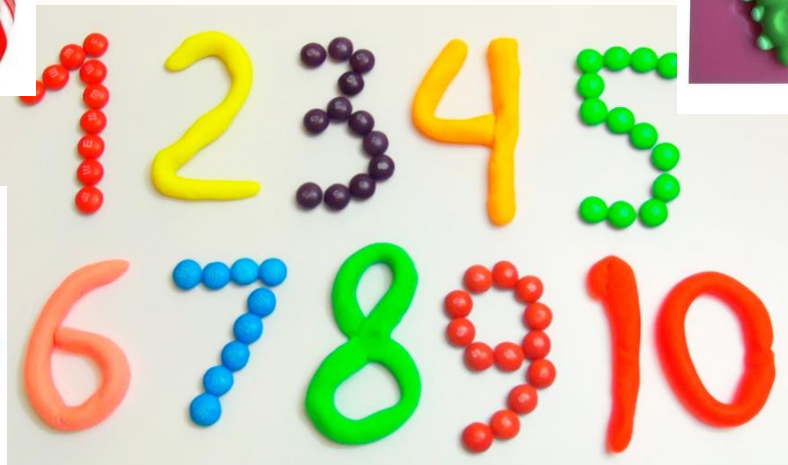


# Aula 14 Goodies\*



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

# SEMINÁRIO

## ***HIDDEN MARCOV MODELS OF ANIMALS MOVEMENT AND BEHAVIOR***

*Dr. Theo Michelot*

### *Ur* **Abstract:**

*Large telemetry data sets are collected on wild animals, and they can be used to learn about their behaviour, and their response to environmental features. Hidden Markov models (HMMs) provide a flexible and rigorous statistical framework to describe animal movement as the result of behavioural states. From an HMM fitted to telemetry data, we can infer the sequence of unobserved behavioural states over the duration of the study, and movement parameters for each behaviour. HMMs are relatively fast to fit, usually taking only a few minutes for tens of thousands of data points. The effects of environmental covariates can be modelled, to learn about the drivers of behavioural switches. I will describe the general formulation of HMMs, and then illustrate the models with various data sets from marine and terrestrial species.*

**Local: Faculdade Ciências Universidade de Lisboa**

**Bloco C/2 Piso 2 Sala: 2.214**

**5 novembro 2019 (3.<sup>a</sup> feira) – 14:00**

lista candidata ao Núcleo de Estudantes de Biologia da FCUL - Eleições: sexta feira



Vamos estar no C2 até quarta (a responder a quaisquer dúvidas que haja)

3 - Por tempo de antena entende-se o espaço de programação própria da responsabilidade do titular do direito, facto que deve ser expressamente mencionado no início e no termo de cada programa.



Ciências  
ULisboa



PRÉMIO  
VOLUNTARIADO  
UNIVERSITÁRIO  
2019



O **Prémio de Voluntariado Universitário** visa reconhecer os projetos de voluntariado promovidos por estudantes do Ensino Superior, com benefícios para a sociedade e igualmente para o desenvolvimento das suas competências pessoais.

#### Quais as categorias de Prémio?

- **Menção Honrosa "IES + Voluntária"** | vamos homenagear a Instituição de Ensino Superior que apresentar o maior número de candidaturas!
- **Apoio Financeiro** | 4.000€ para os 3 projetos vencedores; 1.000 para os restantes 7 finalistas; e 1.000€ no âmbito do prémio Comunicação ao vídeo finalista com maior número de votos online
- **Mentoria Banco** | de um diretor do Banco Santander, capaz de te ajudar a identificar oportunidades, esclarecer dúvidas e definir planos de ação
- **Apoio à Comunicação** | pela divulgação da cerimónia de entrega de prémios e nossas redes

As candidaturas devem ser submetidas até **14 de novembro de 2019** e devem ser feitas online.

**Mais informações**



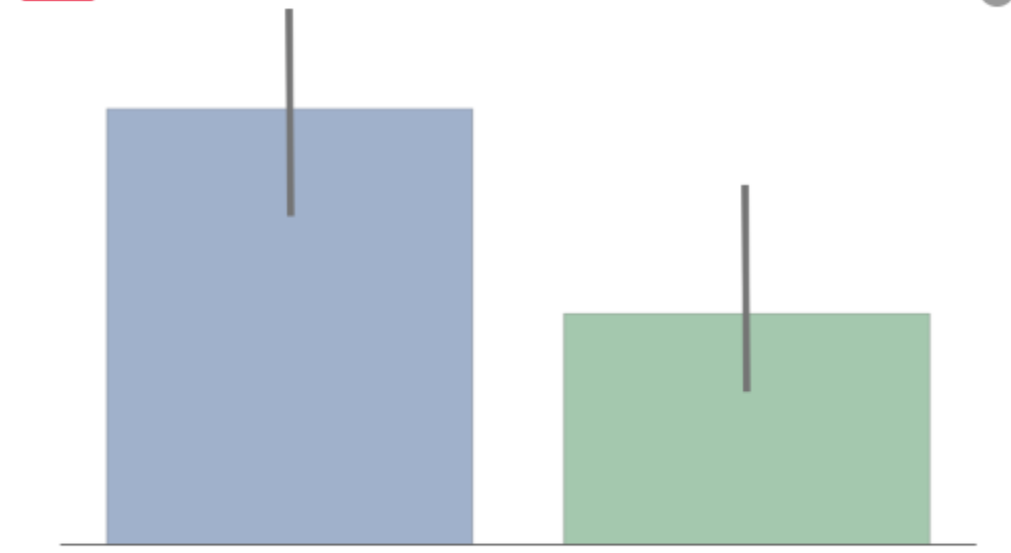
# Why Overlapping Confidence Intervals mean Nothing about Statistical Significance



Prasanna Parasurama [Follow](#)

Mar 17, 2017 · 4 min read

[Save](#)



*“The confidence intervals of the two groups overlap, hence the difference is not statistically significant” — A lot of People*

Esta semana, na TP, ficha TP6.pdf  
(sobre comparações com 3 ou mais amostras)

SEXTA FEIRA TP – Não vou conseguir dar a aula TP

Quem conseguir vir já hoje,  
15:00-17:00,  
optimo

–

fica já despachado!

Quem não conseguir, na sexta dia 15, aula a dobrar...  
se puderem, claro.

Caso contrário, marcarei uma aula extra para compensar.

Ecología Numérica - Aula Teórica 14 – 04-11-2019

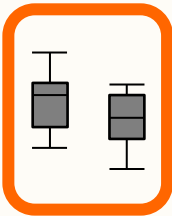


If we gather more and more data and establish more and more associations, however, we will not finally find that we know something. We will simply end up having more and more data and larger sets of correlations.

