

Fundação Universidade Estadual de Maringá - UEM
Centro de Ciências Exatas
Departamento de Estatística - DES

Métodos Não-Paramétricos
Curso de Estatística

Prof. Dr. Robson M. Rossi

Sumário

1	Conceitos em Métodos Não-Paramétricos	1
1.1	Razões e restrições para o uso	1
1.2	Conceitos e definições	2
1.2.1	Teste de hipóteses	2
1.2.2	Teste equivalentes	3
1.2.3	Estatísticas de ordem	3
1.2.4	Distribuição nula	4
1.2.5	Distribuição livre	4
1.2.6	Escalas de Medidas	4
2	Testes aplicáveis a uma amostra	6
2.1	Normalidade dos dados	6
2.2	Teste Qui-quadrado para aderência	7
2.2.1	Estatística teste de Karl Pearson	8
2.2.2	Estimativas de máxima verossimilhança	9
2.3	Teste Kolmogorov-Smirnov	15
2.4	Teste de Shapiro-Wilk para normalidade	19
2.4.1	Aproximações para o teste Shapiro-Wilk	20
2.5	Testes gráficos	22
2.5.1	Q-Q plots	22
2.5.2	P-P plots	22
3	Testes aplicáveis a duas amostras - Dados pareados	27
3.1	Teste dos Sinais	27
3.2	Teste de McNemar	30
3.3	Teste Kappa - Concordância ou de replicabilidade	35
3.4	Teste de Wilcoxon	39
4	Testes aplicáveis a duas amostras - Dados independentes	46
4.1	Teste Qui-quadrado	46
4.1.1	Tabelas de contingência	46
4.1.2	Teste de independência e de homogeneidade	46
4.1.3	Teste Qui-Quadrado em Tabelas 2x2	53
4.2	Teste exato de Fisher	54
4.3	Teste da Mediana	61
4.4	Teste de Mann-Whitney	62
4.4.1	Pequenas amostras ($n_i \leq 20$)	63
4.4.2	Grandes amostras ($n_i > 20$)	64
4.5	Teste de Kolmogorov-Smirnov para duas amostras independentes	67

4.5.1	Pequenas amostras ($n_i \leq 40$)	67
4.5.2	Grandes amostras ($n_i > 40$)	70
5	Testes aplicáveis a k amostras - Dados pareados	76
5.1	Teste de Cochran	76
5.2	Teste de Friedman	81
5.2.1	Grandes amostras ($n > 4$)	81
5.2.2	Pequenas amostras ($n \leq 4$)	83
5.2.3	Comparações múltiplas	83
6	Testes aplicáveis a k amostras - Dados independentes	89
6.1	Teste de Kruskal-Wallis para k amostras	89
6.1.1	Grandes amostras ($n_i > 6$)	89
6.1.2	Pequenas amostras ($n_i \leq 6$)	90
6.1.3	Comparações múltiplas	95
7	Análise de Correlação	102
7.1	Teste de Spearman	102
7.1.1	Grandes amostras ($n \geq 10$)	104
7.1.2	Pequenas amostras ($n < 10$)	104
7.2	Teste de Kendall	108
7.2.1	Pequenas amostras ($n \leq 10$)	108
7.2.2	Grandes amostras ($n > 10$)	110
	Referências Bibliográficas	113
	Formulários	115
	Tabelas	118

Capítulo 1

Conceitos em Métodos Não-Paramétricos

1.1 Razões e restrições para o uso

O objetivo deste capítulo é ilustrar o papel que conceitos e técnicas estatísticas têm na formação do estatístico, na sua prática profissional e de modo especial no avanço do conhecimento nesta área específica.

Este objetivo será alcançado principalmente através da apresentação e discussão dos resultados de estudos relevantes. Através destes exemplos, argumenta-se que conceitos estatísticos têm importante papel a desempenhar, tanto no estudo da saúde de populações, como nos tratamentos individuais.

Na prática experimental é comum o surgimento de situações em que não se tem o conhecimento da distribuição dos dados, ou mesmo que o tenha, os mesmos podem não ter a distribuição normal. A alta dispersão ou até mesmo assimetria, ou ambos, fazem com que o trato com os dados seja dificultado.

Muitos casos destes existem técnicas de transformações nos dados que produzem novos dados cuja distribuição se aproxime da normal. Caso isto não seja possível as técnicas não-paramétricas são uma alternativa para análise estatística.

Um teste não-paramétrico é aquele cujo modelo não especifica condições sobre os parâmetros da população da qual a amostra foi obtida.

Principais razões e restrições para o uso:

- i) São menos exigentes, restritivos, do que os testes paramétricos, pois dispensam a normalidade dos dados;
- ii) Em geral, as probabilidades das afirmativas são probabilidades exatas, salvo que se usam algumas aproximações para grandes amostras;
- iii) Independem da forma da população da qual a amostra foi obtida;
- iv) Os cálculos são relativamente fáceis;
- v) Alguns testes não-paramétricos permitem trabalhar com dados de diferentes populações, ao contrário dos paramétricos;
- vi) São úteis nos casos em que é difícil estabelecer uma escala de valores quantitativos para os dados, por exemplo, *escalas ordinais quantitativas* (de ansiedade - psiquiatria) ou *escalas ordinais qualitativas* (ruim-bom-ótimo);

- vii) São mais eficientes que os paramétricos quando não há pressuposição de normalidade - Os testes paramétricos são mais poderosos se os pressupostos são verificados;
- viii) Em geral não levam a magnitude dos dados - Há perda de informação, por exemplo, valores: 2, 3.5 e 2 quando escoreizados ficam: 2, 3 e 1.

1.2 Conceitos e definições

Os testes não-paramétricos são boas opções para situações em que ocorrem violações dos pressupostos básicos necessários para a aplicação de um teste paramétrico. Por exemplo, para testar a diferença de dois ou mais grupos quando a distribuição subjacente é assimétrica ou dos dados foram coletados em uma escala ordinal.

Às vezes, é preciso comparar duas populações. Por exemplo, imagine que um pesquisador obteve, para um grande número de crianças, a idade em que cada uma delas começou a falar. Para verificar se meninos e meninas aprendem a falar na mesma idade, o pesquisador terá que comparar os dados dos dois sexos.

Outras vezes, é preciso comparar condições experimentais. Por exemplo, para saber se um tratamento tem efeito positivo, organizam-se dois grupos de unidades: um grupo recebe o tratamento em teste (*é o grupo tratado*), enquanto o outro não recebe o tratamento (*é o grupo controle*).

O efeito do tratamento é dado pela comparação dos dois grupos.

Frequentemente, para comparação de dois tratamentos observam-se características ou mede-se o valor da variável resposta de interesse que as características importantes dos indivíduos que integram o mesmo par sejam tão semelhantes quanto possível. O tratamento é administrado a um dos elementos do par e o outro é mantido como controle. A vantagem do procedimento é clara. Os indivíduos no par são semelhantes, exceto no que se refere ao tratamento recebido. Em algumas situações o par é constituído do mesmo indivíduo em duas ocasiões diferentes (amostras dependentes).

A suposição de Normalidade é necessária em alguns testes estatísticos. Nesta seção apresentaremos alguns testes de normalidade assim como algumas técnicas gráficas, afim de que possamos verificar a normalidade dos dados (em caso afirmativo, optamos por testes paramétricos).

1.2.1 Teste de hipóteses

A parte final de uma análise estatística, o teste de hipótese (T.H.), se dá pela tomada de decisão relativa a um problema de interesse, digamos, verificar se há associação entre um fator de risco (tabagismo) e uma doença (câncer) em experimentação médica.

As etapas para um T.H. são:

- (1) Enunciar as **Hipóteses**: Por exemplo:

Tabela 1.1: *Associação entre Doença (Câncer) e Fator de risco (Tabagismo).*

Câncer	Tabagismo		Total
	Sim	Não	
Casos	80 (84%)	5 (14%)	85
Controles	15	30	45
Total	95 (100%)	35 (100%)	130

H_0 : Não há evidências de associação entre Fator de Risco (Tabagismo) e Doença (Câncer pulmonar);

H_a : Há evidências de associação entre Fator de Risco e Doença;

(2) Fixar o *nível de significância* (α) do teste:

α : Erro tipo I: Rejeitar H_0 quando é verdadeira;

α : 1%, 5%, 10%, ...

(3) **Escolha da Estatística teste adequada:** χ^2 ; Fisher; McNemar etc. Dependerá do tipo de situação.

(4) **Obtenção dos valores críticos:** Valores tabelados - $\chi^2_{\alpha\%}$;

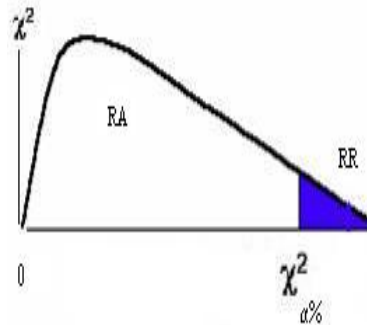


Figura 1.1: *Distribuição Qui-quadrado.*

(5) **Obtenção da Estatística calculada**, baseada nos modelos: χ^2_c (ver cap. IX);

(6) **Conclusão e tomada de decisão**

- Se $\chi^2_c > \chi^2_{\alpha\%}$, isto é, se $\chi^2_c \in RR$ (Região de Rejeição), então rejeitamos H_0 , logo diremos que ao nível $\alpha\%$ de significância, existem evidências de que a doença está associada ao fator de risco;
- Se $\chi^2_c \leq \chi^2_{\alpha\%}$, isto é, se $\chi^2_c \in RA$ (Região de Aceitação), não rejeitamos H_0 , logo não há evidências de associação entre fator de risco e doença.

1.2.2 Teste equivalentes

Dois testes são equivalentes ($T_1 \equiv T_2$) quando têm o mesmo poder, isto é, um rejeita H_0 quando o outro rejeita e vice-versa. Exemplo: Wilcoxon e sua extensão, Friedman.

1.2.3 Estatísticas de ordem

Sejam X_1, X_2, \dots, X_n observações, constituindo uma amostra. Os valores $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ obtidos das observações, quando organizadas em ordem crescente, constituem as estatísticas de ordem. São denominadas "Postos" ou "Ranks" quando os valores assumem suas posições.

Exemplo 1.1. Seja $X : \{23, 15, 5, 41, 10\}$. Em ordem tem-se: $X^* : \{5, 10, 15, 23, 41\}$ cujos ranks são: $X^R : \{1, 2, 3, 4, 5\}$;

Exemplo 1.2. No caso de empate, toma-se a média dos postos, digamos que: $X^* : \{5, 10, 10, 15, 23, 23\}$, tem-se: $X^R : \{1, 2.5, 2.5, 4, 5.5, 5.5\}$. **Obs.:** No programa R utilize o comando **rank(x)**.

1.2.4 Distribuição nula

É a distribuição de uma estatística quando H_0 é verdadeira. Sejam, por exemplo, duas populações P_1 e P_2 representadas, respectivamente, pelas amostras X_1, X_2, \dots, X_n e Y_1, Y_2, \dots, Y_n . Se considerarmos $P_1 \equiv P_2$ (equivalentes), é de se esperar em uma classificação conjunta das observações X_i e Y_i , que elas se distribuam casualmente.

1.2.5 Distribuição livre

Uma estatística W tem distribuição livre se a sua distribuição não depende da população básica.

Exemplo 1.3. Se $X \sim D_1$ e $Y \sim D_2$ conhecidas ou não, então $W = X + Y \sim D_0$, isto é, W é de livre distribuição.

Uma de suas características essenciais é a falta de conhecimento do pesquisador a respeito da forma de distribuição da população da qual foi selecionada a amostra. Seus pressupostos são, de uma maneira geral, poucos, limitando-se à independência das observações e à continuidade básica da variável estudada. *Os testes de livre distribuição podem ser aplicados quando a variável estudada se situa a nível nominal ou ordinal.*

Exemplo 1.4. Exemplo de livre distribuição.

Sejam $X_i \sim D_1$ e $Y_i \sim D_2$ com $i = 1, 2$.

Assim teremos os seguintes arranjos possíveis na classificação conjunta das quatro observações:

Arranjo	Ordem (O_i)				$W = \sum O_i^Y$
	1	2	3	4	
1	X	X	Y	Y	7
2	X	Y	X	Y	6
3	X	Y	Y	X	5
4	Y	X	X	Y	5
5	Y	X	Y	X	4
6	Y	Y	X	X	3

Todos igualmente prováveis ($p = \frac{1}{6}$). Seja $W = \sum O_i^Y$ (soma das ordens de Y em cada arranjo), logo a distribuição de W será:

$$\begin{array}{c|ccccc} W & 3 & 4 & 5 & 6 & 7 \\ \hline p(W) & \frac{1}{6} & \frac{1}{6} & \frac{2}{6} & \frac{1}{6} & \frac{1}{6} \end{array} \Rightarrow \begin{array}{l} \sum p_i = 1 \\ 0 \leq p_i \leq 1 \end{array} \Rightarrow W \text{ é f.d.p.}$$

W é de distribuição livre, pois independe da distribuição básica das observações X e Y .

1.2.6 Escalas de Medidas

Nos teste não-paramétricos são empregados as seguintes escalas:

1. **Nominal:** Utiliza números apenas como meio de distinguir elementos ou suas propriedades em diferentes classes ou categorias:

$$Ex_1 : \left\{ \begin{array}{l} \text{sim: } 1 \\ \text{não: } 0 \end{array} \right. \quad \text{e} \quad Ex_2 : \left\{ \begin{array}{l} A : 1 \\ B : 2 \\ C : 3 \end{array} \right. .$$

2. **Ordinal:** Classificam os elementos, numa ordem crescente ou decrescente:

$$Ex_1 : \begin{cases} \text{Ruim} : 1 \\ \text{Bom} : 2 \\ \text{Ótimo} : 3 \end{cases} .$$

Observação: Em geral as escalas de Likert são usadas com essa classificação ordinal (*Rensis Likert*). Tais escalas de resposta psicométrica são usadas habitualmente em questionário e/ou em pesquisas de opinião. Ao responderem a um questionário baseado nesta escala, os perguntados especificam seu nível de concordância com uma afirmação.

Exemplo: Questão - Assinale sua opinião sobre o aborto:

- 1 () Não concordo totalmente
- 2 () Não concordo parcialmente
- 3 () Indiferente
- 4 () Concordo parcialmente
- 5 () Concordo totalmente

3. **Intervalo de medida:** envolve uma unidade de medida e contém um ponto convencional de referência, “o zero”. Ex. medidas de temperaturas.
4. **Relação de medidas:** Leva em conta a ordem, o intervalo e a relação entre duas medidas. Se dizemos que “esta quantidade é três vezes aquela”, estamos utilizando uma escala de relação de medidas. Seu emprego se verifica em medidas de: peso, altura, imc, rendimentos etc.

Exercício 1.1. Considerando que uma variável X é tal que segue a distribuição Binomial, em que X é o número de sucessos em duas observações, isto é, $n = 2$. Considere que:

$$a_i : \begin{cases} 0, & \text{se a tentativa é um fracasso;} \\ 1, & \text{se a tentativa é um sucesso, } P(a=1)=1/4 \end{cases}$$

então para os 2^2 possíveis resultados de $(a_1; a_2)$, em que $x = a_1 + a_2$, obtenha a distribuição nula de x . **R.** $P(x=0)=1/36$, $P(x=1)=1/48$ e $P(x=2)=1/16$.

Capítulo 2

Testes aplicáveis a uma amostra

2.1 Normalidade dos dados

Na prática não podemos fazer suposições arbitrárias a respeito da distribuição de probabilidade dos dados, logo usamos de técnicas para fazer uma verificação da adequação do modelo suposto.

Suponha que X_1, X_2, \dots, X_n seja uma a.a. selecionada de uma f.d.p. desconhecida, e que se deseje testar a hipótese:

H_0 : Os dados seguem a distribuição Normal.

A verificação pode ser baseada nas características do modelo, em gráficos especiais e em testes de aderência.

Para a distribuição Normal, em especial, sabemos que: $\mu \pm \sigma = 68,26\%$; $\mu \pm 2\sigma = 95,45\%$ e $\mu \pm 3\sigma = 99,73\%$.

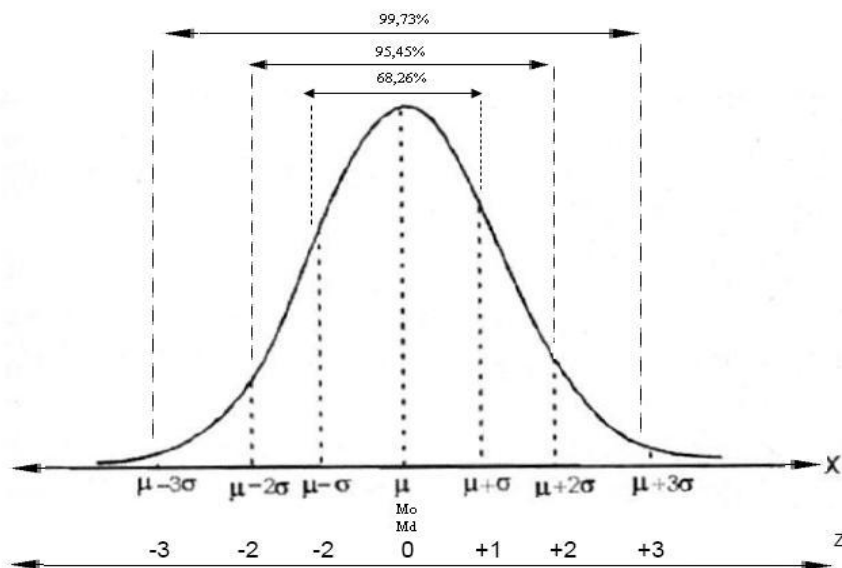


Figura 2.1: *Distribuição Normal.*

Grandes discrepâncias entre as frequências relativas observadas e as probabilidades teóricas acima citadas, indicam uma possível falta de adequação ao modelo gaussiano:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}, \quad -\infty < x < +\infty$$

onde μ : média e σ : desvio-padrão (Anexo - Tabela Função Distribuição da Normal Reduzida).

Um dos testes de aderência a seguir, podem ser utilizados:

Testes de aderência

- 1. *Qui-quadrado* (χ^2);
- 2. *Kolmogorov-Smirnov* ($K-S$) [μ e σ supostamente conhecidos];
- 3. *Lilliefors* é uma variação do $K-S$ [μ e σ estimados pelos dados];
- 4. *Shapiro Wilks* (W) [μ e σ estimados pelos dados].

Este último, tem sido muito utilizado para dados amostrais devido a sua capacidade de adaptação a uma variada gama de problemas sobre avaliação de normalidade de acordo com o teste de hipótese:

$$\begin{cases} H_0 : & \text{Os dados são normalmente distribuídos;} \\ H_a : & \text{Os dados não são normalmente distribuídos.} \end{cases}$$

Outros critérios:

- Técnicas Gráficas
- 1. *Normal-plot*;
- 2. *Q-Q-plot*;
- 3. *P-P-plot*.

2.2 Teste Qui-quadrado para aderência

Suponha uma grande população que consiste de itens de k diferentes tipos, e seja p_i a probabilidade de que um item selecionado aleatoriamente seja do tipo i ($i = 1, 2, \dots, k$).

Assuma que $0 < p_i < 1$ e $\sum_{i=1}^k p_i = 1$. Seja $p_1^0, p_2^0, \dots, p_k^0$ números específicos tais que $0 < p_i^0 < 1$ e $\sum_{i=1}^k p_i^0 = 1$ e suponha as seguintes hipóteses a serem testadas:

$$\begin{cases} H_0 : & p_i = p_i^0 \text{ para } i = 1, 2, \dots, k; \\ H_a : & p_i \neq p_i^0 \text{ para ao menos um valor de } i. \end{cases}$$

Assuma que uma a.a. de tamanho n está para ser selecionada desta população, isto é, n observações independentes estão para ser selecionadas, e existe uma probabilidade p_i que qualquer observação particular será do tipo i .

Para $i = 1, 2, \dots, k$ deveremos denotar por n_i o número de observações na amostra aleatória que são do tipo i .

Portanto n_1, n_2, \dots, n_k são inteiros não negativos tal que $\sum_{i=1}^k n_i = n$. Quando a hipótese nula é verdadeira, o **número esperado** de observações do tipo i é np_i^0 .

A diferença entre o número real de observações n_i e o número esperado, tenderá a ser menor quando H_0 é verdadeira. Em outras palavras, H_0 será rejeitada quando as magnitudes das diferenças ($n_i - np_i^0$) forem relativamente grandes.

2.2.1 Estatística teste de Karl Pearson

O matemático inglês, Karl Pearson (1857-1936) demonstrou que se H_0 for verdadeira, para $n \rightarrow \infty$, os graus de liberdade de Q^2 , dado por:

$$Q^2 = \sum_{i=1}^k \frac{(n_i - np_i^0)^2}{np_i^0}, \quad (2.1)$$

convergem para os graus de liberdade de uma distribuição χ^2 com $(k - 1)$ graus de liberdade, isto é:

$$Q^2 \sim \chi_{k-1}^2 \text{ (Anexo - Tabela de Valores percentuais da distribuição } \chi^2 \text{)}.$$

Este teste é chamado de Teste χ^2 (Qui-quadrado) de Aderência.

Quando cada valor esperado np_i^0 não for muito pequeno, a distribuição χ^2 será uma boa aproximação para a distribuição real de Q^2 . Especificamente, a aproximação será muito boa se $np_i^0 \geq 5$ para $i = 1, 2, \dots, k$.

Observações:

- Se $X = \chi_k^2$ tal que

$$f(x) = \frac{1}{\Gamma(\frac{1}{2})} \left(\frac{x}{2}\right)^{\frac{k}{2}-1} e^{-\frac{x}{2}}, \quad x > 0,$$

então

$$\chi_k^2 = \sum_{i=1}^k z_i^2, \text{ com } z_i \sim N(0, 1);$$

- $\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$;

Exemplo 2.1. *Teste de hipótese sobre uma proporção.*

Suponha que a proporção p de itens defeituosos em uma grande população de itens manufaturados é desconhecida e que as hipóteses a serem testadas são:

$$\begin{cases} H_0 : p = 0,10 \\ H_a : p \neq 0,10 \end{cases}$$

Suponha também que uma amostra aleatória de 100 itens, 16 são defeituosos, isto é, $n_1 = 16$ e $n_2 = 100 - 16 = 84$. Como só existem dois tipos de itens, a saber, defeituosos e não-defeituosos, então $k = 2$.

O valor do número esperado de defeituosos:

$$np_1^0 = 100(0,10) = 10 \text{ (se } H_0 \text{ for verdadeira)}$$

e

$$nq_1^0 = 90 \text{ (número de peças boas).}$$

Como $Q^2 \sim \chi_{k-1}^2$, então:

$$Q_c^2 = \sum_{i=1}^k \frac{(n_i - np_i^0)^2}{np_i^0} = \frac{(16 - 10)^2}{10} + \frac{(84 - 90)^2}{90} = 3,6 + 0,4 = 4.$$

$$g.l. : k - 1 = 2 - 1 = 1 \rightarrow \chi_{5\%,(1)}^2 = 3,84.$$

Como $Q_c^2 > \chi_{5\%,(1)}^2$, rejeita-se a hipótese nula.

Conclusão: Ao nível de 5% de significância existem evidências de que $p \neq 10\%$.

2.2.2 Estimativas de máxima verossimilhança

A função de verossimilhança $L(\theta)$ pode ser utilizada para obter o estimador de máxima verossimilhança quando se deseja verificar a aderência dos dados a um modelo proposto. Neste caso, o procedimento é conhecido como **teste de hipóteses compostas**. Para os valores observados $n_1, n_2, n_3, \dots, n_k$, $L(\theta)$ será dada por:

$$L(\theta) = \prod_{i=1}^k [\pi_i(\theta)]^{n_i}$$

e portanto

$$\log[L(\theta)] = l(\theta) = \sum_{i=1}^k n_i \log \pi_i(\theta),$$

assim $\hat{\theta}$, o e.m.v. (estimador de máxima verossimilhança) de θ , será o valor para o qual $L(\theta)$ é máxima.

Exemplo 2.2. Estimador de máxima verossimilhança para $k=3$, considerando uma Binomial.

Quando $k = 3$, e H_0 especifica que as probabilidades p_1, p_2 e p_3 podem ser representadas nas formas: $p_1 = \theta^2$; $p_2 = 2\theta(1 - \theta)$ e $p_3 = (1 - \theta)^2$, então o log da função de verossimilhança de θ , será dado por:

$$\begin{aligned} L(\theta) &= \prod_{i=1}^3 [p_i]^{n_i} \\ l(\theta) &= n_1 \log(\theta^2) + n_2 \log[2\theta(1 - \theta)] + n_3 \log(1 - \theta)^2 \\ &= (2n_1 + n_2) \log \theta + (2n_3 + n_2) \log(1 - \theta) + n_2 \log 2 \end{aligned}$$

onde derivando em relação a θ , igualando a zero e isolando-o, tem-se seu e.m.v.:

$$\hat{\theta} = \frac{2n_1 + n_2}{2n}.$$

O valor da estatística Q_c^2 pode ser calculada utilizando os valores dos n_i 's observados, e em seguida comparada a estatística teste $\chi_{\alpha\%, k-1-s}^2$ segundo R. A. Fisher (1924), onde s = número de parâmetros estimados pela amostra. Assim $\chi_{\alpha\%, k-1-s}^2 = \chi_{5\%, 3-1-1}^2 = \chi_{5\%, 1}^2 = 3,84$.

Exercício 2.1. Verifique ao nível de 5% se as quantidades abaixo seguem a distribuição binomial. Obtenha via e.m.v., o parâmetro θ : probabilidade de sucesso (Obs.: $\hat{\theta} = \frac{\bar{x}}{n}$).

X	0	1	2	3	$\Rightarrow n = 100$
n_i	35	40	20	5	

R. $\hat{\theta} = 0,0095$; $Q_c^2 = 0,9633 < Q_{5\%; (4-1-1=2)}^2 = 5,99$ - **Não rejeita-se H_0 .**

Exemplo 2.3. Seja X uma v.a. onde $0 < x < 1$, com uma f.d.p. desconhecida. Uma a.a. de 100 observações é retirada desta distribuição e deseja-se testar que é proveniente do intervalo $(0,1)$, isto é, se $X \sim U_{(0,1)}$.

Este problema é um caso não-paramétrico, já que a distribuição de X pode ser qualquer distribuição contínua no intervalo $(0,1)$. Dividiremos este intervalo em 20 subintervalos de amplitudes iguais a 0,05. Se a distribuição real é uma uniforme, então a probabilidade de qualquer observação particular cair dentro do i -ésimo subintervalo é $p_i = \frac{1}{20}$, para $i = 1, 2, \dots, 20$. Sendo $n = 100$, o número esperado de observações em cada subintervalo é $n_i p_i^0 = 100 \frac{1}{20} = 5$. Assim temos de acordo com (2.1) temos:

$$Q_c^2 = \frac{1}{5} \sum_{i=1}^{20} (n_i - 5)^2 \sim \chi_{19}^2.$$

Observação: Este método pode ser aplicado para qualquer distribuição contínua.

Em algumas situações onde a distribuição dos dados é desconhecida, uma das primeiras providências é tentar identificar o comportamento da variável com um modelo, de acordo com as características dos dados.

O procedimento conhecido como “Teste de Aderência”, avalia a distribuição candidata, ao qual será rejeitada ou não. Tal teste é o Qui-quadrado.

$$\chi_c^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}, \quad (2.2)$$

onde o_i : valor observado e e_i : valor esperado.

