Testes de Hipótese

Danilo A C Souto

Pacotes utilizados

```
#install.packages("TeachingDemos")
library(TeachingDemos)
```

Resumo

Sempre há duas hipóteses sendo:

H0 a hipótese Nula e HA a hipótese Alternativa. (também pode ser vista como H1)

$$\mu = \text{ Média Populacional}$$
$$\bar{X} = \text{ Média Amostra}$$
$$\sigma = \text{ Desvio Padrão Populacional}$$
$$s = \text{ Desvio Padrão Amostral}$$
$$n = \text{ Tamanho da amostra}$$

$$H_0 : \bar{X} = \mu \text{ Hipótese Nula}$$
$$H_A : \bar{X} \neq \mu \text{ Hipótese Alternativa}$$

$$c = \text{ Índice de confiança}$$
$$p = \text{ Significância}$$
$$c = 1 - p$$

Passos para execução

Os passos para resolver o problemas são quase sempre os mesmos

1) Formula-se as hipóteses;
2) Selecione o teste apropriado;
3) Veja se será **Unilateral** ou **Bilateral**;
4) Calcula a estatística de teste;
5) De acordo com o nível de **confiança** desejado, compare o resultado na tabela apropriada;
6) Elabore uma resposta.

Agora quanto aos testes:

Teste Z

- Número de amostras > 30; ou
- Desvio Padrão Populacional conhecido;

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Caso contrário:

Teste T

- Número de amostras < 30; ou
- Desvio Padrão *desconhecido*;

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

Exemplo 1

O instituto de normas e medidas determina que um determinado produto químico tem que ter a quantidade de 72mg. Fabricantes além da tolerância de (5%) devem ser multados. Sabe-se que o desvio da linha é de 2mg Foram coletadas as seguintes amostras:

74.3 73.5 73.5 73.2 72.5 74.9 73.6 71.1 73.8 72.6

Deve-se multar o fabricante?

Resposta 1

1) Formula-se as hipóteses;

$$\sigma = 2mg$$
$$\mu = 72mg$$
$$H_0 : \bar{X} = \mu$$
$$H_A : \bar{X} \neq \mu$$
$$c = 95\%$$

2) Selecione o teste apropriado;

Teste Z = pois o desvio padrão é populacional

3) Veja se será **Unilateral** ou **Bilateral**;

Bilateral

4) Calcula a estatística de teste;

```r
mu <- 72
sigma<- 2
amostras1 <- c(74.3, 73.5, 73.5, 73.2, 72.5, 74.9, 73.6, 71.1, 73.8, 72.6)
media <- mean(amostras1)
n <- length(amostras1)

z <- (media - mu)/(sigma/sqrt(n))
```

O valor de Z = 2.0554805

5) De acordo com o nível de **confiança** desejado, compare o resultado na tabela apropriada;

k2 = 0,975

k1 = 0,025

Como a tabela é simetrica |k| = 1,96

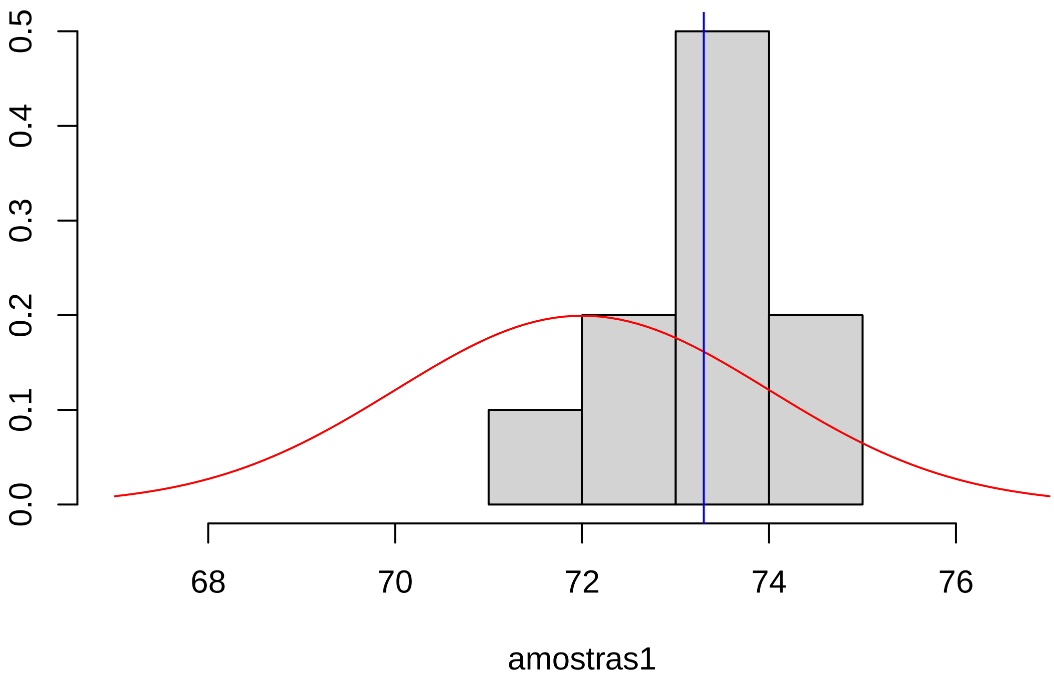
6) Elabore uma resposta.

