

# TPC 9 Teste de Hipóteses - Qui-Quadrado

André Filipe Gomes Silvestre N°104532 CDB1

---

## Exercício

Num inquérito sobre óculos de sol foram colocadas várias questões aos inquiridos. Para além de características sociodemográficas (Typeo, 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 “estilo dos óculos de sol” – variável *type*; e, **2.** a questão *are\_RB*, que indica se os óculos são ou não da marca *SoleMio*.

Pretende-se saber se existe associação entre o estilo dos óculos de sol e o facto de serem da marca RB.

Seguir os seguintes passos:

1. Construir a tabela de contingência com as frequências absolutas, com o estilo em linha e a marca em coluna; adicionar as margens à tabela.
2. Obter a tabela de contingência com as frequências relativas por coluna (distribuição das respostas para o estilo, no grupo dos óculos RB e no grupo dos óculos Não RB)
3. Representar graficamente a informação
4. Definir as hipóteses em teste e identificar a estatística de teste
5. Obter os resultados do teste
6. Mostrar a tabela dos resíduos estandardizados e identificar as associações relevantes.
7. Escrever parágrafo final que resuma os resultados encontrados em 1-6 e responda à questão colocada: saber se existe associação entre o estilo dos óculos de sol e o facto de serem da marca RB (e de que forma)

**Nota:** Os dados estão guardados num ficheiro de formato RDS, “Estudo\_Oculos\_Sol.rds”. Para lerem os dados devem usar a função “readRDS”.

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

## Resolução

### 1. Tabela de contingência com as Frequências Absolutas

Com o estilo em linha e a marca em coluna

```
tab1<-table(bd_olhos_sol$type,bd_olhos_sol$are_RB) # crosstabs freq abs
# var1 em linha
# var2 em coluna

# Tabela com margens
# tab1_sum <- addmargins(tab1,margin = seq_along(dim(tab1)), FUN = sum)

knitr::kable(tab1_sum,format="markdown",digits=3)
```

---

	No	Yes	sum
Classic	161	93	254
Futuristic	23	2	25
Modern	267	50	317
Sport style	37	7	44
sum	488	152	640

---

### 2. Tabela de contingência com as Frequências Relativas por coluna

Distribuição das respostas para o estilo, no grupo dos óculos RB e no grupo dos óculos Não RB

```
tab1_prop <- prop.table(tab1) # freq relativas

# Frequências relativas com margens
tab1_propsum <- addmargins(tab1_prop, FUN = sum)

knitr::kable(t(tab1_propsum),format="markdown",digits=3)
```

---

	Classic	Futuristic	Modern	Sport style	sum
No	0.252	0.036	0.417	0.058	0.762
Yes	0.145	0.003	0.078	0.011	0.238
sum	0.397	0.039	0.495	0.069	1.000

---

```
# Tabela com frequências relativas por Type (soma 1 em linha)
tab1_byType <- prop.table(tab1,margin = 1) # freq relativas
# margin = 1 --> em linha

# para gráfico tem de ser em coluna
tab1_byType_col<-t(tab1_byType)
tab1_byType_col_propsum <- addmargins(tab1_byType_col, FUN = sum)
```

### 3. Representação Gráfica

```
cores<-c('#ff4040','#2e8b57')

# preparar a área
plot(1,
     xlim = c(0,2.5), ylim = c(0,1),
     type = "n",                # vazio
     main = "Gráfico Type por Are_RB", # título
     ylab = "", xlab = "",      # sem nomes
     xaxt = "n")               # sem marcas eixo x

barplot(tab1_byType_col,        # att dados org em colunas
        col = cores,           # cores a usar
        width = 0.45,          # largura das barras
        add=TRUE)              # para dar espaço p legenda

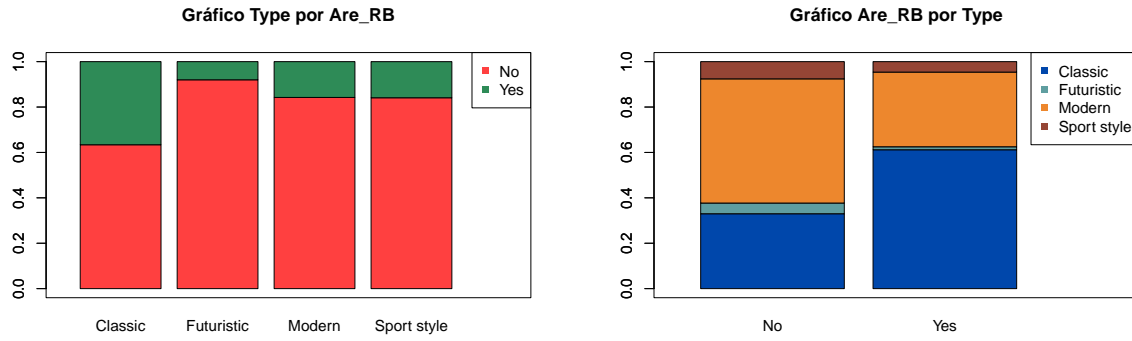
legend("topright",
       legend = rownames(tab1_byType_col),
       pch = 15,
       col = cores)

# Representação gráfica - are_RB por type -----
# preparar a área
tab1_by_areRB <- prop.table(tab1,margin = 2)

cores<-c("#0047ab","#5f9ea0","#ed872d", "#954535")
plot(1,
     xlim = c(0,2.5), ylim = c(0,1),
     type = "n",                # vazio
     main = "Gráfico Are_RB por Type", # título
     ylab = "", xlab = "",      # sem nomes
     xaxt = "n")               # sem marcas eixo x

barplot(tab1_by_areRB,         # att dados org em colunas
        col = cores,           # cores a usar
        width = 0.8,           # largura das barras
        # para dar espaço p legenda
        add=TRUE)

legend("topright",
       legend = rownames(tab1_by_areRB),
       pch = 15,
       col = cores)
```



