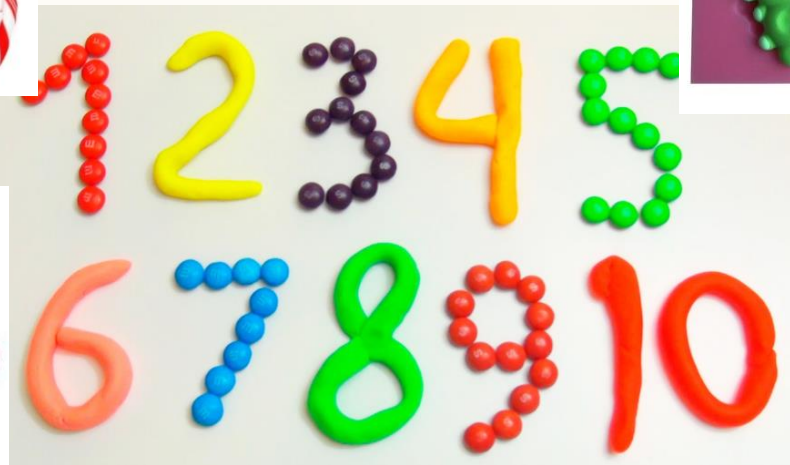
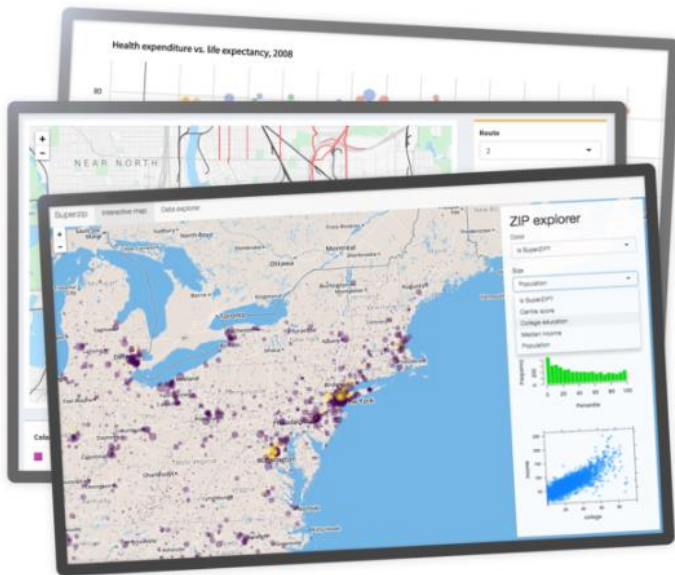


# Aula 27 Goodies\*



\* Goodies related to animals, plants and numbers...



# Shiny

## Easy web applications in R

Shiny is an open source R package that provides an elegant and powerful web framework for building web applications using R. Shiny helps you turn your analyses into interactive web applications without requiring HTML, CSS, or JavaScript knowledge.

[CLICK HERE TO VIEW THE SHINY SHOWCASE](#)

# Correlation Game

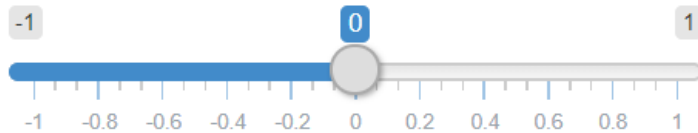
## Difficulty

Easy

## Show:

- Averages
- Standard deviation line
- Ellipse

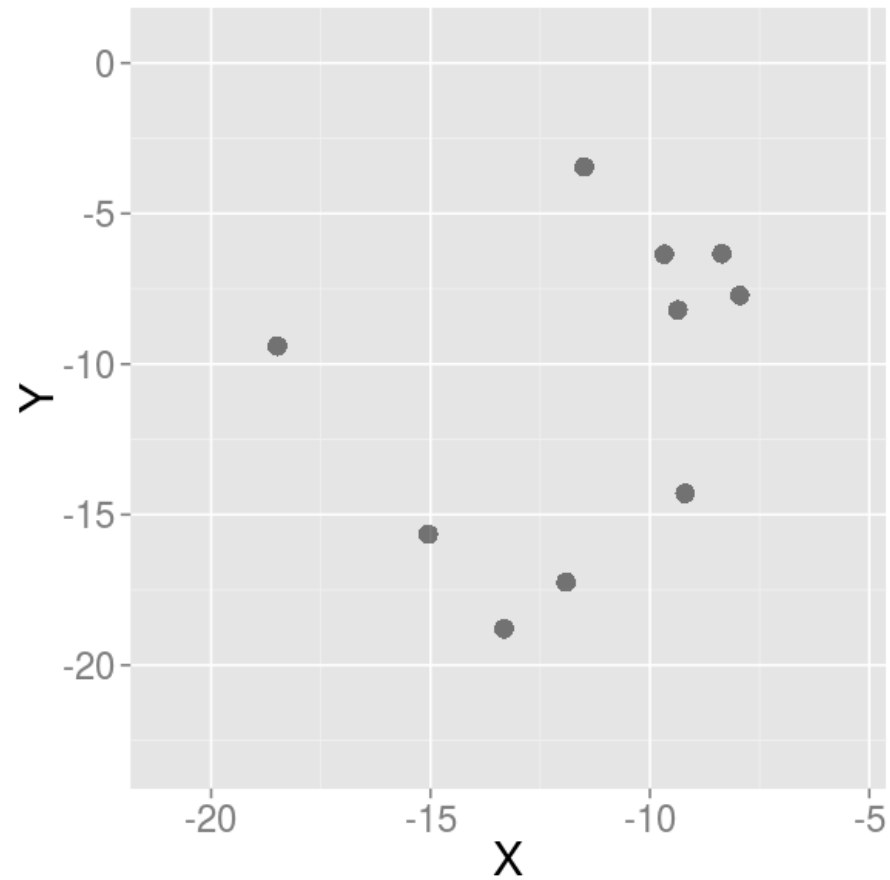
The correlation between X and Y is...



Mark your answer and click 'Submit!'

Submit

New Plot

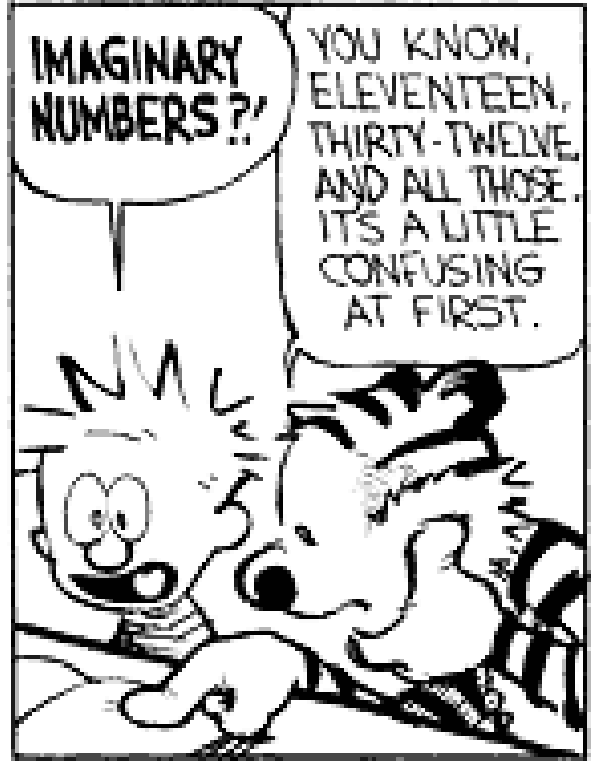
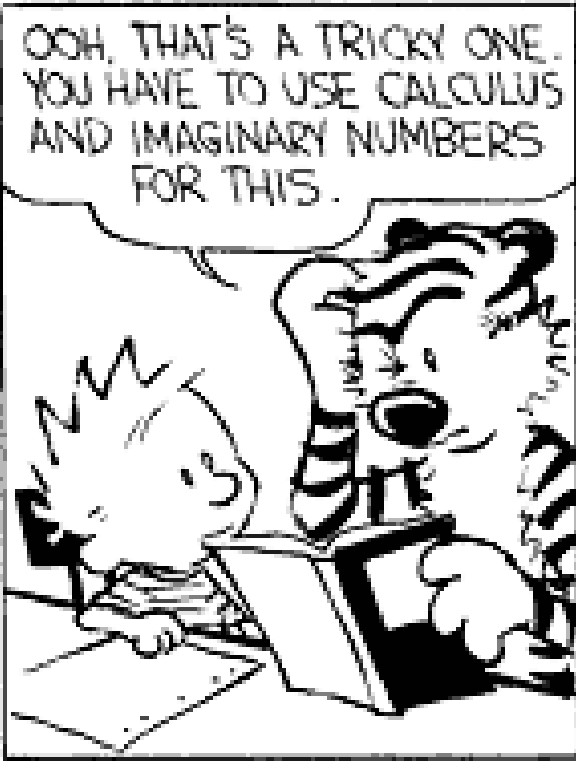


# Hex memory game

Find matching hex!

