```r
#libraries
library(here)
library(tibble)
library(ggplot2)
```

> Num inquérito sobre óculos de sol foram colocadas várias questões aos inquiridos. Para além de características sociodemográficas (sexo, idade e nível de educação), perguntou-se o tipo de óculos de sol que possuíam, quando tinham sido adquiridos, onde tinham sido adquiridos, quanto tinham custado e se eram da marca SoleMio(SM/RB). Para este TPC, irão apenas analisar duas questões: 1. O indicador "Importância do Preço na compra de óculos de sol" – variável Price; e, 2. a questão "are_RB", que indica se os óculos são ou não da marca SoleMio Para além destas questões, ainda foram colocadas outras que originaram a construção de um conjunto de indicadores, cada um numa escala contínua de 0 a 10 – fatores que influenciam a compra de óculos de sol.
>
> Os "Fatores que influenciam a compra de óculos de sol" são variáveis que assumem valores reais no intervalo 0-10, onde 0 corresponde a "nada importante" e 10 corresponde a "extremamente importante".

```r
df<- readRDS(here('tpc7', 'Estudo_Oculos_Sol.rds'))
tibble(df)
```

```
## # A tibble: 640 x 16
##     nquest sex      age educ     type  when_~1 where~2 cost  are_RB
   nb_own Pub.Mk
##      <int> <fct>  <int> <fct>    <fct> <fct>   <fct>   <fct> <fct>   <
   int>  <dbl>
##  1       1 Male      38 Tertiary Spor~ 2+ yea~ Optica~ 200E~ No
          1   7.51
##  2       2 Male      37 Tertiary Clas~ 2+ yea~ Optica~ 200E~ Yes
         2   5.80
##  3       3 Male      33 Tertiary Clas~ Last y~ Optica~ 100E~ No
          1   4.76
##  4       4 Female    25 Profiss~ Spor~ Last y~ Optica~ At l~ No
          3   0.91
##  5       5 Female    34 Seconda~ Mode~ This y~ Optica~ 100E~ No
          3   4.66
##  6      12 Female    43 Profiss~ Clas~ 2+ yea~ Optica~ 100E~ No
          1   4.85
##  7      13 Male      20 Seconda~ Spor~ Last y~ Sports~ Less~ No
          1   5.80
##  8      14 Male      22 Tertiary Clas~ This y~ Optica~ 200E~ No
          3   8
##  9      15 Male      23 Profiss~ Clas~ Last y~ Street~ Less~ No
          1   6.9
## 10      16 Female    24 Tertiary Mode~ This y~ Optica~ 200E~ No
          2   5.70
## # ... with 630 more rows, 5 more variables: Quality <dbl>, Ergonomy <
   dbl>,
## #   Price <dbl>, Style <dbl>, Will_buy_RB <fct>, and abbreviated
   variable names
## #   1: when_bought, 2: where_bought
```

1

> Será que a importância concedida ao preço está, em termos médios, acima do ponto intermédio da escala (i.e. 5)?

Escrever as hipóteses

$$H_0 : \mu \leq \mu_0 \quad (1)$$
$$H_a : \mu > \mu_0 \quad (2)$$

, sendo $\mu_0 = 5$.

Escolher o teste adequado

Como n é suficientemente grande ($n = 640$) e o desvio padrão σ é desconhecida, o teste adequado será:

$$T = \frac{\overline{X} - \mu_0}{{}^{s}/\sqrt{n}} \overset{\circ}{\cap} \mathsf{n}(0, 1) \tag{3}$$

$$\overline{X} \overset{\circ}{\cap} \mathsf{n}(\mu_0, {}^{s}/\sqrt{n}) \tag{4}$$

, sendo $\overline{X}$ a média amostral, S o desvio padrão amostral, e n o tamanho da amostra.

