Bioestatística - Teste de hipóteses

Luciane Alcoforado

Universidade Federal Fluminense
Instituto de Matemática e Estatística
Departamento de Estatística

1 Teste T bilateral com uma amostra

1.1 Exemplo 1:

Um biólogo deseja estudar o efeito de certa substância no tempo de reação de seres vivos a um certo tipo de estímulo. Um experimento é desenvolvido com cobaias que são inoculadas com a substância e submetidas a um estímulo elétrico com seus tempos de reação (em segundos) anotados. Os tempos obtidos foram: 9.1, 9.3, 7.2, 7.5, 13.3, 10.9, 7.2, 9.9, 8.0, 8.6. Admitindo- se que o tempo de reação segue distribuição normal com média 7.6 segundos, verificar se o tempo médio sofre alteração por influência da substância, no nível de 5% de significância.

1.2 Formulação

H_0: tempo médio de reação é de 7.6 segundos

H_1: tempo médio de reação é **diferente** de 7.6 segundos

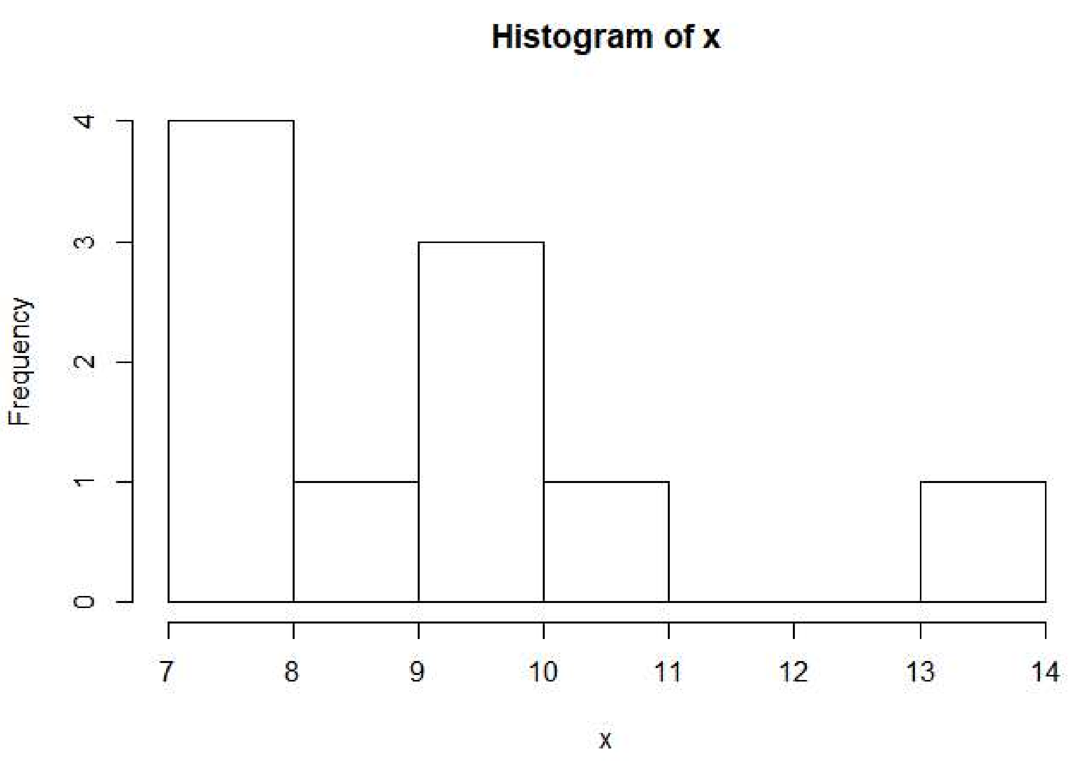
Nível de significância adotado: α = 0.05