Time elapsed: 48







## 8.º JUÍZO CÍVEL DA COMARCA DE LISBOA

Proc. n.º 1313/2002

Conclusão em: 20/11/2008

A fls. 189 vem o Executado apresentar um requerimento intitulado de «oposição à penhora», em que alega ter sido ordenada a penhora de **1/6 do vencimento** que auferir, sendo que atentas as despesas correntes que apresenta ter, o deixa numa situação de grave carência económica.

Conclui, pedindo a isenção de penhora.

Notificado o exequente, vem este requerer a manutenção da penhora.

Foi ordenado a elaboração de um relatório sócio económico do executado, o qual se mostra junto a fls. 213.

Apreciando.

O Tribunal deve ponderar entre o interesse do exequente em ver o seu direito assegurado, e o do executado em cumprir o pagamento da quantia a que se encontra vinculado, interesse esse que tem de ser proporcional.

Pese embora os factos relatados pelo executado, e sendo certo que não competindo ao Tribunal restringi-lo de refazer a sua vida como entender, também não pode o Tribunal prejudicar os compromissos anteriormente por aquele assumidos.

Assim, determina o Tribunal proceder à redução da penhora do vencimento do executado para 1/5 do vencimento.

Notifique.

Lisboa, 2008-12-12.



Ecología Numérica - Aula Teórica 27 – 17-12-2018



Although we often hear that data  
speak for themselves, their voices  
can be soft and sly

— *Frederick Mosteller* —

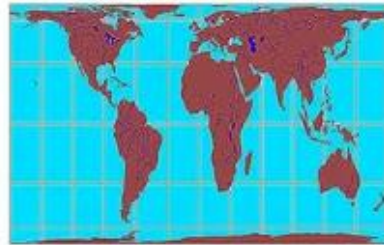
AZ QUOTES

<https://www.azquotes.com/quote/534227>





Mercator Projection



Gall-Peters Projection



Miller Cylindrical Projection



Mollweide Projection



Goode's Homolosine Equal-area Projection



Sinusoidal Equal-Area Projection

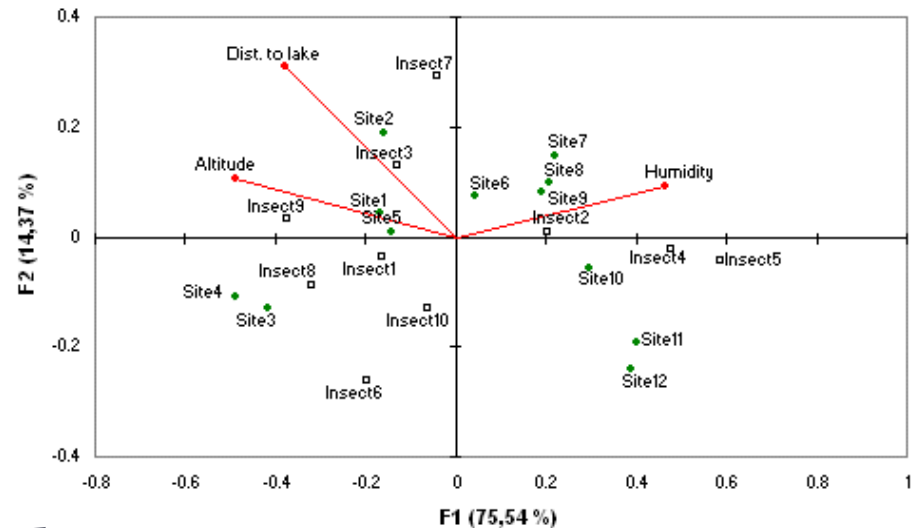


Robinson Projection

Tal como num mapa que representa 3 dimensões em apenas 2, numa análise multivariada há uma distorção da informação quando visualizamos a informação num biplot

Essa distorção será tanto menor quanto maior for a proporção da variância explicada pelos eixos.


CCA Map / Symmetric  
(axes F1 and F2: 89,90 %)



• Sites □ Objects • Variables

<http://geoawesomeness.com/map-distortions/>

[https://help.xlstat.com/customer/en/portal/articles/2062255-canonical-correspondence-analysis-cca-tutorial?b\\_id=9283](https://help.xlstat.com/customer/en/portal/articles/2062255-canonical-correspondence-analysis-cca-tutorial?b_id=9283)



# Canonical Correspondence Analysis (CCA)...

...continued

# Testar a significância dos eixos da CCA

H0: a componente K é apenas fruto do acaso (ou seja, não é relevante fazer a sua interpretação)

(testes de permutações – permutar as linhas das abundancias)

```
> anova(cca.aves,by="axis",step=1000)
Permutation test for cca under reduced model
Forward tests for axes
Permutation: free
Number of permutations: 999

Model: cca(formula = aves[, -1] ~ Perc.veg + Agua + Solo, data = avesamb[, -1])
```

	Df	ChiSquare	F	Pr(>F)	
CCA1	1	0.32804	14.5910	0.002	**
CCA2	1	0.02924	1.3006	0.524	
CCA3	1	0.00142	0.0634	0.999	
Residual	18	0.40468			

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# Forward CCA

Quais as variáveis relevantes: Semelhante a uma análise *stepwise* de *model selection* num modelo de regressão

```
> cca.aves.forward=ordistep(cca(aves[, -1]~1, data=avesamb[, -1]), scope=formula(cca.aves), direction="forward", pstep=1000)
```

```
Start: aves[, -1] ~ 1
```

	Df	AIC	F	Pr(>F)	
+ Perc.veg	1	-9.4512	6.9239	0.005	**
+ Agua	1	-6.4469	3.4872	0.020	*
+ Solo	1	-4.5385	1.5356	0.165	

```
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Step: aves[, -1] ~ Perc.veg
```