Categoria	1	2	...	k
Freq. Observada (o_i)	o_1	o_2	...	o_k
Freq. Esperada (e_i)	e_1	e_2	...	e_k

As as hipóteses a serem testadas são:

$$\begin{cases} H_0 : & \text{D segue o modelo proposto;} \\ H_a : & \text{D não segue o modelo proposto.} \end{cases}$$

Se D for discreta, as categorias são os próprios valores da variável.

No caso contínuo, as categorias são definidas a partir de faixas de valores da variável.

Espera-se que os valores o_i e e_i , respectivos, tenham valores próximos para que o modelo proposto seja “aceito”, isto é, que os dados amostrais sigam o modelo proposto.

É possível demonstrar que para um tamanho de amostra suficientemente grande, o teste segue uma $\chi_{(k-1)}^2$.

Restrições, segundo SIEGEL (1975).

a. Se $k = 2$ então deve ter $e_i \geq 5$;

b. Se $k > 2$, o teste χ^2 não deve ser usado se mais de 20% dos $e_i < 5$ ou se ao menos um $e_i < 1$.

Observação 1) A determinação do número de graus de liberdade (g.l.) depende do número de restrições envolvidas no cálculo da estatística teste. Via de regra, o número de g.l. para um teste de aderência é igual a $k - (n^\circ \text{ de vínculos nos dados amostrais})$. Há pelo menos um vínculo: $\sum e_i = n$, logo $g.l. = (k - 1) - s$, onde: k : n° de categorias ou classes e s : n° de estatísticas usadas para estabelecer as frequências esperadas ou n° de parâmetros que foram estimados.

Por exemplo, para testar a aderência dos dados à distribuição de Poisson, informações sobre a média populacional são retiradas da amostra, logo: $g.l. = k - 1$ (perda de 1 grau de liberdade, usado para estimar a média).

Observação 2) A expressão (2.2) pode ser escrita como:

$$\chi_c^2 = \sum_{i=1}^k \frac{o_i^2}{e_i} - n$$

e com a mesma, implica uma aproximação de distribuições binomiais por normais, conhecida como a **correção de Yates**:

$$\chi_{Yates}^2 = \sum_{i=1}^k \frac{(|o_i - e_i| - 0.5)^2}{e_i} \sim \chi_{(k-1-s)}^2. \quad (2.3)$$

Exemplo 2.4. No experimento de lançar um dado $n = 180$ vezes, verificar se o mesmo é equilibrado.

H_0 : O dado é equilibrado.

Face (categoria)	1	2	3	4	5	6
Freq. Observada (o_i)	20	35	25	35	32	33
Freq. Esperada (e_i)	$\frac{1}{6}n$	$\frac{1}{6}n$	$\frac{1}{6}n$	$\frac{1}{6}n$	$\frac{1}{6}n$	$\frac{1}{6}n$
Freq. Esperada (e_i)	30	30	30	30	30	30

Valor da estatística teste:

$$\chi_c^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i} = \frac{(20 - 30)^2}{30} + \frac{(35 - 30)^2}{30} + \dots + \frac{(33 - 30)^2}{30} = 6,25$$

Como $\chi_{5\%;k-1}^2 = \chi_{5\%;6-1}^2 = \chi_{5\%;5}^2 = 11,07$ é maior do que $6,25 = \chi_c^2$, então 6,25 pertence a região de não-rejeição de H_0 , logo conclui-se ao nível de 5% de significância que não existem evidências de que o dado seja desequilibrado.

Usando o programa computacional R, a linha de comando será:

```
x=c(20,35,25,35,32,33)
p=c(1/6,1/6,1/6,1/6,1/6,1/6)
?chisq.test
chisq.test(x,p=p)
```

```
x=c(89,37,30,28,2)
p=c(0.40,0.20,0.20,0.19,0.01)
chisq.test(x,p=p)
```

Exemplo 2.5. Deseja-se confirmar a afirmação de que a porcentagem de cinzas contidas em carvão produzido por uma certa empresa segue uma distribuição normal. Que decisão deve-se tomar a 5%, baseado nos dados da Tabela (2.1) ?

R. $\chi_{Yates}^2 = 5,90 \sim \chi_{(10-1-2=7)}^2$; p -valor = 0,554 – “Não Rejeita-se H_0 ”

$$\text{Hipóteses: } \begin{cases} H_0 : X \sim N(\mu = 14,5; \sigma = 1,64) \\ H_a : X \sim \text{outra distribuição.} \end{cases}$$

Para se determinar o valor das frequências esperadas e_i :

$$e_i = n \cdot P(x \in \text{categoria } i \mid \text{se } H_0 \text{ for verdadeira})$$

$$e_1 = 250 \cdot P(x < 10,5) = 250 \cdot P(z < -2,44) = 250 \cdot (0,0073) = 1,825 = e_{10} \text{ (pela simetria)}$$

$$e_2 = 250 \cdot P(10,5 \leq x < 11,5) = 250 \cdot (0,026) = 6,50 = e_9$$

\vdots

Tabela 2.1: *Distribuição de cinzas de carvão (%)*.

Cinzas (%)	x_i	Freq. o_i	Freq. e_i
9,5 ┆ 10,5	10	2	1,83
10,5 ┆ 11,5	11	5	6,50
11,5 ┆ 12,5	12	16	19,25
12,5 ┆ 13,5	13	42	39,92
13,5 ┆ 14,5	14	69	57,25
14,5 ┆ 15,5	15	51	57,25
15,5 ┆ 16,5	16	32	39,92
16,5 ┆ 17,5	17	23	19,25
17,5 ┆ 18,5	18	9	6,5
18,5 ┆ 19,5	19	1	1,83
Total	-	250	-

Observação: Como a 1ª e 10ª categorias apresentam frequências esperadas menores que 5, devemos agrupar a 1ª com a 2ª e a 9ª com a 10ª. Desta forma a Tabela (2.1) ficará da forma apresentada na Tabela (2.2), onde $\chi_c^2 = 6,46$ pertencerá a região de não-rejeição de H_0 , já que o

Tabela 2.2: *Distribuição de cinzas de carvão (%)*.

Cinzas (%)	Freq. o_i	Freq. e_i
$\leq 11,5$	7	8,33
11,5 ┆ 12,5	16	19,25
12,5 ┆ 13,5	42	39,75
13,5 ┆ 14,5	69	57,67
14,5 ┆ 15,5	51	57,67
15,5 ┆ 16,5	32	39,75
16,5 ┆ 17,5	23	19,25
$\geq 17,5$	10	8,33
Total	250	250

valor tabelado $\chi_{5\%;(8-1-2)}^2 = 11,07$ será tomado como referência.

Conclui-se ao nível de 5% de significância, que existem evidências de que a % de cinzas no carvão segue uma distribuição normal.

```

xinf=c(0,10.5,11.5,12.5,13.5,14.5,15.5,16.5,17.5,18.5)
xsup=c(10.5,11.5,12.5,13.5,14.5,15.5,16.5,17.5,18.5,30)
# O primeiro (9.5) e o último (19.5) valor das classes foram substituídos
# por 0 e 30, pois deseja-se obter a probabilidade  $\leq 10.5$  e  $\geq 18.5$ 
Fo=c(2,5,16,42,69,51,32,23,9,1)
n=sum(Fo)
p=(pnorm(xsup,mean=14.5,sd=1.64)-pnorm(xinf,mean=14.5,sd=1.64))
Fe=p*n
chisq.test(Fo,p)# Observe o Erro !
Q2Yates=sum((abs(Fo-Fe)-0.5)^2/Fe)
# O Teste Qui-quadrado corrigido deve ser comparado ao valor tabelado
# com (nº de classes - 1 - s) graus de liberdade
Q2Tab=qchisq(0.95,7)
pvalor=1-pchisq(Q2Yates,7)
pvalor

```

Exercício 2.2. Deseja-se estudar a tolerância de um equipamento eletrônico ao número de impactos termo-elétricos. Pelas características de fabricação do equipamento, é possível admitir que a probabilidade de falha seja constante, isto é, após cada impacto, existe uma probabilidade p de que ele falhe. Representando por X a variável aleatória "número de impactos anteriores à falha", pretende-se verificar se o modelo geométrico com $p = 0,4$ é adequado para caracterizar essa variável. A amostra é constituída de 80 equipamentos submetidos a sucessivos impactos termo-elétricos até a ocorrência da primeira falha. Se o modelo geométrico fosse verdadeiro, deveríamos ter, para cada valor da variável X , a frequência observada não muito diferente que seria calculada através da distribuição geométrica que é denominada "frequência esperada". Para avaliar a distância das frequências observadas e esperadas, utilizaremos a distribuição χ^2 , ao nível de 5% de significância. Sabe-se que a f.d.p. geométrica é dada por:

$$\pi(x) = P(X = x) = pq^x, \quad x = 0, 1, 2, \dots, n.$$

As hipóteses a serem testadas são:

$$\begin{cases} H_0 : & \text{Os dados seguem o modelo geométrico com parâmetro } p = 0,4; \\ H_a : & \text{Os dados não seguem o modelo geométrico.} \end{cases}$$

Os valores observados foram:

Impactos	0	1	2	3	4	>4
Freq. Observada (o_i)	30	26	10	5	5	4
Freq. Esperada (e_i)	$80(0,4) = 32$					

R. $\chi_c^2 = 4,24$; $\chi_{5\%,4=(6-1-1)}^2 = 9,49$; "não rejeita-se H_0 ".

Exercício 2.3. Tabele os dados de Exemplo 2.8 (inicie no X mínimo com $h = 11$), teste a hipótese $H_0 : X \sim N\left(\bar{x}, \frac{\sigma}{\sqrt{n}}\right)$ ao nível de 5%.

Exercício 2.4. (Lindström, 1918 & Cochran, 1980) Em um experimento envolvendo o cruzamento de dois tipos de milho, encontraram-se quatro tipos distintos: verde (T_1), dourado (T_2), listras verdes (T_3) e listras verdes e douradas (T_4). De acordo com a hereditariedade mendeliana, as probabilidades de obter esses quatro tipos são: $\frac{9}{16}$, $\frac{3}{16}$, $\frac{3}{16}$ e $\frac{1}{16}$, respectivamente. Em 1301 plantas da segunda geração, Lindström encontrou as seguintes frequências:

Milho	T_1	T_2	T_3	T_4
Freq. Observada (o_i)	773	231	238	59
Freq. Esperada (e_i)	$\frac{9}{16} 1301 = 731,81$			

Utilize o Teste χ^2 para aderência, ao nível de 5% para comprovar ou não a lei de hereditariedade mendeliana.

R. $\chi_c^2 = 9,27$; $\chi_{5\%,3}^2 = 7,815$; “rejeita-se H_0 ”.

```
x=c(773,231,238,59)
p=c(9/16,3/16,3/16,1/16)
chisq.test(x,p=p)
```

2.3 Teste Kolmogorov-Smirnov

Os estatísticos russos Kolmogorov e Smirnov desenvolveram um método, em geral mais poderoso que do χ^2 , para testar a aderência, em que a variável de teste é a maior diferença observada entre a função de distribuição acumulada do modelo e a da amostra.

A função de distribuição acumulada do modelo testado, ou função de repartição, dá as probabilidades acumuladas em cada ponto, ou seja, $F(x) = P(X \leq x)$. A função de distribuição acumulada da amostra corresponderá ao gráfico de frequências relativas acumuladas.

Designaremos essa segunda função por $G(x)$:

- para dados não-agrupados:

$$G(x) = \frac{k}{n}, \quad k = 0, 1, 2, \dots, n;$$

- para dados agrupados:

$$G(x_i) = \frac{Fac_i}{n}, \quad i = 0, 1, 2, \dots, n.$$

O teste consta simplesmente da verificação do valor:

$$d_c = \max |F(x) - G(x)| \quad (2.4)$$

e da comparação com o valor crítico tabelado em função de α (nível de significância) e n (tamanho amostral). Rejeita-se H_0 : “Segue a distribuição de interesse”, caso d_c pertencer a região de rejeição, isto é, se $d_c > d_{\alpha\%;n}$ (Anexo -Tabela de valores críticos da distribuição da Estatística D_n de Kolmogorov-Smirnov).

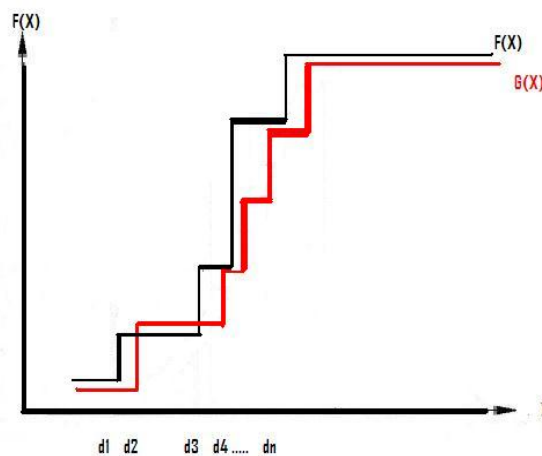


Figura 2.2: Distribuição Kolmogorov-Smirnov.

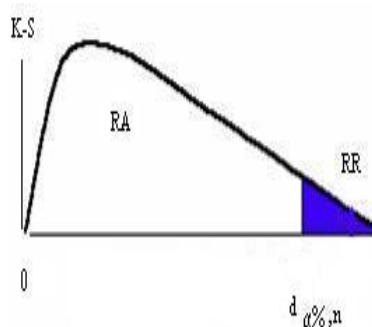


Figura 2.3: Distribuição de Kolmogorov-Smirnov.

Observação: Se $n > 35$, os valores críticos para $\alpha = 5\%$ e $\alpha = 1\%$, respectivamente, serão dados por:

$$\frac{1,36}{\sqrt{n}} \quad \text{e} \quad \frac{1,63}{\sqrt{n}}. \quad (2.5)$$

O método é exato para distribuições contínuas de parâmetros conhecidos, devendo a função $G(x)$ ser construída com base nos valores individuais da amostra. Quando os parâmetros são estimados, o modelo testado é discreto ou os dados são agrupados em classes (caso em que $G(x)$ corresponderá ao polígono de frequências relativas acumuladas), o teste é válido apenas por aproximação.

Exemplo 2.6. Uma amostra de 10 elementos forneceu os seguintes valores:

27,8	29,2	30,6	27,0	33,5
29,5	27,3	25,4	28,0	30,2

Testar a hipótese de que a amostra seja proveniente de uma população Normal de média 30 e desvio-padrão 2, ao nível de 5% de significância.

Tabela 2.3: Tabela de cálculos da estatística d .

i	x	z	F(x) = P(X ≤ x)	G(x)	F(x) - G(x)	
					esquerda	direita
	-	-	-	0,0	0,0107	-
1	25,4	-2,3	0,0107	0,1	0,0332	0,0893
2	27,0	-1,5	0,0668	0,2	0,1115	0,1332
3	27,3	-1,35	0,0885	0,3	0,1643	0,2115
4	27,8	-1,1	0,1357	0,4	0,2413	0,2643
5	28,0	-1,0	0,1587	0,5	0,1554	0,3413
6	29,2	-0,4	0,3446	0,6	0,1987	0,2554
7	29,5	-0,25	0,4013	0,7	0,1602	0,2987
8	30,2	0,1	0,5398	0,8	0,1221	0,2602
9	30,6	0,3	0,6779	0,9	0,0959	0,2221
k=10	33,5	1,75	0,9959	1,0	-	0,0041

De acordo com os resultados apresentados na Tabela (2.3), $d_c = \max |F(x) - G(x)| = 0,3413$ pertence a região de não-rejeição de H_0 , já que $d_{5\%;10} = 0,409$. Portanto, os dados seguem a distribuição Normal com média 30 e desvio-padrão 2 ($p = 0,1526$).

```

x=sort(c(27.8,29.2,30.6,27,33.5,29.5,27.3,25.4,28,30.2))
mean(x)
sd(x)
?ks.test
ks.test(x,'pnorm',30,2)
# Gráfico da função de Distribuição Acumulada empírica
n=length(x)
#n=sum(!is.na(x))#caso tenha 'missing' - este comando elimina-os
y=seq(1,n,1)
plot(sort(x),(1:n)/n,type='s',xlab='x',ylab='F(x)')
lines(sort(x),pnorm(x,mean=mean(x),sd=sd(x)))

```

Exemplo 2.7. Verificar ao nível de 5% e a 1% se os dados da Tabela (2.4) podem ser estudados através da distribuição de Poisson, com $\lambda = 1, 2$.

Tabela 2.4: Dados supostamente com distribuição de Poisson com média 1,2.

x	F _i
0	15
1	25
2	10
3	5
4	4
5	1
Total	60

Sabe-se que se $X \sim \text{Poisson}(\lambda)$, então sua função densidade de probabilidade (Tabela em Anexo da Função de distribuição cumulativa da Poisson) é:

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}; \quad x = 0, 1, 2, \dots$$

logo tem-se os seguintes resultados:

Tabela 2.5: Tabela de cálculos da estatística d para os dados da Tabela 2.4.

x	F _i	F _{ac}	P(X=x)	F(x)	$G(x) = \frac{F_{ac}}{n}$	$ F(x) - G(x) $
0	15	15	0,3012	0,3012	$\frac{15}{60} = 0,2500$	0,0512
1	25	40	0,3614	0,6626	$\frac{40}{60} = 0,6667$	0,0041
2	10	50	0,2169	0,8795	$\frac{50}{60} = 0,8333$	0,0462
3	5	55	0,0867	0,9662	0,9167	0,0495
4	4	59	0,0260	0,9922	0,9833	0,0089
5	1	60	0,0062	0,9984	1,000	0,0016

```
x=c(0,1,2,3,4,5)
Fo=c(15,25,10,5,4,1)
n=sum(Fo)
lambda=1.2
Fx=ppois(x,lambda)
Gx=cumsum(Fo/n)
ks.test(Fx,Gx)
chisq.test(Fx,Gx)
plot(Fx,type='s',xlab='x',ylab='Probabilidade')
lines(Gx,type='s',col='red')
legend(5,0.8,c('Fx','Gx'),lty=c(1,1),col=c('black','red'),bty='n')
```

H_0 : “Os dados seguem a distribuição de Poisson com parâmetro $\lambda = 1,2$ ”

Utilizando as aproximações dadas em (2.5), tem-se:

$$d_{5\%;60} = \frac{1,36}{\sqrt{60}} = 0,18$$

e

$$d_{1\%;60} = \frac{1,63}{\sqrt{60}} = 0,21,$$

como $d_c = 0,0512 < 0,18 < 0,21$, então conclui-se pela não-rejeição de H_0 , isto é, os dados seguem a distribuição de Poisson com parâmetro 1,2, ao nível de 1% e 5%.

```
set.seed(5436)
?rnorm
x=rnorm(50)
y=runif(30)
par(mfrow=c(1,2))
hist(x)
hist(y)
# x e y são da mesma distribuição ?
ks.test(x,y)
```

```
z=x+2
# z é uma transformação linear de x
ks.test(z,'pgamma',3,2)
```

2.4 Teste de Shapiro-Wilk para normalidade

Este teste é mais uma alternativa para verificar a normalidade dos dados.

O teste de Lilliefors, implementado em alguns softwares, também é utilizado para verificar a normalidade nos dados.

O teste Shapiro-Wilk é recomendado para amostras menores que 50. Já o K-S seria para valores maiores.

Estatística teste

Considere uma amostra aleatória $X^{(1)} \leq X^{(2)} \leq \dots \leq X^{(n)}$ ordenada, associada a uma função de distribuição desconhecida $F(x)$.

$$T_{S-W} = \frac{1}{D} \left[\sum_{i=1}^{k=\frac{n}{2}} a_i (X^{n-i+1} - X^i) \right]^2 \quad (2.6)$$

onde

$$D = \sum_{i=1}^{k=\frac{n}{2}} (x_i - \bar{x}_i)^2 = s^2(n-1).$$

X^i : a i -ésima ordem da estatística, onde obtemos pela Tabela S-W, os coeficientes a_i (Anexo - Tabela de coeficientes de Shapiro-Wilk) para k aproximadamente $\frac{n}{2}$.

A estatística T é basicamente o quadrado de um coeficiente de correlação, onde o coeficiente de correlação de Pearson é calculado entre a estatística ordenada X^i na amostra e o escore a_i , que representa o que a estatística ordenada deveria parecer se a população é Normal. Portanto:

- se T estiver próximo de “1” a amostra se comporta como uma amostra Normal;
- se T estiver distante de “1” a amostra se comporta como uma amostra diferente da distribuição Normal;

Hipóteses:

H_0 : $F(x)$ é uma função de distribuição normal;

H_a : $F(x)$ não é normal.

Rejeita-se H_0 ao nível α de significância, se $T_{S-W} < T_{\alpha\%,n}$ (Anexo - Tabela dos valores críticos de Shapiro-Wilk).

Exemplo 2.8. 50 números são selecionados aleatoriamente de uma lista telefônica e o teste S-W será usado para verificar a normalidade dos dados já ordenados, ao nível de 5% de significância.

23	23	24	27	29	31	32	33	33	35
36	37	40	42	43	43	44	45	48	48
54	54	56	57	57	58	58	58	58	59
61	61	62	63	64	65	66	68	68	70
73	73	74	75	77	81	87	89	93	97

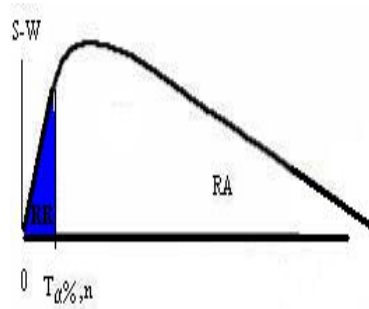


Figura 2.4: Distribuição de Shapiro-Wilk.

Solução: $n = 50$; $\bar{x} = 55,04$; $s = 19,005$, logo $D = 19,005^2(49) = 17697,92$;

Para a obtenção de $\sum_{i=1}^k a_i (X^{n-i+1} - X^i)$, $k = \frac{n}{2} = \frac{50}{2} = 25$ classes:

i	a_i (Anexo - Tabela pg. 17)	$X^{n-i+1} - X^i$	
1	0,3751	74	$\Rightarrow X^{50-1+1} - X^1 = 97 - 23 = 74$
2	0,2574	70	$\Rightarrow X^{50-2+1} - X^2 = 93 - 23 = 70$
\vdots	\vdots	\vdots	
24	0,0104	1	
25	0,0035	1	

Desta forma

$$\sum_{i=1}^k a_i (X^{n-i+1} - X^i) = 0,3751(74) + 0,2574(70) + \dots + 0,0035(1) = 130,63$$

logo

$$T_{S-W} = \frac{(130,63)^2}{17697,92} = 0,9642.$$

O valor tabelado $T_{\alpha\%,n} = T_{5\%,50} = 0,947 < 0,9642 = T_{S-W}$, logo não rejeitamos H_0 , e concluímos que os dados seguem a distribuição normal.

```
x=c(23,23,24,27,29,31,32,33,33,35,36,37,40,42,43,43,44,45,48,
48,54,54,56,57,57,58,58,58,58,59,61,61,62,63,64,65,66,68,68,70,
73,73,74,75,77,81,87,89,93,97)
?shapiro.test
shapiro.test(x)
qqnorm(x)
qqline(x,col=2)
```

2.4.1 Aproximações para o teste Shapiro-Wilk

- Se $3 \leq n \leq 6$

$$W_{(2)} = \ln \left[\frac{T - d_n}{1 - T} \right] \sim N(0, 1) \quad (2.7)$$

onde T é obtido em (2.6) e d_n é tabelado (Anexo - Tabela de converção de Shapiro-Wilk).

Exemplo 2.9. Considere $n = 5$ e $T_{SW} = 0,885$, então $W_{(2)} = \ln \left[\frac{0,885 - 0,5521}{1 - 0,885} \right] = +1,06 < Z_{2,5\%} = +1,96$, logo não rejeitamos H_0 , e concluímos que os dados seguem a distribuição normal.

- Se $7 \leq n \leq 50$

$$G = W_{(3)} = b_n + c_n \ln \left[\frac{T - d_n}{1 - T} \right] \sim N(0, 1) \quad (2.8)$$

onde b_n , c_n e d_n são valores tabelados (Anexo - Tabela de converção de S-W).

Exemplo 2.10. (continuação do Exemplo anterior)

como $n = 50$ e

$$G = -7,677 + 2,212 \ln \left[\frac{0,9642 - 0,1436}{1 - 0,9642} \right] = -0,7488$$

é um valor que pertence a região de não-rejeição de H_0 (já que em nível de 5% de significância $Z_c = \pm 1,96$), concluímos que os dados seguem a distribuição normal.

```
set.seed(434)
x=rnorm(100,mean=5,sd=3)
y=runif(100,min=2,max=4)
par(mfrow=c(2,2))
hist(x)
hist(y)
qqnorm(x)
qqline(x,col=2)
qqnorm(y)
qqline(y,col=2)
shapiro.test(x)
shapiro.test(y)
```

Exercício 2.5. O retorno sob investimentos em 12 meses para 20 ações selecionadas aleatoriamente são as seguintes:

```
9,1  5,0  7,3  7,4  5,5  8,6  7,0  4,3  4,7  8,0
4,0  8,5  6,4  6,1  5,8  9,5  5,2  6,7  8,3  9,2
```

(a) teste a hipótese nula de normalidade usando Shapiro-Wilk (T); **R.** $T_{S-W} = 0,9545 > 0,905 = T_{5\%,20}$, **não rejeita-se H_0 .**

(b) teste a hipótese nula de normalidade usando Shapiro-Wilk (G); **R.** $G = -0,18 \in RA$.

(c) teste a hipótese nula de normalidade ($X \sim N(\bar{x}; s)$) usando Kolmogorov-Smirnov; **R.** $d_{\max} = 0,111 \in RA$.

```
x=sort(c(9.1,5,7.3,7.4,5.5,8.6,7,4.3,4.7,
8,4,8.5,6.4,6.1,5.8,9.5,5.2,6.7,8.3,9.2))
shapiro.test(x)
qqnorm(x)
qqline(x,col=2)
ks.test(x,'pnorm',mean(x),sd(x)) # Comparando K-S com S-W
```

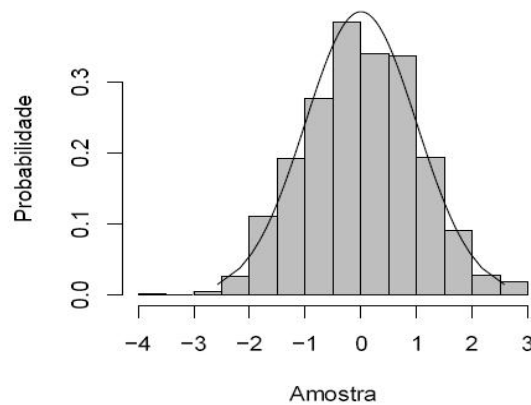



Figura 2.5: *Distribuição de uma amostra de dados (Histograma).*

2.5 Testes gráficos

Alguns testes gráficos são utilizados como auxiliares na determinação de normalidade ou aderência dos dados a uma distribuição. Inicialmente um simples histograma pode nos indicar se a amostra segue uma distribuição normal.

```
set.seed(10)
amostra=rnorm(1000)
hist(amostra,main=' ',prob=TRUE,ylab='Probabilidade',col='gray')
x=seq(-3,3,len=1000)
fx=dnorm(x,0,1)
lines(x,fx,lty=1,col=2)
```

Outras técnicas mais utilizados são:

2.5.1 Q-Q plots

Consistem em representar em um plano cartesiano os quantis amostrais observados (x_i) versus os quantis esperados da distribuição normal (m_i : valores médios das estatísticas de ordem da normal padrão). Quanto mais alinhados os pontos maior será o indício de normalidade.

