# Testes de Aderência

*Pedro Menezes de Araújo, Pedro Pires Costa*

*13 de julho de 2017*

## Introdução

Os testes de aderência (ou bondade de ajuste) são muito utilizados na rotina estatítica clássica. Eles servem para testar se uma determinada amostra aleatória é gerada por uma distribuição específica. São muito úteis, pois em boa parte dos casos não se sabe a distribuição de probabilidade da população no qual a amostra é proveniente. Tal informação é necessária, por exemplo, para decidir quais testes de hipóteses podem ser usados em análises posteriores. O presente trabalho tem por objetivo apresentar e comparar os teste de aderência Qui-quadrado, Kolomogorov-Smirnov e o teste de Cramér-von Mises através de simulações no ambiente R.

Serão geradas 1000 amostras aleatórias com distribuição $Normal(0,1)$, $exp(1)$ e $U[0,1]$ de tamanhos, 10, 100, 1000, 10000, sendo aplicadas em cada teste com nível de significância $\alpha = 0.1$, tendo em vista que nesse tipo deste $H_0$ é a hipótese do interesse (seguir a distribuição pré-especificada pelo pesquisador), com finalidade de comparar a taxa de acerto de cada teste.

## Teste Qui-quadrado

O teste qui-quadrado é bastante utilizado para testar a independência entre variáveis categóricas, mas pode ser adaptado para testar a aderência de amostra aleatória em relação a uma distribuição. O teste tenta captar a diferença entre valor observado e o valor esperado em intervalos *****. Uma restrição para o teste é que os valores esperados $e_{i,j}$ sejam maiores ou iguais a 5.

Seja $X_1, ..., X_n$ amostra aleatória de uma variável $X$ e $Y$ ****. As hipóetes para o teste são:

$$H_0 : X\ segue\ a\ distribuição\, Y$$

$$H_1 : X\ não\ segue\ a\ distribuição\, Y$$

Com estatística de teste:

$$\frac{(o_i - e_i)^2}{e_i} \sim \chi^2_{n-2}$$

Não há uma implementação exata da adaptação do teste qui-quadrado no R para a sua versão de aderência. No entanto, é possível implementa-la de forma autônoma. Para o caso contínuo, visando respeitar a condição da frequência esperada ser maior ou igual a 5, tempos:

$$e_i = p \cdot n \Rightarrow p\ \cdot n \geq 5 \Rightarrow p \geq \frac{5}{n}$$

Optamos por manter a frequência observada igual a 5. Logo, iremos buscar por intervalos na função densidade que compreendam o valor de probabilidade

$$p = \frac{5}{n}$$

# Teste Kolmogorov-Smirnov

O teste de Kolomogorov-Smirnov () é um teste que tem por finalidade comparar se duas amostras aleatórias seguem a mesma distribuição de probabilidade. Ele é formulado através da função de distribuição acumulada empírica, analisando a distância estre as duas funções. Uma das aplicações do teste é como teste de aderência.
Sejam $X_1, ..., X_n$ amostra aleatória com função distribuição acumulada $F$ e $Y_1, ...Y_m$ amostra aleatória com função distribuição acumulada $G$. As hipóteses utilizadas para a construção são:

$$H_0 : F = G$$

$$H_1 : F \neq G$$

E a estatística de teste é:

$$D_{n,m} = max\{\mid F_n(x) - G_m(y) \mid\}$$

No R o teste pode ser feito usando a função **ks.test**. Um exemplo simples, onde queremos testar se a amostra segue uma distribuição $Normal(40, 2)$:

```
ks.test(rnorm(400, mean=40, sd=2), 'pnorm', 40, 2)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  rnorm(400, mean = 40, sd = 2)
## D = 0.043241, p-value = 0.4431
## alternative hypothesis: two-sided
```

No teste acima a estatística $D_{400}$ obteve valor 0.047, e o p-valor 0.3234, logo, como o esperado, não rejeita-se a hipótese de normalidade.

# Teste Cramér-von Mises

O teste de Cramér-von Mises ou cretério de Cramér-von Mises é uma alternativa ao teste de Kolmogorov-Smirnov. O teste também compara a igualdade de distribuições de probabilidae usando a função de distribuição empírica. Para a versão de teste de aderência (também é possível comparar amostras) segue-se que: Seja $x_1, ..., x_n$ amostra aleatória em ordem crescente com função de distribuição empírica $F_n$, a estatística do teste será:

$$T = n\omega^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[ \frac{2i-1}{2n} - F(x_i) \right]^2$$

Sendo $\omega^2$ igual a:

$$\int_{-\infty}^{\infty} [F_n(x) - F^*(x)]^2 dF^*(x)$$

, aonde $F^*$ é a função de distribuição teórica. O p-valor pode é computado usando uma distribuição exata para essa estatística. Exemplo no R usando o pacote 'goftest':

```r
library(goftest)
cvm.test(rnorm(400), 'pnorm')
```

```
##
##  Cramer-von Mises test of goodness-of-fit
##  Null hypothesis: Normal distribution
##
## data:  rnorm(400)
## omega2 = 0.11199, p-value = 0.5295
```

O teste retorna o valor de $\omega^2$ e o p-valor.

# Comparação entre os testes

# Conclusão