

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
INSTITUTO DE CIÊNCIAS EXATAS  
DEPARTAMENTO DE ESTATÍSTICA

Estatística não paramétrica básica no *software*  
R: uma abordagem por resolução de problemas

Magda Carvalho Pires

Matheus Barros Castro

Zaba Valtuille Lieber

Thais Pacheco Menezes

Raquel Yuri da Silveira Aoki

Agosto de 2018

## Prefácio

---

Este material foi elaborado a partir das notas de aula da Profa. Magda Carvalho Pires na disciplina EST080 – Estatística não paramétrica (oferecida pelo Departamento de Estatística da UFMG ao Curso de Graduação em Estatística), sendo resultado dos projetos de Iniciação à Docência dos alunos Matheus Barros de Castro, Zaba Valtuille Lieber, Thais Pacheco Menezes e Raquel Yuri da Silveira Aoki.

O objetivo principal do texto é apresentar a solução de problemas envolvendo a aplicação dos métodos estatísticos não paramétricos mais comuns através do *software* R (<https://cran.r-project.org/>). Não se pretende, portanto, fornecer explicações teóricas sobre os métodos abordados (para tanto, indicamos os livros de Lehmann e D’Abrera [1975], Conover [1999], Siegel e Castellan [2006], Agresti [2007] e Triola [2010]). Os exemplos foram retirados desses livros, do material do Prof. Paulo Guimarães da UFPR ([www.coordest.ufpr.br/wp-content/uploads/2016/08/aluno-2016-np.pdf](http://www.coordest.ufpr.br/wp-content/uploads/2016/08/aluno-2016-np.pdf)) dos tutoriais dos pacotes utilizados (<https://cran.r-project.org/web/packages/>) e também de websites, como o R-bloggers ([www.r-bloggers.com](http://www.r-bloggers.com)) e o Portal Action ([www.portalaction.com.br](http://www.portalaction.com.br)).

Sugestões e possíveis correções podem ser enviadas para [magda@est.ufmg.br](mailto:magda@est.ufmg.br).

# Índice

---

Capítulo 1 – Testes de Aderência .....	1
1.1 Teste Binomial (binom.test) .....	1
1.2 Teste Qui-quadrado (chisq.test) .....	2
1.3 Teste de kolmogorov-Smirnov (ks.test) .....	3
1.4 Teste de Lilliefors (lillie.test).....	4
1.5 Teste Shapiro-Wilk (Ryan-Joiner) (shapiro.test) .....	5
1.6 Teste Anderson-Darling (ad.test).....	5
Capítulo 2 – Testes para Comparação de dois tratamentos (populações).....	7
2.1 Teste do sinal (SIGN.test).....	7
2.2 Teste de Postos com Sinais de Wilcoxon (wilcox.test) .....	8
2.3 Teste exato de Fisher (fisher.test) .....	9
2.4 Teste Qui-Quadrado (chisq.test).....	11
2.5 Teste da mediana (fisher.test) .....	12
2.6 Teste de Wilcoxon\ Mann-Whitney (wilcox.test) .....	13
2.7 Teste de Kolmogorov-Smirnov (ks.test).....	14
Capítulo 3 – Testes para Comparação para mais populações.....	16
3.1 Teste Qui-Quadrado (chisq.test).....	16
3.2 Teste da Mediana (chisq.test).....	17
3.3 Teste por postos de Kruskal-Wallis (1 fator)(kruskal.test).....	18
3.4 Teste de Levene (leveneTest) .....	18
3.5 Teste de Friedman (friedman.test).....	20
Capítulo 4 – Coeficientes e Testes de Concordância e correlação .....	22
4.1 Coeficiente de Correlação de Spearman (cor.test) .....	22
4.2 Coeficiente de Correlação Posto-Ordem de Kendall (T) (cor.test).....	23
4.3 Coeficiente de Kappa (kappa2) .....	24
4.4 Coeficiente de Kappa Múltiplo (kappam.fleiss) .....	25
Referências .....	27

# Capítulo 1 – Testes de Aderência

---

## 1.1 Teste Binomial (binom.test)

Utilizado em amostras com variáveis dicotômicas, testa se a proporção de sucesso observada na amostra ( $\hat{p}$ ) pertence a uma população com um determinado valor de  $p$ .

### **Exemplo 1.1**

Em uma amostra de tamanho 20 foram observados 5 sucessos. Testar se  $p > 0.2$  com nível de significância de 5%.

```
#-----  
binom.test(5, 20, p = 0.2, alternative = c("greater"), conf.level = 0.95)
```

*Parâmetros: O valor 5 representa o número de sucesso observados; 20 é o tamanho da amostra;  $p=0.2$  é a hipótese nula; "greater" para a hipótese alternativa onde testamos se  $p > 0.2$  (usaríamos "two.sided" caso  $p \neq 0.2$  e "less" caso  $p < 0.2$ );  $conf.level=0.95$  para construir um intervalo de confiança.*

*Hipóteses:  $\begin{cases} H_0: p = 0.2 \\ H_1: p > 0.2 \end{cases}$*

### **Resultado:**

```
Exact binomial test  
data: 5 and 20  
number of successes = 5, number of trials = 20, p-value = 0.3704  
alternative hypothesis: true probability of success is greater than 0.2  
95 percent confidence interval:  
 0.1040808 1.0000000  
sample estimates:  
probability of success  
 0.25
```

*Resultado comentado: No resultado do teste são apresentados em sequência o número de sucessos, o número de tentativas e o p-valor calculado. Abaixo é apresentada a hipótese nula do teste realizado, seguida pelo intervalo de confiança calculado. Por fim é apresentada a estimativa da probabilidade de sucesso calculada a partir da amostra.*

Conclusão: Não há evidências suficientes para rejeitarmos a hipótese nula ( $p\text{-valor} = 0.3704$ ).