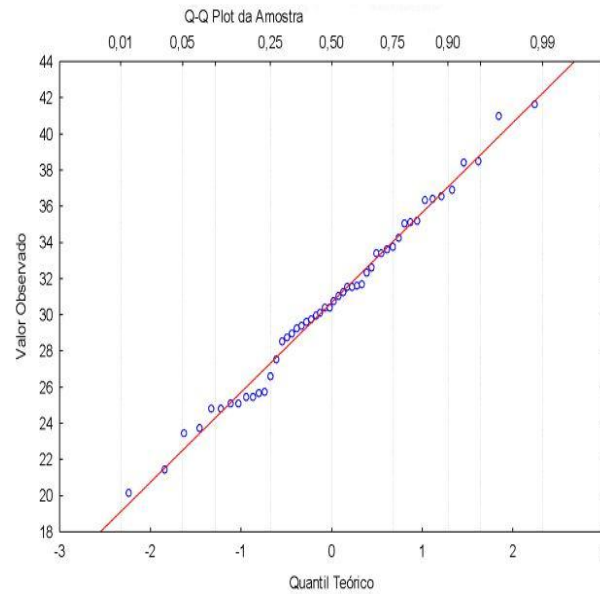
```
x=sort(c(27.8,29.2,30.6,27,33.5,29.5,27.3,25.4,28,30.2))
qqnorm(x)
qqline(x,col=2)
```

2.5.2 P-P plots

Alternativo, que substitui os quantis observados e esperados da distribuição normal pelas respectivas probabilidades acumuladas da distribuição normal. Existem outras técnicas gráficas: Q-Q norm, por exemplo, é muito similar ao Q-Q plot.

Exercício 2.6. *Suponha uma amostra aleatória de 5.000 itens, onde 35 são defeituosos. Teste ao nível de 5% que apenas 1% dos itens são defeituosos (Utilize o teste χ^2). **R.** $\chi_c^2 = 4,55$ - **rejeita-se H_0 .***

Exercício 2.7. *O número de defeitos por unidade observado em uma amostra de 100 aparelhos de*

Figura 2.6: Gráfico *QQ-Plot*.

televisão produzidos em uma linha de montagem apresentou a seguinte distribuição de frequências:

Nº de defeitos:	0	1	2	3	4	5	6	7
Nº de aparelhos:	25	35	18	13	4	2	2	1

Verificar ao nível de 5% de significância, se o número de defeitos por unidade segue razoavelmente uma distribuição de Poisson. **R.** $\chi_c^2 = 18,69 > 12,59 = \chi_{5\%,6}^2$ - **rejeita-se H_0 .**

Exercício 2.8. Ao analisar o número de um determinado tipo de célula observado em 400 campos de um hemocitômetro, um pesquisador assinalou os seguintes resultados. Verificar se esses dados podem ser considerados como tendo distribuição de Poisson. **R.** $\chi^2 = 12,82 < 15,51 = \chi_{5\%,8}^2$ - **Não rejeita-se H_0 .**

Nº de células:	0	1	2	3	4	5	6	7	8	9
Nº de campos:	75	103	121	54	30	13	2	1	0	1

Exercício 2.9. Em um determinado jogo esportivo, 200 pessoas foram selecionadas aleatoriamente e perguntou-se quantos dos 4 jogos prévios eles tinham assistido. Teste ao nível de 5% que a hipótese que estes 200 valores observados podem ser vistos como uma amostra de uma distribuição binomial, tal que as probabilidades são as seguintes:

$$P_0 = (1 - \theta)^4; P_1 = 4\theta(1 - \theta)^3; P_2 = 6\theta^2(1 - \theta)^2; P_3 = 4\theta^3(1 - \theta); P_4 = \theta^4.$$

Nº de jogos vistos previamente	Nº de pessoas
0	33
1	67
2	66
3	15
4	19

Obtenha o e.m.v de θ . **R.** $\hat{\theta} = 0,4$; $\chi_c^2 = 47,81 > 7,815 = \chi_{5\%,3}^2$ - **Rejeita-se H_0 .**

Exercício 2.10. Determine se as notas de uma grande turma de Psicologia introdutória podem ser aproximadas por uma distribuição normal com média 50,5 e desvio-padrão 10. Use um nível de significância de 5%.

Notas	F_i
< 25	14
25 - 30	18
30 - 35	22
35 - 40	20
40 - 45	40
45 - 50	30
50 - 55	22
55 - 60	20
60 - 65	2
65 - 70	6
70 - 75	0
75 - 80	4
≥ 80	2

R. $\chi_c^2 = 179,5 \gg 12,6 = \chi_{5\%,(6)}^2$ - **rejeita-se H_0 .**

Exercício 2.11. Num ensaio com bananeiras, realizado pelo Dr. Jairo Ribeiro de Mattos, da E.S.A. da ESALQ, foram obtidos os seguintes pesos médios (em kg) de cacho:

13,9 17,7 17,9 18,3 18,5 18,9 19,4 19,8 20,2 20,6 21,1 21,3 21,7
21,9 22,0 22,2 22,7 22,8 23,2 23,3 23,4 23,8 24,4 24,4 24,9 -

a) Testar ao nível de 5% se é possível estudar os dados através de uma distribuição normal de média 20 e variância 5,25. Utilize o teste de Shapiro-Wilk. **R.** $T_c = 0,95 > 0,918 = T_{5\%}$ - **Não rejeita-se H_0 .**

b) Refaça o teste, utilizando uma aproximação. **R.** $G_c = -0,64 < |\pm 1,96| = Z_{5\%}$ - **Não rejeita-se H_0 .**

```
x=c(13.9,17.7,17.9,18.3,18.5,18.9,19.4,19.8,20.2,20.6,
21.1,21.3,21.7,21.9,22.0,22.2,22.7,22.8,23.2,23.3,
23.4,23.8,24.4,24.4,24.9)
ks.test(x,'pnorm',20,sqrt(5.25))
shapiro.test(x)
```

Exercício 2.12. Os tempos entre as chegadas de 8 fregueses consecutivos ao balcão de informações de uma loja foram, em minutos:

1,0 0,6 1,2 0,6 1,8 0,3 0,5 2,0

Teste por meio de *K-S* a hipótese de que os tempos entre as chegadas obedecem a uma distribuição exponencial ao nível de 5% de significância [dados : $f(x) = \lambda e^{-\lambda x}$; $\mu(x) = \frac{1}{\lambda}$; $F(x) = 1 - e^{-\lambda x}$].

R. $d_c = 0,268 < 0,457 = d_{5\%,8}$ - **Não rejeita-se H_0** ($p = 0,611$).

```
x=sort(c(1,0.6,1.2,0.6,1.8,0.3,0.5,2))
lambda=1/mean(x)
ks.test(x,'pexp',lambda)
```

Exercício 2.13. Num ensaio com animais em confinamento, foram obtidos os pesos médios (em kg) após o tratamento de uma dieta experimental:

13,9 17,7 17,9 18,3 18,5 18,9 19,4 19,8 20,2 20,6 21,1 21,3
 21,7 21,9 22,0 22,2 22,7 22,8 23,2 23,3 23,4 23,8 24,4 24,4
 24,9

Teste a possibilidade de estudar os dados através de uma distribuição normal com média 20 e variância 5,25, por meio do método de Kolmogorov-Smirnov ao nível de 5% de significância. **R.** **Não, pois $d_c = 0,29 > 0,27 = d_{5\%,25}$ - “Rejeita-se H_0 ”.**

Exercício 2.14. Utilizando os dados de tempo de vida de insetos, testar ao nível de 5% de significância a normalidade por meio dos testes (utilize um software estatístico para análise):

a) Qui-quadrado;

b) Kolmogorov-Smirnov; ($p = 0,08198$ - **Não Rejeita H_0**)

c) Shapiro-Wilk (T); ($T=0,8594$; $W_{(3)}=3,64$; $p = 0,00022$ - **“Rejeita-se H_0 ”**)

8,5209 4,1871 2,5163 1,9133 8,7796 5,9117 0,7608 12,0372 2,6035 1,6889
 5,6255 6,3604 5,0679 3,0310 1,1281 1,3850 12,5783 2,0292 0,5953 0,4451
 3,6014 7,8288 1,3829 1,9344 0,8642 8,5144 4,9774 0,5759 1,5033 0,4750
 1,0414 0,3009 1,7809 2,5638 5,3587 2,3066 1,5298 8,1052 3,1507 8,6276

Comando 2.1. *Trabalhando dados da distribuição Weibull*

$$T \sim \text{Weibull}(\text{shape} = \beta; \text{escala} = \mu); \quad f(t) = \left(\frac{\beta}{\mu}\right) \left(\frac{t}{\mu}\right)^{\beta-1} \exp\left(-\left(\frac{t}{\mu}\right)^{\beta}\right)$$

```
t=sort(rweibull(3000,shape=2,scale=5))
y=dweibull(t,shape=2,scale=5,log=FALSE)
#calcula o valor na f.d.p. f(x) na escala natural - ver ?dweibull
plot(t,y,type='l',ylab='Probabilidade')

Weibull.Normal=function(semente,n,beta,mi)
{
  set.seed(semente)
  par(mfrow=c(2,2))
  t=rweibull(n,beta,mi)
  hist(t,main='Distribuição Weibull',xlab='Tempo',ylab='Frequência')
  library(MASS)
  pars=fitdistr(t,'weibull') # ver ?fitdistr
  beta=pars$estimate[1]
  mi=pars$estimate[2]
  hist(t,prob=T,nclass=20,main=' ',xlab='Tempo',ylab='Probabilidade')
  lines(sort(t),dweibull(sort(t),beta,mi))
  lines(sort(t),dnorm(sort(t),mean(t),sd(t)),col='red')
  t2=sort(t)
  plot(qweibull(ppoints(t2),shape=beta,scale=mi),t2,xlab='Tempo',
  ylab='Tempo-Ajustado',main='P-PPlot')
  abline(lsf(qweibull(ppoints(t2),shape=beta,scale=mi),t2))
  qqnorm(t)
  qqline(t,col=2)
  s=shapiro.test(t)
  cat('\n','parâmetros estimados:')
  list(pars,s)
}

# Weibull.Normal(semente=135,n=1000,beta=2,mi=5)
```

Capítulo 3

Testes aplicáveis a duas amostras - Dados pareados

Podemos distinguir três tipos de pareamento: *auto-pareamento*, *pareamento natural* e *pareamento artificial*.

O *auto-pareamento* ocorre quando o indivíduo serve como seu próprio controle, como na situação em que um indivíduo recebe duas drogas administradas em ocasiões diferentes. Outra situação é a que um tratamento é administrado e as variáveis de interesse são observadas antes e depois do programa. Finalmente, a comparação de dois órgãos no mesmo indivíduo, como braços, pernas, olhos, narinas, segundo alguma característica estudada também constitui um auto-pareamento.

O *pareamento natural* consiste em formar pares tão homogêneos quanto possível, controlando os fatores que possam interferir na resposta, sendo que o pareamento aparece de forma natural. Por exemplo, em experimentos de laboratório pode-se formar pares de cobaias selecionadas da mesma ninhada; em investigações clínicas, gêmeos univitelinos são muito usados.

No *pareamento artificial* escolhe-se indivíduos com características semelhantes, tais como, idade, sexo, nível sócio-econômico, estado de saúde ou, em geral, fatores que podem influenciar de maneira relevante a variável resposta.

3.1 Teste dos Sinais

A prova dos sinais tem sua denominação devida ao fato de utilizar como dados sinais “+” e “-”, ao invés de medidas quantitativas. Particularmente útil nos trabalhos de pesquisa em que é impossível ou inviável a obtenção de uma mensuração quantitativa, mas é possível estabelecer *postos* em relação a cada um dos dois membros de cada par.

A prova dos sinais é aplicável no caso de duas amostras relacionadas quando o pesquisador deseja determinar se duas condições são diferentes. A única suposição que a prova dos sinais exige, é que a variável em estudo tenha distribuição básica contínua. A prova não faz qualquer suposição sobre a forma da distribuição das diferenças, nem supõe que todos os indivíduos tenham sido extraídos da mesma população. Os diferentes pares podem provir de populações diferentes com respeito a idade, sexo, inteligência etc; a única exigência é que, dentro de cada par, o pesquisador tenha conseguido um nivelamento quanto às variáveis extrínsecas importantes.

Método

A hipótese de nulidade comprovada pela “prova dos sinais” é:

$$H_0 : P(X_A > X_B) = P(X_A < X_B) = \frac{1}{2}$$

ou

$$H_0 : \text{“A diferença mediana é zero” ou “Antes = Depois” ou “Não houve alteração”}$$

onde X_A é o julgamento, ou escore, sob uma das condições (ou após o tratamento) e X_B é o julgamento, ou escore, sob a outra condição (ou antes do tratamento). Isto é, X_A e X_B são dois escores para um par.

Pequenas amostras ($n < 20$)

Neste caso a Tabela da distribuição binomial é utilizada com o valor próprio de n e $p = 1/2$.

$p = P(X \geq x)$, x = número de sinais “+” ou $p = P(Y \leq y)$, número de sinais “-”.

Exemplo 3.1. *É possível que a avaliação de um imóvel, feita por um avaliador, seja mais alta do que a avaliação feita por um concorrente. Para testar essa suposição, pede-se a ambos os avaliadores que avaliem uma amostra de nove imóveis; suas estimativas (em R\$) são:*

<i>Propriedade</i>	<i>Avaliador 1</i>	<i>Avaliador 2</i>	<i>Sinal</i>
1	79.250,00	77.380,00	+
2	110.230,00	106.450,00	+
3	93.750,00	95.100,00	-
4	88.360,00	86.000,00	
5	121.300,00	118.960,00	
6	124.820,00	122.630,00	
7	105.220,00	101.710,00	
8	106.660,00	109.850,00	
9	83.500,00	80.760,00	

Aplique o teste dos sinais, ao nível de 5% de significância, para testar as hipóteses:

$$H_0 : \tilde{\mu}_D = 0 = (\tilde{\mu}_1 - \tilde{\mu}_2)$$

$$H_A : \tilde{\mu}_D > 0$$

onde $\tilde{\mu}_D$ é a mediana da população de diferenças (primeiro avaliador menos segundo avaliador).

Solução

Primeiramente cada par de valores é substituído por um sinal “+” se o primeiro valor é maior do que o segundo, e por um sinal “-” caso contrário:

Propriedade	1	2	3	4	5	6	7	8	9
Sinal	+	+	-	+	+	+	+	-	+

Assim, $x = 7$ (total de “+” é a estatística teste), e a Tabela de distribuições binomiais (Anexo - Tabela de Função distribuição cumulativa da Binomial), para $n = 9$ e $p = 0,5$, a probabilidade $P(x \geq 7) = 1 - P(x < 7) = 1 - 0,9102 = 0,09$ (este é o p-valor). Como $p = 0,09 > 0,05 = \alpha\%$, não rejeitamos H_0 . Os dados não confirmam a suposição de que as estimativas do primeiro avaliador sejam, superiores às do segundo avaliador.

Observação: Se houver “0” então $n^* = n -$ número de zeros.

Grandes amostras ($n \geq 20$, geralmente np e $n(1-p) > 5$)

O valor da variável aleatória binomial pode ser aproximada usando:

$$Z_c = \frac{x - 0,5n}{\sqrt{0,25n}} \sim N(0, 1). \quad (3.1)$$

Exemplo 3.2. São apresentadas, a seguir, as classificações, por dois supervisores, do desempenho de uma amostra aleatória de empregados de uma grande companhia, em uma escala de 1 a 100:

Supervisor 1	88	69	97	60	82	90	65	77	86	79	65	95	88	91	68	77	74	85
Supervisor 2	73	67	81	73	78	82	62	80	81	79	77	82	84	93	66	76	74	78
Sinais	+	+	+	-	+	+	+	-	+	0	-	+	+	-	+	+	0	+

Aplicando o teste dos sinais (supor $n \geq 20$) ao nível de 2%, verifique se a diferença entre as medianas dos dois conjuntos de classificações 80,5 e 78,0, respectivamente, para supervisor 1 e 2, pode ser atribuída ao acaso.

Solução

Observação: Pares de mesma classificação devem ser ignorados.

As hipóteses são:

$$H_0 : \tilde{\mu}_D = 0$$

$$H_A : \tilde{\mu}_D > 0$$

Como $n = 18$ então $n^* = 18 - 2 = 16$. A estatística calculada será:

$$Z_c = \frac{12 - 0,5(16)}{\sqrt{0,25(16)}} = 2$$

Assim, com $x = 12$ e $n = 16$, então $Z_c = 2$ ($p = 0,02275 \times 2 = 0,0455$). Para $\alpha = 2\%$, tem-se $z = \pm 2,33$, e como Z_c está entre eles, a hipótese nula não é rejeitada; em outras palavras, a diferença entre as avaliações medianas dos dois avaliadores não é significativa.

Exercício 3.1. São apresentados, a seguir, os números de ovos postos por uma lagarta antes e após o uso de um pesticida.

Antes	Depois
232	189
255	249
249	251
265	230
236	218
240	233
249	236

Utilize o teste dos sinais ao nível de 1% para testar a hipótese de que o pesticida não foi eficaz. Tome como alternativa $\tilde{\mu}_D > 0$. **R. $p = 0,0625$**

Exercício 3.2. Os valores a seguir são medidas de glicemia de pacientes em dois equipamentos distintos.

Aparelho I	142	166	105	179	205	117	182	155	172	199	211	168	143
Aparelho II	171	190	88	184	244	119	186	166	168	231	221	155	170
Aparelho I	166	168	156	220	208	147	155	163	102	124	182	211	
Aparelho II	171	192	158	241	206	160	165	171	90	124	193	224	

Utilize o teste dos sinais ao nível de 5% para testar a hipótese de que as mensurações dos aparelhos não diferem. Tome como alternativa $\tilde{\mu}_D > 0$. **R. $p = 0,004372$**

Exercício 3.3. Doze casais de gêmeos idênticos foram submetidos a testes psicológicos para medir de algum modo a quantidade de agressividade na personalidade de cada pessoa. Estamos interessados em comparar os pares de gêmeos para ver se o gêmeo primogênito tende a ser mais agressivo do que o outro. Os resultados estão a seguir, onde o escore mais alto indica mais agressividade. Teste, por meio dos sinais, ao nível de 5% de significância. **R. $p = 0,2744$**

Filho	Nível de agressividade											
Primogênito:	86	71	77	68	91	72	77	91	70	71	88	87
Segundo gêmeo:	88	77	76	64	96	72	65	90	65	80	81	72

3.2 Teste de McNemar

O Exemplo 3.3 ilustra a necessidade de desenvolvimento de um teste específico para a **situação de dados pareados em que a resposta é dicotômica**.

Exemplo 3.3. Suponhamos que dois patologistas examinaram, separadamente, o material de 100 tumores e os classificaram como benignos ou malignos. A questão de interesse é saber se os patologistas diferem nos seus critérios de decisão.

Tabela 3.1: Classificação de dois patologistas (A e B) quanto à malignidade de tumores.

Diagnóstico de B	Diagnóstico de A		Total
	Malignos	Benignos	
Malignos	9	1	10
Benignos	9	81	90
Total	18	82	100

É importante observar que a unidade de análise aqui é o *tumor*, avaliado por dois patologistas. Embora tenham sido feitas 200 análises, o total de tumores é, na realidade, apenas 100.

Além disto, alguns tumores serão claramente mais malignos do que outros e, portanto, a hipótese fundamental na construção do teste de probabilidade constante de malignidade não é razoável aqui. Isto explica a necessidade de desenvolvimento de teste específico, isto é, para dados pareados.

Considerando que os dados não são ordinais com duas categorias: “0” ou “1”, isto é, cada X_i ou é 0 ou é 1 e Y_i ou é 0 ou é 1, então a questão que surge é: podemos detectar uma diferença entre a probabilidade de (0,1) e a probabilidade (1,0)? Esta questão surge quando estamos em uma situação em que queremos testar no mesmo sujeito **a mudança de estado**: Antes-Depois ou até mesmo se houve diferenças em relação aos percentuais em grupos dados como pareados.

Os dados a serem analisados no processo de comparação podem ser resumidos no formato da Tabela 3.2.

Tabela 3.2: Apresentação de dados obtidos em uma classificação de dados pareados.

Antes	Depois		Total
	Sucesso	Fracasso	
Sucesso	a	b	n_1
Fracasso	c	d	n_2
Total	m_1	m_2	n

Se p_1 e p_2 são as probabilidades de sucesso nos grupos controle e tratamento, respectivamente, a hipótese de interesse é:

$$H_0 : p_1 = p_2$$

$$H_a : p_1 \neq p_2$$

No planejamento tipo Antes-Depois como $b+c$ representa o número total de elementos que acusaram modificação, a perspectiva, sob H_0 , seria que $\frac{b+c}{2}$ acusassem modificações em um sentido, e $\frac{b+c}{2}$ acusassem modificações em outro sentido. Assim, em cada direção é esperado $\frac{b+c}{2}$ modificações.

Em outras palavras, $\frac{b+c}{2}$ é a frequência esperada, sob H_0 , tanto na célula b como da c. As hipóteses são:

$$H_0 : \text{As mudanças não são significativas (P(A,B) = P(B,A))}$$

$$H_a : \text{As mudanças são significativas. (P(A,B) < P(B,A) ou P(A,B) > P(B,A))}$$

Consideremos agora dois casos:

- $n = b + c \leq 20$: Neste caso empregamos o teste binomial com $p_o = 1/2$, identificando a estatística B daquele teste com a frequência observada, ou seja, $B = b$. Verificamos que nestas condições, este nada mais é do que um “teste do sinal”, condicional, se considerarmos os pares (A,B) como positivos e (B,A) como negativos.

$$P[B \leq \text{Bin}(\alpha; n; p_o)] = \alpha$$

Observação: O p -valor aqui é correspondente ao teste unilateral, sendo o dobro se fosse um teste bilateral.

Exemplo 3.4. Dois supermercados disputam a preferência dos consumidores de uma cidade. Um deles (A), para aumentar o seu número de fregueses, lança a campanha publicitária, através de concursos com vários brindes. A Tabela 3.3 apresenta os resultados.

Tabela 3.3: Opiniões de 100 fregueses dos supermercados A e B antes e depois de uma campanha publicitária.

Antes da campanha	Depois da campanha		Total
	A	B	
A	37	3	40
B	13	47	60
Total	50	50	100

A campanha foi eficiente a 5% de significância?

Solução:

H_o : A campanha não foi eficiente ou $P(A,B) \geq P(B,A)$

H_a : A campanha foi eficiente ou $P(A,B) < P(B,A)$.

De acordo com o teste dos sinais, $n = 13 + 3 = 16 < 20$ ($P[B \leq \text{Bin}(\alpha; n; p_o)] = \alpha$); ver Tabela Binomial (5%; 16; 0,5), já que $P(X \leq b = 3) = 0,0106$ (unilateral) que é inferior a 5%, logo rejeitamos H_0 .

- $n = b + c > 20$; Neste caso a Estatística Teste será a de McNemar.

Exercício 3.4. A Tabela a seguir apresenta dados, de pesquisadores de ciências políticas, coletados de casais que responderam sobre as suas aprovações ou não de um dos senadores que eles votaram. As frequências das células representam o número de pares (marido e mulher) que aprovam ou não suas escolhas. Pode-se dizer, em nível de 5% que houve diferença significativa? Em outras palavras, a proporção de esposas que responderam sim é a mesma que a proporção de maridos que responderam sim? **R. $p = 0,1509$**

Tabela 3.4: Aprovações do marido e da esposa.

Aprovação do Marido	Aprovação da Esposa		Total
	Sim	Não	
Sim	20	5	25
Não	10	10	20
Total	30	15	45

A Estatística Teste de McNemar (B/C):

A estatística teste para o teste de McNemar, sem uso de correção, segue uma Qui-quadrado com 1 g.l. e é usualmente escrita como: $\sum \frac{(o_i - e_i)^2}{e_i}$ tal que $e_i = \frac{b+c}{2}$:

$$\chi_{McN}^2 = \frac{\left(b - \frac{b+c}{2}\right)^2}{\frac{b+c}{2}} + \frac{\left(c - \frac{b+c}{2}\right)^2}{\frac{b+c}{2}} = \frac{(b-c)^2}{b+c} \sim \chi_{(1);\alpha\%}^2 \quad (3.2)$$

A Estatística Teste com correção de Yates (1934):

$$\chi_{McN}^2 = \frac{(|b-c| - 1)^2}{b+c} \sim \chi_{(1);\alpha\%}^2. \quad (3.3)$$

A hipótese nula deverá ser rejeitada quando $\chi_{McN}^2 > \chi_{(1);\alpha\%}^2$.

Exemplo 3.5. Amigdalectomia e doença de Hodgkin.

Johnson & Johnson (1972), ao analisarem retrospectivamente a história clínica de pacientes de doença de Hodgkin, um tipo de câncer no tecido linfóide, não encontraram evidências que sustentassem a hipótese de que a amigdalectomia aumenta a suscetibilidade à doença, pela remoção da "barreira linfática" representada pela amígdala. Neste estudo, usaram um planejamento do tipo caso-controle pareado. A conclusão relatada no artigo foi feita, entretanto, com uma análise apropriada para dados provenientes de grupos independentes. Logo após a publicação, vários autores observaram a impropriedade da análise e, utilizando as informações do artigo, fizeram a análise correta, que é baseada nos dados da Tabela 3.5.

Tabela 3.5: Distribuição de pacientes com e sem doença de Hodgkin em um estudo caso-controle pareado segundo à amigdalectomia.

Doença de Hodgkin	Controle		Total
	Operados	Não operados	
Operados	26	15	41
Não operados	7	37	44
Total	33	52	85

H_0 : Não há associação entre amigdalectomia e doença de Hodgkin.

Importante: Observe a disposição dos dados na tabela quando os dados são pareados.

O valor da estatística do teste de McNemar é:

$$\chi_{McN}^2 = \frac{(|15 - 7| - 1)^2}{15 + 7} = 2,23 \text{ (p-valor : 0,1356)}.$$

Este valor deve ser comparado com 3,84 para um nível de significância de 5%. Ou seja, com uma confiança de 95% acreditamos que não há associação entre a doença de Hodgkin e a amigdalectomia. Esta conclusão não está de acordo com trabalhos anteriores, como o de Vianna *et al.* (1971).

```
x=matrix(c(26,7,15,37),nr=2)
?mcnemar.test
mcnemar.test(x)
```

Exercício 3.5. Antes de um debate televisionado nacionalmente entre dois candidatos a presidência, selecionou-se uma amostra aleatória de 100 pessoas e estas relataram suas escolhas como segue na Tabela 3.6. 84 pessoas foram favoráveis ao candidato Democrático, e os 16 restantes ao Republicano. Após o debate as mesmas 100 pessoas expressaram suas preferências novamente. Das pessoas que foram inicialmente favoráveis ao Democrático, exatamente 1/4 deles mudaram de ideia, e também 1/4 das pessoas que inicialmente eram favoráveis ao Republicano passaram para o Democrático. Utilize o teste de McNemar para verificar ao nível de 5% se as intenções de votos da população não se alteraram com o debate.

Tabela 3.6: Resultado de um debate televisionado.

Antes	Depois		Total
	Democrático	Republicano	
Democrático	63	21	84
Republicano	4	12	16
Total	67	33	100

R. $\chi_{McN}^2 = 10,24$ ($p = 0,00137$).

```
x=matrix(c(63,4,21,12),nr=2)
mcnemar.test(x)
```

Exercício 3.6. *Uma agência de publicidade deseja determinar a eficiência de uma campanha que está fazendo para induzir os eleitores a aprovar a construção de calçadas numa área suburbana. Duzentos moradores foram escolhidos aleatoriamente e entrevistados sobre o assunto, antes do início da campanha, e foram novamente entrevistados três semanas após o início da campanha, com os seguintes resultados.*

Tabela 3.7: *Resultado de um campanha.*

Antes	Depois		Total
	Pró	Contra	
Pró	20	30	50
Contra	90	60	150
Total	110	90	200

Pode-se dizer, ao nível de 1% que a campanha foi eficiente?