— *Kenneth Waltz* —

AZ QUOTES

<https://www.azquotes.com/quote/1116537>



anova multifactorial

---

## Casos particulares de ANOVA:

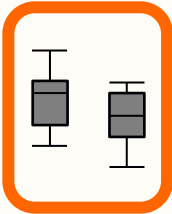
Há diversas situações que configuram casos especiais de ANOVA, em geral, resultantes da falta de independência da dimensão temporal e espacial, tais como:

ANOVA hierárquica

ANOVA por blocos

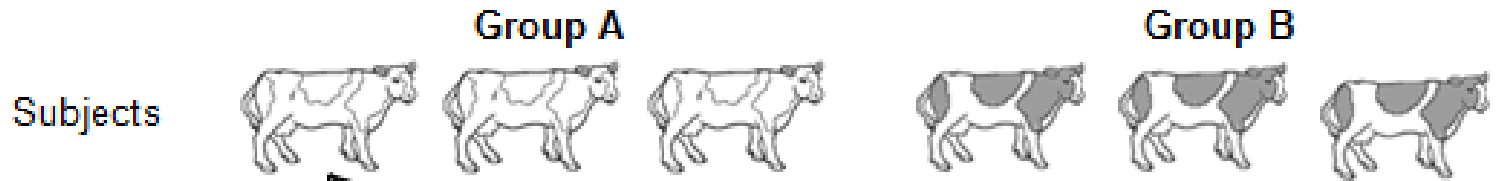
ANOVA para medidas repetidas





anova multifactorial

# ANOVA hierárquica



Subjects

Group A

Group B

Tissue samples

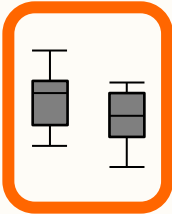
```
aov(qPCR~grupo+vaca/tissue, data=dados)
```

Tissue is nested in vaca

Reverse transcriptase

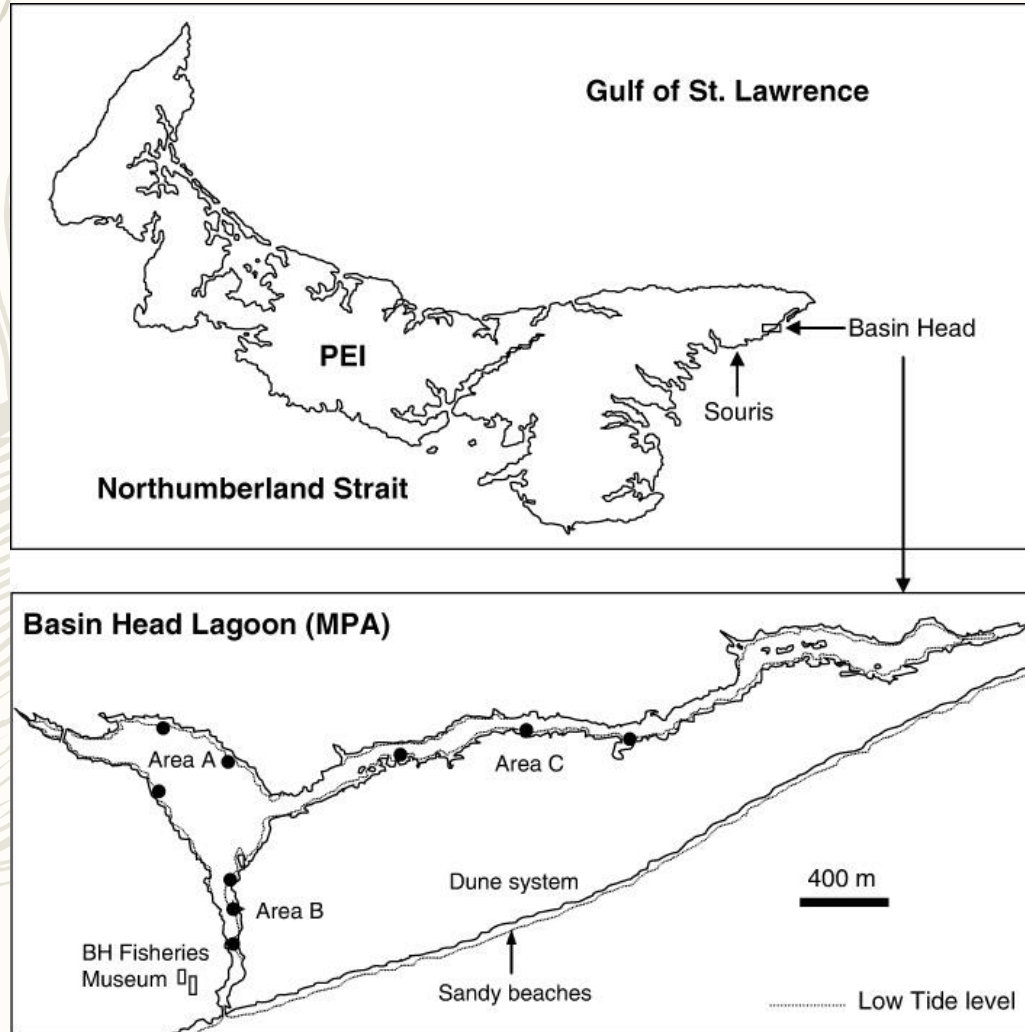
qPCR

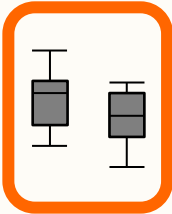
We could have many levels in the hierarchy, but that complicates stuff



