.6931      0.000      3.871
## 2      f    karl           3.045      0.0000      1.386      5.762
## 3      m    karl           5.204      0.0000      0.000      5.182
## 4      f    karl           0.000      0.0000      0.000      3.497
## 5      f    karl           4.942      0.0000      0.000      5.328
## 6      m    karl           0.000      0.0000      1.099      3.664
## Vert_Number Displ_Act Time_in_Centre
## 1          2.485      1.609           0.000
## 2          4.078      1.386           1.946
## 3          4.025      1.946           1.099
## 4          1.792      1.386           0.000
## 5          3.401      2.398           1.386
## 6          2.303      2.197           0.000
```

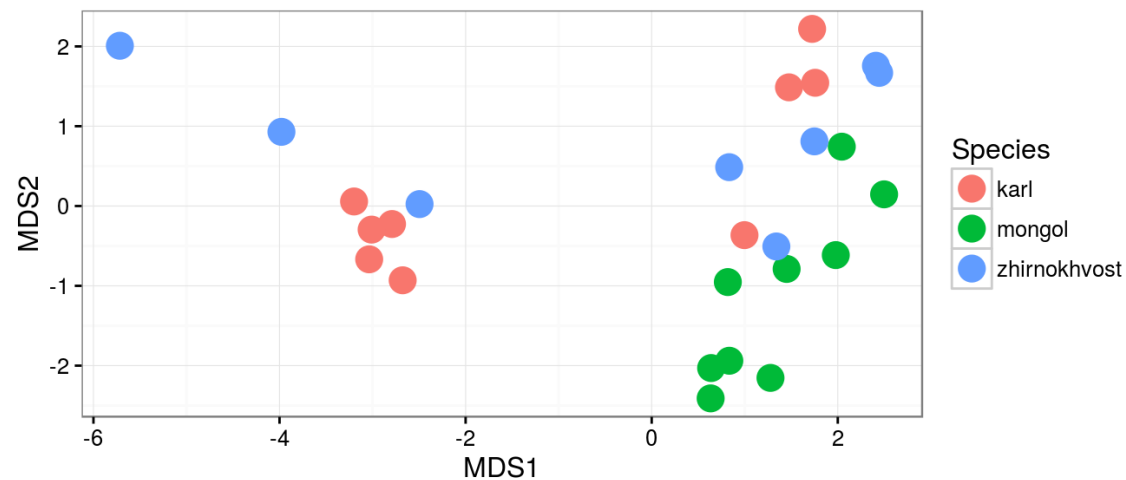


## Решение

```
library(ggplot2)
library(vegan)
theme_set(theme_bw())
mds_pesch <- metaMDS(log_pesch[, 3:ncol(pesch)], distance = "euclidian")
```

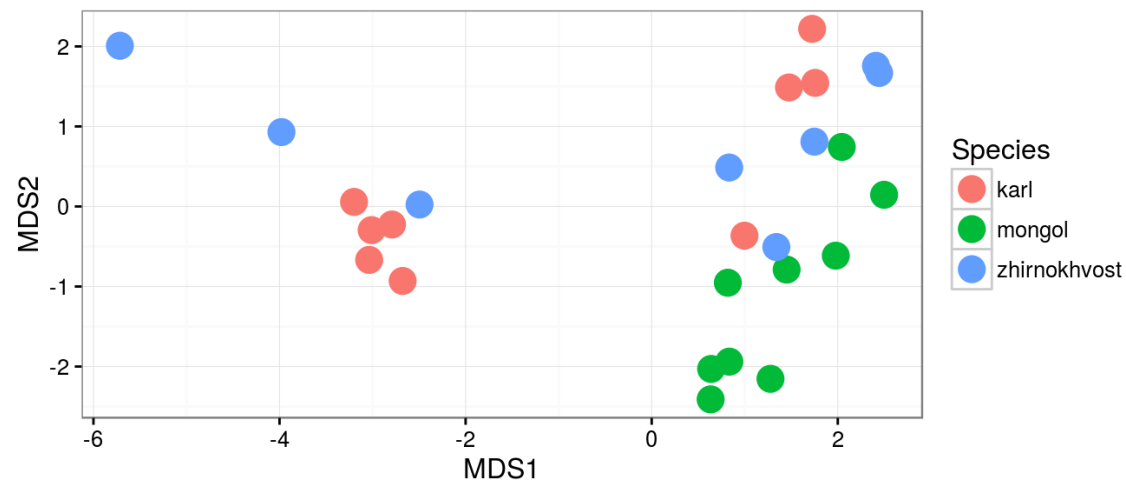
```
## Run 0 stress 0.0593
## Run 1 stress 0.0593
## ... procrustes: rmse 0.000352  max resid 0.00151
## *** Solution reached
```