	Df	AIC	F	Pr(>F)	
+ Agua	1	-14.667	7.3756	0.005	**
+ Solo	1	-14.400	7.0576	0.005	**

```
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

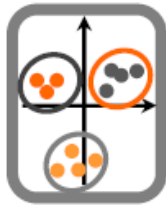
```
Step: aves[, -1] ~ Perc.veg + Agua
```

	Df	AIC	F	Pr(>F)	
+ Solo	1	-12.874	0.1697	0.99	

→ O “Solo”, depois de juntar a “Agua”, já não parece ser util



**ESCALAMENTO  
MULTIDIMENSIONAL  
E  
ANÁLISE DISCRIMINANTE**



# Ecologia Numérica

Componente Teórica - Prática

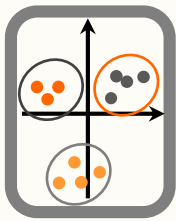
Ficha de trabalho **11**

1. Recolheram-se dados de abundância de várias espécies de anfíbios em vários distritos de Portugal (*DataTP11anfíbios.csv*). Efectue um escalamento multidimensional com base em distâncias euclidianas e retire as principais conclusões.
2. Efectue um escalamento multidimensional não-métrico aos dados da alínea anterior e comente os resultados obtidos.
3. Recolheu dados relativos a densidades de várias espécies de peixes em duas áreas distintas (1 e 2) (*DataTP11peixesgrupos.csv*). Realize uma análise discriminante e comente os resultados obtidos.
4. Efectue uma análise discriminante aos dados *DataTP11solea.csv*, referentes a variáveis morfométricas de 3 espécies de linguado: *Solea solea*, *Solea senegalensis* e *Solea lascaris*. Discuta os resultados obtidos.
5. Aplique técnicas de classificação (classificação hierárquica, com base na distância euclideana e no método da mínima variância) e de ordenação (análise de componentes principais) aos dados da alínea anterior (*DataTP11solea2.csv*). Compare os resultados e comente as principais diferenças obtidas.

Para trabalhar de forma independente, soluções serão disponibilizadas – poderá haver perguntas na parte teórica mas não vos será pedido nada no exame teórico-prático sobre escalamento multidimensional ou análise discriminante.

# ESCALAMENTO MULTIDIMENSIONAL





ordenação

---

## Escalamento multidimensional (MDS)

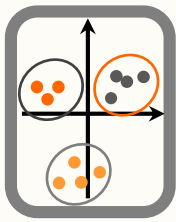
- Conjunto de técnicas que efetua uma ordenação das observações num espaço de dimensão reduzida (usualmente num plano), tentando conservar o mais possível as distâncias

originais;

→ Que podemos escolher !

- Num diagrama de MDS as relações de proximidade entre as várias observações aproximam-se o mais possível das suas distâncias originais.



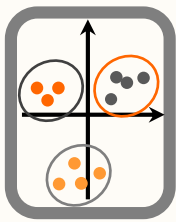


ordenação

---

## Escalamento multidimensional (MDS)

- Na generalidade das situações não existem formas de relacionar directamente os eixos com as variáveis originais, o que muitas vezes pode dificultar a interpretação dos padrões observados.

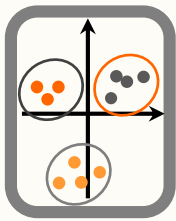


ordenação

---

## Escalamento multidimensional (MDS)

- Podem ser consideradas várias medidas de distância entre as observações;
- Existem duas abordagens distintas: métrica e não métrica (ordinal).

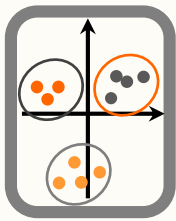


ordenação

---

## Escalamento multidimensional (MDS)

- O desempenho da análise pode ser avaliado:
  1. Subjectivamente, pela análise dos diagramas (avaliação dos padrões de dispersão dos pontos face à matriz inicial dos dados e medidas de distância),
  2. Através do valor de *stress* – indica a proporção da variância das disparidades não explicada pelo modelo de MDS (é como se fosse uma medida da diferença entre as distâncias originais e as distâncias no biplot)

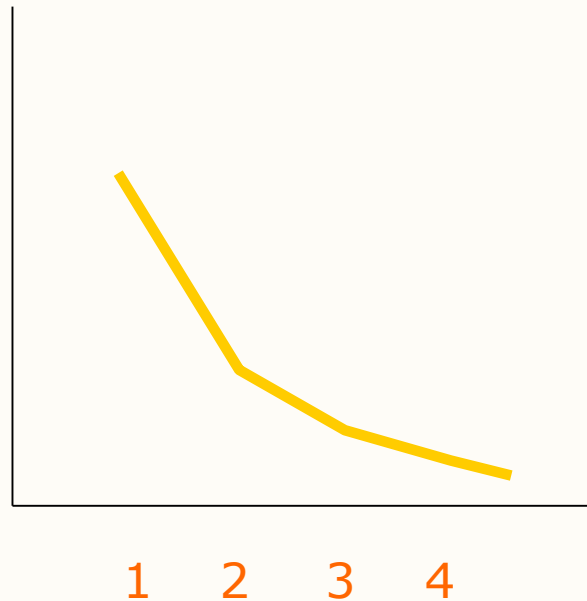


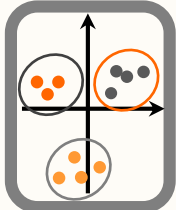
ordenação

# Escalamento multidimensional (MDS)

Stress (proporção da variância não explicada pelo modelo de MDS)

Stress

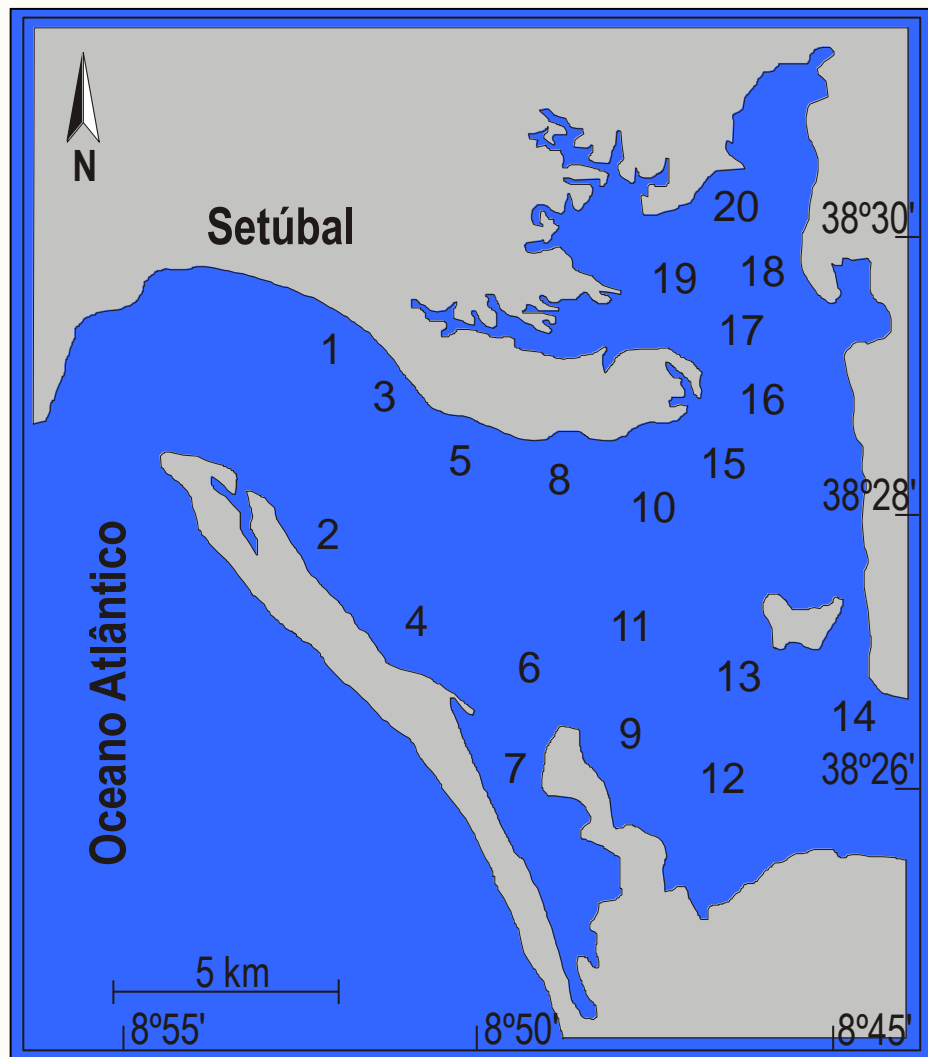


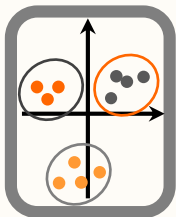


## ordenação

### Um exemplo:

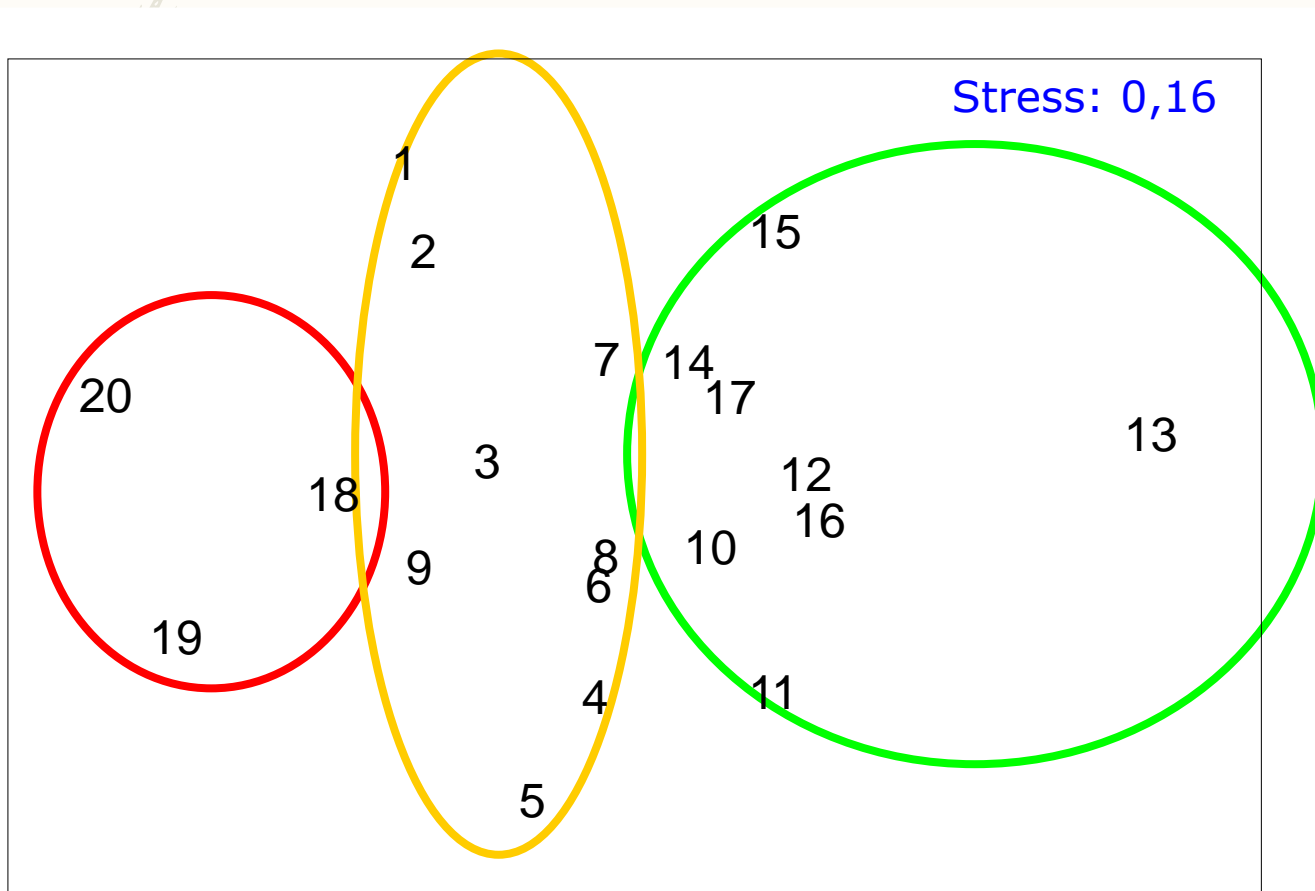
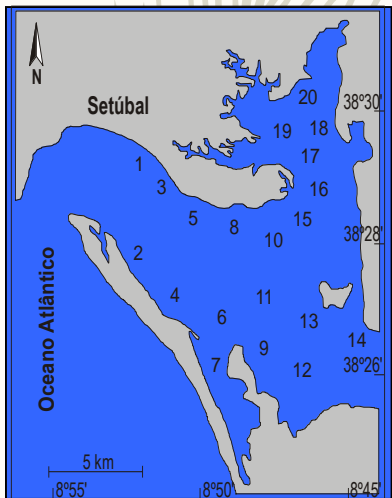
Dados de abundâncias de espécies de peixes em 20 estações de amostragem no estuário do Sado.





ordenação

## Escalamento multidimensional (MDS)



# Escalamento multidimensional (MDS)

