# - Testes Qui-quadrado -

# Aderência e Independência

Agradecimentos às professoras
Monica Sandoval e Elisete Aubin

# 1. Testes de Aderência

Objetivo: Testar a adequabilidade de um modelo probabilístico a um conjunto de dados observados.

**Exemplo 1**: Segundo Mendel (geneticista famoso), os resultados dos cruzamentos de ervilhas amarelas redondas com ervilhas verdes enrugadas ocorrem na proporção de 9:3:3:1, ou seja, seguem uma distribuição de probabilidades dada por:

| Resultado | Amarela redonda | Amarela enrugada | Verde redonda | Verde enrugada |
|---|---|---|---|---|
| Probabilidade | 9/16 | 3/16 | 3/16 | 1/16 |

Uma amostra de **556 ervilhas** resultantes de cruzamentos de ervilhas amarelas redondas com ervilhas verdes enrugadas foi classificada da seguinte forma:

| Resultado | Amarela redonda | Amarela enrugada | Verde redonda | Verde enrugada |
|---|---|---|---|---|
| Frequência observada | 315 | 101 | 108 | 32 |

Há evidências de que os resultados desse experimento estão de acordo com a distribuição de probabilidades proposta por Mendel?

4 categorias para os resultados dos cruzamentos:

***Amarelas redondas (AR), Amarelas enrugadas (AE), Verdes redondas (VR), Verdes enrugadas (VE).***

Segundo Mendel, a probabilidade de cada categoria é dada por:

| | *AR* | *AE* | *VR* | *VE* |
|---|---|---|---|---|
| Probabilidades: (de Mendel) | 9/16 | 3/16 | 3/16 | 1/16 |

No experimento, 556 ervilhas foram classificadas segundo o tipo de resultado, fornecendo a tabela a seguir:

| Tipo de resultado | Frequência observada |
|---|---|
| ***AR*** | 315 |
| ***AE*** | 101 |
| ***VR*** | 108 |
| ***VE*** | 33 |
| Total | 556 |

Objetivo: Verificar se o modelo probabilístico proposto é adequado aos resultados do experimento.

Se o modelo probabilístico for adequado, a ***frequência esperada*** de ervilhas do tipo $\boldsymbol{AR}$, dentre as 556 observadas, pode ser calculada por:

$$556 \text{ x } P(AR) = 556 \text{ x } \mathbf{9/16} = 312{,}75$$

Da mesma forma, temos para o tipo $\boldsymbol{AE}$,

$$556 \text{ x } P(AE) = 556 \text{ x } \mathbf{3/16} = 104{,}25$$

Para o tipo $\boldsymbol{VR}$ temos

$$556 \text{ x } P(VR) = 556 \text{ x } \mathbf{3/16} = 104{,}25$$

E para o tipo $\boldsymbol{VE}$,

$$556 \text{ x } P(VE) = 556 \text{ x } \mathbf{1/16} = 34{,}75$$

Podemos expandir a tabela de frequências dada anteriormente:

| Tipo de resultado | Frequência observada | Frequência esperada (por Mendel) |
|---|---|---|
| ***AR*** | 315 | 312,75 |
| ***AE*** | 101 | 104,25 |
| ***VR*** | 108 | 104,25 |
| ***VE*** | 32 | 34,75 |
| **Total** | 556 | 556 |

→ Pergunta: Podemos afirmar que os valores observados estão suficientemente próximos dos valores esperados, de tal forma que o modelo probabilístico proposto por Mendel é adequado aos resultados desse experimento?

# Testes de Aderência – Metodologia

Considere uma tabela de frequências, com $\boldsymbol{k \geq 2}$ categorias de resultados:

| Categorias | Frequência Observada |
|---|---|
| 1 | $O_1$ |
| 2 | $O_2$ |
| 3 | $O_3$ |
| $\vdots$ | $\vdots$ |
| $k$ | $O_k$ |
| Total | $n$ |

em que $\boldsymbol{O_i}$ é o total de indivíduos ***observados*** na categoria $i$, $i = 1,\ldots,k$.

Seja $p_i$ a probabilidade associada à categoria $i$, $i$ = 1,..., $k$.

O objetivo do teste de aderência é testar as hipóteses

$$H_0 : p_1 = p_{o1} , \quad .... , p_k = p_{ok}$$

$H_1$: existe pelo menos uma diferença

sendo $\boldsymbol{p_{oi}}$ a probabilidade especificada para a categoria $i$, $i$ = 1, ..., $k$, **fixada através do modelo probabilístico de interesse**.

Se $\boldsymbol{E_i}$ é o total de indivíduos ***esperados*** na categoria $i$, quando a hipótese $H_0$ é verdadeira, então:

$$E_i = n \times p_{oi},\ i = 1, ...,k$$

Expandindo a tabela de frequências original, temos

| **Categorias** | **Frequência observada** | **Frequência esperada, sob $H_0$** |
|---|---|---|
| 1 | $O_1$ | $E_1$ |
| 2 | $O_2$ | $E_2$ |
| 3 | $O_3$ | $E_3$ |
| $\vdots$ | $\vdots$ | $\vdots$ |
| $k$ | $O_k$ | $E_k$ |
| Total | $n$ | $n$ |

Quantificação da distância entre as colunas de frequências:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Estatística do teste de aderência

Supondo $H_0$ verdadeira,

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \sim \chi_q^2, \quad \text{aproximadamente,}$$

sendo que $q = k - 1$ representa o número de graus de liberdade.

→ Em outras palavras, se $H_0$ é verdadeira, a v.a. $\chi^2$ tem distribuição aproximada qui-quadrado com $q$ graus de liberdade.