```
mds_pesch <- as.data.frame(mds_pesch$points)
mds_pesch$Species <- pesch$Species
ggplot(mds_pesch, aes(x = MDS1, y = MDS2, colour = Species)) + geom_point(size = 5)
```



## Различаются ли виды песчанок по набору поведенческих признаков?

```
## Run 0 stress 0.0593
## Run 1 stress 0.0593
## ... procrustes: rmse 0.000375  max resid 0.00163
## *** Solution reached
```

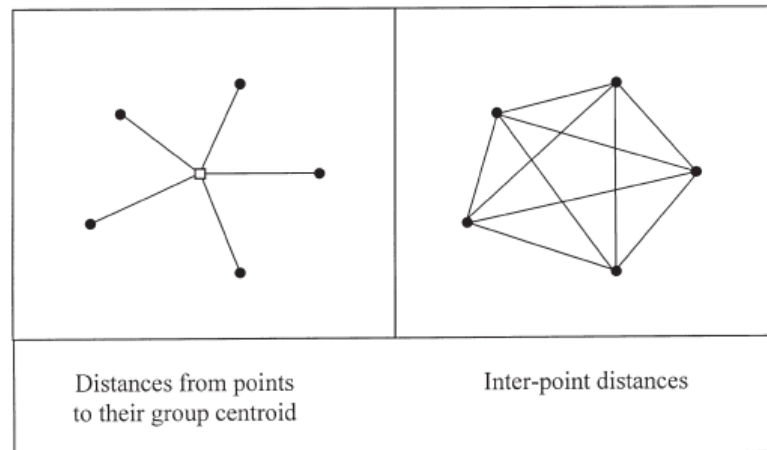


- Мы можем проверить это при помощи PERMANOVA

**PERMANOVA**

## Сумма квадратов отклонений объектов от центроидов равна сумме квадратов взаимных расстояний, деленной на число объектов

Для Евклидовых расстояний эта закономерность была известна давно (например, Kendall, Stuart 1963).

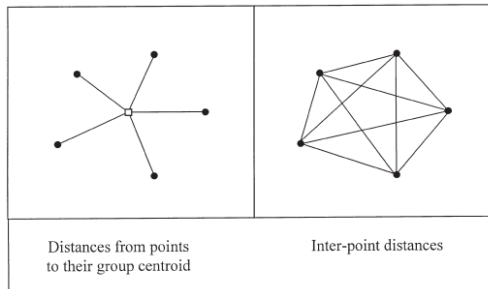


В случае Евклидова расстояния (именно его имплицитно использует MANOVA) центроиды найти очень просто — это средние значения соответствующих координат. Поэтому обычно сначала непосредственно вычисляли центроиды, и затем — сумму квадратов отклонений от них.

Для других мер сходства-различия центроиды найти гораздо сложнее. Например, для коэффициента Брея-Куртиса (не метрика), среднее значение не будет соответствовать центроиду.

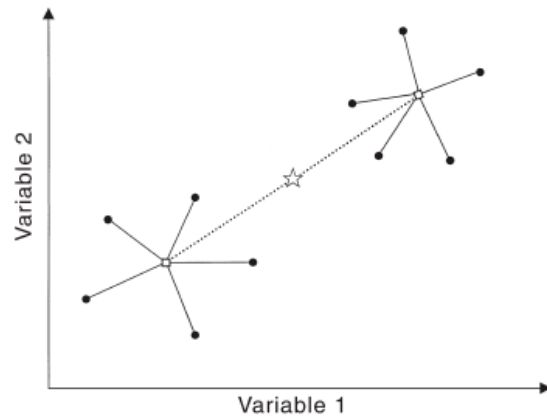
Anderson, 2001

## Марти Антерсон показала, что можно обойтись без вычисления центроидов и для других расстояний



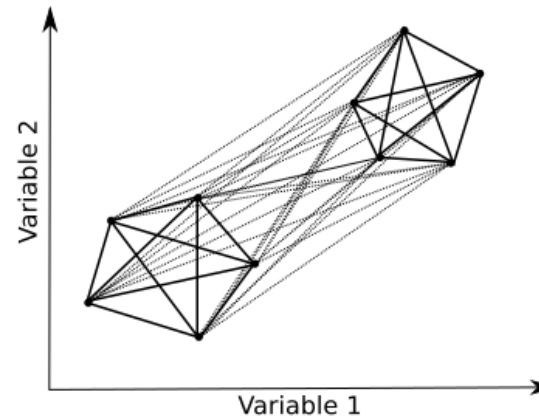
Anderson, 2001

MANOVA (Евклидово расстояние)

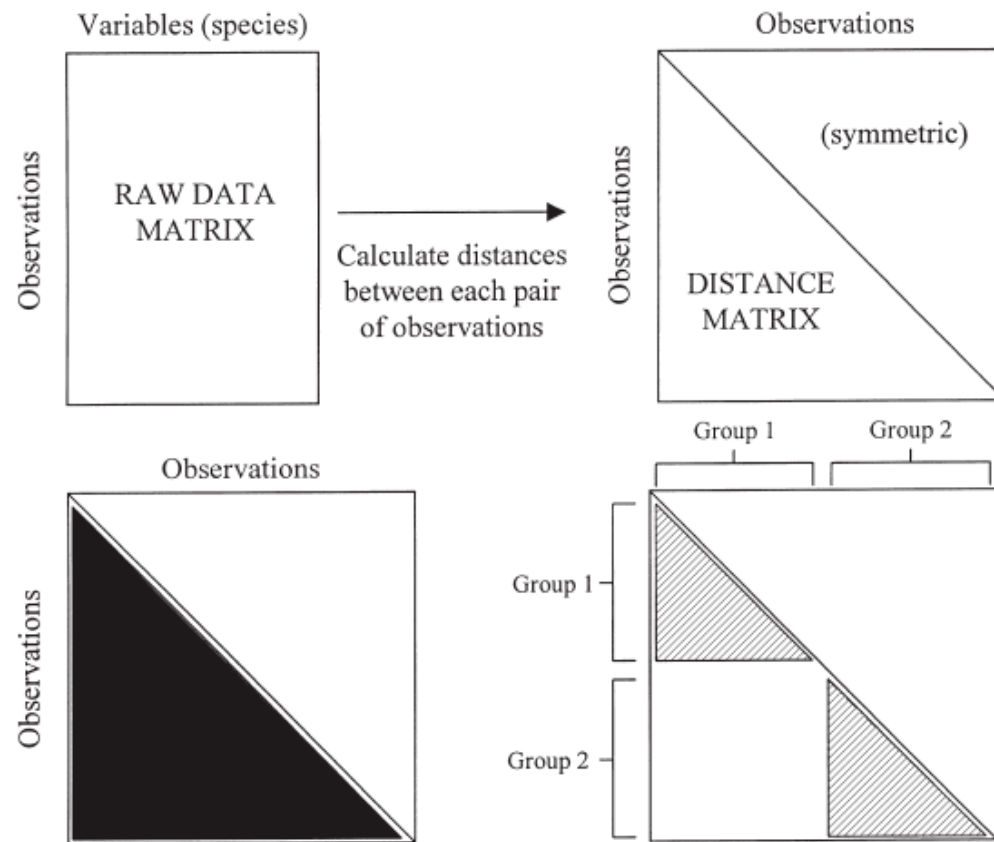


Anderson, 2001

