

РАЗНООБРАЗИЕ И СТРУКТУРА СООБЩЕСТВ: ВВЕДЕНИЕ В МЕТОДЫ АНАЛИЗА

ВСЕ, ЧЕГО НЕЛЬЗЯ ВЫРАЗИТЬ В ЦИФРАХ,
НЕ НАУКА, А ПРОСТО МНЕНИЕ

Роберт Хайнлайн

ЗАКОНЫ МАТЕМАТИКИ, ИМЕЮЩИЕ КАКОЕ-ЛИБО
ОТНОШЕНИЕ К РЕАЛЬНОМУ МИРУ, НЕНАДЕЖНЫ;
А НАДЕЖНЫЕ МАТЕМАТИЧЕСКИЕ ЗАКОНЫ
НЕ ИМЕЮТ ОТНОШЕНИЯ К РЕАЛЬНОМУ МИРУ

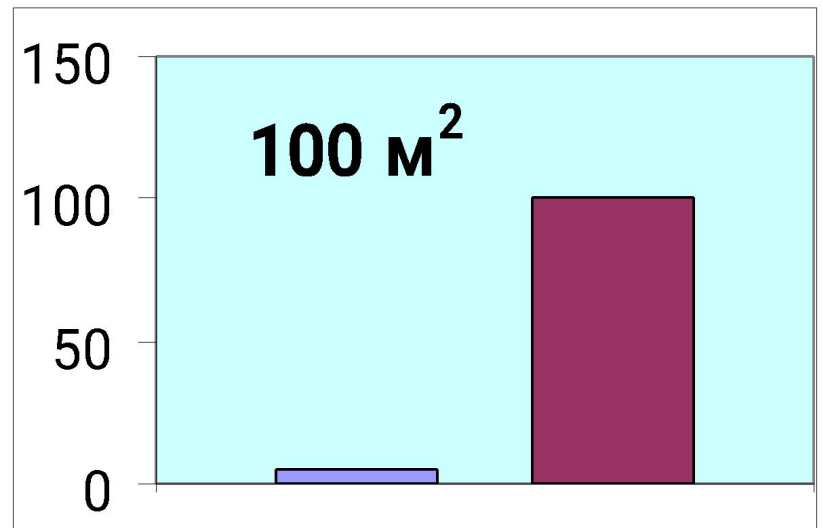
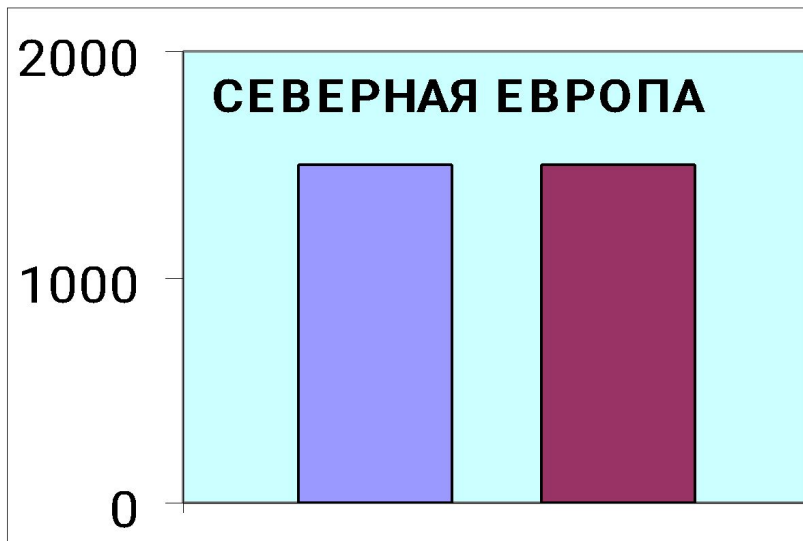
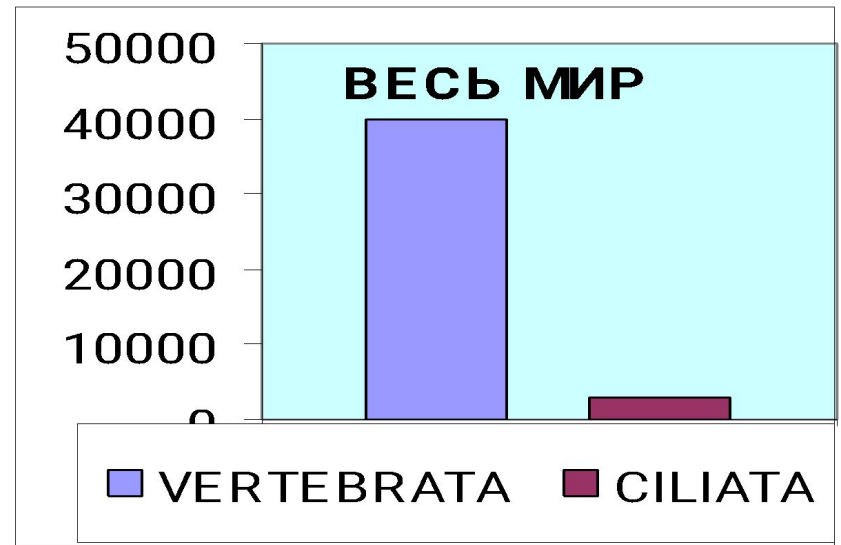
Альберт Эйнштейн

ГОРАЗДО ЛЕГЧЕ ЧТО-ТО ИЗМЕРИТЬ, ЧЕМ ПОНЯТЬ,
ЧТО ИМЕННО ВЫ ИЗМЕРЯЕТЕ

Дж. Салливен

ВИДОВОЕ БОГАТСТВО: КАК ЕГО ОЦЕНИТЬ?

- КАКАЯ ГРУППА БОГАЧЕ ВИДАМИ: ИНФУЗОРИИ ИЛИ ПОЗВОНОЧНЫЕ?



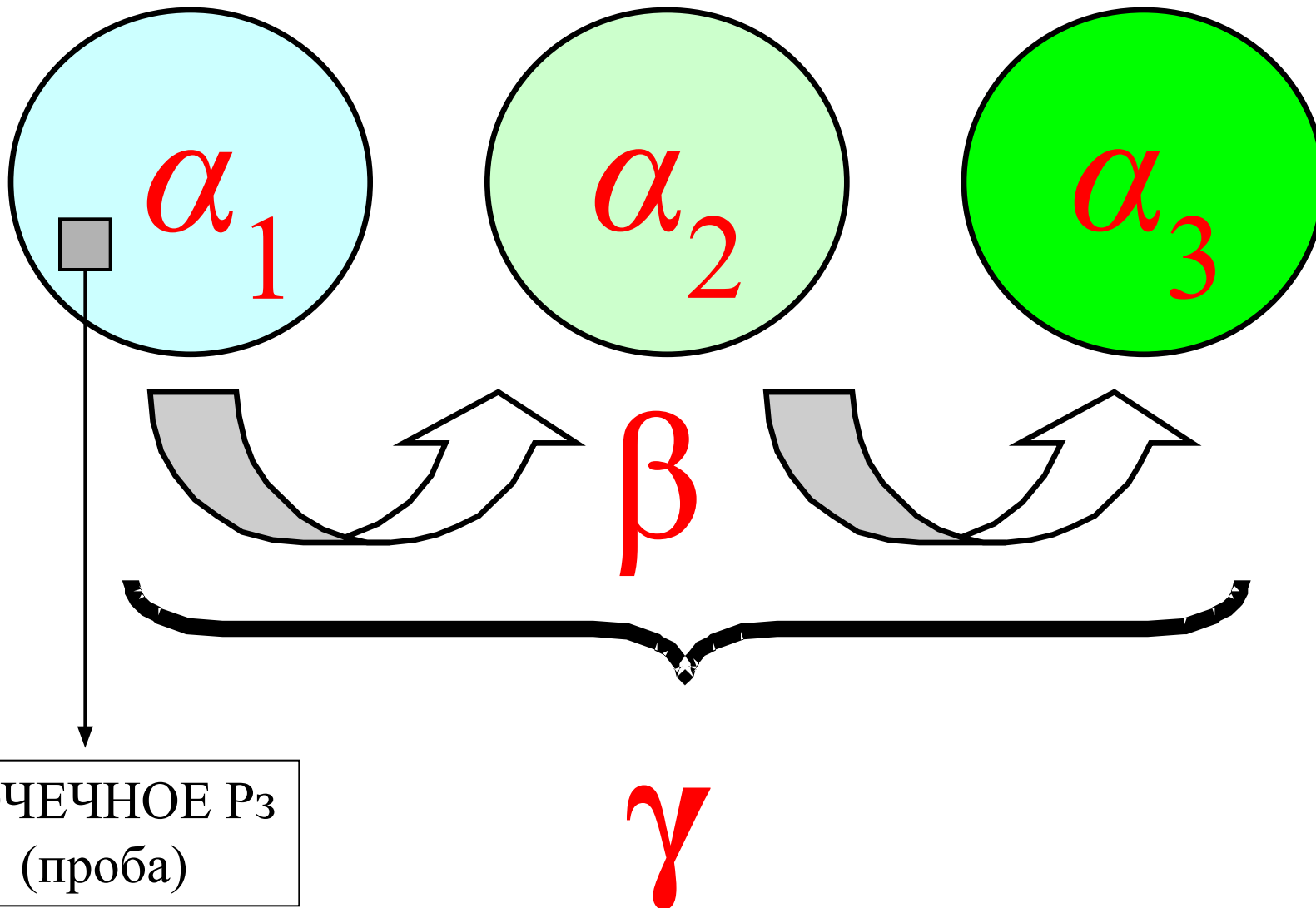
ВИДОВОЕ БОГАТСТВО ЗАВИСИТ ОТ МАСШТАБА



ЕГО НУЖНО НОРМИРОВАТЬ. КАК?

- НА ПЛОЩАДЬ (НА m^2 ? НА ГЕКТАР?)
- НА РАЗМЕР ОСОБИ?
- НА ЧИСЛО ОСОБЕЙ?

