```r
m_log <- function(message) {
    cat(message, '\n')
}

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

Exemplo 1 - Diferença entre duas médias

Possuindo os vetores de dados

A empresa Acme desenvolveu uma nova bateria. O engenheiro no comando afirma que a nova bateria opera continuamente por pelo menos 7 minutos a mais que a bateria antiga.

Para testar a afirmação, a empresa seleciona uma amostra aleatória simples de 100 baterias novas e 100 velhas. As baterias velhas rodam continuamente por 190 minutos com um desvio padrão de 20 minutos; as baterias novas, 200 minutos com um desvio padrão de 40 minutos.

Verifique a afirmação do engenheiro com um nível de significância de 5%.

Hipóteses:

$$H_0 : \mu_1 - \mu_2 \leq 7$$

$$H_a : \mu_1 - \mu_2 > 7$$

```r
set.seed(1924831250)

old_batteries <- rnorm(100, m=190, sd=20)
new_batteries <- rnorm(100, m=200, sd=40)

tb <- data.frame(
    'old batteries' = old_batteries,
    'new batteries' = new_batteries
)

summary(tb)
```

```
 old.batteries   new.batteries
 Min.   :149.5   Min.   :122.4
 1st Qu.:179.3   1st Qu.:175.7
 Median :193.6   Median :202.5
 Mean   :192.3   Mean   :202.7
 3rd Qu.:203.9   3rd Qu.:230.1
 Max.   :237.8   Max.   :284.9
```

```r
(var_test <- var.test(new_batteries, old_batteries))

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

```
    F test to compare two variances

data:  new_batteries and old_batteries
F = 3.9435, num df = 99, denom df = 99, p-value = 5.122e-11
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 2.653342 5.860942
sample estimates:
ratio of variances
          3.943486
```

FALSE

```r
(ans <- t.test(new_batteries, old_batteries,
               alternative='greater',
               mu=7, var.equal=is_var_equal))

n_log('t_c', ans$statistic)
n_log('p-value', ans$p.value)
n_log('degrees of freedom', ans$parameter)
```

```
    Welch Two Sample t-test

data:  new_batteries and old_batteries
t = 0.83867, df = 146.18, p-value = 0.2015
alternative hypothesis: true difference in means is greater than 7
95 percent confidence interval:
 3.629729      Inf
sample estimates:
mean of x mean of y
 202.7294  192.2683



t_c : 0.84
p-value : 0.2
degrees of freedom : 146.18
```

```r
par(mfrow=c(1, 1), pin=c(6, 4))

alpha = 0.05
deg_of_freedom = ans$parameter
```

```
t_c = ans$statistic

curve(dt(x, deg_of_freedom), xlim=c(-5, 5), ylab='')

critical_t = qt(alpha, deg_of_freedom, lower.tail=FALSE)
x_rej = seq(critical_t, 4, len=1e3)
y_rej = dt(x_rej, deg_of_freedom)

polygon(c(critical_t, x_rej), c(0, y_rej),
        col=adjustcolor('red', alpha.f=0.2),
        border='red')

x_acc = seq(-4, critical_t, len=1e3)
y_acc = dt(x_acc, deg_of_freedom)

polygon(c(critical_t, x_acc), c(0, y_acc),
        col=adjustcolor('darkgreen', alpha.f=0.1),
        border='darkgreen')

points(critical_t, 0, pch=18, col='red')
text(critical_t, -0.01, col='red',
     label='t')
points(t_c, 0, pch=18, col='darkgreen')
text(t_c, -0.01, col='darkgreen',
     label='tc')
par(mfrow=c(1, 1), pin=c(6, 4))

curve(dt(x, deg_of_freedom), xlim=c(-5, 5), ylab='')

x_rej = seq(t_c, 4, len=1e3)
y_rej = dt(x_rej, deg_of_freedom)

polygon(c(t_c, x_rej), c(0, y_rej),
        col=adjustcolor('darkblue', alpha.f=0.2),
        border='darkblue')

x_acc = seq(-4, t_c, len=1e3)
y_acc = dt(x_acc, deg_of_freedom)

polygon(c(t_c, x_acc), c(0, y_acc),
        col=adjustcolor('darkgreen', alpha.f=0.1),
        border='darkgreen')

points(t_c, 0, pch=18, col='darkblue')
text(t_c, -0.01, col='darkblue',
```