PERMANOVA (любой коэффициент)

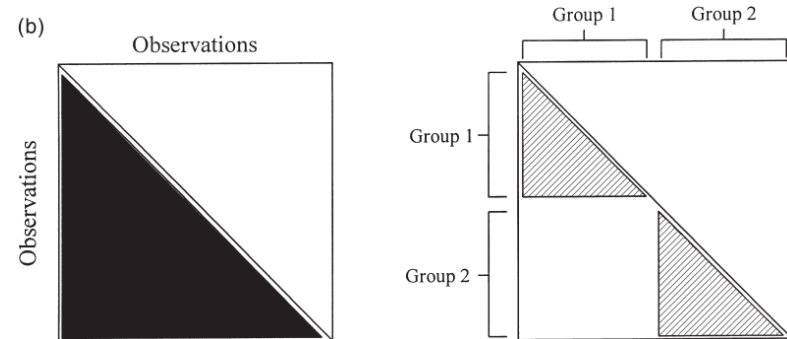


## Можно непосредственно из матрицы любых коэффициентов сходства-различия найти и общую и внутригрупповые суммы квадратов



Anderson, 2001

## Разложение дисперсии становится очень простым



Anderson, 2001

Пусть всего  $N$  элементов, принадлежащих  $a$  группам по  $n$  элементов в каждой,  $d$  - расстояние между  $i$ -тым и  $j$ -тым объектами,  $\epsilon$  - 1 если объекты  $i$  и  $j$  из одной группы и 0, если из разных.

$$SS_{total} = \frac{1}{N} \sum \sum d_{ij}^2$$

Сумма квадратов взаимных расстояний — это сумма квадратов субдиагональных элементов, деленная на число объектов  $N$ .

$$SS_{within} = \frac{1}{n} \sum \sum d_{ij}^2 \epsilon_{ij}$$

Внутригрупповая сумма квадратов — это сумма всех сумм квадратов расстояний между элементами для каждой группы, деленная на  $n$  число объектов в группе.

Тогда межгрупповая сумма квадратов  $SS_{between} = SS_{total} - SS_{within}$

## Псевдо-F статистика

$$F = \frac{SS_{between}/(a - 1)}{SS_{within}/(N - a)}$$

Для оценки значимости псевдо-F используется пермутационная процедура:

- Случайным образом перетасовываются строки исходной матрицы
- После каждого акта пермутации вычисляется  $F_{perm}$
- Уровень значимости
- Внимание! Для гнездового дизайна процедура пермутации имеет свои особенности (обсудим позднее).

$$p = \frac{N_{F_{perm} \geq F}}{N_{permutations}}$$



## Применим метод PERMANOVA, реализованный в функции `adonis()`