УРОВНИ РАЗНООБРАЗИЯ по УИТТЕКЕРУ



НАКОПЛЕНИЕ ВИДОВ («КРИВАЯ СБОРЩИКА»)

Число
ВИДОВ
 S

ВИДОВОЕ БОГАТСТВО

$$d_{MENH} = \frac{S}{\sqrt{N}}$$

$$d_{MARG} = \frac{S-1}{\log N}$$

$$d = \frac{\log S}{\log N}$$

ЧИСЛО ВИДОВ:

- на пробу: S_{sample}
- на n особей: $ES(n)$

Объем выборки, N
(пробы или особи)

ОЦЕНКИ β -РАЗНООБРАЗИЯ

□ ПО СООТНОШЕНИЮ α - и γ -РАЗНООБРАЗИЯ:

$$\beta = \gamma - \alpha \text{ (аддитивный метод)}$$

$$\beta = 1 - \alpha / \gamma \text{ (мультипликативный метод)}$$

НЕДОСТАТОК: ЗАВИСИТ ОТ ОБЪЕМА ВЫБОРКИ

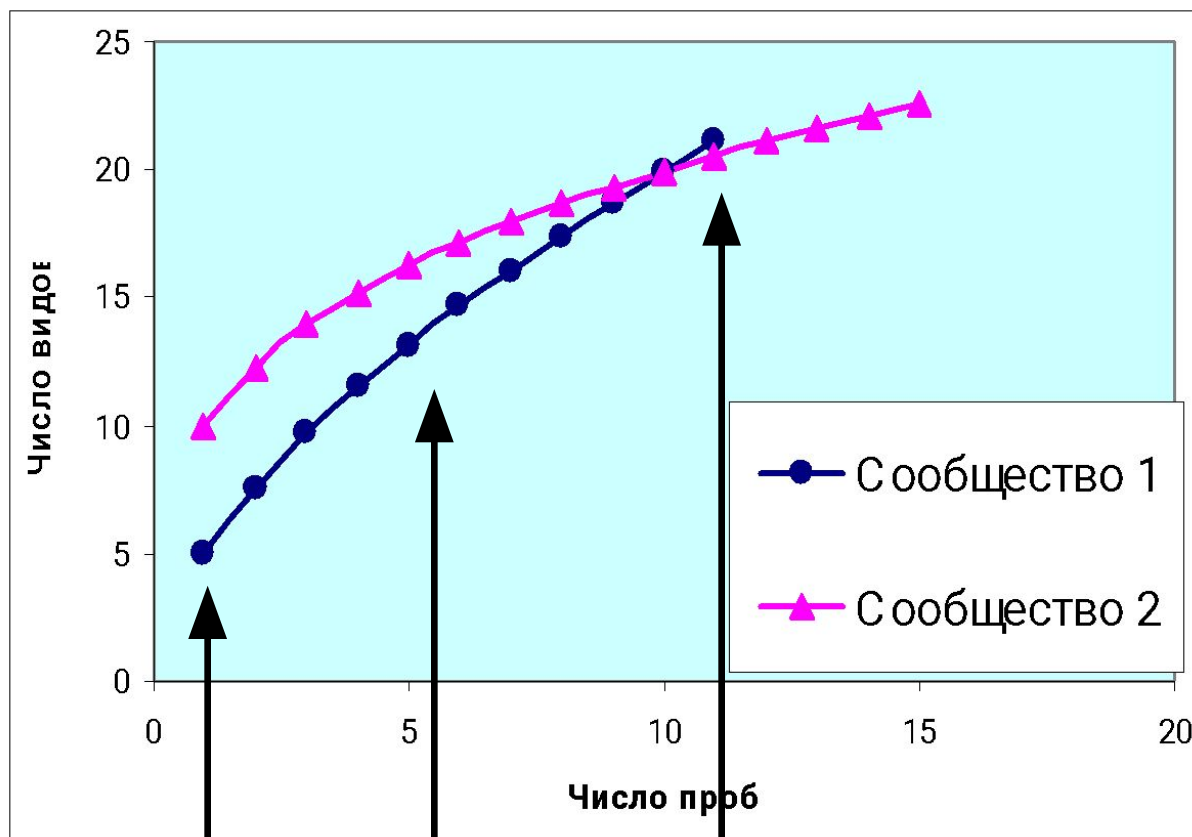
□ НА ОСНОВЕ ИНДЕКСОВ СХОДСТВА: $\beta = 1 - SIMIL$

НЕДОСТАТОК: ЗАВИСИТ ОТ ИНДЕКСА

□ ПО СКОРОСТИ РОСТА КРИВОЙ НАКОПЛЕНИЯ ВИДОВ

НЕДОСТАТКИ: 1) ЗАВИСИТ ОТ СПОСОБА АППРОКСИМАЦИИ КРИВОЙ; 2) НАДО СЧИТАТЬ ВРУЧНУЮ

СРАВНЕНИЕ КРИВЫХ НАКОПЛЕНИЯ ВИДОВ



РАСЧЕТ ПО
СТЕПЕННОЙ
АППРОКСИМАЦИИ:

$$S = a N^b$$

$$\log S = \log a + b \log N$$

$$a = \alpha;$$

$$b = \beta \text{ (угол наклона)}$$

$$\alpha_2 > \alpha_1$$

$$\beta_1 > \beta_2$$

$\gamma_1 = \gamma_2$ (при
сопоставимых объемах
выборки)

ОЦЕНКИ «ПОЛНОГО» ЧИСЛА ВИДОВ ПО ВЫБОРКЕ

Пусть взято N проб, вид найден в n проб.

Встречаемость такого вида

(вероятность найти его в пробе): n/N

Вероятность НЕ найти его в пробе: $1-n/N$

Вероятность пропустить такой вид (не найти ни в одной из проб): $(1-n/N)^N$

Метод Chao (оценка полного числа видов с учетом «пропущенных»):

$$S_{\text{ПОЛН}} = S + \frac{S_1(S_1 - 1)}{2(S_2 + 1)}$$

ТАКСОНОМИЧЕСКОЕ РАЗНООБРАЗИЕ

ПРОБА 1:



ПРОБА 2:



Индекс Среднего
Таксономического Разнообразия: $\Lambda^+ = \left(2 \sum \omega_{ij} \right) / (S (S - 1))$

АНАЛИЗ СХОДСТВА

R ВИДОВ \times Q ПРОБ



R \times R ВИДОВ



Q \times Q ПРОБ

ТРЕБОВАНИЯ К МЕРАМ СХОДСТВА

- МАСШТАБ: $0 \leq S \leq 1$
- СИММЕТРИЯ: $S_{AB} = S_{BA}$
- ВЕРХНИЙ ПРЕДЕЛ: $S_{AA} = 1$
- НИЖНИЙ ПРЕДЕЛ:
 $S_{AB} = 0$, ЕСЛИ $A \cap B = \emptyset$

СХОДСТВО ДЛЯ КАЧЕСТВЕННЫХ ДАННЫХ (ЕСТЬ/НЕТ)

4-ХПОЛЬНАЯ ТАБЛИЦА

		I ПРОБА	
		+	-
II ПРОБА	+	a	b
	-	c	d

$$R_I = a + c \quad R_{II} = a + b$$

$$R_{I+II} = a + b + c$$

$$R_{\text{ОБЩ}} = a + b + c + d$$

ИНДЕКСЫ СХОДСТВА ДЛЯ КАЧЕСТВЕННЫХ ДАННЫХ

- СЪЁРЕНСЕНА-ЧЕКАНОВСКОГО

$$S_{s-cz} = \frac{2a}{2a + b + c} \quad \left(= \frac{a}{\frac{(a+b) + (a+c)}{2}} \right)$$

ЧУВСТВИТЕЛЕН

ПРИ НИЗКОМ СХОДСТВЕ ($a < b, c$)

- ЖАККАРА $S_J = \frac{a}{a + b + c}$

ЧУВСТВИТЕЛЕН ПРИ

ВЫСОКОМ СХОДСТВЕ ($a > b, c$)

ИНДЕКСЫ СХОДСТВА ДЛЯ КАЧЕСТВЕННЫХ ДАННЫХ

- СИМПСОНА
$$S_{SIM} = \frac{a}{a + \min(b, c)}$$

НЕЧУВСТВИТЕЛЕН К РАЗЛИЧИЯМ В ДЛИНЕ СПИСКОВ ($a+b \gg a+c$)

- БАРОНИ-УРБАНИ и БЮССЭ

$$S_{BUB} = \frac{a + \sqrt{ad}}{a + b + c + \sqrt{ad}}$$

ЧУВСТВИТЕЛЕН К КЛЕТКЕ d !

ИНДЕКСЫ СХОДСТВА ДЛЯ КОЛИЧЕСТВЕННЫХ ДАННЫХ

- БРЕЯ-КЁРТИСА (BRAY-CURTIS):

$$S_{B-C} = 1 - \frac{\sum_R |n_{1i} - n_{2i}|}{\sum_R (n_{1i} + n_{2i})} \quad \left(= \sum_R \min(p_{1i}; p_{2i}) \right)$$

- ПИАНКИ (PIANKA) (ЧУВСТВИТЕЛЕН К РАЗЛИЧИЯМ В ДОМИНАНТАХ)

$$S_{PI} = \frac{\sum n_{1i} \cdot n_{2i}}{\sqrt{\sum n_{1i}^2 \cdot \sum n_{2i}^2}}$$

- ЭВКЛИДОВО РАССТОЯНИЕ:

$$D_{EU} = \sqrt{\sum (n_{1i} - n_{2i})^2}$$

МЕТОДЫ МНОГОМЕРНОЙ СТАТИСТИКИ (много объектов со многими признаками)

РАЗБИЕНИЕ НА КЛАССЫ

ТРЕБУЕТСЯ *a priori*

КЛАССИФИКАЦИЯ
(кластер-анализ)

НЕ ТРЕБУЕТСЯ

ДААННЫЕ О ФАКТОРАХ
СРЕДЫ ИСПОЛЬЗУЮТСЯ?