#### 4. Hipóteses em teste + Estatística de teste

X - Estilo de Óculos de Sol (variável *type*)

Y - São ou Não da marca *SoleMio* (variável *are\_RB*)

##### Hipóteses em teste

$H_0$  : O estilo de óculos de sol é independente da marca RB

$H_1$  : Existe relacionamento entre o estilo de óculos de sol e a marca RB

ou, teoricamente,

$$H_0 : \forall (i, j) \in \{1 : r\} \times \{1 : c\} : p_{ij} = p_{i.} \times p_{.j}$$

$$H_1 : \exists (i, j) \in \{1 : r\} \times \{1 : c\} : p_{ij} \neq p_{i.} \times p_{.j}$$

##### Teste Qui-Quadrado

##### Estatística de teste

$$ET = \sum_{i=1}^3 \sum_{j=1}^3 \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \sim \chi^2_{(r-1)(c-1)}$$

Sendo ambas variáveis fatores, o estudo do relacionamento entre elas será feito através da análise do respetivo cruzamento (tabela de contingência), com a subsequente aplicação do **Teste Qui-quadrado de Pearson**.

## 5. Testar a independência das variáveis *type* e *are\_RB*

### Teste Qui-Quadrado

```
teste<-chisq.test(tab1)
teste
```

```
##
## Pearson's Chi-squared test
##
## data:  tab1
## X-squared = 39.268, df = 3, p-value = 1.523e-08
```

### Tabela com **Frequências Absolutas** VS **Frequências Absolutas Esperadas**

```
tab1_esperadas<-teste$expected
# tab1_esperadas_ <- addmargins(tab1_esperadas, FUN = sum)
```

	No	Yes	sum
Classic	161	93	254
Futuristic	23	2	25
Modern	267	50	317
Sport style	37	7	44
sum	488	152	640

	No	Yes	sum
Classic	193.675	60.325	254
Futuristic	19.062	5.938	25
Modern	241.713	75.287	317
Sport style	33.550	10.450	44
sum	488.000	152.000	640

## 6. e 7.

### Tabela com **Resíduos de Pearson** (teste[residuals]) e **Resíduos Estandarizados** (teste[stdres])

	No	Yes
Classic	-2.348	4.207
Futuristic	0.902	-1.616
Modern	1.627	-2.914
Sport style	0.596	-1.067

	No	Yes
Classic	-6.204	6.204
Futuristic	1.888	-1.888
Modern	4.698	-4.698
Sport style	1.267	-1.267

Como  $p - value = 1.522694 \times 10^{-8} < \alpha$  de referência ( $\alpha = 0.05$ ), então rejeita-se a  $H_0$ .

Ou seja, existe evidência estatística que revele associação entre o estilo dos óculos de sol e o facto de serem da marca RB, isto é, relacionamento entre as variáveis.

Isto significa que existem divergências significativas entre as frequências observadas e as frequências esperadas (ou seja, o que esperaríamos observar numa situação de independência).

## TC

**Forma do Relacionamento:** Valores Estandarizados acima de (aprox) 2: associações

Os óculos *Solemio* associam-se a estilo *clássico*; outras marcas de óculos associam-se a estilo *moderno* (valores destacados no quadro dos resíduos) isto é também visível e interpretável a partir dos gráficos, em especial o da direita.