anova multifactorial

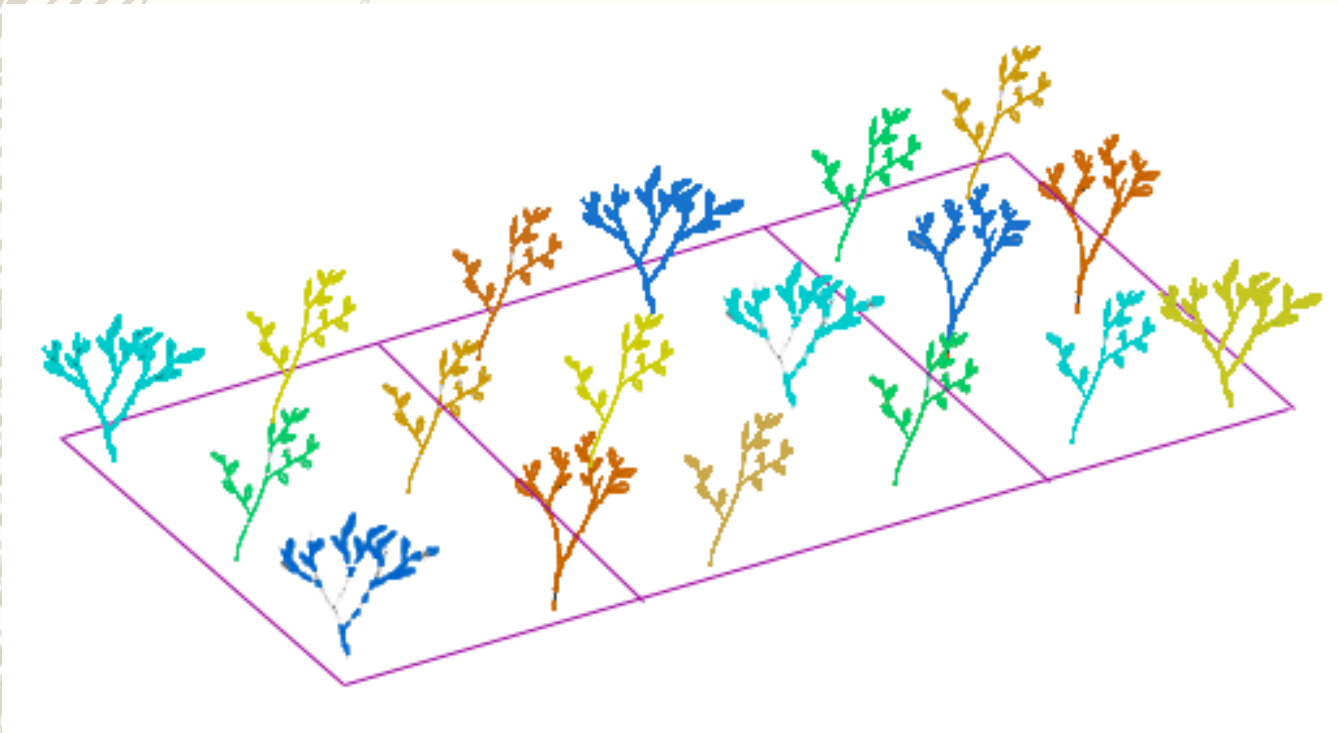
# ANOVA hierárquica

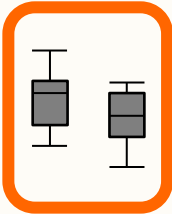




anova multifactorial

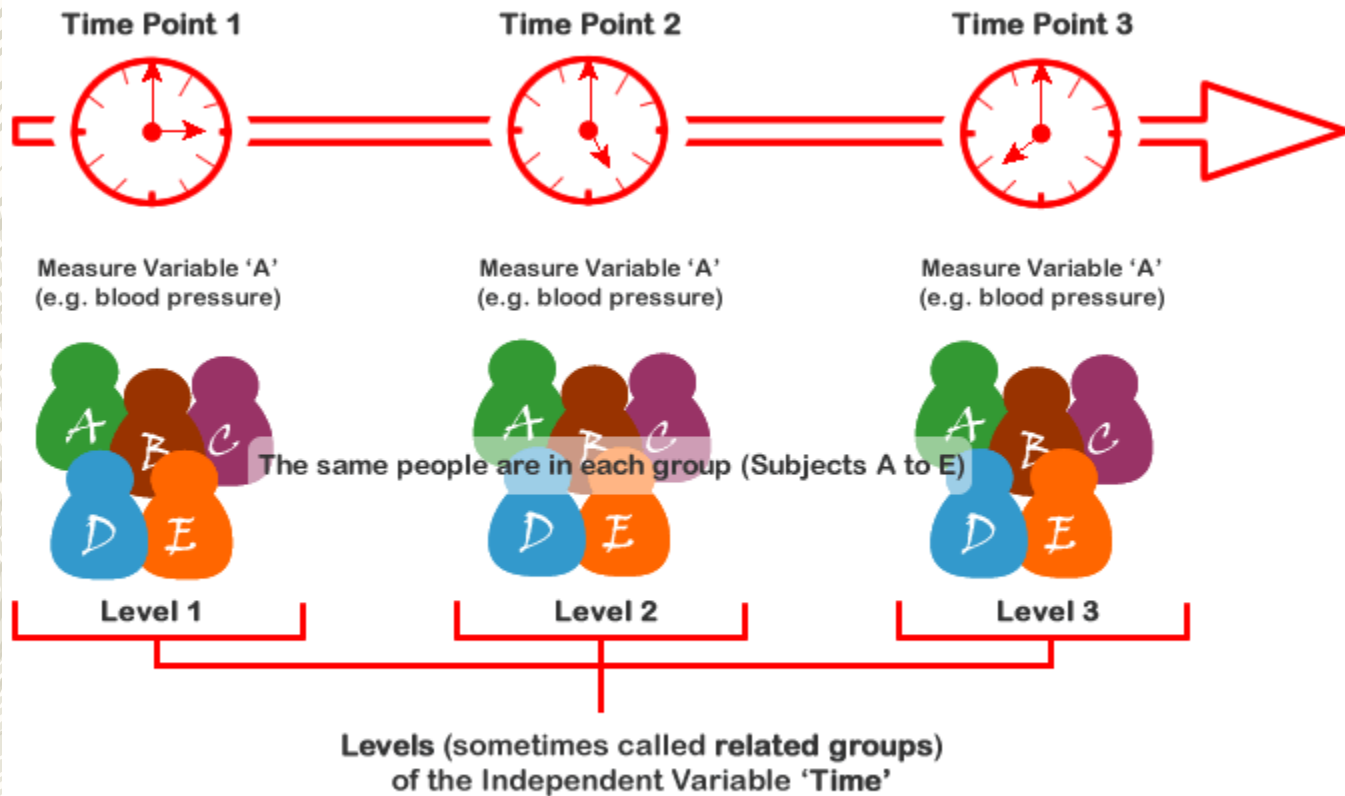
# ANOVA por blocos

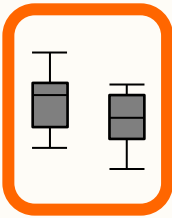




anova multifactorial

# ANOVA para medidas repetidas





anova multifactorial

---

## Casos particulares de ANOVA:

A forma de cálculo das somas dos quadrados e, portanto, das estatísticas  $F$ , é diferente das ANOVAS simples ou multifactoriais simples.

A implementação destas análises permite inferir sobre estes efeitos resultantes da falta de independência.

# ANOVA hierárquica

```
aov (Y ~ X1 + X2 / X3, data = dados)
```

Y depende de X1 e X2, e de X2 que está *nested* (aninhado) em X3

# ANOVA medidas repetidas ou blocos

```
aov (Y ~ X1 + Error (X2), data = dados)
```

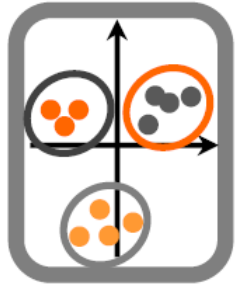
Y depende de X1 e de X2 (e X2 é um efeito aleatório, como um bloco)

Aqui ficam apenas a saber que estas alternativas existem para dados não independentes.

Se as quiserem implementar terão de procurar mais detalhes por vocês mesmos (2 exemplos na TP).

De qualquer forma, existem metodologias genéricas cujo objetivo é lidar com estes casos específicos bem como todos os outros que, não sendo casos especiais, podem aparecer na prática.

São os modelos de regressão, que abordamos de seguida.



# Ecologia Numérica

Componente Teórica - Prática

Ficha de trabalho **6**

6. Recolheu amostras a Norte e a Sul de Peniche, em várias praias da costa portuguesa sujeitas a diferente grau de hidrodinamismo (calmo e exposto), e foram quantificadas as densidades de cracas nas zonas rochosas (DataTP6cracas.csv). Admitindo que os pressupostos são verificados, efectue uma **análise de variância hierárquica**. Quais as principais conclusões?

Os replicados, que podem vir de zonas calmas ou expostas, estão aninhados praias.

7. Administrou uma substância em ratos de laboratório e pretende agora medir o seu efeito na concentração de uma enzima no sangue. Efectuou medições em 10 ratos, duas horas e 24 horas após a administração da substância (DataTP6ratos.csv). Admitindo que os pressupostos são satisfeitos, efectue **uma análise de variância que seja adequada e interprete os resultados**.

Medidas repetidas = Rato é um “bloco”



Muitos dos procedimentos clássicos de inferência estatística não são mais do que casos particulares de modelos de regressão.