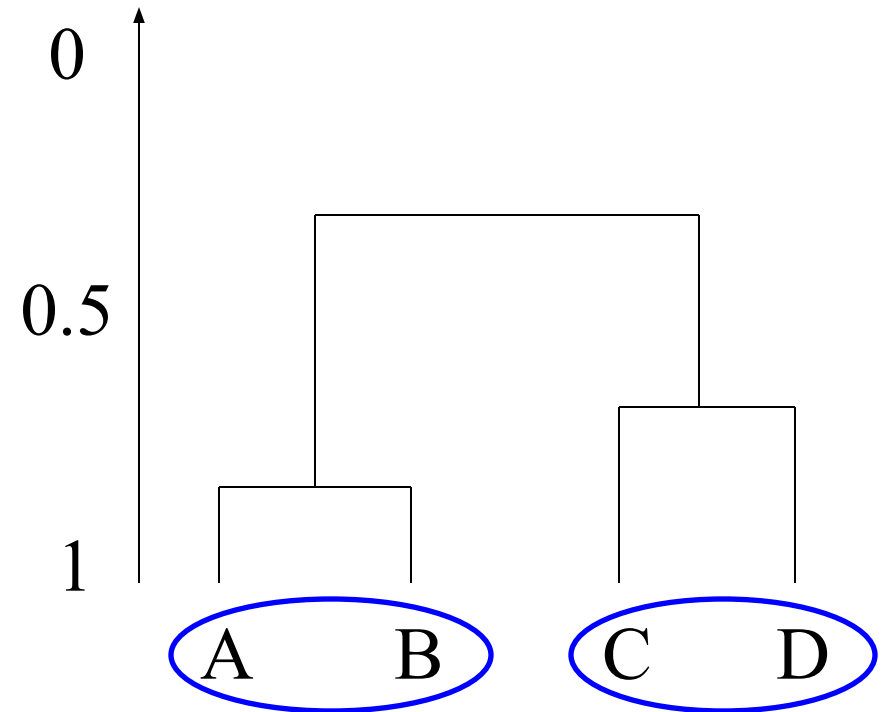
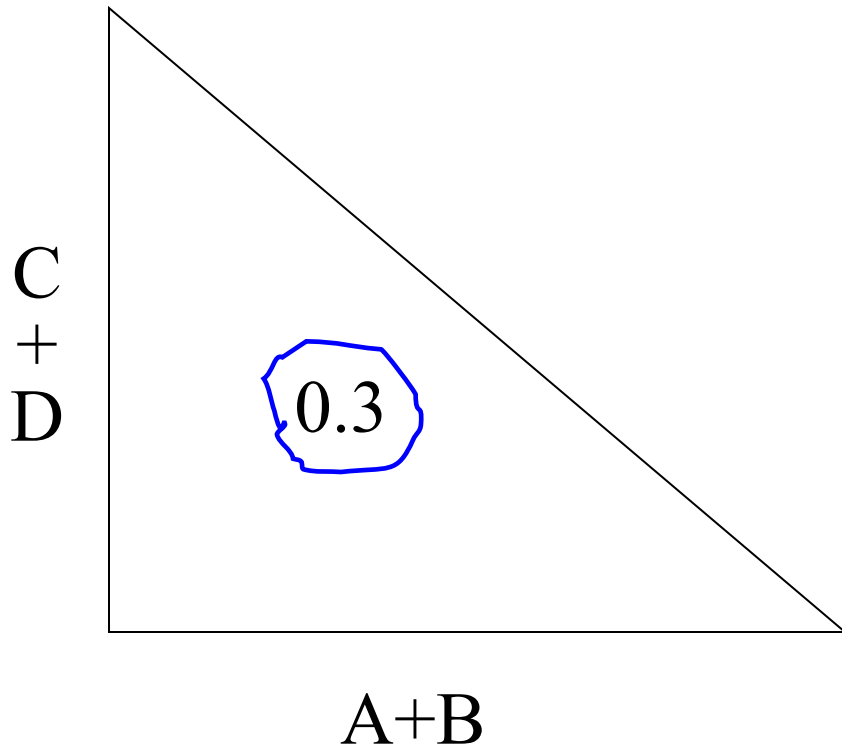
ДА

ПРЯМОЙ ГРАДИЕНТНЫЙ
АНАЛИЗ (ССА, RSM)

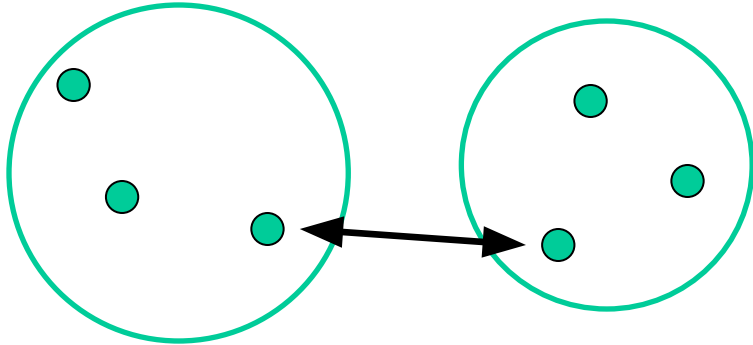
НЕТ

ОРДИНАЦИЯ
(РСА, СА, DСА, MDS)

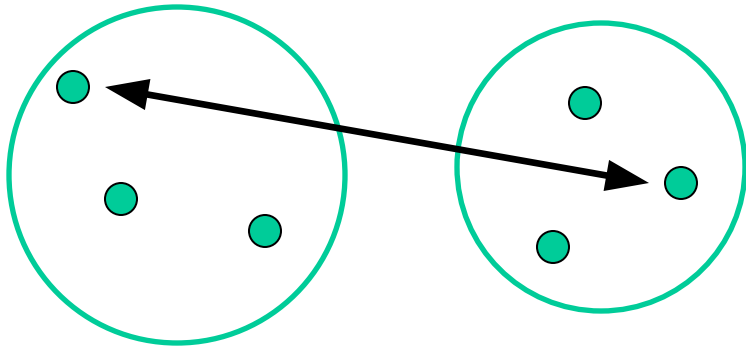
ИЕРАРХИЧЕСКАЯ КЛАССИФИКАЦИЯ: КЛАСТЕР-АНАЛИЗ (CLUSTER-ANALYSIS)



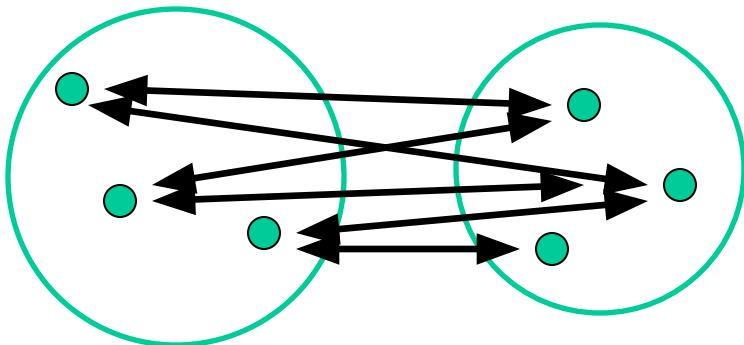
СПОСОБЫ ОБЪЕДИНЕНИЯ ГРУПП ОБЪЕКТОВ



МЕТОД
БЛИЖАЙШЕГО СОСЕДА
(SINGLE LINKAGE)



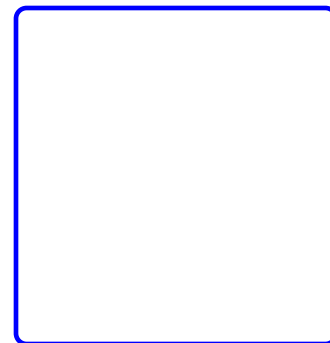
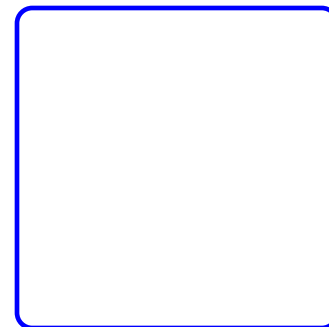
МЕТОД
ДАЛЬНЕГО СОСЕДА
(COMPLETE LINKAGE)



МЕТОД СРЕДНЕГО
ПРИСОЕДИНЕНИЯ
(GROUP AVERAGE)

МЕТОД
БЛИЖАЙШЕГО
СОСЕДА

МЕТОД
ДАЛЬНЕГО
СОСЕДА



МЕТОДЫ ОРДИНАЦИИ

ИСХОДНАЯ ИДЕЯ: ЛЮБОЙ ОБЪЕКТ С n ПРИЗНАКАМИ
МОЖНО ПРЕДСТАВИТЬ КАК ТОЧКУ В n -МЕРНОМ ПРОСТРАНСТВЕ

ЗАДАЧИ:

- ОТРАЗИТЬ ВЗАИМНОЕ ПОЛОЖЕНИЕ ОБЪЕКТОВ
- УМЕНЬШИТЬ РАЗМЕРНОСТЬ ПРОСТРАНСТВА ПРИЗНАКОВ
- ВЫЯВИТЬ «СКРЫТУЮ СТРУКТУРУ» ДАННЫХ

МОДЕЛЬ СВЯЗИ ПРИЗНАКОВ
МЕЖДУ СОБОЙ И СО «СКРЫТЫМИ ФАКТОРАМИ»