Tamanho da amostra em teste: $n = 10$

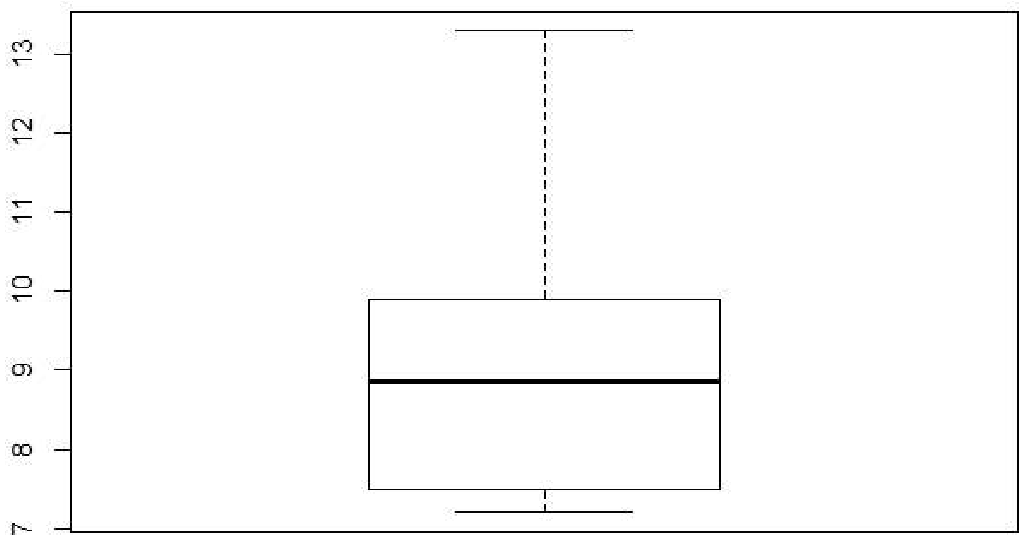
1.3 Comandos no R

1.3.1 Preliminares

```
#Primeiro criamos um vetor com os tempos
x=c(9.1,9.3,7.2,7.5,13.3,10.9,7.2,9.9,8,8.6)
#fazemos um histograma e boxplot
hist(x)
```




```
boxplot(x)
```




```
#os dados não são simétricos.
#vamos usar o teste de shapiro para confirmar a hipótese de normalidade.
#Se p-valor > 0.05 a hipótese de normalidade está confirmada (é o que queremos);
#caso contrário estaremos violando o pressuposto do teste t.

shapiro.test(x)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.89, p-value = 0.2
```

O pressuposto da normalidade para a variável em análise é atendida de acordo com o teste de shapiro (p-valor > 0.05 confirma a hipótese de normalidade)

1.3.2 Aplicando o teste

```
#Vamos aplicar um teste t

#H0:mu=7.6 e H1: mu!=7.6

alfa=0.05 #nivel de significância
```

```
t.test(x, mu=7.6)
```

```
##
##  One Sample t-test
##
## data:  x
## t = 2.5, df = 9, p-value = 0.03
## alternative hypothesis: true mean is not equal to 7.6
## 95 percent confidence interval:
##   7.734 10.466
## sample estimates:
## mean of x
##       9.1
```

```
#Regra de decisão
teste=t.test(x, mu=7.6)
teste$p.value
```

```
## [1] 0.03472
```

```
teste$p.value>alfa #para aceitar H0 resposta deve ser TRUE
```

```
## [1] FALSE
```

A conclusão do teste foi de rejeitar a hipótese nula ao nível de significância de 5%.

Neste contexto quer dizer que o tempo de reação é significativamente diferente de 7.6 s, com nível de significância de 5%.

2 Teste T unilateral com uma amostra

2.1 Exemplo 2:

Uma firma comercial sustenta que seus cigarros contêm não mais que 30 mg de nicotina. Uma amostra de 25 cigarros forneceu os seguintes valores

```
##  [1] 31.7 31.2 32.3 31.0 33.4 31.1 33.6 38.3 29.2 32.5 29.9 32.0 30.4 37.2
## [15] 29.5 30.0 36.4 29.0 33.7 33.0 34.7 30.8 28.8 30.0 33.5
```

Considerando a distribuição normal de probabilidades para a quantidade de nicotina, no nível de 5% de significância, os dados contestam ou não a afirmação do fabricante?

2.2 Formulação

H_0: quantidade de nicotina do cigarro é de 30 mg

H_1: quantidade de nicotina do cigarro é *maior que* 30 mg

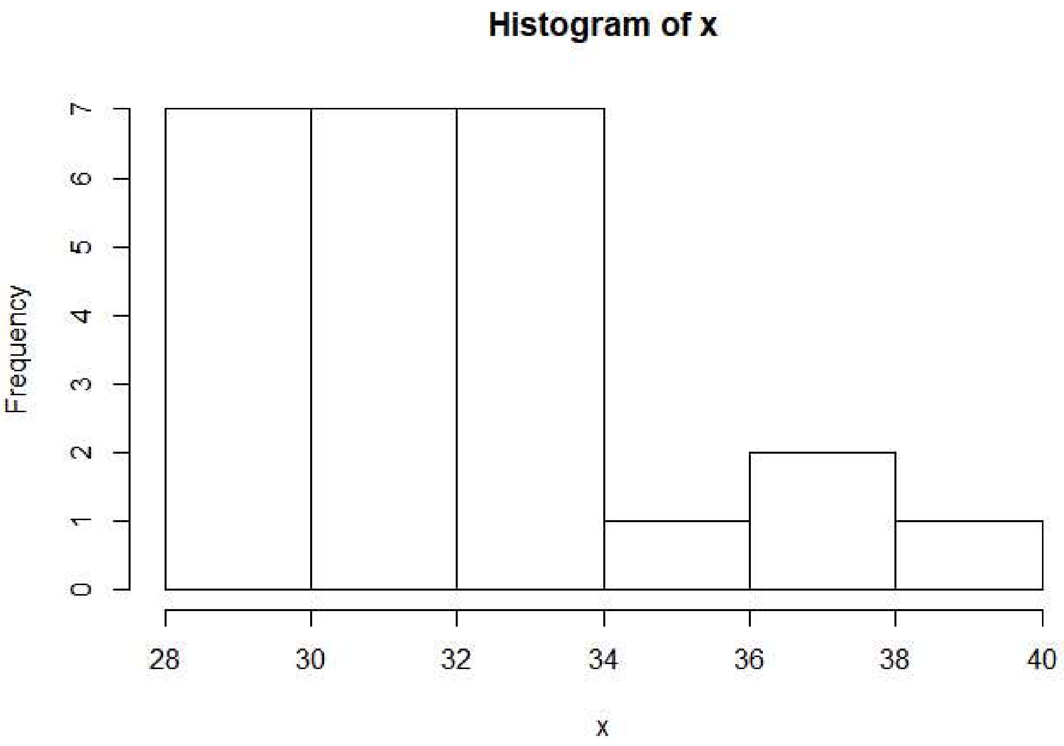
Nível de significância adotado: α = 0.05

