

```
#libraries
library(here)
library(tibble)
library(ggplot2)
library(latex2exp)
```

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 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.

Para este TPC, irão apenas analisar duas questões: 1. O indicador “Importância da Ergonomia na compra de óculos de sol” –variável Ergonomy; e, 2. a questão “are\_RB”, que indica se os óculos são ou não da marca SoleMio

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”.

```
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
##   <int> <fct> <int> <fct> <fct> <fct> <fct> <fct> <fct> <fct> <fct>
##   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
```

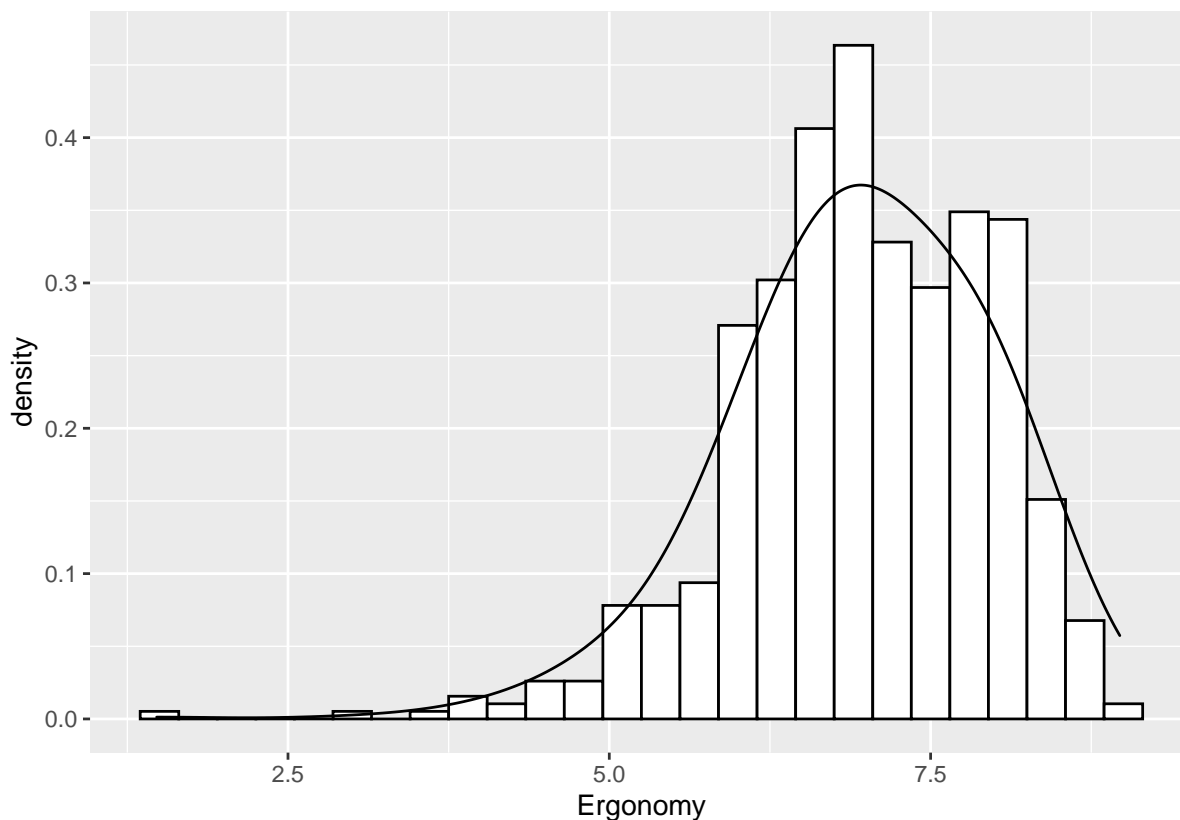
## 1

Pretende-se estimar a importância média concedida à Ergonomia (variável Ergonomy) enquanto fator de influência na compra de óculos de sol, através de um intervalo de confiança apropriado, a 99% de confiança.