**IMPORTANTE.:** Este resultado é válido para ***n grande*** e para

$E_i \geq 5$, $i = 1, \ldots, k$.

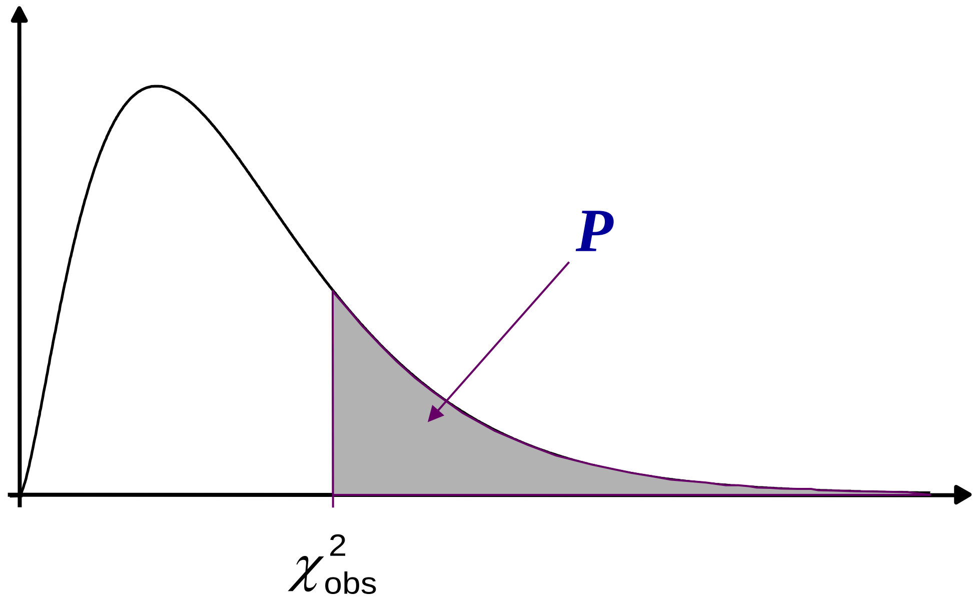
## Regra de decisão:

Pode ser baseada no **nível descritivo** ou **valor *P***, neste caso

$$P = P\,(\,\chi^2_q \geq \chi^2_{obs}),$$

em que $\chi^2_{\text{obs}}$ é o valor calculado, a partir dos dados, usando a expressão apresentada para $\chi^2$.

Graficamente:




Se, para $\alpha$ fixado, obtemos $P \leq \alpha$, **rejeitamos a hipótese $H_0$.**

## Exemplo (continuação): **Cruzamentos de ervilhas**

Hipóteses:

$H_0$: O modelo probabilístico proposto por Mendel é adequado.

$H_1$: O modelo proposto por Mendel não é adequado.

De forma equivalente, podemos escrever:

$H_0$: $P(AR)$ =9/16; $P(AE)$ = 3/16; $P(VR)$ = 3/16; $P(VE)$ = 1/16.

$H_1$: ao menos uma das igualdades não se verifica.

A tabela seguinte apresenta os valores observados e esperados (calculados anteriormente).

| Resultado | $O_i$ | $E_i$ |
|---|---|---|
| $AR$ | 315 | 312,75 |
| $AE$ | 101 | 104,25 |
| $VR$ | 108 | 104,25 |
| $VE$ | 32 | 34,75 |
| Total | 556 | 556 |

Cálculo do valor da estatística do teste ($k$ = $4$):

$$\chi^2_{obs} = \sum_1^4 \frac{(O_i - E_i)^2}{E_i} = \frac{(315-312{,}75)^2}{312{,}75} + \frac{(101-104{,}25)^2}{104{,}25} + \frac{(108-104{,}25)^2}{104{,}25} + \frac{(32-34{,}75)^2}{34{,}75} =$$

$$= 0{,}016 + 0{,}101 + 0{,}135 + 0{,}218 = 0{,}470.$$

Usando a distribuição de qui-quadrado com $q$ = $k$-1 = 3 graus de liberdade, o nível descritivo é calculado por $P = P(\chi^2_3 \geq 0{,}470) = 0{,}925.$

➡ Conclusão: Para $\alpha$ = 0,05, como ***P* = 0,925 > 0,05**, não há evidências para rejeitarmos a hipótese $H_0$, isto é, ao nível de significância de 5%, concluímos o modelo de probabilidades de Mendel se aplica aos resultados do experimento.

O cálculo do ***nível descritivo P*** pode ser feito pelo ***Rcmdr***, via menu, através do seguinte caminho:

***Distribuições → Distribuições contínuas → Distribuição Qui-Quadrado → Probabilidades da Qui-Quadrado → Cauda Superior***

Inserindo o **valor 0,470** e o número de **graus de liberdade igual a 3**, o valor ***P*** será igual a **0,925431**.

Também é possível utilizar tabela da distribuição de quiquadrado para obter, de forma aproximada, o valor do nível descritivo.

# 2. Testes de Independência

Objetivo: Verificar se existe independência entre duas variáveis medidas nas mesmas unidades experimentais.