Como Z é maior que k,

$$|Z| > |k|$$
$$2,05 > 1,96$$

Rejeita-se H0. A média da amostra e da população não são iguais. O fabricante pode vir a ser multado se não se adequar.

```r
hist(amostras1, freq = F, ylim = c(0,0.5) ,xlim=c(67,77), main = "", ylab = "")
curve(expr = dnorm(x, mean= mu , sd = sigma), col = "red"  , from = 67, to = 77, add = T)
abline(v=media, col= "blue")
```



Exemplo 2

O instituto de normas e medidas determina que um determinado produto químico tem que ter a quantidade de 72mg. Fabricantes com desvios fora de 2mg (5%) devem ser multados. Foram coletadas as seguintes amostras

70.3 71.3 76.7 72.2 72.4 77.1 73.4 68.2 69.9 70.7 Deve-se multar o fabricante?

```r
amostras2 <- c(70.3, 71.3, 76.7, 72.2, 72.4, 77.1, 73.4, 68.2, 69.9, 70.7)
mu <- 72
sigma <- 2
media <- mean(amostras2)

cat("\n x_bar: ")
```

```
##
##  x_bar:
```

```r
media
```

```
## [1] 72.22
```

```r
n <- length(amostras2)
#teste Z
z <- (media - mu)/(sigma/sqrt(n))
#z
cat("\n z: ")
```

```
##
##  z:
```

```r
z
```

```
## [1] 0.3478505
```

```r
#conf
cat ("\nintervalo de confiança: ")
```

```
##
## intervalo de confiança:
```

```r
abs(1.96 )
```

```
## [1] 1.96
```

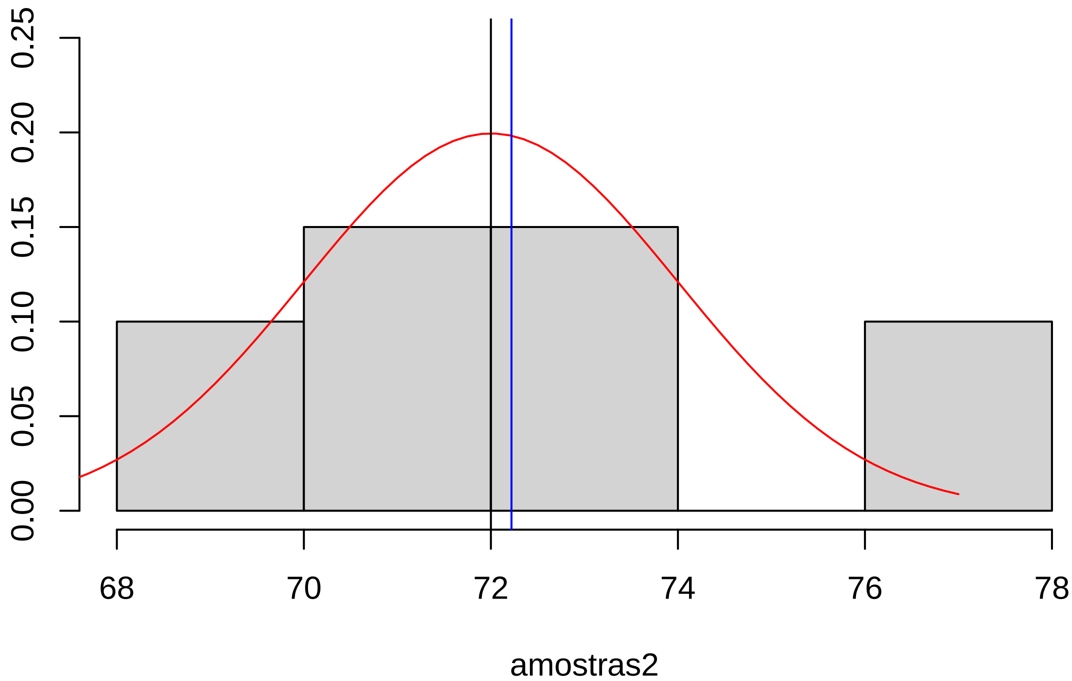
```r
cat("\nrejeita H0? ")
```

```
##
## rejeita H0?
```

```r
abs(z) > abs(1.96)
```