```
library(vegan)
permanova_pesch <- adonis(log_pesch[3:9] ~ log_pesch$Species, method = "euclidean")
permanova_pesch
```

```
##
## Call:
## adonis(formula = log_pesch[3:9] ~ log_pesch$Species, method = "euclidean")
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##              Df SumsOfSqs MeanSqs F.Model    R2 Pr(>F)
## log_pesch$Species  2         60   30.02   3.58 0.237 0.012 *
## Residuals        23        193    8.39         0.763
## Total            25        253         1.000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Мы видим традиционную для ANOVA таблицу результатов. Что здесь что?

## Результаты PERMANOVA

```
##
## Call:
## adonis(formula = log_pesch[3:9] ~ log_pesch$Species, method = "euclidean")
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##              Df SumsOfSqs MeanSqs F.Model    R2 Pr(>F)
## log_pesch$Species  2      60   30.02   3.58 0.237 0.012 *
## Residuals        23     193    8.39         0.763
## Total            25     253         1.000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Поведение разных видов песчанок в тесте открытое поле достоверно различалось (PERMANOVA,  $p = 0.012$ )

# **Условия применимости PERMANOVA**

## Условия применимости PERMANOVA

1. Этот метод чувствителен к различиям в степени разброса значений в пределах групп (гетероскедастичность).
2. Желательно использование сбалансированных комплексов (с равным объемом групп):