### Definir a variável em estudo

A variável em estudo será a importância da ergonomia na compra de óculos de sol, a variável Ergonomy.

```
erg <- df$Ergonomy
ergPlot <- ggplot(df, aes(x = Ergonomy)) + geom_histogram(aes(y=..density
  ..), binwidth=0.3, fill="white", color="black") + geom_density(kernel="
  gaussian", bw=0.5)
ergPlot
```



### Identificar o parâmetro a estimar

Pretende-se estimar a média  $\mu$  da variável Ergonomy, ou seja, a importância média concedida à ergonomia na compra de óculos de sol, através de um intervalo de confiança.

### Escolher a variável fulcral conveniente

Como a média amostral  $\bar{X}$  é estimador de  $\mu$ , a variável fulcral  $Z$  será:

$$Z = \frac{\bar{X} - \mu}{s/\sqrt{n}} \overset{\circ}{\sim} \mathbf{n}(0, 1) \quad (1)$$

, sendo  $s$  o desvio padrão amostral e  $n$  o tamanho da amostra, para  $n > 30$ .

**Identificar o intervalo teórico**

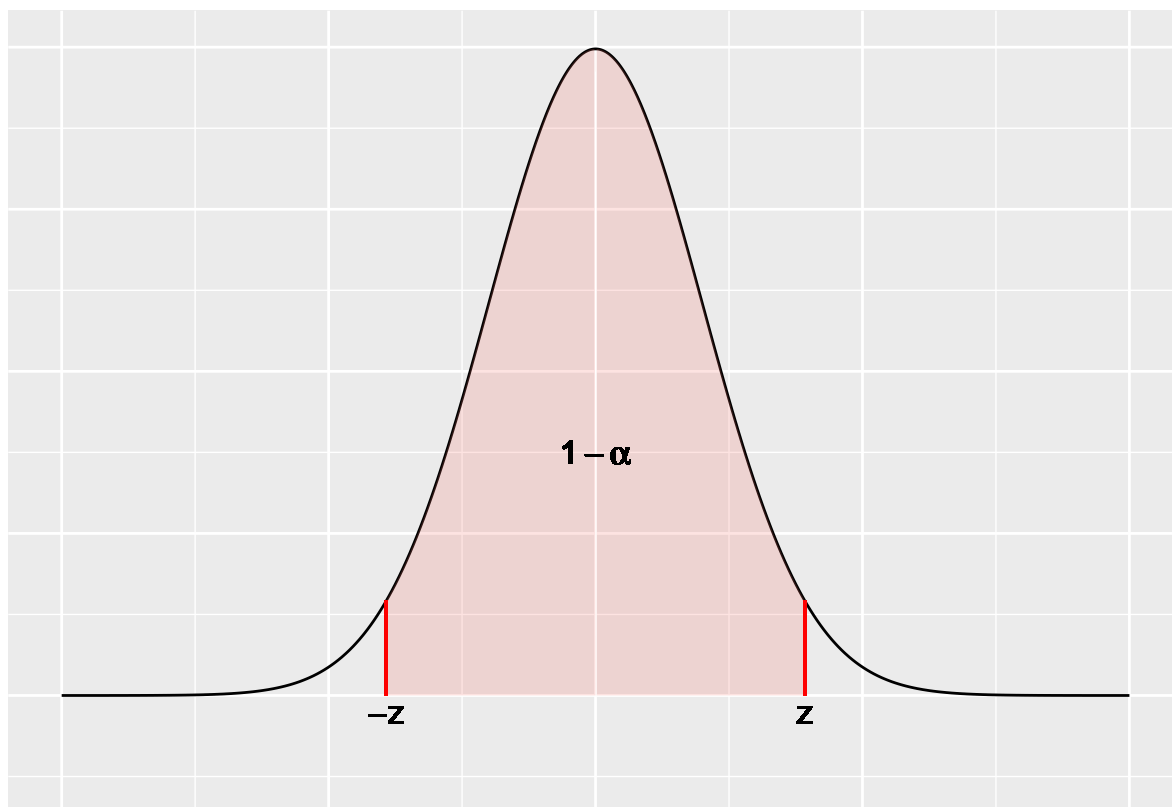
$$P[-z < Z < z] = 1 - \alpha \quad (2)$$