```
anfi <- read.csv2("DataTP11anfios.csv")  
row.names(anfi)=anfi[,1]  
deucanfi<-dist(anfi[, -1])  
dmananfi2<-dist(anfi[, -1],method="manhattan")  
print(deucanfi,digits=2)
```

Esta linha não é relevante,  
estava a testar outra distância  
e ficou aqui esquecida!

```
##      EVOR BEJA FARO SETU LISB LEIR CBRA VISE GUAR AVEI COIM PORT BRAG BRCA  
## BEJA 20.4  
## FARO 12.9 10.8  
## SETU 42.6 36.1 32.8  
## LISB 36.5 42.1 33.8 27.8  
## LEIR 44.5 54.9 45.1 39.8 18.1  
## CBRA 40.5 47.4 38.5 31.5  7.5 12.3  
## VISE 41.8 50.2 40.7 34.0 12.0  7.7 6.0  
## GUAR 42.0 43.8 37.0 26.3  9.8 25.2 13.4 18.2  
## AVEI 51.5 48.7 44.1 21.2 22.6 35.8 25.8 29.1 16.0  
## COIM 67.0 59.9 57.6 28.1 40.9 52.1 43.3 46.0 34.1 19.0  
## PORT 64.6 58.9 55.9 26.3 37.5 47.9 39.6 42.0 31.3 15.8 5.3  
## BRAG 68.2 60.2 58.4 29.0 42.7 54.3 45.1 48.1 35.6 21.1 3.9 8.9  
## BRCA 61.5 54.7 52.3 26.1 35.9 49.0 39.1 42.4 28.4 13.7 8.3 9.2 9.6  
## VREA 40.0 44.3 39.5 50.2 32.3 44.2 35.5 39.7 30.7 43.6 59.8 58.4 60.3 52.2
```

# Escalamiento multidimensional (MDS)

