

```
n_log <- function(name, number, round_coef=2) {
  cat(name, ': ', round(number, round_coef), '\n')
}
```

## Exemplo 1 - Diferença entre duas médias

### Possui acesso ao vetor de dados

A tabela a seguir contém dados sobre 60 participantes do 7º exame da coorte de descendência no Framingham Heart Study.

	Homens			Mulheres		
Características	$n$	$\bar{x}$	$s$	$n$	$\bar{x}$	$s$
Pressão sanguínea sistólica	6	117.5	9.7	4	126.8	12.0
Pressão sanguínea diastólica	6	72.5	7.1	4	69.5	8.1
Colesterol sérico total	6	193.8	30.2	4	215.0	48.8
Peso	6	196.9	26.9	4	146.0	7.2
Altura	6	70.2	1.0	4	62.6	2.3
Índice de massa corporal	6	28.0	3.6	4	26.2	2.0

Construa um intervalo de confiança de 95% para a diferença entre as médias de **pressão sanguínea sistólica** entre homens e mulheres utilizando estes dados.

```
set.seed(123091287)
```

```
men_data <- rnorm(6, m=117.5, sd=9.7)
women_data <- rnorm(4, m=126.8, sd=12.0)
```

```
summary(men_data)
summary(women_data)
n_log('men sd', sd(men_data))
n_log('women sd', sd(women_data))
```

```
ans <- var.test(men_data, women_data)
```

```
ifelse(ans$p.value < 0.05,
  is_var_equal <- FALSE,
  is_var_equal <- TRUE)
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
108.5 109.6 112.7 118.2 125.1 137.4
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
```

```

114.4  116.7  123.3  127.4  133.9  148.5

men sd : 11.97
women sd : 15.45

TRUE

(ans <- t.test(men_data, women_data, var.equal=is_var_equal))
n_log('confidence interval', ans$conf.int)

Two Sample t-test

data:  men_data and women_data
t = -1.0607, df = 8, p-value = 0.3198
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -29.08299  10.75753
sample estimates:
mean of x mean of y
 118.1882  127.3510

confidence interval : -29.08 10.76

```

## Sem acesso ao vetor de dados

```

# Values
x_bar_1 <- 117.5 # mean(men_data)
x_bar_2 <- 126.8 # mean(women_data)
s_1 <- 9.7 # sd(men_data)
s_2 <- 12.0 # sd(women_data)
n_1 <- 6
n_2 <- 4
alpha <- 0.05

# F test
f_c <- s_1^2 / s_2^2
f_p <- 2 * pf(f_c, df1=n_1 - 1, df2=n_2 - 2)
n_log('f test p-value', f_p)

# t test for variance equality
s_c <- sqrt(((n_1 - 1)*s_1^2 + (n_2 - 1)*s_2^2) / (n_1 + n_2 - 2))
s_combined <- s_c * sqrt((n_1 + n_2) / (n_1 * n_2))

deg_of_freedom <- n_1 + n_2 - 2

```

```
critical_t <- qt(alpha / 2, deg_of_freedom, lower.tail=F)
n_log('critical t', critical_t)

me <- critical_t * s_combined
pos_ci <- x_bar_1 - x_bar_2 + me
neg_ci <- x_bar_1 - x_bar_2 - me

ci <- c(neg_ci, pos_ci)

n_log('confidence interval', ci)

f test p-value : 0.61
critical t : 2.31
confidence interval : -25.11 6.51
```

## Exemplo 2 - Diferença entre duas proporções

Um pesquisador da área médica conjectura que fumar pode resultar em rugas ao redor dos olhos. O pesquisador recrutou 150 fumantes e 250 não fumantes para participar em um estudo observacional e constatou que 95 dos fumantes e 105 dos não fumantes vieram a ter rugas proeminentes ao redor dos olhos (baseado em uma pontuação de rugas padronizada administrada por uma pessoa que não sabia quem era ou não fumante.). Alguns resultados do estudo se encontram na tabela a seguir.

	Fumantes	Não fumantes
Tamanho da amostra	150	250
Proporção amostral de rugas proeminentes	$\frac{95}{150} = 0.63$	$\frac{105}{250} = 0.42$
Desvio padrão	0.0394	0.0312

Como os fumantes se comparam com os não fumantes? Faça um intervalo de confiança de 95%.

```
frequencies <- c(95, 105)
populations <- c(150, 250)

prop.test(frequencies, populations, correct=FALSE)

2-sample test for equality of proportions without continuity
correction

data: frequencies out of populations
X-squared = 17.067, df = 1, p-value = 3.609e-05
```

```

alternative hypothesis: two.sided
95 percent confidence interval:
 0.1148942 0.3117725
sample estimates:
   prop 1    prop 2 
0.6333333 0.4200000

```

Na unha

```

# Values
alpha = 0.05
p_1 <- 0.63
p_2 <- 0.42
n_1 <- 150
n_2 <- 250

# z test
sigma_combined <- sqrt((p_1*(1 - p_1)) / n_1 + (p_2*(1 - p_2)) / n_2)
critical_z <- qnorm(alpha / 2, lower.tail=F)
me <- critical_z * sigma_combined
pos_ci <- p_1 - p_2 + me
neg_ci <- p_1 - p_2 - me

ci <- c(neg_ci, pos_ci)

n_log('confidence interval', ci)
confidence interval : 0.11 0.31

```