Definir a região crítica e não crítica

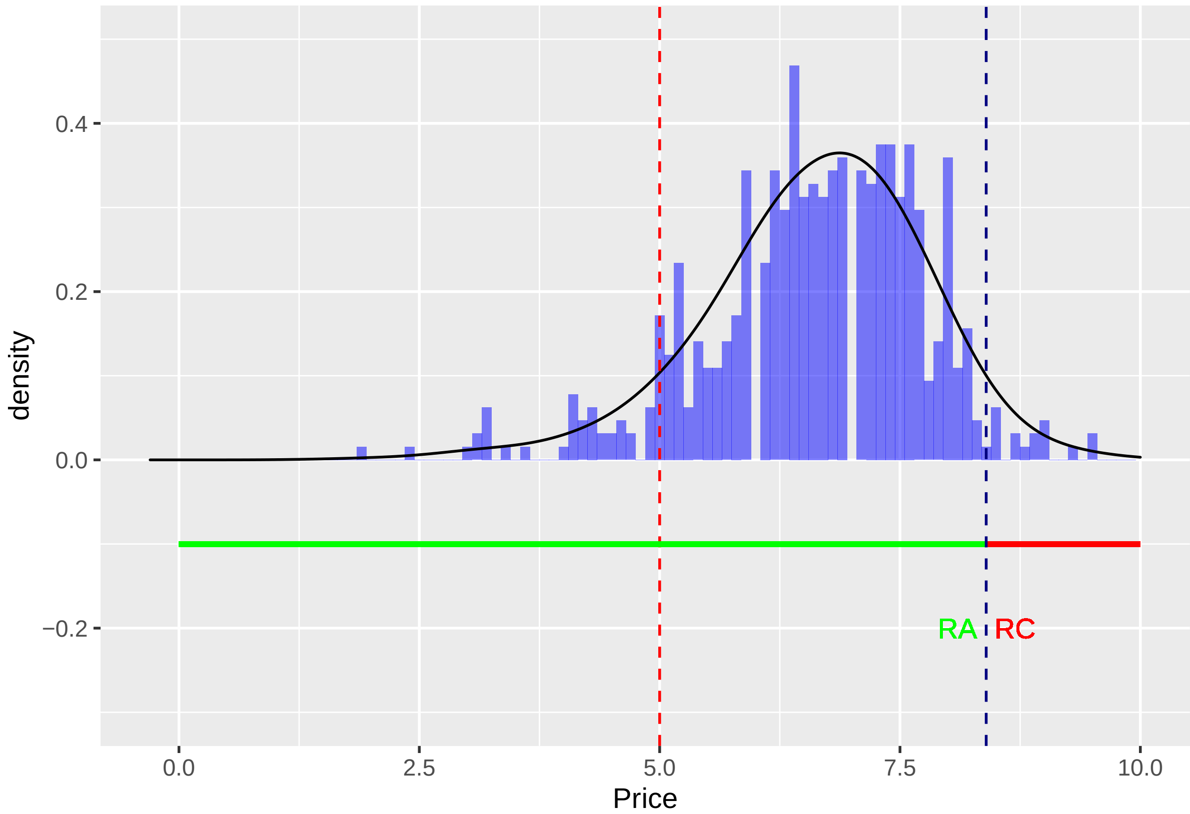
Este teste designa-se por teste unilateral à direita, confome a Eq. 1, e o nível de significância será $\alpha = 0.05$, conforme o enunciado. Assim, a região crítica RC e a região de aceitação RA serão:

$$\mathsf{RC} = \{\overline{X} : \overline{X} \geq \bar{\mathsf{x}}_c\} = [0; \bar{\mathsf{x}}_c] \tag{5}$$

$$\mathsf{RA} = \{\overline{X} : \overline{X} < \bar{\mathsf{x}}_c\} =]\bar{\mathsf{x}}_c; 10] \tag{6}$$

, sendo $\bar{\mathsf{x}}_c = \phi^{-1}(\alpha)$.

```
pricePlot <- ggplot(df, aes(x = Price)) +
  geom_histogram(aes(y = ..density..), binwidth=0.1, fill = 'blue', alpha
      = 0.5) +
  geom_density(kernel="gaussian", bw=0.5) +
  geom_vline(xintercept = 5, color = 'red', linetype = 'dashed')
xc <- qnorm(1-0.05, mean = mean(df$Price), sd = sd(df$Price))
#xc <- qnorm(1-0.05, mean = 5, sd = sd(df$Price)/sqrt(nrow(df)))
pricePlot +
  geom_segment(x = xc, xend = 10, y = -0.1, yend = -0.1, color = 'red',
      size = 1) +
  geom_segment(x = 0, xend = xc, y = -0.1, yend = -0.1, color = 'green',
      size = 1) +
  geom_text(x = xc-0.3, y = -0.2, label = 'RA', color = 'green') +
  geom_text(x = xc+0.3, y = -0.2, label = 'RC', color = 'red') +
  geom_vline(xintercept = xc, color = 'navy', linetype = 'dashed') +
  xlim(-0.3, 10) + ylim(-0.3, 0.5)
```