```
mdsanfi<- cmdscale(deucanfi, eig = TRUE)
```

```
mdsanfi
```

```
## $points
```

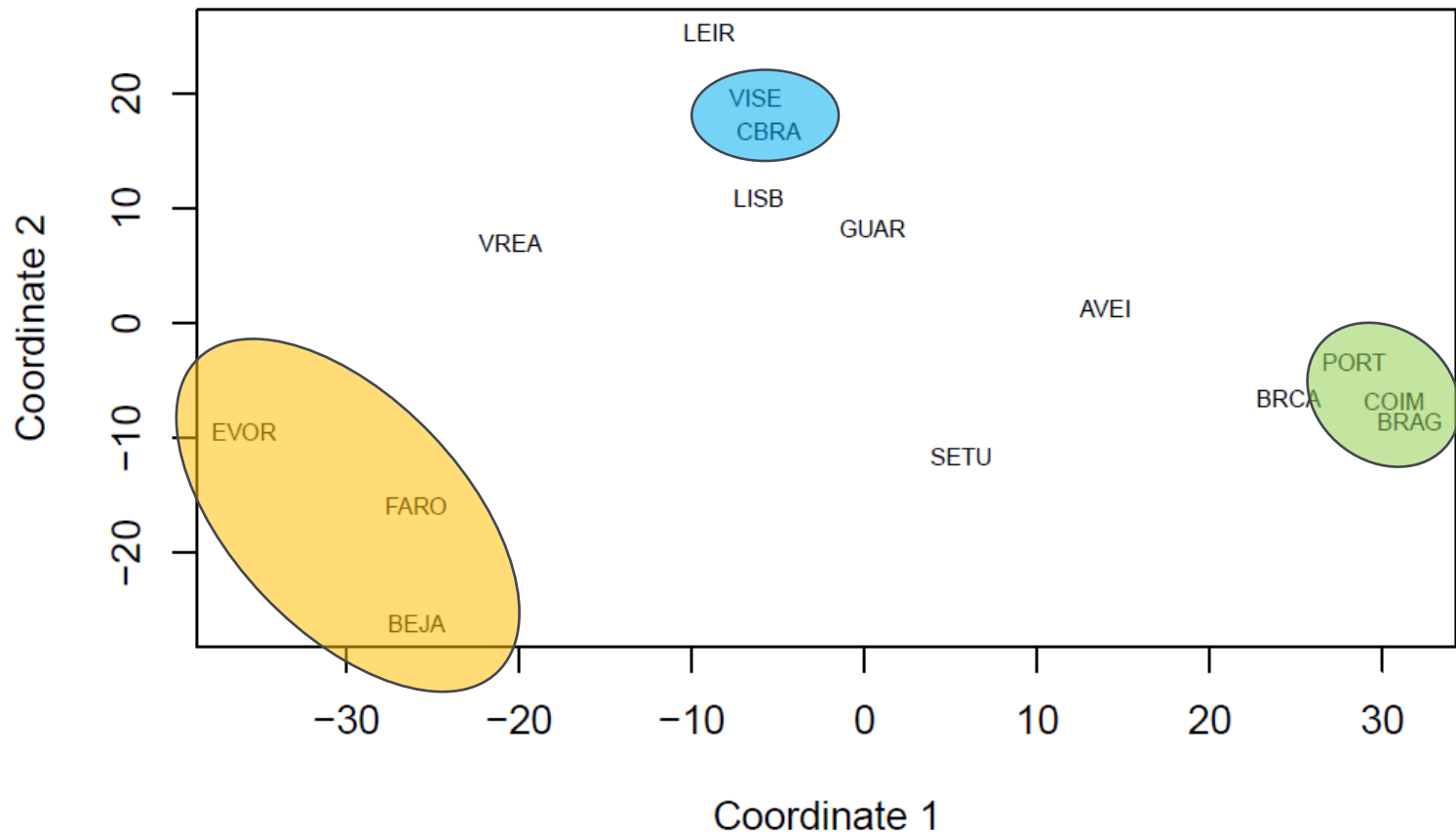
```
##           [,1]      [,2]
## [1,] -35.9372068  -9.447106
## [2,] -25.9322705 -26.132662
## [3,] -25.9645802 -15.938199
## [4,]  5.6126292 -11.697835
## [5,] -6.1061455  10.813623
## [6,] -8.9848858  25.213559
## [7,] -5.5366438  16.597692
## [8,] -6.2936806  19.570766
## [9,]  0.4735304   8.222119
## [10,] 13.9550085   1.242643
## [11,] 30.6908939  -6.783789
## [12,] 28.3591494  -3.467991
## [13,] 31.5838299  -8.565286
## [14,] 24.5776694  -6.527845
## [15,] -20.4972974  6.900310
##
```

```
## $eig
```

```
## [1] 6.820783e+03  2.864971e+03  1.171632e+03  6.150470e+01  1.776065e+01
## [6] 4.014496e+00  9.548624e-13  1.591834e-13  1.492092e-13  1.479654e-13
## [11] 5.156001e-14 -1.019713e-14 -4.809168e-14 -1.853059e-13 -6.608920e-13
```



```
x <- mdsanfi$points[,1]
y <- mdsanfi$points[,2]
plot(x,y, xlab="Coordinate 1",ylab="Coordinate 2",type="n")
text(x,y, labels= anfi[,1],cex=0.6)
```



# Escalamiento multidimensional Não-Métrico(NMDS)

```
library(MASS)
nmmdsanfi<- isoMDS(deucanfi)
```

```
nmmdsanfi
```

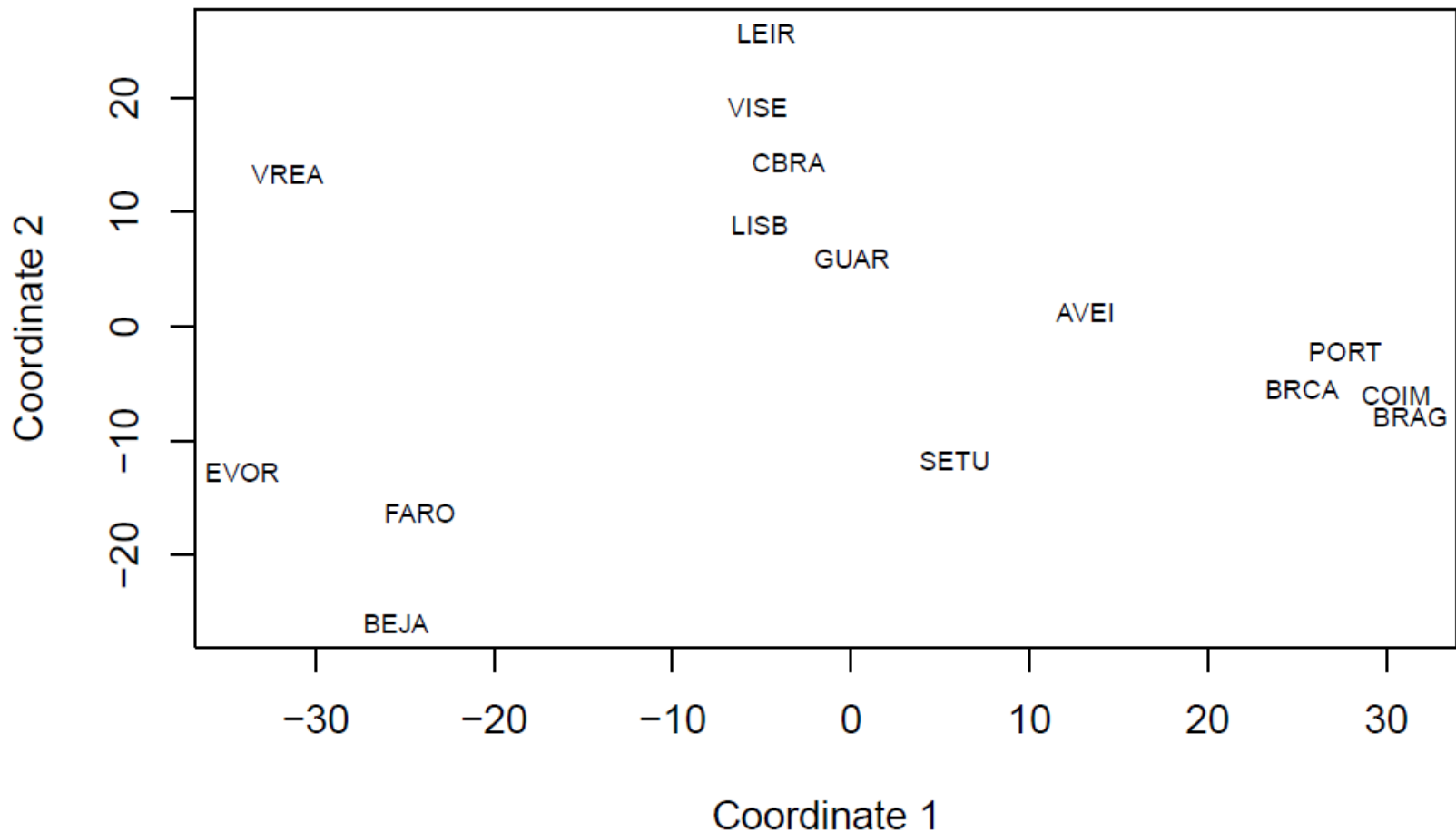
```
## $points
```

```
##           [,1]      [,2]
## [1,] -34.11519011 -12.716941
## [2,] -25.49278236 -25.989141
## [3,] -24.15893226 -16.309516
## [4,]  5.81708937 -11.759853
## [5,] -5.10029304  8.858192
## [6,] -4.75141021  25.687765
## [7,] -3.48416531  14.291690
## [8,] -5.22080235  19.179415
## [9,]  0.05379224  5.977481
## [10,] 13.13845563  1.171623
## [11,] 30.54658981 -6.067311
## [12,] 27.70084748 -2.225513
## [13,] 31.34666571 -7.962404
## [14,] 25.28447070 -5.484271
## [15,] -31.56433532 13.348786
##
```