Exemplo 2: A Associação de Imprensa do Estado de São Paulo fez um levantamento com 1300 leitores, para verificar se a preferência por leitura de um determinado jornal é independente do nível de instrução do indivíduo. Os resultados obtidos foram:

| | **Tipo de Jornal** | | | | |
|---|---|---|---|---|---|
| **Grau de instrução** | **Jornal *A*** | **Jornal *B*** | **Jornal *C*** | **Outros** | **Total** |
| **1º Grau** | 10 | 8 | 5 | 27 | 50 |
| **2º Grau** | 90 | 162 | 125 | 73 | 450 |
| **Universitário** | 200 | 250 | 220 | 130 | 800 |
| **Total** | 300 | 420 | 350 | 230 | 1300 |

Vamos calcular proporções segundo os totais das colunas (poderiam também ser calculadas pelos totais das linhas). Temos a seguinte tabela:

| | Tipo de Jornal | | | | |
|---|---|---|---|---|---|
| **Grau de instrução** | **Jornal *A*** | **Jornal *B*** | **Jornal *C*** | **Outros** | **Total** |
| **1º Grau** | 3,33% | 1,90% | 1,43% | 11,74% | 3,85% |
| **2º Grau** | 30,00% | 38,57% | 35,71% | 31,74% | 34,62% |
| **Universitário** | 66,67% | 59,52% | 62,86% | 56,52% | 61,54% |
| **Total** | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |

$\Rightarrow$ O que representam as porcentagens na colunas?

Distribuição de grau de instrução por tipo de jornal.

$\Rightarrow$ Independentemente da preferência por um tipo de jornal:

**3,85% dos leitores têm o 1º Grau,**

**34,62% têm o 2º Grau e**

**61,54% são universitários.**

Sob ***independência*** entre grau de instrução e preferência por um tipo de jornal, o número esperado de leitores que têm:

- 1º Grau e preferem $A$ é igual a 300 x 0,0385=11,54 (= 300x50/1300),
- 2º Grau e preferem $A$ é 300 x 0,3462 = 103,85 (= 300x450/1300),
- Universitários e preferem $A$ é 300 x 0,6154 = 184,62 (= 300x800/1300).

| | **Tipo de Jornal** | | | | |
|---|---|---|---|---|---|
| **Grau de instrução** | **Jornal $A$** | **Jornal $B$** | **Jornal $C$** | **Outros** | **Total** |
| **1º Grau** | 10 **11,54** **(3,85%)** | 8 **16,15** **(3,85%)** | 5 **13,46** **(3,85%)** | 27 **8,85** **(3,85%)** | **50** (3,85%) |
| **2º Grau** | 90 **103,85** **(34,62)%** | 162 **145,38** **(34,62%)** | 125 **121,15** **(34,62%)** | 73 **79,62** **(34,62%)** | **450** (34,62%) |
| **Universitário** | 200 **184,62** **(61,54%)** | 250 **258,46** **(61,54%)** | 220 **215,38** **(61,54%)** | 130 **141,54** **(61,54%)** | **800** (61,54%) |
| **Total** | 300 | 420 | 350 | 230 | 1300 |

As diferenças entre os valores observados e os esperados não são muito pequenas. Preferência por um tipo de jornal e grau de instrução parecem não ser *independentes*.

# Testes de Independência – Metodologia

Em geral, os dados referem-se a mensurações de duas características ($A$ e $B$) feitas em $n$ unidades experimentais, que são apresentadas conforme a seguinte tabela:

| $A \setminus B$ | $B_1$ | $B_2$ | ... | $B_s$ | **Total** |
|---|---|---|---|---|---|
| $A_1$ | $O_{11}$ | $O_{12}$ | ... | $O_{1s}$ | $O_{1.}$ |
| $A_2$ | $O_{21}$ | $O_{22}$ | ... | $O_{2s}$ | $O_{2.}$ |
| ... | ... | ... | ... | ... | ... |
| $A_r$ | $O_{r1}$ | $O_{r2}$ | ... | $O_{rs}$ | $O_{r.}$ |
| **Total** | $O_{.1}$ | $O_{.2}$ | ... | $O_{.s}$ | $n$ |

Hipóteses a serem testadas – **Teste de independência:**

$H_0$: $A$ e $B$ são variáveis independentes

$H_1$: As variáveis $A$ e $B$ não são independentes

→ Quantas observações devemos esperar em cada casela, se $A$ e $B$ forem independentes?

$$E_{ij} = \frac{O_{i.} \times O_{.j}}{n}$$

Distância entre os valores observados e os valores esperados sob a suposição de independência:

$$\chi^2 = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Supondo $H_0$ verdadeira,

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi_q^2$$

aproximadamente, sendo $\boldsymbol{q = (r - 1) \times (s - 1)}$ o número de **graus de liberdade**.

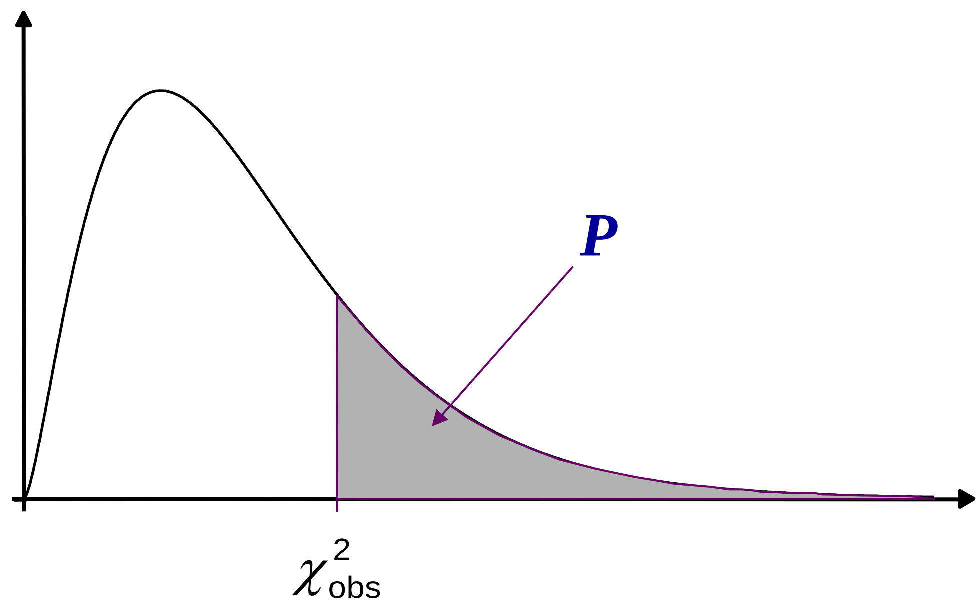
## Regra de decisão:

Pode ser baseada no **valor *P*** (nível descritivo), neste caso

$$P = P(\chi_q^2 \geq \chi_{obs}^2),$$

em que $\chi^2_{\text{obs}}$ é o valor calculado, a partir dos dados, usando a expressão apresentada para $\chi^2$.

Graficamente:




Se, para $\alpha$ fixado, obtemos $P \leq \alpha$, **rejeitamos a hipótese $H_0$ de independência.**

**Exemplo (continuação):** Estudo da independência entre preferência por um tipo de jornal e grau de instrução. 1300 eleitores foram entrevistados ao acaso.

**Hipóteses:** $H_0$: As variáveis preferência por um tipo de jornal e grau de instrução são independentes.

$H_1$: Existe dependência entre as variáveis.

Tabela de valores observados, sob $H_0$

| Grau de instrução | Tipo de Jornal | | | | |
|---|---|---|---|---|---|
| | Jornal A | Jornal B | Jornal C | Outros | Total |
| 1º Grau | 10 | 8 | 5 | 27 | 50 |
| 2º Grau | 90 | 162 | 125 | 73 | 450 |
| Universitário | 200 | 250 | 220 | 130 | 800 |
| Total | 300 | 420 | 350 | 230 | 1300 |

Exemplo do cálculo dos valores esperados sob $H_0$ (independência):

• Número esperado de leitores que têm 1º Grau e preferem o jornal $A$:

$$E_{11} = \frac{300 \times 50}{1300} = 11{,}54$$ .

# Tabela de valores observados e esperados (entre parênteses)

| | Tipo de Jornal | | | | |
|---|---|---|---|---|---|
| **Grau de instrução** | **Jornal $A$** | **Jornal $B$** | **Jornal $C$** | **Outros** | **Total** |
| **1º Grau** | **10** (11,54) | **8** (16,15) | **5** (13,46) | **27** (8,85) | **50** |
| **2º Grau** | **90** (103,85) | **162** (145,38) | **125** (121,15) | **73** (79,62) | **450** |
| **Universitário** | **200** (184,62) | **250** (258,46) | **220** (215,38) | **130** (141,54) | **800** |
| **Total** | **300** | **420** | **350** | **230** | **1300** |

2º Grau e prefere jornal $B$:

$$E_{22} = \frac{420 \times 450}{1300} = 145{,}38$$

Universitário e prefere outros jornais:

$$E_{34} = \frac{230 \times 800}{1300} = 141{,}54$$

Lembre-se:

$$E_{ij} = \frac{O_{i.} \times O_{.j}}{n_{..}}$$

# Cálculo da estatística de qui-quadrado:

| | Tipo de Jornal | | | | |
|---|---|---|---|---|---|
| Grau de instrução | Jornal $A$ | Jornal $B$ | Jornal $C$ | Outros | Total |
| 1º Grau | 10 (11,54) | 8 (16,15) | 5 (13,46) | 27 (8,85) | **50** |
| 2º Grau | 90 (103,85) | 162 (145,38) | 125 (121,15) | 73 (79,62) | **450** |
| Universitário | 200 (184,62) | 250 (258,46) | 220 (215,38) | 130 (141,54) | **800** |
| **Total** | **300** | **420** | **350** | **230** | **1300** |

$$\chi^2_{obs} = \frac{(10-11{,}54)^2}{11{,}54} + \frac{(8-16{,}15)^2}{16{,}15} + \frac{(5-13{,}46)^2}{13{,}46} + \frac{(27-8{,}85)^2}{8{,}85}$$

$$+ \frac{(90-103{,}85)^2}{103{,}85} + \frac{(162-145{,}38)^2}{145{,}38} + \frac{(125-121{,}15)^2}{121{,}15} + \frac{(73-79{,}62)^2}{79{,}62}$$

$$+ \frac{(200-184{,}62)^2}{184{,}62} + \frac{(250-258{,}46)^2}{258{,}46} + \frac{(220-215{,}38)^2}{215{,}38} + \frac{(130-141{,}54)^2}{141{,}54}$$

$$= 0{,}21 + 4{,}12 + 5{,}32 + 37{,}25 + 1{,}85 + 1{,}90 + 0{,}12 + 0{,}55 + 1{,}28 + 0{,}28 + 0{,}10 + 0{,}94$$

$$= 53{,}91.$$

## Determinação do **número de graus de liberdade:**

- Categorias de Grau de instrução: $s$ = 3
- Categorias de Tipo de jornal: $r$ = 4

$\Rightarrow$ **$q$ = ($r$ – 1)×($s$ – 1)=3×2 = 6**

O nível descritivo (valor $P$): $P = P\,(\,\chi^2_6 \geq 53{,}910) < 0{,}0001$

Supondo $\alpha =$ 0,05, temos $\boldsymbol{P} < \boldsymbol{\alpha}$ .