## 1.2 Teste Qui-quadrado (chisq.test)

Utilizado para testar se a frequência observada na amostra difere significativamente da frequência esperada especificada por uma distribuição de probabilidade.

### Exemplo 1.2

*Abaixo, temos o número observado de falhas mecânicas, por hora, em uma linha de montagem, a partir de um experimento com duração de 40 horas. Um engenheiro afirma que o processo descrito, seguem uma distribuição de Poisson com média igual a 3.2. Testar com  $\alpha=0.05$ .*

```
#-----  
#Iremos construir o vetor de probabilidades esperado  
fo=c(0,6,8,11,7,4,3,1,0) #correspondentes a 0,1,2,3,4,5,6,7 e +7 falhas  
fe1 =c(0,1,2,3,4,5,6,7)  
fe=dpois(fe1, 3.2) #calculo da frequência esperada de 0 a 7 falhas  
x=1-sum(fe) #calculo da frequência esperada de 'mais de 7 falhas'  
fe=c(fe,x) #adição de 'mais de 7 falhas' ao numero esperado de falhas  
chisq.test(fo,p=fe)
```

Parâmetros: O primeiro parâmetro é um vetor com os dados observados, o segundo parâmetro ( $p$ ) é um vetor com as probabilidades esperadas (sob  $H_0$ ).

Hipóteses:  $\begin{cases} H_0: \text{As falhas seguem distribuição de Poisson}(3.2) \\ H_1: \text{As falhas não seguem distribuição de Poisson}(3.2) \end{cases}$

Resultado:

```
Chi-squared test for given probabilities  
data: fo  
X-squared = 3.1434, df = 8, p-value = 0.925
```

Resultado comentado: No resultado é apresentada a estatística de teste calculada, o número de graus de liberdade 'df' e o p-valor do teste.

Conclusão: Não há evidências para rejeitar  $H_0$  ( $p=0,925$ ), ou seja, não refutamos a ideia que os dados seguem uma distribuição de Poisson(3.2).

### 1.3 Teste de kolmogorov-Smirnov (ks.test)

Avalia se os dados amostrais se aproximam razoavelmente de uma determinada distribuição.

#### Exemplo 1.3

Testar se os dados em y seguem uma distribuição de Poisson(1.2)

```
#-----  
y=c(15,25,10,5,4,1)  
ks.test(y, "ppois", 1.2)
```

Parâmetros: O parâmetro 'ppois' representa a distribuição de Poisson com lambda igual a 1.2. Caso a distribuição de interesse fosse a Gamma, por exemplo, usaríamos 'pgamma' e seguida colocaríamos os parâmetros da Gamma separados por vírgula. O mesmo é feito para as demais distribuições.

Hipóteses:  $\begin{cases} H_0: Y \text{ segue distribuição de Poisson}(1.2) \\ H_1: Y \text{ não segue distribuição de Poisson}(1.2) \end{cases}$

#### Resultado:

```
One-sample Kolmogorov-Smirnov test  
data: y  
D = 0.8256, p-value = 5.63e-05  
alternative hypothesis: two-sided
```

Resultado comentado: O resultado apresenta em 'data' a variável onde os dados foram retirados, em 'D' a estatística de teste e o 'p-value' representa o p-valor. Temos também uma referência à hipótese alternativa utilizada.

Conclusão: Há fortes evidências para rejeitarmos a hipótese nula ( $p=0,000$ ), ou seja, os dados não são provenientes de uma distribuição Poisson de parâmetro 1.2.

Observação: O teste K-S só deve realizado quando não há observações empatadas.

## 1.4 Teste de Lilliefors (lillie.test)

Esse teste é utilizado para verificar a normalidade dos dados.

*Nota: Para a realização deste teste é necessária a instalação do pacote 'nortest'.*

### **Exemplo 1.4**

*Testar se os dados guardados em x seguem a distribuição Normal.*

```
#-----  
ifelse(!require(nortest),install.packages("nortest",  
dependencies=TRUE),1)  
  
require(nortest)  
  
x=c(1.90642, 2.10288, 1.52229, 2.61826, 1.42738, 2.22488, 1.69742, 3.15435,  
1.98492, 1.99568)  
  
lillie.test(x)
```

**Parâmetros:** *O vetor com os dados a serem testados*

**Hipóteses:**  $\begin{cases} H_0: \text{Os dados seguem distribuição Normal} \\ H_1: \text{Os dados não seguem distribuição Normal} \end{cases}$

**Resultado:**

```
Lilliefors (Kolmogorov-Smirnov) normality test  
  
data: x  
  
D = 0.1771, p-value = 0.5012
```

**Resultado comentado:** *O resultado apresenta a estatística de teste calculada em 'D' e o p-valor resultante.*

**Conclusão:** *Não há evidências para rejeitarmos  $H_0$  (valor-p = 0.5012), concluindo que dados seguem a distribuição normal.*

## 1.5 Teste Shapiro-Wilk (Ryan-Joiner) (shapiro.test)

O teste de shapiro-wilk é outro exemplo de teste de normalidade, sendo mais indicado para amostras menores ou iguais a 50.