```
## $stress
```

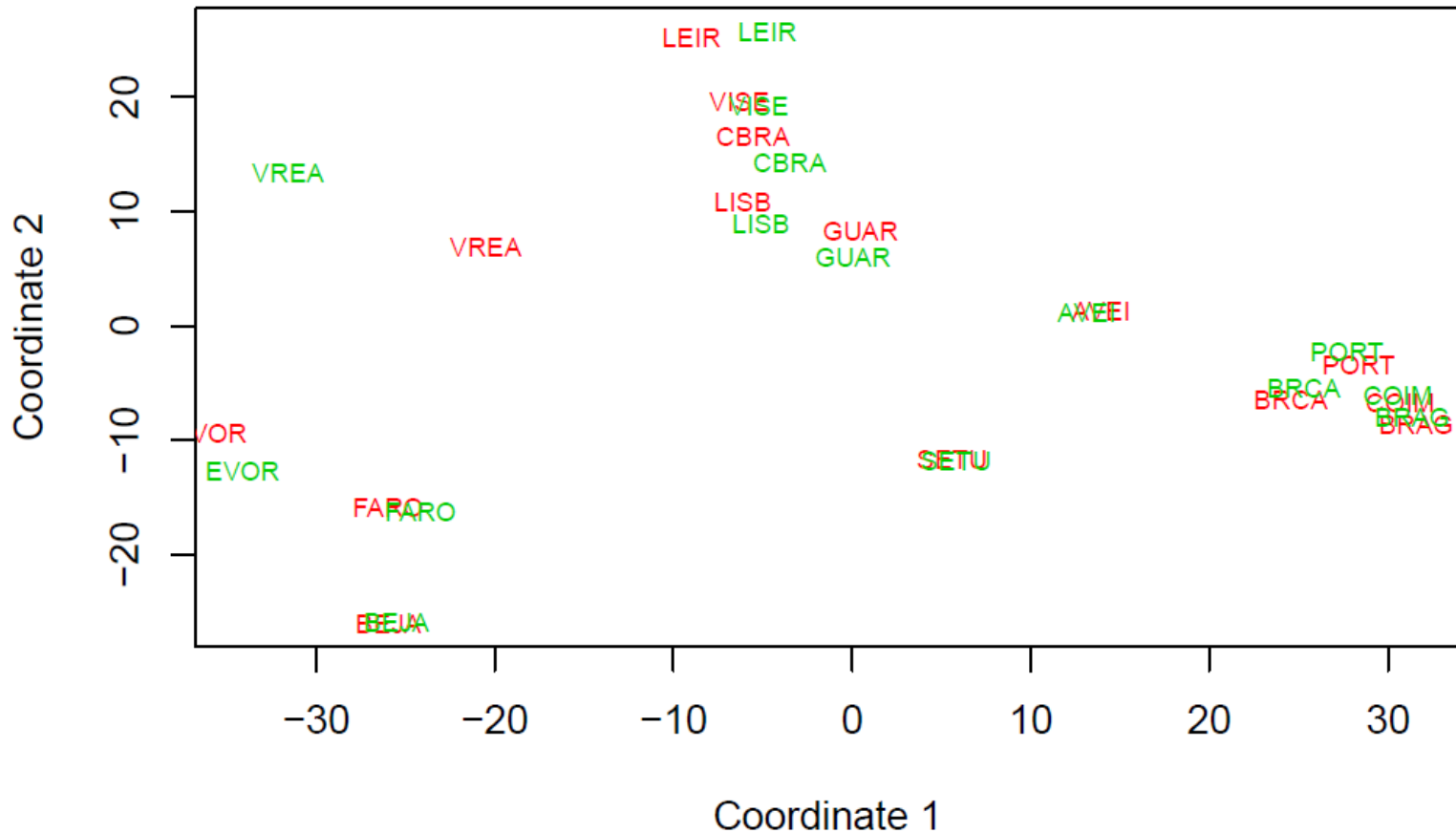
```
## [1] 6.26112
```

```
x2<-nmmdsanfi$points[,1]
y2<-nmmdsanfi$points[,2]
plot(x2,y2,xlab="Coordinate 1",ylab="Coordinate 2",type="n")
text(x2,y2, labels= anfi[,1],cex=0.7)
```



# Not a big difference is there ? MDS vs NMDS

