# QUESTIONÁRIO 11 - TESTES DE ADERÊNCIA E INDEPENDÊNCIA

*Jorge Bazán e Patrícia Stülp*

**Exercício 1:** Um hospital do interior paulista deseja saber se os atendimentos de suas UPAs são uniformemente distribuídos durante os dias da semana, a fim de organizar melhor sua equipe de atendimento. Durante certo período foram coletados os seguintes dados.

| Seg | Ter | Qua | Qui | Sex | Sab | Dom |
|---|---|---|---|---|---|---|
| 0.143 | 0.126 | 0.145 | 0.129 | 0.142 | 0.146 | 0.169 |

Calcule a estatística do teste adequada e obtenha a conclusão do mesmo ao nível de significância de 2%.

1) (Considere três casas decimais) Valor da estatística do teste.
2) Rejeita H0 ou não rejeita H0?

Respostas corretas:

1) Neste caso, temos as seguintes hipóteses

   H0: Os dados seguem o modelo proposto

   H1: Os dados não seguem tal modelo

   cuja estatística de teste é dada por

$$Q^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i},$$

   em que $Q^2 \sim \chi^2_{k-q-1}$.

   Note que $o_i$ são os dados apresentados na tabela acima. Já $e_i$ são os valores esperados. Assumindo-se que H0 seja verdadeira, então os dados provém de uma distribuição Uniforme e, com isso, espera-se que a frequência seja a mesma em todos os casos ($\frac{1}{7} = 0.143$), ou seja,

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $o_i$ | 0.143 | 0.126 | 0.145 | 0.129 | 0.142 | 0.146 | 0.169 |
| $e_i$ | 0.143 | 0.143 | 0.143 | 0.143 | 0.143 | 0.143 | 0.143 |

   Portanto,

$$Q^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} = 0.008$$

2) Usando o R para determinar se vamos rejeitar H0 ou não:

```
alpha = 0.02

# regiao critica (Teste unilateral)
k  = length(obs)  # tamanho dos dados
q  = 2
RC = round(qchisq(1 - alpha, k - q - 1), 3); RC

# 11.668
```

Ou seja, para $\alpha = 0.02$ e $k - q - 1 = 4$ graus de liberdade, temos que RC $= 11.668 >$ $\mathrm{Q}^2 = 0.008$ e, portanto, não há evidências estatísticas para rejeitar H0. Dessa forma, a um nível de sigificância de 2% há indícios de que os dados seguem uma distribuição Uniforme.

**Exercício 2:** Uma empresa de comunicação no estado de São Paulo fez um levantamento com 1431 usuários de seus recursos midiáticos. A empresa deseja verificar se a preferência por um determinado canal de informação para se inteirar das notícias é independente do nível de instrução do indivíduo. Os dados obtidos foram:

| Grau de Instrução | Internet | Televisão | Rede Social | Outros |
|---|---|---|---|---|
| Fundamental | 10 | 46 | 29 | 23 |
| Médio | 76 | 122 | 117 | 154 |
| Superior | 210 | 244 | 151 | 178 |

Ao testar a hipótese de pesquisa, considerando um nível de significância de 10%, obtemos:

1) (Considere três casas decimais) Valor da estatística do teste.

2) As variáveis são dependentes ou independentes?

Respostas corretas:

1) Neste caso, temos as seguintes hipóteses

   H0: As variáveis são independentes

   H1: As variáveis não são independentes

   cuja estatística de teste é dada por

$$Q^2 = \sum_{i=1}^{r} \sum_{i=1}^{s} \frac{(o_{ij} - e_{ij})^2}{e_{ij}},$$

   em que $Q^2 \sim \chi^2_{(r-1)(s-1)}$.