Assim, temos evidências para rejeitar a independência entre as variáveis grau de instrução e preferência por tipo de jornal ao nível de 5% de significância, i.é, a preferência por um jornal depende do grau de instrução do leitor.

**Os cálculos podem ser feitos diretamente no *Rcmdr*:**

***Estatísticas → Tabelas de Contingência → Digite e analise tabela de dupla entrada***

## Saída do *Rcmdr*:

```
data:  .Table
X-squared = 53.9099, df = 6, p-value = 7.692e-10


> .Test$expected # Expected Counts
          1         2         3          4
1  11.53846  16.15385  13.46154   8.846154
2 103.84615 145.38462 121.15385  79.615385
3 184.61538 258.46154 215.38462 141.538462

> round(.Test$residuals^2, 2) # Chi-square Components
     1    2    3     4
1 0.21 4.12 5.32 37.25
2 1.85 1.90 0.12  0.55
3 1.28 0.28 0.10  0.94
```

**Exercício 1:** Deseja-se verificar se o número de acidentes em uma estrada muda conforme o dia da semana. O número de acidentes observado para cada dia de uma semana escolhida aleatoriamente foram:

| Dia da semana | No. de acidentes |
|---|---|
| Seg | 20 |
| Ter | 10 |
| Qua | 10 |
| Qui | 15 |
| Sex | 30 |
| Sab | 20 |
| Dom | 35 |

$\Rightarrow$ O que pode ser dito?

$\Rightarrow$ Qual é o teste de hipóteses apropriado?

Hipóteses a serem testadas:

$H_0$: O número de acidentes não muda conforme o dia da semana;

$H_1$: Pelo menos um dos dias tem número diferente dos demais.

Se $p_i$ representa a probabilidade de ocorrência de acidentes no $i$-ésimo dia da semana,

$H_0$: $p_i$ = 1/7 para todo $i$ = 1,…, 7
$H_1$: $p_i \neq$ 1/7 para pelo menos um valor de $i$.

Total de acidentes na semana: $n$ =140.

Logo, se $H_0$ for verdadeira,

$$E_i = 140 \times 1/7 = 20, \; i = 1,\ldots,7,$$

ou seja, esperamos 20 acidentes por dia.

| Dia da semana | Nº. de acidentes observados ($O_i$) | Nº. esperado de acidentes ($E_i$) |
|---|---|---|
| Seg | 20 | 20 |
| Ter | 10 | 20 |
| Qua | 10 | 20 |
| Qui | 15 | 20 |
| Sex | 30 | 20 |
| Sab | 20 | 20 |
| Dom | 35 | 20 |

Cálculo da estatística de qui-quadrado:

$$\chi^2_{obs} = \sum_{1}^{7} \frac{(O_i - E_i)^2}{E_i} = \frac{(20-20)^2}{20} + \frac{(10-20)^2}{20} + \frac{(10-20)^2}{20} + \frac{(15-20)^2}{20} + \frac{(30-20)^2}{20} + \frac{(20-20)^2}{20} + \frac{(35-20)^2}{20} = 27{,}50$$

Neste caso, temos $\chi^2 \sim \chi_6^2$, aproximadamente.

O nível descritivo é dado por $P = P\,(\chi_6^2 \geq 27{,}50) \cong 0{,}00012$,

que pode ser obtido no ***Rcmdr*** pelo caminho (via menu):

***Distribuições → Distribuições contínuas → Distribuição Qui-Quadrado → Probabilidades da Qui-Quadrado → Cauda Superior***

(inserir o valor **27,50** e o número de graus de liberdade igual a **6**).

Conclusão: Para $\alpha$ = 0,05, temos que ***P* = 0,0001 < $\alpha$.**

Assim, há evidências para **rejeitarmos** $H_0$, ou seja, concluímos ao nível de significância de 5% que o número de acidentes não é o mesmo em todos os dias da semana.

**Exercício 2:** 1237 indivíduos adultos classificados segundo a pressão sanguínea ($mm\ Hg$) e o nível de colesterol ($mg/100cm^3$).

Verificar se existe independência entre essas variáveis.

| Colesterol | Pressão | | | Total |
|---|---|---|---|---|
| | < 127 | 127 a 166 | > 166 | |
| < 200 | 117 | 168 | 22 | 307 |
| 200 a 260 | 204 | 418 | 63 | 685 |
| > 260 | 67 | 145 | 33 | 245 |
| Total | 388 | 731 | 118 | 1237 |

Hipóteses:

$H_0$: Pressão sanguínea e nível de colesterol são independentes;

$H_1$: Nível de colesterol e pressão sanguínea são variáveis dependentes

## ***Rcmdr***: *Estatísticas → Tabelas de Contingência → Digite e análise tabela de dupla entrada*

**Saída do *Rcmdr*:**

```
data:  .Table
X-squared = 13.5501, df = 4, p-value = 0.008878

> .Test$expected # Expected Counts
          1          2          3
1  96.29426  181.4204  29.28537
2 214.85853  404.7979  65.34357
3  76.84721  144.7817  23.37106

> round(.Test$residuals^2, 2) # Chi-square Components
      1    2    3
1  4.45 0.99 1.81
2  0.55 0.43 0.08
3  1.26 0.00 3.97
```

Para $\alpha$ = 0,05, temos $\boldsymbol{P < \alpha}$. Assim, temos evidências para rejeitar a hipótese de independência entre as variáveis pressão sanguínea e nível de colesterol ao nível de 5% de significância.

**Exercício 3:** Uma indústria, desejando melhorar o nível de seus funcionários em cargos de chefia, montou 3 cursos experimentais de inglês utilizando 2 metodologias distintas ($MA$, $MB$). Os dados referentes ao conceito obtido no curso ($A$, $B$ ou $C$) e metodologia utilizada estão na tabela no slide a seguir:

(a) Identifique as variáveis em estudo. Classifique-as.