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

```r
hist(amostras2, freq = F, ylim = c(0,0.25) , main = "", ylab="")
curve(expr = dnorm(x, mean= mu , sd = sigma), col = "red"  , from = 62, to = 77, add = T)
abline(v=mu)
abline(v=media, col= "blue")
```



z.test()

```r
library(TeachingDemos)
?z.test()
mu <- 72
sigma <- 2
z.test(x= amostras1 ,mu=mu, sd=sigma, alternative = "two.sided" )
```

```
##
##  One Sample z-test
##
## data:  amostras1
## z = 2.0555, n = 10.00000, Std. Dev. = 2.00000, Std. Dev. of the sample
## mean = 0.63246, p-value = 0.03983
## alternative hypothesis: true mean is not equal to 72
## 95 percent confidence interval:
##  72.06041 74.53959
## sample estimates:
## mean of amostras1
##              73.3
```

```r
library(TeachingDemos)
?z.test()
mu <- 72
sigma <- 2
z.test(x= amostras2 ,mu=mu, sd=sigma, alternative = "two.sided" )
```

```
##
##  One Sample z-test
##
## data:  amostras2
## z = 0.34785, n = 10.00000, Std. Dev. = 2.00000, Std. Dev. of the sample
## mean = 0.63246, p-value = 0.728
## alternative hypothesis: true mean is not equal to 72
## 95 percent confidence interval:
```

```
##  70.98041 73.45959
## sample estimates:
## mean of amostras2
##             72.22
```

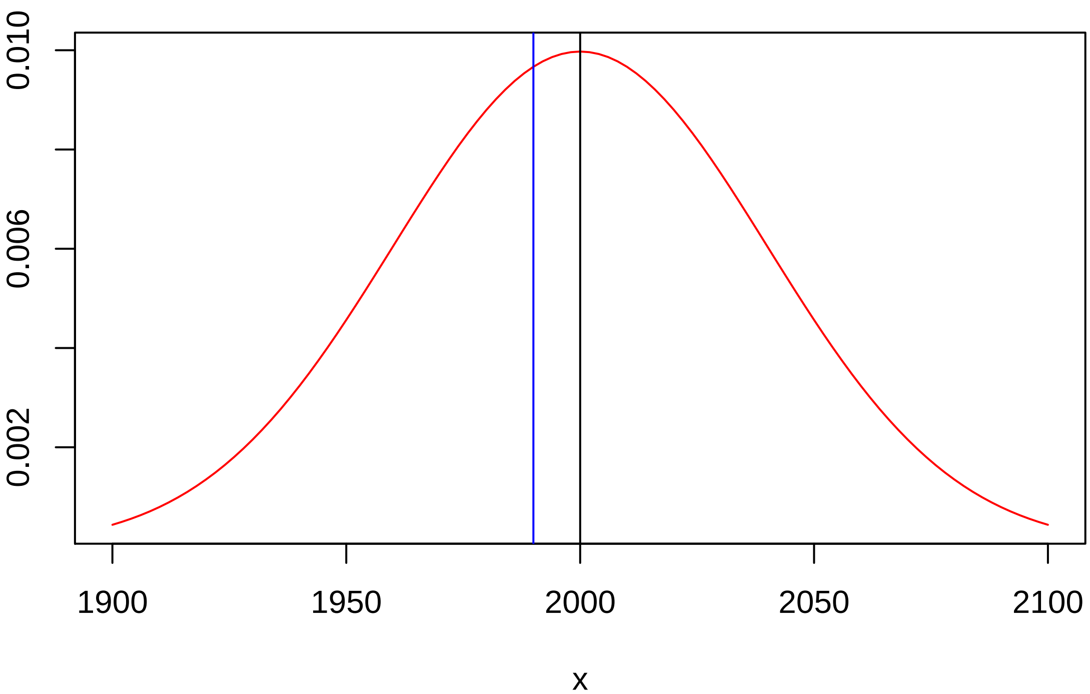
No segundo exemplo ficou muito acima dos 5% (72,8%) portanto não se rejeita H0.