$$P\left[-z < \frac{\bar{X} - \mu}{s/\sqrt{n}} < z\right] = 1 - \alpha \quad (3)$$

$$P\left[\bar{X} - \frac{s \cdot z}{\sqrt{n}} < \mu < \bar{X} + \frac{s \cdot z}{\sqrt{n}}\right] = 1 - \alpha \quad (4)$$

$$]I_{1-\alpha}[\mu = \left[\bar{X} - \frac{s \cdot z}{\sqrt{n}}; \bar{X} + \frac{s \cdot z}{\sqrt{n}}\right[ \quad (5)$$

, sendo  $z$  o valor crítico da distribuição normal padrão, e  $(1 - \alpha)$  o nível de confiança.



**Calcular os valores amostrais necessários**

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad (6)$$

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} \quad (7)$$

$$z = \phi^{-1}(\phi(z)) \quad (8)$$

$$= \phi^{-1}((1 - \alpha) + x/2) \quad (9)$$

$$= \phi^{-1}(0.995) \quad (10)$$

```
mu <- mean(erg)
s <- sd(erg)
n <- length(erg)
z <- qnorm(0.995)
tibble(mu, s, n, z)
```

```
## # A tibble: 1 x 4
##   mu      s      n      z
##   <dbl> <dbl> <int> <dbl>
## 1  6.95 0.976  640  2.58
```

**Construir o intervalo concreto**

$$]I_{0.995}[\mu = \left[ \bar{X} - \frac{s \cdot z}{\sqrt{n}}; \bar{X} + \frac{s \cdot z}{\sqrt{n}} \right] \quad (11)$$

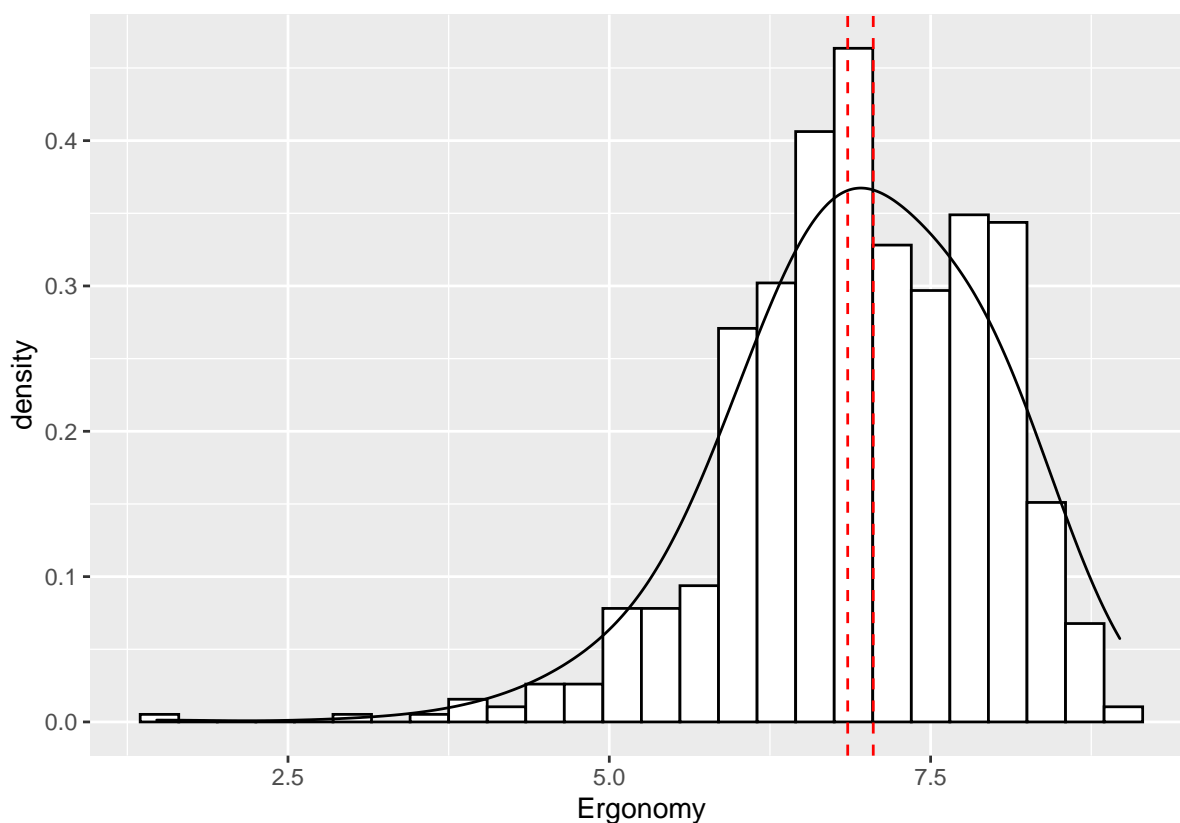
$$= \left[ 6.95 - \frac{0.98 \cdot 2.58}{\sqrt{640}}; 6.95 + \frac{0.98 \cdot 2.58}{\sqrt{640}} \right] \quad (12)$$

```
limits <- c(mu - (s*z/sqrt(n)), mu + (s*z/sqrt(n)))
limits
```

```
## [1] 6.855070 7.053837
```