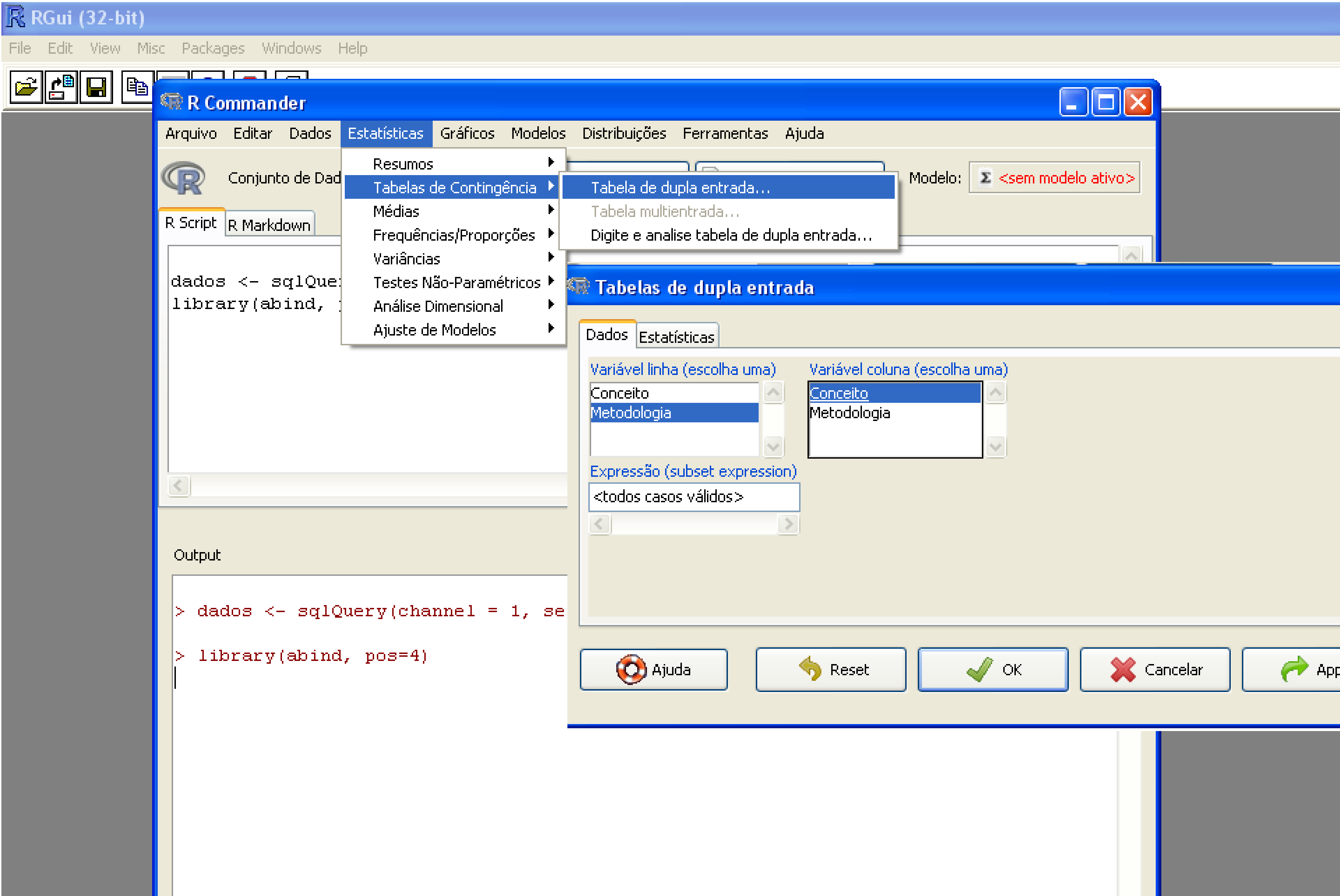
(b) Construa uma tabela de contingência para as variáveis "metodologia" e "conceito".

(c) Conclua se existe associação entre essas variáveis ($\alpha$ = 10%).

# Dados:

| Funcionário | Metodologia | Conceito |
|---|---|---|
| 1 | *MA* | *A* |
| 2 | *MA* | *B* |
| 3 | *MB* | *A* |
| 4 | *MB* | *B* |
| 5 | *MA* | *A* |
| 6 | *MA* | *B* |
| 7 | *MA* | *C* |
| 8 | *MB* | *B* |
| 9 | *MB* | *B* |
| 10 | *MA* | *B* |
| 11 | *MB* | *C* |
| 12 | *MB* | *A* |
| 13 | *MB* | *B* |
| 14 | *MB* | *A* |
| 15 | *MB* | *C* |
| 16 | *MA* | *A* |
| 17 | *MA* | *B* |
| 18 | *MB* | *C* |
| 19 | *MA* | *C* |
| 20 | *MB* | *C* |
| 21 | *MB* | *A* |
| 22 | *MA* | *C* |
| 23 | *MB* | *C* |
| 24 | *MA* | *A* |
| 25 | *MA* | *B* |
| 26 | *MB* | *B* |
| 27 | *MA* | *A* |
| 28 | *MB* | *C* |
| 29 | *MA* | *A* |
| 30 | *MA* | *B* |
| 31 | *MA* | *A* |
| 32 | *MA* | *A* |
| 33 | *MB* | *B* |
| 34 | *MB* | *B* |
| 35 | *MA* | *A* |
| 36 | *MA* | *A* |
| 37 | *MA* | *A* |
| 38 | *MB* | *B* |
| 39 | *MB* | *C* |
| 40 | *MB* | *C* |