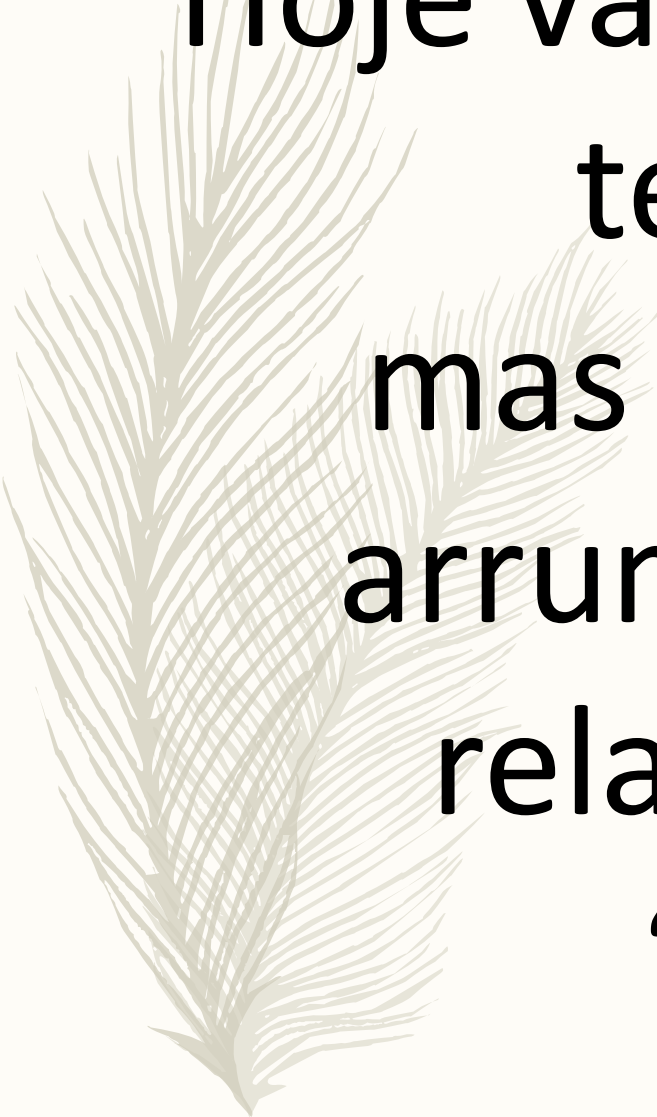
O teste de t-student é uma regressão de uma variável dependente em função de um factor com apenas dois níveis

No caso de uma ANOVA, temos uma regressão de uma variável dependente em função de um factor.

No caso de uma ANOVA a dois factores, temos uma regressão de uma variável dependente em função de dois factores.

No caso de uma ANCOVA, temos uma regressão de uma variável dependente em função de um factor e de uma variável quantitativa.





Hoje vamos entrar num  
tema novo,  
mas antes, vamos  
arrumar a casa em  
relação a temas  
“antigos”

## Análise de variância – Implementação em R

```
#simple anova  
Anova1=aov(y~f1)
```

```
#anova two way  
Anova2W=aov(y~f1*f2)
```

```
#anova 2 way without interaction  
Anova2WsemInteracção==aov(y~f1+f2)
```

```
#anova k way  
AnovaKway=aov(y~f1*f2*...*fk)
```

```
#Ancova  
Ancova1=aov(y~x1+f2)
```

```
#Hierarchical Anova - f2 nested inside f1  
Anova1H=aov(y~f1/f2) or Anova1H=aov(y~ f1 +f1:f2)
```

```
#Repeated measurements anova - f1 measured several times in each level of f2  
Anova1R=aov(y~f1+Error(f2/f1)) Example here
```

y: variável resposta  
f: um factor  
x: variável quantitativa)

# Testes de hipóteses in a nutshell

Hipótese nula ou  $H_0$ : geralmente uma igualdade ou ausência de efeito

Hipótese alternativa ou  $H_1$ : a negação de  $H_0$ , geralmente o que queremos demonstrar

bilateral -  $H_0: \mu=0$  vs.  $H_1: \mu \neq 0$

unilateral -  $H_0: \mu \geq 0$  vs.  $\mu > 0$  ou  $\mu \leq 0$  vs.  $\mu < 0$

Erro tipo I: rejeitar  $H_0$  sendo  $H_0$  verdade (nível de significância do teste  $\alpha$ )

Erro tipo II: não rejeitar  $H_0$  sendo  $H_0$  falsa ( $\beta$ )

$1 - \beta$  = potência do teste = rejeitar  $H_0$  sendo  $H_0$  falsa

Nível de significância, ou  $\alpha$ :

- o risco de cometer um erro de tipo que estamos dispostos a correr
- Definido à priori e independente dos dados
- Valores usuais: 0.01, 0.05, 0.1

Estatística de teste – quantidade que depende dos (i.e. é uma função dos) dados e cuja distribuição, sob  $H_0$  (i.e. sendo  $H_0$  verdadeira), é conhecida

# O procedimento:

Formular as hipóteses, recolher a(s) amostra(s) e calcular a estatística de teste

Comparar estatística de teste com a distribuição sob  $H_0$ .

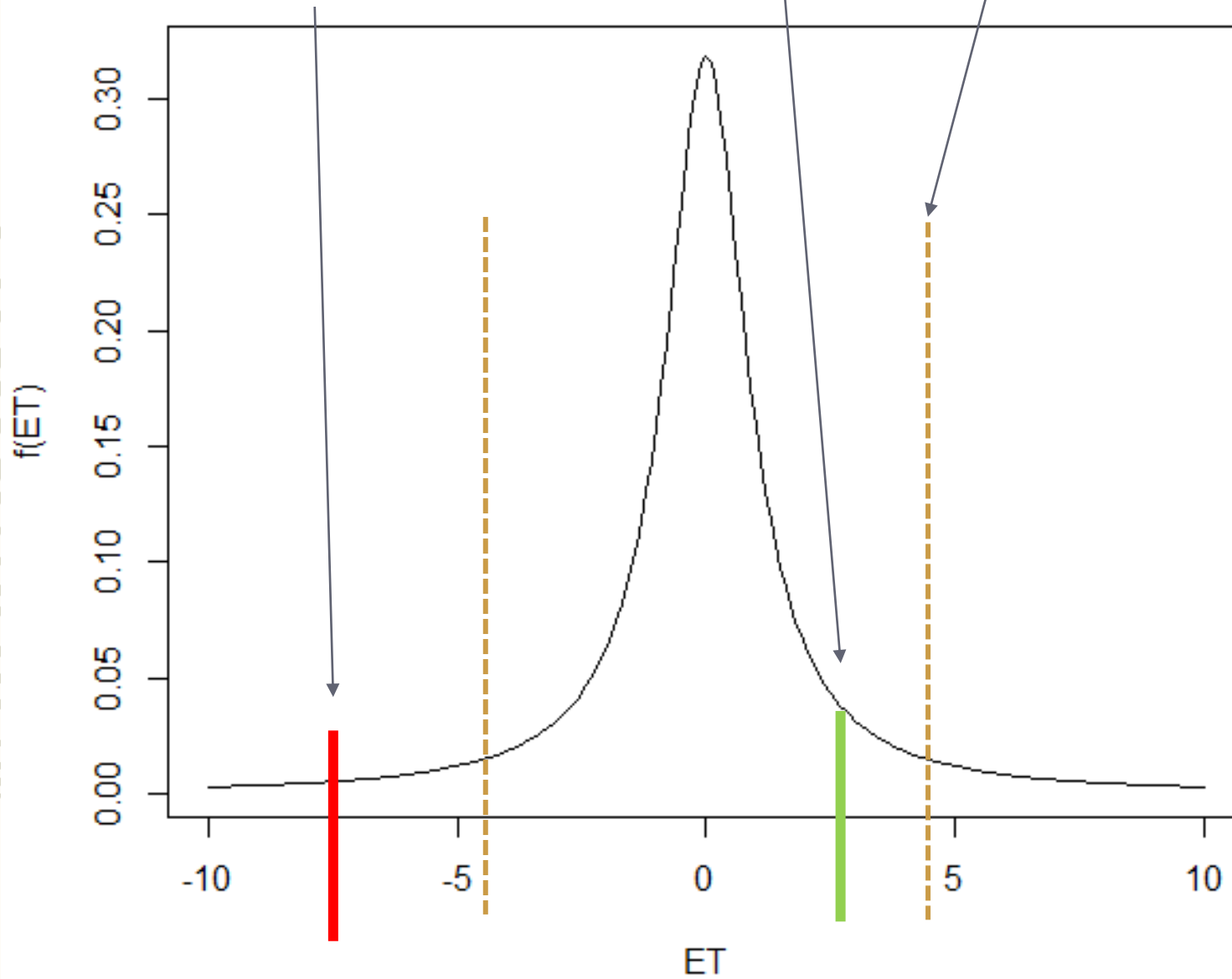
1. Se o valor for extremo, rejeitar  $H_0$
2. Se não for, não rejeitar  $H_0$

Mas como decidir se um valor é extremo?