Tamanho da amostra em teste: $n = 25$

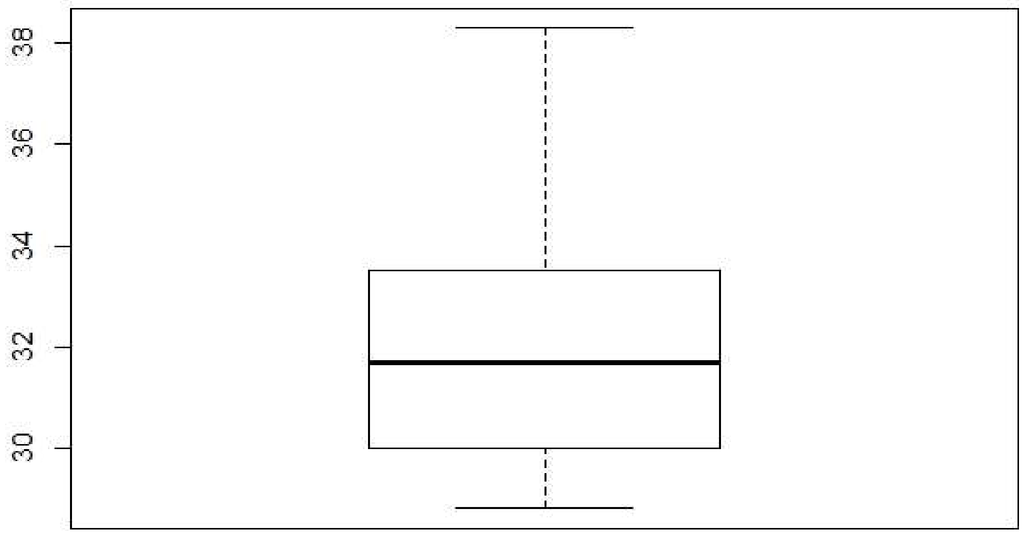
2.3 Comandos no R

2.3.1 Preliminares

```r
#Primeiro criamos um vetor com os valores
x= c(31.7, 31.2, 32.3, 31.0, 33.4, 31.1, 33.6, 38.3, 29.2, 32.5,
29.9, 32.0, 30.4, 37.2, 29.5, 30.0, 36.4, 29.0, 33.7, 33.0,
34.7, 30.8, 28.8, 30.0, 33.5)
#fazemos um histograma e boxplot e o teste de shapiro
hist(x)
```




```
boxplot(x)
```




```
shapiro.test(x)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.93, p-value = 0.08
```

De acordo com o teste de shapiro, a hipótese de normalidade é __________ pois p-valor ____ 0.05.

2.3.2 Aplicando o teste

```
#Vamos aplicar um teste t

#H0:mu=30 e H1: mu>30

alfa=0.05 #nivel de significância

t.test(x, mu=30, alternative = "greater")
```

```
##
##  One Sample t-test
##
## data:  x
## t = 4.2, df = 24, p-value = 0.0002
## alternative hypothesis: true mean is greater than 30
## 95 percent confidence interval:
##  31.25   Inf
## sample estimates:
## mean of x
##     32.13
```

```
#Regra de decisão
teste=t.test(x, mu=30)
teste$p.value
```

```
## [1] 0.0003426
```

