TPC 7: Parametrização dos Modelos - Turma 1

André Filipe Gomes Silvestre Nº104532 CDB1

Exercício

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 Publicidade e Marketing na compra de óculos de sol" − variável *Pub.Mk*; e, **2.** a questão "*are_RB*", que indica se os óculos são ou não da marca *SoleMio*.

1. Pretende-se estimar a importância média concedida à Publicidade e Marketing (variável *Pub.Mk*) enquanto fator de influência na compra de óculos de sol, através de um intervalo de confiança apropriado, a 99% de confiança.

   Passos a seguir:

   1. Definir a variável em estudo
   2. Identificar o parâmetro a estimar
   3. Escolher a variável fulcral conveniente para a estimação
   4. Identificar o intervalo teórico (estimador)
   5. Calcular os valores amostrais necessários
   6. Construir o intervalo concreto (estimativa)
   7. Interpretar o intervalo obtido.

2. Repetir os passos 1 a 7 acima descritos para estimar a proporção de pessoas que possuem óculos da marca *Solemio* (variável *are_RB*).

---

1. Resolução

```r
# Leitura do ficheiro Estudo_Oculos_Sol.rds
bd_oculos_sol <-readRDS("Estudo_Oculos_Sol.rds")
```

No caso em estudo, a Variável Aleatória em estudo é

> **X -** Importância concedida, pelos inquiridos, à Publicidade e Marketing (variável *Pub.Mk*) enquanto fator de influência na compra de óculos de sol.

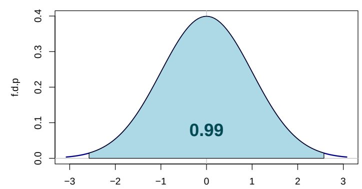
- O parâmetro a estimar é μ
- Nível de confiança a considerar: 99%

Dada que a amostra tem dimensão $n = 640$ e não de conhece o desvio-padrão da distribuição (σ), podemos concluir que a **Variável Fulcral** a usar é

$$VF = \frac{\overline{X} - \mu}{\frac{S'}{\sqrt{n}}} \sim N(0, 1)$$

```r
# Dimensão da Amostra
n<-length(bd_oculos_sol$Pub.Mk)
n<-nrow(bd_oculos_sol) # Alternativa

# Preparar o espaço
conf<-0.99                    # nível de confiança
cauda<-1-conf                 # peso das caudas - alpha
q<- qnorm(conf+cauda/2)       # quantil apropriado

lim_x<-c(-q-0.5,q+0.5)
lim_y<-c(0,dnorm(0))

  plot(1,
     xlim = lim_x,
     ylim = lim_y,
     type = "n",
     ylab = "f.d.p", xlab = "")

# Add x and y-axis lines
abline(h = 0, col="grey")
abline(v = 0, col="grey")

# Desenhar a função
# A função a incluir será a densidade de probabilidade
curve(dnorm(x),
      from =lim_x[1], to = lim_x[2],
      col = "darkblue",
      lwd = 2,
      add=TRUE)

# Marcar a área pretendida
```

```r
pontos_x<-seq(-q,q,0.01)
pontos_y<-dnorm(pontos_x)
cord.x <- c(-q,pontos_x,q)
cord.y <- c(0,pontos_y,0)
polygon(cord.x,cord.y,col='lightblue')
text(x=0,
     y=max(cord.y)/5,
     labels="0.99",
     col="#004953",
     adj=0.5,cex=1.8, font=2)
```



Para IC a **99%**, o erro de estimação será então $erro = z\frac{S'}{\sqrt{(n)}}$, onde z é o quantil de probabilidade 0.995 de uma $N(0,1)$ (pelo TLC), σ não conhecido e n, dimensão da amostra, é elevada $n= $.

Fazendo a partir da *Normal*, o quantil a considerar será $z : P[-z < Z < z] = 0.99$, ou seja $z : P[Z < z] = 0.995$, dado por $qnorm(0.995) = 2.5758$.

$$P\left[\bar{X} - \frac{0.995}{\sqrt{n}}S' < \mu < \bar{X} + \frac{0.995}{\sqrt{n}}S'\right] = 0,99$$

Logo o o intervalo **teórico** (estimador) é dado por

$$]I_{0.99}[_{\mu} \quad = \quad \bar{X} \; \pm \; erro \quad = \quad \bar{X} \; \pm \; \frac{2.576}{\sqrt{n}}S' \quad = \quad \left]\; \bar{X} \; - \; 2.576 \; \frac{S'}{\sqrt{n}} \quad , \quad \bar{X} \; + \; 2.576 \; \frac{S'}{\sqrt{n}} \right[$$

Pelo que o intervalo **concreto** (estimativa) será:

```r
media<-mean(bd_oculos_sol$Pub.Mk)        # média da variável
dp<-sd(bd_oculos_sol$Pub.Mk)             # desvio-padrão da variável
erro<-q*dp/sqrt(nrow(bd_oculos_sol))     # erro amostral
LB<-media-erro                           # limite inferior do IC
UB<-media+erro                           # limite superior do IC
round(c(LB,UB),4)
```

```
## [1] 4.860 5.204
```

O IC será então $]I_{0.99}[_{\mu}{}^{*} =]4.86, 5.2[$.