**Interpretar o intervalo**

```
ergPlot + geom_vline(xintercept = limits, color = "red", linetype = "dashed")
```



Eu confio com 99% de certeza que a média da importância da ergonomia na compra de óculos de sol está entre 6.86 e 7.05 valores de importância.

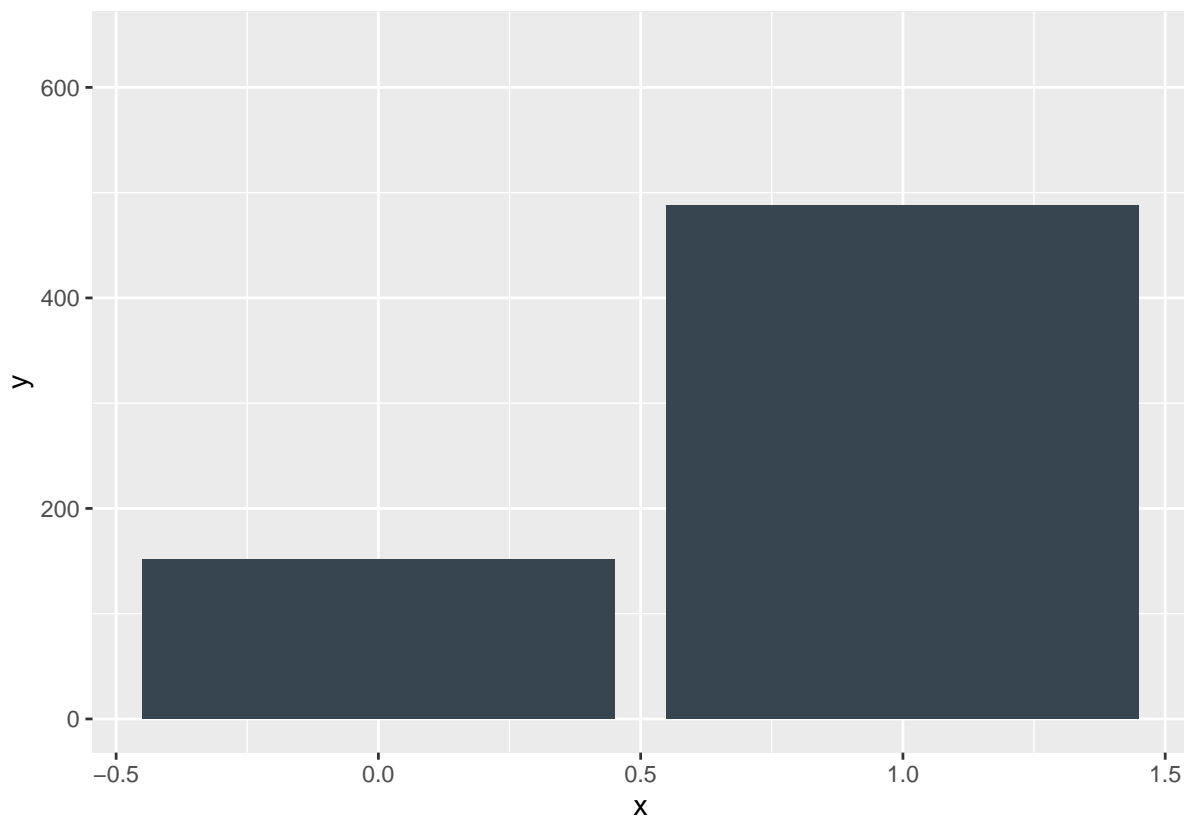
## 2

Repetir os passos 1 a 7 acima descritos para estimar a proporção de pessoas que possuem óculos da marca Solemio (variável are\_RB)

### Definir a variável em estudo

A variável em estudo serão as pessoas que possuem óculos da marca Solemio.

```
poss <- sum(df$are_RB == "Yes")
n <- nrow(df)
ggplot(data.frame(x=c(0,1), y=c(poss, n-poss)), aes(x=x, y=y)) +
  geom_bar(stat="identity", fill="#36454F") +
  ylim(0, n)
```



### Identificar o parâmetro a estimar

Pretende-se estimar o parâmetro  $p$  da distribuição de Bernoulli, a proporção de pessoas que possuem óculos da marca Solemio.

### Escolher a variável fulcral conveniente

Como a média amostral  $\bar{X}$  é estimador de  $p$ , a variável fulcral  $P$  será:

$$P = \frac{Y}{n} \overset{\circ}{\sim} n(p, \frac{p(1-p)}{n}) \quad (13)$$

$$P' = \frac{\frac{Y}{n} - p}{\sqrt{\frac{Y \cdot (n-Y)}{n^3}}} \overset{\circ}{\sim} n(0, 1) \quad (14)$$

,sendo  $Y$  a quantidade de pessoas que possuem óculos da marca, e  $n$  o tamanho da amostra, para  $n > 30$ .

**Identificar o intervalo teórico**

$$]I_{1-\alpha}[p = \left[ \frac{Y}{n} - z \cdot \sqrt{\frac{Y \cdot (n - Y)}{n^3}}; \frac{Y}{n} + z \cdot \sqrt{\frac{Y \cdot (n - Y)}{n^3}} \right] \quad (15)$$

**Calcular os valores amostrais necessário**

$$Y = \sum_{i=1}^n X_i \quad (16)$$

O nosso z foi demonstrado em (10).

```
Y <- sum(df$are_RB == "Yes")
tibble(Y, z)
```

```
## # A tibble: 1 x 2
##       Y      z
##   <int> <dbl>
## 1    152  2.58
```

**Construir o intervalo concreto**

$$]I_{0.995}[p = \left[ \frac{152}{640} - 2.58 \cdot \sqrt{\frac{152 \cdot (640 - 152)}{640^3}}; \frac{152}{640} + 2.58 \cdot \sqrt{\frac{152 \cdot (640 - 152)}{640^3}} \right] \quad (17)$$