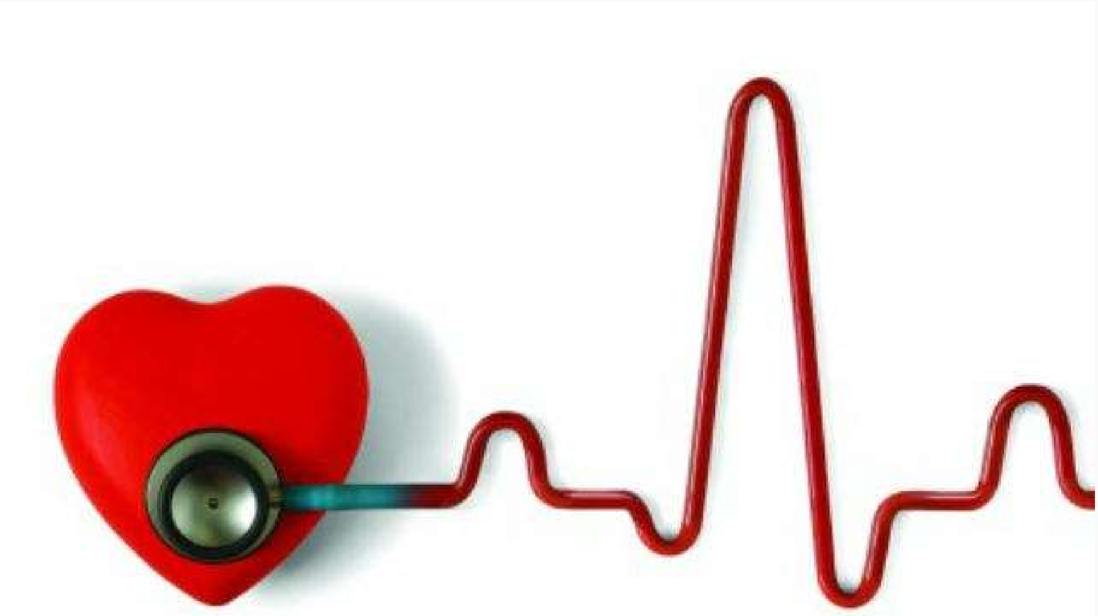
```
teste$p.value>alfa #para aceitar H0 resposta deve ser TRUE
```

```
## [1] FALSE
```

Conclusão do teste: rejeitar H_0, isto é, quantidade de nicotina é significativamente maior do que 30 mg, com nível de significância de 5%.

3 Teste T para a Comparação de Médias de Duas Populações Normais Independentes com Variâncias Desconhecidas e Iguais

3.1 Exemplo 3:



(Ref. pag 89 - Padovani) Um estudo sobre hipertensão induzida por gravidez considerou um grupo de 23 mulheres com essa disfunção recebendo baixa dose de aspirina e um segundo, com 24 mulheres nas mesmas condições, que receberam placebo. A pressão sanguínea arterial dos grupos está descrita no quadro a seguir.

```
grupo=c(rep("A",23), rep("B",24))
set.seed(260);pressao=round(c(rnorm(23,109,7),rnorm(24,111,8)),1)
dados=data.frame(grupo,pressao)
dados
```

```
##    grupo pressao
## 1      A   109.1
```

```
## 2      A   101.0
## 3      A   109.6
## 4      A   103.9
## 5      A   109.2
## 6      A   107.3
## 7      A   106.4
## 8      A   105.9
## 9      A    98.9
## 10     A    99.4
## 11     A   116.1
## 12     A   131.4
## 13     A   112.9
## 14     A   118.8
## 15     A    99.2
## 16     A   108.4
## 17     A   113.5
## 18     A    94.2
## 19     A   116.8
## 20     A   119.1
## 21     A   103.3
## 22     A   111.2
## 23     A   118.9
## 24     B   103.2
## 25     B   105.6
## 26     B   101.1
## 27     B   106.6
## 28     B   108.1
## 29     B   100.9
## 30     B   114.9
## 31     B   121.8
## 32     B   117.9
## 33     B   115.0
## 34     B   104.4
## 35     B   104.2
## 36     B   109.1
## 37     B   101.8
## 38     B   114.1
## 39     B   112.7
## 40     B   109.9
## 41     B   108.7
## 42     B   118.5
## 43     B    96.9
## 44     B   111.5
## 45     B   118.4
## 46     B   100.7
## 47     B   119.3
```

No nível de significância 5%, os grupos diferem quanto à pressão arterial sanguínea? As hipóteses gerais do teste t de Student para amostras independentes são detalhadas a seguir.

3.2 Formulação

H_0: pressão arterial média do grupo 1 é igual a do grupo 2

H_1: pressão arterial média do grupo 1 é *diferente* do grupo 2

Nível de significância adotado: α = 0.05

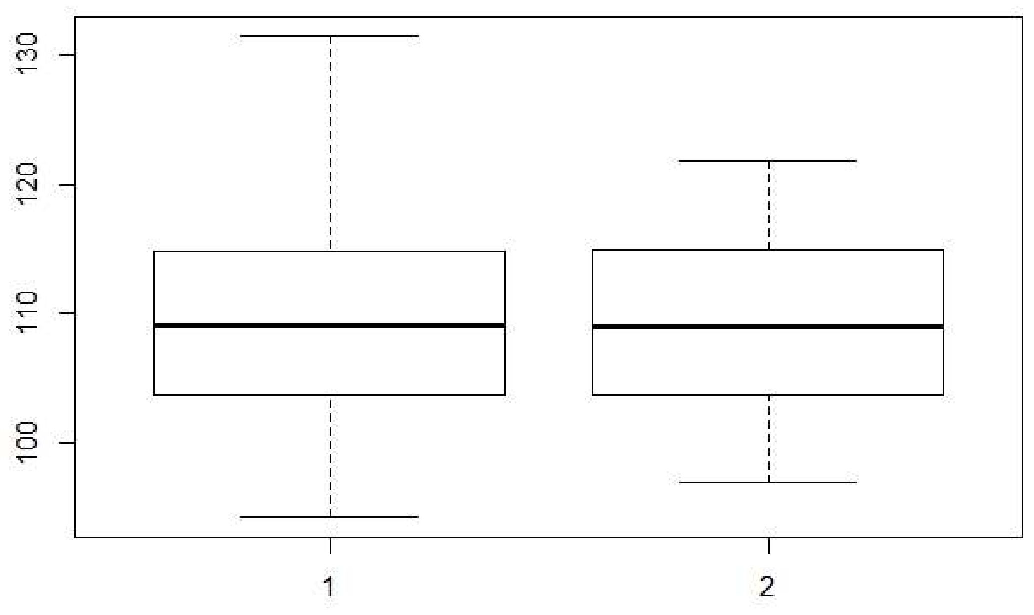
Tamanho da amostra em teste: $n_1 = 23$ e $n_2 = 24$