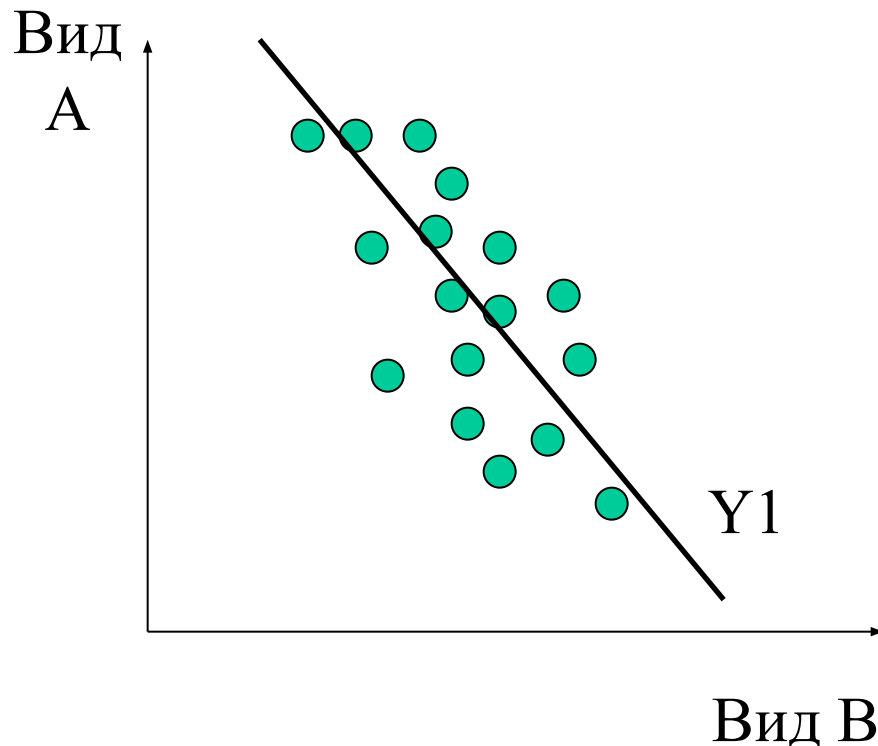
СПОСОБ ПОСТРОЕНИЯ
ОСЕЙ ОРДИНАЦИИ



ОПРЕДЕЛЕНИЕ
ПОЛОЖЕНИЯ ОБЪЕКТОВ
В НОВЫХ ОСЯХ

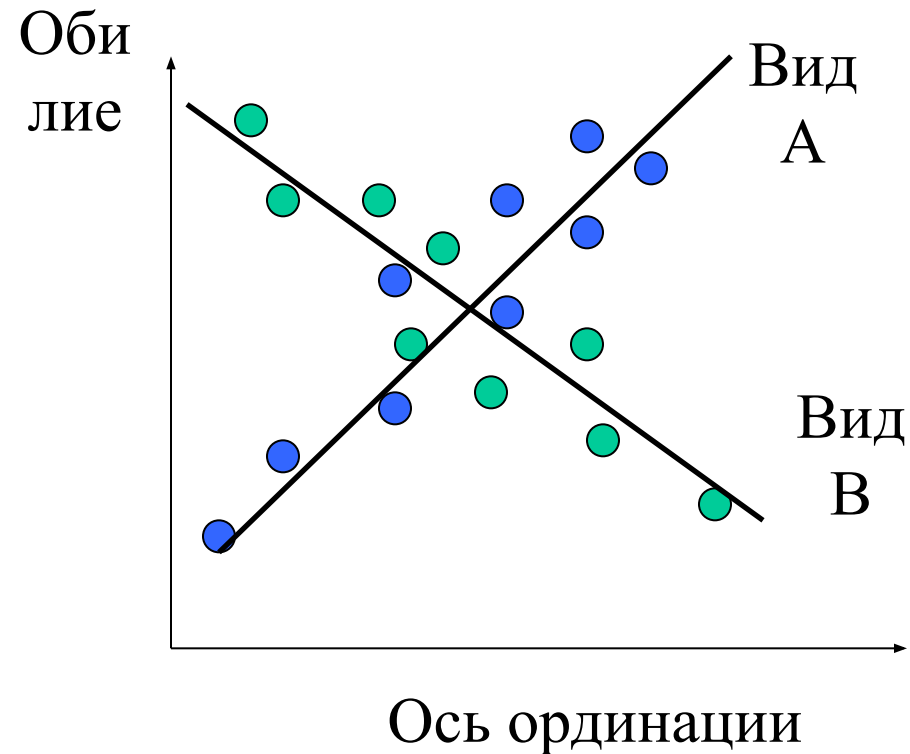
МЕТОД ГЛАВНЫХ КОМПОНЕНТ (PRINCIPAL COMPONENT ANALYSIS, PCA)

ПРЕДПОЛАГАЕТСЯ, ЧТО
ПРИЗНАКИ СВЯЗАНЫ
МЕЖДУ СОБОЙ ЛИНЕЙНО



ТОГДА ОСИ - ЛИНЕЙНЫЕ
КОМБИНАЦИИ ПРИЗНАКОВ

$$Y_1 = a_1 X_A + b_1 X_B$$

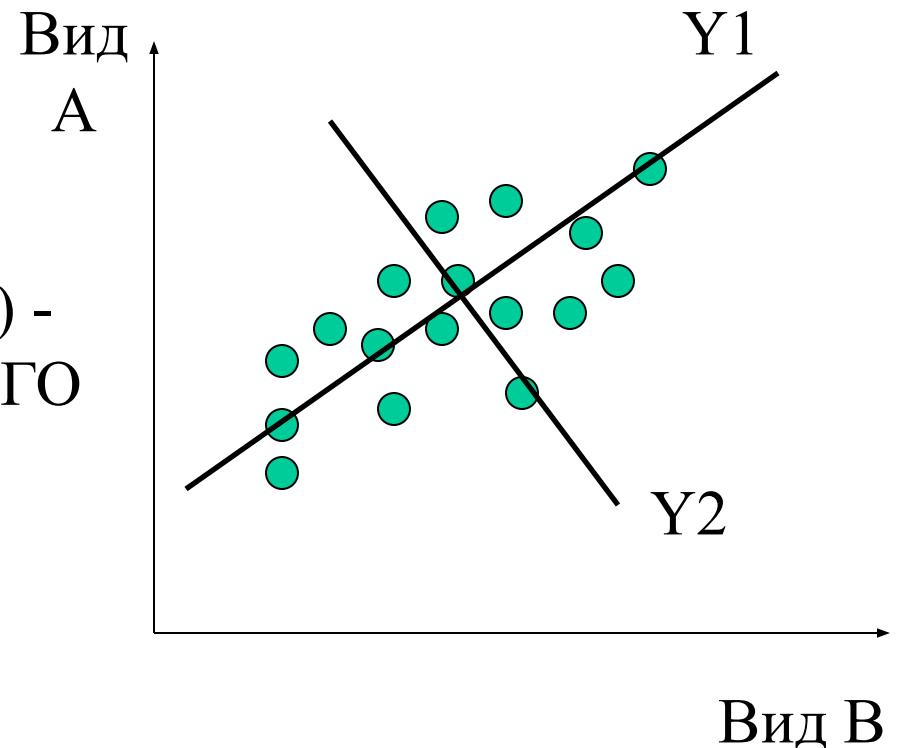


МЕТОД ГЛАВНЫХ КОМПОНЕНТ (PRINCIPAL COMPONENT ANALYSIS)

ОСИ - ЛИНЕЙНЫЕ
КОМБИНАЦИИ ПРИЗНАКОВ

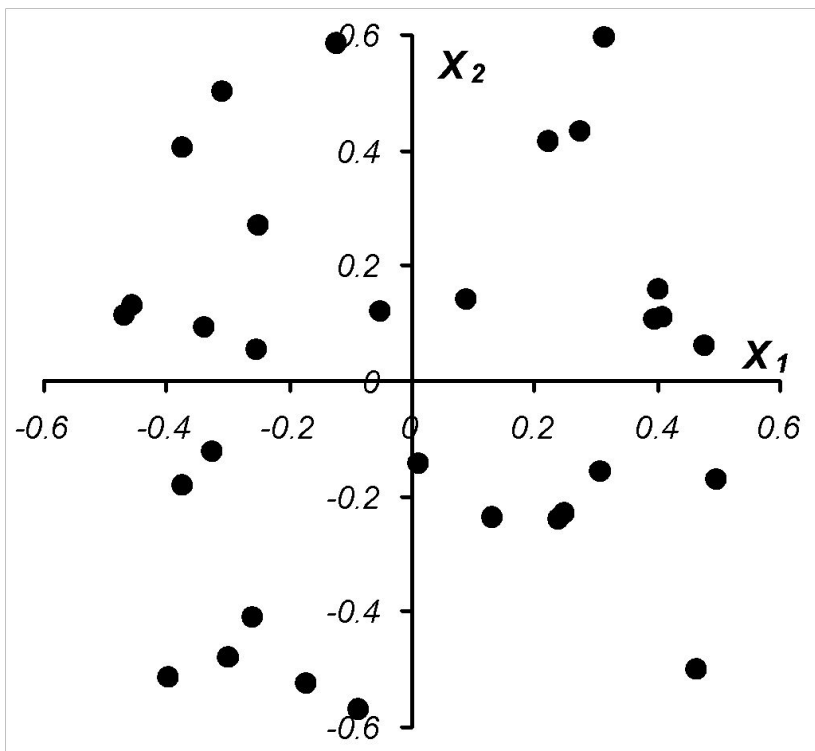
ПЕРВАЯ ОСЬ (КОМПОНЕНТА) -
НАПРАВЛЕНИЕ НАИБОЛЬШЕГО
РАЗБРОСА ТОЧЕК

$$Y1 = a_1 X_A + b_1 X_B$$

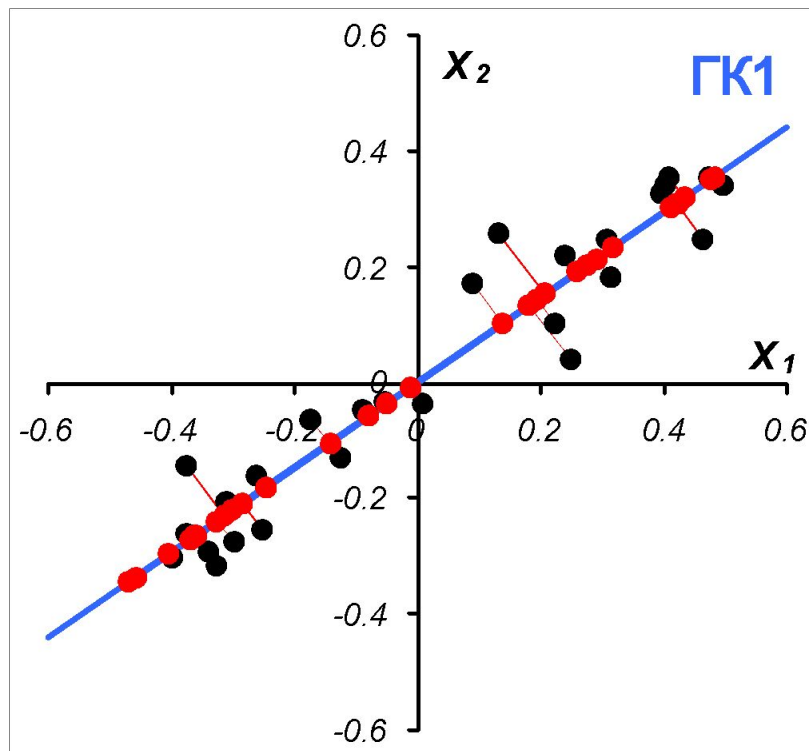


МЕТОД ГЛАВНЫХ КОМПОНЕНТ (PRINCIPAL COMPONENT ANALYSIS)

