



Passo-a-passo

ETAPA 12. TESTE QUI-QUADRADO

Prof. Pedro Feliú

INTRODUÇÃO

Na presente etapa 12, realizaremos um teste Qui-quadrado. Buscaremos apontar alguns aspectos intuitivos centrais para a interpretação do teste que realizaremos nesta etapa, que é menos recheada de comandos do Rstudio e equilibrada com a parte substantiva do teste estatístico que propomos realizar. Vamos utilizar um banco de dados novo, chamado “Entrada_Venezuela_CN”.

O Teste Qui-quadrado

Muitas vezes, principalmente nas ciências humanas, utilizamos variáveis que não são mensuradas numericamente, podendo apenas ser alocadas em categorias (variáveis categóricas). Por exemplo, o sexo da pessoa (feminino ou masculino) ou a religião (budista, espírita, islâmico, etc.). **É importante ressaltar que o elemento pesquisado deve se encaixar em apenas uma categoria.** O teste Qui-quadrado permite a análise de amostras **independentes**, com a variável resposta categórica (qualitativa). O objetivo central deste teste é verificar se duas variáveis estão associadas, isto é, **existe associação entre duas variáveis se o conhecimento de uma altera a probabilidade de algum resultado da outra.** Por exemplo, podemos dizer que há uma associação entre a propensão de uma pessoa ir à praia e o clima. Vale ressaltar que encontrar uma **associação não significa estabelecer uma relação causal.** Desta forma, medir o grau de associação entre duas variáveis qualitativas é o objetivo deste teste.

Para tanto, utilizaremos a votação no Congresso Nacional do Brasil, Câmara e Senado, acerca da aprovação da entrada da Venezuela no MERCOSUL em 2007. **Queremos analisar o nível de significância da associação entre votar sim ou não à entrada do vizinho caribenho e dois fatores separadamente: 1 - o pertencimento ou não do legislador à coalizão de governo; 2 – a ideologia do partido político (esquerda ou direita).** Analisamos, portanto, a associação entre duas variáveis categóricas duas vezes (realizaremos dois testes Qui-quadrado). A variável dependente é sempre o voto do legislador (que pode ser sim ou não). Vale ressaltar que os dados se encaixam em apenas uma das categorias dispostas, ou seja, um legislador não pode votar sim e não no mesmo projeto, como também não pode pertencer e não pertencer à coalizão de governo ao mesmo tempo, por exemplo. Vamos agora iniciar os primeiros passos para realizar o teste do Qui-quadrado, depois retomamos algumas características do teste:

PASSO 1: Importação do banco

Esse passo possui um conteúdo já conhecido. Iremos importar ao Rstudio o banco de dados denominado “Entrada_Venezuela_CN”. Por conveniência chamaremos o objeto de “venezuela”.

```
library(foreign)
venezuela<-
read_xls("C:/Users/Paulo/Documents/Documents/CursoR_Apolo/Bases de
dados/Entrada_Venezuela_CN.xls")
```

Iniciação no R com exemplos de Política Internacional

PASSO 2: Criar *labels* para as variáveis do banco de dados venezuela

```
venezuela$Ideologia <- factor(venezuela$Ideologia,  
  levels = c(0,1),  
  labels = c("Direita", "Esquerda"))
```

```
venezuela$Coalizão <- factor(venezuela$Coalizão,  
  levels = c(0,1),  
  labels = c("Oposição", "Governo"))
```

```
venezuela$Voto <- factor(venezuela$Voto,  
  levels = c(0,1),  
  labels = c("Não", "Sim"))
```

Os comandos acima criam rótulos (labels) para as três variáveis que vamos utilizar (Ideologia, Coalizão e Voto). Todas as variáveis originais são dicotômicas e computadas como “0” e “1”. Por isso usamos na função “factor”, a opção “levels = c(0,1)”. Em seguida no comando, adicionamos a opção “labels” com os nomes que daremos para cada número nessa ordem 0,1. Assim, a ideologia 0 é direita e a 1 esquerda. A coalizão 0 é oposição e a 1 governo. E o voto 0 é não e o voto 1 é sim. Olhem o banco de dados como ficou. Aprendemos a usar rótulos nas variáveis. Vamos agora usar a conhecida função “attach” abaixo e depois prosseguir com o teste.