```
plot(x2,y2,xlab="Coordinate 1",ylab="Coordinate 2",type="n")  
text(x,y, labels= anfi[,1],cex=0.7,col=2)  
text(x2,y2, labels= anfi[,1],cex=0.7,col=3)
```



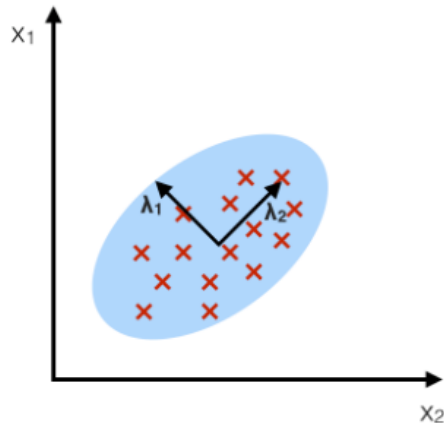


# ANÁLISE DISCRIMINANTE

# Análise Discriminante

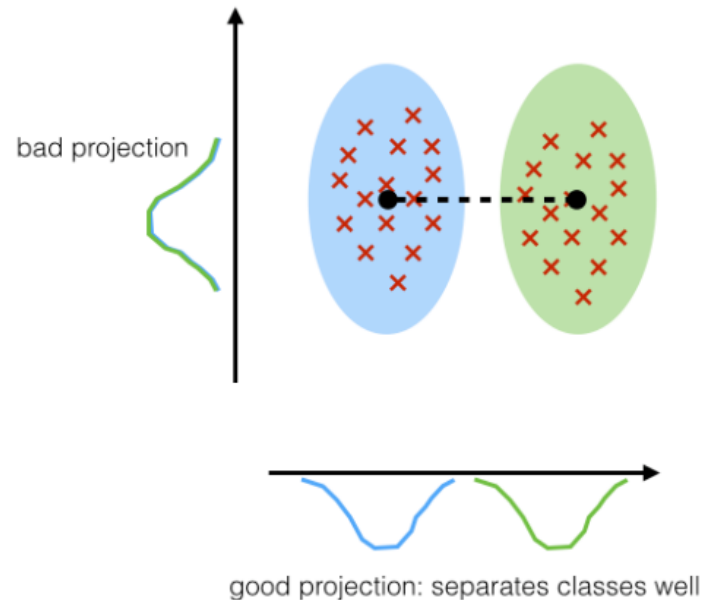
## PCA:

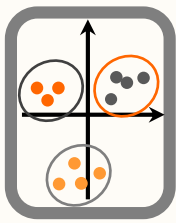
component axes that maximize the variance



## LDA:

maximizing the component axes for class-separation





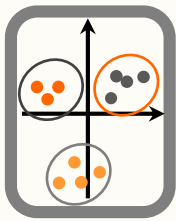
ordenação

---

## Análise discriminante

### *Âmbito:*

- Discriminar diferentes grupos de observações de acordo com um conjunto de variáveis independentes;
- A variável dependente é nominal e será descrita por um conjunto de variáveis independentes. (paralelismo com a regressão logística, ou multinomial)



ordenação

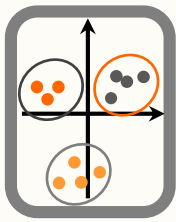
---

## Análise discriminante

A função discriminante que permite classificar as observações nos vários grupos é dada por:

$$Z = W_1 X_1 + W_2 X_2 + \dots + W_n X_n$$



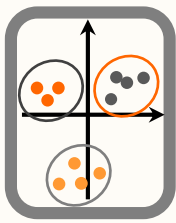


ordenação

---

## Análise discriminante

- A análise discriminante permite igualmente testar hipóteses sobre a igualdade das médias dos vários grupos.
- Os procedimentos de teste são semelhantes a uma (M)ANOVA.

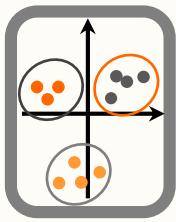


ordenação

---

## Análise discriminante

- Tem como pressupostos a multinormalidade e a homogeneidade de variâncias.
- Quando utilizada no âmbito dos testes de hipóteses e inferência estatísticas estes devem ser cumpridos, se utilizada como técnica exploratória com a performance avaliada pela sua capacidade de predição, esta questão não é essencial.



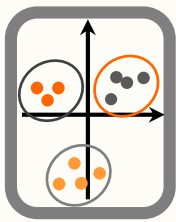
ordenação

---

## Interpretação dos resultados

*Discriminant weights:* (semelhantes aos coeficientes de regressão) são obtidos para cada variável independente e permitem avaliar a magnitude de variação.

*Discriminant loadings:* correlações entre as variáveis independentes e as funções discriminantes.



ordenação

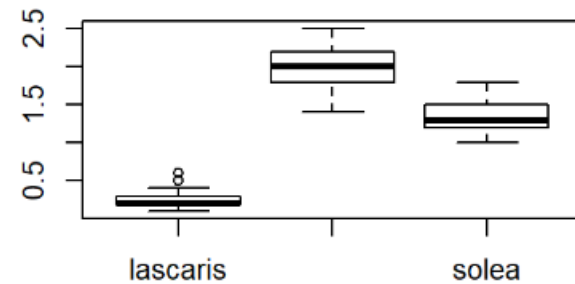
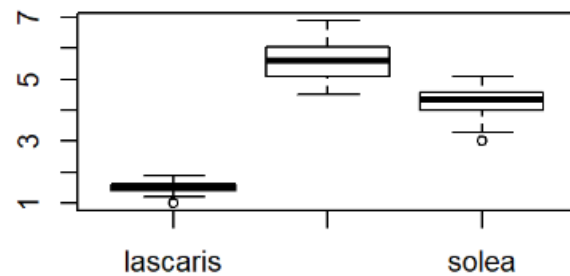
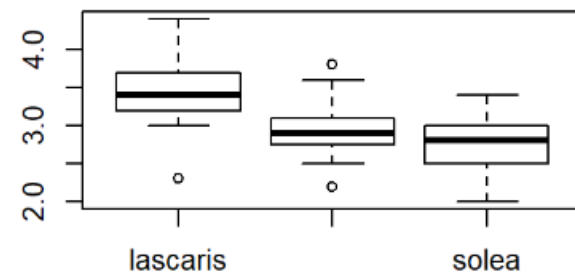
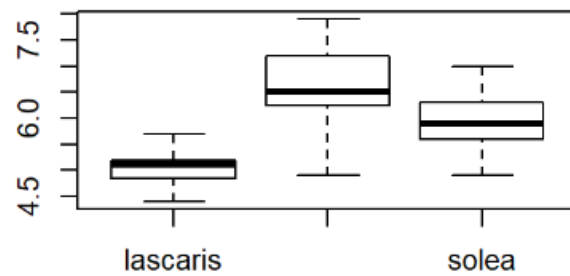
---

## Validação dos resultados

- O desempenho do modelo é avaliado usualmente comparando os valores previstos e observados para a variável dependente, i.e. percentagem de classificações correctas (geralmente usando uma *confusion matrix*).
- Muitas vezes é utilizada uma outra matriz de observações para validar o modelo.

```
lingua <- read.csv("DataTP11solea.csv", header=TRUE, sep=";")
```

```
par(mfrow=c(2,2))  
with(lingua,boxplot(m1~especie))  
with(lingua,boxplot(m2~especie))  
with(lingua,boxplot(m3~especie))  
with(lingua,boxplot(m4~especie))
```



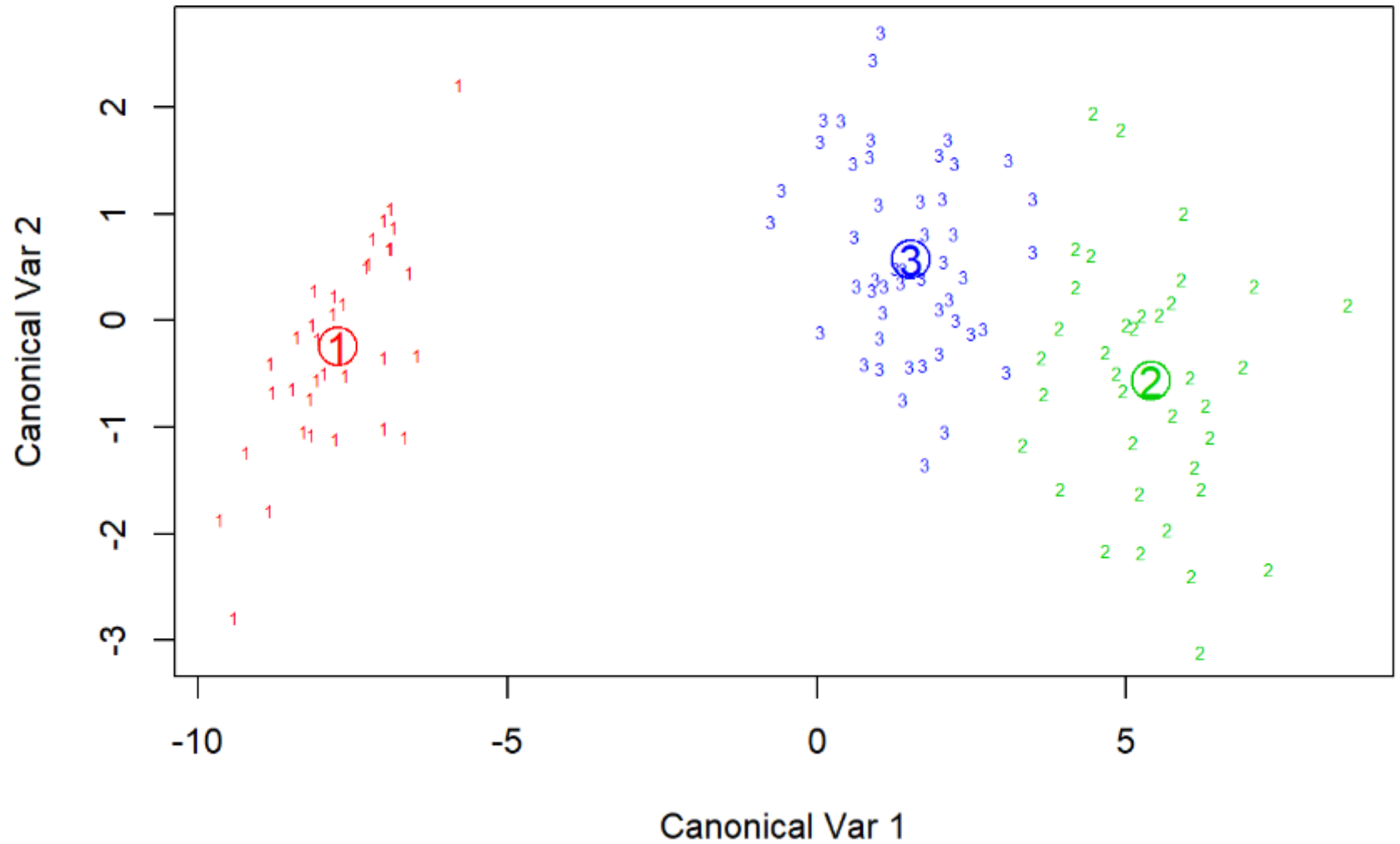
We now implement a discriminat analysis using again `function fda` in package `mda`.