R. $\chi^2_{McN} = 29,01$ ($p = 0$).

Assim como o teste do sinal nas formas bilateral e unilateral, o teste de McNemar poderia ser apresentado nas duas formas. A maneira mais fácil de realizar um teste de McNemar unilateral é exatamente usando o teste do sinal unilateral. O teste de McNemar e suas variações são discutidas por Bennett e Underwood (1970), Ury (1975), Mantel e Fleiss (1975) e McKinlay (1975).

3.3 Teste Kappa - Concordância ou de replicabilidade

O coeficiente de Kappa é utilizado para verificar a concordância entre os diagnósticos de dois especialistas. Observe a Tabela (3.8) abaixo:

Tabela 3.8: *Concordância entre o diagnóstico de dois especialistas.*

Diagnóstico Especialista 2	Diagnóstico Especialista 1		Total
	Presente (+)	Ausente (-)	
Presente (+)	a (++)	b (+-)	$n_1 = a + b$
Ausente (-)	c (-+)	d (-)	$n_2 = c + d$
Total	$m_1 = a + c$	$m_2 = b + d$	n

Cálculos auxiliares:

- Proporção de concordância observada:

$$p_o = \frac{a + d}{n}$$

- Proporção de concordância esperada ou casual:

$$p_e = \frac{n_1 \cdot m_1 + n_2 \cdot m_2}{n^2}$$

- Coeficiente:

$$\begin{aligned} Kappa &= \frac{p_o - p_e}{1 - p_e} \\ &= \frac{2(ad - bc)}{n_1 \cdot m_2 + n_2 \cdot m_1} \end{aligned} \quad (3.4)$$

Classificação segundo Byrt T. (1993):

Kappa	Concordância
$k \leq 0,2$	péssima
$0,2 < k \leq 0,4$	razoável
$0,4 < k \leq 0,6$	boa
$0,6 < k \leq 0,8$	muito boa
$0,8 < k \leq 1$	excelente

O teste pode ser generalizado para um número de linhas maior do que 2:

		Julgador A				Total
		Categoria				
	Categoria	1	2	...	r	
Julgador B	1	n_{11}	n_{12}	...	n_{1r}	$\mathbf{n}_{1.}$
	2	n_{21}	n_{22}	...	n_{2r}	$\mathbf{n}_{2.}$
	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
	r	n_{r1}	n_{r2}		n_{rr}	$\mathbf{n}_{r.}$
	Total	$\mathbf{n}_{.1}$	$\mathbf{n}_{.2}$		$\mathbf{n}_{.r}$	\mathbf{n}

$$p_o = \frac{1}{n} \sum_{i=1}^r n_{ii}$$

e

$$p_e = \frac{1}{n^2} \sum_{i=1}^r (n_{i.}) (n_{.i}),$$

logo, o coeficiente de Kappa ($-1 \leq k \leq +1$) será:

$$k = \frac{n \sum_{i=1}^r n_{ii} - \sum_{i=1}^r [t_{l_i} t_{c_i}]}{n^2 - \sum_{i=1}^r [t_{l_i} t_{c_i}]}, \quad (3.5)$$

onde:

r : n° de linhas da matriz;

n : total de observações;

x_{ii} : n° de observações na linha i e coluna i ;

t_{l_i} : total marginal da linha i ;

t_{c_i} : total marginal da coluna i .

Exemplo 3.6. Em uma determinada experiência, foi avaliado o grau de lesão do tecido hepático, em 20 cobaias às quais foi administrada uma certa substância tóxica. Os resultados dos exames efetuados por dois patologistas foram o seguinte:

Tabela 3.9: *Teste Kappa para a concordância entre patologistas quanto ao grau de lesão do tecido hepático.*

Cobaia	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Patologista 1	+	+	+	+	+	+	-	+	+	+	+	+	+	-	-	+	-	+	+	-
Patologista 2	+	+	-	+	+	+	-	-	+	+	-	+	+	-	-	-	-	-	+	-

+ : presença de lesão hepática

- : ausência de lesão hepática.

Com base nesses resultados, é possível construir a tabela abaixo:

Tabela 3.10: *Concordância entre o diagnóstico de dois especialistas.*

Diagnóstico	Diagnóstico Patologista 1		Total
Patologista 2	Presente (+)	Ausente (-)	
Presente (+)	10	0	10
Ausente (-)	5	5	10
Total	15	5	20

$$\text{Hipóteses: } \begin{cases} H_0 : \text{Não concordam,} \\ H_a : \text{Concordam.} \end{cases}$$

Então tem-se:

$$p_o = \frac{a + d}{n} = \frac{10 + 5}{20} = 0,75$$

$$p_e = \frac{n_1 \cdot m_1 + n_2 \cdot m_2}{n^2} = \frac{15 \cdot 10 + 5 \cdot 10}{20^2} = \frac{200}{400} = 0,5$$

$$Kappa = \frac{p_o - p_e}{1 - p_e} = \frac{0,75 - 0,5}{1 - 0,5} = 0,5 \quad (Z = 2,582; p = 0,0049).$$

Conclui-se que existe uma concordância apenas moderada entre os diagnósticos dos patologistas.

Exercício 3.7. Em uma determinada situação hospitalar, foi avaliado o nível de infecção pulmonar em 10 pacientes analisados por dois especialistas distintos. Com base nos resultados apresentados abaixo, verifique se existe concordância (classifique) entre os diagnósticos dos especialistas (Justifique). +: infecção pulmonar

Paciente	1	2	3	4	5	6	7	8	9	10
Especialista 1	+	+	+	-	+	-	-	-	+	-
Especialista 2	+	-	-	-	+	-	-	+	-	-

R. K = 0,2; p = 0,245.

Exercício 3.8. Uma clínica investigou a concordância entre dois de seus patologistas e os resultados são apresentados a seguir.

Resultados de exames de dois patologistas quanto ao indício de câncer de estômago em 40 cobaias (ratos de laboratório).

Cobaia	1	2	3	4	5	6	7	8	9	10
Patologista 1	+	-	-	-	+	+	-	+	-	-
Patologista 2	-	-	-	+	+	+	-	-	+	-

	11	12	13	14	15	16	17	18	19	20
+	-	+	+	+	+	+	+	+	+	-
-	+	+	-	+	-	+	-	+	+	-

Cobaia	21	22	23	24	25	26	27	28	29	30
Patologista 1	+	-	-	-	-	+	+	+	-	+
Patologista 2	+	+	-	+	+	+	-	-	+	+

	31	32	33	34	35	36	37	38	39	40
-	-	+	-	-	-	+	+	+	+	-
-	+	+	-	-	-	-	-	+	+	-

Ao nível de 5% de significância, qual o resultado? qual o grau de concordância?

R. $K = 0,1022$; $p = 0,2579$.

Exercício 3.9. A seguir tem-se os Resultados de medidas de dois questionários distintos aplicados em pacientes de uma clínica psiquiátrica. Em nível de 5%, verifique se as ferramentas de diagnóstico clínico são duas-a-duas igualmente eficazes e também quantifique sua concordância.

Questionário	Score de Depressão : Se > 7 é doente															
HAD	7	8	6	3	1	9	4	7	6	5	4	6	8	12	4	
BECK ₁	4	7	1	4	2	11	6	8	11	11	0	5	6	10	1	
BECK ₂	10	6	12	8	6	15	5	1	20	18	8	6	5	13	0	

R. $K_{HAD \times Beck_1} = 0,21$; $K_{HAD \times Beck_2} = 0,03$; $K_{Beck_1 \times Beck_2} = 0,35$.

3.4 Teste de Wilcoxon

Trata-se de um teste não-paramétrico para comparar **dois tratamentos quando os dados são obtidos através do esquema de pareamento em respostas contínuas**. A prova de Wilcoxon, de 1945, (Químico e Estatístico americano, Frank Wilcoxon - 1892-1965) avalia a grandeza das diferenças quando comparados postos de observações. Dada a grandeza das diferenças observadas, atribui-se maior valor para a maior diferença encontrada, diminuindo este valor de acordo com as menores diferenças existentes.

As hipóteses a serem testadas são:

$$\text{Hipóteses: } \begin{cases} H_0 : \tilde{\mu}_1 = \tilde{\mu}_2 : \text{“Antes = Depois”}, \\ H_a : \tilde{\mu}_1 \neq \tilde{\mu}_2; \end{cases}$$

ou considerando os totais dos ranks, T_i :

$$\text{Hipóteses: } \begin{cases} H_0 : T_1 = T_2, \\ H_a : T_1 \neq T_2. \end{cases}$$

Procedimento

1. Calcular d_i : diferença entre as observações

$$d_i = x_i^{(2)} - x_i^{(1)}$$

2. Ignorar os sinais e atribuir postos
3. Calcular a soma dos postos (T_+ e T_-).
4. Obter o valor da estatística calculada

$$T_c = \min\{|T_-|; |T_+|\} \quad (3.6)$$

5. Obter o valor da estatística tabelada : $T_{\alpha;n}$ (Anexo - Tabela de Valores críticos da distribuição T de Wilcoxon);

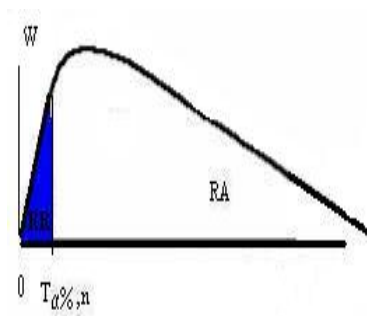


Figura 3.1: *Distribuição de Wilcoxon.*

6. Concluir pela rejeição de H_0 se $T_c \leq T_{\alpha;n}$.

Pequenas Amostras ($n \leq 20$)

No confronto de dois grupos pareados quando desejamos identificar se existe diferença significativa entre os mesmos quanto às medidas encontradas, empregamos o teste de Wilcoxon quando a variação dos valores apresenta-se de forma acentuada.

A aplicação deste teste pressupõe que as duas amostras sejam casualizadas e dependentes, e que as variáveis em confronto sejam contínuas.

A metodologia do teste consiste em se proceder à ordenação dos valores das amostras e, posteriormente, atribuir aos mesmos seus “ranks”. Em seguida, obtém-se os totais do *ranks* da amostra de menor tamanho, consultando-se a tabela do referido teste.

Exemplo 3.7. *Foi realizado um ensaio clínico em que foram utilizadas duas drogas A e B. Com a droga A foram tratados oito pacientes e oito com a droga B. Os níveis de anticorpos correspondentes estão apresentados da Tabela 3.11:*

Tabela 3.11: *Níveis de anticorpos.*

Tratamento		d=B-A	rank de d
A	B	-	-
7,4	9,1	1,7	5
12,3	9,7	-2,6	6
11,8	19,2	7,4	8
16,4	22,4	6,0	7
1,9	1,5	-0,4	1
3,0	3,5	0,5	2,5
6,8	7,8	1,0	4
20,4	19,9	-0,5	2,5

Observação: No R o comando usado é **rank(d)**.

$H_0: \tilde{\mu}_A = \tilde{\mu}_B$ vs $H_a: \tilde{\mu}_A \neq \tilde{\mu}_B$.

Tem-se:

Soma dos ranks (-) : $|T_-|$: 9,5

Soma dos ranks (+): $|T_+|$: 26,5

assim

$$T_c = \min\{9, 5; 26, 5\} = 9, 5$$

consultando-se a tabela de Wilcoxon, verifica-se que $T_{\alpha\%;n} = T_{5\%;8} = 3$ (bilateral), logo como $T_c = 9,5 > 3 = T_{\alpha\%;n}$ ($p = 0,2620$), não rejeitamos H_0 , ou seja, que os valores comparados, referentes às drogas A e B, não apresentam-se com diferença significativa. Conclui-se, então, que os valores de níveis de anticorpos em relação aos dois tipos de drogas se comportam de forma semelhante.

```
A=c(7.4,12.3,11.8,16.4,1.9,3,6.8,20.4)
B=c(9.1,9.7,19.2,22.4,1.5,3.5,7.8,19.9)
?wilcox.test
wilcox.test(A,B,paired=TRUE)
```

Exercício 3.10. Um estudo foi conduzido para investigar o uso de oxigenação extracorpórea (ECMO), um sistema mecânico para oxigenar o sangue no tratamento de recém-nascidos com falha respiratória neonatal. Imagina-se que o uso desse procedimento possa reduzir o rendimento do ventrículo esquerdo de uma criança, diminuindo a quantidade de sangue bombeado ao corpo. Assim, há o interesse de comparar a dimensão do ventrículo esquerdo antes e durante o uso do ECMO. A Tabela 3.12 mostra os dados relevantes de uma amostra de 15 bebês que sofrem de dificuldade respiratória. Teste ao nível de 5% de que a diferença mediana é nula, isto é, se a dimensão do ventrículo esquerdo antes e durante o uso do ECMO é a mesma. **R. $T=18,5 > T_{5\%;11} = 10$; p -valor no R = 0,2463 - “Não rejeita-se H_0 ”.**

Tabela 3.12: Dimensão do Ventrículo Esquerdo para uma amostra de bebês que sofrem de falha respiratória.

Tratamento		d=Durante-Antes	rank de d
Antes	Durante	-	-
1,6	1,6		
2,0	2,0		
1,2	1,2		
1,6	1,6		
1,6	1,5		
1,7	1,6		
1,6	1,5		
1,6	1,7		
1,6	1,4		
1,7	1,5		
1,0	1,3		
1,5	1,8		
1,5	1,8		
1,4	1,8		
1,5	2,0		

Observação: Uma vez que uma diferença cujo resultado **0** não é ordenada e existem quatro delas no conjunto de dados, o tamanho da amostra será reduzido para $n = 11$.

Grandes Amostras ($n > 20$)

Nos casos de grandes amostras, estas apresentam valores com distribuição normal e, portanto, a comparação das medidas será realizada através de determinação de um valor W^* .

São enunciadas naturalmente as hipóteses. H_0 será rejeitada se o valor de W^* for maior ou igual a Z_α para um nível de significância, α , pré-determinado.

Assim, para a comparação de conjuntos de medidas, provenientes de duas amostras, será necessário utilizar a expressão de W^* :

$$W^* = \frac{T_c - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \sim N(0,1) \quad (3.7)$$

em que:

n : número de elementos da menor amostra;

$E(T) = \frac{n(n+1)}{4}$ e $Var(T) = \frac{n(n+1)(2n+1)}{24}$ e $N(0,1)$: Normal Padrão.

Exemplo 3.8. Evolução do tratamento com tianeptina.

A tianeptina é um fármaco antidepressivo do grupo dos tricíclicos. Sua ação antidepressiva potencial foi demonstrada em estudos pré-clínicos através de testes em animais. Rocha (1995) relata os resultados de um ensaio realizado com o objetivo de investigar o efeito da tianeptina. Participam deste ensaio pacientes de Belo Horizonte, Rio de Janeiro e Campinas. Sucintamente, o ensaio consistiu em administrar a droga e observar a evolução da droga, e quantificar a depressão através da escala de Montgomery-Asberg (MADRS), em que os valores maiores indicam maior gravidade de depressão. O escore foi obtido para cada paciente antes e depois do tratamento, 42 dias após o início do ensaio. Pelo planejamento adotado, os pacientes não diferiam em termos de depressão no início do ensaio. Assim, uma evidência sobre o efeito da tianeptina pode ser verificada comparando-se os escores antes da droga e ao fim de 42 dias. Os dados são apresentados na Tabela 3.13:

Tabela 3.13: *Tratamento com tianeptina.*

Tianeptina					
Primeiro dia (m=21)	Rank	Último dia (n=20)	Rank	d_i	Rank $ d_i $
24	14	6	2,5	-18	12,5
46	39	33	28,5	-13	8
26	17	21	11,5	-5	4,5
44	38	26	17	-18	12,5
27	20,5	10	7	-17	10,5
34	32	29	22,5	-5	4,5
33	28,5	33	28,5	0	*
25	15,5	29	22,5	+4	2
35	34	37	36	+2	1
30	24	15	8,5	-15	9
38	37,5	2	1	-36	19
38	37,5	21	11,5	-17	10,5
31	25	7	4	-24	16,5
27	20,5	47	40	+20	14
34	32	*	*	*	*
36	35	15	8,5	-21	15
18	10	8	5	-10	7
25	15,5	22	13	-3	3
34	32	6	2,5	-28	17
33	28,5	9	6	-24	16,5
32	26	26	17	-6	6
Soma	-	562	293	-	$T_- = 172$
de Ranks:					$T_+ = 23$
Mediana:	33	21	-	-	-

Sendo assim temos: $|T_-| = 172$ e $|T_+| = 23$, então

$$T_c = \min\{172; 23\} = 23$$

De acordo com a Tabela de Wilcoxon (bilateral),

$$T_{\alpha;n} = T_{5\%;20} = 52$$

Como $23 < 52$, rejeitamos H_0 . logo existem evidências ao nível de 5% de significância de que o tratamento teve efeito.

Supondo n grande teremos:

$$W^* = \frac{23 - \frac{19(19+1)}{4}}{\sqrt{\frac{19(19+1)(2 \cdot 19 + 1)}{24}}} = \frac{23 - 95}{\sqrt{617,5}} = -2,897 \left(\frac{\text{p-valor}}{2} = 0,001866 \right).$$

Considerando que o valor de $W^* = -2,897$ está na área de rejeição de H_0 , para o valor de $Z_{5\%} = \pm 1,96$, rejeitamos H_0 , ou seja, os valores dos dois conjuntos apresentam diferenças significativas ($p = 2 \times 0,001866 = 0,0037$).

Observação: O default do R é $p = 0,002072$, pois o teste é bilateral.

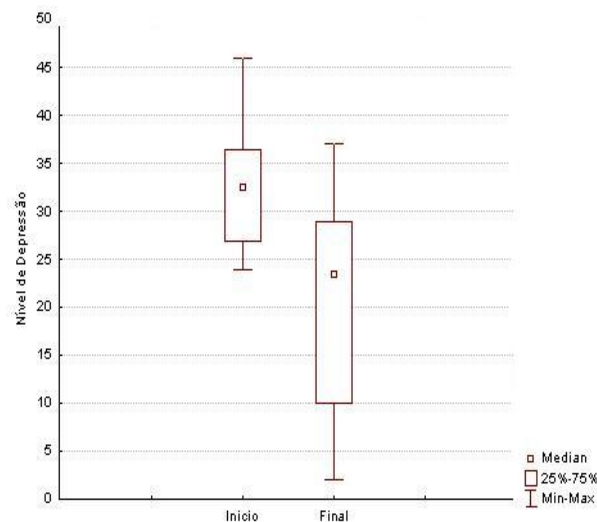


Figura 3.2: Tratamento com Tianeptina.

Observação: Observando os ranks totais (Antes = 263 e Depois = 143) ou as medianas (Antes = 32,5 e Depois = 23,5), podemos concluir que a tianeptina reduziu os níveis de depressão. Para quantificar esta redução em termos de média, sugere-se utilizar um procedimento inferencial, por meio de técnicas estatísticas mais sofisticadas como, por exemplo, reamostragem via Bootstrap ou geração de amostra via inferência Bayesiana.

```
dia.inicial=c(24,46,26,44,27,34,33,25,35,30,38,38,31,27,36,18,25,34,33,32)
dia.final=c(6,33,21,26,10,29,33,29,37,15,2,21,7,47,15,8,22,6,9,26)
wilcox.test(dia.inicial,dia.final,paired=TRUE)
```

Exercício 3.11. Doze casais de gêmeos idênticos foram submetidos a testes psicológicos para medir de algum modo a quantidade de agressividade na personalidade de cada pessoa. Estamos interessados em comparar os pares de gêmeos para ver se o gêmeo primogênito tende a ser mais agressivo do que o outro. Os resultados estão a seguir, onde o escore mais alto indica mais agressividade. Teste ao nível de 5% de significância.

Filho	Nível de agressividade											
Primogênito:	86	71	77	68	91	72	77	91	70	71	88	87
Segundo gêmeo:	88	77	76	64	96	72	65	90	65	80	81	72

R. T = 24,5, p = 0,4765 (Bilateral) e p = 0,2382 (Unilateral)

```
g1=c(86,71,77,68,91,72,77,91,70,71,88,87)
g2=c(88,77,76,64,96,72,65,90,65,80,81,72)
wilcox.test(g1,g2,paired=TRUE)
wilcox.test(g1,g2,paired=TRUE,alternative='greater')
```

Exercício 3.12. Em um teste na água do mar, uma bola enrijecida (endurecida) é prensada em um material testado sob uma resistência padrão. O diâmetro do recorte (entrada, dente, reentrância) esférica é então medida. Duas bolas de aço são disponíveis (uma de cada um dos dois fabricantes) e suas performances serão comparadas em 22 pedaços de material. Cada pedaço do material será testado duas vezes, uma vez com cada bola. Os dados estão logo a seguir. Teste ao nível de 5% a hipótese de igualdade de tratamentos.

Diâmetro X	73	43	47	53	58	47	52	38	61	56	56	34	55	65	75
Diâmetro Y	51	41	43	41	47	32	24	43	53	52	57	44	57	40	68
Diâmetro X	37	45	50	60	55	58	62	72							
Diâmetro Y	41	42	50	45	52	55	48	62							

R. T = 25,5; Z = 2,20; p = 0,028.

Exercício 3.13. Avalie o resultado de uma dieta de 5 semanas ao nível de 5% de significância.

Peso Inicial	202	189	149	186	149	200	220	190	164	161	162	171	193
Peso Final	204	177	154	169	140	200	214	189	167	150	155	172	183
Peso Inicial	163	187	178	218	181	140	168						
Peso Final	158	184	192	210	166	143	164						

R. T = 32, Z = 1,59, p = 0,12.

Exercício 3.14. Um experimento tem como objetivo investigar o uso do medicamento amiloride como terapia para os pacientes com fibrose cística. Acredita-se que esse medicamento possa auxiliar a melhorar o fluxo de ar nos pulmões, ao atrasar a perda da função pulmonar frequentemente associada com a doença. A capacidade vital forçada (FVC-forced vital capacity), é o volume de ar que uma pessoa pode expelir dos pulmões em 6 segundo; Gostaríamos de comparar a redução da FVC ocorrida em um período de 25 semanas de aplicação do medicamento com a redução ocorrida nos mesmos pacientes em período de tempo similar durante o tratamento com placebo. Teste ao nível de 5% de significância utilizando o teste dos sinais. **R. T = 19, Z = -2,10, p = 0,035** (o

Tabela 3.14: Redução na FVC (ml)

Grupo	Redução na FVC (ml)														
Placebo	224	80	75	541	74	85	293	-23	525	-38	508	225	525	1023	
Medicamentoso	213	95	33	440	-32	-28	445	-178	367	140	323	10	65	343	

medicamento reduz a perda da função pulmonar).

Exercício 3.15. Em uma determinada situação hospitalar, foi avaliado o nível de infecção pulmonar em pacientes analisados por dois especialistas distintos. Com base nos resultados apresentados na Tabela abaixo, verifique se existe concordância entre os diagnósticos dos especialistas (Justifique e Classifique). **R. Kappa = 0,4, Z = 1,83, p = 0,0339** (concordância leve).

Tabela 3.15: Diagnóstico de câncer.

Paciente	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Esp.1	+	+	+	-	+	-	-	-	+	-	-	-	+	+	-	+	-	-	+	+
Esp.2	+	-	-	-	+	-	-	+	-	-	-	-	-	+	-	+	-	+	+	+

Exercício 3.16. Um estudo foi conduzido para investigar o uso de oxigenação extracorpórea (ECMO) - um sistema mecânico para oxigenar o sangue - no tratamento de recém-nascidos com falha respiratória neonatal. Imagina-se que o uso desse procedimento possa reduzir o rendimento do ventrículo esquerdo de uma criança, diminuindo a quantidade de sangue bombeado no corpo. Teste ao nível de 5% se a dimensão do ventrículo esquerdo antes e durante o uso do ECMO é a mesma.

LVD Antes	1,6	2,0	1,2	1,6	1,6	1,7	1,6	1,6	1,6	1,7	1,0	1,5	1,5	1,4	1,5
LVD Durante	1,6	2,0	1,2	1,6	1,5	1,6	1,5	1,7	1,4	1,5	1,3	1,8	1,8	1,8	2,0

R. p = 0,213; com aproximação normal p = 0,198.

```
antes=c(1.6,2,1.2,1.6,1.6,1.7,1.6,1.6,1.6,1.7,1,1.5,1.5,1.4,1.5)
depois=c(1.6,2,1.2,1.6,1.5,1.6,1.5,1.7,1.4,1.5,1.3,1.8,1.8,1.8,2)
wilcox.test(antes,depois,paired=TRUE)
```

```
boxplot(Ozone~Month,data=airquality,xlab='Mês',ylab='Nível de Ozônio')
wilcox.test(Ozone~Month,data=airquality,subset=Month%in%c(5,7))
# Compara os grupos: Mês 5 e 7 com relação ao nível de Ozônio
```


Capítulo 4

Testes aplicáveis a duas amostras - Dados independentes

4.1 Teste Qui-quadrado

Neste tópico estudaremos a **relação entre duas variáveis categóricas** (ou qualitativas), onde as observações podem ser classificadas em uma das várias categorias (níveis ou células) mutuamente exclusivas.

O problema de mensuração do grau de associação entre dois conjuntos de escores é de caráter bem diferente do teste da simples existência de uma associação em determinada população. Naturalmente, há interesse em avaliar o grau de associação entre dois conjuntos de escores referentes a um grupo de indivíduos. Mas é talvez de muito maior interesse podermos afirmar que determinada associação observada em uma amostra de escores indica, ou não, probabilidade de associação entre as variáveis na população da qual se extraiu a amostra (Siegel, 1956).

4.1.1 Tabelas de contingência

Segundo o dicionário Webster o termo “contingency” significa *the quality or state of having a close connection or relationship*. O termo “contingency table” se refere ao fato de que as tabelas construídas são usadas para testar a existência de relações entre duas variáveis. Observamos que infelizmente este termo em português não tem o mesmo significado. Segundo o Aurélio, “contingência” significa qualidade do que é contingente, incerteza se uma coisa vai acontecer ou não e “contingente” significa que pode ou não suceder, eventual, incerto. Logo, em estatística, quando nos referimos a palavra “contingência” estamos tomando o mesmo significado da língua inglesa.

O teste chi-quadrado (χ^2) pode ser usado para avaliar a relação entre duas variáveis qualitativas. Este teste é um teste não-paramétrico, que é muito útil, pois não precisa da suposição de normalidade das variáveis para analisar o grau de associação entre as duas variáveis, porém este teste é menos poderoso que o teste paramétrico.

Dentro deste teste devemos distinguir dois tipos de testes: o de **independência** e o de **heterogeneidade**. Vejamos quando estamos tratando de um teste de independência e quando estamos tratando com um teste de homogeneidade, através de exemplos.

4.1.2 Teste de independência e de homogeneidade

Suponha que uma educadora quer analisar se existe relação entre a participação ativa dos pais dos alunos nos deveres extra-escolares e o desempenho dos alunos em Matemática. Suponha que ela categoriza o desempenho dos alunos em três grupos: baixo, médio, alto e, do mesmo modo,

categoriza a participação dos pais em dois grupos: *participação ativa*, *participação fraca*. Suponha que ela deseje trabalhar com 300 crianças.

Neste caso a educadora pode delinear sua pesquisa de duas formas:

Caso 1. Selecionar uma amostra de crianças aleatoriamente e examinar em que célula cada uma está alocada, logo o único valor fixo será o total geral que será de 300. Mas os totais de colunas e de linhas serão frutos da pesquisa, portanto, aleatórios, neste caso estamos frente a um teste de independência de variáveis “ou de associação”.