### **Exemplo 1.5**

*Testar se os dados guardados em x seguem a distribuição Normal.*

```
#-----  
require(nortest)  
  
x=c(1.90642, 2.10288, 1.52229, 2.61826, 1.42738, 2.22488, 1.69742, 3.15435,  
1.98492, 1.99568)  
  
shapiro.test(x)
```

**Parâmetros:** *O vetor com os dados a serem testados*

**Hipóteses:**  $\begin{cases} H_0: \text{Os dados seguem distribuição Normal} \\ H_1: \text{Os dados não seguem distribuição Normal} \end{cases}$

**Resultado:**

```
Shapiro-Wilk normality test  
  
data: x  
  
W = 0.9267, p-value = 0.4162
```

**Resultado comentado:** *O resultado apresenta a estatística de teste calculada em 'W' e o p-valor resultante.*

**Conclusão:** *Não há evidências para rejeitar  $H_0$  ( $p=0,4162$ ), ou seja, os dados são provenientes de uma distribuição normal.*

## 1.6 Teste Anderson-Darling (ad.test)

Esse teste também é para testar a normalidade dos dados e é mais poderoso para amostras maiores que 50.

*Nota: Para a realização deste teste é necessária à instalação do pacote 'nortest'.*



### **Exemplo 1.6**

*Testar se os dados são normais.*

```
#-----  
ifelse(!require(nortest),install.packages("nortest",  
dependencies=TRUE),1)  
  
require(nortest)  
  
x=c(1.90642, 2.10288, 1.52229, 2.61826, 1.42738, 2.22488, 1.69742, 3.15435,  
1.98492, 1.99568)  
  
ad.test(x)
```

**Parâmetros:** *O vetor com os dados a serem testados.*

**Hipóteses:**  $\begin{cases} H_0: \text{Os dados seguem distribuição Normal} \\ H_1: \text{Os dados não seguem distribuição Normal} \end{cases}$

**Resultado:**

```
Anderson-Darling normality test  
  
data: x  
  
A = 0.3417, p-value = 0.4144
```

**Resultado comentado:** *O resultado apresenta a estatística de teste calculada em 'A' e o p-valor resultante.*

**Conclusão:** *Não há evidências para rejeitar  $H_0$  ( $p=0,4144$ ), ou seja, os dados são provenientes de uma distribuição normal.*

**Observação:** Apresentamos diferentes testes de normalidades aplicados à mesma amostra, mas a escolha por um deles deve ser baseada no tamanho amostral: o teste de Anderson-Darling é mais poderoso para amostras maiores que 50, enquanto o de Shapiro-Wilk é mais poderoso para amostras menores ou iguais a 50. O teste K-S (Lilliefors) é o menos poderoso.

## Capítulo 2 – Comparação de dois grupos

---

### 2.1 Teste do sinal (SIGN.test)

Utilizado para identificar se dois conjuntos de dados possuem a mesma medida de tendência central para experimentos pareados, testando se a mediana da diferença entre os grupos é 0. Calculamos a diferença entre as observações e atribuímos sinais de '+' ou '-' para resultados positivos e negativos respectivamente.

Nota: É necessária a instalação do pacote 'BSDA' para a realização desse teste.

#### **Exemplo 2.1**

*Deseja-se verificar se existe diferença entre os tempos de taxiamento de decolagem e os tempos de taxiamento de pouso para o Voo 21 da American Airlines de Nova York para Los Angeles. Utilize o teste do sinal com nível de significância de 5%.*

```
ifelse(!require(BSDA),install.packages("BSDA",dependencies=T),1)
require(BSDA)
decolagem<-c(13,20,12,17,35,19,22,43,49,45,13,23)
pouso<-c(13,4,6,21,29,5,27,9,12,7,36,12)
SIGN.test(decolagem, pouso, md=0, alternative="two.sided")
```

**Parâmetros:** *pouso e decolagem são os dados coletados (amostras pareadas). Md é hipótese nula sobre a mediana da diferença das observações, o valor padrão é 0. Alternative indica a hipótese alternativa a ser realizada pelo teste.*

#### **Hipóteses:**

$$\begin{cases} H_0: \text{Não existe diferença nos tempos de taxiamento de pouso e decolagem} \\ H_1: \text{Existe diferença nos tempos de taxiamento de pouso e decolagem} \end{cases}$$

Resultado:

Dependent-samples Sign-Test			
data: decolagem and pouso			
S = 8, p-value = 0.2266			
alternative hypothesis: true median difference is not equal to 0			
95 percent confidence interval:			
-3.574545 32.085455			
sample estimates:			
median of x-y			
8.5			
Achieved and Interpolated Confidence Intervals:			
	Conf.Level	L.E.pt	U.E.pt
Lower Achieved CI	0.8540	0.0000	16.0000
Interpolated CI	0.9500	-3.5745	32.0855
Upper Achieved CI	0.9614	-4.0000	34.0000

Resultado comentado: Os resultados apresentam a estatística de teste 'S' calculada, o p-valor do teste, a hipótese alternativa sob estudo, o intervalo de confiança calculado para a mediana das diferenças (com base em interpolação) e a estimativa da mediana das diferenças. Além disso o resultado também apresenta os intervalos de confiança inferior, interpolado e superior construídos com base nos dados.

Conclusão: Não há evidências para rejeitar  $H_0$  ( $p=0,2266$ ), ou seja, não há evidências de que existe diferença nos tempos de taxiamento de pouso e decolagem do Voo 21 da American Airlines.

## 2.2 Teste de Postos com Sinais de Wilcoxon (wilcox.test)

Teste utilizado para comparar se dois grupos possuem a mesma medida de tendência central. Esse teste leva em consideração a magnitude das diferenças entre os pares. Devido a este fato é mais poderoso, pois dá mais peso a diferenças maiores entre os pares.