# *Rcmdr*: Construção da tabela de contingência (ou tabela de frequencias conjuntas)



```
> .Table             Conceito
        Metodologia  A  B  C
                 MA 11  6  3
                 MB  4  8  8

> rowPercents(.Table) # Row Percentages
              Conceito
Metodologia  A  B  C Total Count
         MA 55 30 15   100    20
         MB 20 40 40   100    20

X-squared = 5.8251, df = 2, p-value = 0.05434

> .Test$expected # Expected Counts
             Conceito
Metodologia   A B   C
         MA 7.5 7 5.5
         MB 7.5 7 5.5

> round(.Test$residuals^2, 2) # Chi-square Components
             Conceito
Metodologia    A    B    C
         MA 1.63 0.14 1.14
         MB 1.63 0.14 1.14
```

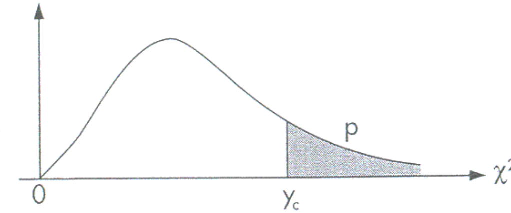
Conclusão: $\alpha = 0{,}10 \Rightarrow P < \alpha$.
Assim, os dados indicam que, ao nível de 10% de significância, o conceito depende da metodologia.

**Tabela IV** — Distribuição Qui-quadrado

$Y \sim \chi^2(\nu)$

Corpo da tabela dá os valores $y_c$ tais que $P(Y > y_c) = p$.

Para valores $\nu > 30$, use a aproximação normal dada no texto.