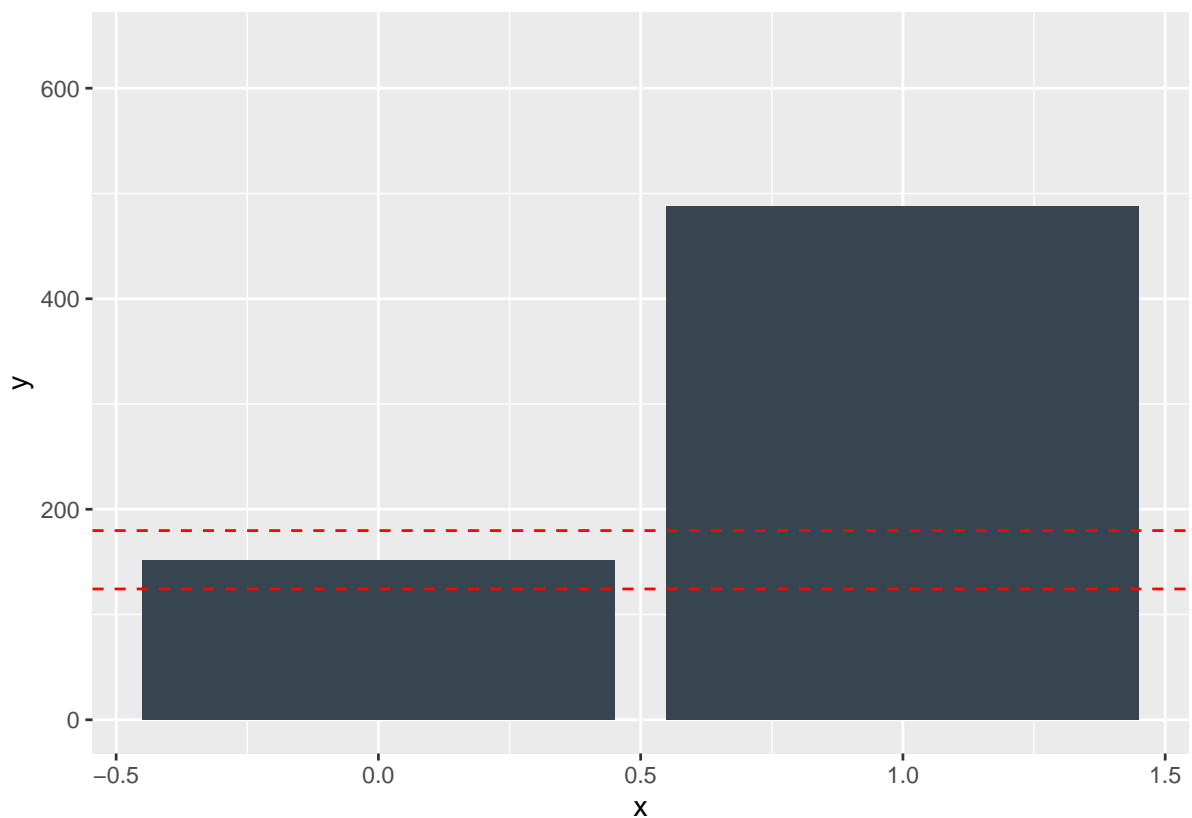
```
limits2 <- c(Y/n - z*sqrt(Y*(n-Y)/n^3), Y/n + z*sqrt(Y*(n-Y)/n^3))
limits2
```

```
## [1] 0.194171 0.280829
```



## Interpretar o intervalo

```
ggplot(data.frame(x=c(0,1), y=c(poss, n-poss)), aes(x=x, y=y)) +  
  geom_bar(stat="identity", fill="#36454F") +  
  ylim(0, n) +  
  geom_hline(yintercept = limits2[1]*n, color = "red", linetype = "  
    dashed") +  
  geom_hline(yintercept = limits2[2]*n, color = "red", linetype = "  
    dashed")
```



Eu confio com 99% de certeza que a proporção de pessoas que não possuem óculos da marca Solemio está entre 0.19 e 0.28.