Данные без структуры

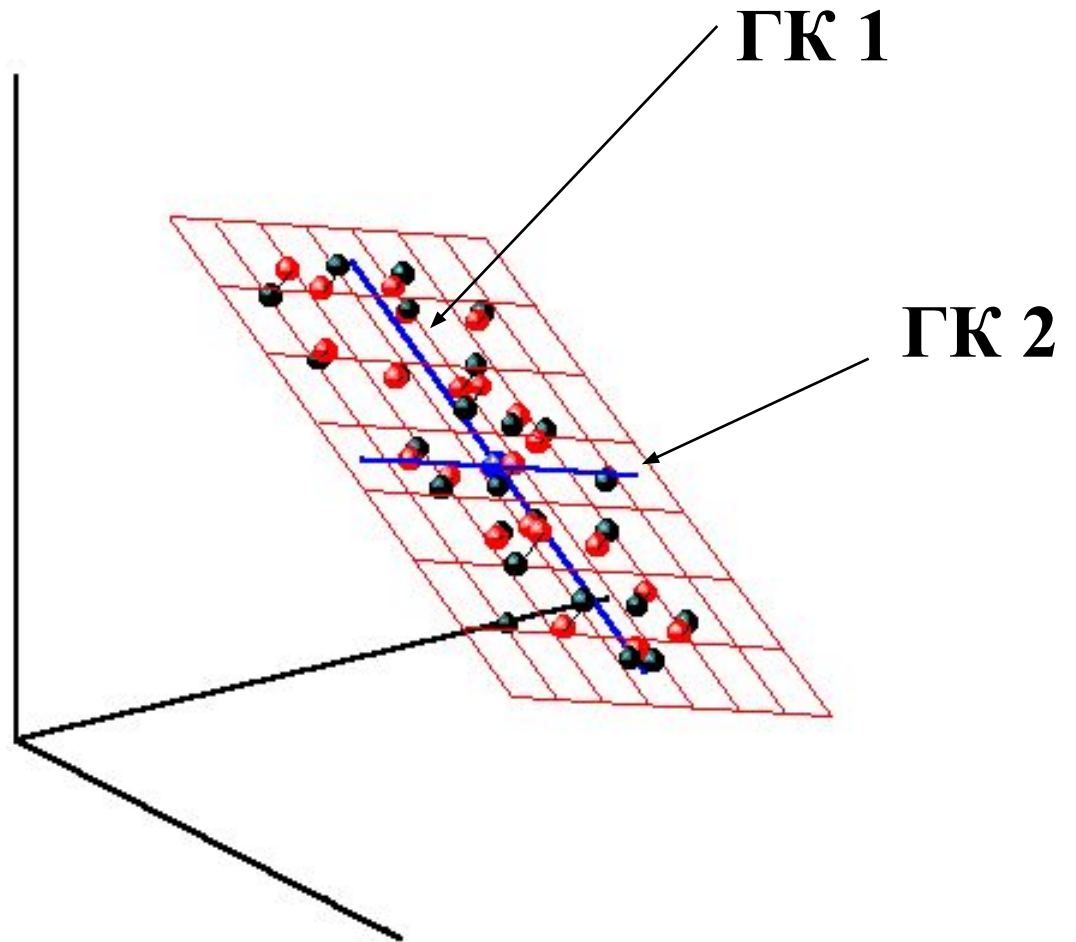


Данные со скрытой структурой



МЕТОД ГЛАВНЫХ КОМПОНЕНТ (PRINCIPAL COMPONENT ANALYSIS)

ВТОРАЯ ОСЬ -
НАПРАВЛЕНИЕ
НАИБОЛЬШЕГО
РАЗБРОСА ТОЧЕК,
ПЕРПЕНДИКУЛЯР-
НОЕ ПЕРВОЙ



ПРИМЕР: ОРДИНАЦИЯ СТАНЦИЙ ПО ФАКТОРАМ СРЕДЫ

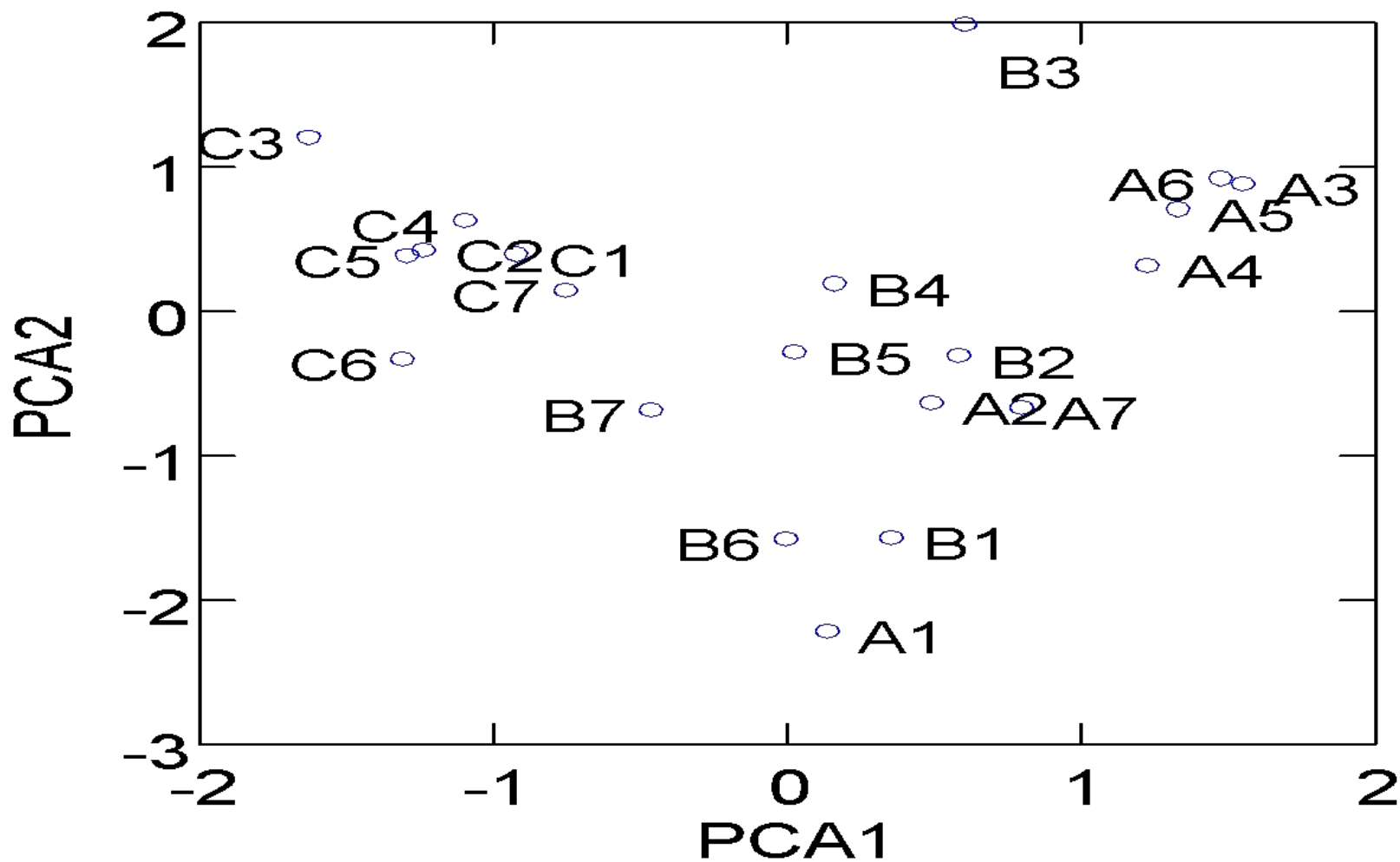
НАГРУЗКА

(ВКЛАД В КОМПОНЕНТУ)

ПЕРЕМЕННАЯ 1-АЯ (44%) 2-АЯ (19%)

РАЗМЕР ЧАСТИЦ	0.936	0.245
РАЗБРОС РАЗМЕРОВ	0.967	0.073
ВРЕМЯ ОСУШЕНИЯ	0.911	-0.212
ВЛАЖНОСТЬ ГРУНТА	-0.671	0.066
МОЩН. СОВР. ОСАДКА	-0.609	0.364
PH	-0.561	0.058
СТЕПЕНЬ СОРТИРОВКИ	0.555	-0.747
% ИЛОВОЙ ФРАКЦИИ	-0.522	-0.673
АЭРИРОВАННЫЙ СЛОЙ	0.137	0.723
EH	0.216	0.564