Logo a tabela de contingência será:

Tabela 4.1: *Exemplo de tabela de contingência.*

Participação dos pais	Desempenho do aluno em Matemática			Total
	Baixo	Médio	Alto	
Ativa				Aleatório
Fraca				Aleatório
Total	Aleatório	Aleatório	Aleatório	300

\Rightarrow

Teste de Independência

Porém ela pode fixar o número de alunos de acordo com seu desempenho.

$$\text{Hipóteses : } \begin{cases} H_0 : & \text{Não existe associação entre Participação dos pais e Desempenho do aluno em Matemática;} \\ H_a : & \text{Existe associação entre Participação dos pais e Desempenho do aluno em Matemática;} \end{cases}$$

Caso 2. Pegar uma amostra aleatória de tamanho 100 de cada grupo de alunos, logo os totais das colunas serão fixos, mas os totais das linhas serão aleatórios e assim estaremos frente a um teste de homogeneidade:

Logo a tabela de contingência será:

Tabela 4.2: *Exemplo de tabela de contingência.*

Participação dos pais	Desempenho do aluno em Matemática			Total
	Baixo	Médio	Alto	
Ativa				Aleatório
Fraca				Aleatório
Total	100 (fixo)	100 (fixo)	100 (fixo)	300

\Rightarrow

Teste de Heterogeneidade

$$\text{Hipóteses : } \begin{cases} H_0 : & P_{Baixo} = P_{Medio} = P_{Alto} \\ H_a : & P_i \neq P_j, \text{ para algum } i \neq j \end{cases}$$

Os valores totais, das colunas e das linhas, são chamados de “totais marginais”. **Quando os totais marginais variam livremente, o teste de associação é chamado de independência (ou de associação), e quando um dos conjuntos, linha ou coluna é fixado pelo pesquisador então é chamado de teste de homogeneidade (ou heterogeneidade).** Isso vai depender do pesquisador. No exemplo da educadora, observemos que para ela é muito mais fácil fixar o número de alunos segundo seu desempenho, do que fixar pela participação dos pais, que, a priori será quase impossível.

Teste de Independência

Apresentaremos a lógica do teste com um exemplo bastante simples.

Exemplo 4.1. *Associação entre Gênero e Tipo de comercial de TV.*

Suponha que 125 crianças foram expostas a três tipos de comerciais de TV, sobre cereais para café da manhã. Após a exposição foi solicitado a cada criança para indicar qual dos comerciais ela gostou mais. O que se deseja saber é se a escolha do comercial está relacionado ao gênero da criança: pois suspeita-se de que o gênero pode estar influenciando na escolha do comercial. Os dados podem ser apreciados na Tabela 4.3.

Tabela 4.3: *Número de crianças segundo tipo de comercial escolhido e gênero.*

Gênero	Tipo de comercial			Total
	A	B	C	
Meninos	30	29	16	75
Meninas	12	33	5	50
Total	42	62	21	125

Analisando atentamente a Tabela 4.3, composta por valores absolutos, percebemos:

- a amostra está composta por mais meninos do que meninas,
- nos comerciais A e C o número de meninos é maior do que meninas, e
- no comercial B essa relação se inverte.

Contudo, essa análise fica prejudicada pela composição da amostra, que tem mais meninos do que meninas. Portanto, a primeira coisa a fazer é analisar as estruturas percentuais, mostradas na Tabela 4.4, ou seja retirar a influência da amostragem.

Tabela 4.4: *Porcentagem para os valores absolutos de crianças por tipo de comercial escolhido e gênero apresentados na tabela anterior.*

Gênero	Tipo de comercial			Total
	A	B	C	
Meninos	71%	47%	76%	60%
Meninas	29%	53%	24%	40%
Total	100%	100%	100%	-

Observe cuidadosamente a Tabela 4.4, onde 60% da amostra é composta por meninos. Se a preferência das crianças pelos comerciais independe do gênero, esperaríamos que a estrutura percentual para cada comercial ficasse em torno de 60% para os meninos e 40% para as meninas, desvios grandes destes percentuais estariam mostrando evidências de que existe alguma relação entre essas variáveis. Essa inspeção intuitiva, também, pode ser feita analisando a estrutura dentro de cada gênero como mostra a Tabela 4.5.

Analisando a Tabela 4.5 observamos que as meninas tem uma forte preferência pelo comercial B, enquanto que os meninos se dividem entre o comercial A e B.

Assim, intuitivamente percebemos que existe interferência do gênero na preferência, agora precisamos saber até que ponto essas diferenças se devem ao acaso, ou a existência de associação entre as duas variáveis:

Tabela 4.5: *Porcentagem de crianças por gênero e tipo de comercial escolhido.*

Gênero	Tipo de comercial			Total
	A	B	C	
Meninos	40%	39%	21%	100%
Meninas	24%	66%	10%	100%
Total	33%	50%	17%	-

X: preferência pelo comercial (A, B e C) \rightarrow qualitativa;

Y: gênero (meninos, meninas) \rightarrow qualitativa.

Hipótese nula (H_0): A preferência pelo comercial independe do gênero da criança;

Hipótese alternativa (H_a): A preferência pelo comercial depende do gênero da criança (ou, o gênero interfere na preferência pelo comercial);

Analogamente, temos

$$\text{Hipóteses : } \begin{cases} H_0 : & \text{Independência entre as variáveis;} \\ H_a : & \text{Dependência ou associação entre as variáveis.} \end{cases}$$

Como deveriam ser os valores a serem observados se as variáveis fossem independentes?, ou dito de outra forma, sob a hipótese de nulidade, de independência de variáveis, como deveriam ser os valores a serem observados? A lógica nos diz que esses valores devem estar muito próximos da estrutura percentual global. Esses valores são chamados de valores esperados.

Tabela 4.6:

Valores percentuais					e	Valores absolutos				
Gênero	Tipo de comercial			Total		Gênero	Valores esperados			Total
	A	B	C				A	B	C	
Meninos	60%	60%	60%	60%	Meninos	25	37	13	75	
Meninas	40%	40%	40%	40%	Meninas	17	25	8	50	
Total	42	62	21	125	Total	42	62	21	125	

Cada valor esperado será calculado supondo que a estrutura percentual global se mantenha em cada coluna:

Calculando os valores esperados, sobre a suposição de independência:

Valor esperado menino, comercial A: 60% de 42 = 25,2

Valor esperado menino, comercial B: 60% de 62 = 37,2

...

O mesmo teria acontecido se fixarmos primeiro o comercial:

Valor esperado comercial A, menino: 33,7% de 75 = 25,2

Valor esperado comercial A, menina: 33,7% de 50 = 16,8

...

Tanto faz fixar a linha ou a coluna pois:

$$E_i = \frac{T_l \times T_c}{T} = T_l \times \frac{T_c}{T} = T_c \times \frac{T_l}{T}$$

Por exemplo, calculemos o valor esperado da primeira linha e primeira coluna:

$$E_i = \frac{75 \times 42}{125} = 75 \times \frac{42}{125} = 42 \times \frac{75}{125} = 25,2$$

Assim calculando os valores esperados para todas as células temos:

Tabela 4.7:

Gênero	Tipo de comercial						Total
	A		B		C		
Meninos	$o_i = 30 \quad e_i = 25,2$		$29 \quad 37,2$		$16 \quad 12,6$		75
	$d_i = +4,8$		-8,2		+3,4		
Meninas	$12 \quad 16,8$		$33 \quad 24,8$		$5 \quad 8,4$		50
	-4,8		+8,2		-3,4		
Total	42		62		21		125

Dentro de cada célula, no canto superior esquerdo colocamos o valor observado, no canto superior direito o valor esperado (sob a hipótese de independência) e, na parte inferior, a distância entre o observado e o esperado. Logo, se as variáveis fossem independentes, as distâncias entre os valores observados e esperados deveriam ser muito pequenas, caso contrário haverá indícios de dependência. A pergunta agora é: quando a distância é pequena ou grande? Para isto devemos calcular o valor (χ^2) Qui-quadrado da amostra:

$$\chi_{amostra}^2 = \sum_{i=1}^k \frac{(\text{observado} - \text{esperado})^2}{\text{esperado}} = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i} \quad (4.1)$$

que terá uma distribuição chi-quadrado com graus de liberdade igual ao produto do número de linhas menos um vezes o número de colunas menos um, isto é:

$$\chi_{amostra}^2 \sim \chi_{\nu}^2 \Rightarrow \begin{array}{l} \nu : \text{graus de liberdade} \\ \nu : (n^{\circ} \text{ colunas} - 1) \times (n^{\circ} \text{ linhas} - 1) \end{array}$$

Para testar as hipóteses, temos a estatística teste qui-quadrado com correção de continuidade de Yates:

$$\chi_{cY}^2 = \sum_{i=1}^k \frac{(|O_i - E_i| - \frac{1}{2})^2}{E_i} \sim \chi_{[L-1 \times C-1; (1-\alpha)\%]}^2 \quad (\text{Tabelas } L \times C) \quad (4.2)$$

No nosso exemplo sem a correção de Yates, temos:

$$\begin{aligned} \chi_{amostra}^2 &= \frac{(+4,8)^2}{25,2} + \frac{(-8,2)^2}{37,2} + \frac{(+3,4)^2}{12,6} + \frac{(-4,8)^2}{16,8} + \frac{(+8,2)^2}{24,8} + \frac{(-3,4)^2}{8,4} \\ &= 0,914 + 1,808 + 0,917 + 1,371 + 2,711 + 1,376 \\ &= 9,09818 \quad (\text{p-valor} : 0,0106) \end{aligned}$$

$$\text{onde } v = (2 - 1) \times (3 - 1) = 1 \times 2 = 2$$

Para aceitar ou rejeitar a hipótese devemos conhecer o valor de $\chi_{calculado}^2$ na Tabela Qui-quadrado, com dois graus de liberdade. Para $\alpha = 5\%$, o valor crítico é 5,991, como o valor da amostra é maior que o valor crítico, logo rejeitamos a hipótese nula, concluindo que o gênero interfere na preferência pelos comerciais.

No caso de dispormos de um pacote estatístico, este, via de regra, calcula o p -valor, nesse caso é só comparar esse valor com o nível de significância desejado. No nosso caso o p -valor = 0,01058, ou seja, rejeitamos ao nível de 5% mas não ao nível de 1%.

```
M=matrix(c(30,12,29,33,16,5),nr=2,nc=3)
chisq.test(M) # sem correção de Yates
```

Observação 1: O comando para construção de tabelas cruzadas no *R* é “`xtabs`” e para realizar o teste χ^2 pode-se utilizar o comando “`summay(xtabs())`”.

Observação 2: Quando as variáveis são independentes (H_0), as frequências observadas tendem a ficar perto das esperadas. Neste caso, o valor de χ^2 deve ser pequeno.

Limitações do teste χ^2 :

1. Infelizmente, o teste qui-quadrado não permite concluir como se dá a relação, uma vez que ele testa apenas a hipótese geral de que as duas variáveis são independentes. Examinando a distância entre valor observado e esperado, por exemplo, observamos que as meninas tem uma maior preferência pelo comercial B, porém não podemos concluir nada;
2. Uma outra limitação do teste qui-quadrado é que o valor esperado de células não deve ser menor ou igual a 5 e devido a isso torna vulnerável a estatística. Nesse caso usamos outra estratégia: o teste χ^2 de Yates;
3. Se muitos valores ($>20\%$ ou $>30\%$ dependendo do livro texto) esperados ($e_i \leq 5$) são menores ou iguais a 5 ou se pelo menos $e_i < 1$, então utilizamos o teste exato de Fisher.

Teste de Homogeneidade

Quando testamos independência de variáveis, o pesquisador só controla o tamanho total da amostra, mas os totais para cada coluna e linha são aleatórios.

Exemplo 4.2. No caso do exemplo anterior, os pesquisadores selecionaram aleatoriamente 125 crianças, das quais 75 eram meninos e 50 meninas. Ele não fixou o número de meninos e o número de meninas.

Vejam os um exemplo de teste de homogeneidade. Retomemos o exemplo inicial e suponhamos que a professora fixou o tamanho dentro de cada grupo de alunos e os resultados foram os seguintes:

Tabela 4.8: Número de alunos segundo seu desempenho em Matemática e participação dos pais nas atividades extra-escolares.

Participação dos pais	Desempenho em Matem.			Total
	Baixo	Médio	Alto	
Ativa	5	25	70	100
Fraca	95	75	30	200
Total	100	100	100	300

$$\begin{cases} H_0 : p_{11} = p_{12} = p_{13} \rightarrow \text{igualdade de proporções;} \\ H_a : p_{1i} \neq p_{1j}; \text{ para algum } i \neq j \rightarrow \text{existe pelo menos uma proporção diferente.} \end{cases}$$

A hipótese nula esta testando que a proporção de alunos com baixo desempenho é igual a proporção de alunos médio e igual a proporção de alunos com desempenho alto quando seus pais participam ativamente das atividades extra-escolares, contra a hipótese alternativa que indica que existe pelo menos uma proporção diferente.

O teste é idêntico ao teste de independência.

Exemplo 4.3. *Comparação entre faculdades Pública e Privada.*

A Tabela (4.9) a seguir apresenta os resultados de um teste de conhecimento padronizado aplicado a uma amostra aleatória de estudantes da escola particular de ensino superior e a uma amostra de estudantes selecionados aleatoriamente da escola pública de ensino superior.

Tabela 4.9: *Escore de estudantes de faculdades pública e particular.*

Escola	Escore do teste				Total
	0-275	276-350	351-425	426-500	
Privada	6	14	17	9	46
Pública	30	32	17	3	82
Total	36	46	34	12	128

Para testar a hipótese nula de que a distribuição dos escores do teste é a mesma para estudantes de escolas públicas e privadas, o teste para diferenças de probabilidades é usado. Uma região crítica de tamanho aproximado $\alpha = 5\%$, corresponde a valores de χ^2 maiores que 7,815, obtido da distribuição Qui-quadrado com $(L - 1)(C - 1) = (2 - 1)(4 - 1) = 3$ graus de liberdade.

Pode-se verificar após alguns cálculos utilizando a fórmula de Pearson que $\chi_c^2 = 17,3$ e $\chi_{Yates}^2 = 11,8$.

Como $11,8 > 7,815$, a hipótese nula é rejeitada. De fato, a hipótese nula poderia ter sido rejeitada usando um nível de significância menos que 0,001, assim, $\hat{\alpha} = 0,001$.

Conclusão: Os escores dos testes são distribuídos diferentemente entre os estudantes de faculdade pública e privada.

Observações:

- Neste exemplo, observa-se que os dados possuem uma escala de medida ao menos ordinal, uma escala mais forte do que a escala de medida nominal considerada mais apropriada para o teste usado. Realmente um teste não-paramétrico mais poderoso baseado em ranks - o teste de Mann-Whitney, poderia ter sido usado para a comparação antes do agrupamento numa tabela de contingência. Entretanto, os dados foram suficientes para resultar numa clara decisão reduzida usando este teste, e assim o teste mais poderoso não foi necessário.

```
M=matrix(c(6,30,14,32,17,17,9,3),nr=2,nc=4)
chisq.test(M)
```

Exercício 4.1. *Considere uma amostra aleatória de estudantes de uma certa universidade. Cada estudante foi classificado de acordo com o curso em que está matriculado e se ele se graduou numa escola superior no Estado ou fora do Estado. Os resultados estão na Tabela (4.10).*

Tabela 4.10: *Distribuição de estudantes Fora e Dentro do Estado quanto ao curso escolhido.*

	Curso				Total
	Engenharia	Artes e Ciências	Economia Doméstica	outros	
No Estado	16	14	13	13	56
Fora do Estado	14	6	10	8	38
Total	30	20	23	21	94

Teste a hipótese nula de que a universidade na qual cada estudante está matriculado é independente de ser de Fora ou de Dentro do Estado ao nível de 5% de significância. **R.** $\chi_c^2 = 1,52$, não se rejeita H_0 .

4.1.3 Teste Qui-Quadrado em Tabelas 2x2

Comparar dois grupos através do resultado observado em uma variável dicotômica é um problema comum na pesquisa médica, aparecendo com frequência em todos os tipos de estudos clínicos.

A variável de interesse é a ocorrência de um evento, como o desenvolvimento de uma doença de certo atributo, por exemplo, albinismo. O problema de comparações das probabilidades de ocorrência do evento ou do atributo nos dois grupos (p_1 e p_2) é formulado através das hipóteses:

$$\text{Hipóteses : } \begin{cases} H_0 : p_1 = p_2 & \text{"não há diferença significativa entre os grupos"} \\ H_a : p_1 \neq p_2 & \text{"há diferença significativa entre os grupos"} \end{cases}$$

A Tabela 4.11 apresenta dados genéricos de uma situação envolvendo a comparação de dois grupos e que a resposta de interesse é dicotômica: a ocorrência ou não de um evento.

Tabela 4.11: *Distribuição quanto à ocorrência de um evento.*

Grupo	Ocorrência do Evento		Total
	Sim	Não	
I	a	b	$n_1 = a + b$
II	c	d	$n_2 = c + d$
Total	$m_1 = a + c$	$m_2 = b + d$	$n = n_1 + n_2$

A estatística teste será:

$$\chi_c^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i} \sim \chi_{\alpha\%;1}^2 \quad (4.3)$$

ou, para Tabelas 2×2 :

$$\chi_c^2 = \frac{n(ad - bc)^2}{n_1 n_2 m_1 m_2} \sim \chi_{\alpha\%;1}^2 \quad (\text{Tabelas } 2 \times 2) \quad (4.4)$$

Há restrições para aplicação do χ^2 em tabelas contingência, logo a correção de Yates, F. (1934) é necessária.

Esta correção deve ser feita quando:

- $n < 40$;
- $20 < n < 40$ e $E_i \leq 5$ para algum i ;
- $n > 40$ e $E_i \leq 5$ para algum i ;

Para testar as hipóteses acima citadas temos a estatística teste qui-quadrado com correção de continuidade de Yates, temos

$$\chi_{cY}^2 = \frac{n(|ad - bc| - \frac{n}{2})^2}{n_1 n_2 m_1 m_2} \sim \chi_{\alpha\%;1}^2 \quad (\text{Tabelas } 2 \times 2) \quad (4.5)$$

Exemplo 4.4. *Estudo sobre a associação entre o uso corrente de contraceptivos e o infarto de miocárdio.*

Shapiro *et al.* (1979) observaram os resultados entre pacientes com idade entre 30 e 34 anos e estes são mostrados na Tabela 4.12.

Entre os casos, ou seja, entre as pacientes que tiveram um infarto do miocárdio, a proporção de uso recente de contraceptivos é $p_1 = \frac{9}{21} = 0,43$ e entre os controles, $p_2 = \frac{33}{423} = 0,08$. A diferença

Tabela 4.12: Distribuição de uso de contraceptivo oral segundo grupo que sofreu ou não infarto do miocárdio.

Grupo	Uso recente		Total
	Sim	Não	
Casos	9 (1,99)	12	21
Controles	33	390	423
Total	42	402	444

entre estas duas proporções (0,35) parece indicar que o uso de contraceptivos orais é mais frequente entre os casos (infartos). Mas será que este resultado não ocorreu por mero acaso?

A expressão $\chi^2_{cY} = 24,76 > 3,84 = \chi^2_{1;95\%}$, nos indica, com alto grau de confiança (95%) afirmar que existe associação entre o uso de contraceptivos orais e infarto do miocárdio para pacientes entre 30 e 34 anos.

```
M=matrix(c(9,33,12,390),nr=2)
chisq.test(M) # com correção de Yates automática
```

4.2 Teste exato de Fisher

Nos casos em que formamos uma tabela de contingência com formato 2×2 , com **pequeno número de observações** ($n < 20$ ou **próximo**) e, conseqüentemente, com frequências observadas em cada casela muito baixas, a literatura apresenta a utilização do *teste exato de Fisher* (Fisher, R. A.; 1922 - Estatístico inglês, biólogo e geneticista), no qual estimamos, **a partir da menor frequência contida na tabela**, a probabilidade de ocorrência deste valor e de uma frequência menor ainda, fazendo-se $p = \sum_{i=0}^n p_i$, em que n é a menor frequência verificada na tabela.

Em uma tabela de contingência 2×2 , com os totais marginais fixos, apresentada como a Tabela 4.11, as frequências observadas têm distribuição hipergeométrica e a probabilidade de ocorrência i será dada por:

$$p_i = \frac{n_1!n_2!m_1!m_2!}{a!b!c!d!n!} \quad (4.6)$$

O método

Seja A o evento que representa a seguinte tabela de contingência: onde os totais nas linhas são

Tabela 4.13: Tabela de contingência genérica.

Linha	Coluna		Total
	C ₁	C ₂	
L ₁	x_1	$n_1 - x_1$	n_1
L ₂	x_2	$n_2 - x_2$	n_2
Total	$x_1 + x_2$	$n - x_1 - x_2$	n

considerados fixos.

Então a probabilidade de A quando H_0 : “Não há associação entre as variáveis”, é verdadeira será dada por:

$$P(A) = \binom{n_1}{x_1} \binom{n_2}{x_2} p^{x_1+x_2} (1-p)^{n-x_1-x_2}, \quad 0 < p < 1.$$

Agora seja B o evento cujos totais nas colunas são $(x_1 + x_2)$ e $(n - x_1 - x_2)$. Desde que a probabilidade de estar na coluna C_1 é denotada por p , temos uma situação onde a distribuição binomial se aplica, assim

$$P(B) = \binom{n}{x_1 + x_2} = \frac{n!}{(x_1 + x_2)!(n - (x_1 + x_2))!} p^{x_1 + x_2} (1 - p)^{n - x_1 - x_2}.$$

Estamos interessados na probabilidade da ocorrência do evento A dado que os totais coluna são fixados, isto é, dado que B é verdadeiro:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Considere o fato que se A ocorre (obtemos a tabela 2x2 especificada pelo evento A), B automaticamente ocorre os totais coluna são $(x_1 + x_2)$ e $(n - x_1 - x_2)$. O evento $A \cap B$ ocorre se, e somente se, o evento A ocorre. Assim, temos que $P(A \cap B) = P(A)$, logo:

$$P(A|B) = \frac{\binom{n_1}{x_1} \binom{n_2}{x_2}}{\binom{n}{x_1 + x_2}} = \frac{P(A)}{P(B)}$$

e por fim:

$$P(A|B) = \frac{n_1!n_2!(x_1 + x_2)!(n - x_1 - x_2)!}{n!x_1!x_2!(n_1 - x_2)!(n_2 - x_2)!} = \frac{n_1!n_2!m_1!m_2!}{a!b!c!d!n!}. \quad (4.7)$$

Para aplicar o teste devemos somar a probabilidade dessa tal ocorrência do evento, com a probabilidade da ocorrência mais extrema, isto é, o p-valor é a soma das probabilidades hipergeométricas para os resultados ao menos tão favoráveis à hipótese alternativa quanto os resultados observados e, portanto, igual a probabilidade hipergeométrica que x_1 é ao menos tão grande quanto o valor observado.

Se por exemplo, os totais linha e coluna são iguais a 2, existem 3 possíveis tabelas:

$$\begin{array}{ccc} \begin{array}{|c|c|} \hline 2 & 0 \\ \hline 0 & 2 \\ \hline \end{array} & \begin{array}{c} \text{Total} \\ 2 \\ 2 \end{array} & \Rightarrow \frac{\binom{2}{0}\binom{2}{0}}{\binom{4}{2}} = \frac{1}{6} \\ \hline \begin{array}{|c|c|} \hline 1 & 1 \\ \hline 1 & 1 \\ \hline \end{array} & \begin{array}{c} \text{Total} \\ 2 \\ 2 \end{array} & \Rightarrow \frac{\binom{2}{1}\binom{2}{1}}{\binom{4}{2}} = \frac{2}{3} \\ \hline \begin{array}{|c|c|} \hline 0 & 2 \\ \hline 2 & 0 \\ \hline \end{array} & \begin{array}{c} \text{Total} \\ 2 \\ 2 \end{array} & \Rightarrow \frac{\binom{2}{0}\binom{2}{2}}{\binom{4}{2}} = \frac{1}{6} \end{array}$$

Exemplo 4.5. *Caso 1 - presença de valor nulo na tabela (o cálculo é direto).*

Tabela 4.14: Associação entre Reação e Enzima.

Reação	Enzima		Total
	Presente	Ausente	
+	5	1	6
-	0	3	3
Total	5	4	9 (< 20)

Deseja-se verificar se existe associação entre Reação (+ ou -) e Enzima (presente ou ausente). Como $n = 20$ é baixo, a estatística de Fisher é sugerida. Considerando um nível de significância $\alpha = 5\%$, tem-se:

H_0 : Não há associação entre Reação e Enzima.

$$p = \frac{6!3!5!4!}{9!} \cdot \frac{1}{5!1!0!3!} = 0,0476 < 5\%$$

logo rejeita-se H_0 . Observe que neste caso, um único cálculo produziu a probabilidade de rejeição de H_0 , sendo a mesma verdadeira, isto é, o p-valor (teste unilateral).

Exemplo 4.6. *Caso 2. - ausência de valor nulo na tabela.*

Neste caso o cálculo das probabilidades é análogo, mas em seguida, construímos outra tabela 2x2, subtraindo uma unidade dos valores das células da diagonal com menor número de casos. As duas unidades subtraídas são adicionadas aos valores das células da outra diagonal, a fim de que todos os totais marginais não sejam alterados. A partir dessa tabela obtém-se nova probabilidade e assim sucessivamente até que uma das células seja "nula".

$$\text{Hipóteses : } \begin{cases} H_0 : p_1 \leq p_2 \text{ "não há associação"} \\ H_a : p_1 > p_2 \end{cases}$$

Tabela 4.15: Associação entre Reação e Enzima. Tabela 1.

Reação	Enzima		Total
	Presente	Ausente	
+	5	3	8
-	2	5	7
Total	7	8	15

Observe que as marginais são fixas !

$p_1 = 5/8 = 63\%$ vs $p_2 = 2/7 = 29\%$;

$$p_2 = \frac{8!7!7!8!}{5!3!2!5!15!} = 0,18275$$

Tabela 4.16: Associação entre Reação e Enzima. Tabela 2.

Reação	Enzima		Total
	Presente	Ausente	
+	6	2	8
-	1	6	7
Total	7	8	15

$$p_1 = \frac{8!7!7!8!}{6!2!1!6!15!} = 0,0305$$

Tabela 4.17: Associação entre Reação e Enzima. Tabela 3.

Reação	Enzima		Total
	Presente	Ausente	
+	7	1	8
-	0	7	7
Total	7	8	15

$$p_0 = \frac{8!7!7!8!}{7!1!0!7!15!} = 0,00124$$

A soma dos p_i é o próprio p-valor (unilateral) do teste, isto é, $p = 0,18275 + 0,0305 + 0,00124 = 0,2145 > 5\%$, logo não rejeitamos H_0 .

Exemplo 4.7. Um estudo foi realizado para verificar a existência de associação entre o tipo de tratamento e mortalidade por AIDS. A Tabela 4.18 apresenta os dados.

Tabela 4.18: Associação entre Tipo de tratamento e Mortalidade por AIDS.

Tratamento	Mortalidade		Total
	Sim	Não	
A	7	5	12
B	1	9	10
Total	8	14	22

Hipóteses : $\begin{cases} H_0 : p_A \leq p_B & \text{"não há associação entre tipo de tratamento e mortalidade"} \\ H_a : p_A > p_B \end{cases}$

$p_A = 58\%$ vs $p_B = 10\%$;

Sendo $p = \sum_{i=0}^1 p_i$, temos

$$p_1 = \frac{12!10!8!14!}{7!5!1!9!22!} = 0,024.$$

e

Tabela 4.19: Associação entre Tipo de tratamento e Mortalidade por AIDS.