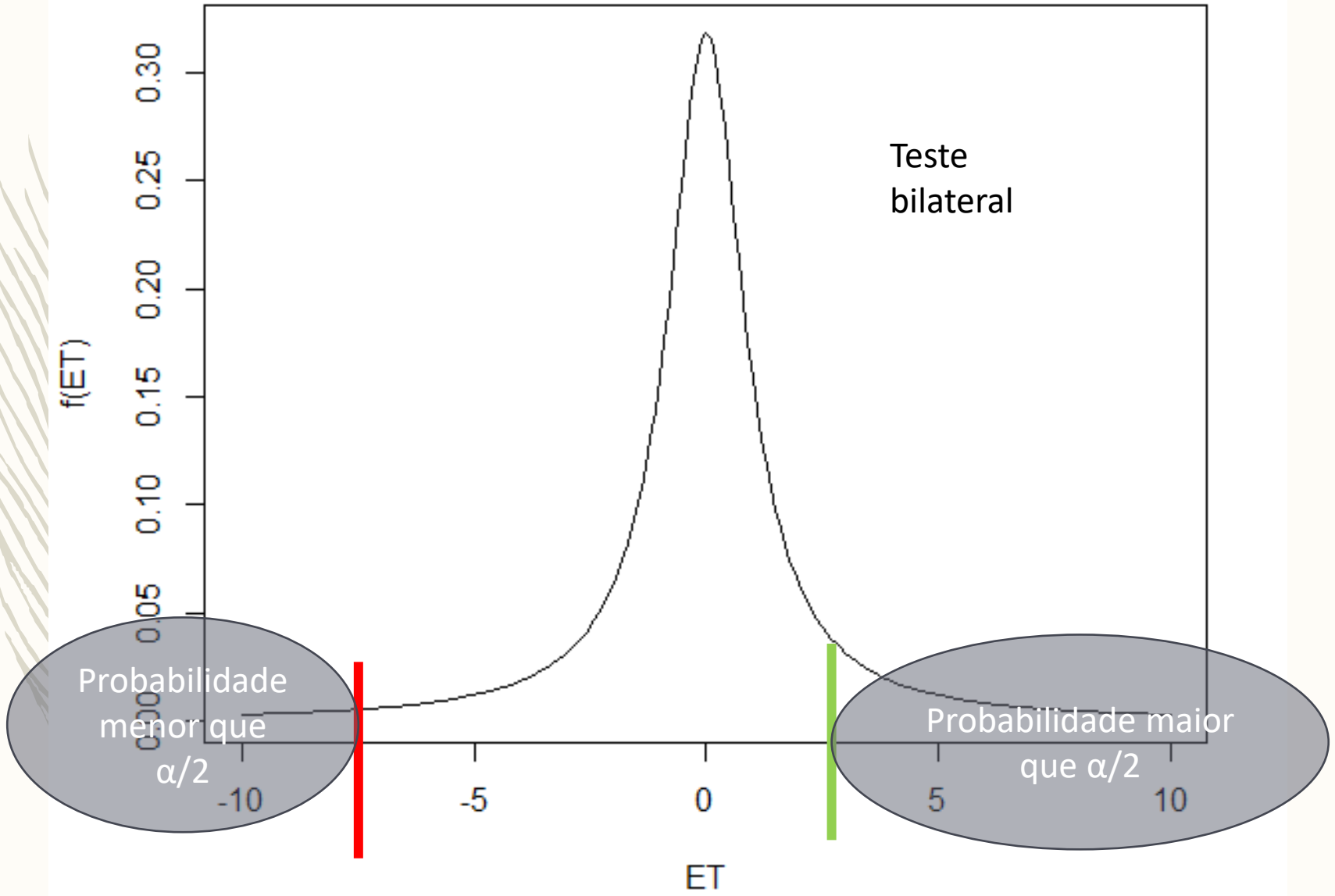
De duas formas equivalentes:

1. Usando valores criticos, que dependem do  $\alpha$  desejado
2. Calculando o P-Value, que comparamos com o  $\alpha$  desejado

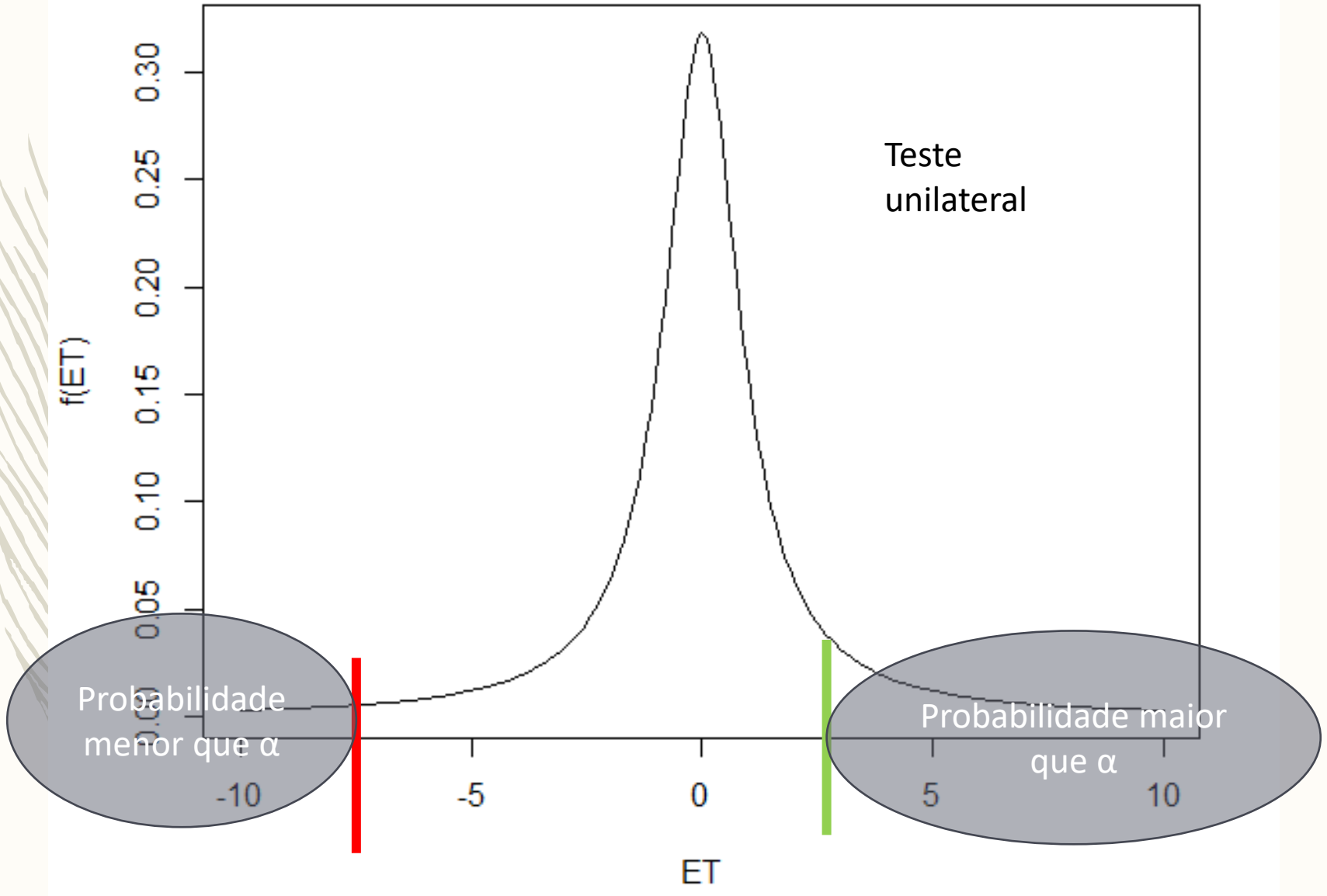
Usando valores críticos: se ET for mais extrema que o **valor crítico** para o nível de significancia  $\alpha$ , **rejeitar**  $H_0$ , caso contrário, **não rejeitar**  $H_0$



