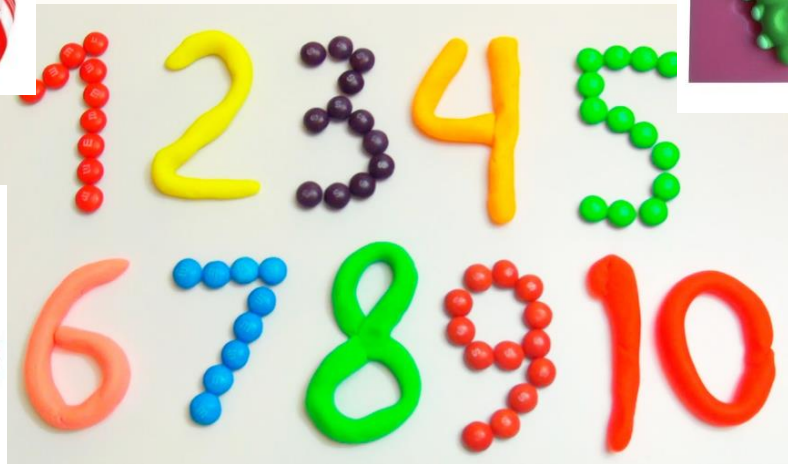



# Goodies\*




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


<https://twitter.com/BestleySophie/status/1184029125746384896?s=03>




- Home
- Explore
- Notifications
- Messages
- Bookmarks
- Lists
- Profile
- More

Tweet

 Tweet

**Sophie Bestley**  
@BestleySophie

Beautiful and compelling graphic for this

**Kevin Pluck** @kevpluck · Oct 12

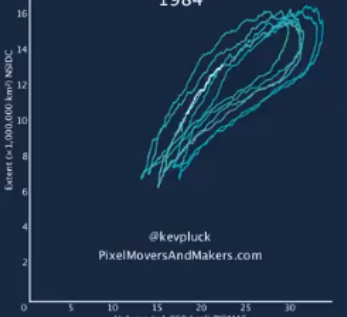
Arctic sea ice volume vs. extent 1979 - 2019.

PixelMoversAndMakers.com

#GlobalWarming #ClimateChange

Arctic Sea Ice Volume and Extent

1984








@kevpluck  
PixelMoversAndMakers.com

0:27 86K views


9:51 AM · Oct 15, 2019 · Twitter for iPhone


6 Retweets 8 Likes



Search Twitter

**Relevant people**

**Sophie Bestley**  
@BestleySophie  
@IMASUTAS @UTAS\_@mpred\_lab  
#womeninSTEMM #womeninpolarsci

**Kevin Pluck**  
@kevpluck  
#TeamMuskOx @TheHurricaneBot You can buy me a coffee! [Ko-fi.com/kevpluck](http://Ko-fi.com/kevpluck)

**Trends for you**

Trending in Portugal

**Trump**  
2.21M Tweets

Trending in Portugal

**CMTV**

Trending in Portugal

**Portugal**  
48.3K Tweets

Trending in Portugal

**Girl**  
1.06M Tweets

Trending in Portugal

**Sofia**  
42.7K Tweets

[Show more](#)

Terms Privacy policy Cookies Ads info

More © 2019 Twitter, Inc.

<https://pixelmoversandmakers.com/>

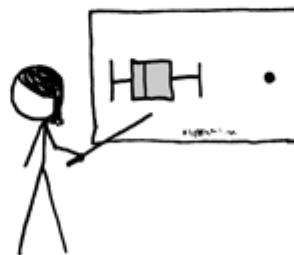
CAN MY BOYFRIEND  
COME ALONG?



I'M NOT YOUR  
BOYFRIEND!  
/ YOU TOTALLY ARE.  
I'M CASUALLY  
DATING A NUMBER  
OF PEOPLE.




BUT YOU SPEND TWICE AS MUCH  
TIME WITH ME AS WITH ANYONE  
ELSE. I'M A CLEAR OUTLIER.



YOUR MATH IS  
IRREFUTABLE.

FACE IT—I'M  
YOUR STATISTICALLY  
SIGNIFICANT OTHER.





DON'T BECOME A  
STATISTIC, RALPH.  
YOU KNOW HUMANS  
EAT EIGHT OF US IN  
THEIR SLEEP EVERY  
YEAR.

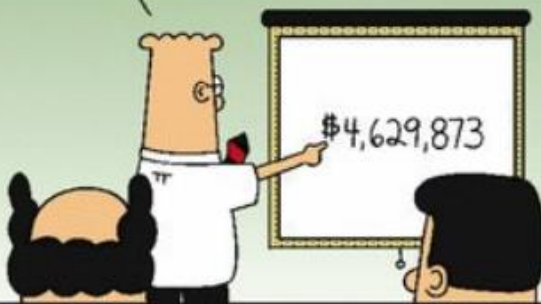


CARTOONSTOCK  
.com

Search ID: nfkn2188

W. FAKES

I DIDN'T HAVE ANY  
ACCURATE NUMBERS  
SO I JUST MADE UP  
THIS ONE.



www.dilbert.com scottadams@aol.com

STUDIES HAVE SHOWN  
THAT ACCURATE  
NUMBERS AREN'T ANY  
MORE USEFUL THAN THE  
ONES YOU MAKE UP.