### **Exemplo 2.2**

Os dados a seguir são das colheitas de espigas de milho (em libras por acre) de dois diferentes tipos de sementes (normais e secadas no forno) que foram usados em lotes adjacentes. Testar se existe diferença entre a colheita utilizando ambas as sementes.

```
#-----  
normal=c(1903, 1935, 1910,2496, 2108, 1961,2060, 1444, 1612, 1316, 1511)  
secada=c(2009, 1915, 2011, 2463, 2180, 1925, 2122, 1482, 1542, 1443, 1535)  
wilcox.test(normal,secada, paired=TRUE,alternative = c("two.sided"))
```

Parâmetros: 'normal' e 'secada' são vetores com os dados coletados; 'paired=TRUE' indica que os dados são pareados (o valor FALSE é utilizado para outro teste relacionado a postos); alternative = c("two.sided") diz a respeito da hipótese alternativa, indicando que ela é bilateral.

Hipóteses:  $\begin{cases} H_0: \text{Não há diferença entre os grupos} \\ H_1: \text{Há diferença entre os grupos} \end{cases}$

Resultado:

```
Wilcoxon signed rank test  
data: normal and secada  
V = 15, p-value = 0.123  
alternative hypothesis: true location shift is not equal to 0
```

Resultado comentado:

Conclusão: Não há evidências para rejeitar  $H_0$  ( $p=0.123$ ), ou seja, parece não existir diferença significativa entre as duas sementes.

## **2.3 Teste exato de Fisher (fisher.test)**

Utilizado em tabelas de contingência 2x2 para comparar 2 grupos de duas amostras independentes e pequenas.

### **Exemplo 2.3**

Verificar ao nível de 5% de significância se a proporção de divórcios amigáveis é maior na classe alta quando comparado com a classe média.

```
#-----
#Os dados sobre divórcios são alocados a dois vetores, chamados 'amigavel'
#e 'naoamigavel'. O primeiro elemento de cada vetor são o número de
#divórcios da classe alta e o segundo elemento, o número de divórcios da
#classe baixa

amigavel=c(3,2)

naoamigavel=c(2,3)

divorcio=cbind(amigavel,naoamigavel)

divorcio

fisher.test(divorcio, alternative = "greater", conf.int = TRUE, conf.level
= 0.95)
```

**Parâmetros:** *Matriz 2x2 no formato de uma tabela de contingência. Alternative indica a hipótese alternativa do teste realizado, também são válidos os valores "less" e "two.sided", conf.int indica que o intervalo de confiança deve ser construído para a razão de chances e conf.level indica o nível de confiança a ser utilizado para a construção do intervalo.*

**Hipóteses:**

$\{H_0: \text{Prop de divórcios amigáveis na classe alta} = \text{Prop de divórcios amigáveis na classe baixa}$   
 $H_1: \text{Prop de divórcios amigáveis na classe alta} > \text{Prop de divórcios amigáveis na classe baixa}$

**Resultado:**

```
Fisher's Exact Test for Count Data

data:  divorcio

p-value = 0.5

alternative hypothesis: true odds ratio is greater than 1

95 percent confidence interval:

 0.1541449      Inf

sample estimates:

odds ratio

 2.069959
```

**Resultado comentado:** *São apresentados o p-valor do teste, a hipótese alternativa em consideração, o intervalo de confiança construído baseado na hipótese alternativa e a estimativa da razão de chances com base na tabela de contingência.*

**Conclusão:** Não há evidências para rejeitar  $H_0$  ( $p=0,50$ ), ou seja, não há evidências de que o número de divórcios amigáveis da classe alta seja maior que o número de divórcios amigáveis da classe média.

## 2.4 Teste Qui-Quadrado (chisq.test)

Usado em tabelas 2x2 somente quando  $N > 20$  e se todas as frequências esperadas são maiores ou iguais a 5. Quando utilizado em matrizes com dimensões maiores que 2x2, só é usado se pelo menos 20% das células tem frequência esperada maior que 5 e nenhuma delas tem frequência esperada menor que 1.

### **Exemplo 2.4**

*Desejamos testar se a altura define se uma pessoa será um líder ou não.*

```
#-----  
# Construímos uma tabela 2x3 com o nível de liderança nas colunas (líder,  
#moderado ou não classificado) e a altura da pessoa (baixo ou alto) nas  
#linhas  
baixo=c(12,22,9)  
alto=c(32,14,6)  
lider=cbind(baixo,alto)  
chisq.test(lider)
```

**Parâmetros:** Matriz com o conjunto de dados para se testar se existe associação entre as linhas e as colunas.

**Hipóteses:**  $\begin{cases} H_0: \text{Não existe influência da altura sobre ser um líder ou não} \\ H_1: \text{A altura da pessoa influencia em ser um líder ou não} \end{cases}$

**Resultado:**

```
Pearson's Chi-squared test  
  
data: lider  
X-squared = 10.7122, df = 2, p-value = 0.004719
```

**Resultado comentado:** O resultado apresenta a estatística de teste calculada, os graus de liberdade e o p-valor do teste.

**Conclusão:** Ao nível de 5% de significância, há fortíssimas evidências para rejeitarmos a hipótese nula ( $p\text{-valor} = 0.004719$ ), concluindo que a altura é um fator que pode definir se uma pessoa será um líder ou não.

## 2.5 Teste da mediana (fisher.test)

O teste se baseia no cálculo da possibilidade de que os dois grupos provenham de populações com a mesma mediana. O nível de mensuração deve no mínimo ser ordinal. Em sua realização, através dos dados, montamos uma nova tabela 2x2 e a partir dela realizamos o teste Exato de Fisher ou Qui-quadrado.