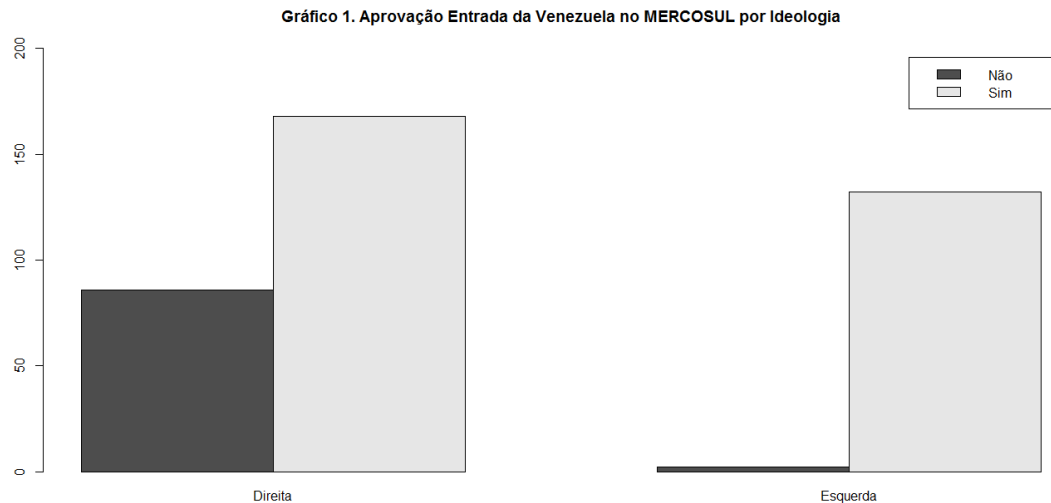
attach (Venezuela)

PASSO 3: Geração de uma tabela 2 x2 e um gráfico de barras

Vamos agora preparar os dados para realizar o primeiro teste Qui-quadrado, averiguando a existência ou não de associação entre o voto do legislador e a ideologia do partido político do mesmo.

```
t1 = table(Voto,Ideologia)  
barplot(t1, beside=TRUE, legend=TRUE, ylim=c(0,200),main="Gráfico 1.  
Aprovação Entrada da Venezuela no MERCOSUL por Ideologia")
```

Iniciação no R com exemplos de Política Internacional



No gráfico 1 acima o eixo vertical revela o número de deputados. Realizaremos o primeiro teste Qui-quadrado para confirmação ou rejeição da hipótese principal (H_1):

$$H_0: \pi_{\text{Direita}} = \pi_{\text{Esquerda}}$$

$$H_1: \pi_{\text{Direita}} \neq \pi_{\text{Esquerda}}$$

A hipótese principal, denominada H_1 , dispõe que o comportamento do legislador (o seu voto) é afetado pela ideologia política do partido, ou seja, não é igual o voto dos esquerdistas e direitistas. A hipótese nula, denominada H_0 , dispõe que tanto faz o legislador ser de direita ou esquerda no que tange o seu voto nesta matéria, o comportamento pode ser considerado igual.

O teste Qui-quadrado é estimado por meio da seguinte fórmula: $X^2 = \sum \frac{(O-E)^2}{E}$, onde X^2 é o resultado do teste; O – é o observado de cada célula, ou seja, cada cruzamento de linha com coluna da tabela t1 criada anteriormente; E – é o resultado esperado de cada célula, representando a hipótese nula (H_0) de independência entre as variáveis e pode ser obtido por:

$$E = \frac{(total\ da\ linha) \times (total\ da\ coluna)}{(total\ geral)}$$

Vale lembrar que esse cálculo corresponde à tabela t1 2x2. Vamos agora para o passo 4, onde o Rstudio calculará o X^2 (Qui-quadrado).

PASSO 4: Estimação do Qui-quadrado

Utilizaremos a função **chisq.test** para estimar o Qui-quadrado (X^2) por meio da tabela “t1” criada, entre o voto e a ideologia. Segue o comando:

```
chisq.test(t1)
```

```
Pearson's Chi-squared test with Yates' continuity correction  
data: t1  
X-squared = 50.571, df = 1, p-value = 1.149e-12  
> |
```

Para interpretar o resultado acima definimos cada item mostrado pelo Rstudio:

X-squared: é o resultado do X^2 , apontado anteriormente, neste caso 50,571, um valor alto.

df: é *degrees of freedom*, ou seja, graus de liberdade (gl em português). $df = (L-1).(C-1)$, onde L é o número de linhas e C é o número de colunas da tabela “t1”. Como temos duas linhas e duas colunas, fica: $df=(2-1).(2-1)=1$. A forma da distribuição Qui-quadrado torna-se menos assimétrica à medida que cresce o número de graus de liberdade.

p-value: ou **probabilidade de significância** (p-valor). O p-value é a probabilidade da estatística Qui-quadrado acusar um valor maior ou igual do que o valor de X^2 , calculado sob uma base amostral. As referências para o p-value são 99% (0.01), 95% (0.05) e 90% (0.10). Como neste caso o p-value é 1.149e-12 (ou 0.0000001149), muito menor que 0.01, 0.05 e 0.10, podemos concluir que há significância estatística no teste, aceitando a hipótese primeira (H_1), ou seja, há associação entre o voto do legislador e a ideologia política do seu partido. Quando os dados revelam um X^2 pequeno e, conseqüentemente, um valor de p grande (p-value), o teste aceita H_0 e rejeita H_1 .