> Com 99% de confiança podemos inferir que a verdadeira importância média dada pelos inquiridos à Publicidade e Marketing (variável *Pub.Mk*) enquanto fator de influência na compra de óculos de sol está entre os 4.86 e os 5.204

EXTRA

Nota Adicional: Usando o t.test, os resultados seriam os seguintes.

```
teste_pubmkt<-t.test(bd_oculos_sol$Pub.Mk,conf.level = 0.99)
LB1<-teste_pubmkt$conf.int[1]
UB1<-teste_pubmkt$conf.int[2]
c(LB1,UB1)
```

```
## [1] 4.859438 5.204498
```

Comparando os valores obtidos, verifica-se que os IC são praticamente iguais.

```
tab1<-rbind(c(LB,UB),c(LB1,UB1))
rownames(tab1)<-c("Normal","t-Student")
colnames(tab1)<-c("LB","UB")
tab1<-round(tab1,4)

library(knitr)
kable(head(tab1), align = 'c', booktabs = TRUE)
```

	LB	UB
Normal	4.8600	5.2040
t-Student	4.8594	5.2045

2. Resolução

No caso em estudo, a Variável Aleatória em estudo é

X - Pessoas que possuem óculos da marca *Solemio* (variável *are_RB*)

Esta é uma população de *Bernoulli* em que o sucesso, p, é *ter os óculos da marca Solemio*

- O parâmetro a estimar é p (proporção)

```
# Coluna como fator
bd_oculos_sol$are_RB<-as.factor(bd_oculos_sol$are_RB)

# Contagens por níveis do fator
freq_absoluta<-table(bd_oculos_sol$are_RB)
freq_absoluta
```

```
##
##  No Yes
## 488 152
```

```
# Percentagens por níveis do fator
proporcao_percent<-prop.table(freq_absoluta)
round(proporcao_percent,2)
```

```
##
##   No  Yes
## 0.76 0.24
```

```
# Nível de confiança
conf<-0.99
caudas<-1-conf

# Obter z
z<-qnorm(conf+caudas/2)
```

$\hat{p} = \bar{X} \rightarrow$ proporção de sucessos da amostra

Pelo ***Teorema do Limite Central***

$$\frac{\overline{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \ \dot{\sim} \ N(0,1)$$

Em que

$$P\left[-z < \frac{\overline{X} - p}{\sqrt{\frac{p(1-p)}{n}}} < z\right] = 0,99$$

Assim,

$$\text{erro} \ = \sqrt{\frac{p(1-p)}{n}}$$

Então, usando o estimador da variável

$$\frac{\overline{X}(1-\overline{X})}{n}$$

Não é estabelecido outro nível de confiança, logo vamos assumir o mesmo, 99%, pelo que se mantém o quantil necessário, $z : P[-z < Z < z] = 0.99$, ou seja $z : P[Z < z] = 0.995$, dado por $qnorm(0.995) = 2.5758$.

Logo o o intervalo **teórico** (estimador) é dado por

$$]I_{0.99}[_p \ = \ \overline{X} \pm erro \ = \ \overline{X} \pm 2.576\sqrt{\frac{\overline{X}(1-\overline{X})}{n}} = \ \left]\overline{X} \ - \ 2.576\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}, \overline{X} \ + \ 2.576\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}\right[$$

Pelo que o intervalo **concreto** (estimativa) será:

```r
# Cálculo do erro-padrão do estimador (standard error of estimator) se_prop
# sqrt (x_barra*(1-x_barra)/n)

# x_barra é a proporção de inquiridos com os óculos da marca Solemio
n<- sum(freq_absoluta)
n
```

```
## [1] 640
```

```r
se_prop <- as.numeric(sqrt(proporcao_percent["Yes"]*(1-proporcao_percent["Yes"])/n))

# Cálculo do erro de estimação
erro<-z*se_prop
erro
```

```
## [1] 0.04332902
```

```r
# Cálculo do Intervalo de confiança em %
LB<-as.numeric(proporcao_percent["Yes"])-erro
UB<-as.numeric(proporcao_percent["Yes"])+erro

round(c(LB,UB),4)
```

```
## [1] 0.1942 0.2808
```

O IC será então $]I_{0.99}[_p{}^* =]0.194, 0.281[$.

> Com 99% de confiança podemos afirmar que a verdadeira proporção/percentagem de pessoas que possuem óculos de sol da marca *Solemio* (variável *are_RB*) está entre os 19.42% e os 28.08%