

# Comparações entre duas condições dependentes e independentes

*Paulo S. P. Silveira (paulo.silveira@fm.usp.br)*  
*Koichi Sameshima (koichi.sameshima@fm.usp.br)*  
*José O. Siqueira (jose.siqueira@fm.usp.br)*

## Contents

<b>Objetivos</b>	<b>2</b>
<b>Preparação</b>	<b>2</b>
<b>Distribuição <math>t</math></b>	<b>2</b>
Funções para distribuições em R . . . . .	3
distribuição binomial . . . . .	3
distribuição normal . . . . .	5
a distribuição $t$ . . . . .	8
<b>Raciocínio inferencial</b>	<b>10</b>
<b>Métodos Robustos</b>	<b>12</b>
<b>Teste <math>t</math> para uma condição</b>	<b>12</b>
- situação . . . . .	12
- hipóteses (planejamento) . . . . .	13
- coleta dos dados . . . . .	13
- estatística descritiva . . . . .	13
- estatística inferencial . . . . .	17
<b>Reverendo o raciocínio</b>	<b>23</b>
<b><math>t</math> pareado (duas condições dependentes)</b>	<b>23</b>
-situação . . . . .	23
- planejamento . . . . .	24
- coleta dos dados . . . . .	24
- estatística descritiva . . . . .	25
- estatística inferencial . . . . .	29
<b>teste <math>t</math> para duas condições independentes (teste <math>t</math> de Welch)</b>	<b>31</b>
- situação . . . . .	31
- planejamento . . . . .	32
- coleta dos dados . . . . .	32
- estatística descritiva . . . . .	32
- estatística inferencial . . . . .	37
significância estatística . . . . .	38
significância prática . . . . .	43
<b>Conceitos adicionais</b>	<b>44</b>
Teste $t$ sem os dados brutos . . . . .	44
teste $t$ com <i>bootstrapping</i> e tamanho de efeito . . . . .	44
algumas manobras úteis . . . . .	45

construção de dois <i>boxplots</i> , lado a lado . . . . .	45
guardar o gráfico para um arquivo . . . . .	48
guardar a saída textual em um arquivo . . . . .	48
intervalo de confiança robusto . . . . .	49
<b>Sobre os métodos tradicionais</b>	<b>50</b>
- o teste $t$ de Student . . . . .	50
- o (não) uso de testes não-paramétricos . . . . .	51

v20200327.1510

## Objetivos

- reconhecer e mencionar propriedades da distribuição  $t$ .
- reconhecer as indicações e aplicar um teste  $t$  para uma condição.
- reconhecer as indicações e aplicar um teste  $t$  pareado (condições dependentes).
- reconhecer as indicações e aplicar um teste  $t$  independente (condições independentes).
- definir hipóteses estatísticas nula e alternativa.

## Preparação

Os exemplos aqui apresentados estão disponíveis. Caso queira usá-los, crie um projeto, coloque o arquivo desta aula e os seguintes arquivos na pasta do mesmo:

- Animacao\_t\_central.R
- Animacao\_t\_nao\_central.R
- Nifedipina.R
- Violencia\_estadios.R
- Violencia\_estadios.xlsx
- Nutricao.R
- Nutricao.xlsx

## Distribuição $t$

É uma distribuição de probabilidades que considera graus de liberdade ( $\nu$ , letra grega *ni*).

Sob  $H_0$  é semelhante à distribuição normal padronizada, centrada em  $t = 0$ , mas com suas caudas mais “pesadas” (desvio-padrão  $> 1$ ).

Os graus de liberdade dependem do tamanho da amostra; quanto menor, mais pesadas são as caudas e, portanto, diferenças teriam que ser maiores para rejeitarmos a hipótese nula.



Experimente `Animacao_t_central.R` para ver o aspecto da distribuição  $t$  e observe:

- sob  $H_0$  a distribuição  $t$  é centrada em zero.
- quando as caudas têm maior área, então o valor crítico, que define  $\alpha$ , afasta-se.
- aproxima-se da distribuição normal se  $\nu \rightarrow \infty$ .

---

## Funções para distribuições em R

R dispõe de uma pequena família de funções básicas para cada tipo de distribuição. Estes pequenos conjuntos são análogos uns aos outros conjuntos, facilitando o aprendizado.

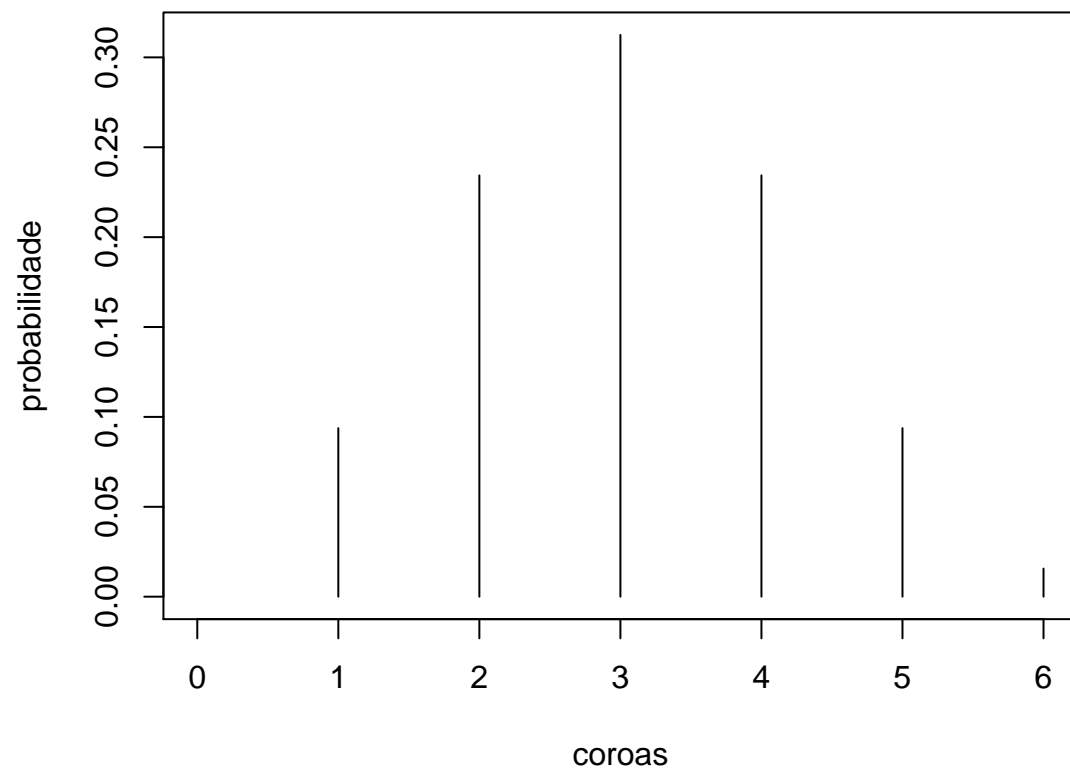
### distribuição binomial

A família de funções para a distribuição binomial é:

- `dbinom(x, size, prob, log = FALSE)`
- `pbinom(q, size, prob, lower.tail = TRUE, log.p = FALSE)`
- `qbinom(p, size, prob, lower.tail = TRUE, log.p = FALSE)`
- `rbinom(n, size, prob)`

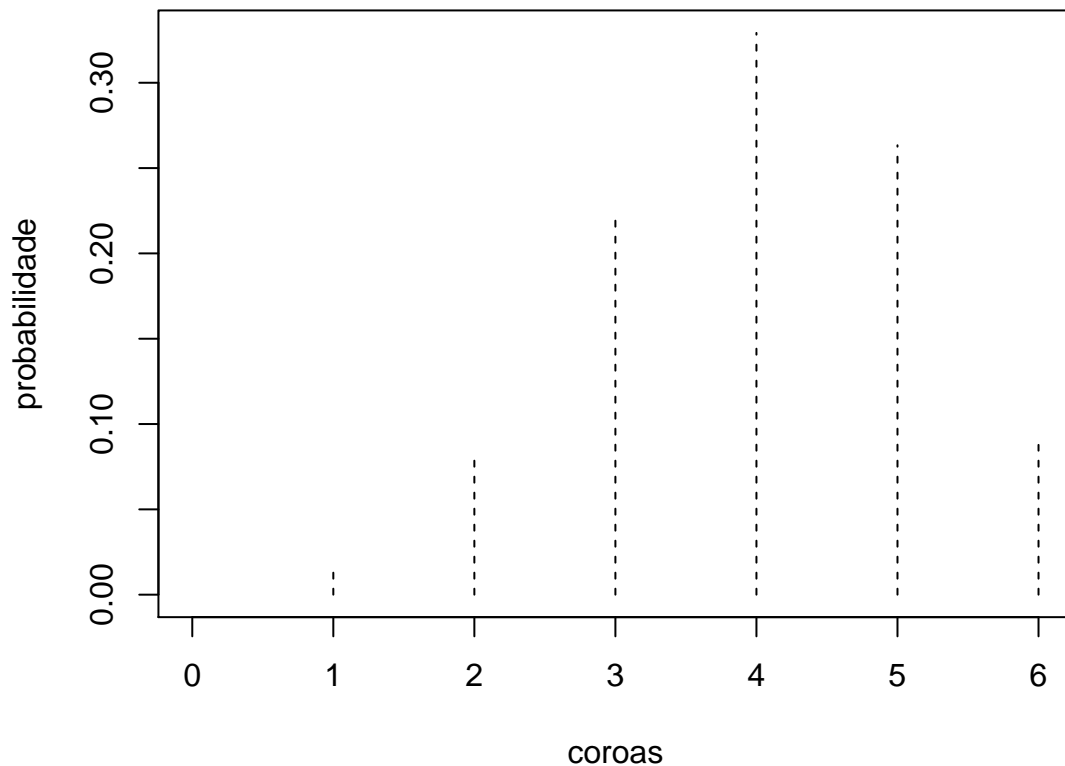
que dependem do número de eventos (*size*) e sucesso de cada evento (*prob*). Esta distribuição serve, por exemplo, para modelar uma moeda. Por exemplo, uma moeda bem balanceada tem 50% de probabilidade de sair coroa. Caso fosse testada com 6 jogadas, esperamos que 3 coroas sejam o mais provável, e o gráfico com a distribuição de probabilidades pode usar o seguinte código:

```
jogadas <- seq(0:6)
probabilidades <- dbinom(x=jogadas, size=6, prob=0.5)
plot(jogadas, probabilidades,
     xlab="coroas", ylab="probabilidade",
     xlim=c(0,6), type="h")
```



Para uma moeda desbalanceada, viciada para dar  $\frac{2}{3}$  de coroas, esperamos 4 coroas em 6 jogadas, como vemos alterando o parâmetro *prob*:

```
jogadas <- seq(0:6)
probabilidades <- dbinom(x=jogadas, size=6, prob=2/3)
plot(jogadas, probabilidades,
     xlab="coroas", ylab="probabilidade",
     xlim=c(0,6), type="h", lty=2)
```



A curva produzida para a moeda desbalanceada é a mesma daquela obtida pela moeda balanceada, apenas transladada para a direita. A dificuldade para distinguir uma moeda da outra está em observar que a moeda de 50% tem certa probabilidade de sair com 4 coroas, bem como a desbalanceada pode obter 3.

### distribuição normal

A distribuição normal tem conjunto de funções similar às da binomial:

- `dnorm(x, mean = 0, sd = 1, log = FALSE)`
- `pnorm(q, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)`
- `qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)`
- `rnorm(n, mean = 0, sd = 1)`

na qual necessitamos da média (*mean*) e desvio-padrão (*sd*) para sua caracterização. É usada com variáveis quantitativas contínuas, portanto os gráficos devem usar linhas contínuas.

Por comparação exploraremos distribuições normais com médias de 3 e 4 e desvio-padrão de 1.22.



Este desvio-padrão não foi escolhido ao acaso. O desvio-padrão de uma binomial é dado por:

$$sd_{binomial} = \sqrt{n \cdot p \cdot (1 - p)}$$

onde  $n$  é o número de jogadas,  $p$  é a probabilidade do evento. Então, para o exemplo acima,  $sd_{binomial} = \sqrt{6 \cdot 0.5 \cdot (1 - 0.5)} \approx 1.22$ .

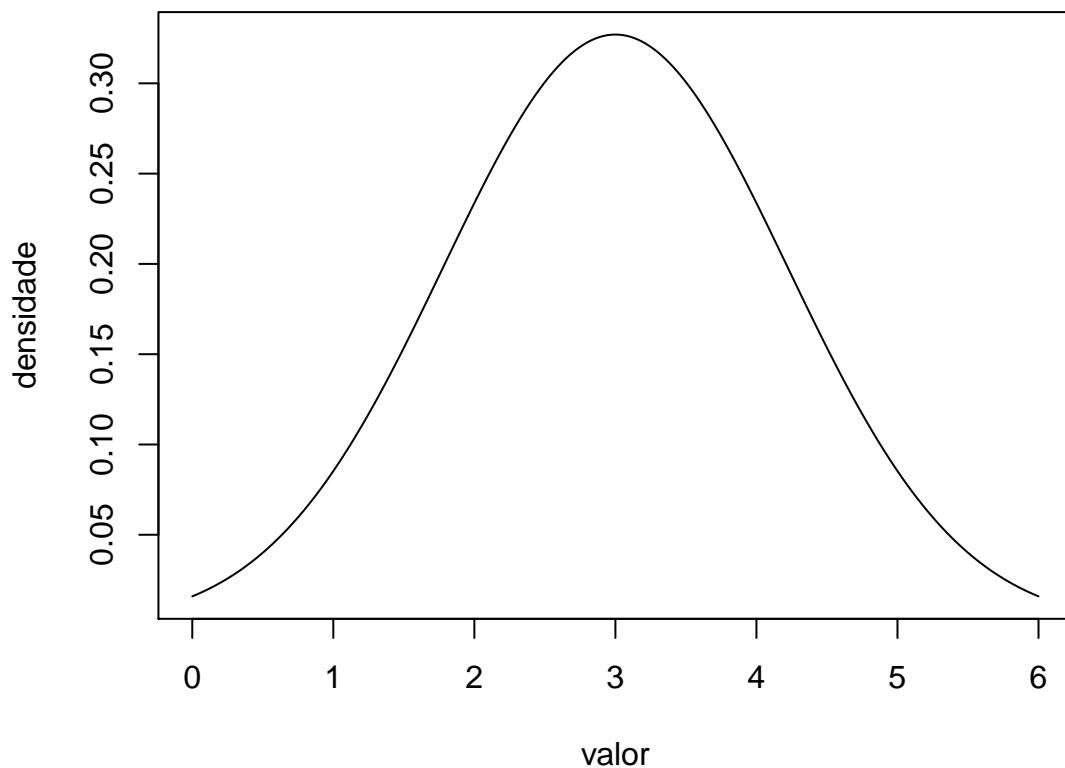
Para este exemplo procuramos mostrar uma distribuição normal com formato similar à binomial do exemplo anterior.

---

O código para ilustrar distribuições normais é muito similar aos mostrados para as distribuições binomiais, mas usando a função `dnorm()` no lugar de `dbinom()`.

Para mostrar uma distribuição normal com média de 3 podemos usar:

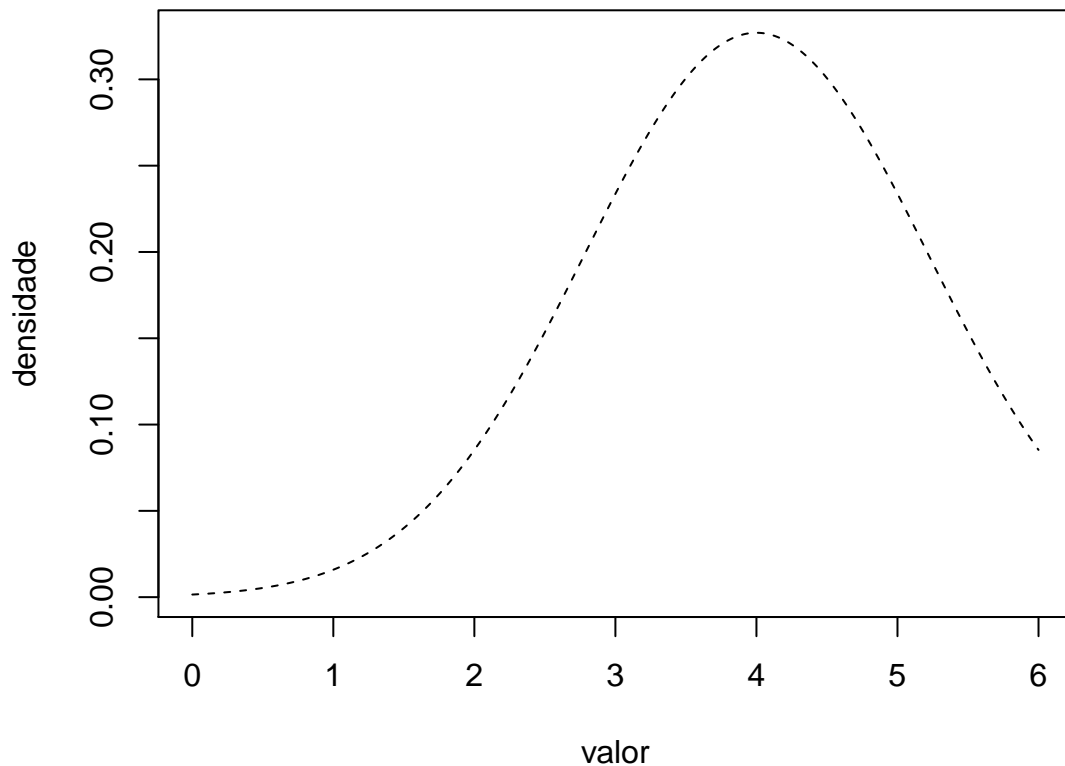
```
valores <- seq(from=0, to=6, by=0.01)
densidades <- dnorm(x=valores, mean=3, sd=1.22)
plot(valores, densidades,
     xlab="valor", ylab="densidade",
     type="l")
```



e para média de 4:

```
valores <- seq(from=0, to=6, by=0.01)
densidades <- dnorm(x=valores, mean=4, sd=1.22)
plot(valores, densidades,
```