  - Во-первых, в этом случае PERMANOVA устойчив к гетерогенности дисперсии (Anderson, Walsh, 2013)
  - Во-вторых, если данные сбалансированы, то можно легко реализовать *post hoc* тест. В алгоритме, который реализован в `adonis()`, такого ограничения нет, но за это приходится платить отсутствием функции для *post hoc* анализа.

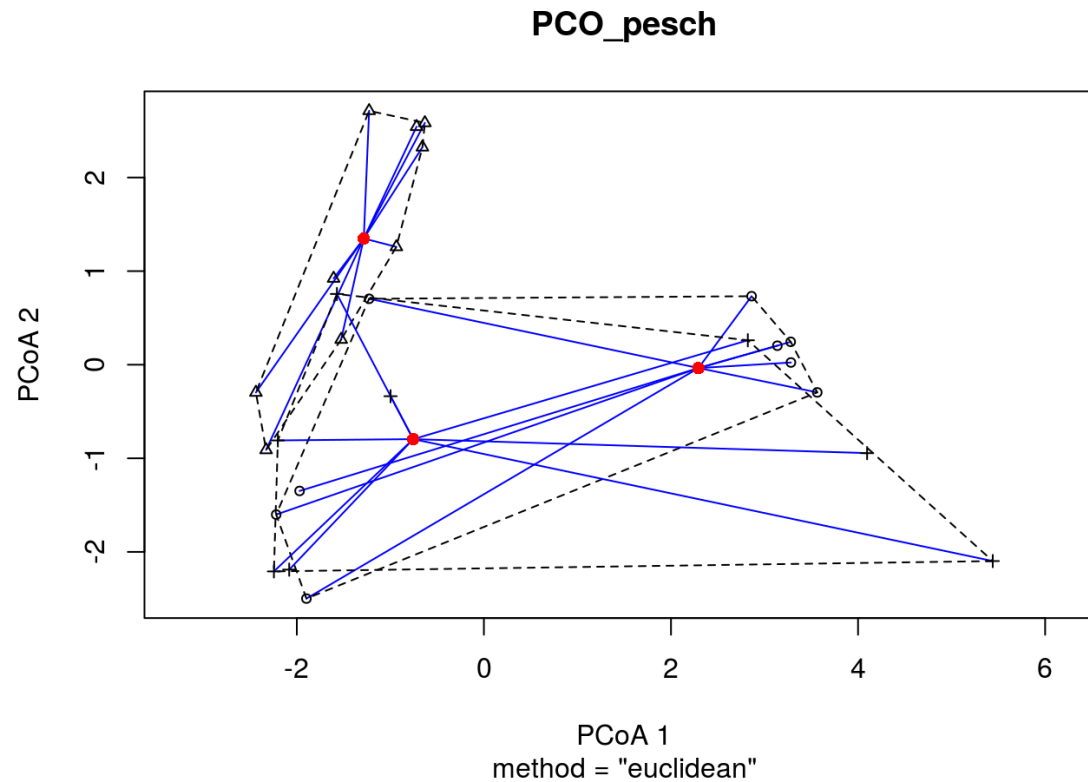
## Проверка равенства внутригрупповых дисперсий

Для проверки равенства внутригрупповых дисперсий в пакете `vegan` специальной функции нет, но можно использовать функцию `betadisper()`, которая изначально предназначена для сравнения  $\beta$ -разнообразия сообществ в местах с разными условиями

Эта функция вычисляет внутригрупповые центроиды и координаты точек в пространстве главных координат (*Principal coordinates analysis* = *PCoA* = *metric MDS*)

```
dist_pesch <- vegdist(log_pesch[,3:ncol(pesch)], method = "euclidian")
PCO_pesch <- betadisper(dist_pesch, log_pesch$Species)
```

Объект, возвращаемый `betadisper()`, позволяет также нарисовать наши объекты в пространстве PCO



## Достоверность различий отклонений от центроидов в разных группах проверяется с помощью процедуры PERMDISP2

Процедура PERMDISP2 реализована в пакете `vegan` в функции `anova.betadisper()`. Это многомерный аналог теста Левина на гомогенность дисперсий в группах, который, например, используется для проверки условий применимости дисперсионного анализа.

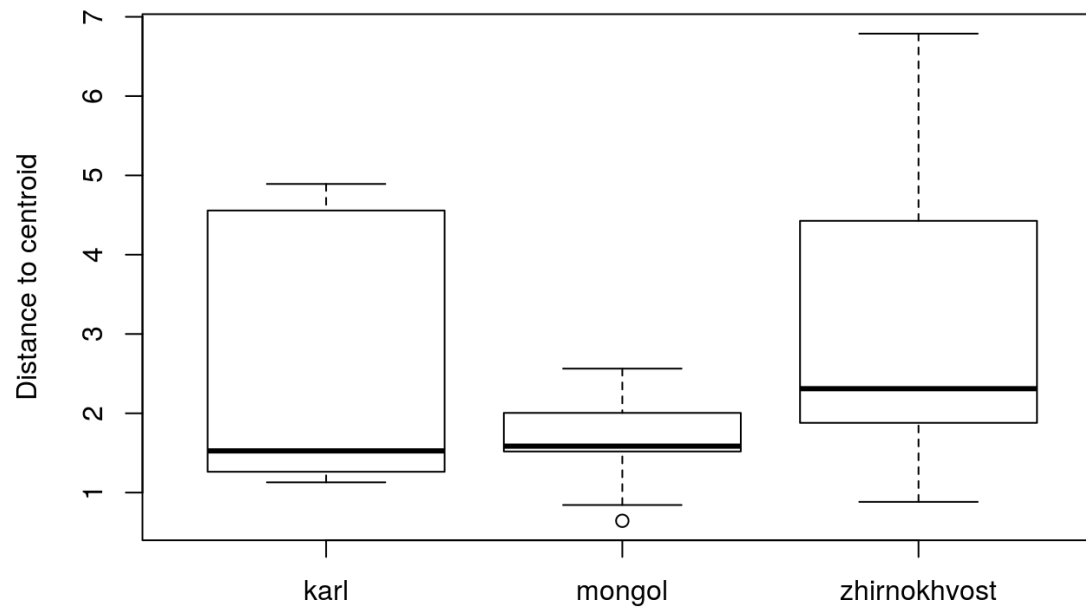
```
anova(PCO_pesch)
```