5808 ©2008 Scott Adams, Inc./Dist. by UFS, Inc.

HOW  
MANY  
STUDIES  
SHOWED  
THAT?

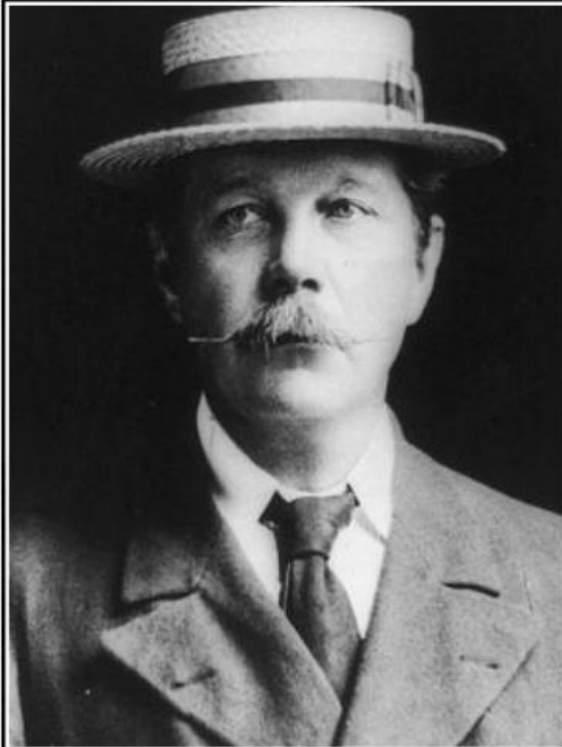


EIGHTY-  
SEVEN.





Ecología Numérica - Aula Teórica 12 – 23-10-2018



It is a capital mistake to theorize  
before one has data. Insensibly one  
begins to twist facts to suit theories,  
instead of theories to suit facts.

— *Arthur Conan Doyle* —

AZ QUOTES

<https://www.azquotes.com/quote/367106>

Esta semana não há aula TP na quarta (Ciências Research Day) e na sexta (Feriado). A vossa missão é realizar a FT5 (já no FENIX), e a presença em aula será substituída pela entrega por email de um relatório dinâmico, com o título FT5?????, onde o ????? será o vosso número de aluno. Data limite de entrega: 03 de Novembro de 2019.



## Ecologia Numérica

Componente Teórica - Prática

Ficha de trabalho **5**

1. Efectue um **teste t** para avaliar se o comprimento médio dos gastrópodes é igual a 27 mm, partindo do princípio de que são satisfeitos os pressupostos exigidos para realizar esta análise (DataTP5gastropodes.csv).

2. Efectue o **teste que lhe pareça adequado** para avaliar se a densidade populacional em Concelhos de determinado Distrito é igual a 60 habitantes por km<sup>2</sup> (DataTP5habitantes.csv).

3. Efectue um teste que lhe permita averiguar se as médias das duas populações a que se referem as amostras são iguais (DataTP5concelhos.csv).

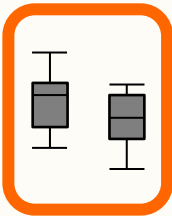
4. Efectue o **teste de Mann-Whitney** para avaliar se o comprimento total dos machos é igual ao das fêmeas (DataTP5comprimento.csv).

5. Efectue o **teste t para amostras emparelhadas** para avaliar se os comprimentos da barbatana peitoral direita são iguais ao da esquerda (DataTP5barbatanas.csv).

6. Admita que os pressupostos de normalidade e homocedasticidade não são cumpridos no caso anterior. Efectue um **procedimento alternativo** que lhe permita realizar um teste para avaliar a mesma hipótese.

7. Efectue o teste que julgar adequado aos dados DataTP5densidades.csv para avaliar se as amostras em causa são provenientes da mesma população. Explore os resultados obtidos e apresente-os tal como se pretendesse incluí-los numa apresentação ou publicação científica.