Tratamento	Mortalidade		Total
	Sim	Não	
A	8	4	12
B	0	10	10
Total	8	14	22

assim

$$p_0 = \frac{12!10!8!14!}{8!4!0!10!22!} = 0,0015.$$

O valor de p será $0,024 + 0,0015 = 0,0255$ (p -valor = 0,0263 - unilateral).

Como este p é menor que o nível de significância, para $\alpha = 0,05$ a decisão correta será rejeitar H_0 , isto é, pode-se concluir que **há diferença** quanto à mortalidade em relação ao tipo de tratamento, sendo B mais eficaz, já que $P(\text{Morte}|B) = 10\%$ e $P(\text{Morte}|A) = 58\%$.

```
F=matrix(c(7,1,5,9),nr=2)
fisher.test(F,alternative='greater')
# A opção 'greater' é usada para um teste unilateral
```

Observação: O comando “fisher.test” também pode ser utilizado para tabelas de ordem maiores do que 2×2 .

Exemplo 4.8. *Ensaio com ratos modificados geneticamente.*

Suponha um grupo de dezesseis ratos, divididos em dois grupos, Experimental (tratados) e Normal (não tratados). O grupo experimental é formado por 9 animais geneticamente modificados, por apresentarem uma disfunção pancreática com diminuição da capacidade de produção de insulina. Imagine que, após um ano e meio em ambiente controlado, o número de ratos vivos do grupo experimental e do normal seja o seguinte (Tabela 4.20):

Tabela 4.20: *Estudo em ratos sobre disfunção pancreática.*

Grupo	Sobrevida + 1,5 ano		Total
	Vivos	Mortos	
Normal	5	2	7
Experimental	1	8	9
Total	6	10	16

Pode-se notar que aproximadamente $p_1 = 71\%$ dos ratos normais ainda permaneciam vivos, enquanto a sobrevivência para o grupo Experimental é de apenas $p_2 = 11\%$. Considerando um teste unilateral, tem-se as seguintes hipóteses:

$$\text{Hipóteses} : \begin{cases} H_0 : p_1 \leq p_2 \\ H_a : p_1 > p_2 \end{cases}$$

onde o p-valor é $p = 0,0245$, sendo H_0 rejeitada a 5% de significância.

$$\begin{aligned} p_1 &= \frac{7!9!6!10!}{5!2!1!8!16!} = 0,0236 \\ p_0 &= \frac{7!9!6!10!}{6!1!0!9!16!} = 0,00087 \\ p &= 0,00087 + 0,0236 \simeq 0,0245 \end{aligned}$$

No caso de proceder a um teste bilateral para testar as hipóteses a seguir, o procedimento se altera:

$$\text{Hipóteses} : \begin{cases} H_0 : p_1 = p_2 \\ H_a : p_1 \neq p_2 \end{cases}$$

O teste exato de Fisher consiste em elaborar, com base nos totais marginais do fator discriminante da tabela original, duas outras tabelas, que serão denominadas tabelas ou matrizes extremas, X_1 e X_2 . A matriz extrema 1 é a matriz em que todos os animais mortos pertencem ao grupo normal (são 10 mortos, mas o grupo normal tem 7 ratos, logo, os outros 3 que morreram pertencem ao grupo experimental) e todos que sobraram do grupo experimental estão vivos (todos os vivos, pois sobraram $9 - 3 = 6$ ratos). Desta forma, da Tabela 4.20 resulta na Matriz X_1 (Tabela 4.21):

A matriz extrema 2, X_2 , corresponde a uma tabela com os mesmos totais marginais, mas com todos os animais vivos no grupo normal. Então, pela matriz X_2 (Tabela 4.22):

Observe que existe um óbito no grupo normal, pois o número de ratos do grupo normal é maior que o de ratos que sobreviveram. Obviamente, os 9 mortos que restaram pertencem ao grupo experimental. Nenhum rato do grupo experimental sobreviveu.

Tabela 4.21: *Matriz X₁*.

Grupo	Sobrevida + 1,5 ano		Total
	Vivos	Mortos	
Normal	0	7	7
Experimental	6	3	9
Total	6	10	16

Tabela 4.22: *Matriz X₂*.

Grupo	Sobrevida + 1,5 ano		Total
	Vivos	Mortos	
Normal	6	1	7
Experimental	0	9	9
Total	6	10	16

Finalmente, o valor de significância para o teste é calculado segundo a fórmula apresentada anteriormente:

Para a matriz original:

$$F_0 : \frac{7!.9!.10!.6!}{2!.5!.8!.1!.16!} = 0,02360$$

Para a matriz X₁ :

$$F_{X_1} : \frac{7!.9!.10!.6!}{0!.7!.6!.3!.16!} = 0,0105$$

Para a matriz X₂ :

$$F_{X_2} : \frac{7!.9!.10!.6!}{6!.1!.0!.9!.16!} = 0,0009$$

Finalmente,

$$p = F_0 + F_{X_1} + F_{X_2} = 0,0236 + 0,0105 + 0,0009 = 0,035 \text{ ou } 3,5\%.$$

Ou seja, a afirmação de que a sobrevida dos ratos geneticamente alterados é diferente que a dos ratos normais envolve uma probabilidade de erro de 3,5%. Portanto, ao nível de 5% de significância, rejeita-se a hipótese de nulidade (ou seja, de que as vidas dos ratos normais e dos transgênicos não são iguais).

Observação: O Teste Exato de Fisher para tabelas com dimensão acima de 2x2, pode ser obtido e o mesmo é utilizado quando ocorrem frequências nas células da tabela iguais a zero e/ou frequências pequenas a ponto de não satisfazerem as exigências necessárias à aplicação do teste qui-quadrado. Tal teste não será visto somente na aula prática, visto que seu cálculo manual é complexo.

```
F=matrix(c(5,1,2,8),nr=2)
fisher.test(F)
```

```
F=matrix(c(0,2,4,4,1,1,4,2,0,0,3,3),nr=3)
fisher.test(F)
```

Exercício 4.2. Verifique se o sexo dos filhos anômalos depende do sexo do genitor anômalo, ao nível de 5%.

Tabela 4.23: Associação entre filhos e pais anômalos.

Genitor Anômalo	Filhos		Total
	Normais	Anômalos	
Pai	3	9	12
Mãe	6	4	10
Total	9	13	22

R. $p = 0,1099$

```
F=matrix(c(3,6,9,4),nr=2)
fisher.test(F,alternative = 'greater')
```

Exercício 4.3. (a) Verifique ao nível de 5%, por meio de um teste unilateral, se bebês expostos à nicotina (proveniente da mãe) interfere no desenvolvimento de uma determinada anomalia.

Tabela 4.24: Associação entre Nicotina e desenvolvimento de anomalia.

Exposição	Anomalia		Total
	Sim	Não	
Sim	12	2	14
Não	2	4	6
Total	14	6	20

R. $p = 0,0374$

(b) Refaça por meio de um teste bilateral.

R. $p = 0,0374$

```
F=matrix(c(12,2,2,4),nr=2)
fisher.test(F,alternative = 'greater')
fisher.test(F)
```

4.3 Teste da Mediana

Um caso especial do teste χ^2 de Pearson, a ideia, no teste da Mediana é semelhante ao caso do Teste dos Sinais (dados pareados), sendo que aqui é **adequado ao caso de dados não-pareados, especialmente se as amostras forem de tamanhos diferentes**. A linha de raciocínio é a seguinte: se as populações são idênticas, a mediana do conjunto formado pelas duas amostras fornece uma boa estimativa da mediana da distribuição comum. Além disso, a proporção de valores abaixo da mediana geral deve tender a ser a mesma nas duas amostras. O fato de a mediana estimada ser um valor aleatório complica um pouco a sistemática de teste, mas para amostras grandes, podemos com bons resultados, reduzir o problema ao teste da igualdade das proporções de valores abaixo (ou acima) da mediana nas duas amostras. Se o número total de valores envolvidos for ímpar, a mediana coincidirá com um dos valores, e deverá ser desconsiderada na análise. Se ambas as amostras tiverem dez ou mais elementos ($n \geq 10$), a condição geral de aproximação da distribuição binomial pela normal estará satisfeita e poderemos, com razoável aproximação:

$$Z_c = \frac{p_1 - p_2}{\sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim N(0, 1) \quad (4.8)$$

onde p_1 e p_2 são frequências relativas de valores abaixo (ou acima) da mediana estimada em cada amostra, p : a estimativa da proporção comum, com base nas duas amostras reunidas. Entretanto, se $p_1 = p_2 = 0,5$, podemos diretamente fazer $p = 0,5$ na expressão e reescrevê-la como:

$$Z_c = \frac{2(p_1 - p_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim N(0, 1) \quad (4.9)$$

valor que será testado bilateralmente ao nível de significância desejado.

Exemplo 4.9. *Comparação entre Procedência de melões.*

Foram pesados (kg) 11 melões de procedência A e 16 de procedência B, escolhidos ao acaso. Os pesos obtidos são dados a seguir. Ao nível $\alpha = 5\%$, há evidência de que a distribuição dos pesos varia com a procedência?

Procedência A	1,23	1,29	1,12	1,26	1,24	1,26	1,08	1,47	1,33	1,41	1,19
Procedência B	1,42	1,26	1,34	1,21	1,25	1,30	1,58	1,32	1,29	1,15	1,45
	1,27	1,28	1,22	1,32	1,37						

Solução

Supondo que não temos elementos para fazer hipóteses sobre as distribuições populacionais, usaremos o teste da mediana. Após ordenar os 27 dados, a mediana geral será o 14º valor, e igual a 1,28. Excluindo este valor (que é de procedência B), temos, considerando p como a frequência relativa de valores abaixo da mediana:

$$Md^c = \text{mediana}\{A \cup B\}$$

$$p_i = \frac{\#x_i < Md^c}{n_i}$$

$$H_0 : p_A = p_B, \text{ ou "os pesos não variam com a procedência".}$$

$$p_A = \frac{7}{11} \simeq 0,636 \text{ e } p_B = \frac{6}{15} = 0,4$$

$$\therefore Z_c = \frac{2(0,636 - 0,4)}{\sqrt{\frac{1}{11} + \frac{1}{15}}} \simeq 1,19 < |\pm 1,96| = Z_{2,5\%}, \quad p\text{-valor} = 0,23446$$

logo em nível de 5% de significância, não há evidência de que a distribuição dos pesos varie com a procedência.

Exercício 4.4. (a) Verifique, pelo teste da mediana, em nível de 5% de significância, se os dois grupos de valores que seguem podem ser considerados provenientes de populações distintas.

Grupo 1:	23	25	29	19	45	22
	20	40	43	17	27	26
	39	40	25	26	49	34
Grupo 2:	31	46	23	21	37	33
	51	28	32	31	40	25

R. $\tilde{\mu}_D = 30$; $p_1 = \frac{11}{18} = 0,61$, $p_2 = \frac{4}{12} = 0,33$, $Z_c = 1,49$ “**Não rejeita-se H_0** ”. **$p = 0,1362$**

(b) Considere p , desconhecido e determine seu valor tal que H_0 seja rejeitada. Sugestão: utilize a equação 4.8.

R. **$p = 0,1754$ ou $p = 0,8246$.**

4.4 Teste de Mann-Whitney

Inicialmente proposto pelo alemão Deuchler, Gustav (1914) e posteriormente aprimorado por Wilcoxon, F. (1945) para amostras de tamanhos iguais, sua extensão para tamanhos amostrais distintos se deu por Mann, Henry e Whitney, Donald Ranson em 1947).

Este teste corresponde a mais uma alternativa para a comparação de **duas amostras independentes cujas resposta é contínua**, utilizando, como os demais testes não-paramétricos, os números naturais para classificação conjunta dos valores observados. Portanto, o posto de um valor de um conjunto de n valores corresponde a um número natural que indicará a sua posição no conjunto anteriormente ordenado (posto, *score* ou *rank*), isto é, todas as N observações recebem uma pontuação através dos números naturais $1, 2, 3, 4, \dots, n$. Assim, ao menor valor se dará o número 1, e assim sucessivamente até o valor maior, que receberá a maior pontuação. Quando ocorre a presença de valores iguais no conjunto, considera-se um ponto médio, não afetando o posto seguinte. Portanto, num conjunto de seis valores já ordenados $\{7-12-18-18-19-23\}$ os postos serão $\{1-2-3,5-3,5-5-6\}$, respectivamente.

Quanto ao procedimento mais adequado para a aplicação do teste, é baseado no cálculo de U_1 e U_2 , sendo:

$$U_i = n_1 n_2 + \frac{n_i(n_i + 1)}{2} - T_i, \quad i = 1, 2. \quad (4.10)$$

Os valores n_1 e n_2 são os tamanhos das duas amostras de T_1 e T_2 , que correspondem as somas dos pontos (postos) **atribuídos aos valores das duas amostras conjuntamente**.

4.4.1 Pequenas amostras ($n_i \leq 20$)

$$U_c = \min\{U_1; U_2\} \quad (4.11)$$

Obtemos uma estatística tabelada : $U_T = U_{\alpha\%, n_1, n_2}$ (Anexo Tabela de valores críticos do teste U de Mann-Whitney) e concluímos para a rejeição de H_0 se $U_c \leq U_T$.

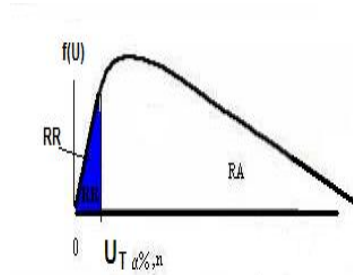


Figura 4.1: Distribuição de Mann-Whitney.

Exemplo 4.10. Testar se a diferença entre as médias dos escores que compõem as duas seguintes amostras aleatórias independentes é significativa ao nível de 5%.

<i>Amostra 1</i>	<i>Amostra 2</i>
10 (12)	10 (12)
4 (4,5)	10 (12)
1 (1,5)	8 (8,5)
2 (3)	7 (7)
4 (4,5)	9 (10)
8 (8,5)	-
1 (1,5)	-
5 (6)	-
$n_1 = 8$ ($T_1 = 41,5$)	$n_2 = 5$ ($T_1 = 49,5$)

$$U_1 = n_1 n_2 + \frac{n_i(n_i + 1)}{2} - T_1 = 8 \times 5 + \frac{8(8 + 1)}{2} - 41,5 = 34,5$$

$$U_2 = 8 \times 5 + \frac{5(5 + 1)}{2} - 49,5 = 5,5$$

O valor tabelado $U_T = U_{5\%, 8, 5} \simeq 7$ é maior que $U_c = \min\{34,5; 5,5\} = 5,5$ ($p = 0,02$), logo rejeita-se H_0 .

```
x=c(10,4,1,2,4,8,1,5)
y=c(10,10,8,7,9)
wilcox.test(x,y,correct=FALSE,paired=FALSE)
```

4.4.2 Grandes amostras ($n_i > 20$)

Neste caso o teste pode ser aplicado por aproximação normal:

$$\mu^* = \frac{n_1 n_2}{2}, \text{ sendo } n_i = \min\{n_1; n_2\}$$

e

$$\sigma^* = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}},$$

considerando $w = \min\{U_1; U_2\}$ tem-se a expressão:

$$Z = \frac{w - \mu^*}{\sigma^*} \sim N(0, 1). \quad (4.12)$$

Exemplo 4.11. Um estudo no setor de psiquiatria do departamento de medicina da UEM, teve como objetivo verificar se os níveis de depressão medidos na escala HAD, diferem significativamente ($\alpha = 5\%$) entre Homens e Mulheres. Os dados de uma amostra de 22 indivíduos de cada gênero é apresentada na Tabela. Sabe-se que $n_M = 22$, $n_H = 22$, $T_M = 455,5$ e $T_H = 490,5$. O valor de U_M e de U_H , respectivamente serão:

$$U_M = 22 \times 22 + \frac{22(22+1)}{2} - 455,5 = 281,5,$$

$$U_H = 22 \times 22 + \frac{22(22+1)}{2} - 490,5 = 246,5.$$

O teste segue com os cálculos:

$$\mu^* = \frac{22 \times 22}{2} = 242 \text{ e } \sigma^* = \sqrt{\frac{22 \cdot 22 (22 + 22 + 1)}{12}} = 42,6.$$

Sendo assim, $w = \min\{281,5; 246,5\} = 246,5$, tem-se:

$$Z = \frac{246,5 - 242}{42,6} = -0,11 \text{ (p-valor bilateral} = 0,54).$$

Como o valor de $|Z|$ é maior do que $|Z_\alpha|$, quando $\alpha = 5\%$, ou seja, o valor 1,96, rejeitamos H_0 . Considera-se que as amostras diferem entre si ao nível de 5% de significância.

Observação: Podem existir parametrizações diferente par determinados programas computacionais. Os comandos a seguir, pode-se observar que os resultados são diferentes dos vistos anteriormente, isto é, $U = 202,5$ ($p = 0,4872$).

```
m=c(1,5,2,4,3,6,10,1,0,2,5,1,6,7,15,7,16,9,9,13,10,18)
h=c(2,7,10,18,15,0,3,3,5,18,11,8,10,8,1,10,7,5,5,3,15)
wilcox.test(m,h,correct=FALSE,paired=FALSE)
```

Exercício 4.5. Participaram de um concurso de elegância infantil 38 crianças de 10 anos. O número de meninos era igual ao de meninas. Essas crianças dispunham de um farto e variado guarda-roupa, devendo elas próprias fazer as escolhas. As notas que elas receberam estão indicadas a seguir:

Meninos: 10,26,35,99,12,16,94,61,58,65,43,38,51,70,92,77,84,89,90;

Meninas: 18,27,36,100,79,54,52,57,41,44,50,67,63,72,88,96,85,91,86.

Realize o teste estatístico adequado para afirmar ou negar, com $\alpha = 5\%$, que os critérios de escolha diferem conforme o sexo (não há normalidade nos dados);

R. $W = T_{\text{masc}} = 354$, $T_{\text{fem}} = 387$, $z = -0,4817$, não rejeita-se H_0 .

Exercício 4.6. São apresentados, a seguir, os aumentos de peso (em libras) de duas amostras independentes de perus alimentados com duas rações diferentes, mantidos em condições idênticas. Utilize a aproximação normal ao nível de 5%, para testar se existe diferença entre os tratamentos.

Ração 1:	16,3	10,1	10,7	13,5	14,9	11,8	14,3	10,2
	12,0	14,7	23,6	15,1	14,5	18,4	13,2	14,0
Ração 2:	21,3	23,8	15,4	19,6	12,0	13,9	18,8	19,2
	15,3	20,1	14,8	18,9	20,7	21,1	15,8	16,2

R. $W = T_{r1} = 181,5$, $T_{r2} = 346,5$, $z = -3,11$ ($p = 0,0019$), **rejeita-se H_0 .**

Exercício 4.7. Em um experimento, planejado para estimar os efeitos da inalação prolongada de óxido de cádmio, 15 cobaias serviram de sujeitos para o experimento, enquanto 10 serviram para o controle. A variável de interesse foi o nível de hemoglobina depois do experimento. Deseja-se saber se podemos concluir, ao nível de 5% de significância, que a inalação prolongada de óxido de cádmio diminui o nível de hemoglobina, conforme os seguintes dados:

Expostos:	14,4	14,2	13,8	16,5	14,1	16,6	15,9	15,6	14,1	15,3
	15,7	16,7	13,7	15,3	14,0					
Não-expostos:	17,4	16,2	17,1	17,5	15,0	16,0	16,9	15,0	16,3	16,8
(controles)										

R. $W = T_{\text{exp}} = 145$, $T_{n-\text{exp}} = 180$, $z = -2,77$ ($p = 0,0055$), **rejeita-se H_0 .**

Exercício 4.8. *Análise de selênio no soro, líquido folicular e número de embriões na área zootécnica.*

Tabela 4.25: *Análise de selênio no soro, amostras coletadas no primeiro e no último dia do experimento e de líquido folicular, no último dia do experimento.*

Animal	Grupo	Soro (A: inicial)	Soro (B: final)	Líquido folicular (C)
7080	Tratado	0,094	0,110	0,130
7051	Tratado	0,050	0,080	-
4398	Tratado	0,096	0,076	0,090
6958	Tratado	0,040	0,130	0,080
6930	Tratado	0,036	0,110	0,190
média	-	0,063	0,101	0,123
7042	Controle	0,062	0,100	0,170
7058	Controle	0,068	0,068	-
1817	Controle	0,094	0,190	0,190
3794	Controle	0,038	0,046	0,050
70	Controle	0,026	0,036	0,150
média	-	0,058	0,088	0,140

(a) (Tabela 4.25) Teste ao nível de 5% se existe diferença entre tratados antes e depois, com relação ao nível de soro;

(b) (Tabela 4.25) Teste ao nível de 5% se existe diferença entre tratados e controles com relação ao nível de soro antes, depois, e com relação ao nível de líquido folicular.

Tabela 4.26: *Análise de Totais de embriões viáveis e inviáveis, nível de sódio e Líquido folicular.*

Animal	Grupo	Viável	Inviável	D-7	Sódio inicial	Sódio final
7080	Tratado	22	5	16	0,094	0,110
7051	Tratado	13	4	8	0,050	0,080
4398	Tratado	94	19	46	0,096	0,076
6958	Tratado	1	-	-	0,040	0,130
6930	Tratado	46	17	18	0,036	0,110
7042	Controle	32	11	16	0,062	0,100
7058	Controle	3	4	4	0,068	0,068
1817	Controle	9	-	6	0,094	0,190
3794	Controle	26	3	8	0,038	0,046
70	Controle	45	4	32	0,026	0,036

(c) (Tabela 4.26) Teste ao nível de 5% se existe diferença entre tratados antes e depois, com relação ao nível de soro;

(d) (Tabela 4.26) Teste ao nível de 5% se existe diferença entre tratados e controles com relação ao total viável, inviável, e com relação ao D-7.

4.5 Teste de Kolmogorov-Smirnov para duas amostras independentes

A prova de Kolmogorov-Smirnov de duas amostras comprova se as mesmas foram extraídas da mesma população. A prova bilateral é sensível a qualquer diferença nas distribuições das quais se extraíram as amostras - diferenças de localização, na dispersão, na assimetria etc. A prova unilateral é utilizada para determinar se os valores da população da qual se extraiu uma das amostras são, ou não, estocasticamente maiores do que os valores da população que originou a outra amostra, por exemplo, para testar a predição de que os escores de um grupo experimental serão "melhores" do que os escores do grupo controle.

A presente metodologia diz respeito à **concordância entre duas distribuições cumulativas**, assim como na metodologia utilizada para uma amostra.

Método

Para aplicar a prova de Kolmogorov-Smirnov de duas amostras, construímos uma distribuição de frequência cumulativa para cada amostra de observações, utilizando os mesmos intervalos para ambas as distribuições. Para cada intervalo, subtraímos uma função da outra. A prova focaliza o maior desses desvios observados.

Seja $S_{n_1}(X)$ = função acumulada observada para uma das amostras, isto é $S_{n_1}(X) = \frac{k}{n_1}$, onde k : número de escores não superiores à X , e seja $S_{n_2}(X)$ = função acumulada observada para a outra amostra, $S_{n_2}(X) = \frac{k}{n_2}$.

A estatística teste será:

$$D_c = \max \{|S_{n_1}(X) - S_{n_2}(X)|\} \sim D_{\alpha\%}. \quad (4.13)$$

A distribuição amostral de D é conhecida, achando-se, por outro lado, tabeladas as probabilidades associadas à ocorrência de valores tão grandes quanto um D observado sob hipótese de nulidade: que "as duas amostras provenham de mesma distribuição", isto é

$$H_0 : F_1(X) \equiv F_2(X).$$

A hipótese nula será rejeitada ao nível $\alpha\%$ de significância se D_c exceder o valor tabelado, isto é, se $D_c > D_{\alpha\%,n}$.

4.5.1 Pequenas amostras ($n_i \leq 40$)

Quando $n_1 = n_2$, e quando tanto n_1 como n_2 não superam 40, pode-se utilizar diretamente valores críticos tabelados para as comparações. Assim, por exemplo, em uma prova unilateral com $n = 14$, se $D_c > 7$ (Anexo - Tabela de valores críticos do teste de Kolmogorov-Smirnov para duas amostras), rejeitamos H_0 ao nível de 5% (teste unilateral).

Exemplo 4.12. *Efeito de primazia em estudantes.*

Lepley (1934) comparou o aprendizado serial de 10 alunos da 7ª série com o aprendizado serial de 10 alunos do 1º colegial, para comprovar a hipótese de que o efeito de primazia é menos predominante no aprendizado de estudantes mais jovens. O efeito de primazia é a tendência para reter a matéria aprendida no começo de determinada série mais facilmente do que a matéria aprendida no fim daquela série. Lepley comprovou sua hipótese comparando o percentual de erros cometidos pelos dois grupos na primeira metade da série, prevendo que o grupo mais velho cometeria relativamente menos erros do que o grupo mais jovem, ao evocar a primeira metade da série.

% de erros										
7ª Série	39,1	41,2	45,2	46,2	48,4	48,7	55,0	40,6	52,1	47,2
1º Col	35,2	39,2	40,9	38,1	34,4	29,1	41,8	24,3	32,4	32,6
<i>Mediana : $\tilde{\mu}$ \bar{x} s</i>										
7ª Série	46,7			46,37		5,08				
1º Col	34,8			34,8		5,48				

As hipóteses a serem testadas são:

$$H_0 : \tilde{\mu}_{7a} \leq \tilde{\mu}_{1a}$$

$$H_a : \tilde{\mu}_{7a} > \tilde{\mu}_{1a}$$

7ª Série (X)	1ª Série (Y)	S_x	S_y	$ S_x - S_y $
-	24,3	0	1/10	0,1
-	29,1	0	2/10	0,2
-	32,4	0	3/10	0,3
-	32,6	0	4/10	0,4
-	34,4	0	5/10	0,5
-	35,2	0	6/10	0,6
-	38,1	0	7/10	0,7
39,1	-	1/10	7/10	0,6
-	39,2	1/10	8/10	0,7
40,6	-	2/10	8/10	0,6
-	40,9	2/10	9/10	0,7
41,2	-	3/10	9/10	0,6
-	41,8	3/10	10/10	0,7
45,2		4/10	10/10	0,6
46,2		5/10	10/10	0,5
47,2		6/10	10/10	0,4
48,4		7/10	10/10	0,3
48,7		8/10	10/10	0,2
42,1		9/10	10/10	0,1
55,0		10/10	10/10	0,0

Note que a maior discrepância entre as duas séries é $D_c = \max\{|S_x - S_y|\} = 0,7$. Considerando um nível de significância: 5% e $n_1 = n_2 = 10$, então como $D_{\alpha\%,10} = 0,6$ com $D_c = 0,7$, nossa decisão será em rejeitar H_0 , e concluir que os alunos do 1º colegial cometem proporcionalmente menos erros do que os da 7ª série, ao evocar a primeira metade da série ($\tilde{\mu}_{7a} > \tilde{\mu}_{1a}$).

Outra metodologia que pode ser utilizada é a de Hollander e Wolfe (1973), que consideraram a estatística J_3 tal que:

$$J_3 = \frac{n_1 n_2}{d} D_c$$

onde d : máximo divisor comum de n_1 e n_2 .