```
## Analysis of Variance Table
##
## Response: Distances
##      Df Sum Sq Mean Sq F value Pr(>F)
## Groups    2    9.7    4.86    2.05  0.15
## Residuals 23   54.5    2.37
```

В нашем случае достоверных различий разброса внутригрупповых расстояний не выявлено

## Для визуализации можно нарисовать боксплот

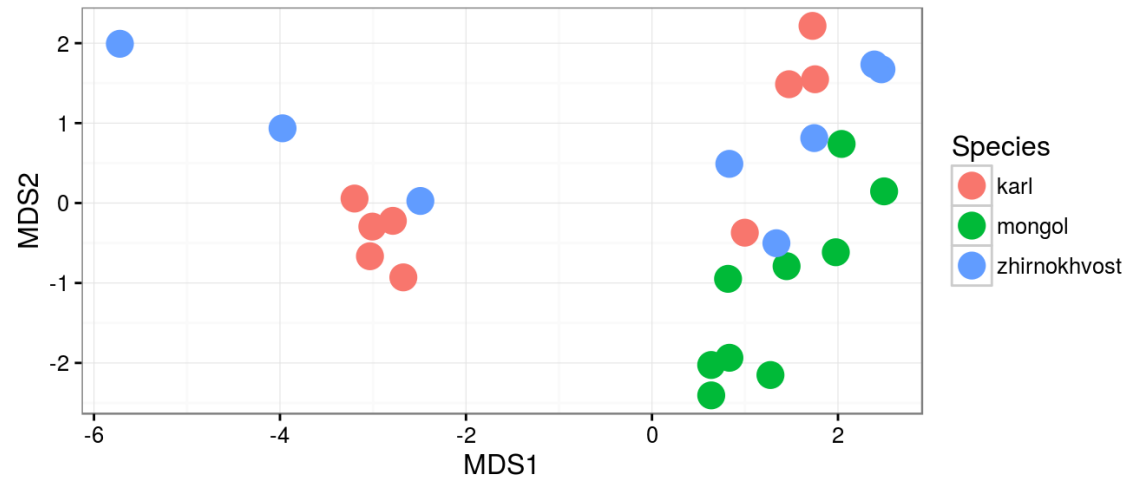
`boxplot(PC0_pesch)`





**Более подробная интерпретация  
результатов perMANOVA**

## Post hoc тесты в perMANOVA



На приведенной ординации видно, что точки, соответствующие Монгольским песчанкам сегрегированы от остальных.

Для выявления попарных различий нужны попарные сравнения.

## Post-hoc тест

Внимание! В пакете `vegan` пост хок тест не реализован. Но мы можем сделать простейшую его версию самостоятельно.

Проведем попарные сравнения между группами, то есть

- Карликовые VS Монгольские
- Карликовые VS Жирнохвостые
- Монгольские VS Жирнохвостые