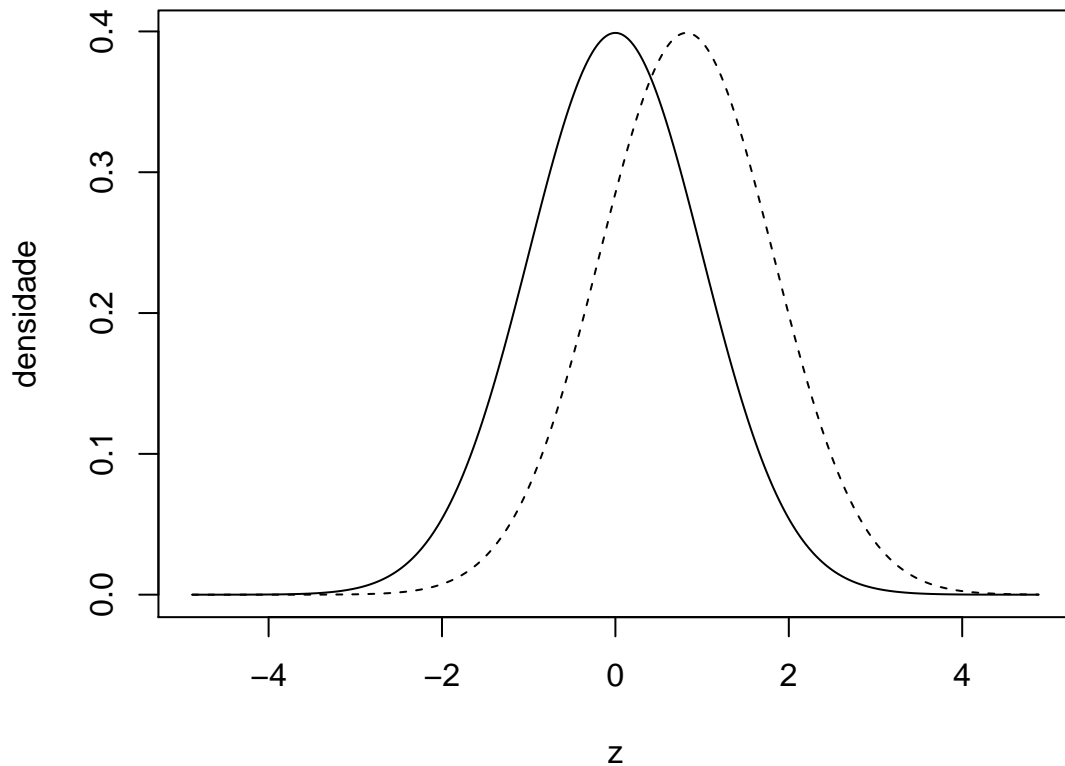
```
xlab="valor", ylab="densidade",
type="l", lty=2)
```



Observamos, novamente, a curva transladada para a direita quando a média aumenta de 3 para 4. O problema estatístico é análogo a distinguir uma moeda balanceada de uma viciada em  $\frac{2}{3}$ : saber se duas distribuições com médias numericamente diferentes podem ser tratadas como diversas.

A decisão estatística, considerando  $H_0 : \mu_A = \mu_B$ , é tomada com base nas distribuições normais padronizadas: é o caso de um teste  $z$ , e as duas distribuições exemplificadas aqui corresponderiam aproximadamente a:

```
z <- seq(from=-4*1.22, to=4*1.22, by=0.01)
H0_z <- dnorm(x=z, mean=0, sd=1)
plot(z, H0_z,
     xlab="z", ylab="densidade",
     type="l")
delta <- (4-3)/1.22
H1_z <- dnorm(x=z, mean=delta, sd=1)
lines(z, H1_z, lty=2)
```



Esta representação ilustra normais padronizadas no intervalo de  $\pm 4$  desvios-padrão. A distribuição de referência, que tinha média de 4 (correspondendo a  $H_0 : \mu_A = \mu_B$ , linha sólida) foi centrada em zero. A distribuição correspondente à média de 4 ( $H_1 : \mu_A \neq \mu_B$ , linha pontilhada) está a aproximadamente a 0.82 unidades de desvio-padrão acima da média de referência ( $(4 - 3)/1.22 \approx 0.8196$ ).

### a distribuição $t$

Para a distribuição  $t$  as funções são:

- `dt(x, df, ncp, log = FALSE)`
- `pt(q, df, ncp, lower.tail = TRUE, log.p = FALSE)`
- `qt(p, df, ncp, lower.tail = TRUE, log.p = FALSE)`
- `rt(n, df, ncp)`

Como as distribuições normais, as funções  $t$  são para variáveis quantitativas contínuas. Já são padronizadas e, portanto, não aparece média e desvio-padrão entre seus parâmetros. Em vez disto, aparecem duas novidades:

- os graus de liberdade ( $df$ , degrees of freedom), já discutidos, e
- o parâmetro de não centralidade ( $ncp$ ).

No caso,  $df$  é relacionado com o tamanho da amostra ( $df = n - 1$ ), e  $ncp$  define a translação da distribuição  $t$ , representando  $H_1$ . A curva correspondente a  $H_0$  tem  $ncp = 0$ .

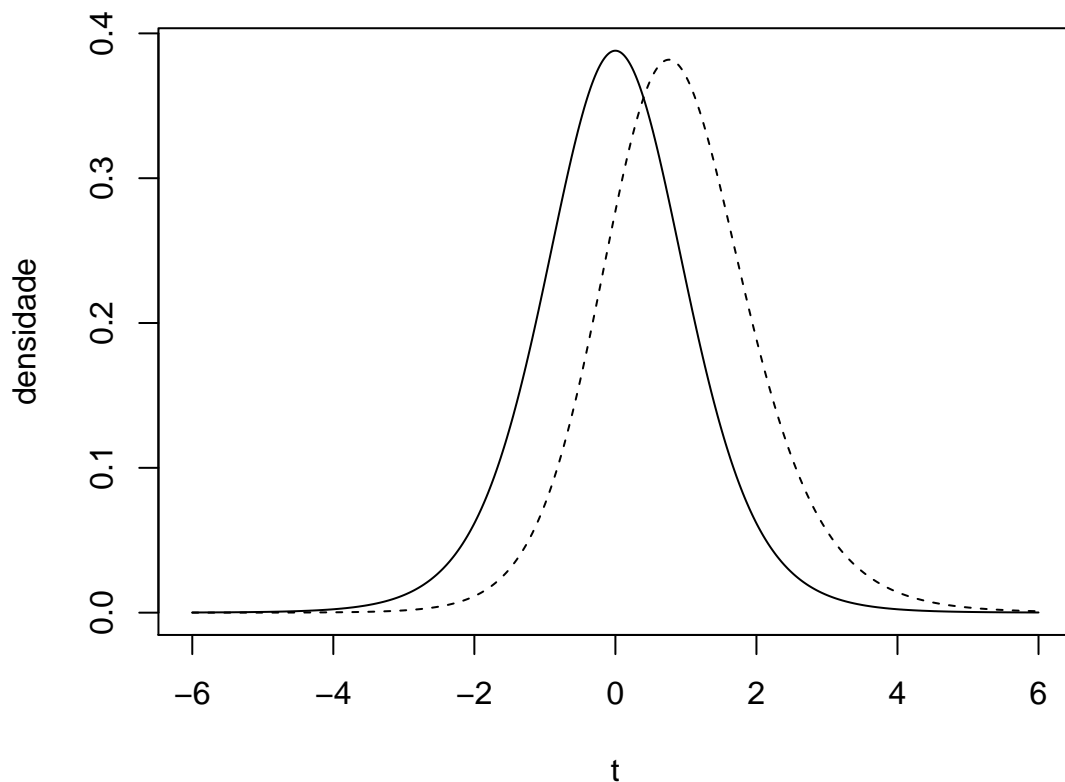
O código similar ao das normais padronizadas é:



```

t <- seq(from=-6, to=6, by=0.01)
H0_t <- dt(x=t, df=10-1, ncp=0)
plot(t, H0_t,
      xlab="t", ylab="densidade",
      type="l")
H1_t <- dt(x=t, df=10-1, ncp=0.82)
lines(t, H1_t, lty=2)

```

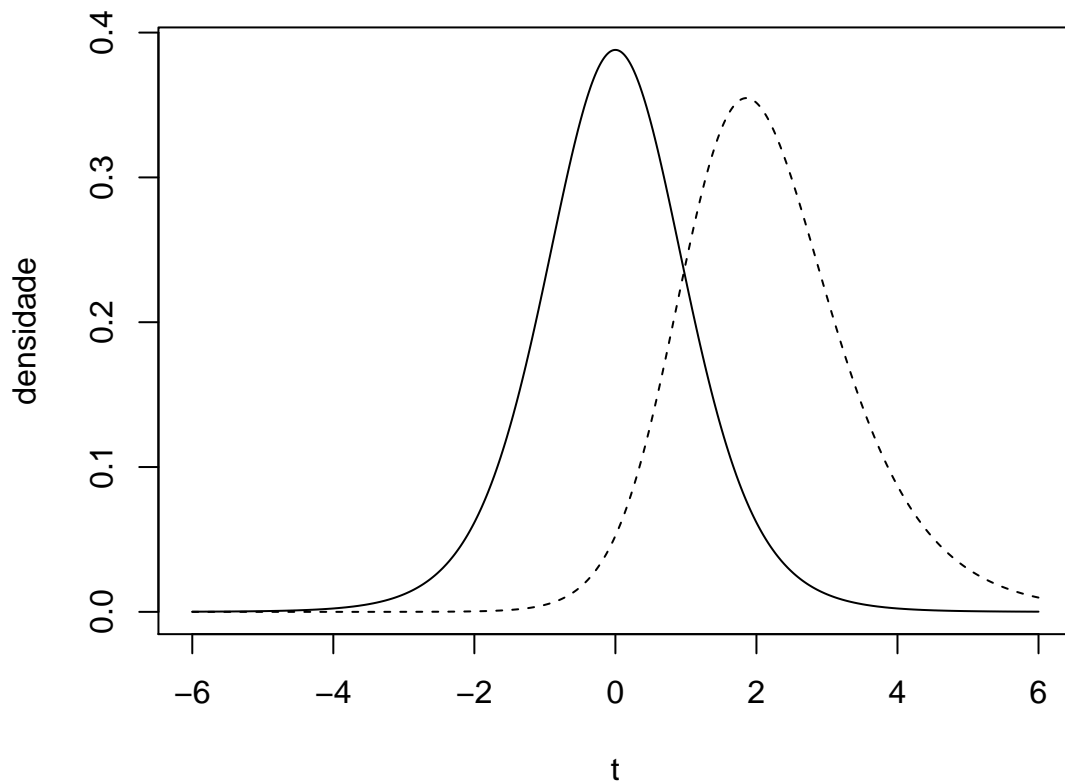


Algo além da translação ocorreu quando *ncp* não é zero: observe que a curva pontilhada é ligeiramente mais baixa que a de linha sólida. Mais difícil de perceber é que a curva pontilhada é assimétrica. É mais visível com valor maior de *ncp*:

```

t <- seq(from=-6, to=6, by=0.01)
H0_t <- dt(x=t, df=10-1, ncp=0)
plot(t, H0_t,
      xlab="t", ylab="densidade",
      type="l")
H1_t <- dt(x=t, df=10-1, ncp=2)
lines(t, H1_t, lty=2)

```



Na estatística  $t$ , então, as curvas que representam a hipótese alternativa são transladas mas, também, assimétricas.

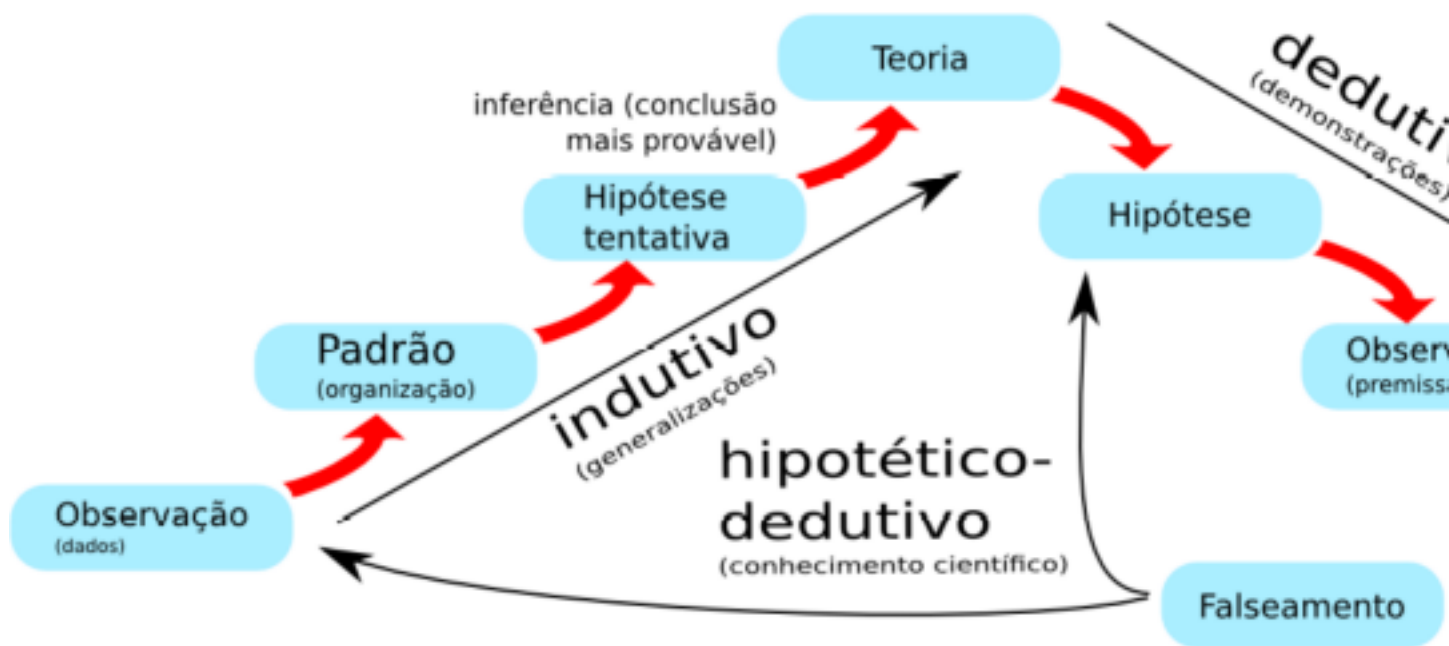
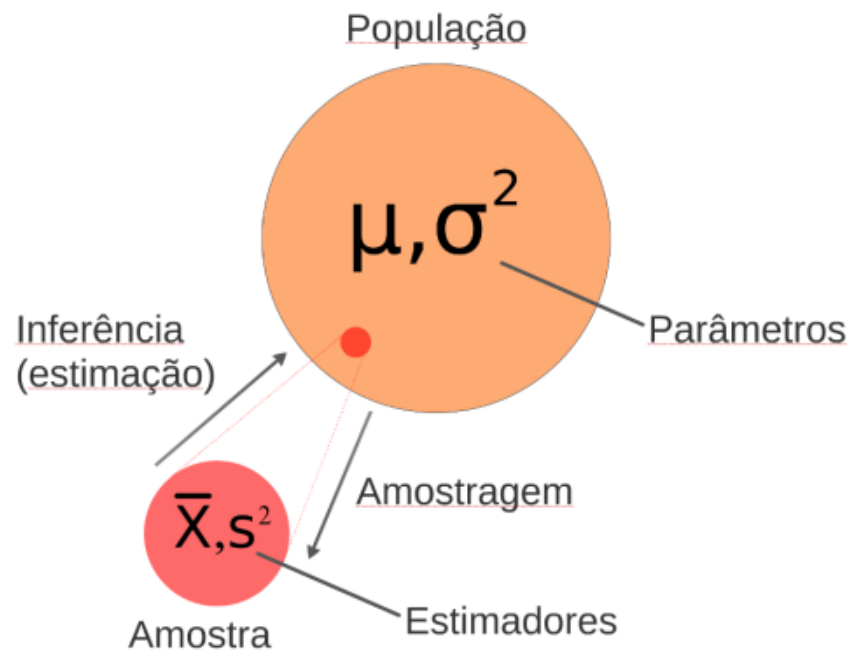


O código disponível em `Animacao_t_nao_central.R` compara a curva observada sob  $H_0$  na animação anterior com diversas curvas representando  $H_1$ s. Observe, sob  $H_1$ , assimetria das distribuições, e as áreas correspondentes a  $\beta$  e ao poder do teste  $(1 - \beta)$ .

---

## Raciocínio inferencial

Análise estatística inferencial é o processo de estimar características de uma população a partir de uma amostra, através de teste da hipótese nula, usando seus estimadores.



## Métodos Robustos



<https://performancedrive.com.au/icon-land-rover-defender-90-6-2-chev-v8-1606/>

## Teste $t$ para uma condição

### - situação

Suspeita-se de que um medicamento vasodilatador (Nifedipina) para Hipertensão Arterial, amplamente receitado, esteja aumentando a frequência cardíaca dos pacientes.

É sabido que a frequência cardíaca na população normal tem Distribuição Normal com média 70 bpm.