3.3 Comandos no R

3.3.1 Preliminares

```
x1=dados[1:23,2]
x2=dados[24:47,2]

boxplot(x1,x2)
```




```
shapiro.test(x1)
```

```
##
##  Shapiro-Wilk normality test
##
```

```
## data:  x1
## W = 0.97, p-value = 0.7
```

```
shapiro.test(x2)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x2
## W = 0.96, p-value = 0.5
```

```
#vamos testar a homogeneidade da variância
var.test(x1,x2)
```

```
##
##  F test to compare two variances
##
## data:  x1 and x2
## F = 1.5, num df = 22, denom df = 23, p-value = 0.4
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.6344 3.4582
## sample estimates:
## ratio of variances
##              1.475
```

```
#pvalor >0.05 aceita a homogeneidade da variância

#Aceitamos a homogeneidade da variância
#Agora vamos realizar o teste t com variâncias iguais
```

De acordo com o teste de shapiro, a hipótese de normalidade é __________ pois p-valor ____ 0.05.

De acordo com o teste da variância, a homogeneidade da variância é ____________ pois p-valor ______ 0.05.

3.3.2 Aplicando o teste

```
#H0:m1-m2=0; H1:m1 difere m2


t.test(x1,x2)
```

```
##
##  Welch Two Sample t-test
##
## data:  x1 and x2
## t = -0.027, df = 43, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.667  4.544
## sample estimates:
## mean of x mean of y
##     109.3     109.4
```

```
#pvalor >0.05 aceita-se a H0

#Logo conclui-se não haver diferença significativa entre as médias.
```

Como p-valor foi maior do que o nível de significância (p-valor > 0.05) conclui-se que a pressão arterial sanguínea do grupo 1 que tomou aspirina não difere significativamente do grupo 2 que tomou placebo.

4 Teste T para a Comparação de Médias de Duas Populações Normais Independentes com Variâncias Desconhecidas e Desiguais

4.1 Exemplo 4:

(Ref. pag 91 - Padovani) Acredita-se que o nível médio de carboxihemoglobina dos fumantes seja mais alto do que o nível médio dos não fumantes. A seguir são apresentados os resultados amostrais de dois grupos. Como o conjunto de dados é muito grande, vamos exibir apenas os primeiros valores.

```
grupo=c(rep("NF",121), rep("F",75))
set.seed(2);carbox=round(c(rnorm(121,1.8,1),rnorm(75,4.1,1.6)),1)
carbox[which(carbox<0)]=1
dados=data.frame(grupo,carbox)
head(dados)
```

```
##   grupo carbox
## 1    NF    0.9
## 2    NF    2.0
## 3    NF    3.4
## 4    NF    0.7
```

```
## 5     NF     1.7
## 6     NF     1.9
```

No nível de significância 5%, o grupo de fumantes apresenta nível médio de carboxihemoglobina maior do que no grupo de não fumantes? As hipóteses gerais do teste t de Student para amostras independentes são detalhadas a seguir.

4.2 Formulação

H_0: **nível médio de carboxihemoglobina dos não fumantes é igual a dos fumantes**

H_1: **nível médio de carboxihemoglobina dos não fumantes é *menor do que (less)* a dos fumantes**