### Exemplo 2.5

Verifique se os salários do RH( Recursos Humanos) provêm de populações com mediana diferente dos salários da CQ (controle de qualidade).

```
#-----  
rh= c(4,3,8,3,5,7,2)  
cq= c(11,10,7,6,5,8,9,10)  
a=c(rh,cq)  
median(a)  
salario= matrix(c(1,6,5,3),nrow = 2,dimnames = list(Mediana = c(">  
mediana", "<= mediana"),departamento = c("RH", "CQ")))  
salario  
fisher.test(salario)
```

Estrutura do teste: o vetor 'a' é a união dos dois vetores cujas medianas queremos comparar. Depois de calcular a mediana de 'a', o próximo passo é montar uma matriz. No comando montar a matriz, que no caso chamará 'salario', colocamos um vetor com os dados. Esses dados são obtidos da seguinte maneira: O primeiro elemento é o número de salários acima da mediana do vetor 'rh', o segundo elemento é o número de salários abaixo ou igual a mediana, também do vetor 'rh'. O terceiro elemento é o número de salários acima da mediana do vetor 'cq' e o quarto elemento é o número de salários abaixo ou igual a mediana do vetor 'cq'. O comando 'nrow=2' dividirá o vetor em dois, formando então uma matriz 2x2. O comando 'dimnames' muda o nome das linhas e colunas, nessa ordem respectivamente. Primeiro colocamos um nome que define todas as linhas e dentro de um vetor c, os nomes de cada linha. Os mesmos passos são usados para mudar o nome das colunas.

Parâmetros: Uma matriz 2x2 cujas linhas são os grupos em estudo e as colunas contém o número de observações acima e abaixo da mediana dos dados em conjunto respectivamente.

Hipóteses: 
$$\begin{cases} H_0: \text{Mediana}(RH) = \text{mediana}(CQ) \\ H_1: \text{Mediana}(RH) \neq \text{mediana}(CQ) \end{cases}$$

### Resultado:

```
Fisher's Exact Test for Count Data

data:  salario

p-value = 0.1189

alternative hypothesis: true odds ratio is not equal to 1

95 percent confidence interval:

 0.001827082 1.768053629

sample estimates:

odds ratio

 0.1189474
```

Resultado comentado: São apresentados o p-valor do teste, a hipótese alternativa em estudo, o intervalo de confiança construído para a razão de chances e a estimativa para a razão de chances calculada com base na tabela informada.

Conclusão: Não há evidências para rejeitar  $H_0$  ( $p=0,1189$ ), ou seja, não existe diferença significativa entre as medianas dos salários dos dois departamentos.

## 2.6 Teste de Wilcoxon\ Mann-Whitney (wilcox.test)

Utilizado para verificar se dois grupos pertencem à mesma população. Na prática, verifica-se se há evidências para afirmar que valores de um grupo A são superiores aos valores do grupo B.

### Exemplo 2.6

*Desejamos testar se os dois grupos pertencem a uma mesma população.*

```
a = c(6, 8, 2, 4, 4, 5)
b = c(7, 10, 4, 3, 5, 6)

wilcox.test(a,b, correct=FALSE alternative = "two.sided")
```

Parâmetros: 'a' e 'b' representam os grupos que queremos testar. No comando, o parâmetro 'correct=FALSE', indica que não queremos que seja aplicada a correção de continuidade.

Hipóteses:  $\begin{cases} H_0: \text{Os dois grupos possuem a mesma medida de tendência central} \\ H_1: \text{Os grupos estão centrados em pontos diferentes} \end{cases}$



Resultado:

```
Wilcoxon rank sum test
data: a and b
W = 14, p-value = 0.5174
alternative hypothesis: true location shift is not equal to 0
```

Resultado comentado: O teste apresenta a estatística de teste 'W' calculada e o p-valor do teste, além disso é apresentada a hipótese alternativa em estudo.

Conclusão: Não há evidências para rejeitar  $H_0$  ( $p=0,5174$ ), concluindo que os dois grupos são da mesma população.

## 2.7 Teste de Kolmogorov-Smirnov (ks.test)

Esse teste é usado para determinar se duas amostras procedem da mesma população ou se uma amostra segue uma distribuição determinada na hipótese nula. A segunda situação foi abordada na Seção 1.3.

### Exemplo 2.7

Deseja-se comparar duas soluções químicas com relação ao grau de P.H. Testar se os dados A e B tem a mesma distribuição ao nível de 5% de significância.

```
#-----
A=c(7.49, 7.35,7.54,7.48,7.48,7.37,7.51,7.50,7.52,7.46)
B=c(7.28,7.35,7.52,7.50,7.38,7.48,7.31,7.22,7.41,7.45)
ks.test(A,B,alternative="two.sided")
```

Parâmetros: A e B são os dados coletados dos dois grupos. Alternative indica a hipótese alternativa do teste, é utilizado 'greater' ou 'less' para testar se a distribuição de A se encontra acima da curva de distribuição de B ou abaixo respectivamente.

### Hipóteses:

$\{H_0: A \text{ distribuição de } A \text{ é igual à distribuição de } B$   
 $\{H_1: A \text{ distribuição de } A \text{ é diferente da distribuição de } B$

Resultado:

Two-sample Kolmogorov-Smirnov test

data: A and B

D = 0.5, p-value = 0.1641

alternative hypothesis: two-sided

Resultado comentado: O resultado apresenta a estatística de teste 'D' calculada, e o p-valor do teste, além da hipótese alternativa sob estudo.

Conclusão: Não há evidências para rejeitar  $H_0$  ( $p=0,1641$ ), concluindo que os dados seguem a mesma distribuição.

## Capítulo 3 – Comparação de três ou mais grupos

---

### 3.1 Teste Qui-Quadrado (chisq.test)

Os dados são frequências em  $r$  categorias discretas. Deseja-se testar se os dados são de uma mesma população ou seguem uma mesma distribuição.