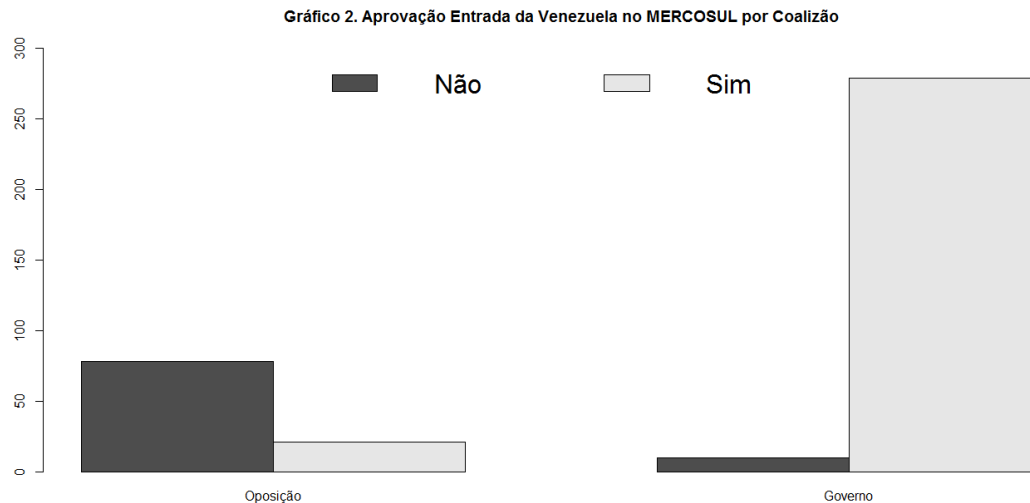
Para absorver melhor o conteúdo desta etapa, que é um pouco mais densa do ponto de vista de seu conteúdo estatístico, vamos realizar mais um teste Qui-quadrado. Vamos realizar no próximo passo os comandos para criação da t2, relacionando “Voto” e “Coalizão”, do mesmo banco de dados.

PASSO 5: Criar tabela t2 e gráfico de barras.

Geraremos uma tabela simples de duas linhas e duas colunas (2x2) com o comando **table**, nomeando esta tabela de “t2”. Na linha dispomos o voto e na coluna a coalizão. Em seguida realizamos um gráfico de barras para ver descritivamente o comportamento de ambas as variáveis (“Voto” e “Coalizão”). Seguem os comandos:

```
t2 = table(Voto,Coalizão)  
barplot(t2, beside=TRUE, legend=TRUE, ylim=c(0,300) , args.legend = list(bty =  
"n", x = "top", cex=2.0, ncol = 2), main="Gráfico 2. Aprovação Entrada da  
Venezuela no MERCOSUL por Coalizão")
```

Iniciação no R com exemplos de Política Internacional



Vejam que no gráfico acima eu usei o comando “args.legend” para mudar a posição, o tamanho e o estilo da legenda. Os comandos `bty = "n"`, `x = "top"`, `cex=2.0`, `ncol = 2` na sequência fazem: tira as linhas em volta da legenda, situa a legenda no topo do gráfico, estabelece o tamanho 2.0 para a legenda (pode ser maior ou menor, basta alterar o número no comando) e estabelece que sejam duas cores na legenda.

PASSO 6: Qui-quadrado voto x coalizão

`chisq.test(t2)`

```
> chisq.test(t2)

Pearson's Chi-squared test with Yates' continuity correction

data:  t2
X-squared = 234.32, df = 1, p-value < 2.2e-16

> |
```

O teste resultou em um Qui-quadrado maior do que anterior, bem alto, com o mesmo df (1), pois a tabela também é 2x2 e um p-value de 2.2e-16 (0.000000000000000022), também muito significativo. Concluímos que há associação entre o voto dos legisladores nesta matéria e o pertencimento ou não à coalizão de governo. O teste revelou que esta associação é ainda mais forte do que a ideologia, sendo a variável principal para entender o posicionamento legislativo na entrada da Venezuela no MERCOSUL em 2007.