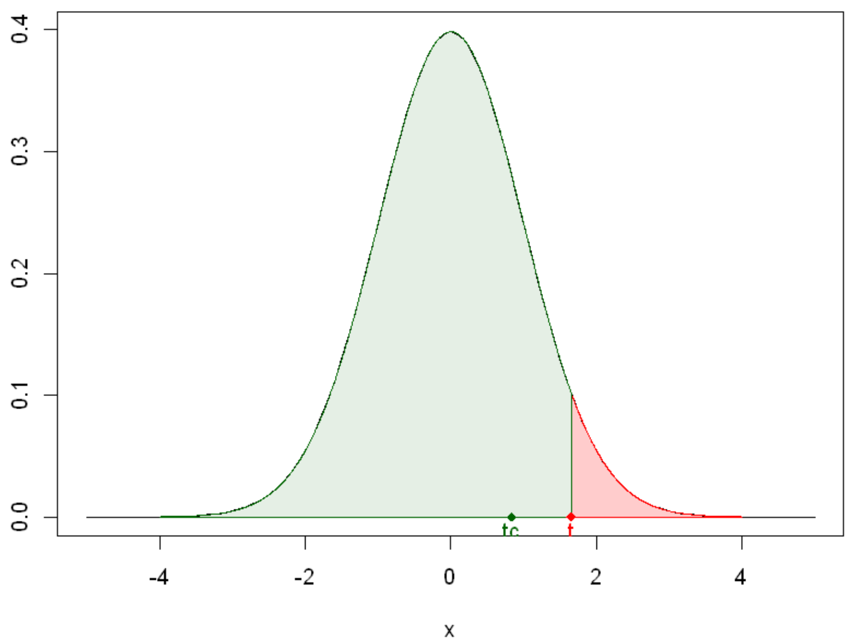

Figure 1: png

```
label='tc')
```

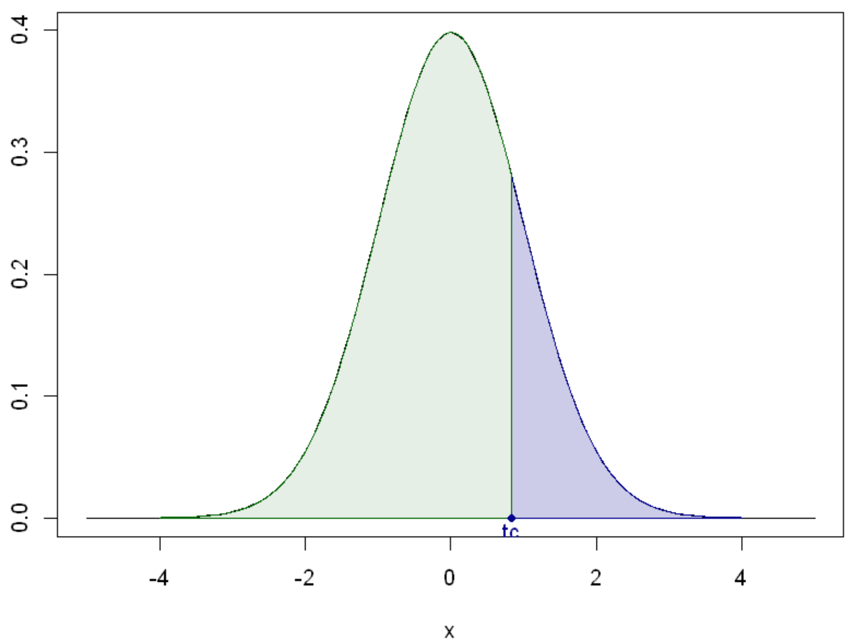

Figure 2: png

Desconhecendo os vetores de dados

```
# Values
s_1 <- 40 # sd(new_batteries)
s_2 <- 20 # sd(old_batteries)
x_bar_1 <- 200 # mean(new_batteries)
x_bar_2 <- 190 # mean(old_batteries)
n_1 <- 100
n_2 <- 100
mu <- 7

# F test
```

```
f_c <- s_1^2 / s_2^2
f_p_value <- 2 * pf(f_c, df1=n_1 - 1, df2=n_2 - 2, lower.tail=F)
n_log('f test p-value', f_p_value)

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

t_c <- ((x_bar_1 - x_bar_2) - mu) / s_combined

t_p_value <- pt(t_c, df=deg_of_freedom, lower.tail=F)
n_log('t test p-value', t_p_value)
```

```
f test p-value : 0
t test p-value : 0.25
```