Usando o P-value: se o P-value, a probabilidade de observar sob H0 uma estatística de teste tão ou mais extrema que a observada, for menor que  $\alpha$ , **rejeitar** h0, caso contrário, **não rejeitar** H0



Usando o P-value: se o P-value, a probabilidade de observar sob  $H_0$  uma estatística de teste tão ou mais extrema que a observada, for menor que  $\alpha$ , **rejeitar**  $H_0$ , caso contrário, **não rejeitar**  $H_0$





A única diferença entre um teste unilateral e bilateral é qual a zona de rejeição, ou seja, qual o valor crítico da estatística de teste, ou, equivalentemente como se calcula o P-value

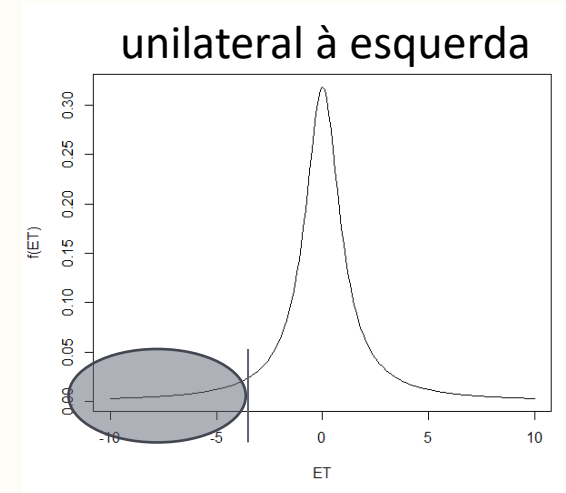
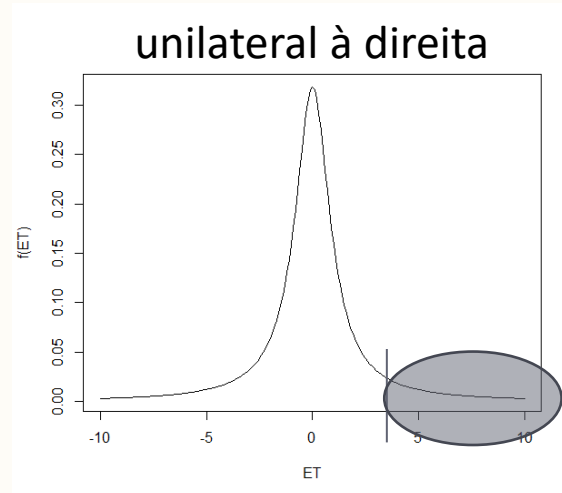
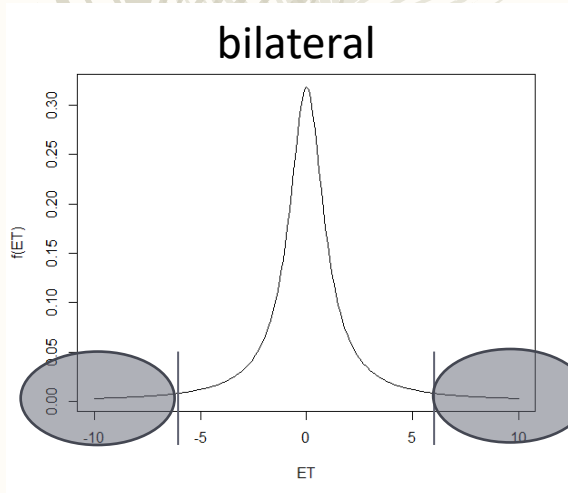
bilateral

$$P\text{-value} = 2 * \min(P(ET > \text{valor observado}) P(ET < \text{valor observado}))$$

unilateral

$$P\text{-value} = P(ET > \text{valor observado}) \text{ num teste unilateral à direita}$$

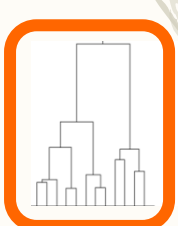
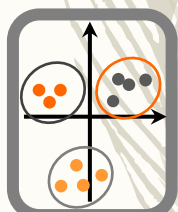
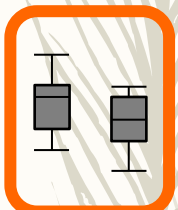
$$P\text{-value} = P(ET < \text{valor observado}) \text{ num teste unilateral à esquerda}$$



# Ecologia Numérica

---

regressão e  
modelos lineares generalizados



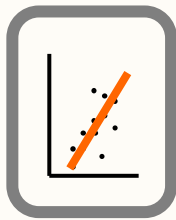
Francis Galton called it regression (toward the mean) because he modeled sons heights using their fathers height and noticed that sons height 'regresses' towards the mean. In other words tall fathers tend to have shorter sons and vice versa.



## regressão e MLG

---

- Que análise efectuar quando temos uma variável dependente e uma ou mais independentes?
- Quais as potencialidades e limitações da análise de regressão?
- Como interpretar os seus resultados?
- Como generalizar a regressão a casos para os quais os pressupostos não são verificados?