Exemplo 3

Um fabricante de molhos anuncia que o conteúdo líquido das embalagens de seu produto é, em média, de 2.000 gramas, com desvio padrão de 40 gramas. A fiscalização de pesos e medidas investigou uma amostra aleatória de 64 latas, verificando uma média de 1990 gramas por embalagem. Dado um nível de significância de 0,05, o fabricante deverá ser multado por vender o produto abaixo do especificado?

Como pode ver o pacote faz o teste inteiro. Ele calcula a estatística Z bilateral e já faz a comparação com o valor de significancia p (p-value). Neste caso como ficou abaixo do estipulado (5%) rejeita-se a H0.

```r
curve(expr = dnorm(x, mean= 2000 , sd = 40), col = "red"  , from = 1900, to = 2100, ylab = "")
abline(v=2000)
abline(v=1990, col= "blue")
```



A média da amostra está um pouco à esquerda, parece não haver inconformidades.

Vamos conferir com um teste Z **unilateral à esquerda.**

$$H_0 : \bar{x} = \mu$$
$$H_A : \bar{x} < \mu$$

```r
z.test(x= 1990, n=64, mu=2000, sd=40, alternative = "less")
```

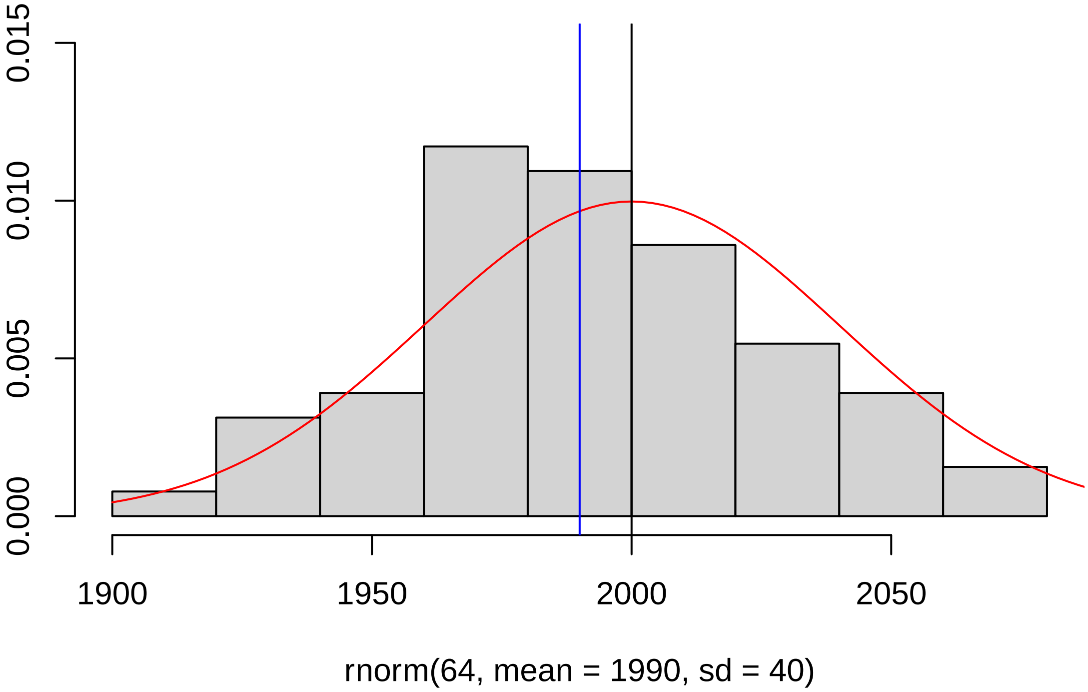
```
##
##  One Sample z-test
##
## data:  1990
```

```
## z = -2, n = 64, Std. Dev. = 40, Std. Dev. of the sample mean = 5,
## p-value = 0.02275
## alternative hypothesis: true mean is less than 2000
## 95 percent confidence interval:
##      -Inf 1998.224
## sample estimates:
## mean of 1990
##         1990
```

Pelo teste estatístico deve-se rejeitar H0. Ou seja a fábrica não cumpre com o padrão estabelecido

Mas e quanto ao gráfico?

```r
set.seed(123)
hist(rnorm(64,mean=1990,sd=40), freq = F, ylim = c(0,0.015) , main = "", ylab = "")
curve(expr = dnorm(x, mean= 2000 , sd = 40), col = "red"  , from = 1900, to = 2100, add = T)
abline(v=2000)
abline(v=1990, col= "blue")
```



Exemplo 4

Um projeto de investimento está sendo avaliado pelo tempo médio de pay-back. Uma situação envolvendo cenários futuros forneceu os seguintes tempos de retorno do investimento (em anos):

2,8 - 4,3 - 3,7 - 6,4 - 3,2 - 4,1 - 4,4- 4,6 - 5,2 - 3,9.