Nível de significância adotado: α **= 0.05**

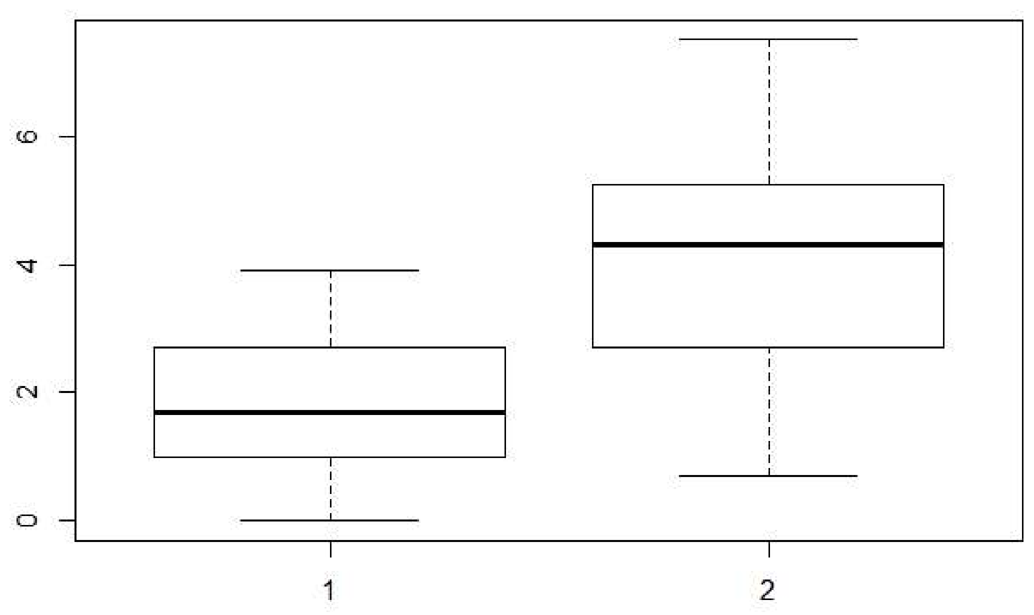
Tamanho da amostra em teste: $n_1 = 121$ **e** $n_2 = 75$

4.3 Comandos no R

4.3.1 Preliminares

```
x1=dados[1:121,2]
x2=dados[122:196,2]

boxplot(x1,x2)
```



```r
shapiro.test(x1)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x1
## W = 0.95, p-value = 0.0004
```

```r
shapiro.test(x2)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x2
## W = 0.97, p-value = 0.07
```

```r
#vamos testar a homogeneidade da variância
var.test(x1,x2)
```

```
##
##  F test to compare two variances
##
## data:  x1 and x2
## F = 0.43, num df = 120, denom df = 74, p-value = 0.00003
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.2787 0.6358
## sample estimates:
## ratio of variances
##             0.4253
```

```r
#pvalor >0.05 aceita a homogeneidade da variância
#pvalor <0.05 rejeita a homogeneidade da variância
```

De acordo com o teste de shapiro, a hipótese de normalidade é __________ pois p-valor ____ 0.05.

De acordo com o teste da variância, a homogeneidade da variância é ____________ pois p-valor ______ 0.05.

Como rejeitamos a homogeneidade da variância, devemos levar isso em conta na aplicação do teste t, acrescentando o argumento var.equal = F.

4.3.2 Aplicando o teste

```
#m1 é o nível médio de carboxihemoglobina dos não fumantes e m2 dos fumantes
#H0:m1-m2=0; H1:m1 - m2 < 0

#Agora vamos realizar o teste t unilateral com variâncias diferentes

t.test(x1,x2, var.equal = F, alternative = "less")
```

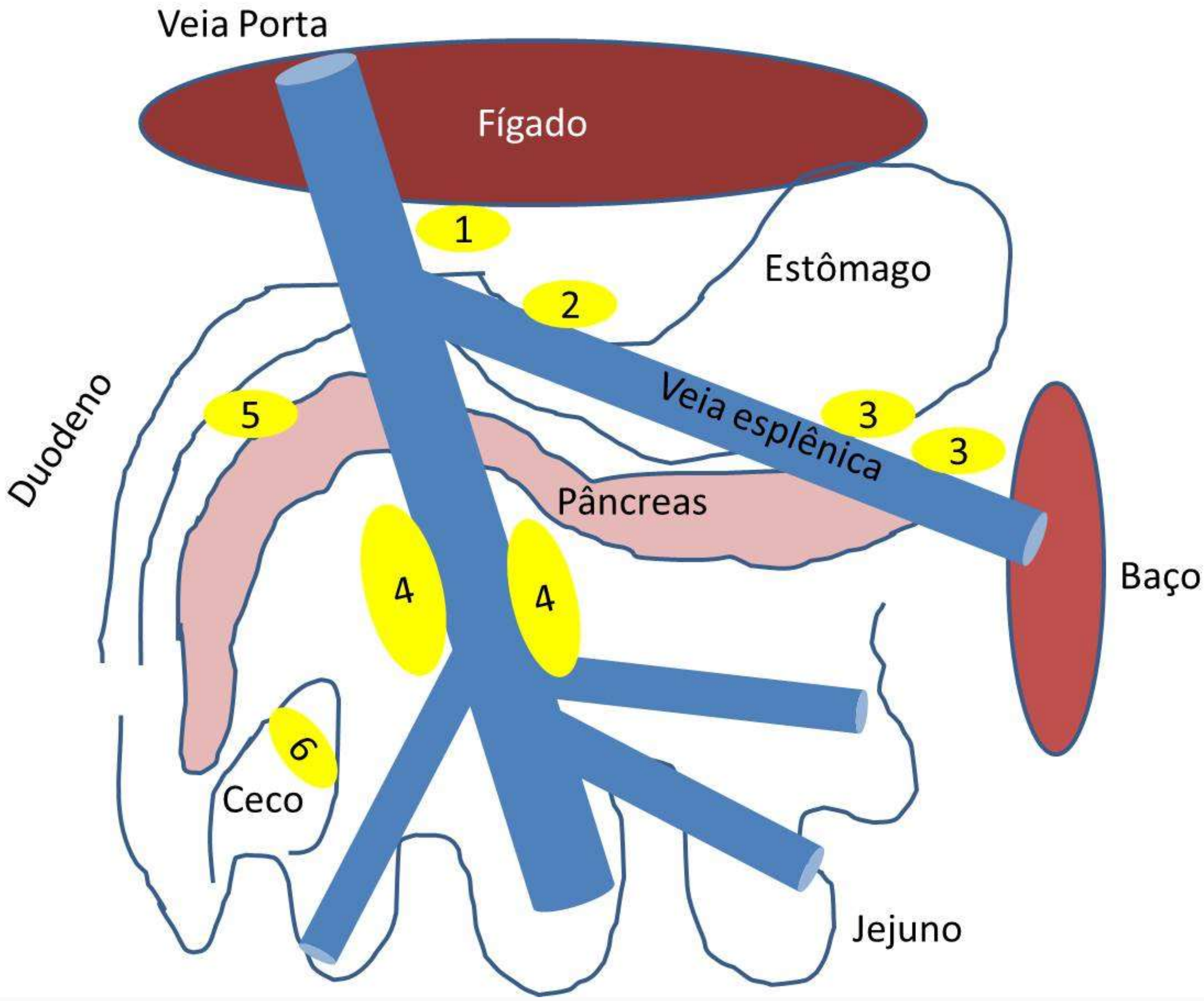
```
##
##  Welch Two Sample t-test
##
## data:  x1 and x2
## t = -10, df = 110, p-value <0.0000000000000002
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##    -Inf -1.797
## sample estimates:
## mean of x mean of y
##     1.910     4.049
```

```
#pvalor >0.05 aceita-se a H0
#pvalor <0.05 rejeita-se a H0
```