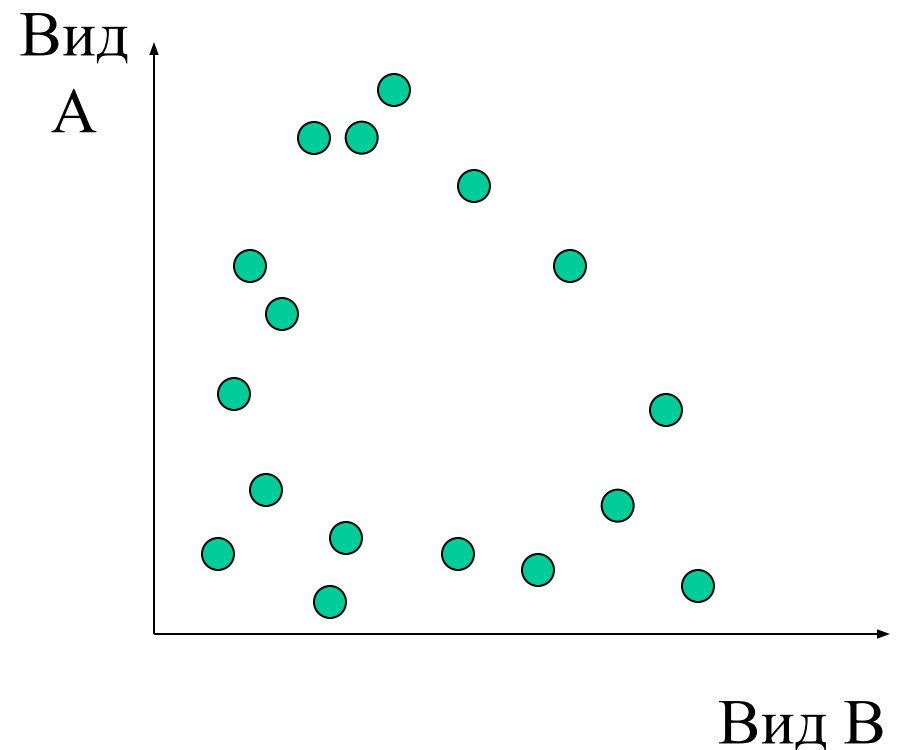
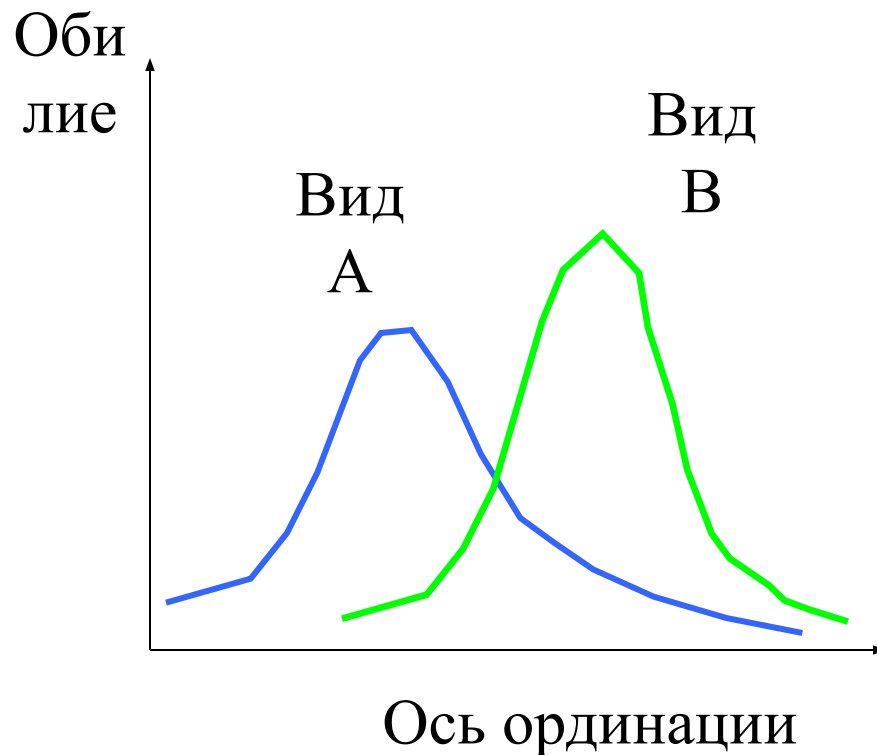
ПРИМЕР: ОРДИНАЦИЯ СТАНЦИЙ ПО ФАКТОРАМ СРЕДЫ



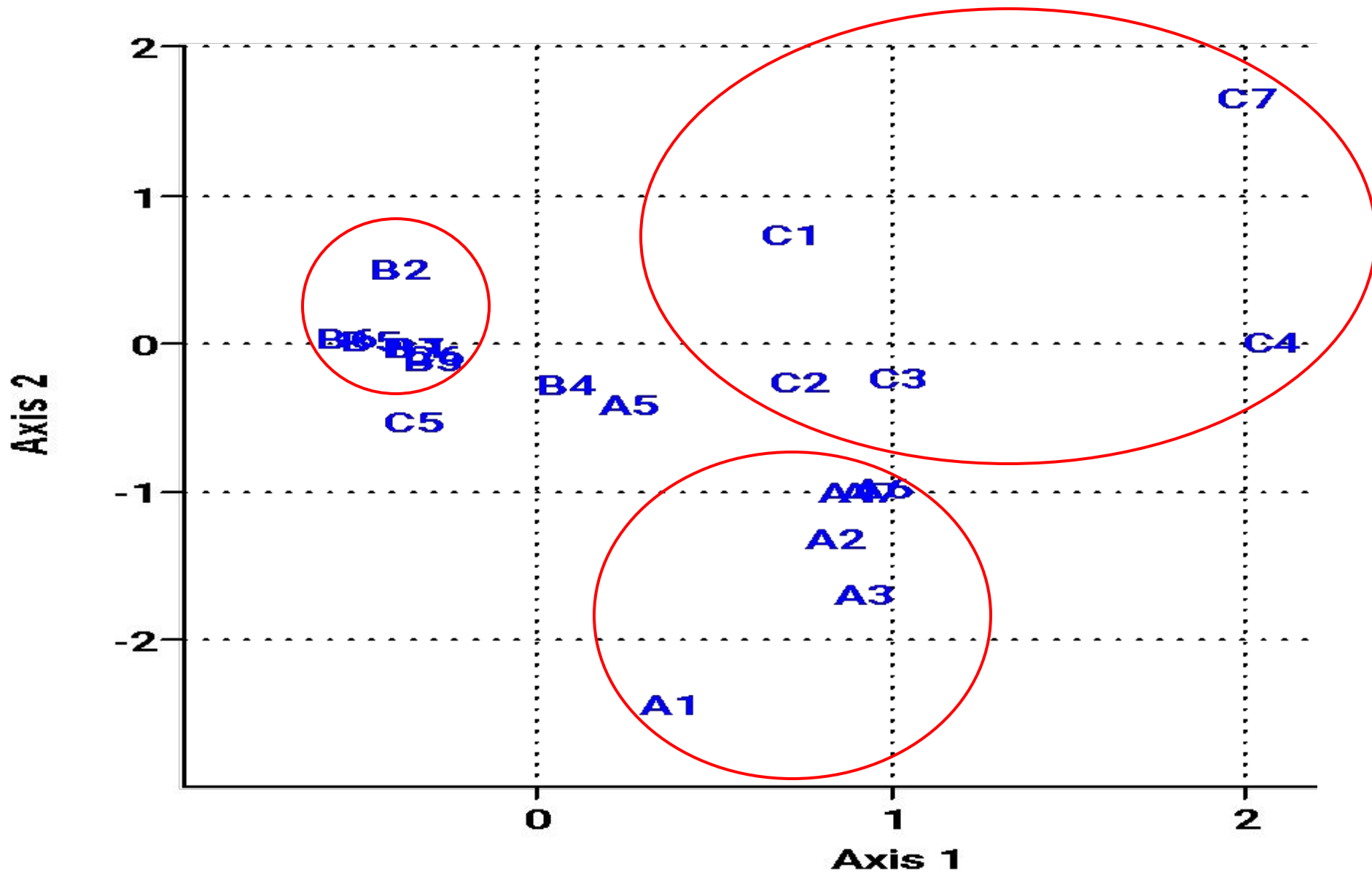
АНАЛИЗ СООТВЕТСТВИЙ (CORRESPONDENCE ANALYSIS, CA-DCA)

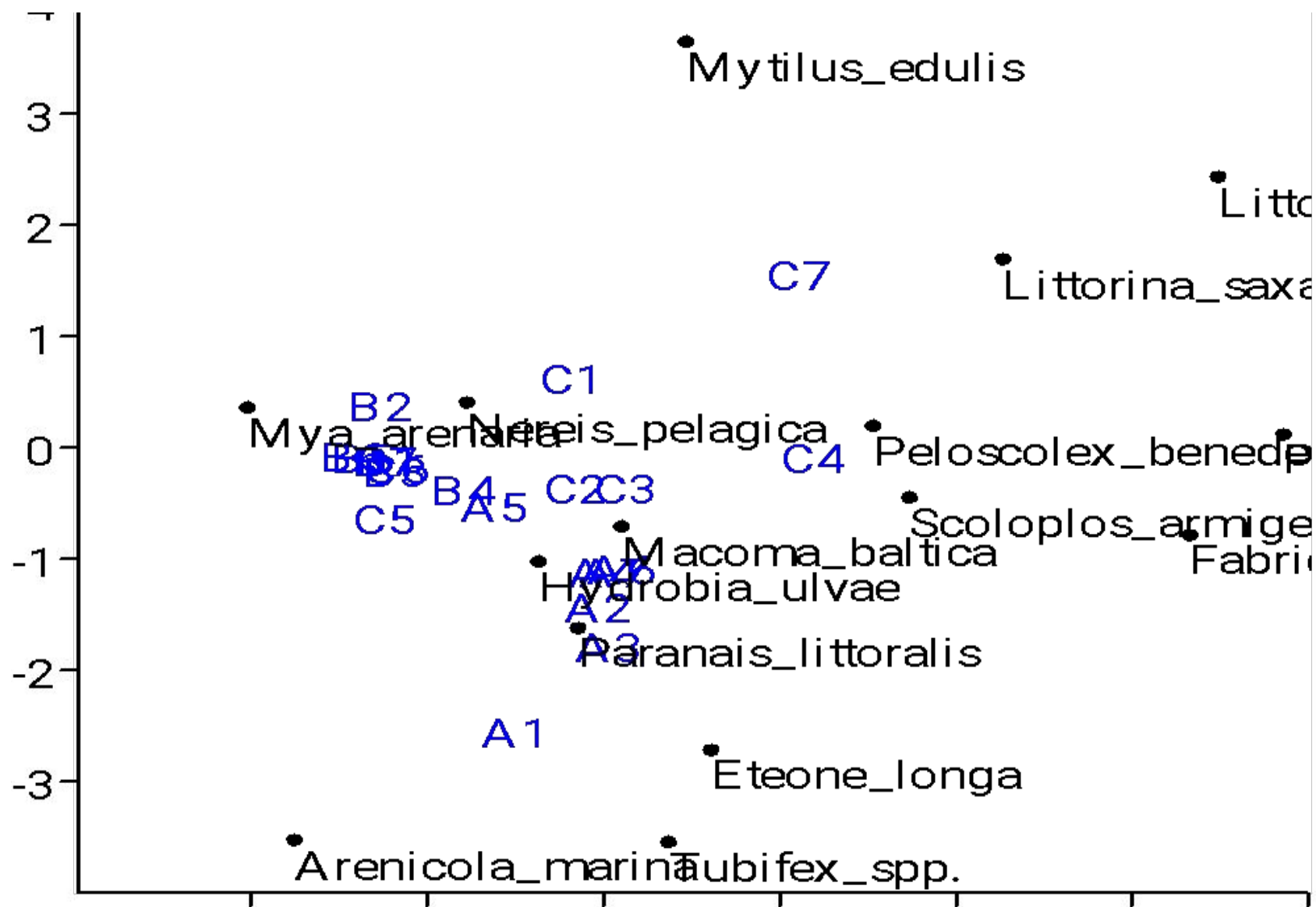
ВИДЫ (ПРИЗНАКИ)
РАСПРЕДЕЛЕННЫ
ВДОЛЬ ОСЕЙ
УНИМОДАЛЬНО