   Note que se as variáveis forem independentes, então a distribuição conjunta delas é igual ao produto de suas marginais, ou seja

| Grau de Instrução | Internet | Televisão | Rede Social | Outros | Total |
|---|---|---|---|---|---|
| Fundamental | 10 (23,51) | 46 (32.72) | 29 (23.58) | 23 (28.19) | 108 |
| Médio | 76 (102.07) | 122 (142.08) | 117 (102.42) | 154 (122.42) | 469 |
| Superior | 210 (170.42) | 244 (237.20) | 151 (170.99) | 178 (204.38) | 783 |
| Total | 296 | 412 | 297 | 355 | 1360 |

Portanto,

$$
\begin{aligned}
Q^2 &= \sum_{i=1}^{r}\sum_{i=1}^{s}\frac{(o_{ij}-e_{ij})^2}{e_{ij}}\\
&= \frac{(10-23,51)^2}{23,51}+\frac{(76-102.07)^2}{102.07}+\frac{(210-170.42)^2}{170.42}\\
&\quad +\frac{(46-32.72)^2}{32.72}+\frac{(122-142.08)^2}{142.08}+\frac{(244-237.20)^2}{237.20}\\
&\quad +\frac{(29-23.58)^2}{23.58}+\frac{(117-102.42)^2}{102.42}+\frac{(151-170.99)^2}{170.99}\\
&\quad +\frac{(23-28.19)^2}{28.19}+\frac{(154-122.42)^2}{122.42}+\frac{(178-204.38)^2}{204.38}\\
&= 50.203
\end{aligned}
$$

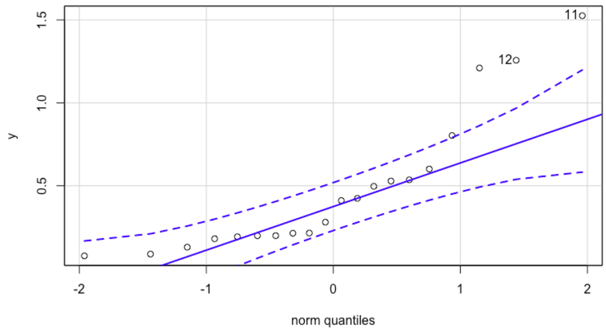
2) Usando o R para determinar se vamos rejeitar H0 ou não:

```
alpha = 0.10

# numero de linhas e colunas
r = nrow(dados); s = ncol(dados)

# regiao critica (Teste unilateral)
qc = round(qchisq(1 - alpha, (r-1)*(s-1)), 3); qc

# 10.645
```

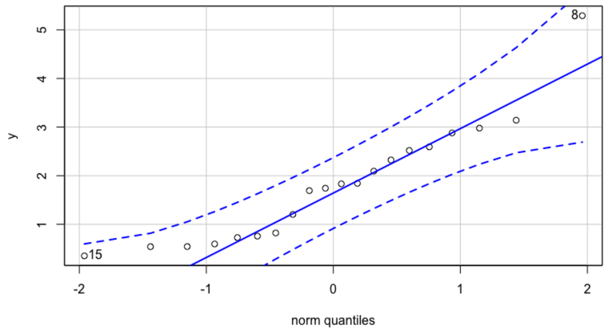
Ou seja, para $\alpha = 0.10$ e $(r-1)(s-1) = 6$ graus de liberdade, temos que qc $= 10.645 <$ $\mathrm{Q}^2 = 50.203$ e, portanto, há evidências estatísticas para rejeitar H0. Dessa forma, a um nível de sigificância de 10% podemos dizer que as variáveis são dependentes.

**Exercício 3:** Os gráficos a seguir apresentam o Qqplot para três conjunto de dados. Para quais deles é razoável assumir normalidade?

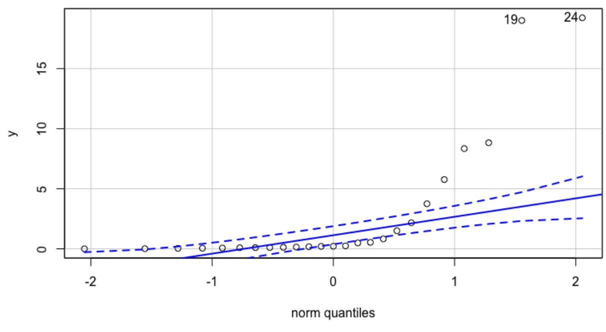
a)




b)




c)




Resposta correta: b, pois todos os pontos encontram-se dentro das bandas de confiança. Dessa forma, podemos dizer que é razoável assumir uma distribuição Normal para os dados.