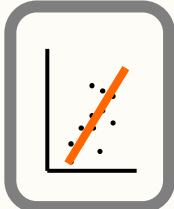
## Análise da regressão

### *Âmbito:*

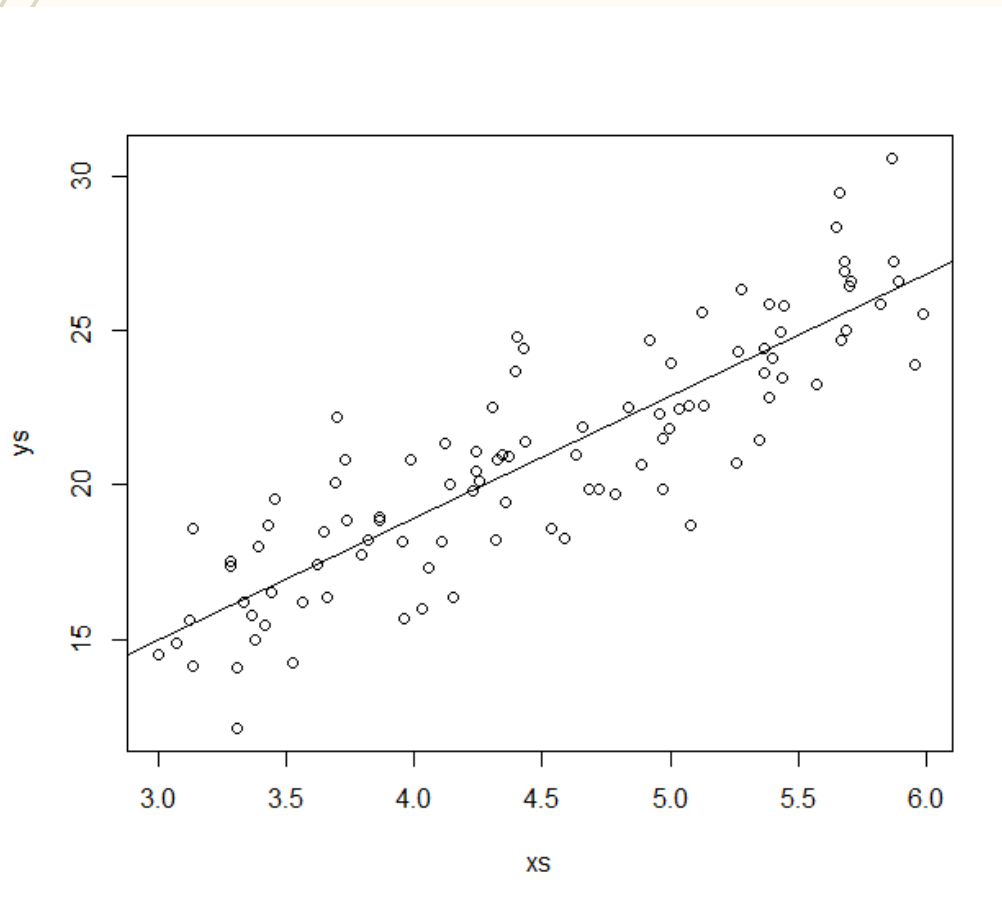
Avaliar as relações entre duas (ou mais) variáveis

### *Objectivos:*

- Modelação dum evento ecológico
- Testes de hipóteses
- Predição (modelo preditivo)



# Análise da regressão





## O modelo linear

$$Y = \beta_0 + \beta_1 X$$

Cada observação é dada por

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$



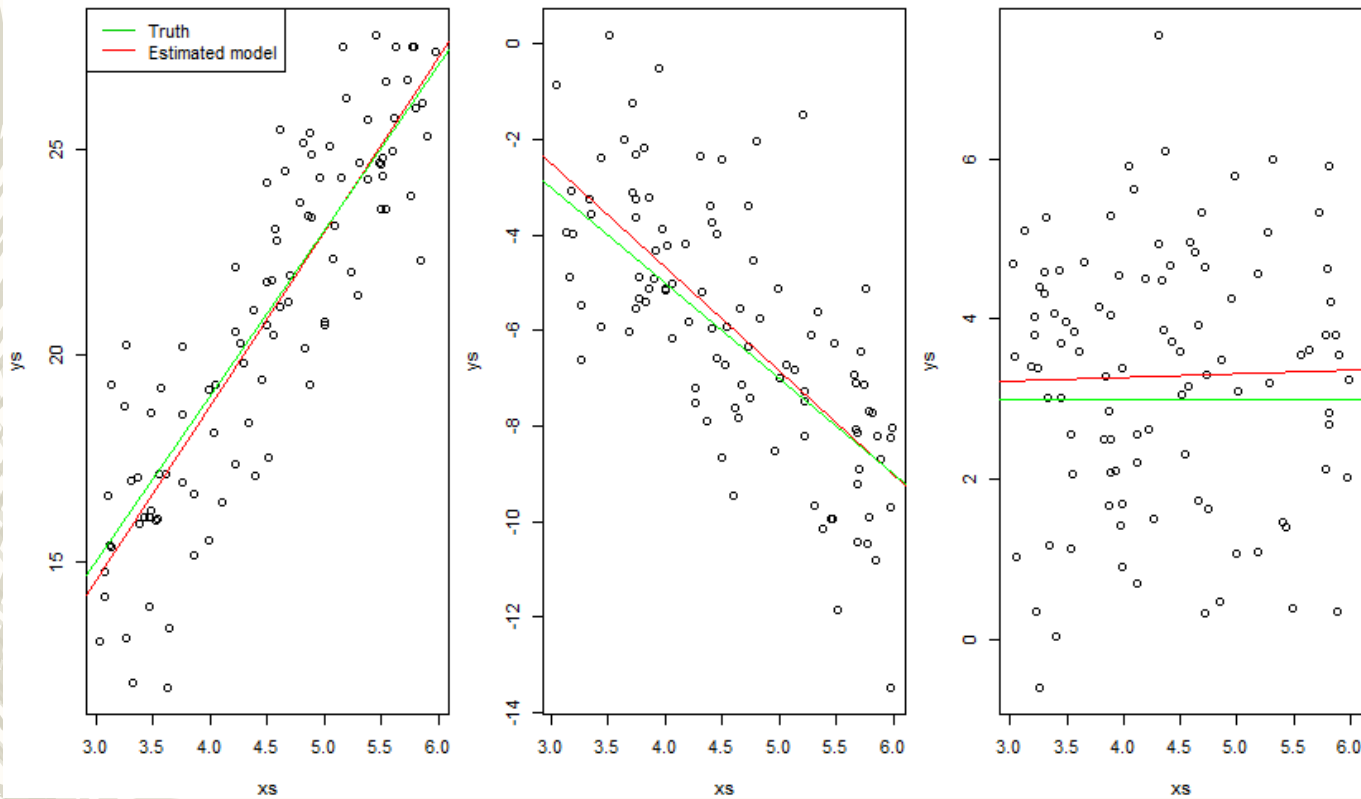
Estimativas dos coeficientes da regressão:  
(estimativas de mínimos quadrados)