Desta forma tem-se:

$$J_3 = \frac{10 \cdot 10}{10} \frac{7}{10} = 7,$$

e sendo $J_3 = 7$ pela Tabela H-W de probabilidades (Anexo - Tabela de valores críticos de Kolmogorov-Smirnov/Hollander-Wolfe), $p = 0,012$, portanto rejeita-se H_0 .

```
k1=c(39.1,41.2,45.2,46.2,48.4,48.7,55,40.6,52.1,47.2)
k2=c(35.2,39.2,40.9,38.1,34.4,29.1,41.8,24.3,32.4,32.6)
?ks.test
ks.test(k1,k2)
```

Exercício 4.9. a) Verifique ao nível de 5% se as populações ao quais as amostras a seguir foram retiradas, são idênticas.

b) Refaça o exercício utilizando a estatística de Hollander e Wolfe.

$X : \{7,6 \ 8,4 \ 8,6 \ 8,7 \ 9,3 \ 9,9 \ 10,1 \ 10,6 \ 11,2\}$

$Y : \{5,2 \ 5,7 \ 5,9 \ 6,5 \ 6,8 \ 8,2 \ 9,1 \ 9,8 \ 10,8 \ 11,3 \ 11,5 \ 12,3 \ 12,5 \ 13,4 \ 14,6\}$

R. a) $D_c = 0,4 < D_{5\%} = 0,533$ "Não rejeito H_0 "; $p = 0,2653$. b) $J_3 = 18$, $p = 0,101$

```
k1=c(7.6,8.4,8.6,8.7,9.3,9.9,10.1,10.6,11.2)
k2=c(5.2,5.7,5.9,6.5,6.8,8.2,9.1,9.8,10.8,11.3,11.5,12.3,12.5,13.4,14.6)
ks.test(k1,k2)
```

Exercício 4.10. Concentração de glicose no sangue.

Em um experimento para comparar a eficácia de dois remédios A e B na redução das concentrações de glicose no sangue, o remédio A foi dado a 25 pacientes e o remédio B a outros 15. As reduções das concentrações de glicose no sangue para os pacientes que receberam o remédio A e B estão apresentadas a seguir. Utilize a aproximação para grandes amostras e teste ao nível de 2%.

Remédio A:	0,04	0,13	0,16	0,28	0,35	0,39	0,40	0,44	0,49
	0,58	0,71	0,72	0,75	0,77	0,83	0,90	0,91	1,12
	1,15	1,20	1,38	1,50	1,54	1,64	1,73		
Remédio B:	0,68	0,86	0,89	1,01	1,07	1,25	1,26	1,31	1,48
	1,59	1,63	1,78	1,81	1,82	1,95			

R. $D_c = 0,533 > 0,496 = D_{5\%}$ ($p = 0,005997$) - "Rejeita-se H_0 ".

Observação: Não há um valor tabelado para $D_{5\%}$, logo uma aproximação dos valores críticos de Kolmogorov-Smirnov pode ser utilizada: $\sqrt{\frac{40}{375}} \simeq 0,33$, então $1,36(0,33) \simeq 0,45$.

4.5.2 Grandes amostras ($n_i > 40$)

Quando tanto n_1 , como n_2 são maiores que 40, pode-se utilizar aproximações para os valores críticos de Kolmogorov-Smirnov para as comparações. Aqui não é necessário que estas quantidades sejam iguais.

Bilateral	$D_{\alpha\%}$	Unilateral	
$\alpha = 1\%$	$1,63k$	$\alpha = 0,5\%$	
$\alpha = 2\%$	$1,52k$	$\alpha = 1\%$	
$\alpha = 5\%$	$1,36k$	$\alpha = 2,5\%$	
$\alpha = 10\%$	$1,22k$	$\alpha = 5\%$	
$\alpha = 20\%$	$1,07k$	$\alpha = 10\%$	

; onde $k = \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$

Exemplo 4.13. Considere $n_1 = 55$ e $n_2 = 60$, e que um pesquisador deseje usar uma prova bilateral ao nível de 5%.

O cálculo do valor tabelado será:

$$D_{5\%} = 1,36k = 1,36\sqrt{\frac{55+60}{55 \cdot 60}} = 0,254.$$

O valor calculado deverá ser maior que $D_{5\%} = 0,254$ para que H_0 possa ser rejeitada.

Exercício 4.11. A tianeptina é um fármaco anti-depressivo. Para verificar sua eficácia, foram selecionados dois grupos, ambos depressivos, e seus escores foram anotados. Um dos grupos foi submetido ao fármaco e o outro não. Após um período de observação os resultados foram apresentados. Teste ao nível de 5% a eficácia da Tianeptina quanto a redução da Depressão. **R.** $U_c = 44 < 64 = U_{5\%}$, "Rejeita-se H_0 ", logo há diferença. $p = 0,00367$

Tabela 4.27: Tratamento de Depressão com Tianeptina.

Grupo 1: Tianeptina :	6	33	21	26	10	29	33	29	37	15	2	21	7	26	13	-
Grupo 2: Placebo :	24	46	26	44	27	34	33	25	35	30	38	38	31	27	34	32

Exercício 4.12. (Pagano et al., pg. 275) Considere as distribuições de escores de idade mental de duas populações de crianças que sofrem de fenilcetonúria (Indivíduos com essa disfunção são incapazes de metabolizar a proteína fenilalamina). Foi sugerido que um elevado nível sérico de fenilalamina aumenta a probabilidade de deficiência mental das crianças. Teste ao nível de 5% que baixa exposição à fenilalamina (< 10 mg/dl) difere da alta exposição (≥ 10 mg/dl) com relação a idade mental. **R.** $T_c = 0 < 47 = T_{5\%}$, "Rejeita-se H_0 ", logo há diferença.

Tabela 4.28: Escores de idade mental em crianças.

Baixa exposição	34,5	37,5	39,5	40	45,5	47	47	47,5	48,7	49	51	51	52
	53	54	54	55	56,5	57	58,5	58,8					
Alta exposição	28	35	37	37	43,5	44	45,5	46	48	48,3	48,7	51	52
	53	53	54	54	55								

Exercício 4.13. A frequência da inversão F do cromossomo III foi estudada em populações urbanas de *Drosophila willistoni*. Os exemplares foram coletados em zonas considerada de alta, média e baixa urbanização, na cidade de Porto Alegre (Valente et al, 1993). Os dados foram comparados entre si e com um grupo-controle não-urbano. Verifique ao nível de 1% de significância se existem evidências de associação entre a inversão F e a Urbanização. Utilize os resíduos padronizados para concluir sobre a possível relação. **R.** $\chi^2_c = 70,3 > 11,35 = \chi^2_{1\%,3}$, "**Rejeita-se H_0** ".

Tabela 4.29: Associação entre Inversão F e Urbanização quando comparadas a um controle.

Inversão F	Urbanização				Total
	Alta	Intermediária	Baixa	Controles	
Sim	63	421	641	223	1349
Não	475	1201	1542	658	3875
Total	538	1622	2183	881	5224
% Inversão F	12%	26%	29%	25%	26%

Exercício 4.14. Crianças hospitalizadas e alimentadas por nutrição endovenosa às vezes apresentam colestase, um bloqueio do fluxo da bile que pode produzir cálculos biliares e outros problemas. Suponha que foi realizado um estudo do tipo caso-controle para avaliar o efeito ($\alpha = 5\%$) de uma infecção grave sobre o risco de colestase em crianças com nutrição parenteral. A Tabela 4.30 mostra o resultado de um estudo feito em 113 crianças hospitalizadas. Como o desfecho, isto é, a colestase, é uma característica rara na população. (Fonte: Carvalho, 1993). **R.** $\chi^2_c = 5,54 < 3,84 = \chi^2_{5\%,1}$, "**Não rejeita-se H_0** ", logo não há associação.

Tabela 4.30: Presença de colestase em crianças com nutrição endovenosa, com e sem infecção grave.

Infecção grave	Colestase	
	Sim	Não
Sim	19	61
Não	1	32

Exercício 4.15. Com o objetivo de avaliar fatores de risco para o câncer intra-epitelial da cérvix uterina, Soares (1998) estudou 43 casos com essa doença e 63 mulheres controles da população de Porto Alegre. A Tabela 4.31 apresenta dados relativos à presença do alelo $DQB1^*03$, do sistema HLA . Teste a associação entre a doença e o fator ao nível de 5%. **R.** $\chi^2_c = 15,36 > 3,84 = \chi^2_{5\%,1}$, "**Rejeita-se H_0** ", logo há associação.

Tabela 4.31: Associação entre câncer intra-epitelial da cérvix uterina e presença do alelo $DQB1^*03$.

Câncer	Presença do alelo	
	$DQB1^*03$	Outro
Casos	33	10
Controles	24	39

Exercício 4.16. Em um estudo genético realizado em macacos amazônicos da subespécie “*Saimiri sciureus ustus*”, Silva et al. (1993) encontraram variação nas frequências de dois tipos de enzimas *glioxalase* (GLO) em animais que vivem nas margens do rio Jamari, em Rondônia. Na Tabela 4.32, estão os números observados de animais com diferentes tipos de enzimáticos, coletados nas margens esquerda e direita desse rio. Verifique se existem evidências de associação entre as variáveis ao nível de 1%.

Tabela 4.32: Associação entre tipos enzimáticos e margem do rio Jamari-Rondônia.

Enzima	Margem do rio	
	Esquerda	Direita
GLO 2	72	74
GLO 2-3	22	3

Exercício 4.17. Verificar se a frequência dos dois tipos de cromossomos Y está relacionado com o grau de mistura racial aparente do touro ao nível de 5%.

Tabela 4.33: Associação entre tipo de cromossomo e contaminação de raça aparente.

Contaminação racial aparente	Tipo do cromossomo	
	Acrocêntrico	Submetacêntrico
Sim	38	16
Não	13	8

Exercício 4.18. Ao investigar a eventual associação entre sinais inflamatórios e velocidade de hemossedimentação, um clínico observou os seguintes resultados em 15 pacientes. Com base nesses resultados, o que se pode concluir a respeito da suposta associação entre as duas características estudadas, ao nível de 5%? **R. $p = 0,031$, “Rejeita-se H_0 ”, logo há associação.**

Tabela 4.34: Associação entre Inflamação e Hemossedimentação.

Inflamação	Hemossedimentação	
	Aumentada	Normal
Sim	6	2
Não	1	6

Exercício 4.19. O coleóptero *chauliognathus flavipes* pode apresentar 10 diferentes padrões para as manchas pretas que ocorrem sobre os élitros (asas), que são amarelos. Machado e Araújo (1994) coletaram insetos dessa espécie em várias localidades do Rio Grande do Sul, nos anos de 1989 e 1990, como mostra a Tabela 4.35. Teste a hipótese de igualdade de proporções entre as localidades ao nível de 5% e a 1%.

Tabela 4.35: Número de indivíduos da espécie *Chauliognathus* com diferentes padrões de manchas nos élitros, coletados em três localidades do Rio Grande do Sul.

Localidade	Padrão do élitros		
	Claro	Intermediário	Escuro
Porto Alegre	67	20	4
São Leopoldo	68	29	19
Caxias do Sul	26	3	6

Exercício 4.20. Verifique se as localidades apresentadas na Tabela 4.36 diferem quanto a frequência dos diferentes grupos sanguíneos do sistema ABO, ao nível de 5%. **R. $p = 0,00833$, “Rejeita-se H_0 ”, logo há diferença nas frequências dos grupos sanguíneos.**

Tabela 4.36: Número de indivíduos da espécie *Chauliognathus* com diferentes padrões de manchas nos élitros, coletados em duas localidades do Rio Grande do Sul.

Localidade	Tipo sanguíneo			
	A	B	AB	O
Cidade I	43	9	1	47
Cidade II	29	17	9	45

Exercício 4.21. Um estudo foi conduzido para investigar se existe uma associação entre doença cardíaca e apnéia (ronco). Utilize os dados da Tabela 4.37 e teste esta hipótese ao nível de 5%.

Tabela 4.37: Associação entre doença cardíaca e apnéia.

Doença cardíaca	Apnéia			
	não	ocasionalmente	quase toda noite	todas as noites
Sim	24	35	21	30
Não	1355	603	192	224

Exercício 4.22. Reestruture os dados para uma tabela 2×2 considerando apenas sim ou não para a apnéia, em seguida teste a mesma hipótese ao nível de 5%.

Exercício 4.23. Diagnóstico e tratamento da Osteoporose: *Osteoporose é uma doença esquelética sistêmica caracterizada por baixa massa óssea e alteração da micro-arquitetura, levando a um aumento da fragilidade óssea e, consequentemente, do risco de fratura. A massa óssea aumenta durante a infância e principalmente adolescência, atingindo seu pico em torno dos 25 anos na mulher e 30 a 35 anos no homem. Após a parada do crescimento ósseo, a formação e a reabsorção ocorrem na mesma proporção, em um processo denominado de remodelação óssea. Com o avanço da idade cronológica, a formação óssea torna-se menor que a reabsorção, levando a uma perda de massa óssea em torno de 1% ao ano. A osteoporose é uma doença assintomática até o paciente apresentar alguma fratura. Assim, com objetivo de detectar pacientes com osteoporose coletou-se uma amostra de 30 pacientes aleatoriamente em um Hospital Universitário. De cada paciente realizou-se um cadastro onde tinha as seguintes variáveis: idade, peso, altura, atividade física, história de fratura materna, ingestão adequada de cálcio, raça e outras como, tabagismo, algumas doenças (hipogonadismo, síndromes disabsortivas) e drogas (corticóide) são fatores de risco que aceleram a perda óssea. Utilize os dados a seguir referente ao assunto Osteoporose e verifique se existe diferença a 5% de significância entre os grupos: Tabagismo e Ingestão, com relação entre a variável: IMC, onde*

$$IMC = \frac{\text{Peso}}{\text{Altura}^2};$$

Tabagismo	1	1	2	2	2	1	1	2	2	1	2	1
Ingestão	1	1	2	2	1	1	1	2	1	1	1	2
Altura	1,65	1,60	1,74	1,62	1,97	1,59	1,64	1,50	1,47	1,64	1,71	1,82
Peso	56,8	45,9	68,7	74,8	74,1	58,9	56,4	59,7	46,9	58,7	96,4	85,7
2	2	2	2	2	1	1	1	1	1	1	2	2
2	1	1	1	1	1	1	1	2	2	2	1	1
1,56	1,54	1,75	1,41	1,80	1,65	1,57	1,79	1,87	1,80	1,70	1,65	1,59
57,4	69,8	95,4	56,8	75,6	84,6	56,7	85,4	75,6	74,2	68,4	62,4	63,1
												58,4

Tabagismo: 1 (sim) - 2 (Não); Ingestão: 1 (sim) - 2 (Não).

O Índice de Massa Corporal é um indicador de obesidade e pode ser classificado como apresentado na Tabela 4.38.

Tabela 4.38: Classificação do IMC segundo a OMS.

Condição	IMC em adultos
Abaixo do peso	< 18,5
Peso normal	18,5 – 25,0
Sobrepeso	25,0 – 30,0
Obeso	>= 30,0

Nos exercícios a seguir, realize o teste de comparações para dois grupos independentes e compare os métodos Mann-Whitney e Kolmogorov-Smirnov.

Exercício 4.24. Os dados abaixo representam o número médio de horas durante as últimas noites de 9 pacientes deprimidos e de 8 pessoas normais monitoradas, medido em noites sucessivas. Testar a hipótese de que o tempo médio de sono noturno é afetado em pacientes deprimidos ao nível de 5% de significância.

Pacientes Deprimidos: 7,1 6,8 6,7 7,5 6,2 6,9 6,5 7,3 7,2

Pacientes Normais: 8,2 7,5 7,7 7,8 8,0 7,4 7,3 6,5

R. $U_c = 10,5 < 17 = U_T$, "Rejeita-se H_0 ", logo há diferença.

Exercício 4.25. A desidrogenase da lactase apresenta os seguintes valores para pacientes acometidos de malária e esquitossomose. Teste a hipótese de que existe diferença significativa entre os dois tratamentos, ao nível de 5% de significância.

MALÁRIA: 450 380 800 1100 430 520 475 450 360

ESQUITOSSOMOSE: 350 280 270 370 280 440 320 370

Exercício 4.26. Em estudo realizado no Hospital de São Geraldo da UFMG, Cronemberg & Calixto (1991) estudaram a capacidade de redução da pressão intraocular das drogas timolol, betaxolol e levobunolol. Para isto, medicaram 10 pacientes com as drogas e compararam pareadamente os resultados com aqueles obtidos quando os pacientes receberam um placebo. As medidas da pressão intraocular, expressar em mmHg, obtida às 6 horas da manhã com os pacientes usando timolol e placebo são apresentadas na tabela a seguir. Verifique ao nível de 5% se o timolol reduz a pressão ocular.

Placebo	22	25	23	18	24	24	17	23	22	23
Timolol	18	20	20	17	16	20	20	20	20	24

R. $U_c = 22 < 23 = U_T$, "Rejeita-se H_0 ", logo há redução.

Exercício 4.27. Anuncia-se que um certo tranquilizante tem efeitos curativos em psicóticos. Um psicólogo clínico, desejando estimar a diferença média de tendência psicótica entre dois grupos, planejou um experimento em que um dos grupos (G_1) recebeu o tranquilizante e o outro (G_2) não recebeu. Testar ao nível de confiança de 95% que o tranquilizante tem efeitos curativos em psicóticos.

G_1 : 2 3 5 7 7 8 8 5 9 1

G_2 : 1 1 1 2 2 3 3 4 3 2

Capítulo 5

Testes aplicáveis a k amostras - Dados pareados

5.1 Teste de Cochran

O teste de *Cochran* é utilizado quando se deseja comparar tratamentos distintos aplicados a uma mesma amostra (medidas repetidas), o que caracteriza a dependência ou o pareamento das unidades amostrais em que a resposta é da forma 0 ou 1, isto é, sucesso ou fracasso. O teste de Cochran é uma extensão do teste de McNemar.

As hipóteses a serem testadas são:

H_0 : Os tratamentos não diferem entre si;

H_a : Pelo menos um par de tratamentos diferem entre si.

Estatística teste:

$$Q_c = \frac{k(k-1) \sum_{j=1}^k (T_j)^2 - (k-1)N^2}{kN - \sum_{i=1}^n (S_i)^2} \sim \chi_{\alpha\%;k-1}^2. \quad (5.1)$$

Sendo:

k : tratamentos;

n : indivíduos;

N : total de um's;

T_j : total de um's na coluna j ;

S_i : total de um's na linha i .

A decisão pela rejeição de H_0 será tomada se $Q_c \geq \chi_{\alpha\%;k-1}^2$, assim existe pelo menos um par de tratamentos que diferem entre si ao nível α de significância.

Exemplo 5.1. Um psiquiatra quer testar a eficiência de 4 tratamentos em 10 sujeitos submetidos a um teste cuja resposta é “sim” ou “não”. Os dados são apresentados na Tabela (5.1).

Tabela 5.1: Teste de Cochran: Dez indivíduos submetidos a quatro tratamentos supostamente homogêneos.

Indivíduos	Tratamentos				S_i	$(S_i)^2$
	T_1	T_2	T_3	T_4		
1	1	1	1	1	4	16
2	1	0	0	1	2	4
3	1	0	1	0	2	4
4	1	1	1	0	3	9
5	0	0	0	0	0	0
6	0	1	1	1	3	9
7	1	0	1	1	3	9
8	0	0	1	0	1	1
9	1	1	1	0	3	9
10	0	1	0	0	1	1
T_j	6	5	7	4	N=22	62
$(T_j)^2$	36	25	49	16	126	-

Assim tem-se:

$$\begin{aligned}
 Q_c &= \frac{4(4-1) \sum_{j=1}^4 (T_j)^2 - (4-1)(22^2)}{4(22) - \sum_{i=1}^{10} (S_i)^2} \\
 &= \frac{4(3)(126) - (4-1)(22^2)}{4(22) - (62)} \\
 &= 2,31 < \chi_{5\%,3}^2 = 7,815 \quad (p = 0,511);
 \end{aligned}$$

logo não se rejeita H_0 . Conclui-se que os tratamentos não diferem entre si ao nível de 5% de significância.

Observação: Em caso de rejeição da hipótese nula, um procedimento de comparações múltiplas deve ser realizado (ver Castellanos e Merino, 1989). Uma alternativa é proceder a comparação 2 a 2 por meio do teste McNemar.

Não há uma função interna do R para o teste de Cochran, entretanto, podemos implementar uma função para o cálculo da estatística, assim como o seu respectivo teste de hipótese da seguinte forma:

```
# Instalar a função primeiro !
cochran.test=function(T,S,alfa)
{
  k=length(T)
  s=length(S)
  N=sum(T)
  Qt=qchisq(1-alfa,df=k-1)
  cat('\n', 'Qtabelado:', round(Qt,2))
  Q=(k*(k-1)*sum(T^2)-(k-1)*N^2)/(k*N-sum(S^2))
  cat('\n', 'Q-Cochran:', round(Q,2))
  if (Q>=Qt) cat ('\n', 'Decisão: Rejeita-se Ho')
  else cat ('\n', 'Decisão: Não Rejeita-se Ho')
  p.valor=1-pchisq(Q,df=(k-1))
  cat('\n', 'p-valor:', p.valor)
}

T=c(6,5,7,4)
S=c(4,2,2,3,0,3,3,1,3,1)
cochran.test(T,S,alfa=0.05)
```

Exemplo 5.2. *Suponha que estejamos interessados em estudar a influência da atitude de um entrevistador sobre as respostas das donas de casa a determinada pesquisa de opinião.*

Poderemos treinar um entrevistador para efetuar suas entrevistas de três maneiras diferentes:
entrevista tipo 1 – demonstrando interesse, cordialidade, entusiasmo;
entrevista tipo 2 – demonstrando formalismo, reserva e cortesia;
entrevista tipo 3 – demonstrando desinteresse, modo abrupto, formalismo áspero.

O entrevistador visitaria três grupos de 18 casas, aplicando aleatoriamente o tipo 1 de entrevista a um grupo, o tipo 2 a outro grupo, o tipo 3 ao terceiro grupo.

Teríamos então 3 amostras relacionadas (correspondentes) com 18 elementos cada uma ($n = 18$). Desejamos comprovar se as diferenças fundamentais nos tipos de entrevista influenciariam o número de respostas afirmativas (“sim”) dadas a determinada pergunta.

Os dados são apresentados na Tabela (5.2).

Etapa 1:

H_0 : A probabilidade de um “sim” é a mesma para os três tipos de entrevista.

H_1 : As probabilidades de um “sim” diferem conforme o tipo de entrevista.

Etapa 2: Estabelecendo o nível de significância: $\alpha = 0,01$.

Etapa 3: Estabelecendo a estatística de teste: Q de Cochran porque os dados se referem a mais de dois grupos relacionados ($k = 3$) e se apresentam dicotomizados sob forma “sim” ou “não”.

Etapa 4: Estabelecendo os valores críticos

para $gl = k - 1 = 3 - 1 = 2$ e nível de significância = 0,01 um valor de 9,21.

Etapa 5: Cálculo da Estatística teste:

Calculando os somatórios dos resultados correspondentes a cada grupo e o somatório dos escores de cada bloco (L_i) temos:

$G1 = 13$ - nº total de respostas sim para entrevista 1

$G2 = 13$ - nº total de respostas sim para entrevista 2

$G3 = 3$ - nº total de respostas sim para entrevista 3

Tabela 5.2: *Influência do pesquisador em pesquisas de opinião.*

Conjunto	Resposta à entrevista 1	Resposta à entrevista 2	Resposta à entrevista 3	L_i	L_i^2
1	0	0	0	0	0
2	1	1	0	2	4
3	0	1	0	1	1
4	0	0	0	0	0
5	1	0	0	1	1
6	1	1	0	2	4
7	1	1	0	2	4
8	0	1	0	1	1
9	1	0	0	1	1
10	0	0	0	0	0
11	1	1	1	3	9
12	1	1	1	3	9
13	1	1	0	2	4
14	1	1	0	2	4
15	1	1	0	2	4
16	1	1	1	3	9
17	1	1	0	2	4
18	1	1	0	2	4
T_j	13	13	3	N=29	63
T_j^2	169	169	9	347	-

após cálculos $Q = 16,7$.

Etapa 6: Como $Q > \chi^2_{\text{crítico}}$ ($16,7 > 9,21$), então rejeita-se H_0 ($p = 0,00024$), ou seja, o número de respostas “sim” difere significativamente em relação aos tipos 1, 2 e 3 de entrevista.

```

T=c(13,13,3)
S=c(0,2,1,0,1,2,2,1,1,0,3,3,2,2,2,3,2,2)
cochran.test(T,S,alfa=0.01)

```

Exercício 5.1. Comparar quatro técnicas de treinamento aplicadas a seis blocos de funcionários do setor de marcação de luz de uma empresa cinematográfica, ao nível de 5% de significância. Cada bloco foi formado de modo que os quatro funcionários que o constituíam eram equivalentes em seu desempenho atual no cargo. Os resultados obtidos na escala de avaliação de desempenho, aplicada após o treinamento, foram os que constam na Tabela (5.3).

Observação: Se $x \geq 7 \Rightarrow$ "1" ou "Alto desempenho".

R. $Q_c = 6,18$ ($p = 0,10$)

Tabela 5.3: Teste em funcionários do setor de marcação de luz de uma empresa cinematográfica.

Blocos	Tratamentos							
	X ₁	Escore	X ₂	Escore	X ₃	Escore	X ₄	Escore
A	7	1	9	1	10	1	6	0
B	8	1	12	1	11	1	5	0
C	6	0	7	1	8	1	4	0
D	3	0	5	0	6	0	6	0
E	9	1	10	1	8	1	2	0
F	4	0	7	1	5	0	9	1
Total	-	3	-	5	-	4	-	1

Exercício 5.2. A seguir tem-se os Resultados de medidas de dois questionários distintos aplicados em pacientes de uma clínica psiquiátrica. Em nível de 5%, verifique se as ferramentas de diagnóstico clínico são igualmente eficazes. **R.** $Q_c = 2,89$ ($p = 0,24$)

Questionário	Score de Depressão : Se >7 é doente														
HAD	7	8	6	3	1	9	4	7	6	5	4	6	8	12	4
BECK ₁	4	7	1	4	2	11	6	8	11	11	0	5	6	10	1
BECK ₂	10	6	12	8	6	15	5	1	20	18	8	6	5	13	0

5.2 Teste de Friedman

O teste de Friedman é recomendado como um substituto do teste F para medidas repetidas, quando procede-se à comparação de k amostras relacionadas ou dependentes cujas observações apresentam valores com acentuadas variações e em cada tratamento são constituídos blocos. Na verdade, procura-se fazer a comparação de tratamentos em que são formados blocos com a intenção de que isto resulte em um pareamento considerável entre os diversos tratamentos.

A hipótese nula a ser testada é:

$$H_0 : T_1 = T_2 = \dots = T_k$$

5.2.1 Grandes amostras ($n > 4$)

A forma de realização do teste é análogo aos demais. Dentro de cada um dos n blocos formados procede-se à classificação das i -ésimas observações em k tratamentos, utilizando-se números naturais.

Assim, tem-se a estatística teste:

$$H_{Fr} = \frac{12}{Nk(k+1)} \sum_{j=1}^k T_j^2 - 3N(k+1) \text{ tal que } H_{Fr} \sim \chi_{\alpha\%;(k-1)}^2, \quad (5.2)$$

onde:

N : n° de linhas;

k : n° de colunas ou tratamentos;

T_j : soma dos postos na coluna j ;

Conclusão: Se $H_{Fr} > \chi_{\alpha\%;(k-1)}^2$, rejeita-se H_0 .

Exemplo 5.3. São prescritos quatro procedimentos técnicos para determinação de certa variável (tempo de sobrevivência, em semanas, de pacientes com câncer). Foram formados cinco blocos e obtidos os seguintes valores:

Tabela 5.4: Dados de quatro procedimentos técnicos.