#### **Exemplo 3.1**

*Deseja-se testar se a proporção de homens e mulheres em 3 partidos difere significativamente.*

```
#-----  
M <- as.table(rbind(c(762, 327, 468), c(484,239,477)))  
dimnames(M) <- list(Sexo=c("M","F"),  
                    Partido=c("Democrat","Independent", "Republican"))  
chisq.test(M)
```

**Parâmetros:** 'M' é a matriz com os dados.

**Hipóteses:**  $\begin{cases} H_0: A \text{ proporção nos 3 partidos é semelhante} \\ H_1: A \text{ proporção nos 3 partidos não é semelhante} \end{cases}$

**Resultado:**

```
Pearson's Chi-squared test  
  
data: M  
X-squared = 30.0701, df = 2, p-value = 2.954e-07
```

**Resultado comentado:** São apresentados a estatística teste Qui-quadrado, os graus de liberdade e o p-valor calculado.

**Conclusão:** *Há fortes evidências para rejeitarmos  $H_0$  ( $p=0,000$ ), concluindo haver alguma diferença entre as proporções de homens e mulheres nos três partidos políticos analisados em questão.*

## 3.2 Teste da Mediana (chisq.test)

Nesse teste averiguamos se os k grupos provem de populações com medianas iguais. O nível de mensuração deve ser, no mínimo, em escala ordinal.

### Exemplo 3.2

*Deseja-se verificar se há diferença significativa entre os escores de QI de sujeitos com baixa exposição ao chumbo, média exposição ao chumbo e alto exposição ao chumbo. Use o nível de significância de 0.05 para testar se as três amostras provêm de populações com medianas iguais.*

```
baixo<-c(85,90,107,85,100,97,101,64)
medio<-c(78,97,107,80,90,83)
alto<-c(93,100,97,79,97)

mediana<-median(c(baixo,medio,alto)) #mediana geral
b<-c(sum(baixo>mediana),sum(baixo<=mediana))
m<-c(sum(medio>mediana),sum(medio<=mediana))
a<-c(sum(alto>mediana),sum(alto<=mediana))
dados<-as.table(cbind(b,m,a))
chisq.test(dados)
```

Parâmetros: Para a realização desse teste, foi criada uma matriz onde as colunas ainda são os k grupos, e as linhas indicam quantos indivíduos de cada grupo estão acima e quantos estão abaixo da mediana geral. Obtida essa matriz, o teste é realizado como no Teste qui-quadrado.

Hipóteses:  $\begin{cases} H_0: \text{Os grupos têm medianas semelhantes} \\ H_1: \text{Pelo menos um grupo tem mediana diferente das demais} \end{cases}$

Resultado:

```
Pearson's Chi-squared test

data: dados

X-squared = 0.8163, df = 2, p-value = 0.6649
```

Resultado comentado: A saída é a mesma do teste qui-quadrado.

Conclusão: Não há evidências para rejeitar  $H_0$  ( $p=0,6649$ ), ou seja, não há evidências suficientes para afirmar que os escores medianos dos grupos são diferentes.

Observação: Esse exemplo para implementação do teste da Mediana é apenas ilustrativo, pois os dados não atendem o requisito do Teste Qui-Quadrado de que todas as frequências esperadas devem ser maiores que 5. Para maiores detalhes, consulte Agresti [2007].

### 3.3 Teste por postos de Kruskal-Wallis (1 fator) (kruskal.test)

Utilizado para testar se k grupos são semelhantes.

#### **Exemplo 3.3**

*Serão utilizados os mesmos dados do exemplo anterior.*

```
baixo<-c(85,90,107,85,100,97,101,64)
medio<-c(78,97,107,80,90,83)
alto<-c(93,100,97,79,97)

QI<-c(baixo,medio,alto)

grupo<-
c(rep("b",length(baixo)),rep("m",length(medio)),rep("a",length(alto)))

kruskal.test(QI~factor(grupo))
```

**Hipóteses:**  $\begin{cases} H_0: \text{Os grupos são semelhantes} \\ H_1: \text{Pelo menos um grupo difere dos demais} \end{cases}$

**Resultado:**

```
Kruskal-Wallis rank sum test

data:  QI by factor(grupo)

Kruskal-Wallis chi-squared = 0.70311, df = 2, p-value = 0.7036
```

**Resultado comentado:** O teste apresenta a estatística de teste, os graus de liberdade e o p-valor calculado.

**Conclusão:** Não há evidências para rejeitar  $H_0$  ( $p=0,7036$ ), ou seja, não há evidências de que os grupos apresentam diferentes escores de QI.

### 3.4 Teste de Levene (leveneTest)

Para a realização de muitos testes, é necessária a suposição de que as amostras tenham a mesma variância. O Teste de Levene tem a função de testar se essas suposições são válidas, analisando se todos os grupos têm a mesma variância.

*Nota: Para a realização deste teste é necessária à instalação do pacote 'car'.*

### **Exemplo 3.4**

*Utilizando os mesmos dados dos exemplos anteriores, testaremos se os grupos têm a mesma variância.*

```
#-----  
Install.packages('car',dependencies = T)  
require(car)  
baixo<-c(85,90,107,85,100,97,101,64)  
medio<-c(78,97,107,80,90,83)  
alto<-c(93,100,97,79,97)  
  
QI<-c(baixo,medio,alto)  
  
grupo<-  
c(rep("b",length(baixo)),rep("m",length(medio)),rep("a",length(alto)))  
  
leveneTest(QI,grupo)
```

*Parâmetros: Os parâmetros do teste são dois vetores, tais como os construídos no teste de Kruskal Wallis.*

*Hipóteses:  $\left\{ \begin{array}{l} H_0: \text{Os grupos possuem a mesma variância} \\ H_1: \text{Pelo menos 1 grupo possui variância diferente} \end{array} \right.$*

*Resultado:*

```
Levene's Test for Homogeneity of Variance (center = median)  
  
      Df F value Pr(>F)  
group  2  0.7152 0.5041  
  
      16
```

*Resultado comentado: São apresentados os graus de liberdade dos grupos e dos resíduos, a estatística de teste F e o p-valor calculado.*

*Conclusão: Não há evidências para rejeitar  $H_0$  ( $p=0,5041$ ), ou seja, não há evidências de que exista diferença entre as variâncias.*

### 3.5 Teste de Friedman (friedman.test)

O teste de Friedman é uma alternativa não paramétrica para o teste anova, quando existem blocos completamente aleatorizados.