$$\beta_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}$$



## Análise da regressão

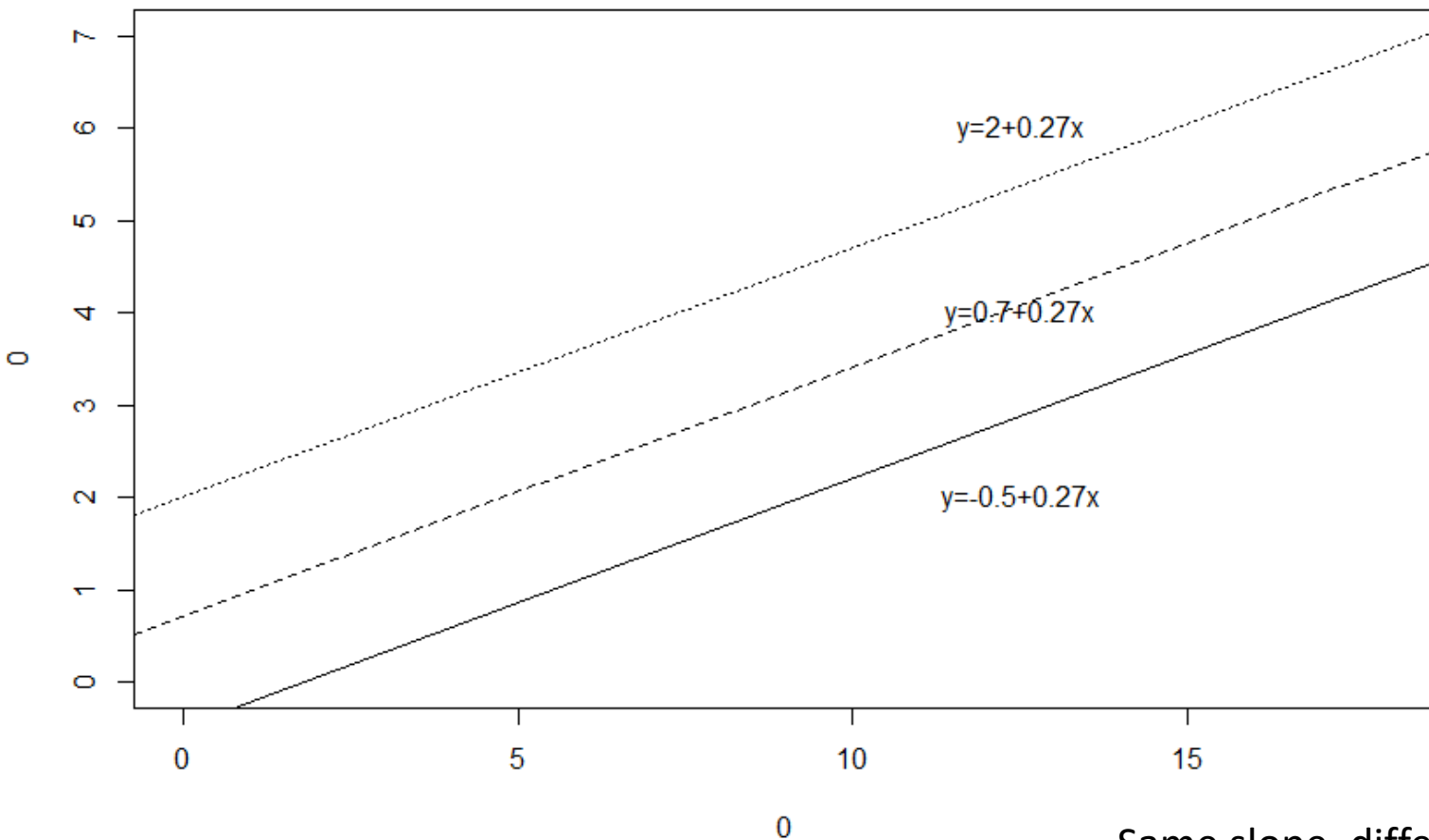


O declive (positivo, negativo ou nulo) determina o tipo de relação!





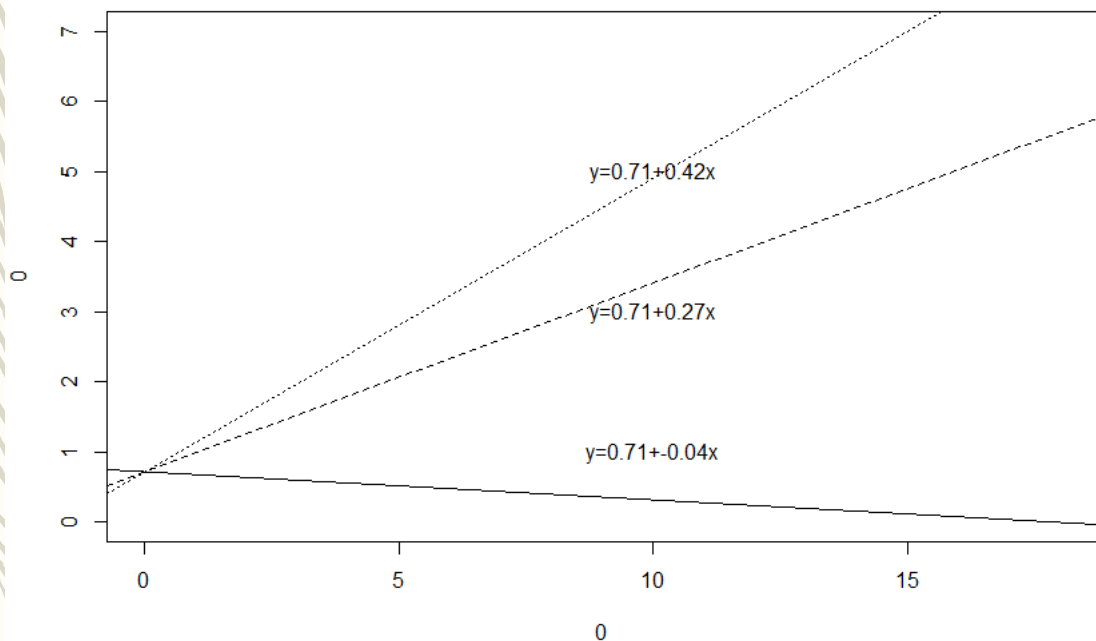
# Análise da regressão



Same slope, different intercepts



# Análise da regressão



Same intercept,  
different slopes



# Testes de hipóteses na análise da regressão

## *Testes F*

Hipóteses:

$$H_0: \beta_1 = 0$$

$H_1: \beta_1$  é diferente de 0

$$SQ_{TOTAL} = SQ_{REGRESSÃO} + SQ_{residual (erro)}$$



## Testes de hipóteses na análise da regressão

$$SQ_{TOTAL} = SQ_{REGRESSÃO} + SQ_{residual (erro)}$$

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2$$



## Testes de hipóteses na análise da regressão

Estatística de teste:

$$F = \frac{\frac{SQ_{REGRESSÃO}}{v_{REGRESSÃO}}}{\frac{SQ_{ERRO}}{v_{ERRO}}} = \frac{QM_{REGRESSÃO}}{QM_{ERRO}}$$

sendo  $gl_{regressão} = p - 1$  ( $p = n^{\circ}$  de coeficientes) e  
 $gl_{erro} = n - 2$  ( $n = n^{\circ}$  observações)

Valor crítico:

$$F_{\alpha, v_{reg}, v_{erro}}$$

Critério de decisão:

Rejeitar  $H_0$  se:

$$F > F_{\alpha, v_{reg}, v_{erro}}$$

Não rejeitar  $H_0$  caso contrário



## Variância explicada pelo modelo de regressão

*O coeficiente de determinação  $R^2$*

$$R^2 = \frac{SQ_{REGRESSÃO}}{SQ_{TOTAL}}$$

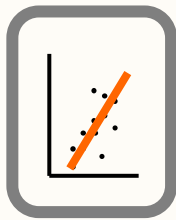
é a porção da variabilidade total explicada pelo modelo de regressão e é uma medida da adequabilidade da relação linear e de fitness do modelo.

$$0 \leq R^2 \leq 1$$

R é o coeficiente de correlação linear entre X e Y



# Regressão Multipla



## Análise da regressão múltipla

### *Âmbito:*

Avaliar a relação entre a variável dependente e múltiplas variáveis independentes

### *Objectivos:*

- Modelação dum evento ecológico
- Testes de hipóteses
- Predição (modelo preditivo)





## Análise da regressão múltipla

Quando existe mais do que uma variável independente

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

cada observação é dada por:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \varepsilon_i$$



## Análise da regressão múltipla

### *Variabilidade explicada pelo modelo*

O coeficiente de determinação  $R^2$  depende de  $p$  e  $n$  e, por isso, uma estimativa mais adequada é o  $R^2$  ajustado, dado por:

$$R_a^2 = 1 - \frac{MS_{RESIDUAL}}{MS_{TOTAL}} = 1 - \frac{n-1}{n-p-1} (1 - R^2)$$



## Testes de hipóteses na análise da regressão

### *Testes F*

Hipóteses:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$H_1$ : Pelo menos um dos  $\beta_i$  é diferente de 0