Exemplo 2 - Diferença entre duas proporções

Suponha que a empresa Acme Drug Company desenvolva um novo medicamento contra resfriado. A empresa afirma que o medicamento é igualmente efetivo para homens e mulheres. Para testar essa afirmação, eles escolheram uma amostra aleatória simples de 100 mulheres e 200 homens de uma população de 100.000 voluntários.

Ao final do estudo, 38% das mulheres pegaram resfriado e 51% dos homens pegaram resfriado. Baseado nestes resultados, podemos rejeitar a afirmação da empresa do medicamento ser igualmente efetivo para homens e mulheres? Use um nível de significância de 5%.

Hipóteses

$$H_0 : \pi_1 - \pi_2 = 0$$

$$H_0 : \pi_1 \neq \pi_2 = 0$$

```
populations <- c(100, 200)
frequencies <- c(0.38 * populations[[1]], 0.51 * populations[[2]])

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

```
    2-sample test for equality of proportions without continuity
    correction

data:  frequencies out of populations
X-squared = 4.5268, df = 1, p-value = 0.03337
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.24768764 -0.01231236
```

```
sample estimates:
prop 1 prop 2
  0.38   0.51
```

Na unha

```
# Values
p_1 <- 0.38
p_2 <- 0.51
n_1 <- populations[[1]]
n_2 <- populations[[2]]
pi_0 <- 0

# z test for equality
p_1_2 <- (n_1 * p_1 + n_2 * p_2) / (n_1 + n_2)
sigma_combined <- sqrt(p_1_2 * (1 - p_1_2) * (1 / n_1 + 1 / n_2))

z_c = ((p_1 - p_2) - pi_0) / sigma_combined
n_log('z_c', z_c)

z_p <- 2 * pnorm(z_c)
n_log('z test p-value', z_p, 3)
```

```
z_c : -2.13
z test p-value : 0.033
```

```
par(mfrow=c(1, 1), pin=c(6, 4))

alpha <- 0.05
critical_z <- qnorm(alpha / 2, lower.tail=F)

curve(dnorm(x), xlim=c(-5, 5), ylab='')

x_rej_1 = seq(-4, -critical_z, len=1e3)
y_rej_1 = dnorm(x_rej_1)

polygon(c(x_rej_1, -critical_z), c(y_rej_1, 0),
        col=adjustcolor('red', alpha.f=0.2),
        border='red')

x_rej_2 = seq(critical_z, 4, len=1e3)
y_rej_2 = dnorm(x_rej_2)

polygon(c(critical_z, x_rej_2), c(0, y_rej_2),
        col=adjustcolor('red', alpha.f=0.2),
        border='red')
```

```r
x_acc = seq(-critical_z, critical_z, len=1e3)
y_acc = dnorm(x_acc)

polygon(c(-critical_z, x_acc, critical_z), c(0, y_acc, 0),
        col=adjustcolor('darkgreen', alpha.f=0.1),
        border='darkgreen')

points(z_c, 0, pch=18, col='red')
text(z_c - 0.1, -0.01, col='red',
     label='zc')

points(critical_z, 0, pch=18, col='red')
text(critical_z, -0.01, col='red',
     label='z')

points(-critical_z, 0, pch=18, col='red')
text(-critical_z + 0.1, -0.01, col='red',
     label='-z')
par(mfrow=c(1, 1), pin=c(6, 4))

curve(dnorm(x), xlim=c(-5, 5), ylab='')

x_rej_1 = seq(-4, z_c, len=1e3)
y_rej_1 = dnorm(x_rej_1)

polygon(c(x_rej_1, z_c), c(y_rej_1, 0),
        col=adjustcolor('darkblue', alpha.f=0.2),
        border='darkblue')

x_rej_2 = seq(-z_c, 4, len=1e3)
y_rej_2 = dnorm(x_rej_2)

polygon(c(x_rej_2, -z_c), c(0, y_rej_2),
        col=adjustcolor('darkblue', alpha.f=0.2),
        border='darkblue')