## - hipóteses (planejamento)

Para verificar essa suspeita, planejou-se obter uma amostra aleatória de 50 pacientes que recebem Nifedipina para se medir a frequência cardíaca.

$$H_0 : \mu_{nifedipina} = \mu_0$$

$$H_1 : \mu_{nifedipina} > \mu_0$$

Adota-se  $\mu_0 = 70$  bpm

---



O teste é unicaudal (só investigamos se há aumento da frequência cardíaca) e a direção é explícita em  $H_1$ , com o símbolo  $>$ . É comum encontrar a anotação da hipótese nula com o símbolo complementar ( $\leq$  neste exemplo):

$$H_0 : \mu_{nifedipina} \leq \mu_0$$

$$H_1 : \mu_{nifedipina} > \mu_0$$

Mas optamos por usar o símbolo de igualdade ( $=$ ) porque mais adequadamente espelha o que se espera de  $H_0$ , a ausência de efeito. Matematicamente, também, é equivalente (GATÁS RR (1978, p. 220-223) Elementos de Probabilidade e Inferência. SP: Atlas.)

---

## - coleta dos dados



A amostra de 50 pacientes forneceu:

72, 74, 70, 70, 69, 71, 72, 71, 69, 74, 71, 71, 70, 73, 69, 68, 68, 71, 71, 72, 70, 69, 73, 69, 71, 70, 72, 73, 70, 72, 67, 72, 67, 68, 69, 72, 70, 70, 70, 71, 74, 67, 69, 71, 71, 73, 71, 71, 70, 71

## - estatística descritiva

- Esquema de 5 pontos de Tuckey:

```
bpm <- c(72, 74, 70, 70, 69, 71, 72, 71, 69, 74, 71, 71, 70, 73, 69, 68, 68,
        71, 71, 72, 70, 69, 73, 69, 71, 70, 72, 73, 70, 72, 67, 72, 67, 68,
        69, 72, 70, 70, 70, 71, 74, 67, 69, 71, 71, 73, 71, 71, 70, 71)
sumario <- summary(bpm)
print(sumario)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
67.00	69.25	71.00	70.58	72.00	74.00

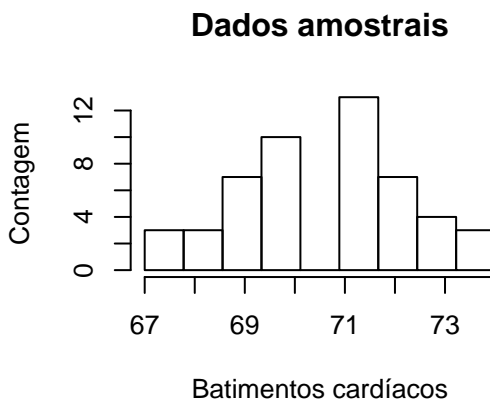
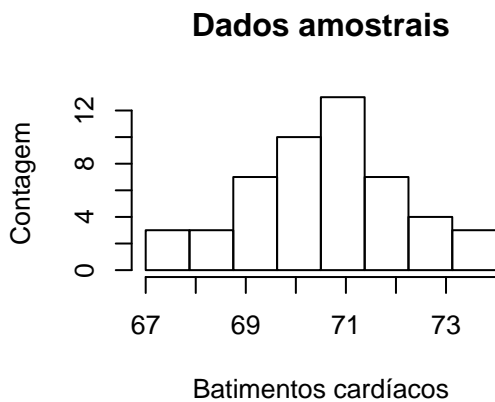
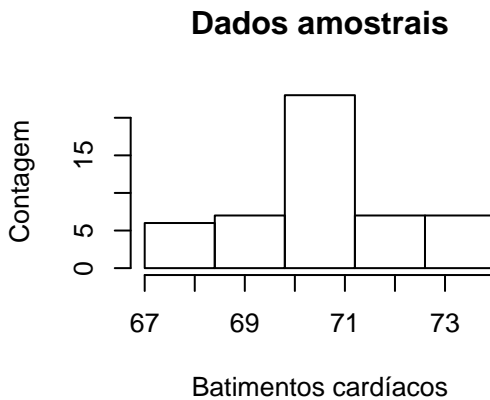
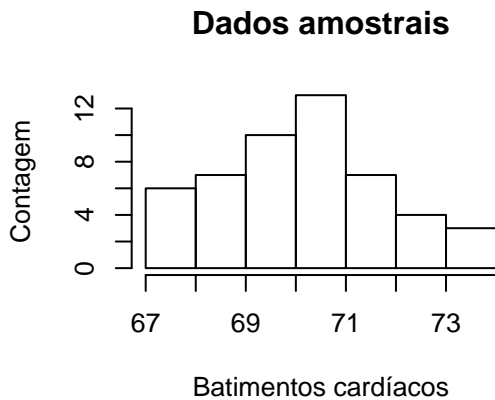
- Histogramas (para ver a seu gosto)

```

par(mfrow=c(2,2))

hist(bpm, freq=TRUE,
     main="Dados amostrais",
     xlab="Batimentos cardíacos", ylab="Contagem")
divisoes <- seq(from=min(bpm),to=max(bpm),by=(max(bpm)-min(bpm))/5)
hist(bpm, freq=TRUE, breaks=divisoes,
     main="Dados amostrais",
     xlab="Batimentos cardíacos", ylab="Contagem")
divisoes <- seq(from=min(bpm),to=max(bpm),by=(max(bpm)-min(bpm))/8)
hist(bpm, freq=TRUE, breaks=divisoes,
     main="Dados amostrais",
     xlab="Batimentos cardíacos", ylab="Contagem")
divisoes <- seq(from=min(bpm),to=max(bpm),by=(max(bpm)-min(bpm))/9)
hist(bpm, freq=TRUE, breaks=divisoes,
     main="Dados amostrais",
     xlab="Batimentos cardíacos", ylab="Contagem")

```



```

par(mfrow=c(1,1))

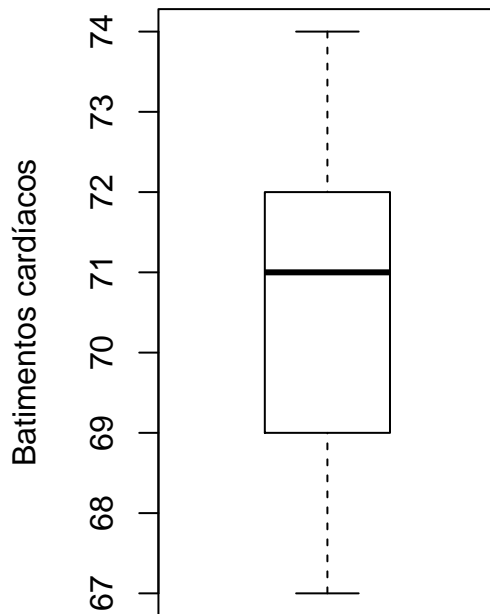
```

- *Boxplot* para ver a distribuição dos dados:

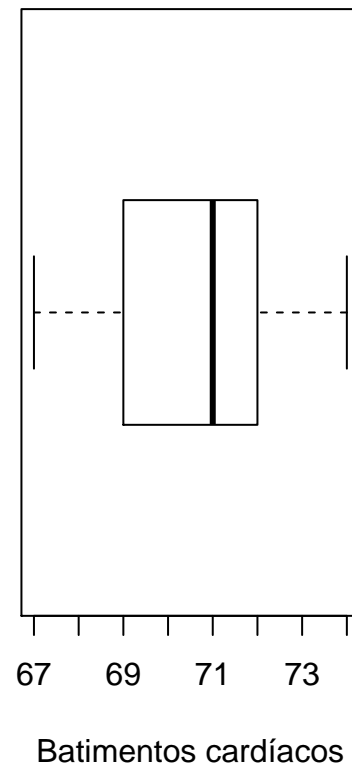
```
par(mfrow=c(1,2))

boxplot (bpm,
        main="Dados amostrais",
        ylab="Batimentos cardíacos", xlab="")
boxplot (bpm, horizontal = TRUE,
        main="Dados amostrais",
        xlab="Batimentos cardíacos", ylab="")
```

**Dados amostrais**



**Dados amostrais**

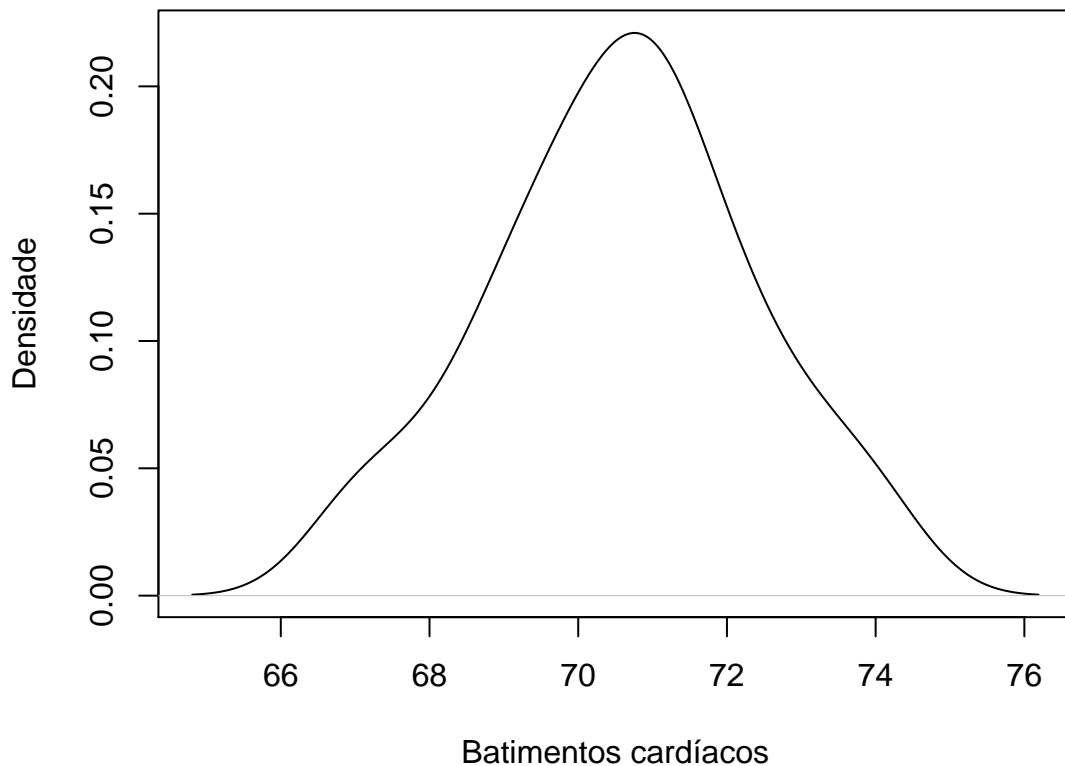


```
par(mfrow=c(1,1))
```

- *Density plot* para ver o formato da distribuição dos dados:

```
densprob <- density(bpm)
plot (densprob,
     main="Distribuição dos dados amostrais",
     xlab="Batimentos cardíacos", ylab="Densidade")
```

## Distribuição dos dados amostrais



no “olhômetro”, parece aproximar-se da distribuição normal?

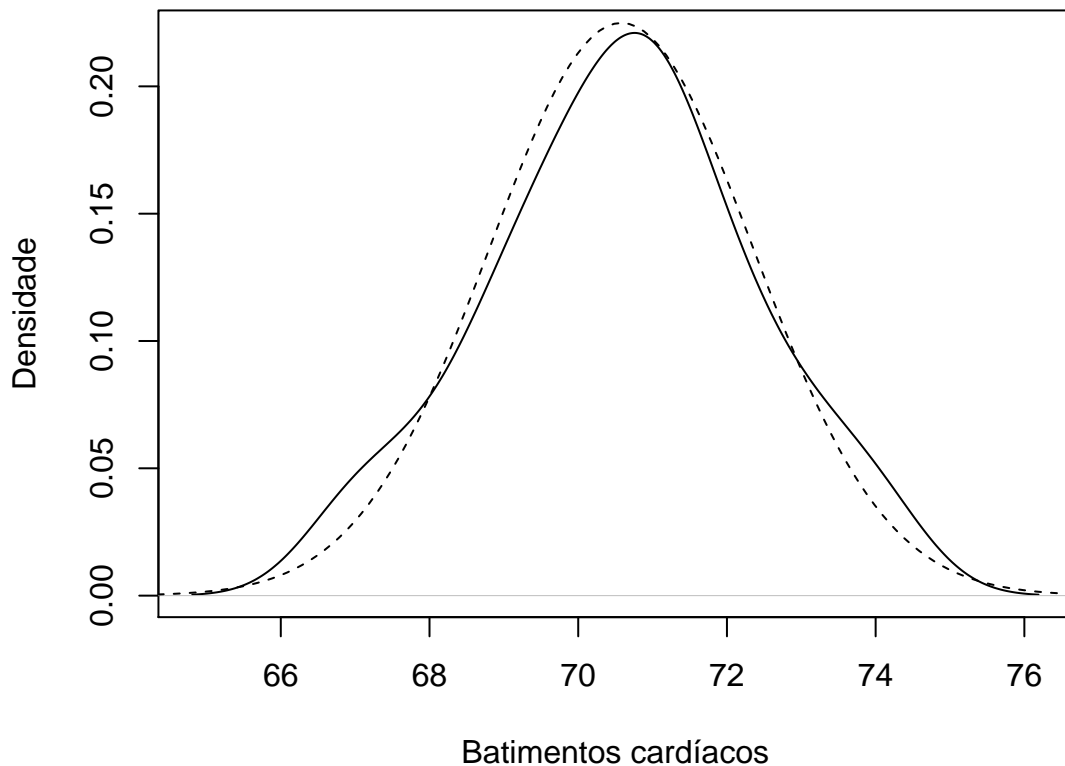
```
# dados
bpm <- c(72, 74, 70, 70, 69, 71, 72, 71, 69, 74, 71, 71, 70, 73, 69, 68, 68,
        71, 71, 72, 70, 69, 73, 69, 71, 70, 72, 73, 70, 72, 67, 72, 67, 68,
        69, 72, 70, 70, 70, 71, 74, 67, 69, 71, 71, 73, 71, 71, 70, 71)

# density plot
densprob <- density(bpm)
plot (densprob,
      main="Distribuição dos dados amostrais",
      xlab="Batimentos cardíacos", ylab="Densidade")

# distribuicao com media +- 4 desvios-padrao
media_bpm <- mean(bpm, na.rm = TRUE)
dp_bpm <- sd(bpm, na.rm = TRUE)
x <- seq(media_bpm-4*dp_bpm,media_bpm+4*dp_bpm,by=0.1)
distribnormal <- dnorm(x, mean=media_bpm, sd=dp_bpm)
lines(x, distribnormal, lty=2)
```



## Distribuição dos dados amostrais



### - estatística inferencial

Para um teste  $t$  para uma condição, unilateral à direita (note o uso de “*greater*”), é necessário fornecer o valor de referência ( $\mu_{pop} <- 70$ ) executado com:

```
bpm <- c(72, 74, 70, 70, 69, 71, 72, 71, 69, 74, 71, 71, 70, 73, 69, 68, 68,  
        71, 71, 72, 70, 69, 73, 69, 71, 70, 72, 73, 70, 72, 67, 72, 67, 68,  
        69, 72, 70, 70, 70, 71, 74, 67, 69, 71, 71, 73, 71, 71, 70, 71)  
mu_pop <- 70  
alfa <- 0.05  
t_out <- t.test(bpm, mu=mu_pop,  
                conf.level = 1-alfa, alternative = "greater")  
print (t_out)
```

One Sample t-test

```
data: bpm  
t = 2.312, df = 49, p-value = 0.01251  
alternative hypothesis: true mean is greater than 70  
95 percent confidence interval:  
 70.15942      Inf  
sample estimates:
```