```
pair <- combn(unique(as.character(log_pesch$Species)), 2)
ncomb <- dim(pair)[2]
x <- log_pesch[, -c(1:2)]
y <- log_pesch$Species
for (i in 1:ncomb) {
  filter <- y %in% pair[, i]
  posthoc <- adonis(x[filter, ] ~ y[filter], method = "euclidean")$aov.tab$Pr[1]
  cat(pair[, i], ": p = ", posthoc, "\n", sep = " ")
}
```

```
## karl mongol : p = 0.001
## karl zhirnokhvost : p = 0.401
## mongol zhirnokhvost : p = 0.005
```

## Введем поправку Бонферрони

Если мы принимаем  $\alpha = 0.05$ , как порог для отвержения  $H_0$ , то для отвержения  $H_0$  при множественных сравнениях  $\alpha$  становится

$$\alpha = \frac{0.05}{\text{Number of comparisons}}$$

Для нашего случая, когда используется 3 сравнения,  $\alpha = \frac{0.05}{3} = 0.017$

То есть за достоверные будем принимать те значения  $p$ , которые будут меньше 0.017, а не 0.05

## Задание

Выясните, какой из признаков сильнее всего различается между видами?

## Решение

Один из вариантов ответа — процедура SIMPER

```
simper_pesch <- simper(log_pesch[, 3:9], log_pesch$Species, permutations = 999)
summary(simper_pesch)
```

```
##
## Contrast: karl_mongol
##
##          contr      sd ratio  av.a  av.b cumsum      p
## Time_to_entrance 0.0861 0.0500  1.72 2.089 2.802  0.253 0.802
## Time_in_Centre   0.0741 0.0364  2.04 0.569 2.686  0.471 0.001 ***
## Quadr_Number     0.0572 0.0353  1.62 4.435 6.015  0.639 0.699
## Vert_Number      0.0374 0.0273  1.37 2.932 3.850  0.749 0.719
## Displ_Act        0.0367 0.0231  1.59 1.913 0.840  0.857 0.598
## Defecation       0.0279 0.0206  1.36 0.398 0.947  0.939 0.103
## Urination        0.0208 0.0141  1.47 0.154 0.719  1.000 0.033 *
##
## Contrast: karl_zhirnokhvost
##
##          contr      sd ratio  av.a  av.b cumsum      p
## Time_to_entrance 0.1089 0.0985  1.106 2.089 2.7794  0.281 0.043 *
## Quadr_Number     0.0841 0.0912  0.922 4.435 4.2944  0.498 0.118
## Displ_Act        0.0667 0.0565  1.180 1.913 0.6931  0.670 0.001 ***
## Vert_Number      0.0638 0.0691  0.923 2.932 2.9993  0.834 0.047 *
## Time_in_Centre   0.0341 0.0305  1.118 0.569 0.8503  0.922 1.000
## Defecation       0.0191 0.0271  0.703 0.398 0.0866  0.971 0.940
## Urination        0.0111 0.0181  0.616 0.154 0.1373  1.000 0.996
##
## Contrast: mongol_zhirnokhvost
##
"""
```

**Более сложные дизайны в  
PERMANOVA**

## Многофакторный ортогональный дизайн в PERMANOVA

Выясним, влияет ли пол и вид полевок на поведение.

Отфильтруем исходные данные (в случае с жирнохвостыми полевками были изучены только самки)

```
log_pesch2 <- log_pesch[log_pesch$Species != "zhirnokhvest", ]
```



## Проведем двухфакторный анализ PERMANOVA

Различается ли поведение песчанок в зависимости от видовой принадлежности и пола?

```
twofact_pesch <- adonis(log_pesch2[,3:ncol(pesch)] ~ log_pesch2$Gender *  
log_pesch2$Species, method = "euclidian")
```

```
twofact_pesch
```

```
##  
## Call:  
## adonis(formula = log_pesch2[, 3:ncol(pesch)] ~ log_pesch2$Gender *  
log_pesch2$Species, method = "euclidian")  
##  
## Permutation: free  
## Number of permutations: 999  
##  
## Terms added sequentially (first to last)  
##  
##
```

	Df	SumsOfSqs	MeanSqs	F.Model
## log_pesch2\$Gender	1	7.2	7.2	1.13
## log_pesch2\$Species	1	45.5	45.5	7.16
## log_pesch2\$Gender:log_pesch2\$Species	1	4.5	4.5	0.72
## Residuals	14	88.9	6.4	
## Total	17	146.1		

```
##
```

	R2	Pr(>F)
## log_pesch2\$Gender	0.049	0.322
## log_pesch2\$Species	0.311	0.001 ***
## log_pesch2\$Gender:log_pesch2\$Species	0.031	0.528
## Residuals	0.608	
## Total	1.000	

## Здесь возможен иерархический дизайн

Различается ли поведение самцов и самок у этих видов песчанок?