Como p-valor foi menor do que o nível de significância (p-valor < 0.05) conclui-se que o nível de carboxihemoglobina entre não fumantes é significativamente menor do que o grupo de fumantes.

5 Teste para a Comparação de Médias de Duas Populações Normais Dependentes (Amostras

Pareadas, Amostras Emparelhadas) e Variâncias Desconhecidas



5.1 Exemplo 5:

(Ref. pag 92 - Padovani) O calibre da veia esplênica é, em média, o mesmo, antes e depois da oclusão da veia porta? Utilizaremos os seguintes dados de cães.

```
antes=c(75, 50, 50, 60, 50, 70)
depois=c(85, 75, 70, 65, 60, 90)
```

5.2 Formulação

H_0: **Não há diferença do calibre da veia esplênica antes e depois da oclusão da veia porta**

H_1: **Há diferença do calibre da veia esplênica antes e depois da oclusão da veia porta**

Nível de significância adotado: α **= 0.05**

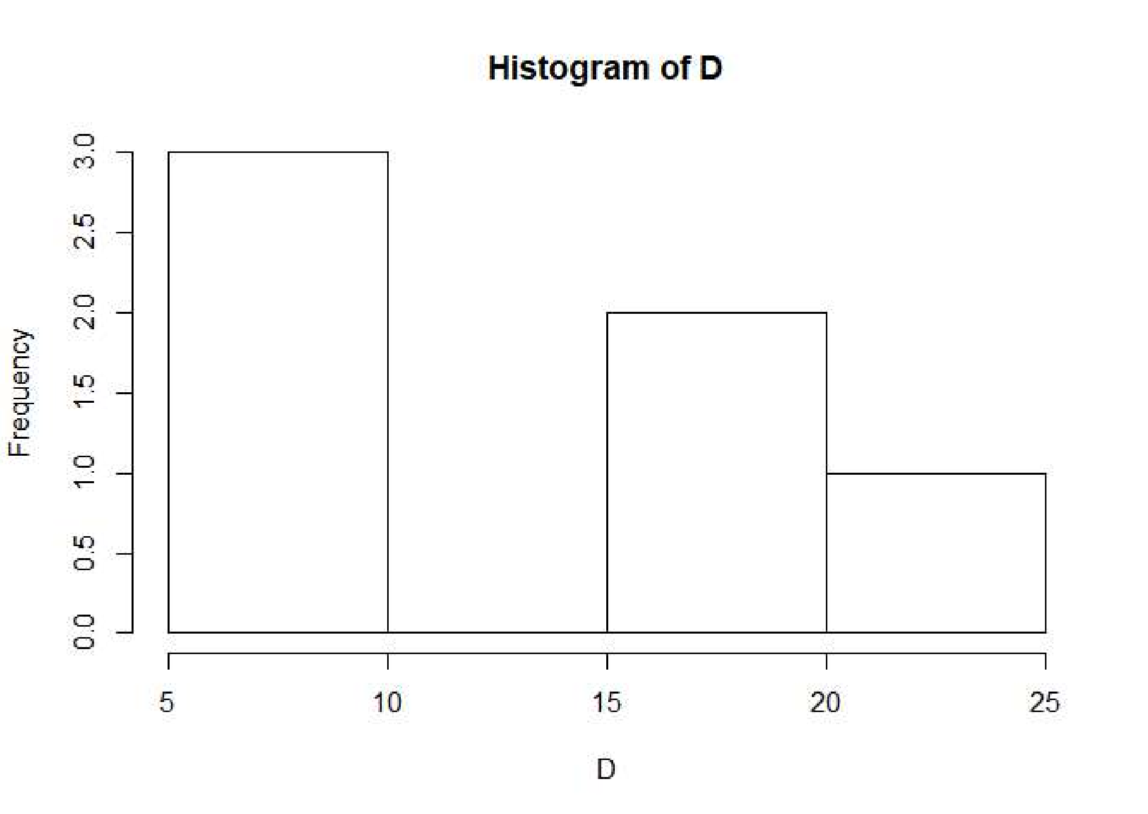
Tamanho da amostra em teste: $n_1 = 6$ **e** $n_2 = 6$