#### Exemplo 3.5

Em um estudo de hipnose, as emoções de medo, felicidade, depressão e calma foram estimuladas (em ordem aleatória) em oito pacientes durante a hipnose. A tabela a seguir apresenta as medidas resultantes do potencial elétrico em minivolts da pele dos pacientes. Cada paciente é considerado um bloco. Teste se existe efeito da emoção induzida nos pacientes no potencial elétrico medido.

	Medo	Felicidade	Depressão	Calma
Paciente 1	23,1	22,7	22,5	22,6
Paciente 2	57,6	53,2	53,7	53,1
Paciente 3	10,5	9,7	10,8	8,3
Paciente 4	23,6	19,6	21,1	21,6
Paciente 5	11,9	13,8	13,7	13,3
Paciente 6	54,6	47,1	39,2	37
Paciente 7	21	13,6	13,7	14,8
Paciente 8	20,3	23,6	16,3	14,8

```
#-----  
emocao=rep(c('medo','felicidade','depressão','calma'),8)  
paciente=rep(c(1:8),each=4)  
resposta=c(23.1,22.7,22.5,22.6,57.6,53.2,53.7,53.1,10.5,9.7,10.8,8.3,23.6  
,19.6,21.1,21.6,11.9,13.8,13.7,13.3,54.6,47.1,39.2,37.0,21.0,13.6,13.7,14  
.8,20.3,23.6,16.3,14.8)  
friedman.test(resposta, groups=emocao, blocks=paciente)
```

Parâmetros: Os parâmetros são três vetores, o primeiro com as respostas, o segundo com os grupos em estudo e o terceiro com os blocos do experimento.

Hipóteses:

$\{H_0: Os\ grupos\ se\ comportam\ de\ forma\ similar$   
 $\{H_1: Pelo\ menos\ um\ dos\ grupos\ difere\ dos\ demais$

Resultado:

```
Friedman rank sum test  
data: resposta, emocao and paciente  
Friedman chi-squared = 6.45, df = 3, p-value = 0.09166
```

Resultado comentado: O resultado apresenta a estatística de teste, os graus de liberdade e o p-valor calculado.

Conclusão: Não há evidências para rejeitar  $H_0$  ( $p=0,09166$ ), ou seja, não há evidências de que o potencial elétrico cutâneo é diferente entre todas as emoções estudadas.



## Capítulo 4 – Coeficientes e Testes de Concordância e correlação

---

Mede a correlação linear entre duas variáveis. Suponha que se deseja verificar uma associação ou correlação entre duas variáveis A e B. Podem existir diversas explicações do porque elas variam conjuntamente, como mudanças em A causam mudanças em B, ou mudanças em B causam mudanças em A, ou mudanças em outras variáveis causam mudanças tanto em A como em B, ou a relação observada é somente uma coincidência.

### 4.1 Coeficiente de Correlação de Spearman (cor.test)

A escala de mensuração das duas variáveis deve ser pelo menos ordinal. Baseia-se na atribuição de postos às observações X e Y e possui um cálculo semelhante ao de Pearson.

#### **Exemplo 4.1**

*Deseja-se calcular a relação entre a diversidade de gafanhotos e o numero de anos após a aplicação de um pesticida.*

```
anos= c(0, 1, 3, 5, 9, 12, 13, 15, 21, 25)
gafanhotos=c(0.00, 0.19, 0.15, 1.49, 1.10, 1.12, 1.61, 1.42, 1.48, 1.92)
cor.test(anos,gafanhotos, method="spearman",alternative="two.sided")
```

*Parâmetros: Para o calculo da correlação, é necessário entrar com os dois vetores que se deseja comparar e depois definir em 'method' o tipo de correlação, no caso, 'sperman', Alternative indica a hipótese alternativa a ser testada. Outros valores possíveis são "greater" ou "less".*

#### Hipóteses:

$\{H_0: \text{As variáveis não são correlacionadas } (\rho_s = 0)$   
 $\{H_1: \text{As variáveis são correlacionadas } (\rho_s \neq 0)$

#### Resultado:

```
Spearman's rank correlation rho
data: anos and gafanhotos
S = 32, p-value = 0.008236
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.8060606
```

Resultado Comentado: O resultado apresenta o p-valor do teste realizado, a hipótese alternativa sob estudo e a estimativa da correlação de spearman.

Conclusão: O valor da estimativa do coeficiente de Spearman é 0,806. Há fortes evidências de que a correlação é significativa. Podemos inferir que há uma correlação positiva e significativa entre diversidade de gafanhotos e o tempo após aplicação de pesticidas.

## 4.2 Coeficiente de Correlação Posto-Ordem de Kendall(T) (cor.test)

A escala de mensuração das duas variáveis deve ser pelo menos ordinal. Assim como o de Spearman, baseia-se na atribuição de postos às observações de X e Y para medir a correlação.

### Exemplo 4.2

Será utilizado o mesmo conjunto de dados anterior, em que deseja-se calcular a relação entre a diversidade de gafanhotos e o numero de anos após a aplicação de um pesticida.

```
anos= c(0, 1, 3, 5, 9, 12, 13, 15, 21, 25)
gafanhotos=c(0.00, 0.19, 0.15, 1.49, 1.10, 1.12, 1.61, 1.42, 1.48, 1.92)
cor.test(anos,gafanhotos, method="kendall",alternative="two.sided")
```

Parâmetros: Os dados são colocados de maneira análoga ao coeficiente de correlação de 'sperman', mudando apenas o tipo de método que deseja-se utilizar, no caso 'kendall'.