```
linguada<-fda(especie~.,data=lingua)
linguada
```

```
## Call:
## fda(formula = especie ~ ., data = lingua)
##
## Dimension: 2
##
## Percent Between-Group Variance Explained:
##      v1      v2
##  99.1 100.0
##
## Degrees of Freedom (per dimension): 5
##
## Training Misclassification Error: 0.02479 ( N = 121 )
```

```
plot(linguada)
```

## Discriminant Plot for predict classes



## Confusion matrix

```
linguada$confusion
```

```
##           true
## predicted  lascaris senegalensis solea
## lascaris      35         0         0
## senegalensis  0         35         2
## solea         0         1        48
```

In this case, only 3 of the 121 individuals were incorrectly classified. The real classification performance should be evaluated against new observations.



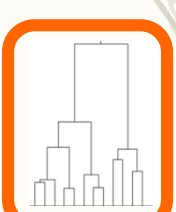
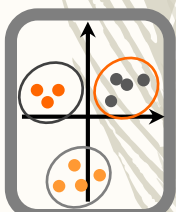
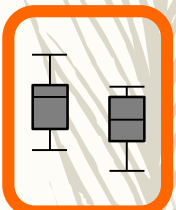
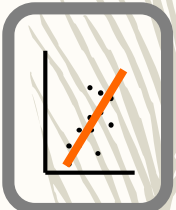
# ecologia numérica

---

Take home messages

Wrap up

Síntese



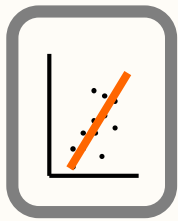
É sempre bom  
voltar ao  
início!

# O que quero que saibam (ou gostaria que soubessem!) fazer no final deste curso

---

- Compreender números e saber interpretá-los num contexto ecológico
- Saber como planear a recolha de dados para responder a uma questão ecológica
- Saber como analisar os dados resultantes
- Ter uma noção básica das principais técnicas de análise de dados ecológicos
- Saber o que não sabem e quando pedir ajuda para analisar um conjunto de dados
- Identificar potenciais problemas em ecologia relacionados com a análise de dados
- Ter sensibilidade e bom senso!!!

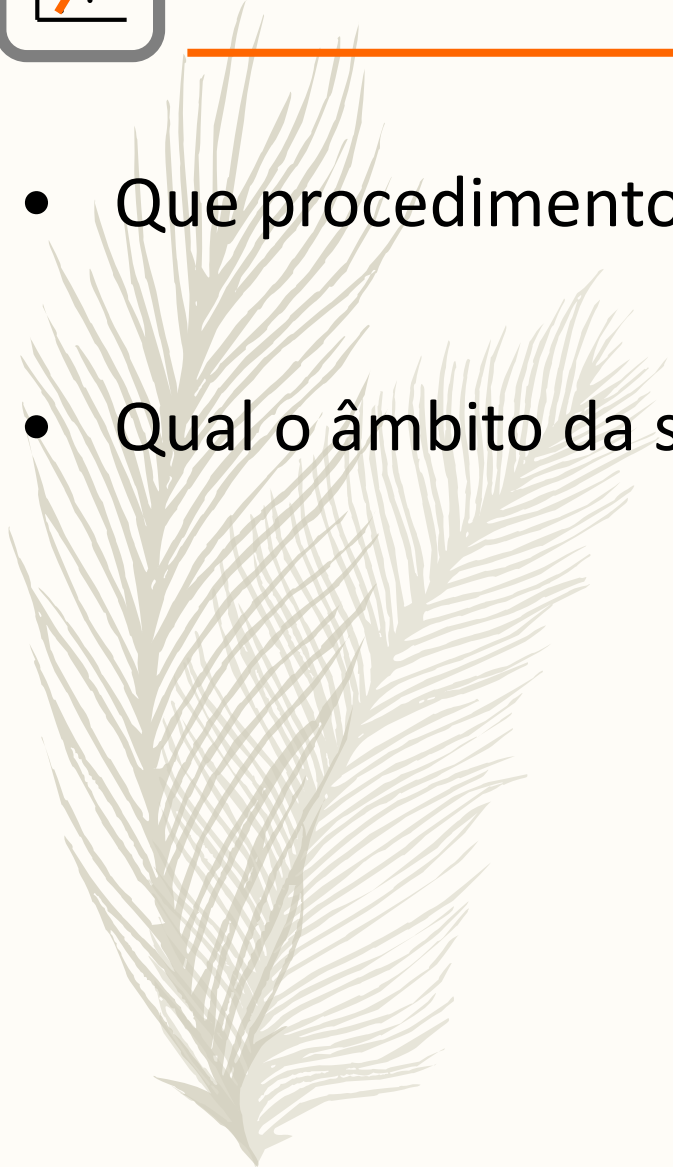
Bonus: usar o R e fazer relatórios dinâmicos em Rmarkdown (uma linha extra no CV)

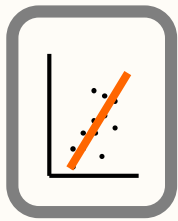


## síntese

---

- Que procedimentos utilizar em diferentes situações?
- Qual o âmbito da sua aplicação?





síntese

---

## Âmbito do estudo

Comparação de amostras

Relações entre variáveis

Evidenciar padrões em dados multivariados



síntese

## Âmbito do estudo

Uma variável  
de interesse

Várias  
variáveis  
de interesse

- Comparação de amostras
- Relações entre variáveis
- Evidenciar padrões em dados multivariados



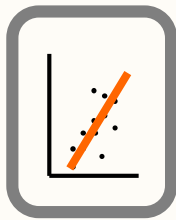
síntese

## Âmbito do estudo

- Comparação de amostras
- Relações entre variáveis
- Evidenciar padrões em dados multivariados

Inferência  
estatística

Análises  
exploratórias

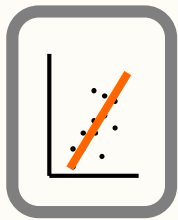


síntese

---

# Principais abordagens metodológicas

- **Comparação de amostras**
  - Testes a 1 e 2 amostras
  - ANOVA & KW
  - ANOVA Multifactorial
- Relações entre variáveis
- Evidenciar padrões em dados multivariados



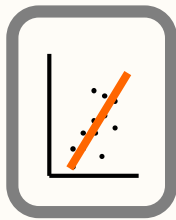
síntese

---

# Principais abordagens metodológicas

- Comparação de amostras
- **Relações entre variáveis**
  - Regressão e correlação
  - Tabelas de contingência (independência)
  - PCA, CA, CCA
- Evidenciar padrões em dados multivariados





síntese

---

# Principais abordagens metodológicas

- Comparação de amostras
  - Relações entre variáveis
  - Evidenciar padrões em dados multivariados
- 

- Classificação
- PCA, CA, CCA, MDS
- Análise discriminante



# CURRENT TRENDS IN ECOLOGICAL STATISTICS ARE DETACHED FROM ECOLOGISTS' STATISTICAL TEACHING

"...Ainda ensinamos estatística como há 20 anos atrás..." –  
Mária do Rosário Oliveira, 2019

TIAGO A. MARQUES



9<sup>TH</sup> NOVEMBER 2019

# SO WHAT SHOULD WE AIM TO TEACH BIOLOGY STUDENTS? (SOME RANDOM THOUGHTS)

- Data: collection, processing, management
- **Think before acting**
- Randomness, variability, confounding
- **Uncertainty is a good thing**
- Models are not truth
- Decisions under uncertainty lead to errors
- R programming, dynamic reports and reproducible research
- Data visualization (if you can't plot it... is it real?)
- Regression models: GLMs, GAMs, GLMMs, GAMMs
- Bayesian paradigm
- The limitations of statistics



# FIM

**98% OF ALL  
STATISTICS  
ARE MADE UP**

PictureQuotes.com

PICTUREQUOTES

82.7% of all statistics are made up  
on the spot.

- Steven Wright -

[www.StubhadEEP.com](http://www.StubhadEEP.com)

**69%** of all  
statistics  
are **made up!**

**93.45%**  
**OF ALL  
STATISTICS  
ARE MADE UP**