5.3 Comandos no R

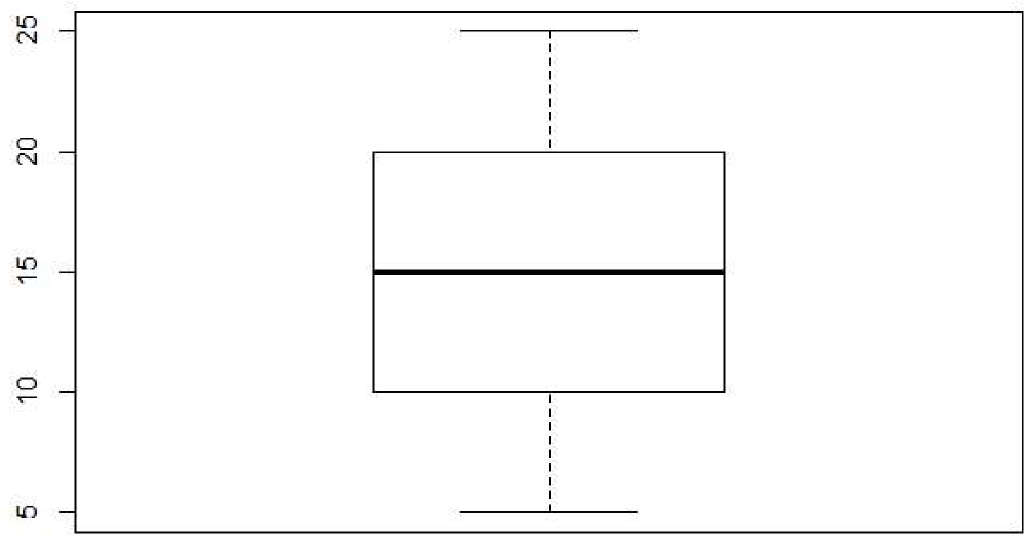
5.4 Preliminares

```
x1=antes
x2=depois
D=x2-x1

hist(D)
```




```
boxplot(D)
```




```
shapiro.test(D)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  D
## W = 0.91, p-value = 0.5
```

```
#D é a diferença entre as medidas antes e depois
#H0: D = 0; H1: D diferente de 0
```

De acordo com o teste de shapiro, a hipótese de normalidade é __________ pois p-valor ____ 0.05.

5.5 Aplicando o teste

```
t.test(x1,x2, paired = T)
```

```
##
##  Paired t-test
##
## data:  x1 and x2
## t = -4.7, df = 5, p-value = 0.005
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -23.129  -6.871
## sample estimates:
## mean of the differences
##                     -15
```

```
#pvalor >0.05 aceita-se a H0
#pvalor <0.05 rejeita-se a H0
```

Como p-valor < 0.05 rejeita-se a hipótese nula. Logo a conclusão é que há diferença significativa entre o calibre da veia esplênica antes e após a oclusão da veia porta.

6 Trabalho para ser iniciado em 06/11/18 e ser entregue via conexão uff até o dia 20/11/18, pdf colocando o nome dos participantes do trabalho.

O trabalho deve seguir os passos apresentados nos exemplos anteriores, observando-se o teste adequado em cada situação e ao final deve apresentar a análise.

6.1 Exercício 1.

Selecionam-se aleatoriamente oito comprimidos diferentes de cada um de dois remédios antigripais concorrentes, e faz-se um teste do conteúdo de acetaminofena em cada um. Os resultados, em mg, são os seguintes:

Dozenol	472	487	506	511	496	524	504	501
Niteze	562	512	494	528	552	508	496	532

Considerando o nível de 5% significância, teste a afirmação de que a quantidade média de acetaminofena é a mesma nas duas marcas.

6.2 Exercício 2.

Um médico deseja saber se uma certa droga reduz a pressão arterial média. Para isso mediu a pressão arterial de 10 voluntários, antes e após a ingestão da droga, obtendo os dados do quadro a seguir.

Voluntários	A	B	C	D	E	F	G	H	I	J
Antes	68	80	90	74	75	69	66	83	87	83
Depois	60	71	88	72	71	70	66	78	85	76

Teste se existe significância (5%) estatística de que a droga realmente reduz a pressão arterial média?