$$\bar{x}_c = \phi^{-1}(\alpha) = \phi^{-1}(0.05) = 8.397 \tag{7}$$

$$\therefore \text{RA} = [0; 8.397] \wedge \text{RC} =]8.397; 10] \tag{8}$$

A regra de decisão será:

1. Se $\overline{X} > 8.397$, rejeitar H_0 (A importância concedida ao preço está, em termos médios, acima do ponto intermédio da escala);
2. Se $\overline{X} \leq 8.397$, não rejeitar H_0 (A importância concedida ao preço não está, em termos médios, acima do ponto intermédio da escala).

Calcular e verificar se o valor do teste(t) está na região crítica ou não, e tomar a decisão

```
t.test(df$Price, mu = 5, alternative = 'greater', conf.level = 0.95)
```

```
##
##  One Sample t-test
##
## data:  df$Price
## t = 38.13, df = 639, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 5
## 95 percent confidence interval:
##  6.553947      Inf
## sample estimates:
## mean of x
##   6.62411
```

Como $\overline{X} = 6.624 \leq 8.397$, não rejeitamos H_0, ou seja, a importância concedida ao preço está em termos médios abaixo do ponto intermédio da escala.

2

> Será que homens e mulheres diferem, em termos médios, na importância concedida ao preço?

Escrever as hipóteses

$$H_0 : \mu_H = \mu_M \equiv \mu_H - \mu_M = 0 \tag{9}$$

$$H_a : \mu_H \neq \mu_M \equiv \mu_H - \mu_M \neq 0 \tag{10}$$

, sendo μ_H a média da importância concedida ao preço pelos homens e μ_M a média da importância concedida ao preço pelas mulheres.

Escolher o teste adequado

Como n_H e n_M são suficientemente grandes ($n_H = 308$ e $n_M = 332$) e o desvio padrão σ é desconhecido, o teste adequado será:

$$T = \frac{\left(\overline{X}_H - \overline{X}_M\right) - (\mu_H - \mu_M)_0}{\sqrt{\frac{S_H^2}{n_H} + \frac{S_M^2}{n_M}}} \mathring{\cap}\, \mathsf{n}(0,1) \tag{11}$$

$$\left(\overline{X}_H - \overline{X}_M\right) \mathring{\cap}\, \mathsf{n}\left(\mu_H - \mu_M, \sqrt{\frac{S_H^2}{n_H} + \frac{S_M^2}{n_M}}\right) \tag{12}$$

Definir a região crítica e não crítica;

Este teste designa-se por teste bilateral, e as regiões crítica e de aceitação são:

$$\text{RC} = \{\overline{X} : \overline{X} \geq \bar{\text{x}}_c\} = [\bar{\text{x}}_c; 10] \cup [-10; -\bar{\text{x}}_c] \tag{13}$$

$$\text{RA} = \{\overline{X} : \overline{X} < \bar{\text{x}}_c\} =] - \bar{\text{x}}_c; \bar{\text{x}}_c[\tag{14}$$

, sendo $\bar{\text{x}}_c = \phi^{-1}(\alpha/2)$ o ponto crítico.

```
xc2 <- qnorm(1-0.025, mean = mean(df$Price), sd = sd(df$Price))
xc2
```

```
## [1] 8.736079
```

Calcular e verificar se o valor do teste(t) está na região crítica ou não, e tomar a decisão

```
t.test(df$Price ~ df$sex, mu = 0, alternative = 'two.sided', conf.level =
    0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  df$Price by df$sex
## t = -1.9884, df = 638, p-value = 0.0472
## alternative hypothesis: true difference in means between group Female
    and group Male is not equal to 0
## 95 percent confidence interval:
##  -0.335158827 -0.002091658
## sample estimates:
## mean in group Female   mean in group Male
##             6.542959             6.711584
```