ЭТО ПРЕДПОЛАГАЕТ, ЧТО
САМИ ПРИЗНАКИ СВЯЗАНЫ
МЕЖДУ СОБОЙ НЕЛИНЕЙНО



ПРИМЕР АНАЛИЗА СООТВЕТСТВИЙ: ОРДИНАЦИЯ СТАНЦИЙ ПО ОБИЛИЮ ВИДОВ





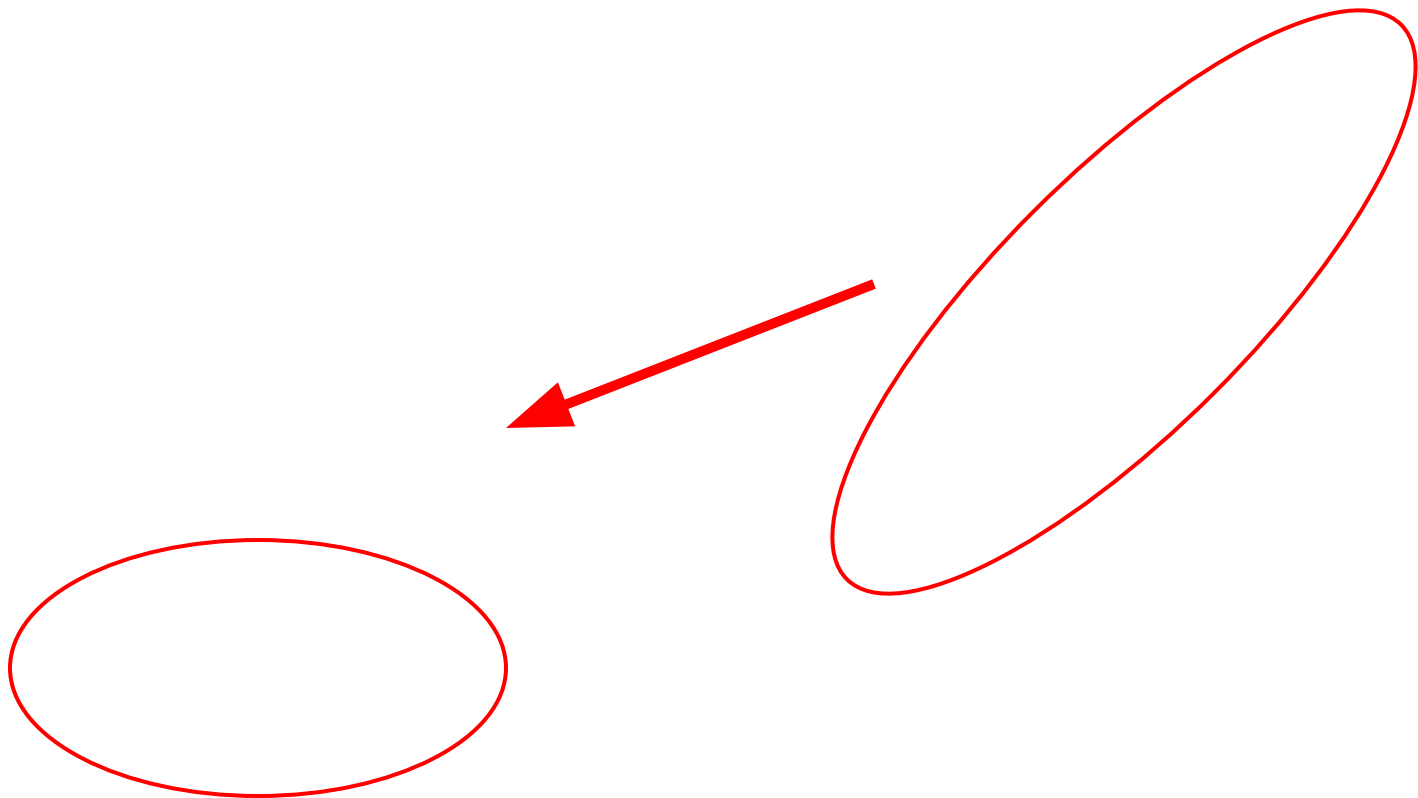
МНОГОМЕРНОЕ ШКАЛИРОВАНИЕ (MULTIDIMENSIONAL SCALING)

ЗАДАЧИ:

- ВИЗУАЛИЗАЦИЯ ДАННЫХ О СХОДСТВЕ
- УМЕНЬШЕНИЕ РАЗМЕРНОСТИ

РАСПОЛАГАЕТ ОБЪЕКТЫ ТАК, ЧТОБЫ
РАССТОЯНИЯ МЕЖДУ НИМИ
СООТВЕТСТВОВАЛИ ВЕЛИЧИНАМ
НЕСХОДСТВА

ПРИМЕР МНОГОМЕРНОГО ШКАЛИРОВАНИЯ: ОРДИНАЦИЯ СТАНЦИЙ ПО ОБИЛИЮ ВИДОВ

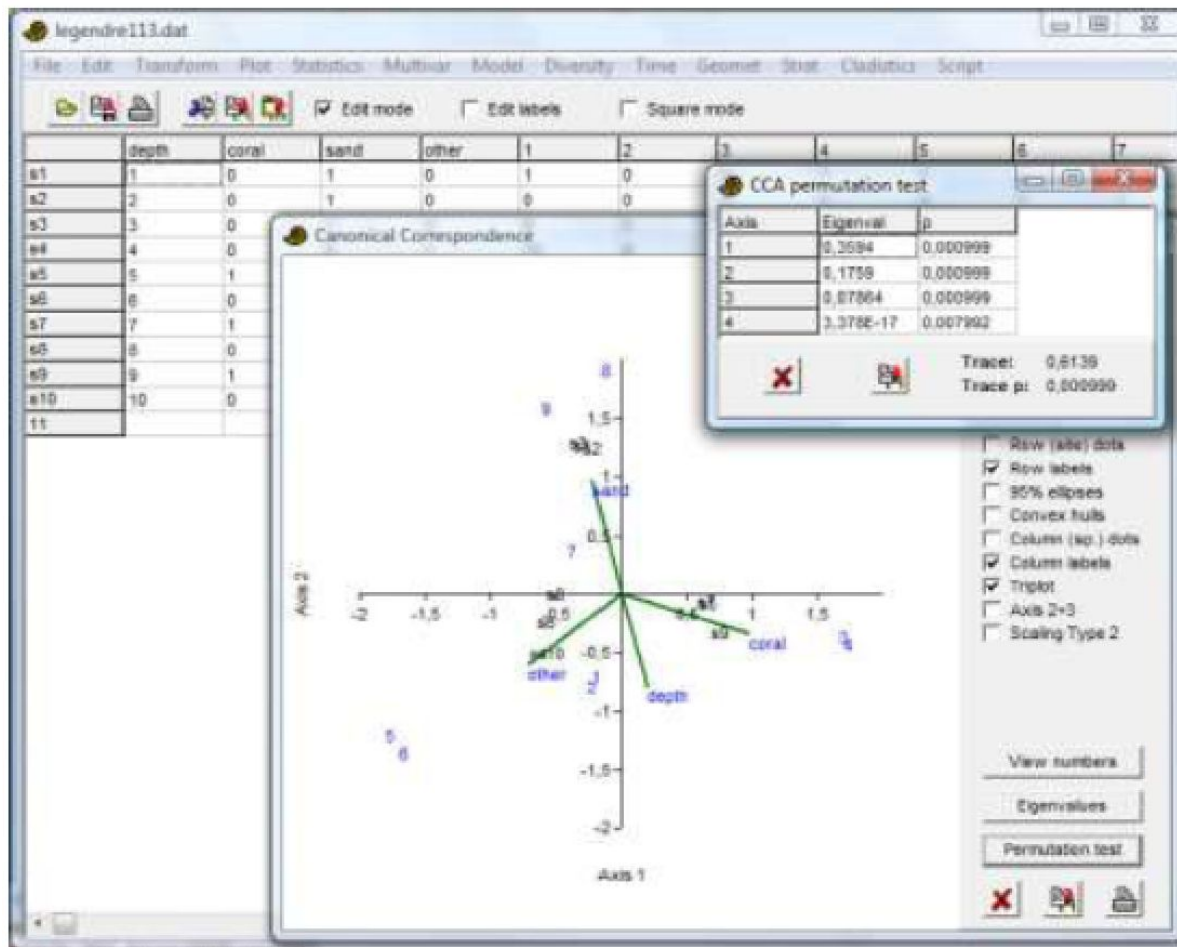


ВЫЯВЛЕНИЕ СВЯЗИ МЕЖДУ СТРУКТУРОЙ СООБЩЕСТВ И ФАКТОРАМИ СРЕДЫ

- Корреляционный анализ (недостатки: нелинейность, выбросы, мультиколлинеарность, множ-ть сравнений)
- Прямой градиентный анализ
(канонический анализ соответствий, CCA)
- Регрессия на матрицах сходства (RSM)
- Группы проб выделяются (ординацией или классификацией) – ANOSIM (проверка гипотез)
- Ординация дает четкие результаты (высок % объясненной дисперсии PCA или DCA), факторов немного – корреляция координат проб на осях с факторами среды

КАНОНИЧЕСКИЙ АНАЛИЗ СООТВЕТСТВИЙ (CANONICAL CORRESPONDENCE ANALYSIS, CCA)

АНАЛОГ ССА, НО ОСИ ОРДИНАЦИИ КОНСТРУИРУЮТСЯ
КАК ЛИНЕЙНЫЕ КОМБИНАЦИИ ФАКТОРОВ СРЕДЫ.