```
mean of x
70.58
```

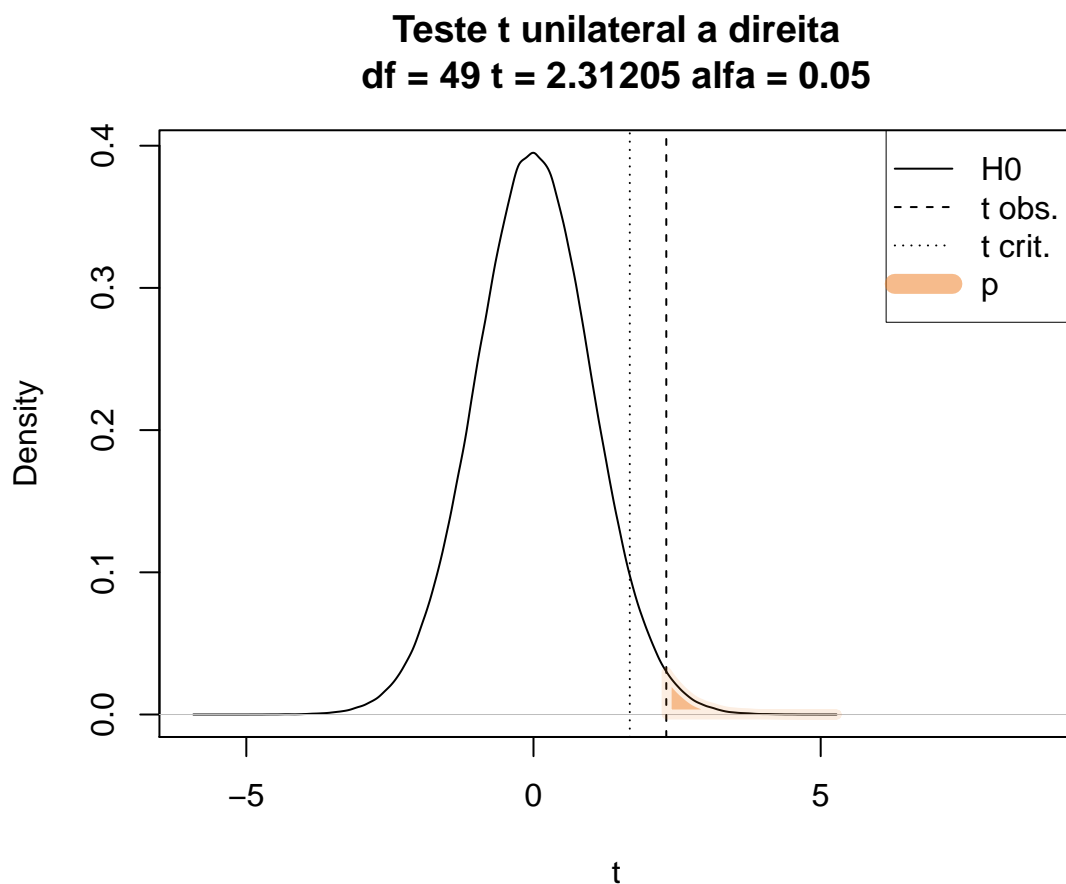
Para a decisão estatística, observe o valor da estatística do teste, guardada em `t_out$statistic=2.3120489` e o valor- $p$  associado em `t_out$p.value=0.0125097`.

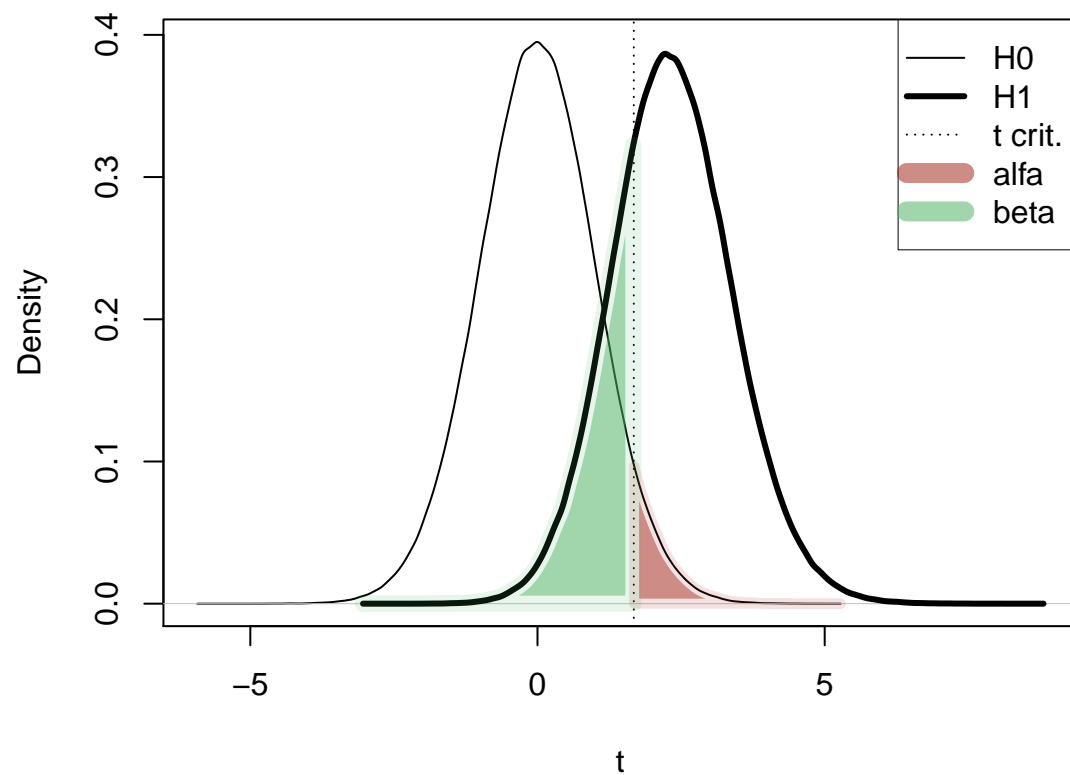
Para  $\alpha = 0.05$ , rejeita-se  $H_0$  e, portanto, o uso de nifedipina está associada ao aumento da frequência cardíaca neste estudo.

Para  $\alpha = 0.01$ , não se rejeita  $H_0$  e, portanto, não há elementos neste estudo para afirmar-se que o uso de nifedipina está associada ao aumento da frequência cardíaca.

**FALTOU escolher  $\alpha$  no planejamento do estudo.**

Disponibilizamos o RScript *Nifedipina.R* que reúne os procedimentos acima, gerando alguns resultados e gráficos adicionais.





BPM :

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
67.00	69.25	71.00	70.58	72.00	74.00

Desvio-padroao = 1.77

Tamanho da amostra = 50

One Sample t-test

```
data: Nifedipina$BPM
t = 2.312, df = 49, p-value = 0.01251
alternative hypothesis: true mean is greater than 70
95 percent confidence interval:
 70.15942      Inf
sample estimates:
mean of x
 70.58
```

Referencial teorico:

valor-p = 0.01250973

d de Cohen = 0.3269731 (Pequeno)

sob  $H_0$ : distribuicao t com media = 0 e d.p. = 1.021055

sob  $H_1$ : distribuicao t com media = 2.348206 e d.p. = 1.049534

Note que:

- o valor  $\beta$  e seu complemento, o poder do teste ( $1 - \beta$ ) nestes gráficos foram computados ‘a posteriori’ e, portanto, são inúteis para a decisão.
- apareceu o **d de Cohen**, medida de tamanho de efeito (significância prática), classificado como “pequeno” de acordo com:

## d de Cohen

Effect size	d	Reference
Very small	0.01	Sawilowsky, 2009
Small	0.20	Cohen, 1988
Medium	0.50	Cohen, 1988
Large	0.80	Cohen, 1988
Very large	1.20	Sawilowsky, 2009
Huge	2.0	Sawilowsky, 2009

Sawilowsky, S (2009) New effect size rules of thumb. Journal of Modern Applied Statistical Methods 8(2): 467-74.



A outra maneira de decidir é utilizar o intervalo de confiança (IC).

Note que a função `t.test()` exigiu o parâmetro  $\alpha$ ; necessário, justamente, para o cálculo do IC.

No exemplo, fornecemos  $\alpha = 0.05$  e foi computado  $IC_{95} = [70.1594209, \text{Inf}]$  (guardado na variável `t_out$conf.int`). O limite superior é infinito (*Inf* significa  $\infty$ ) porque o teste é unilateral. O valor populacional ( $\mu = 70 \text{ bpm}$ ) está fora do intervalo, levando à rejeição de  $H_0$  para este alfa.

Caso o teste fosse executado com  $\alpha = 0.01$  teríamos:

```
bpm <- c(72, 74, 70, 70, 69, 71, 72, 71, 69, 74, 71, 71, 70, 73, 69, 68, 68,
        71, 71, 72, 70, 69, 73, 69, 71, 70, 72, 73, 70, 72, 67, 72, 67, 68,
        69, 72, 70, 70, 70, 71, 74, 67, 69, 71, 71, 73, 71, 71, 70, 71)
mu_pop <- 70
alfa <- 0.01
t_out <- t.test(bpm, mu=mu_pop,
               conf.level = 1-alfa, alternative = "greater")
print(t_out)
```

### One Sample t-test

```
data: bpm
t = 2.312, df = 49, p-value = 0.01251
alternative hypothesis: true mean is greater than 70
```

```
99 percent confidence interval:
 69.97671      Inf
sample estimates:
mean of x
  70.58
```

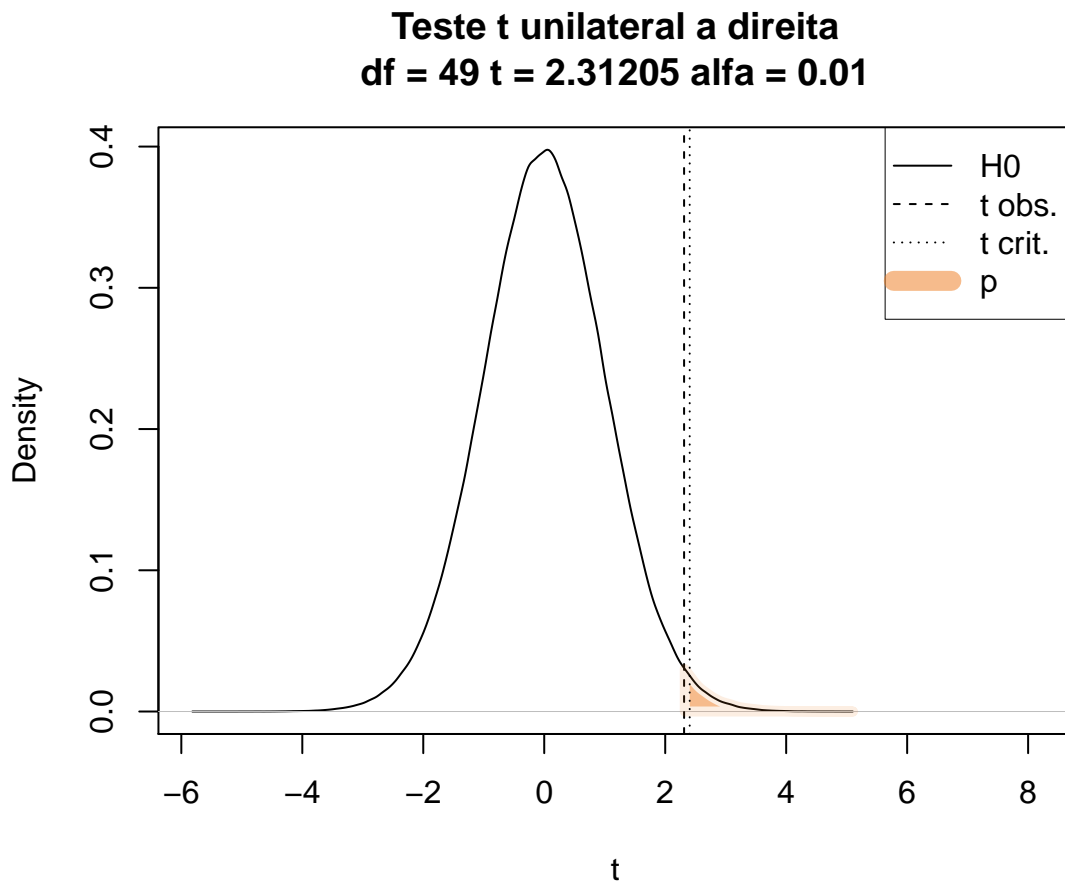
O valor populacional ( $\mu = 70 \text{ bpm}$ ) agora está dentro do intervalo, levando à não rejeição de  $H_0$ .

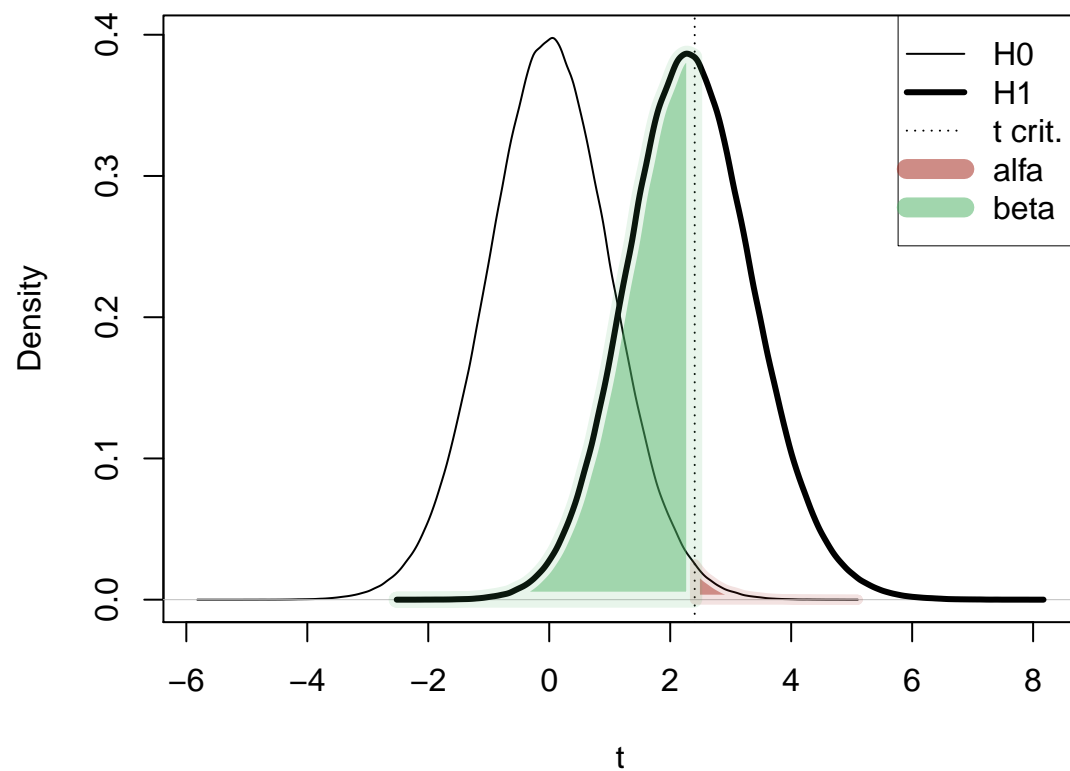
---

Alterando-se, no início de *Nifedipina.R* o valor de alfa:

```
alfa <- 0.01 # nível de significancia adotado
```

a saída altera-se de acordo, mostrando  $p > \alpha$  e a não-rejeição de  $H_0$ :





BPM :

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
67.00	69.25	71.00	70.58	72.00	74.00

Desvio-padroao = 1.77

Tamanho da amostra = 50

One Sample t-test

```
data: Nifedipina$BPM
t = 2.312, df = 49, p-value = 0.01251
alternative hypothesis: true mean is greater than 70
99 percent confidence interval:
 69.97671      Inf
sample estimates:
mean of x
  70.58
```

Referencial teorico:

valor-p = 0.01250973

d de Cohen = 0.3269731 (Pequeno)

sob  $H_0$ : distribuicao t com media = 0 e d.p. = 1.021055

sob  $H_1$ : distribuicao t com media = 2.348206 e d.p. = 1.049534

## Revendo o raciocínio

1. Formular a hipótese de interesse.
2. Fixar um nível de significância ( $\alpha$ ).
3. Escolher e executar o teste estatístico apropriado.
4. Decidir sobre  $H_0$ .

$H_0$  é a hipótese nula (sempre abrange a igualdade).  $H_1$  é a hipótese alternativa (oposta da  $H_0$ ).  $H_0$  pode ser não rejeitada ou rejeitada (a rejeição deve ser baseada em evidências obtidas a partir da amostra).

A decisão estatística sempre envolve possíveis erros:

		“A VERDADE”	
		$H_0$ é Verdadeira	$H_0$ é Falsa
Decisão	Aceitar $H_0$	Decisão Correta	Erro Tipo II ( $\beta$ )
	Rejeitar $H_0$	Erro tipo I ( $\alpha$ )	Decisão Correta

que são:

- $\alpha$ , a probabilidade de erro do tipo I, rejeitar  $H_0$  quando não há efeito,
- $\beta$ , a probabilidade de erro do tipo II, não rejeitar  $H_0$  quando há efeito (mas o valor de  $\beta$  precisa ser estabelecido *a priori* para ter valor no momento da decisão sobre o teste).

A decisão sobre o teste depende da comparação entre a probabilidade de se observar uma diferença sob a hipótese nula, dada uma amostra de tamanho  $n$  e a probabilidade do erro do tipo I ( $\alpha$ ) escolhida previamente:

- se a probabilidade de que a diferença seja observada ao acaso for “grande” ( $p > \alpha$ ), não se rejeita  $H_0$  (só podemos falar em aceitar  $H_0$  se o poder *a priori* for maior que 90%).
- se a probabilidade de que a diferença seja observada ao acaso for “pequena” ( $p < \alpha$ ), rejeita-se  $H_0$ .

## $t$ pareado (duas condições dependentes)

Também chamado de teste  $t$  relacionado, aplica-se tipicamente às situações em que o mesmo indivíduo (ou a mesma unidade experimental) tem uma variável quantitativa medida em dois momentos ou duas condições experimentais.

### -situação

A pesquisadora Yob está interessada na violência de massa durante as partidas de futebol. Ela pensa que a violência do grupo é resultado dos assentos desconfortáveis do estádio.

Adaptado de Dancey & Reidy (2011) Estatística sem matemática para psicologia. 5a edição. Porto Alegre: Penso.



## - planejamento

Por isso, Yob modifica dois estádios diferentes na Inglaterra. Em um estádio coloca assentos bem apertados e desconfortáveis. No outro, instala assentos confortáveis, com muito espaço para as pernas e entre os assentos adjacentes.

A professora organiza uma competição, de modo que um clube jogue metade das partidas em um estádio e a outra metade no outro estádio. Ela prevê que o número de prisões e expulsões será maior no estádio que apresenta os assentos mais desconfortáveis.

- 
- Este é um delineamento entreparticipantes ou intraparticipantes? **intraparticipantes**
  - Que tipo de variável a professora Yob mediu: discreta ou contínua? **discreta**
    - Qual é a variável independente (VI)? **tipo de acomodação nos estádios**
    - Qual é a variável dependente (VD)? **diferença do número de prisões ou expulsões (NPE)**
  - Este é um teste unilateral ou bilateral? **unilateral**
  - Qual é a hipótese nula?  $H_0$  : **a diferença populacional de NPE entre as duas condições é nula.**
  - Qual é a hipótese de pesquisa?  $H_1$  : **a diferença populacional de NPE é maior nos estádios desconfortáveis.**
  - Qual alfa escolhe?  $\alpha = 0.05$
- 

## - coleta dos dados

Ela acompanha um grupo de 12 fãs adolescentes agressivos e grosseiros do clube e registra o número de vezes que cada um é preso ou expulso do estádio.

Aqui disponibilizamos o *RScript* *Violencia\_estadios.R* e destacamos seus principais trechos.

Os dados estão disponíveis na planilha Excel *Violencia\_estadios.xlsx*:

```
library(readxl)
Dtfrm <- read_excel("Violencia_estadios.xlsx", sheet = "dependente")
# diferença entre as condições experimentais
Dtfrm$dif <- Dtfrm$Conforto - Dtfrm$Desconforto
print(Dtfrm)
```

```
# A tibble: 12 x 4
  Adolescente Desconforto Conforto   dif
  <chr>         <dbl>      <dbl> <dbl>
```



1	a	8	3	-5
2	b	5	2	-3
3	c	4	4	0
4	d	6	6	0
5	e	4	2	-2
6	f	8	1	-7
7	g	9	6	-3
8	h	10	3	-7
9	i	7	4	-3
10	j	8	1	-7
11	k	6	4	-2
12	l	7	3	-4

Abra a planilha para verificar como os dados foram guardados e como aparecem ao serem lidos. Note que `read_excel()` lê a aba “dependente” da planilha, na qual cada linha tem as observações de um indivíduo, feitas nas duas condições experimentais.

## - estatística descritiva

Exibe a estatística descritiva:

```
cat("\nEsquema de 5 estatísticas de Tukey & média\n")
cat("\nTamanho da amostra: ",length(Dtfrm$dif),"\n", sep="")
cat("\nNúmero de ocorrências em ",names(Dtfrm)[2],":\n", sep="")
sumario <- summary(Dtfrm[[2]], digits = 3)
print (sumario)
cat("\nNúmero de ocorrências ",names(Dtfrm)[3],":\n", sep="")
sumario <- summary(Dtfrm[[3]], digits = 3)
print (sumario)
cat("Diferença do número de ocorrências (",names(Dtfrm)[3]," - ",names(Dtfrm)[2],"):\n", sep="")
sumario <- summary(Dtfrm$dif, digits = 3)
print (sumario)
```

Esquema de 5 estatísticas de Tukey & média

Tamanho da amostra: 12

Número de ocorrências em Desconforto:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
4.00	5.75	7.00	6.83	8.00	10.00

Número de ocorrências Conforto:

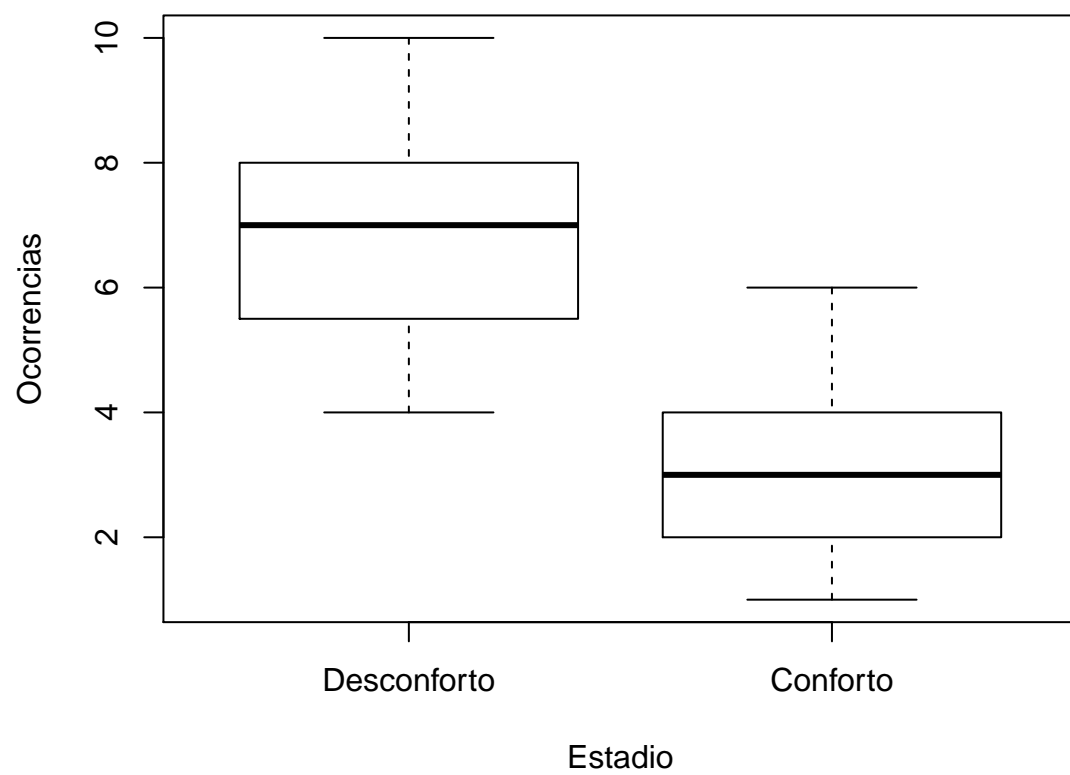
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.00	2.00	3.00	3.25	4.00	6.00

Diferença do número de ocorrências (Conforto - Desconforto):

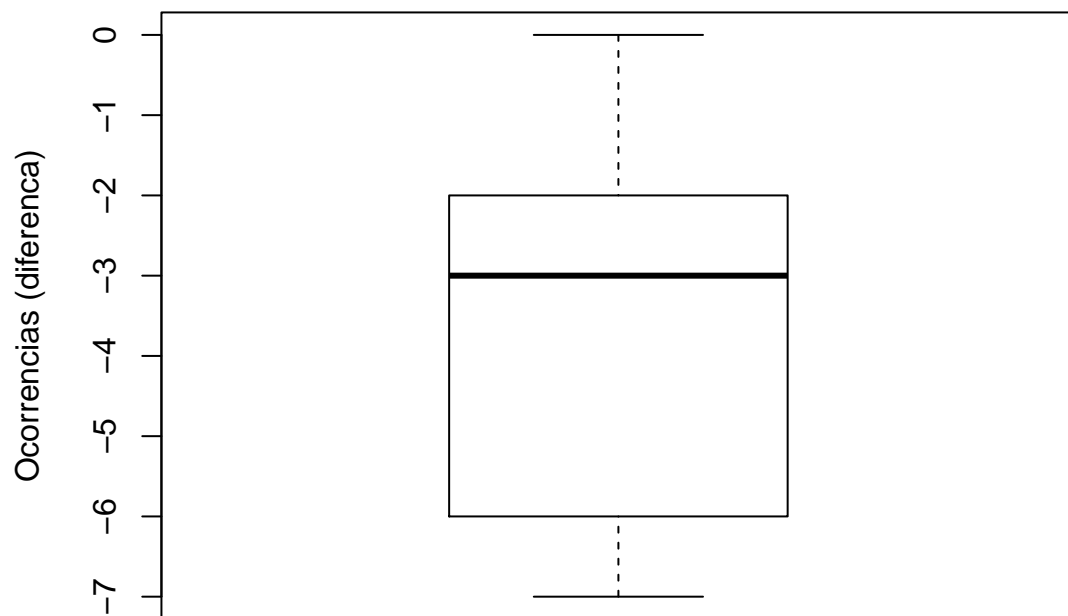
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-7.00	-5.50	-3.00	-3.58	-2.00	0.00

Gera alguns gráficos para estatística descritiva (abra o código R para ver como os gráficos foram gerados):

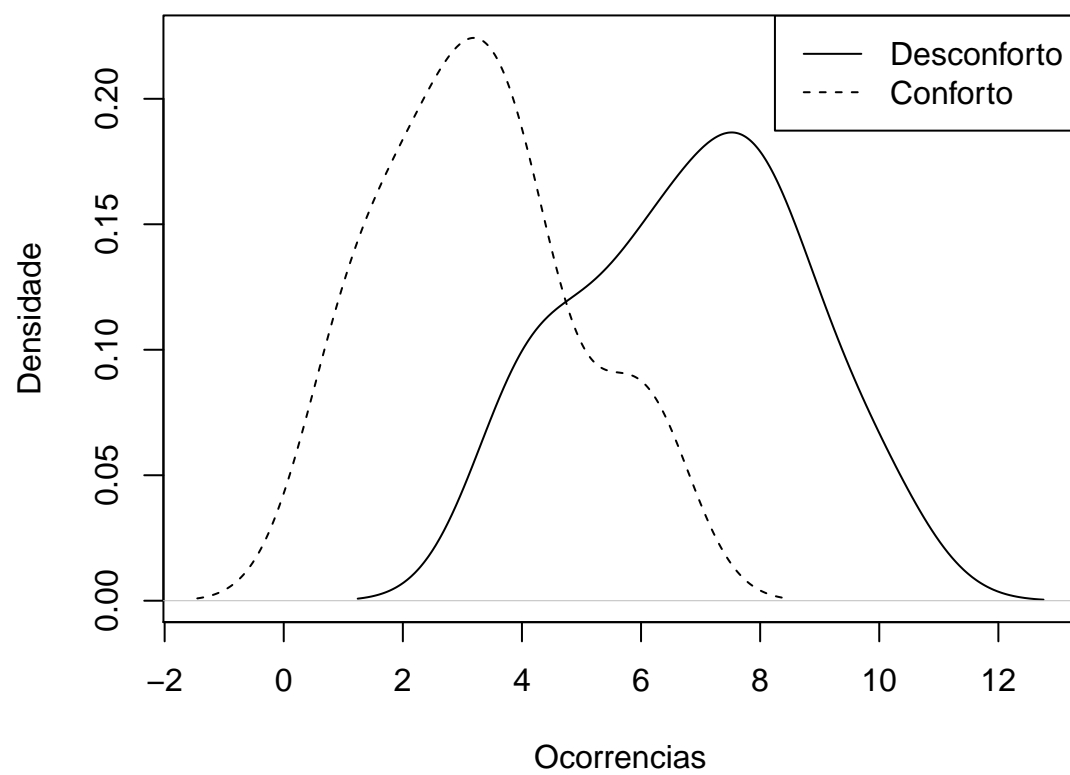
## Distribuição dos dados



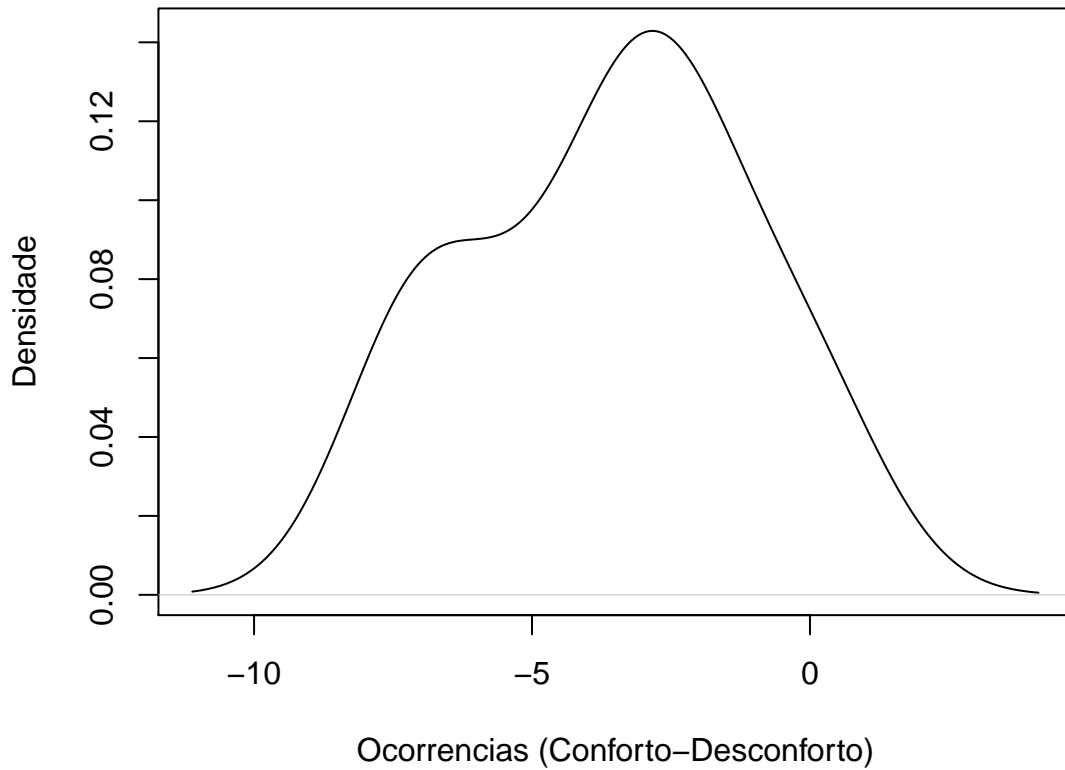
### Diferenças de ocorrências (Conforto-Desconforto)



## Distribuição dos dados



## Distribuição dos dados



### - estatística inferencial

Faz testes de estatística inferencial:

```
corr_test <- cor.test(Dtfm$Desconforto,Dtfm$Conforto)
cat("\nCoeficiente de correlacao de Pearson\n")
print (corr_test)
```

Coeficiente de correlacao de Pearson

Pearson's product-moment correlation

```
data: Dtfm$Desconforto and Dtfm$Conforto
t = 0.04565, df = 10, p-value = 0.9645
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.5641406  0.5835022
sample estimates:
      cor
0.01443426
```

mostrando que os comportamentos do mesmo indivíduo em cada uma das condições experimentais não estão associados.

```
cat("Análise de significancia estatística: valor-p\n")
t_out <- t.test(Dtfrm$dif,mu=0,alternative="less")
print(t_out)
```

Análise de significancia estatística: valor-p

One Sample t-test

```
data: Dtfrm$dif
t = -4.9592, df = 11, p-value = 0.0002147
alternative hypothesis: true mean is less than 0
95 percent confidence interval:
 -Inf -2.285695
sample estimates:
mean of x
-3.583333
```

mostrando que há diferença entre as condições experimentais (rejeição de  $H_0$ , valor- $p < \alpha$ , intervalo de confiança 95% não inclui o valor zero). O teste foi feito com Conforto – Desconforto, encontrando-se números negativos: então NPE é maior em Desconforto; o número de NPE é significativamente maior nos estádios desconfortáveis.

Computa, também, valores para a significância prática:

```
F <- t^2
df <- t_out$parameter
eta2 <- F/(F+df)

# Elis P (2010) The essential guide to effect sizes. Cambridge
if (eta2 < 0.01) {mag_eta2<-c("Desprezível")}
if (eta2>=0.01 && eta2<0.06) {mag_eta2<-c("Pequeno")}
if (eta2>=0.06 && eta2<0.14) {mag_eta2<-c("Intermediário")}
if (eta2>=0.14) {mag_eta2<-c("Grande")}
R2aj <- (F-1)/(F+df)

# tamanho de efeito d de Cohen
dp <- sd(Dtfrm$dif)
m <- t_out$estimate
d <- abs(t_out$statistic)/sqrt(t_out$parameter+1)
# Sawilowsky, S (2009) New effect size rules of thumb. Journal of Modern Applied Statistical Methods 8(
if (d<0.01) {mag_Cohen<-c("Desprezível")}
if (d>=0.01 && d<0.2) {mag_Cohen<-c("Muito pequeno")}
if (d>=0.2 && d<0.5) {mag_Cohen<-c("Pequeno")}
if (d>=0.5 && d<0.8) {mag_Cohen<-c("Intermediário")}
if (d>=0.8 && d<1.2) {mag_Cohen<-c("Grande")}
if (d>=1.2 && d<2) {mag_Cohen<-c("Muito grande")}
if (d>=2) {mag_Cohen<-c("Enorme")}

cat("Análise de significancia prática: tamanho de efeito\n")
cat("\td de Cohen = ",d," (",mag_Cohen,")", "\n",sep="")
cat("\teta^2 = R^2 = ",eta2," (",mag_eta2,")\n",sep="")
```

Análise de significancia prática: tamanho de efeito

$d$  de Cohen = 1.431599 (Muito grande)

$\eta^2 = R^2 = 0.3270348$  (Grande)

A tabela implementada no *RScript* para a intensidade do efeito calculado por  $\eta^2$  ( $\eta^2$ ) é:

Table 2.1 *Cohen's effect size benchmarks*

Test	Relevant effect size	Effect size classes		
		Small	Medium	Large
Comparison of independent means	$d$ , $\Delta$ , Hedges' $g$	.20	.50	.80
Comparison of two correlations	$q$	.10	.30	.50
Difference between proportions	Cohen's $g$	.05	.15	.25
Correlation	$r$	.10	.30	.50
	$r^2$	.01	.09	.25
Crosstabulation	$w$ , $\phi$ , $V$ , $C$	.10	.30	.50
ANOVA	Cohen's $f = \sqrt{\eta^2 / (1 - \eta^2)}$	.10	.25	.40
	$\eta^2$	.01	.06	.14
Multiple regression	$R^2$	.02	.13	.26
	$f^2$	.02	.15	.35

Notes: The rationale for most of these benchmarks can be found in Cohen (1988) at the following pages: Cohen's  $d$  (p. 40),  $q$  (p. 115), Cohen's  $g$  (pp. 147–149),  $r$  and  $r^2$  (pp. 79–80), Cohen's  $w$  (pp. 224–227),  $f$  and  $\eta^2$  (pp. 285–287),  $R^2$  and  $f^2$  (pp. 413–414).

Elis P (2010) The essential guide to effect sizes. Cambridge

## teste $t$ para duas condições independentes (teste $t$ de Welch)

### - situação



<https://www.fns.usda.gov/snap/supplemental-nutrition-assistance-program-education-snap-ed>

O SNAP-Ed (Supplemental Nutrition Assistance Program Education) é um programa baseado em evidências que ajuda as pessoas a terem uma vida mais saudável.

O SNAP-Ed ensina às pessoas que usam ou qualificam para o SNAP uma boa nutrição e como fazer com que o seu dinheiro de alimentação se estenda ainda mais.

Os participantes do SNAP-Ed também aprendem a ser fisicamente ativos.

## - planejamento

Brendon Small e Coach McGuirk fazem com que seus alunos do SNAP-Ed mantenham diários do que comem por uma semana e depois calculam a ingestão diária de sódio em miligramas.

Desde que as classes receberam diferentes programas de educação nutricional, eles querem ver se a ingestão média de sódio é a mesma para as duas turmas.

## - coleta dos dados

Desenvolvemos *Nutricao.R* para as análises descritiva e inferencial. Abaixo destacamos seus techos principais. Neste *RScript* aproveitamos funções de alguns pacotes (verifique se estão instalados em seu computador)