| Graus de liberdade $\nu$ | p = 99% | 98% | 97,5% | 95% | 90% | 80% | 70% | 50% | 30% | 20% | 10% | 5% | 4% | 2,5% | 2% | 1% | 0,2% | 0,1% | Graus de liberdade $\nu$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $0{,}0^{3}16$ | $0{,}0^{3}63$ | 0,001 | 0,004 | 0,016 | 0,064 | 0,148 | 0,455 | 1,074 | 1,642 | 2,706 | 3,841 | 4,218 | 5,024 | 5,412 | 6,635 | 9,550 | 10,827 | 1 |
| 2 | 0,020 | 0,040 | 0,051 | 0,103 | 0,211 | 0,446 | 0,713 | 1,386 | 2,408 | 3,219 | 4,605 | 5,991 | 6,438 | 7,378 | 7,824 | 9,210 | 12,429 | 13,815 | 2 |
| 3 | 0,115 | 0,185 | 0,216 | 0,352 | 0,584 | 1,005 | 1,424 | 2,366 | 3,665 | 4,642 | 6,251 | 7,815 | 8,311 | 9,348 | 9,837 | 11,345 | 14,796 | 16,266 | 3 |
| 4 | 0,297 | 0,429 | 0,484 | 0,711 | 1,064 | 1,649 | 2,195 | 3,357 | 4,878 | 5,989 | 7,779 | 9,488 | 10,026 | 11,143 | 11,668 | 13,277 | 16,924 | 18,467 | 4 |
| 5 | 0,554 | 0,752 | 0,831 | 1,145 | 1,610 | 2,343 | 3,000 | 4,351 | 6,064 | 7,289 | 9,236 | 11,070 | 11,644 | 12,832 | 13,388 | 15,086 | 18,907 | 20,515 | 5 |
| 6 | 0,872 | 1,134 | 1,237 | 1,635 | 2,204 | 3,070 | 3,828 | 5,348 | 7,231 | 8,558 | 10,645 | 12,592 | 13,198 | 14,449 | 15,033 | 16,812 | 20,791 | 22,457 | 6 |
| 7 | 1,239 | 1,564 | 1,690 | 2,167 | 2,833 | 3,822 | 4,671 | 6,346 | 8,383 | 9,803 | 12,017 | 14,067 | 14,703 | 16,013 | 16,622 | 18,475 | 22,601 | 24,322 | 7 |
| 8 | 1,646 | 2,032 | 2,180 | 2,733 | 3,490 | 4,594 | 5,527 | 7,344 | 9,524 | 11,030 | 13,362 | 15,507 | 16,171 | 17,534 | 18,168 | 20,090 | 24,352 | 26,125 | 8 |
| 9 | 2,088 | 2,532 | 2,700 | 3,325 | 4,168 | 5,380 | 6,393 | 8,343 | 10,656 | 12,242 | 14,684 | 16,919 | 17,608 | 19,023 | 19,679 | 21,666 | 26,056 | 27,877 | 9 |
| 10 | 2,558 | 3,059 | 3,247 | 3,940 | 4,865 | 6,179 | 7,267 | 9,342 | 11,781 | 13,442 | 15,987 | 18,307 | 19,021 | 20,483 | 21,161 | 23,209 | 27,722 | 29,588 | 10 |
| 11 | 3,053 | 3,609 | 3,816 | 4,575 | 5,578 | 6,989 | 8,148 | 10,341 | 12,899 | 14,631 | 17,275 | 19,675 | 20,412 | 21,920 | 22,618 | 24,725 | 29,354 | 31,264 | 11 |
| 12 | 3,571 | 4,178 | 4,404 | 5,226 | 6,304 | 7,807 | 9,034 | 11,340 | 14,011 | 15,812 | 18,549 | 21,026 | 21,785 | 23,337 | 24,054 | 26,217 | 30,957 | 32,909 | 12 |
| 13 | 4,107 | 4,765 | 5,009 | 5,892 | 7,042 | 8,634 | 9,926 | 12,340 | 15,119 | 16,985 | 19,812 | 22,362 | 23,142 | 24,736 | 25,472 | 27,688 | 32,535 | 34,528 | 13 |
| 14 | 4,660 | 5,368 | 5,629 | 6,571 | 7,790 | 9,467 | 10,821 | 13,339 | 16,222 | 18,151 | 21,064 | 23,685 | 24,485 | 26,119 | 26,873 | 29,141 | 34,091 | 36,123 | 14 |
| 15 | 5,229 | 5,985 | 6,262 | 7,261 | 8,547 | 10,307 | 11,721 | 14,339 | 17,322 | 19,311 | 22,307 | 24,996 | 25,816 | 27,488 | 28,259 | 30,578 | 35,628 | 37,697 | 15 |
| 16 | 5,812 | 6,614 | 6,908 | 7,962 | 9,312 | 11,152 | 12,624 | 15,338 | 18,418 | 20,465 | 23,542 | 26,296 | 27,136 | 28,845 | 29,633 | 32,000 | 37,146 | 39,252 | 16 |
| 17 | 6,408 | 7,255 | 7,564 | 8,672 | 10,085 | 12,002 | 13,531 | 16,338 | 19,511 | 21,615 | 24,769 | 27,587 | 28,445 | 30,191 | 30,995 | 33,409 | 38,648 | 40,790 | 17 |
| 18 | 7,015 | 7,906 | 8,231 | 9,390 | 10,865 | 12,857 | 14,440 | 17,338 | 20,601 | 22,760 | 25,989 | 28,869 | 29,745 | 31,526 | 32,346 | 34,805 | 40,136 | 42,312 | 18 |
| 19 | 7,633 | 8,567 | 8,906 | 10,117 | 11,651 | 13,716 | 15,352 | 18,338 | 21,689 | 23,900 | 27,204 | 30,144 | 31,037 | 32,852 | 33,687 | 36,191 | 41,610 | 43,820 | 19 |
| 20 | 8,260 | 9,237 | 9,591 | 10,851 | 12,443 | 14,578 | 16,266 | 19,337 | 22,775 | 25,038 | 28,412 | 31,410 | 32,321 | 34,170 | 35,020 | 37,566 | 43,072 | 45,315 | 20 |
| 21 | 8,897 | 9,915 | 10,283 | 11,591 | 13,240 | 15,445 | 17,182 | 20,337 | 23,858 | 26,171 | 29,615 | 32,671 | 33,597 | 35,479 | 36,343 | 38,932 | 44,522 | 46,797 | 21 |
| 22 | 9,542 | 10,600 | 10,982 | 12,338 | 14,041 | 16,314 | 18,101 | 21,337 | 24,939 | 27,301 | 30,813 | 33,924 | 34,867 | 36,781 | 37,659 | 40,289 | 45,962 | 48,268 | 22 |
| 23 | 10,196 | 11,293 | 11,688 | 13,091 | 14,848 | 17,187 | 19,021 | 22,337 | 26,018 | 28,429 | 32,007 | 35,172 | 36,131 | 38,076 | 38,968 | 41,638 | 47,391 | 49,728 | 23 |
| 24 | 10,856 | 11,992 | 12,401 | 13,848 | 15,659 | 18,062 | 19,943 | 23,337 | 27,096 | 29,553 | 33,196 | 36,415 | 37,389 | 39,364 | 40,270 | 42,980 | 48,812 | 51,179 | 24 |
| 25 | 11,524 | 12,697 | 13,120 | 14,611 | 16,473 | 18,940 | 20,867 | 24,337 | 28,172 | 30,675 | 34,382 | 37,652 | 38,642 | 40,646 | 41,566 | 44,314 | 50,223 | 52,620 | 25 |
| 26 | 12,198 | 13,409 | 13,844 | 15,379 | 17,292 | 19,820 | 21,792 | 25,336 | 29,246 | 31,795 | 35,563 | 38,885 | 39,889 | 41,923 | 42,856 | 45,642 | 51,627 | 54,052 | 26 |
| 27 | 12,879 | 14,125 | 14,573 | 16,151 | 18,114 | 20,703 | 22,719 | 26,336 | 30,319 | 32,912 | 36,741 | 40,113 | 41,132 | 43,194 | 44,140 | 46,963 | 53,022 | 55,476 | 27 |
| 28 | 13,565 | 14,847 | 15,308 | 16,928 | 18,939 | 21,588 | 23,647 | 27,336 | 31,319 | 34,027 | 37,916 | 41,337 | 42,370 | 44,461 | 45,419 | 48,278 | 54,411 | 56,893 | 28 |
| 29 | 14,258 | 15,574 | 16,047 | 17,708 | 19,768 | 22,475 | 24,577 | 28,336 | 32,461 | 35,139 | 39,087 | 42,557 | 43,604 | 45,722 | 46,693 | 49,588 | 55,792 | 58,302 | 29 |
| 30 | 14,953 | 16,306 | 16,791 | 18,493 | 20,599 | 23,364 | 25,508 | 29,336 | 33,530 | 36,250 | 40,256 | 43,773 | 44,834 | 46,979 | 47,962 | 50,892 | 57,167 | 59,703 | 30 |
| | p = 99% | 98% | 97,5% | 95% | 90% | 80% | 70% | 50% | 30% | 20% | 10% | 5% | 4% | 2,5% | 2% | 1% | 0,2% | 0,1% | |