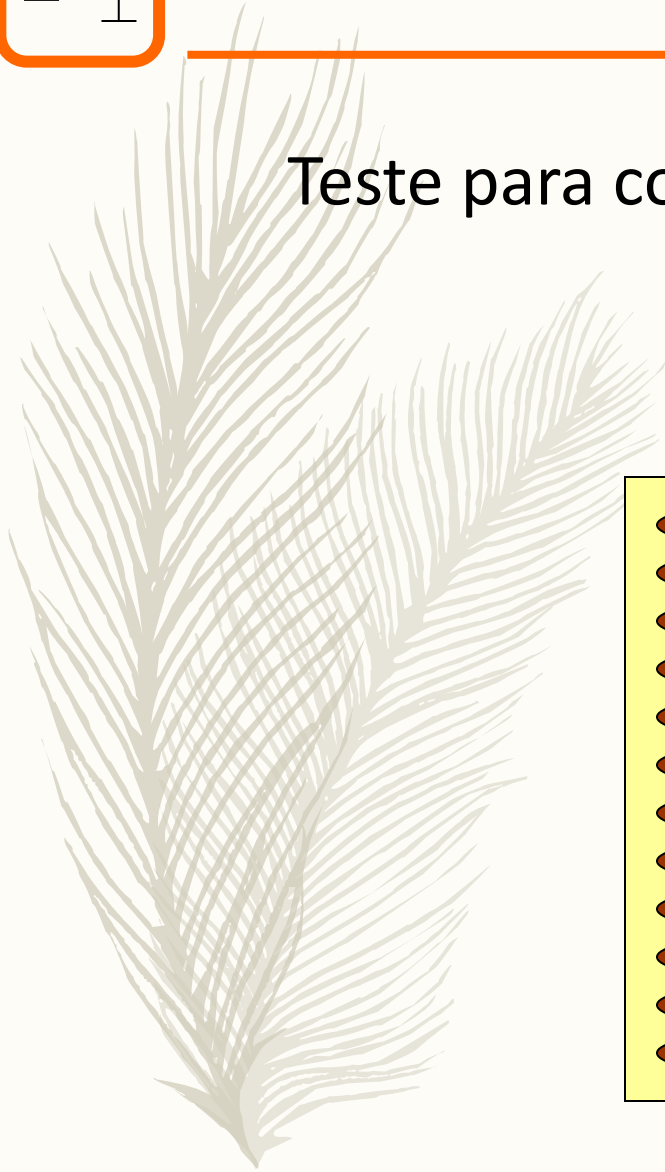
A TP de hoje é de entrada livre. Na quarta e quinta feira estarei disponível para atendimentos por marcação (e-mail). Se for necessário arranharei uma sala para podermos ter aula(s) extra.



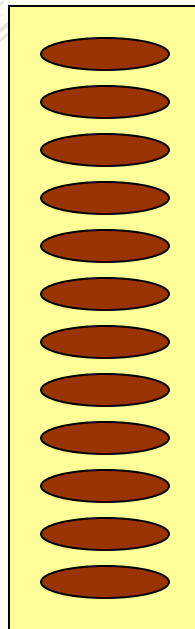
## testes de hipóteses a duas amostras

---

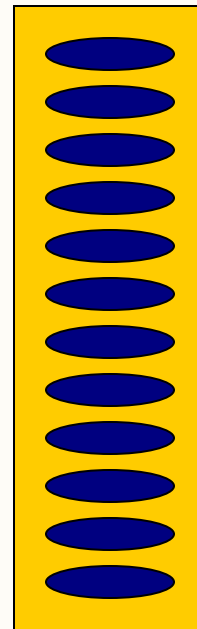
### Teste para comparação de duas amostras



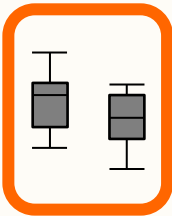
A



B







## testes de hipóteses a duas amostras

### Teste t para 2 amostras

Hipóteses:

$$H_0: \mu_A = \mu_B$$

$$H_1: \mu_A \neq \mu_B$$

Estatística de teste:

$$t = \frac{\bar{X}_A - \bar{X}_B}{s_{\bar{X}_A - \bar{X}_B}}$$

Valor crítico:

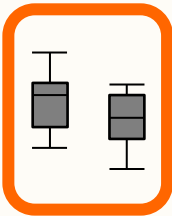
$$t_{\alpha(2), \nu}$$

Critério de decisão:

Rejeitar  $H_0$  se:

$$|t| > t_{\alpha(2), \nu}$$

Não rejeitar  $H_0$ , caso contrário



## testes de hipóteses a duas amostras

### Teste t para 2 amostras

Estatística de teste:

$$t = \frac{\bar{X}_A - \bar{X}_B}{S_{\bar{X}_A - \bar{X}_B}}$$

onde

$$S_{\bar{X}_A - \bar{X}_B} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}},$$


$$s_p^2 = \frac{SS_1 + SS_2}{\nu_1 + \nu_2},$$

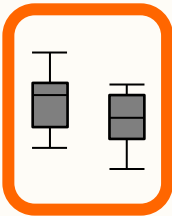
$$\nu_1 = n_1 - 1$$

e

$$\nu_2 = n_2 - 1$$

$$\nu = \nu_1 + \nu_2 = n_1 + n_2 - 2$$


$$\text{SS} = \text{sum of squares} = \sum_{i=1}^n (x_i - \bar{x})^2$$



## testes de hipóteses a duas amostras

### Teste t para 2 amostras

Hipóteses:

$$H_0: \mu_A \leq \mu_B$$

$$H_1: \mu_A > \mu_B$$

$$H_0: \mu_A \geq \mu_B$$

$$H_1: \mu_A < \mu_B$$

Crítérios de decisão:

Rejeitar  $H_0$  se:

$$t > t_{\alpha(1), \nu}$$

Não rejeitar  $H_0$  caso contrário

Rejeitar  $H_0$  se:

$$t < -t_{\alpha(1), \nu}$$

Não rejeitar  $H_0$  caso contrário

```
> set.seed(1234)
> t.test(rnorm(50), rnorm(50))
```

Welch Two Sample t-test

```
data: rnorm(50) and rnorm(50)
t = -3.0732, df = 95.633, p-value = 0.00276
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.9753573 -0.2098077
sample estimates:
mean of x mean of y
-0.4530530 0.1395295
```