Avalie a hipótese de que o tempo médio de retorno seja superior a 4 anos a uma significância de 5%.

$$H_0 : \mu = 4$$

$$H_A : \mu > 4 \qquad t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

$$p = 0,05$$

$$S = ?$$

$$k = ?$$

```r
amostra4 <- c(2.8 , 4.3 , 3.7 , 6.4 , 3.2 , 4.1 , 4.4, 4.6 , 5.2 , 3.9)
t.test(amostra4, conf.level = 0.95, mu = 4, alternative = "greater"  )
```

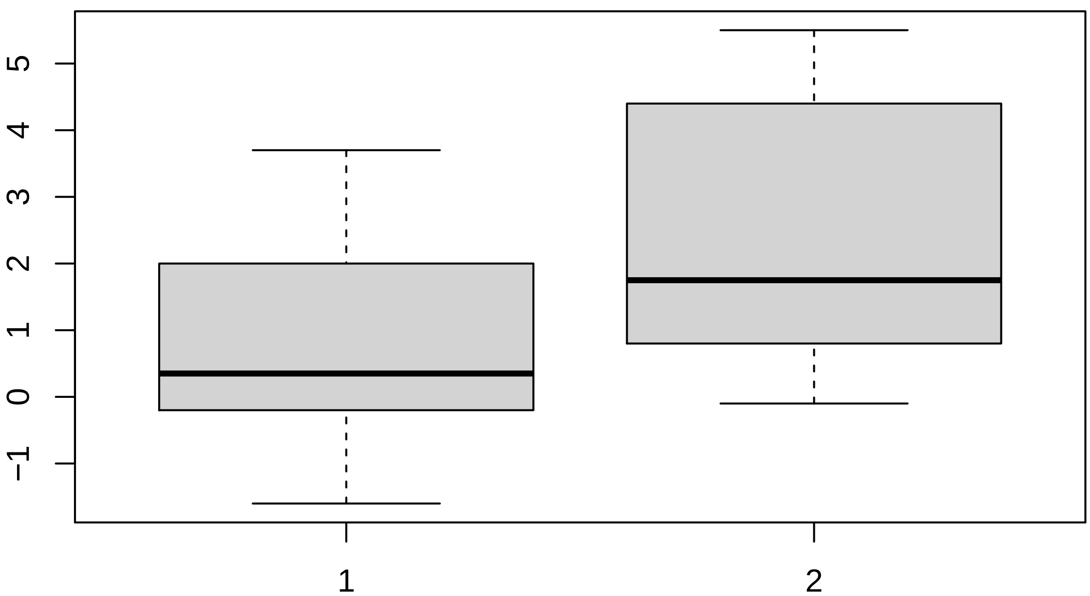
```
##
##  One Sample t-test
##
## data:  amostra4
## t = 0.80778, df = 9, p-value = 0.22
## alternative hypothesis: true mean is greater than 4
## 95 percent confidence interval:
##  3.669977      Inf
## sample estimates:
## mean of x
##      4.26
```

Exemplo 5

Dois medicamentos foram administrado sem gurpos diferentes de paciente para avaliar seu impacto no sono. há diferença entre os grupos? g1: 0.7 -1.6 -0.2 -1.2 -0.1 3.4 3.7 0.8 0.0 2.0 g2: 1.9 0.8 1.1 0.1 -0.1 4.4 5.5 1.6 4.6 3.4

Hipotese: média_g1 = média_g2

```r
g1 <- c(0.7,-1.6,-0.2,-1.2,-0.1,3.4,3.7,0.8,0.0,2.0)
g2 <- c(1.9, 0.8, 1.1, 0.1,-0.1, 4.4, 5.5, 1.6, 4.6, 3.4)
boxplot(g1,g2)
```




```r
var.test(g1,g2, alternative = "two.sided")
```

```
##
##  F test to compare two variances
##
## data:  g1 and g2
## F = 0.79834, num df = 9, denom df = 9, p-value = 0.7427
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
```

```
##  0.198297 3.214123
## sample estimates:
## ratio of variances
##          0.7983426
```

```r
t.test(g1,g2, alternative = "two.sided", conf.level = 0.95, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  g1 and g2
## t = -1.8608, df = 17.776, p-value = 0.07939
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.3654832  0.2054832
## sample estimates:
## mean of x mean of y
##      0.75      2.33
```