Hipóteses:  $\begin{cases} H_0: \text{As variáveis não são correlacionadas } (\tau = 0) \\ H_1: \text{As variáveis são correlacionadas } (\tau \neq 0) \end{cases}$

Resultado:

```
Kendall's rank correlation tau
data: anos and gafanhotos
T = 38, p-value = 0.004687
alternative hypothesis: true tau is not equal to 0
sample estimates:
tau
0.6888889
```

Resultado comentado: O teste apresenta saída similar ao comando anterior.

Conclusão: O valor da estimativa do coeficiente de kendall é 0,688. Há fortes evidências de que a correlação é significativa. Podemos inferir que há uma correlação positiva e significativa entre diversidade de gafanhotos e o tempo após aplicação de pesticidas.

### 4.3 Coeficiente de Kappa (kappa2)

Baseada no número de respostas concordantes, ou seja, o número de casos cujo resultado é o mesmo entre os juizes. Nesse caso, 2 avaliadores classificam  $n$  objetos em  $m$  categorias e o coeficiente mede o grau de concordância do que seria esperado tão somente pelo acaso.

#### Exemplo 4.3

Uma amostra de 30 pacientes foi avaliada por dois psicólogos que classificaram os pacientes em Psicótico, Neurótico ou Orgânico. Encontre o coeficiente de kappa.

*Nota: Para a realização deste teste é necessária a instalação do pacote 'irr'.*

```
install.packages('irr',dependencies=T)

require('irr')

psicologoA =
c('neurotico','psicótico','Orgânico','Orgânico','Orgânico','psicótico','O
rgânico','neurotico','psicótico','psicótico')

psicologoB =
c('neurotico','psicótico','neurotico','Orgânico','Orgânico','psicótico','
Orgânico','Orgânico','neurotico','psicótico')

tabela = cbind(psicologoA,psicologoB)

kappa2(tabela)
```

Parâmetros: Os parâmetros são uma tabela  $2 \times n$  com as conclusões de cada avaliador nas colunas e os indivíduos avaliados nas linhas.

Hipóteses:  $\begin{cases} H_0: \text{Os avaliadores não concordam entre si } (k = 0) \\ H_1: \text{Os avaliadores concordam entre si } (k > 0) \end{cases}$

### Resultado:

Cohen's Kappa for 2 Raters (Weights: unweighted)

Subjects = 10

Raters = 2

Kappa = 0.545

z = 2.46

p-value = 0.0138

Resultado comentado: O resultado apresenta o número de indivíduos classificados, o número de avaliadores, o coeficiente kappa calculado e a estatística de teste utilizada para a verificação se o coeficiente kappa é significativo.

Conclusão: O coeficiente de Kappa obtido foi de 0.545, indicando que existe alguma concordância entre as classificações dos psicólogos. Há evidências de que existe essa concordância entre os avaliadores é significativa ( $p=0,0138$ ).

#### 4.4 Coeficiente de Kappa Múltiplo (kappam.fleiss)

Utilizado para descrever a intensidade da concordância entre juízes ou entre métodos de classificação quando existem mais de dois avaliadores.

#### Exemplo 4.4

Suponha agora que uma amostra de 10 pacientes foi classificada por três psicólogos em psicótico, neurótico ou orgânico. Encontre o coeficiente de Kappa múltiplo.

```
install.packages('irr',dependencies=T)

require('irr')

psicologoA =
c('neurotico','psicótico','Orgânico','Orgânico','Orgânico','psicótico','O
rgânico','neurotico','psicótico','psicótico')

psicologoB =
c('neurotico','psicótico','neurotico','Orgânico','Orgânico','psicótico','
Orgânico','Orgânico','neurotico','psicótico')

psicologoC =
c('neurotico','neurotico','neurotico','psicótico','Orgânico','Orgânico','
Orgânico','neurotico','neurotico','psicótico')

tabela = cbind(psicologoA,psicologoB,psicologoC)

kappam.fleiss(tabela)
```

Parâmetros: O parâmetro é uma matrix  $n \times m$ , cujas linhas são os indivíduos classificados e as colunas são os avaliadores.

Hipóteses:  $\begin{cases} H_0: \text{Os avaliadores não concordam entre si } (K = 0) \\ H_1: \text{Os avaliadores concordam entre si } (K > 0) \end{cases}$

Resultado:

Fleiss' Kappa for m Raters

Subjects = 10

Raters = 3

Kappa = 0.398

z = 3.08

p-value = 0.00209

Resultado comentado: O resultado apresenta os mesmos valores do teste de kappa para 2 avaliadores.

Conclusão: Há fortes evidências para rejeitarmos  $H_0$  e concluirmos que o coeficiente de Kappa é significativamente diferente de 0. O coeficiente de Kappa obtido foi de 0.398.

## Referências

---

Conover, W. J. Practical nonparametric statistics. 2a. ed. New York: John Wiley & Sons, 1999.

Lehmann, E.L.; D'Abrera, H.J.M. Nonparametrics: Statistical Methods Based on Ranks. Holden-Daym, California, 1975. p. 264

Siegel, S., Castellan, Jr., N. J., Estatística não-paramétrica para ciências do comportamento. São Paulo: Bookman (Artmed), 2006.

Agresti, A – An Introduction to Categorical Data Analysis (Wiley Series in Probability and Statistics) – 2ª edição. New York, USA: Wiley, 2007.

Triola, M. F. *Introdução à estatística*. Rio de Janeiro: Livros Técnicos e Científicos, 2013.