Very unexpected !!  
A type I error 😊

```
> t.test(rnorm(50), rnorm(50))
```

Welch Two Sample t-test

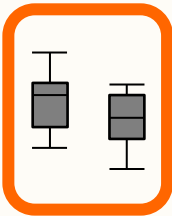
```
data: rnorm(50) and rnorm(50)
t = -0.20502, df = 89.415, p-value = 0.838
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.4546898 0.3696294
sample estimates:
mean of x mean of y
0.01997809 0.06250827
```

```
> t.test(rnorm(50), rnorm(50, mean = 2))
```

Welch Two Sample t-test

```
data: rnorm(50) and rnorm(50, mean = 2)
t = -9.7984, df = 97.796, p-value = 3.381e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.270291 -1.505553
```





## testes de hipóteses a 2 amostras

### Teste sobre diferença no valor médio entre duas amostras

Avaliar se os pressupostos são cumpridos

Não

Transformação  
dos dados

Não

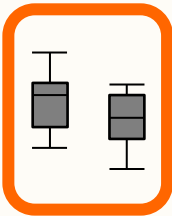
Sim

Testes não paramétricos

Teste de Wilcoxon-Mann-Whitney

Testes paramétricos

Teste de t-student para 2 amostras



## testes de hipóteses a duas amostras

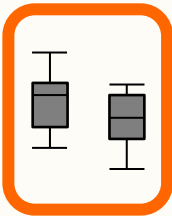
---

### Teste $t$ para 2 amostras

#### Pressupostos do teste $t$ :

- 1 - Resíduos provenientes duma população normal (**importante**: são os resíduos que tem de ser normais, não os dados – no caso de um teste de  $t$ , é quase equivalente, mas não é a mesma coisa!)
- 2 - Variâncias das populações das duas amostras homogéneas

No entanto, o teste  $t$  é bastante robusto, i.e. a sua validade não é grandemente afectada por desvios moderados dos pressupostos.



## testes de hipóteses a duas amostras

*Teste para comparação de duas amostras: abordagem não-paramétrica*

### Teste de Wilcoxon-Mann-Whitney

Hipóteses:

$H_0$ : amostras provenientes da mesma população

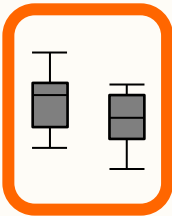
$H_1$ : amostras não provenientes da mesma população

Estatística de teste:

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$D_i^* = X_i - Y_i \quad R_i = \text{ordem atribuída a } D_i \text{ (} D_i = |D_i^*| \text{)}$$

Dado que a atribuição das ordens de forma crescente ou decrescente é arbitrária, temos que calcular  $U'$



## testes de hipóteses a duas amostras

*Teste para comparação de duas amostras: abordagem não-paramétrica*

### Teste de Wilcoxon-Mann-Whitney

Hipóteses:

$H_0$ : amostras provenientes da mesma população

$H_1$ : amostras não provenientes da mesma população

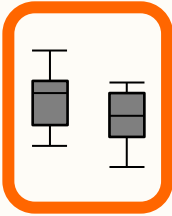
Estatística de teste:

$$U = U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

Soma das ordens da amostra 1 na amostra conjunta ordenada

$$R_1 = \sum_{i=1}^{n_1} R_{i1}$$

Dado que calcular a soma das ordens da amostra 1 ou 2 é uma decisão arbitrária, temos que calcular também o  $U' = U_2 = n_1 n_2 - U_1$



## testes de hipóteses a duas amostras

# Teste de Wilcoxon-Mann-Whitney

*Hipóteses:*

$H_0$ : amostras provenientes da mesma população

$H_1$ : amostras não provenientes da mesma população

Estatística de teste:

$$U = U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$U' = U_2 = n_2 n_1 + \frac{n_2(n_2 + 1)}{2} - R_2 = n_1 n_2 - U$$

$$R_2 = \sum_{i=1}^{n_2} R_{2i}$$



Procedimento:

Temos duas amostras, uma de  $X$ 's e outra de  $Y$ 's

Obtemos a amostra conjunta  $XY=c(X,Y)$

Ordenar os valores da amostra conjunta  $XY$

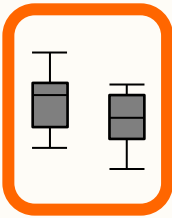
$R_1$  = soma dos ranks da amostra 1

$R_2$  = soma dos ranks da amostra 2

Calcular  $U_1$  e  $U_2$

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

Tirar conclusões em função da significância escolhida e do P-value calculado com base na amostra



testes de hipóteses a duas amostras

---

## Teste de Wilcoxon-Mann-Whitney

Valor crítico:

$$U_{\alpha(2), n_1, n_2}$$

Critério de decisão:

Rejeitar  $H_0$  se:

$$U \vee U' > U_{\alpha(2), n_1, n_2}$$

Não rejeitar  $H_0$  caso contrário

```

> amostra1=c(2.1,3.6,4.3,1.5,6.3,7.4)
> n1=length(amostra1)
> amostra2=c(3.2,4.2,1.3,2.2,0.9,5.1,2.8)
> n2=length(amostra2)
> dados=data.frame(valores=c(amostra1,amostra2),amostra=rep(1:2,times=c(n1,n2)))
> dados$rank=rank(dados$valores)
> dados
  valores amostra rank
1      2.1       1    4
2      3.6       1    8
3      4.3       1   10
4      1.5       1    3
5      6.3       1   12
6      7.4       1   13
7      3.2       2    7
8      4.2       2    9
9      1.3       2    2
10     2.2       2    5
11     0.9       2    1
12     5.1       2   11
13     2.8       2    6