```
nested_pesch <- adonis(log_pesch2[, 3:ncol(pesch)] ~ log_pesch2$Gender, strata =  
log_pesch2$Species, method = "euclidian")
```

```
nested_pesch
```

```
##  
## Call:  
## adonis(formula = log_pesch2[, 3:ncol(pesch)] ~ log_pesch2$Gender,      method =  
"euclidian", strata = log_pesch2$Species)  
##  
## Blocks: strata  
## Permutation: free  
## Number of permutations: 999  
##  
## Terms added sequentially (first to last)  
##  
##          Df SumsOfSqs MeanSqs F.Model    R2 Pr(>F)  
## log_pesch2$Gender 1      7.2   7.21   0.83 0.049  0.27  
## Residuals      16    138.9   8.68           0.951  
## Total          17    146.1           1.000
```

Внимание! Пермутации производятся только в пределах группирующего параметра, указанного в параметре strata

## Задание

- Создайте датафрейм из файла `simulated_data.csv` (Это данные симулированные по алгоритму, приведенному в справке по функции `adonis()`)
- В полученном датафрейме описано обилие двух видов на экспериментальных площадках двух типов: без добавления и с добавлением  $\text{NO}_3$ , по 6 повторнотей в каждом эксперименте. Эксперименты были независимо проведены на 3 полях.
- Оцените, зависит ли структура совместного поселения этих двух видов от концентрации  $\text{NO}_3$ .

## Решение

```
com <- read.csv("data/simulated_data.csv", sep = ',', header = T)
```

```
# Ошибочный дизайн
```

```
com_permanova <- adonis(com[,1:2] ~ com$N03)
```

```
# Правильный дизайн
```

```
com_permanova2 <- adonis(com[,1:2] ~ com$N03, strata = com$field)
```

## Summary

- При одновременном тестировании нескольких гипотез растет вероятность ошибки I рода.
- Чтобы контролировать вероятность ошибки I рода, либо используют более жесткий уровень значимости, либо тестируют сложную гипотезу вместо нескольких простых.
- PERMANOVA дает возможность тестировать сложные гипотезы в отношении явлений, описанных по многим переменным (т.е. на многомерных данных).
- В PERMANOVA можно использовать любые коэффициенты сходства-различия.
- Для применения PERMANOVA требуется равенство разбросов точек между центроидами их групп, но при равных объемах групп анализ устойчив к отклонениям от этого условия.
- При использовании PERMANOVA важно не запутаться в дизайне.

## Другие программы

- *Primer 6.0 + PERMANOVA* Коммерческий продукт. Пожалуй самая удобная программа для применения метода PERMANOVA.
- *PAST* Здесь метод называется NPMANOVA. Из некоммерческих продуктов, пожалуй, самый лучший. R пока отстаёт...
- *Оригинальная программа М. Андерсон (PERMANOVA)*. Можно скачать из сети. К ней прилагается программа *PERMDISP*, предназначенная для проверки на равенство дисперсий. Чудовищно неудобный ввод. Зато можно использовать почти любой дизайн (однофакторный, многофакторный, иерархический, с ковариатами). Важно, что факторы могут рассматриваться и как фиксированные, и как случайные.

## Что почитать

- Anderson, M.J. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26: 32–46.
- Anderson, M.J. 2005. PERMANOVA: a FORTRAN computer program for permutational multivariate analysis of variance. Department of Statistics, University of Auckland, New Zealand.
- Anderson, M.J. (2004). PERMDISP: a FORTRAN computer program for permutational analysis of multivariate dispersions (for any two-factor ANOVA design) using permutation tests. Department of Statistics, University of Auckland, New Zealand.
- Legendre P., Legendre L. (2012) Numerical ecology. Second english edition. Elsevier, Amsterdam.