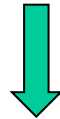


ЭТИ ОСИ
ПОДБИРАЮТСЯ
ТАК, ЧТОБЫ
ОПИСЫВАТЬ
РАСПРЕДЕЛЕНИЕ
ВИДОВ ВДОЛЬ НИХ
(ОТКЛИК ВИДОВ –
УНИМОДАЛЬНЫЙ)

МАТРИЦЫ СХОДСТВА – ПРОВЕРКА ГИПОТЕЗ

ПРОБЛЕМА:

ВЕЛИЧИНЫ СХОДСТВА В МАТРИЦЕ **ВЗАИМОСВЯЗАНЫ**
(ЕСЛИ **A** похоже на **B** и **B** похоже на **C**,
то **A** не может сильно отличаться от **C**)



К ИНДЕКСАМ СХОДСТВА НЕПРИМЕНИМЫ
ОБЫЧНЫЕ МЕТОДЫ ОЦЕНКИ ДОСТОВЕРНОСТИ



РЕШЕНИЕ:

ДЛЯ ОЦЕНКИ НЕСЛУЧАЙНОСТИ КАКОЙ-ЛИБО ВЕЛИЧИНЫ,
СРАВНИВАЕМ ЕЕ РЕАЛЬНОЕ ЗНАЧЕНИЕ С ПОЛУЧЕННЫМИ В
РЕЗУЛЬТАТЕ СЛУЧАЙНЫХ ПЕРЕСТАНОВОК (PERMUTATION
TEST)

ГИПОТЕЗА О СООТВЕТСТВИИ МЕЖДУ ДВУМЯ МАТРИЦАМИ СХОДСТВА (ПЕРЕСТАНОВОЧНЫЙ ТЕСТ МАНТЕЛЯ, MANTEL' PERMUTATION TEST)

2 МАТРИЦЫ СХОДСТВА (ПО РАЗНЫМ ПРИЗНАКАМ)
ДЛЯ ОДНОГО И ТОГО ЖЕ НАБОРА ОБЪЕКТОВ



МЕРА СООТВЕТСТВИЯ - КОРРЕЛЯЦИЯ
МЕЖДУ ПАРАМИ (S_{ij}, P_{ij})

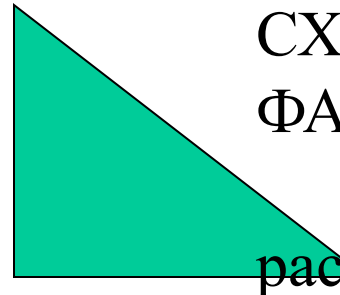
ПРИМЕР:

СООТВЕТСТВУЕТ ЛИ СХОДСТВО
СТАНЦИЙ ПО БЕНТОСУ СХОДСТВУ ПО АБИОТЕ?

1-я матрица:
СХОДСТВО ПО
БЕНТОСУ (индекс
Чекановского)



2-я матрица:
СХОДСТВО ПО
ФАКТОРАМ
(Эвклидово
расстояние)



Rank correlation method: Spearman

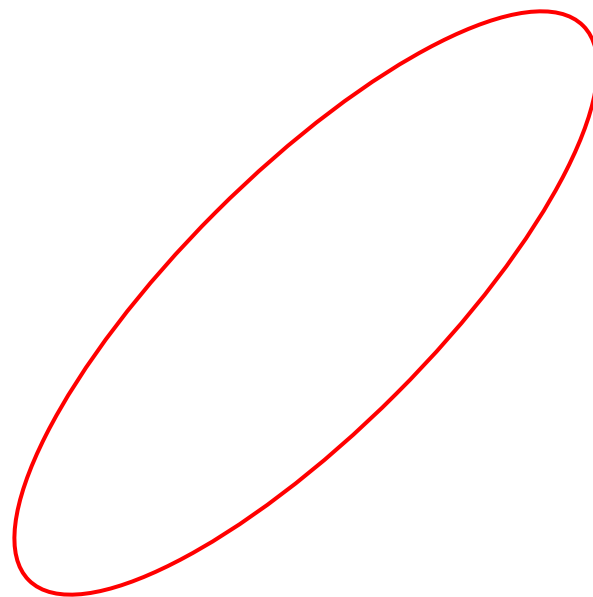
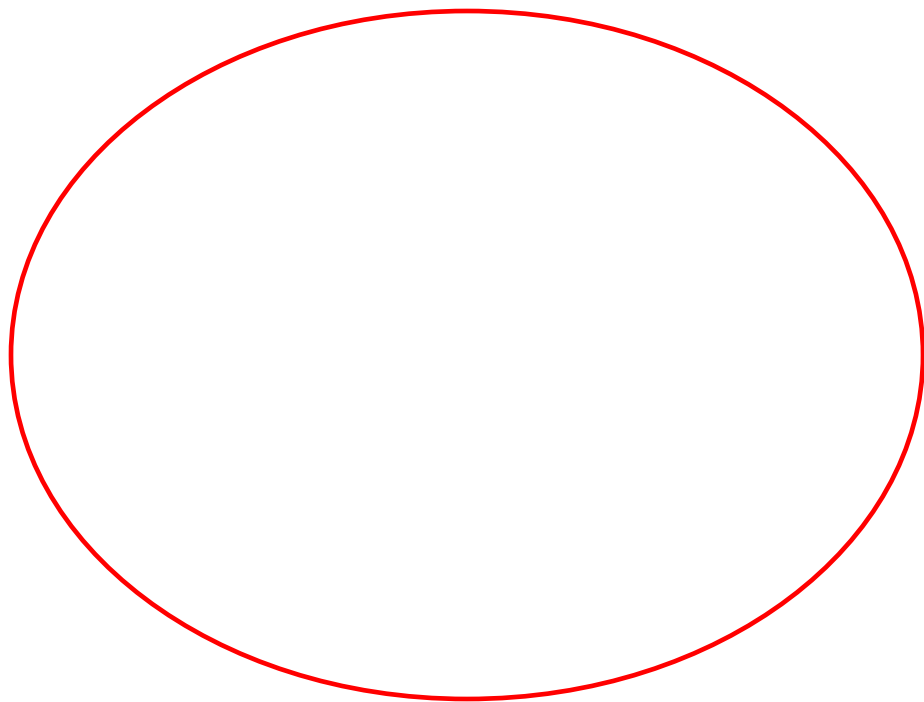
Sample statistic (Rho): **0.257**

Significance level of sample statistic: **0.7 %**

Number of permutations: 999

Number of permuted statistics greater than or equal to Rho: **6**

ОРДИНАЦИЯ СТАНЦИЙ ПО ОБИЛИЮ ВИДОВ (МНОГОМЕРНОЕ ШКАЛИРОВАНИЕ)



НЕСЛУЧАЙНОСТЬ ГРУППИРОВКИ ОБЪЕКТОВ (Analysis Of Similarities, ANOSIM)

ДЕЙСТВИТЕЛЬНО ЛИ СТАНЦИИ ОДНОГО
ГОРИЗОНТА БОЛЕЕ ПОХОЖИ, ЧЕМ СТАНЦИИ
РАЗНЫХ ГОРИЗОНТОВ?

- 1) РАНЖИРУЕМ ВСЕ ВЕЛИЧИНЫ СХОДСТВА
(ПО УБЫВАНИЮ)
- 2) СЧИТАЕМ СРЕДНИЕ РАНГИ:
 - ДЛЯ ВСЕХ ПАР **ВНУТРИ ГРУПП** ($S_{вн}$)
 - ДЛЯ ВСЕХ ПАР **ИЗ РАЗНЫХ ГРУПП** ($S_{меж}$)
- 3) СЧИТАЕМ R-СТАТИСТИКУ:
$$R = \frac{S_{меж} - S_{вн}}{n(n-1)/4}$$

- R меняется от -1 до $+1$
- $R = +1$, если **ВСЕ** пробы из одной группы более схожи, чем **ЛЮБАЯ** пара проб из разных групп
- $R = 0$, если **нет различий** между сходством проб внутри групп и между группами

ОЦЕНКА ДОСТОВЕРНОСТИ:

- СЧИТАЕМ $R_{случ}$ ДЛЯ СЛУЧАЙНЫХ ПЕРЕТАСОВОК ПРОБ ПО ГРУППАМ
- СРАВНИВАЕМ РЕАЛЬНОЕ R СО МНОЖЕСТВОМ $R_{случ}$

Global Test

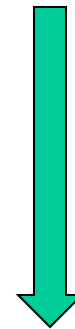
Sample statistic (Global R): 0.558

Уровень значимости R: 0.1%

Число случайных вариантов: 999

Число случайных вариантов, давших значение R,
больше или равное наблюдаемому: 0

РЕАЛЬНОЕ R



Global
R

Pairwise Tests

Groups	R Statistic	Significance Level %	Actual Permutations	Number \geq Observed
A, B	0.828	0.4	999	3
A, C	0.514	0.3	999	2
B, C	0.341	0.5	999	4

ВЫДЕЛЕНИЕ ХАРАКТЕРНЫХ И ДИФФЕРЕНЦИРУЮЩИХ ВИДОВ ДЛЯ ГРУПП СТАНЦИЙ

- **ХАРАКТЕРНЫЕ ВИДЫ:** ОПРЕДЕЛЯЮТ СХОДСТВО
ПРОБ ВНУТРИ ГРУППЫ
- **ДИФФЕРЕНЦИРУЮЩИЕ ВИДЫ:** ОПРЕДЕЛЯЮТ
РАЗЛИЧИЯ МЕЖДУ СТАНЦИЯМИ

Процедура SIMPER
(Similarity percentages - species contributions)

ХАРАКТЕРНЫЕ ВИДЫ

Group A Average similarity: **59.57**

Species	Av.Abund	Contrib%	Cum.%
Hydrobia ulvae	573.58	63.78	63.78
Macoma baltica	403.01	21.17	84.95
Tubifex spp.	106.70	8.94	93.90

Group B Average similarity: **80.52**

Mya arenaria	3130.74	72.68	72.68
Hydrobia ulvae	663.32	13.88	86.56
Macoma baltica	303.65	6.51	93.06

Group C Average similarity: **50.70**

Mya arenaria	1336.32	28.50	28.50
Macoma baltica	645.24	28.19	56.69
Hydrobia ulvae	527.11	27.47	84.16
Littorina spp.	214.29	5.26	89.41
Mytilus edulis	191.32	2.82	92.24

ДИФФЕРЕНЦИРУЮЩИЕ ВИДЫ

Groups A & B Average dissimilarity = **68.22**

	Group A	Group B	
Species	Av.Abund	Av.Abund	Contrib%
Mya arenaria	219.76	3130.74	43.66
Hydrobia ulvae	573.58	663.32	23.70
Tubifex spp.	106.70	17.67	6.79
Arenicola marina	43.08	4.17	4.37

Groups B & C Average dissimilarity = **46.86**

Mya arenaria	3130.74	1336.32	39.77
Macoma baltica	303.65	645.24	15.10
Mytilus edulis	169.59	191.32	10.47
Littorina spp.	20.36	214.29	9.95
Nereis pelagica	279.11	157.68	7.18
Peloscolex benedeni	10.48	80.99	3.56