```

	valores	amostra	rank
1	2.1	1	4
2	3.6	1	8
3	4.3	1	10
4	1.5	1	3
5	6.3	1	12
6	7.4	1	13
7	3.2	2	7
8	4.2	2	9
9	1.3	2	2
10	2.2	2	5
11	0.9	2	1
12	5.1	2	11
13	2.8	2	6

```

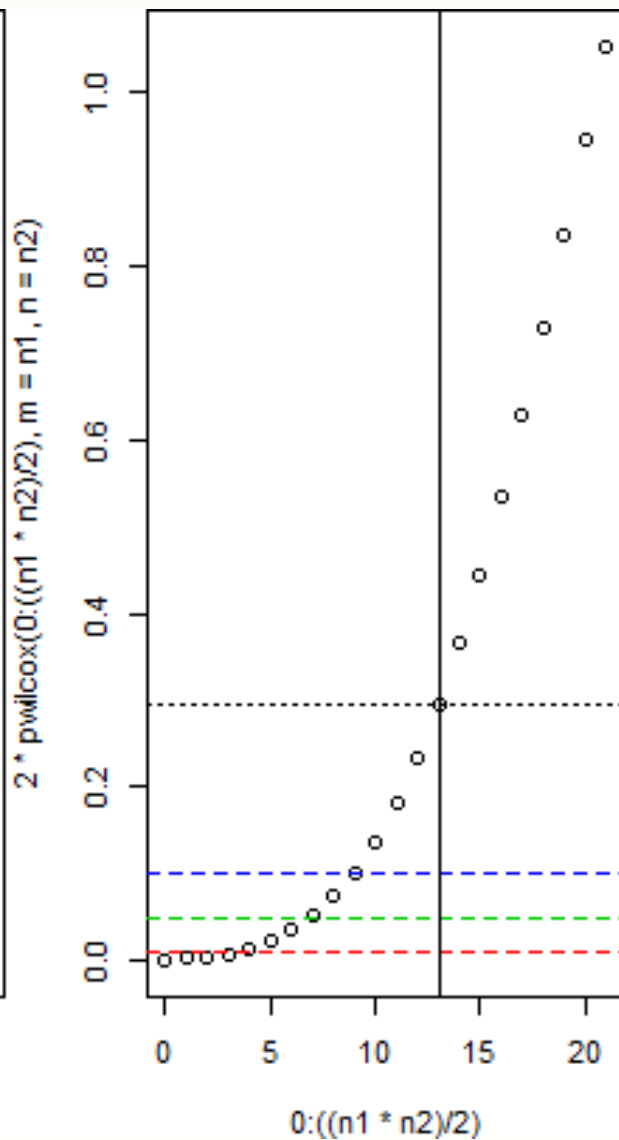
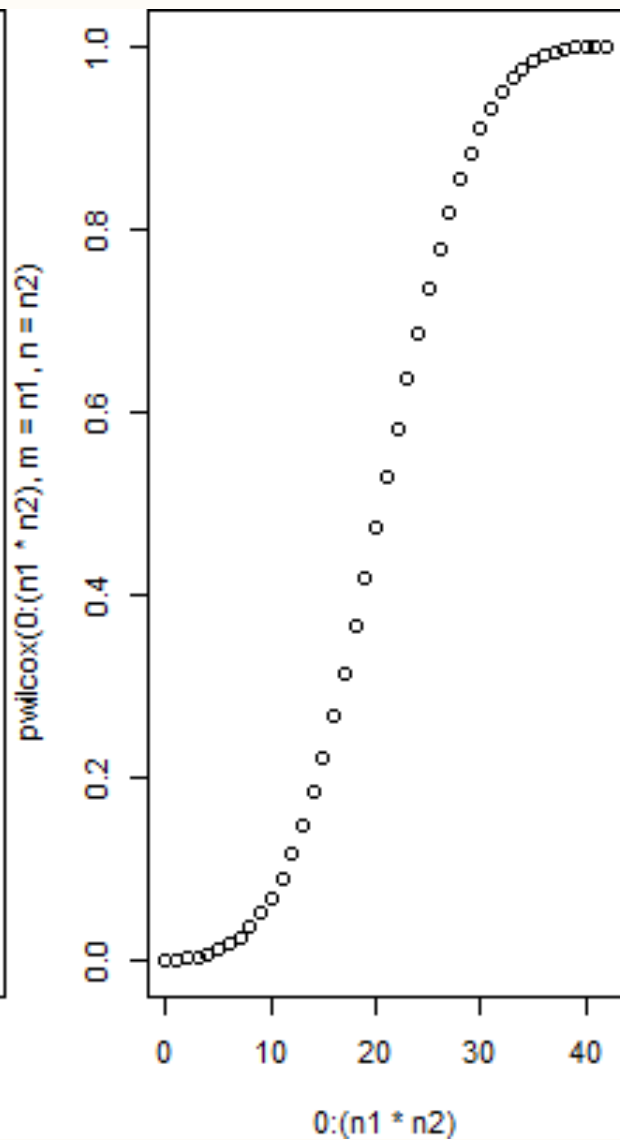
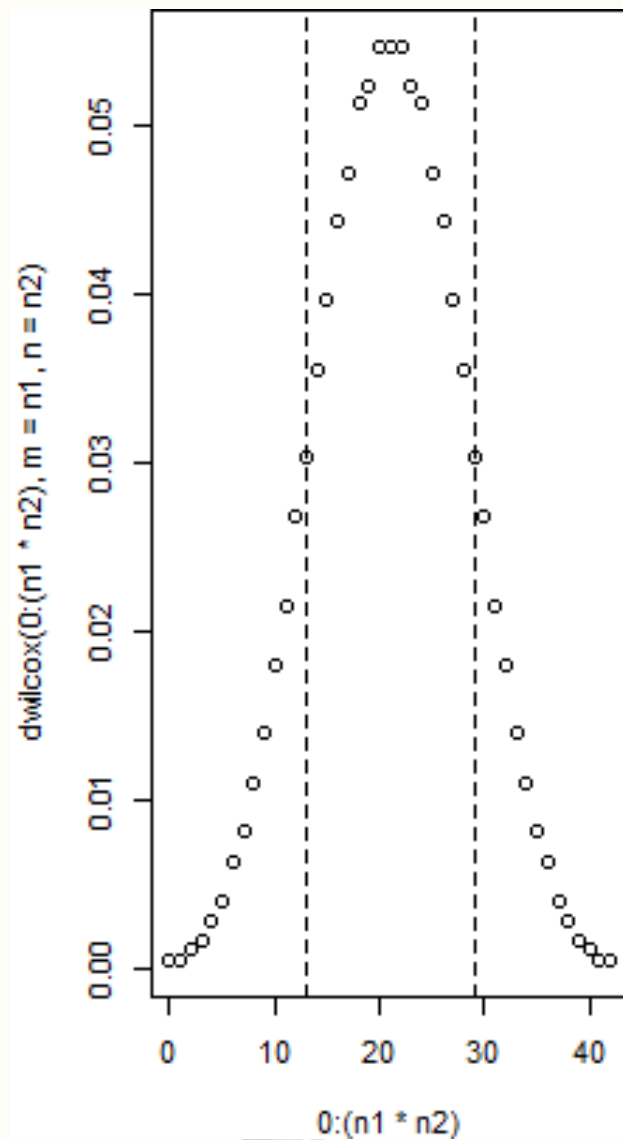
> #get test statistic
> U=n1*n2+(n1*(n1+1))/2-sum(dados$rank[dados$amostra==1])
> Ulinha=n1*n2-U
> U;Ulinha
[1] 13
[1] 29
> #test
> wilcox.test(amostra1,amostra2)

      wilcoxon rank sum test

data:  amostra1 and amostra2
W = 29, p-value = 0.2949
alternative hypothesis: true location shift is not equal to 0

> #p-value by hand
> #2 vezes a probabilidade de ser menor ou igual a U
> 2*pwilcox(U,m=n1, n=n2)
[1] 0.2948718
> #2 vezes a probabilidade de ser maior ou igual a Ulinha
> #que é o mesmo que 2 vezes a P de ser maior que U linha-1
> 2*pwilcox(Ulinha-1,m=n1, n=n2,lower.tail = FALSE)
[1] 0.2948718