x_acc = seq(z_c, -z_c, len=1e3)
y_acc = dnorm(x_acc)

polygon(c(z_c, x_acc, -z_c), c(0, y_acc, 0),
        col=adjustcolor('darkgreen', alpha.f=0.1),
        border='darkgreen')

points(z_c, 0, pch=18, col='darkblue')
text(z_c, -0.01, col='darkblue',
```

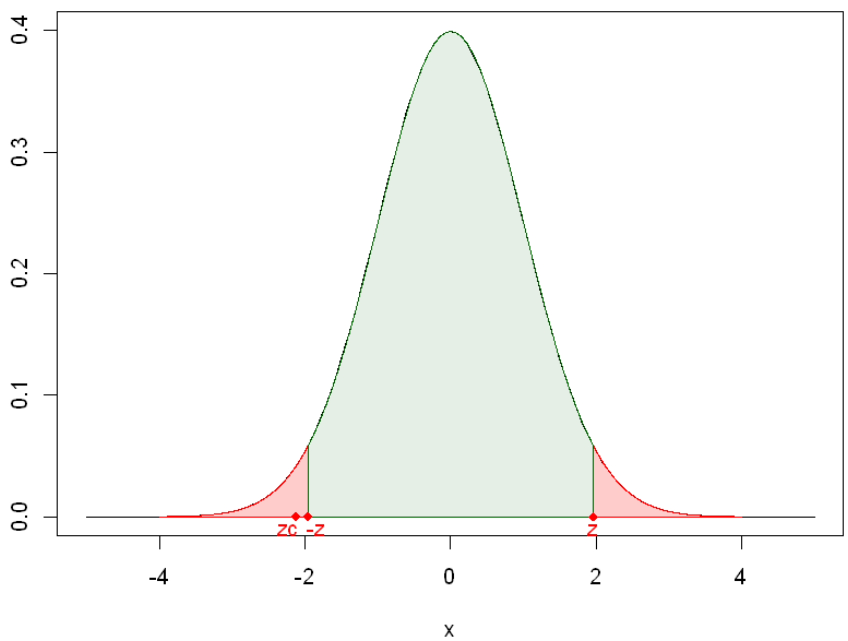

Figure 3: png

```
     label='-zc')

points(-z_c, 0, pch=18, col='darkblue')
text(-z_c, -0.01, col='darkblue',
     label='zc')
```

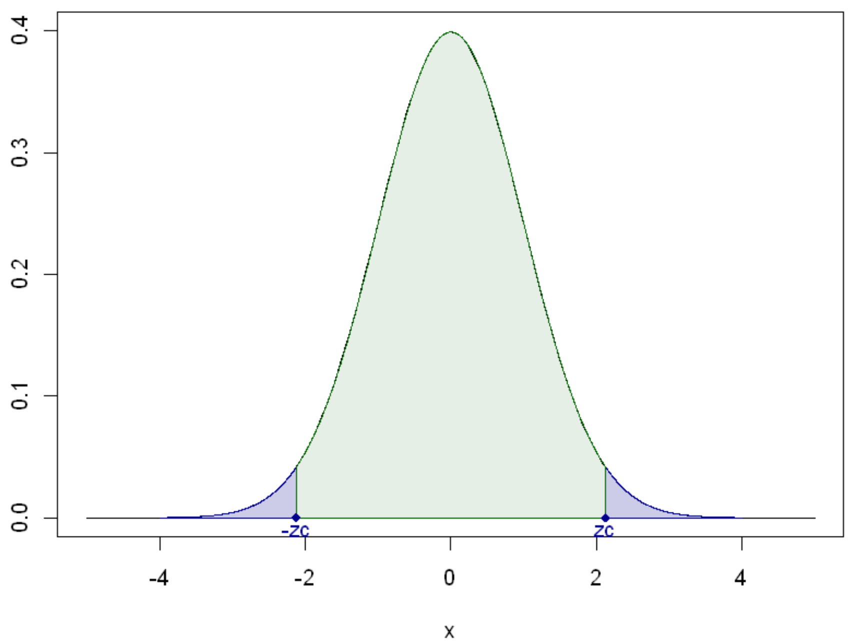

Figure 4: png