```
library(readxl)
library(car)
```

Loading required package: carData

```
library(lattice)
library(ggplot2)
```

Registered S3 methods overwritten by 'ggplot2':

method	from
[.quosures	rlang
c.quosures	rlang
print.quosures	rlang

```
library(rcompanion)
```

Os dados estão na planilha *Nutricao.xlsx*:

```
Dtfm <- read_excel("Nutricao.xlsx")
# os instrutores devem ser tratados como fator
Dtfm$Instructor <- as.factor(Dtfm$Instructor)
Dtfm$Instructor <- factor(Dtfm$Instructor, levels=unique(Dtfm$Instructor))
print(Dtfm)
```

```
# A tibble: 40 x 3
  Instructor Student Sodium
  <fct>      <chr>    <dbl>
1 Brendon Small a      1200
2 Brendon Small b      1400
3 Brendon Small c      1350
4 Brendon Small d       950
5 Brendon Small e      1400
6 Brendon Small f      1150
7 Brendon Small g      1300
8 Brendon Small h      1325
9 Brendon Small i      1425
10 Brendon Small j      1500
# ... with 30 more rows
```

## - estatística descritiva

Verificamos se os dados estão coerentes:



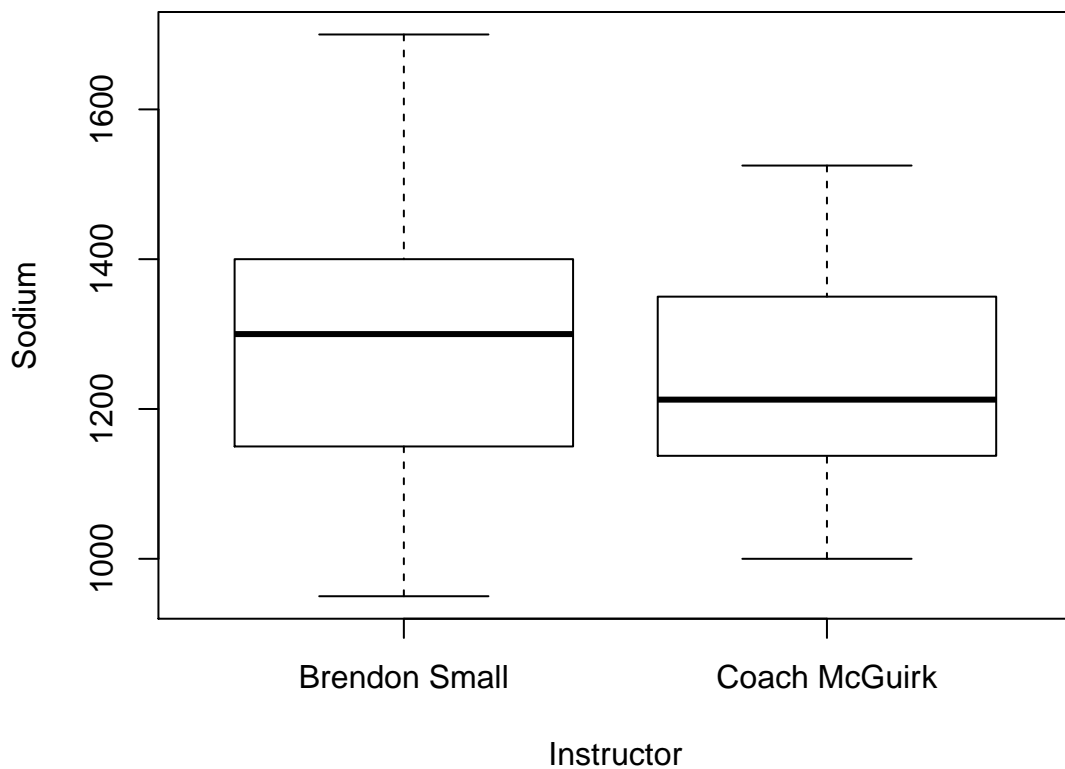
```
res_sodiumbyinstructor <- summary.data.frame(Dtfrm, digits=2)
print (res_sodiumbyinstructor)
```

Instructor	Student	Sodium
Brendon Small:20	Length:40	Min. : 950
Coach McGuirk:20	Class :character	1st Qu.:1150
	Mode :character	Median :1250
		Mean :1267
		3rd Qu.:1362
		Max. :1700

Gráficos sugeridos:

- *boxplot*:

```
grf <- boxplot(Sodium ~ Instructor, data=Dtfrm, ylab=names(Dtfrm)[3],
               xlab="Instructor")
```



```
print(grf)
```

```
$stats
      [,1] [,2]
[1,]  950 1000.0
[2,] 1150 1137.5
[3,] 1300 1212.5
```

```
[4,] 1400 1350.0  
[5,] 1700 1525.0
```

```
$n  
[1] 20 20
```

```
$conf  
      [,1]      [,2]  
[1,] 1211.675 1137.424  
[2,] 1388.325 1287.576
```

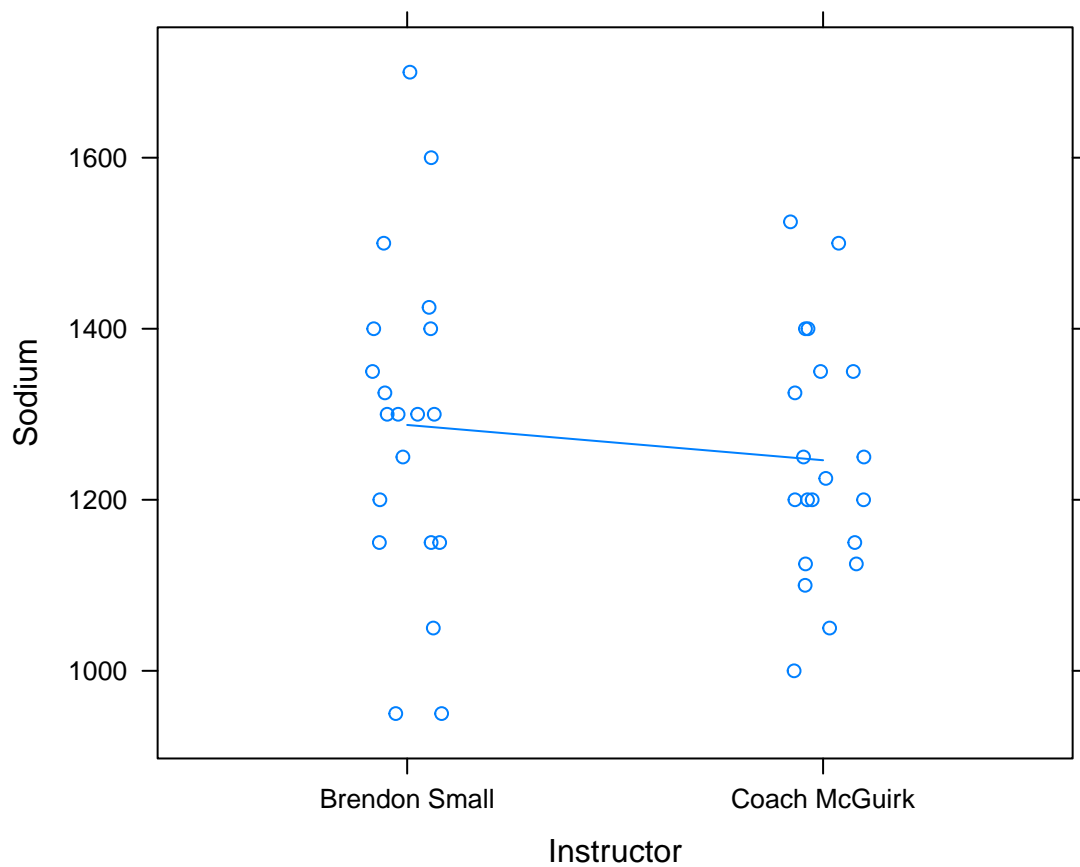
```
$out  
numeric(0)
```

```
$group  
numeric(0)
```

```
$names  
[1] "Brendon Small" "Coach McGuirk"
```

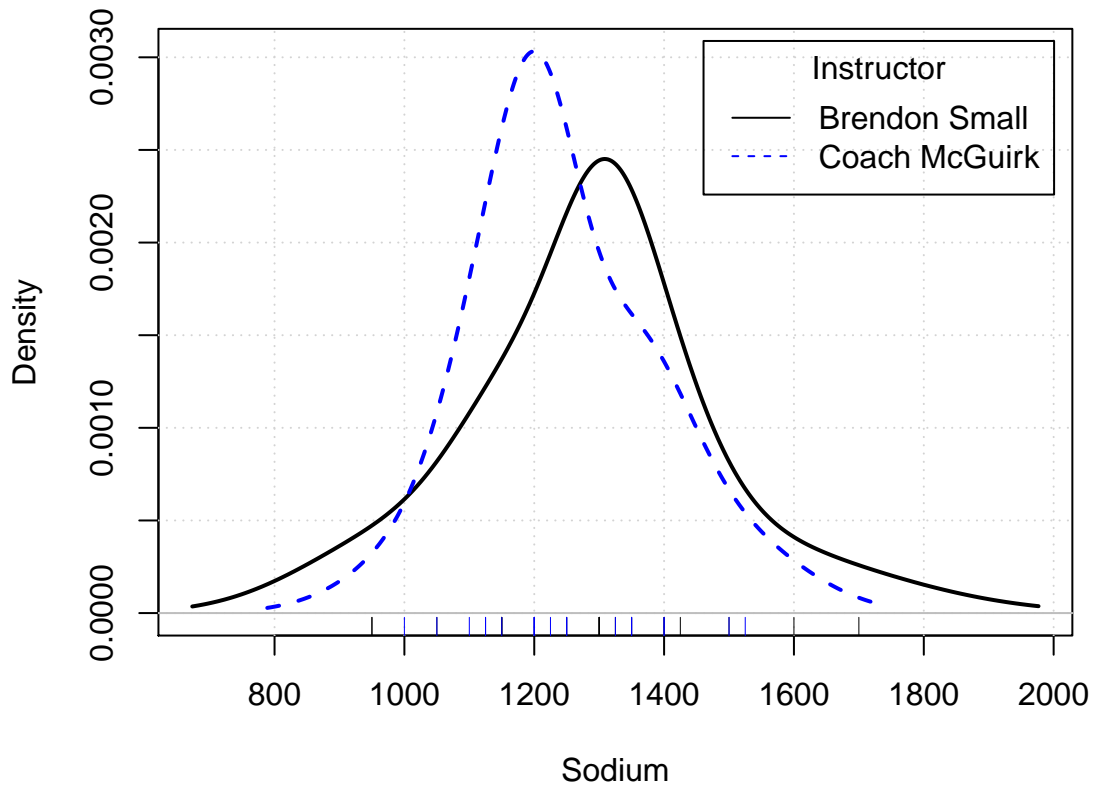
- *xyplot()* do pacote *lattice*:

```
grf <- lattice::xyplot(Sodium ~ Instructor, data=Dtfrm, type=c("p","a"), jitter.x=TRUE)  
print(grf)
```



- uma variante de *density plot* do pacote *car*:

```
grf <- car::densityPlot(Sodium~Instructor, data=Dtfirm, rug=TRUE)
```



```
print(grf)
```

```
$`Brendon Small`
```

```
Call:
```

```
adaptiveKernel(x = x[g == group], bw = if (is.numeric(bw)) bw[group] else bw, adjust = adjust[g
```

```
Data: x[g == group] (500 obs.); Bandwidth 'bw' = 92.23
```

	x		y
Min.	: 673.3	Min.	:3.574e-05
1st Qu.	: 999.2	1st Qu.	:1.735e-04
Median	:1325.0	Median	:4.381e-04
Mean	:1325.0	Mean	:7.619e-04
3rd Qu.	:1650.8	3rd Qu.	:1.218e-03
Max.	:1976.7	Max.	:2.451e-03

```
$`Coach McGuirk`
```

```
Call:
```

```

    adaptiveKernel(x = x[g == group], bw = if (is.numeric(bw)) bw[group] else bw,      adjust = adjust[g
Data: x[g == group] (500 obs.); Bandwidth 'bw' = 70.4

```

	x	y
Min.	: 788.8	Min. :2.804e-05
1st Qu.	:1025.6	1st Qu.:2.137e-04
Median	:1262.5	Median :7.231e-04
Mean	:1262.5	Mean :1.050e-03
3rd Qu.	:1499.4	3rd Qu.:1.708e-03
Max.	:1736.2	Max. :3.032e-03

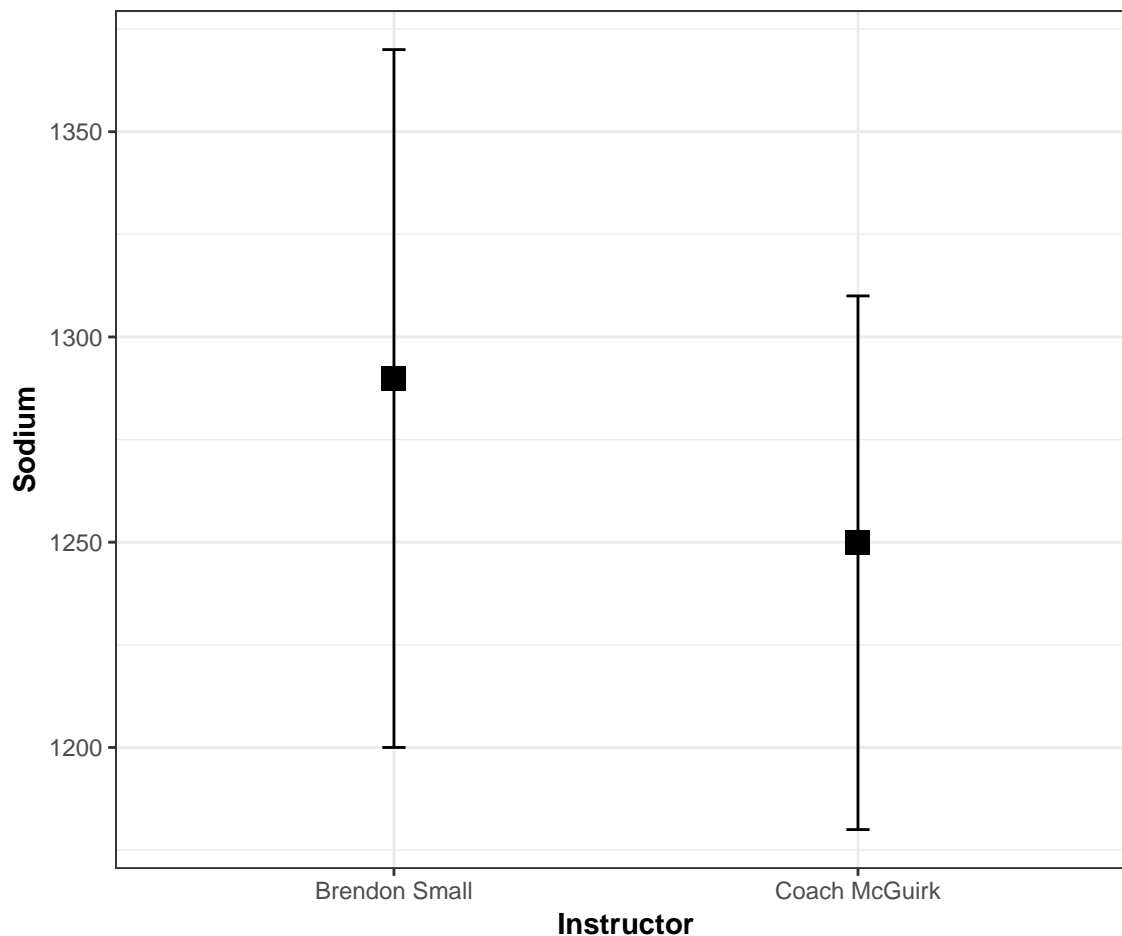
- gráfico com médias e intervalos de confiança separados por grupo, combinando recursos dos pacotes *ggplot2* e *rcompanion*:

```

SumMM <- rcompanion::groupwiseMean(Sodium ~ Instructor,
    data    = Dtfrm,
    conf    = 0.95,
    digits  = 3,
    traditional = FALSE,
    percentile = TRUE)

grf <- ggplot2::ggplot(SumMM, ggplot2::aes(x = Instructor, y = Mean)) +
  ggplot2::geom_errorbar(ggplot2::aes(ymin = Percentile.lower,
    ymax = Percentile.upper),
    width = 0.05, size = 0.5) +
  ggplot2::geom_point(shape = 15,
    size = 4) +
  ggplot2::theme_bw() +
  ggplot2::theme(axis.title = ggplot2::element_text(face = "bold")) +
  ylab(names(Dtfrm)[3])
print(grf)

```



Repare o que acontece com a linha

```
print(grf)
```

Há gráficos que aparecem quando sua função é chamada, e o que é armazenado na variável **grf** é um objeto com informações sobre o gráfico (e.g., `boxplot()`); em outros gráficos, o que retorna e é armazenado em **grf** é o gráfico propriamente dito (e.g., `lattice::xyplot()`). Sempre que for usar gráficos em seus *RScripts*, precisará testar caso a caso.

## - estatística inferencial

Definimos alfa:

```
alfa <- 0.05 # nivel de significancia adotado
```

O teste *t* a ser aplicado é de Satterthwaite (apesar do R exibir como teste de Welch), conforme as seguintes referências:

- Manuais do STATA
- SATTERTHWAIT, FE (1946) Approximate distribution of estimates of variance components. Biometrics Bulletin, 2(6): 110-114 e
- WELCH, BL (1947) The generalization of 'Student's' problem when several different population variances are involved. Biometrika, 34(1/2): 28-35.

## significância estatística

Para operacionalizar o teste, calculamos:

```
# separa os dois instrutores
SodiumBS <- subset(Dtfrm,select=Sodium,subset=Instructor=="Brendon Small",drop=TRUE)
SodiumCM <- subset(Dtfrm,select=Sodium,subset=Instructor=="Coach McGuirk",drop=TRUE)
# dados da amostra
nA <- sum(!is.na(SodiumBS))
nB <- sum(!is.na(SodiumCM))

# significancia estatistica
t_out <- t.test(Sodium ~ Instructor, data = Dtfrm)

# significancia pratica
t <- t_out$statistic # estatistica de teste t
df <- t_out$parameter # graus de liberdade
dfefic <- (df-min(nA,nB))/(nA+nB-2-min(nA,nB))
```

exibimos o resultado:

```
cat("\nTamanho das amostras: \n", sep="")
cat ("\tBrendon Small: n = ", nA, "\n", sep="")
cat ("\tCoach McGuirk: n = ", nB, "\n", sep="")

cat ("\n")
cat("Análise de significancia estatistica: valor-p\n")
cat("Teste t de Satterthwaite\n")
print(t_out)

cat("Eficiencia do numero de graus de liberdade =",dfefic,"\n\n")
```

Tamanho das amostras:

Brendon Small: n = 20

Coach McGuirk: n = 20

Análise de significancia estatistica: valor-p

Teste t de Satterthwaite

Welch Two Sample t-test

data: Sodium by Instructor

t = 0.76722, df = 34.893, p-value = 0.4481

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-67.91132 150.41132

sample estimates:  
 mean in group Brendon Small mean in group Coach McGuirk  
 1287.50 1246.25

Eficiencia do numero de graus de liberdade = 0.8273926

e construímos os gráficos:

```
# distribuicao t sob H0 (central: ncp = 0)
tH0 <- rt(1e6, df)
dtH0 <- density(tH0)
# distribuicao t sob H1, ncp = t
# ncp
F <- t^2 # estatistica de teste F de Fisher
eta2 <- F/(F+df)
f2 <- eta2/(1-eta2) # f de Cohen
ncp <- df*f2 # parametro de nao-centralidade: ncp = F
tH1 <- rt(1e6, df, ncp)
dtH1 <- density(tH1)

# media e dp das dist. t central e nao-central
mediaH0 <- 0
dpH0 <- sqrt(df/(df-2))
beta <- sqrt(df/2)*gamma((df-1)/2)/gamma(df/2)
mediaH1 <- ncp*beta
dpH1 <- sqrt((df*(1+ncp^2)/(df-2))-mediaH1^2)

cat("\nReferencial teorico:\n")
cat("\tvalor-p =", t_out$p.value, "\n")
cat("sob H0: distribuicao t com media = ", mediaH0, " e d.p. = ", dpH0, "\n", sep="")
cat("sob H1: distribuicao t com media = ", mediaH1, " e d.p. = ", dpH1, "\n", sep="")
cat("\n")
```

Referencial teorico:

valor-p = 0.4481092

sob H0: distribuicao t com media = 0 e d.p. = 1.029953

sob H1: distribuicao t com media = 0.6016735 e d.p. = 1.032641

e construímos os gráficos:

```
# graficos
for (g in 1:2)
{
  # limites de x
  min_x <- min(mediaH0-3*dpH0, mediaH1-3*dpH1)
  max_x <- max(mediaH0+3*dpH0, mediaH1+3*dpH1)
  if (g == 1)
  {
    plot(dtH0,
         main=paste("Teste t independente\ndf =", round(df,3), ", t =", round(t,5), ", alfa =", alfa),
         xlab="t",
         xlim=c(min_x,max_x),
         lwd=1, lty=1
    )
  }
}
```

```

}
if (g == 2)
{
  plot(dtH0,
       main=NA,
       xlab="t",
       xlim=c(min_x,max_x),
       lwd=1, lty=1
  )
}
qalfa <- qt(c(alfa/2,1-alfa/2),df,0)
abline(v=qalfa[1], lty = 3)
abline(v=qalfa[2], lty = 3)
if (g==1)
{
  abline(v=abs(t),lwd=1,lty=2)
  abline(v=-abs(t),lwd=1,lty=2)
  # area do valor p
  polx <- dtH0$x[dtH0$x>=abs(t)]; polx <- c(min(polx),polx,max(polx))
  poly <- dtH0$y[dtH0$x>=abs(t)]; poly <- c(0,poly,0)
  polygon(polx,poly,border="#EE802622",col="#EE802688",lwd=5)
  polx <- dtH0$x[dtH0$x<=-abs(t)]; polx <- c(min(polx),polx,max(polx))
  poly <- dtH0$y[dtH0$x<=-abs(t)]; poly <- c(0,poly,0)
  polygon(polx,poly,border="#EE802622",col="#EE802688",lwd=5)
}
if (g==2)
{
  # H1
  lines(dtH1,lwd=3,lty=1)
  # area alfa
  polx <- dtH0$x[dtH0$x<=qalfa[1]]; polx <- c(min(polx),polx,max(polx))
  poly <- dtH0$y[dtH0$x<=qalfa[1]]; poly <- c(0,poly,0)
  polygon(polx,poly,border="#a3261b22",col="#a3261b88",lwd=5)
  polx <- dtH0$x[dtH0$x>=qalfa[2]]; polx <- c(min(polx),polx,max(polx))
  poly <- dtH0$y[dtH0$x>=qalfa[2]]; poly <- c(0,poly,0)
  polygon(polx,poly,border="#a3261b22",col="#a3261b88",lwd=5)
  # area beta
  polx <- dtH1$x[dtH1$x>=qalfa[1] & dtH1$x<=qalfa[2]]; polx <- c(min(polx),polx,max(polx))
  poly <- dtH1$y[dtH1$x>=qalfa[1] & dtH1$x<=qalfa[2]]; poly <- c(0,poly,0)
  polygon(polx,poly,border="#4EB26522",col="#4EB26588",lwd=8)
}

# legenda
if (g==1)
{
  p_txt <- t_out$p.value;
  if (p_txt > 0.001)
  {
    p_txt <- round(p_txt,4)
  } else
  {
    p_txt <- format(format(p_txt, scientific = TRUE, digits = 4))
  }
}

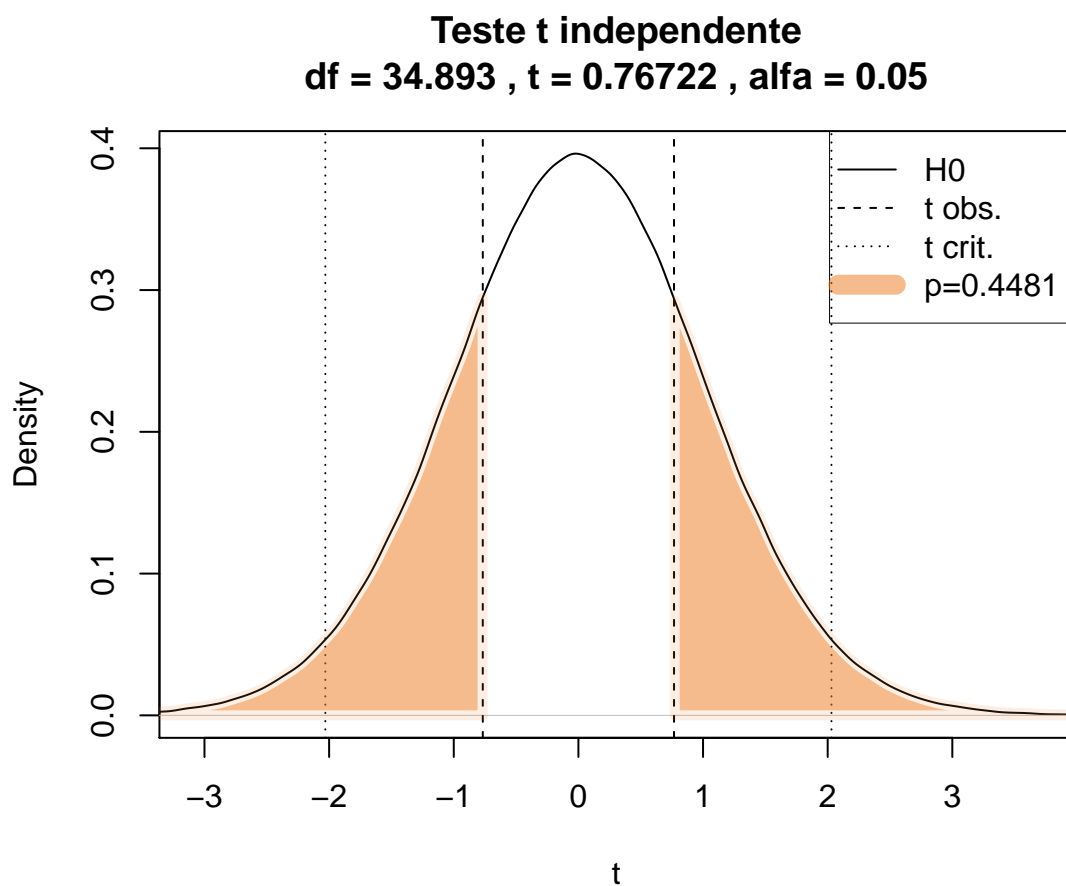
```

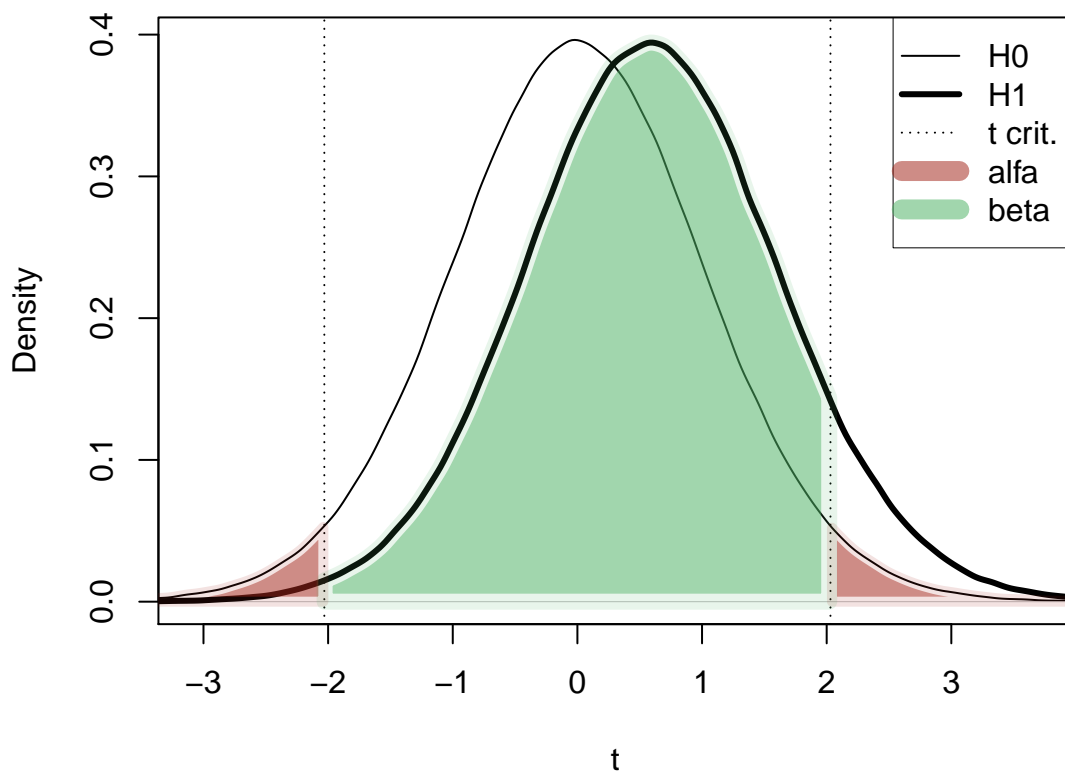


```

legend ("topright",
  c("H0", "t obs.", "t crit.", paste("p=", p_txt, sep="")),
  lwd=c(1,1,1,10),
  lty=c(1,2,3,1),
  pch=NA,
  col=c("black", "black", "black", "#EE802688"),
  box.lwd=0, bg="transparent")
}
if (g==2)
{
  legend ("topright",
    c("H0", "H1", "t crit.", "alfa", "beta"),
    lwd=c(1,3,1,10,10),
    lty=c(1,1,3,1,1),
    pch=NA,
    col=c("black", "black", "black", "#a3261b88", "#4EB26588"),
    box.lwd=0, bg="transparent")
}
}

```





Verifique e interprete a saída. Não rejeitamos  $H_0$ : não há elementos para afirmar que há diferença de resultado, quanto à ingestão de sódio, quando comparamos os dois grupos submetidos a diferentes programas educacionais.



Existe um outro teste  $t$  feito por *bootstrapping* (do package *MKinfer*):

```
library(MKinfer)
```

```
Attaching package: 'MKinfer'
```

```
The following object is masked from 'package:rcompanion':
```

```
quantileCI
```

```
t.boot <- MKinfer::boot.t.test(Sodium ~ Instructor, data=Dtfmr, R=10^6)
print(t.boot)
```

Bootstrapped Welch Two Sample t-test

```
data: Sodium by Instructor
bootstrapped p-value = 0.448
95 percent bootstrap percentile confidence interval:
-61.25 143.75
```

```
Results without bootstrap:
t = 0.76722, df = 34.893, p-value = 0.4481
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-67.91132 150.41132
sample estimates:
mean in group Brendon Small mean in group Coach McGuirk
1287.50 1246.25
```

A novidade são as primeiras linhas. O que aparece após “Results without bootstrap” é o teste de Welch visto acima.

## significância prática

Calculamos os tamanhos de efeito:

```
# Elis P (2010) The essential guide to effect sizes. Cambridge
if (eta2 <0.01) {mag_eta2<-c("Desprezível")}
if (eta2>=0.01 && eta2<0.06) {mag_eta2<-c("Pequeno")}
if (eta2>=0.06 && eta2<0.14) {mag_eta2<-c("Intermediário")}
if (eta2>=0.14) {mag_eta2<-c("Grande")}

d <- abs(t)/sqrt(1/((1/nA)+(1/nB)))
# Sawilowsky, S (2009) New effect size rules of thumb. Journal of Modern Applied Statistical Methods 8(
if (d<0.01) {mag_Cohen<-c("Desprezível")}
if (d>=0.01 && d<0.2) {mag_Cohen<-c("Muito pequeno")}
if (d>=0.2 && d<0.5) {mag_Cohen<-c("Pequeno")}
if (d>=0.5 && d<0.8) {mag_Cohen<-c("Intermediário")}
if (d>=0.8 && d<1.2) {mag_Cohen<-c("Grande")}
if (d>=1.2 && d<2) {mag_Cohen<-c("Muito grande")}
if (d>=2) {mag_Cohen<-c("Enorme")}
g <- d*(1-3/(4*df-1))
if (g<0.01) {mag_Hedges<-c("Desprezível")}
if (g>=0.01 && g<0.2) {mag_Hedges<-c("Muito pequeno")}
if (g>=0.2 && g<0.5) {mag_Hedges<-c("Pequeno")}
if (g>=0.5 && g<0.8) {mag_Hedges<-c("Intermediário")}
if (g>=0.8 && g<1.2) {mag_Hedges<-c("Grande")}
if (g>=1.2 && g<2) {mag_Hedges<-c("Muito grande")}
if (g>=2) {mag_Hedges<-c("Enorme")}

# selecao de modelo
R2aj <- (F-1)/((F-1)+df+1)
omega2 <- (F-1)/((F-1)+df+2)
```

e exibimos o resultado:

```
cat("Análise de significância prática: tamanho de efeito\n")
cat("\td de Cohen = ", d, " (", mag_Cohen, ")", "\n", sep="")
cat("\tg de Hedges = ", g, " (", mag_Hedges, ")", "\n", sep="")
cat("\teta^2 = R^2 = ", eta2, " (", mag_eta2, ")", "\n", sep="")

cat("\nSeleção de modelo:\n")
cat("\tR^2 ajustado = ", R2aj, "\n")
cat("\tomega^2 = ", omega2, "\n")
```

Análise de significância prática: tamanho de efeito

```
d de Cohen = 0.2426174 (Pequeno)
g de Hedges = 0.2373649 (Pequeno)
eta^2 = R^2 = 0.01658973 (Pequeno)
```

Seleção de modelo:

```
R^2 ajustado = -0.01159381
omega^2 = -0.01127601
```

## Conceitos adicionais

### Teste $t$ sem os dados brutos

É muito comum, em publicações, que somente tenhamos acesso às medidas-resumo (número de participantes, média, desvio-padrão e correlação). Nestes casos, os *RScripts* acima não são utilizáveis.

Para fazer os testes  $t$  de Welch, independentes ou relacionados (pareados), quando os dados brutos não estão disponíveis, criamos os seguintes scripts:

- t\_Welch\_unilateral\_esquerda.R
- t\_Welch\_unilateral\_direita.R
- t\_Welch\_bilateral.R
- t\_relacionado\_unilateral\_direita.R
- t\_relacionado\_bilateral.R
- t\_relacionado\_unilateral\_esquerda.R

Estude e aprenda a modificar estes *RScripts* para seu uso.

### teste $t$ com *bootstrapping* e tamanho de efeito

Um código que incorpora o teste  $t$  e um cálculo de  $d$  de Cohen (requer o package *lsr*) mais adequado à versão robusta do teste  $t$  de Welch é:

```
library(readxl)
library(MKinfer)
library(lsr)
Dados <- readxl::read_excel("Nutricao.xlsx")
MKinfer::boot.t.test(Sodium ~ Instructor, data=Dados, R=10^6)
```

Bootstrapped Welch Two Sample t-test

```
data: Sodium by Instructor
bootstrapped p-value = 0.4473
95 percent bootstrap percentile confidence interval:
-61.25 143.75
```

Results without bootstrap:

```
t = 0.76722, df = 34.893, p-value = 0.4481
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-67.91132 150.41132
```

sample estimates:

```
mean in group Brendon Small mean in group Coach McGuirk
1287.50 1246.25
```

```
# d de Cohen para o teste t de Welch
d <- lsr::cohensD(Sodium ~ Instructor, data=Dados, method="unequal")
# Sawilowsky, S (2009) New effect size rules of thumb.
# Journal of Modern Applied Statistical Methods, 8(2): 467-74.
if (0 <= d && d < 0.01) {dc <- "negligible"}
if (0.01 <= d && d < 0.2) {dc <- "very small"}
if (0.2 <= d && d < 0.5) {dc <- "small"}
if (0.5 <= d && d < 0.8) {dc <- "medium"}
if (0.8 <= d && d < 1.2) {dc <- "large"}
if (1.2 <= d && d < 2) {dc <- "very large"}
if (2 <= d && d < Inf) {dc <- "huge"}
cat("Cohen's d = ", d, " (" ,dc, " effect size)", sep="")
```

Cohen's d = 0.2426174 (small effect size)

## algumas manobras úteis

### construção de dois *boxplots*, lado a lado

Um exemplo caricato (apenas 4 medidas em cada grupo) é dado por:

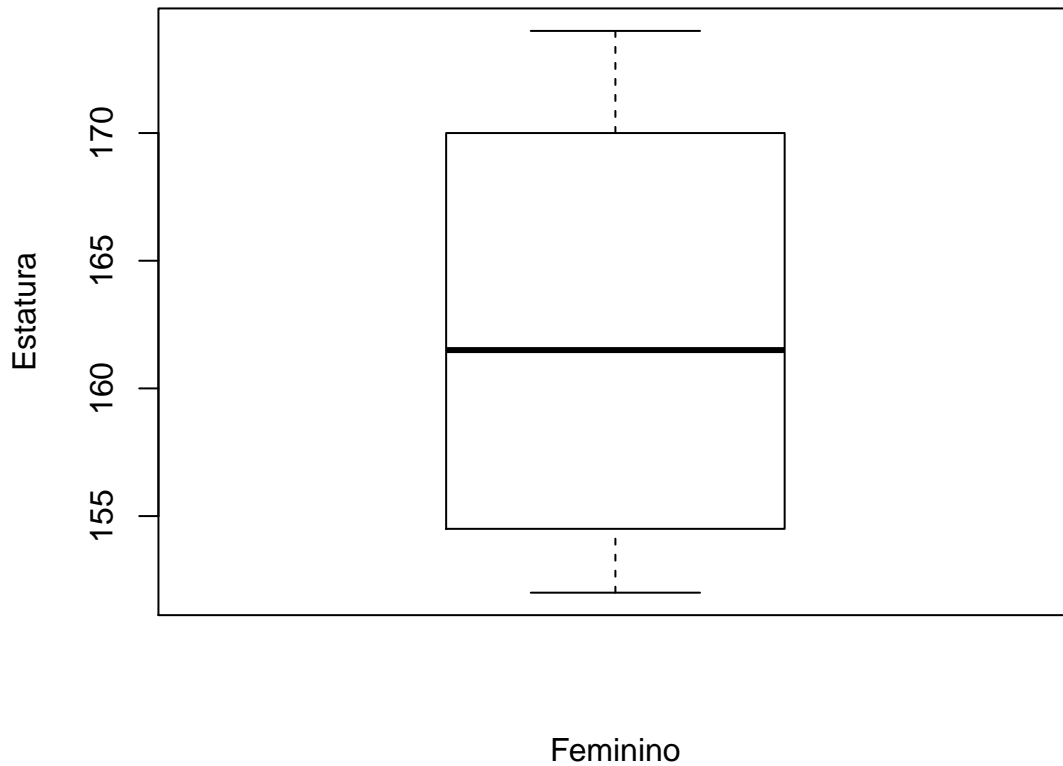
```
estadm <- c(176,183,173,191) # estatura de 4 homens
estatf <- c(157,152,174,166) # estatura de 4 mulheres
# Criando dois boxplot simples
boxplot(estadm, main="Boxplot de estatura de homens adultos (cm)",
        xlab="Masculino", ylab="Estatura")
```

### Boxplot de estatura de homens adultos (cm)



```
boxplot(estatf, main="Boxplot de estatura de mulheres adultas (cm)",  
        xlab="Feminino", ylab="Estatura")
```

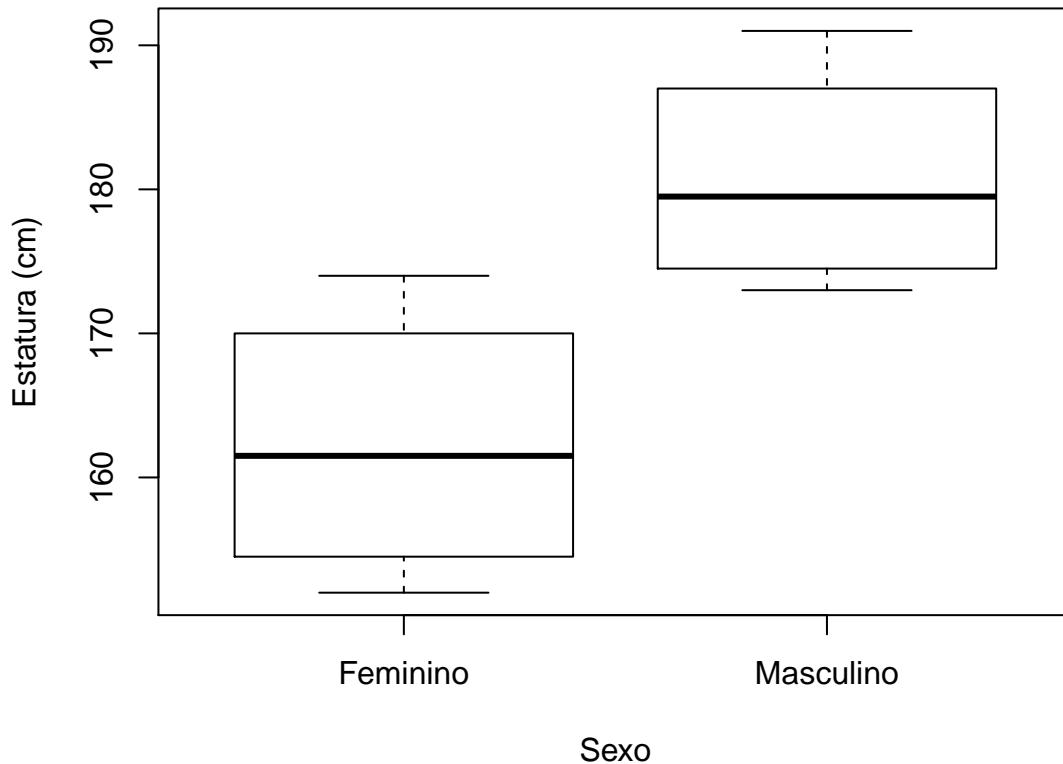
## Boxplot de estatura de mulheres adultas (cm)



Para juntar os dois gráficos, uma possibilidade é construir dois vetores de mesmo tamanho, um com os dados e outro que indique o sexo (Masculino ou Feminino):

```
estatm <- c(176,183,173,191)
estatf <- c(157,152,174,166)
estat <- c(estatm,estatf)
sexo <- rep(c("Masculino","Feminino"),each=4)
boxplot(estat~sexo, main="Boxplot de estatura de adultos (cm)", xlab="Sexo", ylab="Estatura (cm)")
```

## Boxplot de estatura de adultos (cm)



### guardar o gráfico para um arquivo

Experimente criar um arquivo .png:

```
png("estatura_homem_mulher.png")
boxplot(estat~sexo, main="Boxplot de estatura de adultos (cm)", xlab="Sexo", ylab="Estatura (cm)")
dev.off()
```

A função `png()` tem vários outros parâmetros. Veja a documentação do R (com `?png`): especialmente `width` e `height` são fundamentais. Há outros formatos gráficos disponíveis.

### guardar a saída textual em um arquivo

Você pode desviar a saída em tela para um arquivo texto. Experimente a função `sink()`:

```
sink("estatura_homem_mulher.txt")
resumo <- summary(estat)
cat ("Estatura de todos os individuos:\n")
print (resumo)
sink()
```

Precisa ser usada duas vezes: a primeira para abrir o arquivo e a segunda para fechá-lo - `sink()`, sem parâmetros - voltando a exibir as saídas na tela.

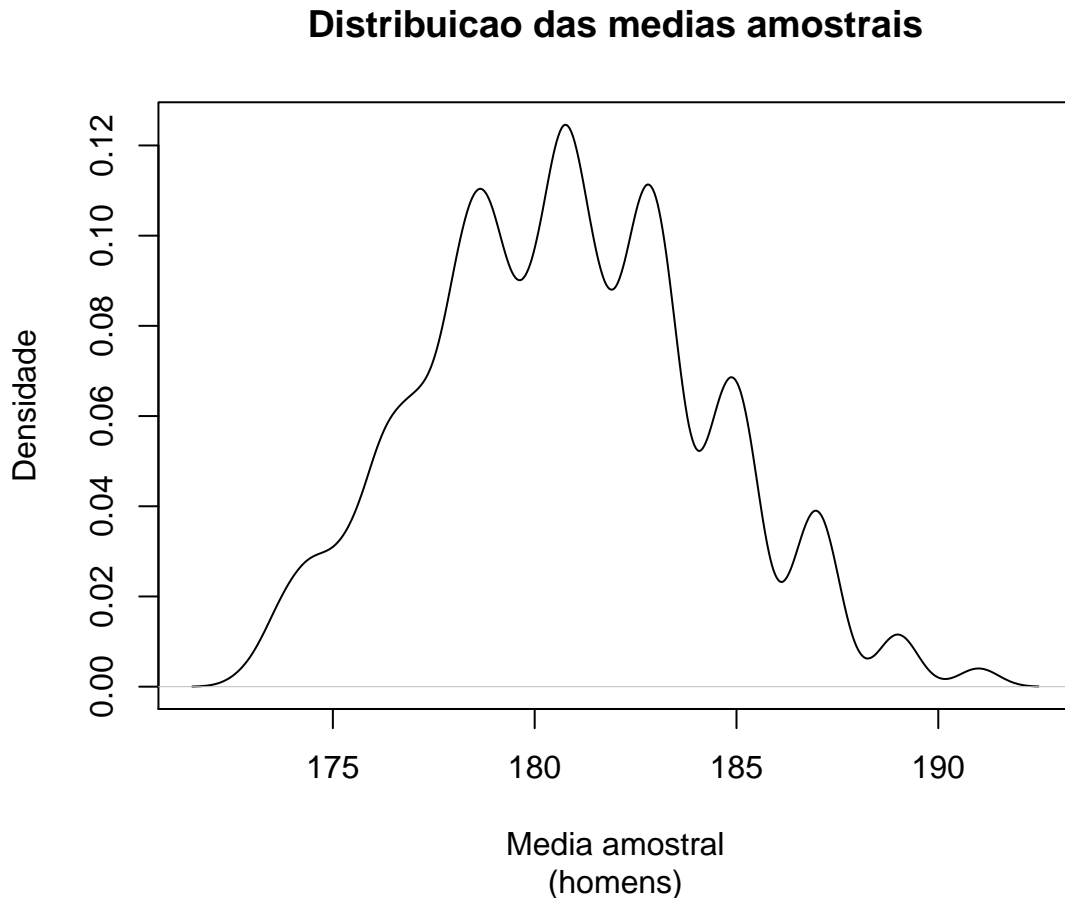


## intervalo de confiança robusto

Na linha de métodos robustos é possível estimar o intervalo de confiança 95% por bootstrapping ([https://en.wikipedia.org/wiki/Bootstrapping\\_\(statistics\)](https://en.wikipedia.org/wiki/Bootstrapping_(statistics))).

Por exemplo, para as estaturas dos homens deste exemplo:

```
rm <- replicate(1e4, mean(sample(estatm, replace=TRUE)))
plot(density(rm), main="Distribuicao das medias amostrais",
     sub="(homens)", xlab="Media amostral", ylab="Densidade")
```



```
qm <- quantile(rm, probs=c(0.025, 0.975))
cat (paste ("Intervalo de confianca 95% para os homens: [",qm[1],", ",qm[2], "]\n",sep=""))
```

Intervalo de confianca 95% para os homens: [174.5, 187.25]

A função `replicate(1e4, mean(sample(estatm, replace=TRUE)))` é uma forma de fazer reamostragem (`sample`), de uma população hipotética com média (`mean`) igual à da amostra de estatura dos homens com reposição (`replace=TRUE`), repetindo (`replicate`) o processo 10.000 vezes (`1e4`).

O intervalo de confiança de 95% é obtido com a função `quantile()`, localizando os valores que deixam 2.5% da área sob as caudas esquerda e direita desta distribuição de probabilidades.

## Sobre os métodos tradicionais



<http://unusual-cars.com/wp-content/uploads/2016/01/Ford-Model-T-1908.jpg>

### - o teste $t$ de Student



[https://en.wikipedia.org/wiki/Student%27s\\_t-test](https://en.wikipedia.org/wiki/Student%27s_t-test)

Este teste foi inicialmente publicado por em 1908 por William Sealy Gosset sob o pseudônimo de Student, e posteriormente aprimorado por Ronald Fisher, que introduziu os graus de liberdade.

Em R, o teste  $t$  default é executado com a correção de Satterthwaite (erroneamente exibido como Welch), mas é possível forçar o teste clássico, adicionando o parâmetro *var.equal*:

#### Two Sample $t$ -test

```
data: Sodium by Instructor
t = 0.76722, df = 38, p-value = 0.4477
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -67.59215 150.09215
sample estimates:
mean in group Brendon Small mean in group Coach McGuirk
                1287.50                1246.25
```

O que muda são os graus de liberdade, aqui correspondendo a um número inteiro igual ao tamanho da amostra subtraído de duas unidades (uma para cada grupo).

## - o (não) uso de testes não-paramétricos

Classicamente, havia várias condições para se poder executar o teste  $t$ . No entanto, em suas versões robustas e com tamanho de amostra suficiente, podemos dispensar os testes prévios de homocedasticidade e normalidade. Quando estas condições não são atendidas, os pesquisadores podem optar pelos métodos equivalentes não paramétricos (e.g., Wilcoxon e Mann-Whitney).

Há respaldo na literatura especializada, sugerindo que atualmente esta pode não ser a melhor opção:

---

Stat Papers (2011) 52:219–231  
DOI 10.1007/s00362-009-0224-x

---

### REGULAR ARTICLE

## The two-sample $t$ test: pre-testing its assumptions does not pay off

Dieter Rasch · Klaus D. Kubinger · Karl Moder

**Abstract** Traditionally, when applying the two-sample  $t$  test, some pre-testing occurs. That is, the theory-based assumptions of normal distributions as well as of homogeneity of the variances are often tested in applied sciences in advance of the tried-for  $t$  test. But this paper shows that such pre-testing leads to unknown final type-I- and type-II-risks if the respective statistical tests are performed using the same set of observations. In order to get an impression of the extension of the resulting misinterpreted risks, some theoretical deductions are given and, in particular, a systematic simulation study is done. As a result, we propose that it is preferable to apply no pre-tests for the  $t$  test and no  $t$  test at all, but instead to use the Welch-test as a standard test: its power comes close to that of the  $t$  test when the variances are homogeneous, and for unequal variances and skewness values  $|\gamma_1| < 3$ , it keeps the so called 20% robustness whereas the  $t$  test as well as Wilcoxon's  $U$  test cannot be recommended for most cases.

**Keywords** Pre-tests · Two-sample  $t$  test · Welch-test · Wilcoxon- $U$  test

RESEARCH ARTICLE

# Why Psychologists Should by Default Use Welch's $t$ -test Instead of Student's $t$ -test

Marie Delacre\*, Daniël Lakens† and Christophe Leys\*

When comparing two independent groups, psychology researchers commonly use Student's  $t$ -tests. Assumptions of normality and homogeneity of variance underlie this test. More often than not, when these conditions are not met, Student's  $t$ -test can be severely biased and lead to invalid statistical inferences. Moreover, we argue that the assumption of equal variances will seldom hold in psychological research, and choosing between Student's  $t$ -test and Welch's  $t$ -test based on the outcomes of a test of the equality of variances often fails to provide an appropriate answer. We show that the Welch's  $t$ -test provides a better control of Type 1 error rates when the assumption of homogeneity of variance is not met, and it loses little robustness compared to Student's  $t$ -test when the assumptions are met. We argue that Welch's  $t$ -test should be used as a default strategy.

**Keywords:** Welch's  $t$ -test; Student's  $t$ -test; homogeneity of variance; Levene's test; Homoscedasticity; statistical power; type 1 error; type 2 error

---



# Nonparametric methods for paired samples

U. Munzel\*

*Department of Medical Statistics, University of Göttingen,  
Humboldtallee 32, 37073 Göttingen, Germany*

The small sample and asymptotic properties of nonparametric tests for paired samples are examined. Linear rank statistics are compared with the paired t-test and the Wilcoxon-signed-rank test in simulation studies. From a minimax point of view the linear rank statistics turn out to be the best. Moreover, it is illustrated that the Wilcoxon-signed-rank test should not be used if it is not clear that the differences of the paired observations have a symmetric distribution.

*Key Words & Phrases:* Asymmetry, Behrens-Fisher problem, paired t-test, rank transform, ties, Wilcoxon-signed-rank test.