```



```
> set.seed(1234)
> x=rnorm(50)
> y=rnorm(50)
> wilcox.test(x,y,exact=TRUE)
```

wilcoxon rank sum test

data: x and y

w = 819, p-value = 0.002743

alternative hypothesis: true location shift is not equal to 0

```
> 2*pnorm(819,50,50)
[1] 0.002743261
```

```
> set.seed(1234)
> x=rnorm(50)
> y=rnorm(50,mean=1)
> wilcox.test(x,y,exact=TRUE)
```

wilcoxon rank sum test

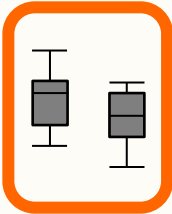
data: x and y

w = 276, p-value = 2.733e-13

alternative hypothesis: true location shift is not equal to 0

```
> 2*pnorm(276,50,50)
[1] 2.732594e-13
```

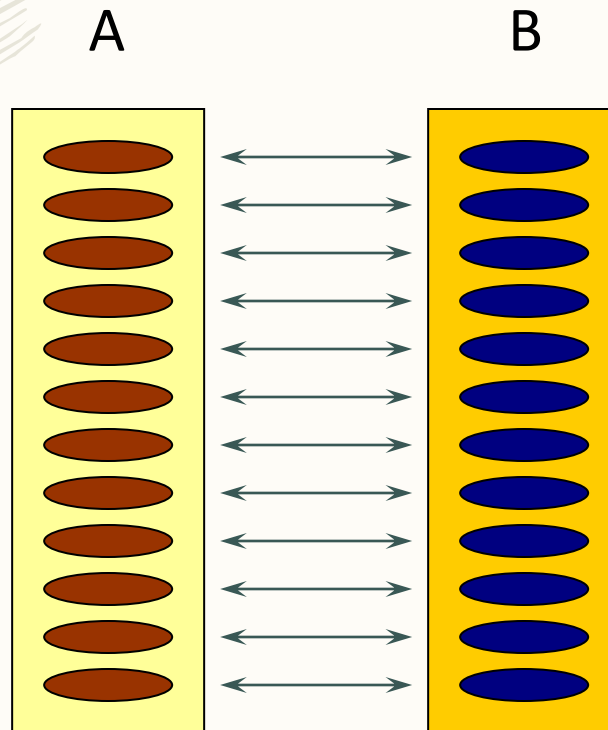


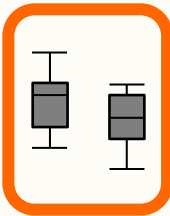


## testes de hipóteses a duas amostras emparelhadas

### Testes para amostras emparelhadas

e.g. comparação de observações não independentes (ou seja, quando há uma estrutura de correlação) – exemplos: 2 folhas em cada árvore, mão esquerda e mão direita, resultados em gémeos, etc...





## testes de hipóteses a duas amostras emparelhadas

### Teste t para amostras emparelhadas

Hipóteses:

$$H_0: \mu_d = 0$$

$$H_1: \mu_d \neq 0$$

$$d_i = X_{1i} - X_{2i}$$

Valor crítico:

$$t_{\alpha(2), \nu}$$

Estatística de teste:

$$t = \frac{\bar{d}}{s_{\bar{d}}}$$

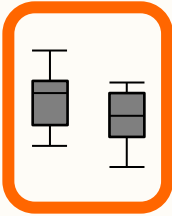
$\nearrow \frac{s_d}{\sqrt{n}}$

Critério de decisão:

Rejeitar  $H_0$  se:

$$|t| > t_{\alpha(2), \nu}$$

Não rejeitar  $H_0$ , caso contrário



## testes de hipóteses a duas amostras emparelhadas

### Teste de Wilcoxon para amostras emparelhadas

Hipóteses:

$$H_0: d=0$$

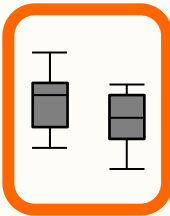
$$H_1: d \neq 0$$

Estatística de teste:

$$T^+ = \sum (R_i \text{ para os quais } D_i^* \text{ são positivos})$$

$$D_i^* = X_i - Y_i \quad R_i = \text{ordem atribuída a } D_i \text{ (} D_i = |D_i^*| \text{)}$$

$$T^+ = \frac{n(n+1)}{2} - T^-$$



## testes de hipóteses a duas amostras emparelhadas

---

# Teste de Wilcoxon para amostras emparelhadas

Hipóteses:

$$H_0: d=0$$

$$H_1: d \neq 0$$

Critério de decisão:

Rejeitar  $H_0$  se:

$$T^+ \vee T^- < T_{\alpha(2),n}$$

Não rejeitar  $H_0$  caso contrário



Xi	Yi	Di*	Sign di
3.2	1.3	2.9	+
3.6	2.8	0.9	+
2.4	3.2	-0.8	-
2.1	1.9	0.2	+
4.2	3.7	0.5	+
5.4	6.1	-0.7	-
4.7	4.6	0.1	+
3.2	3.5	-0.3	-

$$T+ = 8+7+2+4+1=22$$

di	ri
2.9	8
0.9	7
0.8	6
0.2	2
0.5	4
0.7	5
0.1	1
0.3	3

$$T- = 6+5+3=14$$



```
> X=c(3.2,3.7,2.4,2.1,4.2,5.4,4.7,3.2)
> Y=c(0.3,2.8,3.2,1.9,3.7,6.1,4.6,3.5)
> Di=X-Y
> signDi=sign(Di)
> ranks=rank(abs(Di))
> Tplus=sum(ranks[signDi>0])
> Tminus=sum(ranks[signDi<0])
> Tplus;Tminus
[1] 22
[1] 14
> wilcox.test(x=X,y=Y,paired=TRUE)
```

wilcoxon signed rank test

data: X and Y  
V = 22, p-value = 0.6406  
alternative hypothesis: true location shift is not equal to 0

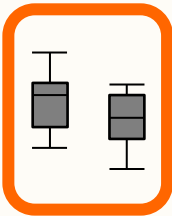
```
> 2*psignrank(q=14, n=8)
[1] 0.640625
```

```
> wilcox.test(rnorm(50),rnorm(50,1),paired=TRUE)
```

wilcoxon signed rank test with continuity correction

data: rnorm(50) and rnorm(50, 1)  
V = 165, p-value = 5.205e-06  
alternative hypothesis: true location shift is not equal to 0

**Desafio:** gerar uma função que mostra os 3 plots usuais para ver o resultado do teste e decidir se rejeitamos ou não  $H_0$



# testes de hipóteses a uma ou duas amostras (emparelhadas ou não)

## Síntese

1 amostra

Teste  $t$

Paramétrico (P)

Teste Wilcoxon

Não P (NP)

2 amostras:

Teste  $t$

P

Teste Mann-Whitney

NP

2 amostras  
emparelhadas:

Teste  $t$

para amostras emparelhadas

P

Teste Wilcoxon

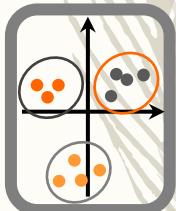
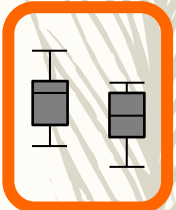
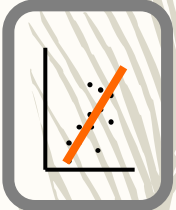
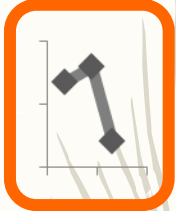
para amostras emparelhadas

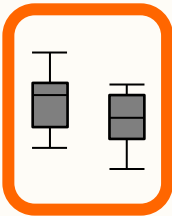
NP

# Ecologia Numérica

---

testes a mais de duas amostras  
análise de variância e  
equivalente não paramétrico

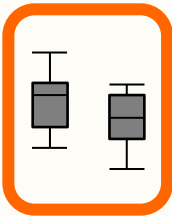




## anova e equivalente não paramétrico

---

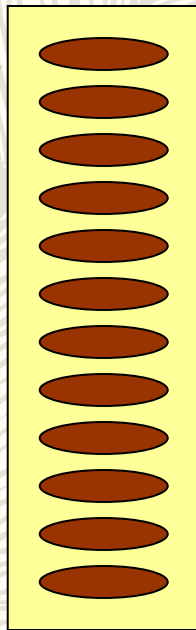
- Quais os testes mais correntes para situações de 3 ou mais amostras?
- Quais as condições para a sua aplicação?
- Como interpretar os seus resultados?



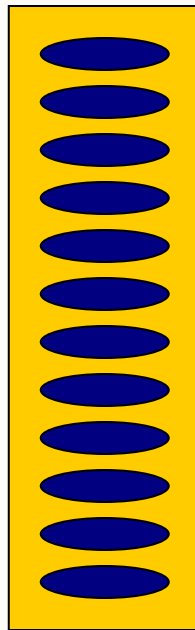
anova e equivalente não paramétrico

## Testes a mais de duas amostras

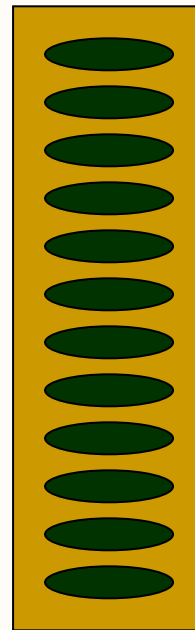
A



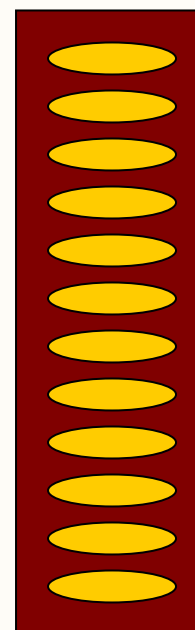
B

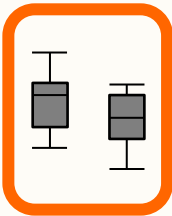


C



D





anova e equivalente não paramétrico

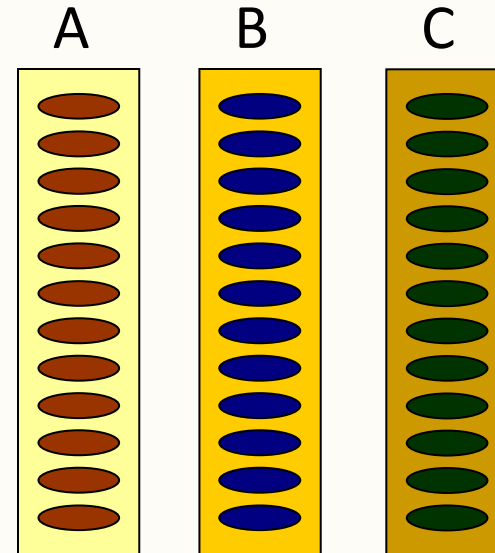
## Testes a mais de duas amostras

Realizar todos os testes de pares de amostras, duas a duas:

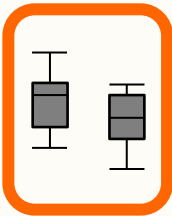
A vs B

B vs C

A vs C



Mas isso não é uma boa ideia...Porque não?



anova e equivalente não paramétrico

## Testes a mais de duas amostras

Testes a duas amostras:

A vs B

B vs C

A vs C

Porque não?

Para o conjunto das 3 hipóteses a probabilidade de correctamente não rejeitar todas as  $H_0$  seria  
 $0.95^3 = 0.86$

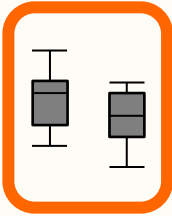
ou seja,  $\alpha = 0.14$  !!!!

Hipótese nula

Hipótese nula

	Hipótese nula	
	Não Rejeitar	Rejeitar
Verdadeira	Não há erro	$\alpha$
Falsa	$\beta$	Não há erro





anova e equivalente não paramétrico

---

## A análise de variância (ANOVA)

$$H_0: \mu_A = \mu_B = \mu_C$$

$H_1$ : Há pelo menos uma média diferente das outras = as médias não são todas iguais

*Fontes de variação:*

Intra-amostra ou intra-grupos (erro)

Entre-amostras ou entre-grupos