Tratamentos			
A	B	C	D
12 (2)	13 (3)	16 (4)	7 (1)
8 (2)	9 (3)	12 (4)	5 (1)
14 (2)	20 (3)	22 (4)	6 (1)
17 (3)	16 (2)	21 (4)	11 (1)
12 (2)	15 (3)	16 (4)	10 (1)
$T_1 = 11$	$T_2 = 14$	$T_3 = 20$	$T_4 = 5$

ATENÇÃO: Observe que os Ranks são colocados nas linhas.

Substituindo-se os valores na expressão no teste, tem-se:

$$\begin{aligned}
 H_{Fr} &= \frac{12}{5(4)(4+1)} \sum_{j=1}^4 T_j^2 - 3(5)(4+1) \\
 &= \frac{12}{100} [121 + 196 + 400 + 25] - 75 \\
 &= 14,04 \quad (p = 0,00285).
 \end{aligned}$$

Como $H_{Fr} \sim \chi^2_{(k-1)}$, então $H_{Fr} \sim \chi^2_{(4-1)} \Rightarrow \chi^2_{3;5\%} = 7,82$ e $\chi^2_{3;1\%} = 11,34$, portanto, considerando que o valor encontrado de $H_{Fr} = 14,04$ é maior do que os valores da tabela, tanto para $\alpha = 5\%$ como para $\alpha = 1\%$, concluímos pela rejeição de H_0 , logo verifica-se que há diferença significativa entre as medidas dos tratamentos. Aparentemente o tratamento C apresentou melhores resultados em relação aos demais grupos.

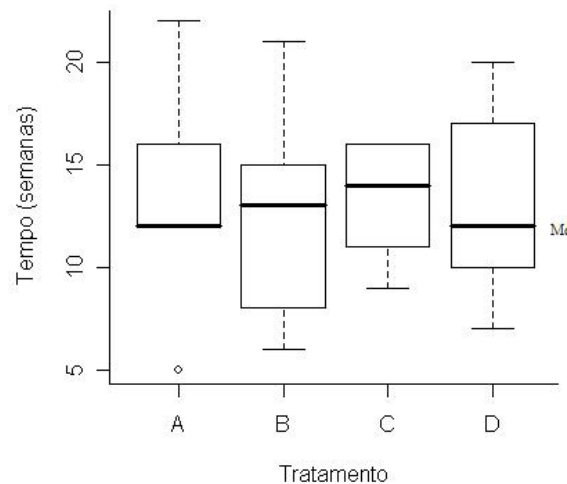


Figura 5.1: Tempos de sobrevida (semanas) de pacientes submetidos a quatro tratamentos distintos.

Para verificar se esta afirmação é verdadeira, isto é, se C realmente é o melhor tratamento, é necessário proceder a um teste de comparações múltiplas.

```
dados=matrix(c(12,8,14,17,12,
13,9,20,16,15,
16,12,22,21,16,
7,5,6,11,10),
nr=5,byrow=FALSE,dimnames=list(1:5,c('A','B','C','D')))
?friedman.test
friedman.test(dados)
Tempo=c(12,8,14,17,12,13,9,20,16,15,16,12,22,21,16,7,5,6,11,10)
Tratamento=rep(c('A','B','C','D'),5)
boxplot(Tempo~Tratamento,xlab='Tratamento',ylab='Tempo(semanas)')
```

5.2.2 Pequenas amostras ($n \leq 4$)

No caso em que n é no máximo 4, utilizamos a Tabela de Friedman (Anexo - Tabela de valores críticos do teste Friedman).

Exemplo 5.4. Suponha que 3 grupos são observados sob 4 condições. Aqui $N = 3$ e $k = 4$. Cada grupo contém 4 indivíduos correspondentes, um associado a cada uma das 4 condições. Suponhamos que os escores obtidos em nosso estudo sejam da Tabela (5.5). Aplicar a prova de Friedman a esses dados para comparar as condições ao nível de 5%.

Tabela 5.5: Escores de três grupos correspondentes sob quatro condições.

Grupo	Condições			
	I	II	III	IV
A	9(4)	4(2)	1(1)	7(3)
B	6(3)	5(2)	2(1)	8(4)
C	9(4)	1(1)	2(2)	6(3)
Total	11	5	4	10

H_0 : Condição I = Condição II = Condição III = Condição IV;

O valor calculado de $H_{Fr} = \chi_r^2 = 7,4$, e da Tabela de Friedman, tem-se que $p = 0,033x2 = 0,06$ (no R este valor é 0,06018), desta forma não rejeitamos H_0 ao nível de 5% de significância. Concluimos que não existe diferença entre as condições.

```
dados=matrix(c(9,6,9,4,5,1,1,2,2,7,8,6),
nr=3,byrow=FALSE,
dimnames=list(c('A','B','C'),c('I','II','III','IV'))))
friedman.test(dados)
```

5.2.3 Comparações múltiplas

A finalidade das comparações múltiplas é complementar os resultados ou as conclusões obtidas pelo teste de Friedman, quando nele se rejeita H_0 ; procurando com isso, localizar as possíveis diferenças entre pares de tratamentos.

Obviamente, os recursos dessas comparações são limitados, pois elas nos permitem apenas confrontar os tratamentos dois a dois, não sendo possível, portanto, a comparação entre grupos de tratamentos.

Pequenas amostras

Para cada par de tratamentos ou grupos, a diferença mínima significativa é dada por:

$$d.m.s = |T_i - T_j|; \quad i \neq j \quad (5.3)$$

será comparada ao valor tabelado Δ_1 (Anexo - Tabela dms para comparações múltiplas de Friedman), ou seja,

$$P(|T_i - T_j| \geq \Delta_1) = \alpha.$$

Para os dados da Tabela (5.5), tem-se os totais dos ranks

$$\begin{aligned} T_1 &= 11, \\ T_2 &= 5, \\ T_3 &= 4 \text{ e} \\ T_4 &= 10, \end{aligned}$$

então,

$$\begin{aligned} |T_1 - T_2| &= 6, \\ |T_1 - T_3| &= 7, \\ |T_1 - T_4| &= 1, \\ |T_2 - T_3| &= 1, \\ |T_2 - T_4| &= 5 \text{ e} \\ |T_3 - T_4| &= 6. \end{aligned}$$

Na Tabela em Anexo, para $k = 4$ (número de tratamentos ou grupos) e $n = 3$ (número de repetições por tratamento/condição), com $\alpha = 0,049$, nos dá $\Delta_1 = 8$.

Concluí-se, assim, que, à taxa $\alpha = 0,049$ ou 5%, não foram constatadas diferenças significativas entre os pares de tratamentos.

Grandes amostras

Quando o número de blocos, ou o número de tratamentos, ou ambos ultrapassam os valores previstos na Tabela Anexa, obtemos a d.m.s. como se segue:

$$d.m.s. = Q \sqrt{\frac{nk(k+1)}{12}}. \quad (5.4)$$

Os valores de Q são tabelados (Anexo - Tabela de valores da Amplitude Q).

Exemplo 5.5. *Comparação entre tratamentos de dados hipotéticos.*

Admitamos um caso hipotético onde $k = 5$ tratamentos e $n = 16$ repetições por tratamento, com os totais dos ranks por tratamentos iguais a: $T_1 = 28$, $T_2 = 45$, $T_3 = 41$, $T_4 = 59$ e $T_5 = 67$.

Como não dispomos de tabela de Δ_1 , para $n = 16$, calculamos a d.m.s. pelo método apropriado às grandes amostras. Se admitirmos, por exemplo, um nível de significância, $\alpha = 5\%$, a Tabela Q nos dará:

$$Q = 3,858$$

e, conseqüentemente:

$$d.m.s. = 3,858 \sqrt{\frac{16(5)(6)}{12}} \simeq 24,4.$$

Por outro lado, temos:

$$\begin{array}{ll} |T_1 - T_2| = 17 & |T_2 - T_4| = 14 \\ |T_1 - T_3| = 13 & |T_2 - T_5| = 22 \\ |T_1 - T_4| = 31^* & |T_3 - T_4| = 18 \\ |T_1 - T_5| = 39^* & |T_3 - T_5| = 26^* \\ |T_2 - T_3| = 4 & |T_4 - T_5| = 8 \end{array}$$

Concluimos que, ao nível de 5%, $T_1 \neq T_4$; $T_1 \neq T_5$ e $T_3 \neq T_5$, já que são maiores que d.m.s. = 24,4.

$$\begin{aligned} T_1 &= 28^a, \\ T_2 &= 45^{abc}, \\ T_3 &= 41^{ab}, \\ T_4 &= 59^{bd} \text{ e} \\ T_5 &= 67^{cd}. \end{aligned}$$

Ao nível de 1% apenas $T_1 \neq T_4$ e $T_1 \neq T_5$, já que são maiores que d.m.s. = 29,11.

Para realizar comparações múltiplas para o Teste de Friedman, a função `friedmanmc` está implementada na livreria `pgirmess` do R.

```
dados=matrix(c(9,6,9,4,5,1,1,2,2,7,8,6),
nr=3,byrow=FALSE,
dimnames=list(c('A','B','C'),c('I','II','III','IV'))
friedman.test(dados)
library(pgirmess)
friedmanmc(dados)
```

Exemplo 5.6. *Concurso de beleza.*

Sete juízes atribuíram ordem a 5 finalistas em um concurso de beleza. As ordens foram de 1 para a melhor e 5 para a pior. Os resultados estão na Tabela (5.6). Pode a hipótese nula de

Tabela 5.6: *Notas de um concurso de beleza.*

Juiz	Finalista				
	A	B	C	D	E
1	5	2	1	3	4
2	5	1	2	3	4
3	3	1	2	4	5
4	2	3	4	1	5
5	3	1	2	5	4
6	4	1	2	3	5
7	4	2	3	1	5

atribuição aleatória das ordens ser rejeitada ao nível de 1% de significância? Realize o teste de comparações múltiplas em caso de rejeição de H_0 **R.** $\chi_r^2 = 15,54$, $\chi_{1\%,4}^2 = 13,277$ - **rejeita-se H_0 .** $B \neq E \Rightarrow |B - E| = 21$, **d.m.s. = 20,42.**

```
dados=matrix(c(5,2,1,3,4,
5,1,2,3,4,
3,1,2,4,5,
2,3,4,1,5,
3,1,2,5,4,
4,1,2,3,5,
4,2,3,1,5),
nr=5,byrow=TRUE,
dimnames=list(c('J1','J2','J3','J4','J5','J6','J7'),
c('A','B','C','D','E'))
friedman.test(dados)
friedmanmc(dados,probs=0.01)
```


Exercício 5.3. Num ensaio sobre adubação nitrogenada de alface, realizado pelo Prof. Salim Simão, da ESALQ, foram considerados os seguintes tratamentos:

Trat 1: Testemunha (ou controle); Trat 2: 5g de salitre/10 d'água; Trat 3: 10g de salitre/10 d'água; Trat 4: 20g de salitre/10 d'água.

A adubação básica foi NPK e a adubação nitrogenada referida foi feita em cobertura. Os resultados de produção (peso de 12 pés, em gramas) foram os apresentados na Tabela (5.7). Verifique pelo teste de Friedman, se houve resposta à adubação nitrogenada, ao nível de 5%. Realize o teste de comparações múltiplas em caso de rejeição de H_0 .

Tabela 5.7: Tratamento com adubação nitrogenada.

Bloco	Tratamento			
	1	2	3	4
1	3,640	4,200	4,700	5,300
2	4,890	4,550	6,020	5,900
3	4,800	5,320	5,250	5,150
4	4,460	5,500	5,580	5,560

Exemplo 5.7. Comparando variedades de cana-de-açúcar.

Num ensaio sobre competição de variedades de cana-de-açúcar, foram obtidos os seguintes resultados de produção em t/ha. Aplique o teste de Friedman aos resultados obtidos, adotando 5%

Tabela 5.8: Produção de cana-de-açúcar em t/ha.

Variedade	Bloco					
	I	II	III	IV	V	VI
1	110,6	119,5	120,1	105,3	130,8	138,1
2	116,7	128,4	131,5	114,8	146,8	155,5
3	140,3	150,0	150,9	114,7	153,9	156,9
4	143,4	153,8	151,5	144,1	154,6	159,3

de significância. Realize o teste de comparações múltiplas em caso de rejeição de H_0 .

$T_1 = 6$; $T_2 = 13$; $T_3 = 17$ e $T_4 = 24$;

$H_{Fr} = 17$; $\chi^2_{51\%,3} = 7,815$ - rejeita-se H_0 . $V_1 \neq V_4$, d.m.s. = 12,74.

Exercício 5.4. *Em um estudo do efeito de três padrões diferentes de reforço sobre a extensão do aprendizado discriminante entre ratos (Grosslight e Radlow, 1956), treinaram-se três amostras correspondentes de 18 ratos sob três tipos de reforço:*

RT: Reforço total; RPN: Reforço Parcial - em que cada sequência de tentativas terminava com uma tentativa não recompensada; RPC: Reforço Parcial - em que cada sequência de tentativas terminava com uma tentativa recompensada;

Ao cabo desse treinamento, mediu-se a extensão do aprendizado pela rapidez com que os diversos ratos adquiriam um hábito. Teste a hipótese H_0 : Os diversos tipos de reforço não tem efeito diferencial, considerando um nível de significância de 5%. Realize o teste de comparações múltiplas em caso de rejeição de H_0 .

Tabela 5.9: *Postos de 18 grupos correspondentes no estudo de transferência de aprendizado após treinamento sob três condições diferentes de reforço.*

Grupo	Tipo de Reforço		
	RT	RPN	RPC
1	1	3	2
2	2	3	1
3	1	3	2
4	1	2	3
5	3	1	2
6	2	3	1
7	3	2	1
8	1	3	2
9	3	1	2
10	3	1	2
11	2	3	1
12	2	3	1
13	3	2	1
14	2	3	1
15	2.5	2.5	1
16	3	2	1
17	3	2	1
18	2	3	1

Exercício 5.5. *Uma amostra de 14 crianças, constituinte de um grupo de pesquisa, foi mensurada em três momentos distintos: P_1 , P_2 e P_3 de sua participação numa tarefa de classe cujo objetivo era tomar os alunos mais dependentes uns dos outros na consecução de uma promoção curricular. Resultaram os seguintes escores-identificação (escores mais altos indicam maior "harmonia grupal"). Teste a diferença significativa em nível de 5% entre as amostras nos diferentes períodos de observação. Realize o teste de comparações múltiplas em caso de rejeição de H_0 .*

Tabela 5.10: *Escores de harmonia grupal em três períodos de observação em alunos.*

P_2	P_3	P_3
40	62	75
51	51	53
45	60	62
41	43	51
30	49	52
42	45	46
65	73	62
68	66	68
50	57	55
61	63	69
55	43	45
48	46	45
70	67	68
50	61	67

Capítulo 6

Testes aplicáveis a k amostras - Dados independentes

6.1 Teste de Kruskal-Wallis para k amostras

O teste de Kruskal-Wallis foi criado como um substituto ao teste F na análise paramétrica.

Ele é utilizado para que se verifique o contraste entre **k amostras independentes**, portanto uma extensão do teste Mann-Whitney.

Requisitos:

- Comparação entre 3 ou mais grupos independentes;
- Dados ordinais (que possam ser ordenados);
- Dados contínuos, onde não há suposição de normalidade.

6.1.1 Grandes amostras ($n_i > 6$)

Os valores obtidos nas diversas amostras diferem entre si e portanto, será uma maneira de verificar se estas diferenças são devidas ao acaso ou se as amostras provêm de populações diferentes.

Da mesma forma que nos outros testes, serão consideradas as hipóteses nula (H_0) e alternativa (H_a), isto é:

H_0 : “não há diferença significativa entre os tratamentos”.

H_a : “há diferença significativa entre os tratamentos”.

Quanto à metodologia usada no teste de Kruskal-Wallis, a expressão é definida por:

$$H_{K-W} = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{T_j^2}{n_j} - 3(N+1) \text{ tal que } H \sim \chi_{\alpha\%, (k-1)}^2 \quad (6.1)$$

em que:

k : n° de tratamentos;

T_j : soma das ordens (todos os dados são considerados simultaneamente) atribuídas ao tratamento $j = 1, 2, \dots, k$;

n_j : o número de observações em cada tratamento;

N : o total de observações em todos os tratamentos $N = \sum_{j=1}^k n_j$;

α : nível de significância.

6.1.2 Pequenas amostras ($n_i \leq 6$)

Quando o número de observações é pequeno ($n_i \leq 6$), faz-se uso da Tabela em Anexo de valores críticos de Kruskal-Wallis, considerando um número $k = 3$ de tratamentos.

Exemplo 6.1. Pretende-se comparar os grupos A, B e C dispostos na Tabela (6.1). Sabe-se que os dados são provenientes de grupos independentes, e os valores são referentes à escores de ansiedade.

Tabela 6.1: Ranks de níveis de ansiedade de três grupos distintos de pacientes.

Postos		
A	B	C
4	2	7
9	8	13
3	10	14
1	11	12
5	6	-
$T_A = 22$	$T_B = 37$	$T_C = 46$

Verifique ao nível de 5% se existe diferença nos escores de ansiedade entre os grupos.

Utilizando a equação (6.1), obtém-se facilmente H_{K-S} :

$$H_{K-W} = \frac{12}{14(15)} \left(\frac{22^2}{5} + \frac{37^2}{5} + \frac{46^2}{4} \right) - 3(15) = 6,41 \quad (p = 0,04065).$$

Assim, com $H_{K-S} = 6,4$ e $n_1 = n_2 = 5$ e $n_3 = 4$, segue que $p < 0,05$, já que para $\alpha = 5\%$, $h = 5,643$ (valor tabelado). desta forma opta-se pela rejeição de H_0 . Pode-se afirmar que ao nível de 5% de significância existem evidências de que pelo menos um par de grupos diferem. O próximo passo, seria realizar um teste de comparações múltiplas para saber qual par causou a rejeição de H_0 .

```
x=c(4,9,3,1,5)
y=c(2,8,10,11,6)
z=c(7,13,14,12)
?kruskal.test
kruskal.test(list(x,y,z))
```

Exemplo 6.2. *Análise de tempo de sobrevida de câncer.*

Analisar o tempo de sobrevida (tempo livre do câncer até reincidência), em semanas, de pacientes atendidos na clínica de abdômen.

Tabela 6.2: *Tempo de sobrevida (semanas).*

Radioterapia ($n_1 = 7$)	Quimioterapia ($n_2 = 8$)	Cirurgia ($n_3 = 8$)
17 (11)	20 (12)	32 (17)
14 (9)	5 (3)	35 (20)
4 (2)	9 (6)	26 (15)
8 (5)	13 (8)	34 (18,5)
29 (16)	34 (18,5)	21 (13)
6 (4)	2 (1)	45 (21)
15 (10)	11 (7)	50 (23)
-	22 (14)	47 (22)
$T_1 = 57,0$	$T_2 = 69,5$	$T_3 = 149,5$

Questões: Há diferença significativa entre os tempos de sobrevivência? Qual o tratamento recomendado baseado no tempo de sobrevida?

Independentemente do número de observações em cada grupo e utilizando os números naturais, procedemos à ordenação dos valores. Assim os valores acima receberiam a seguinte numeração, conforme os valores já entre parênteses.

Então:

$$\begin{aligned}
 H_{K-W} &= \frac{12}{N(N+1)} \sum_{j=1}^k \frac{T_j^2}{n_j} - 3(N+1) \\
 &= \frac{12}{23(23+1)} \sum_{i=1}^3 \frac{T_j^2}{n_j} - 3(23+1) \\
 &= \frac{1}{46} \left[\frac{57^2}{7} + \frac{69,5^2}{8} + \frac{149,5^2}{8} \right] - 72 \\
 &= \frac{1}{46} (3861,7) - 72 \\
 &= 11,95 \text{ (p-valor : 0,0025).}
 \end{aligned}$$

Como $H_{K-W} \sim \chi_{\alpha\%,(k-1)}^2$, então $H \sim \chi_{5\%,(3-1)}^2 \Rightarrow \chi_{5\%,2}^2 = 5,99$ e $\chi_{1\%,2}^2 = 9,21$, portanto, considerando que o valor encontrado de $H_{K-W} = 11,95$ é maior do que os valores tabelados, tanto para $\alpha = 5\%$ como para $\alpha = 1\%$, concluímos pela rejeição de H_0 e consequentemente, pela indicação de que o tratamento cirúrgico se destaca dos demais, pois apresenta maiores valores aos tempos de sobrevida, sendo de 34,5 (mediana) o tempo até reincidência do câncer.

```

x=c(17,14,4,8,29,6,15)
y=c(20,5,9,13,34,2,11,22)
z=c(32,35,26,34,21,45,50,47)
tempo=c(x,y,z)
rank(tempo)
tratamento=factor(rep(1:3,c(length(x),length(y),length(z))),
labels=c('Radioterapia','Quimioterapia','Cirurgia'))
boxplot(tempo~tratamento,xlab='Tratamento',ylab='Tempo(semanas)')
kruskal.test(tempo,tratamento)

```

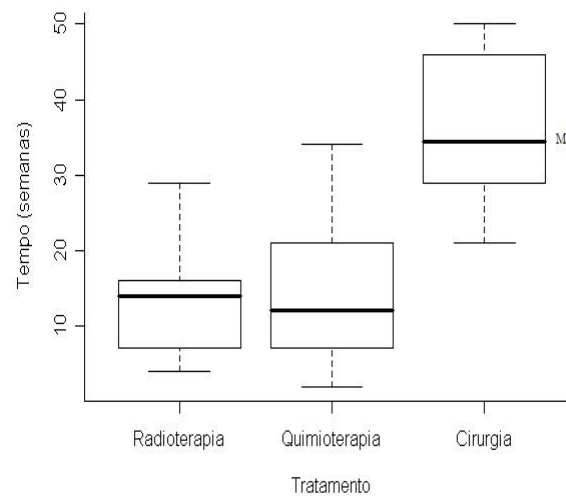


Figura 6.1: Tempos de sobrevivência (semanas) de três grupos de pacientes submetidos a tratamentos distintos.

Exercício 6.1. Considere os seguintes dados disponíveis na Tabela 6.3. Teste ao nível de 5% a hipótese de igualdade entre tratamentos. **R.** $H_{K-W} = 0,06 < 5,099 = \chi^2_{5\%,(2)}$; $p\text{-valor} = 0,9694$

Tabela 6.3: Dados de três tratamentos distintos.

T	TP	TAVP
9,54	11,48	12,25
14,90	16,84	17,61
55,81	58,93	64,41
18,74	20,45	21,19
18,81	20,05	20,64
40,27	38,75	38,34
28,59	29,17	29,61
-	0,79	0,81
0,53	0,59	0,63
325,27	408,89	420,31
125,30	142,44	147,93
2,55	2,91	3,01
$T_T=196$	$T_{TP}=211$	$T_{TAVP}=223$

Observação: Comparando com a mais poderosa prova paramétrica, a prova F, e em condições tais que as suposições básicas da prova F sejam satisfeitas, a prova de Kruskal-Wallis tem eficiência assintótica de $3/\pi = 95,5\%$ (Andrews, 1954). A prova de Kruskal-Wallis é mais eficiente do que a extensão da prova da mediana, porque utiliza mais as informações contidas nas observações convertendo os escores em postos ao invés de simplesmente dicotomizá-las em “acima” ou “abaixo”.

Exercício 6.2. Considere os seguintes dados disponíveis na Tabela 6.4. Teste ao nível de 5% a hipótese de igualdade entre tratamentos, agora para cada variável: PB, EE e Cinzas.

Tabela 6.4: Dados de três tratamentos distintos considerando as variáveis PB, EE e Cinzas.

Tratamento	PB	EE	Cinzas
T	7,75	2,17	6,01
T	7,81	1,97	6,16
T	7,69	2,35	6,18
TP	11,47	1,80	6,51
TP	11,63	1,85	6,63
TP	11,81	1,89	6,79
TAVP	12,88	2,29	7,25
TAVP	13,56	2,34	8,51
TAVP	12,78	2,44	7,52

R. $p\text{-valor}=0,02, 0,051 \text{ e } 0,027$, respectivamente para PB, EE e Cinzas.

No caso de **mais de 20% ocorrências de empates** entre dois ou mais escores, atribui-se a cada um deles a média dos postos respectivos e a estatística teste passa a ser:

$$H_{K-W}^{empates} = \frac{\frac{12}{N(N+1)} \sum_{j=1}^k \frac{T_j^2}{n_j} - 3(N+1)}{1 - \frac{\sum_{i=1}^g T_i^*}{N^3 - N}} \quad \text{tal que } H \sim \chi_{\alpha\%, (k-1)}^2 \quad (6.2)$$

onde $\mathbf{T}^* = \mathbf{t}^3 - \mathbf{t}$ (t sendo o número de observações empatadas em um grupo de escores empatados).

$\sum_{i=1}^g T_i^*$: representa o somatório sobre todos os grupos de empates.

g : número de grupos de empates.

Exemplo 6.3. Efeito do tamanho da cria no peso ao nascer de leitões.

Um pesquisador determinou os pesos (kg) ao nascer, dos 56 leitõezinhos de oito crias diferentes, para comprovar se esse peso é afetado pelo tamanho da cria.

O cálculo do teste Kruskal-Wallis para os dados da Tabela (6.5) sem considerar empates tem como resultado: $H_{K-W} = 18,464$ onde H_0 é rejeitada ao nível de 5%, já que $\chi_{(5\%;8-1)}^2 = \chi_{(5\%;7)}^2 = 14,07$.

Para introduzir a correção para empates, devemos primeiro determinar quantos grupos de empate ocorrem, e quantos escores empatados há em cada grupo. O primeiro empate ocorre entre dois leitõezinhos da cria 7 (ambos pesaram 1,2 kg). Ambos receberam o posto 2,5. Logo, t = número de observações empatadas = 2. Para esta ocorrência

$$T^* = t^3 - t = 8 - 2 = 6.$$

O próximo empate ocorreu entre 4 leitõezinhos, aos quais se atribuiu o posto comum 8,5. Aqui, $t = 4$ e $T^* = 60 = 4^3 - 4$.

Prosseguindo dessa maneira com os dados da Tabela (6.6), verificamos a ocorrência de 13 grupos de empates. Podemos contar o número de observações em cada grupo empatado para determinar os diversos valores de t e calcular o valor de T em cada caso. Eis o resultado dessa contagem:

t	2	4	2	2	4	5	4	4	3	7	6	2	2	Total
T^*	6	60	6	6	60	120	60	60	24	336	210	6	6	$\sum_{i=1}^{13} T_i^* = 960$

Tabela 6.5: *Peso (kg) ao nascimento de baco­rinhos de crias diferentes.*

Tamanho da cria							
1	2	3	4	5	6	7	8
2	3,5	3,3	3,2	2,6	3,1	2,6	2,5
2,8	2,8	3,6	3,3	2,6	2,9	2,2	2,4
3,3	3,2	2,6	3,2	2,9	3,1	2,2	3
3,2	3,5	3,1	2,9	2	2,5	2,5	1,5
4,4	2,3	3,2	3,3	2		1,2	
3,6	2,4	3,3	2,5	2,1		1,2	
1,9	2	2,9	2,6				
3,3	1,6	3,4	2,8				
2,8		3,2					
1,1		3,2					
n ₁ =10	8	10	8	6	4	6	4

Tabela 6.6: *Ranks dos peso ao nascimento de baco­rinhos de crias diferentes.*

Tamanho da cria							
1	2	3	4	5	6	7	8
8,5	52,5	47,5	41	23	36	23	18,5
27,5	27,5	54,5	47,5	23	31,5	12,5	15,5
47,5	41	23	41	31,5	36	12,5	31
41	52,5	36	31,5	8,5	18,5	18,5	4
56	14	41	47,5	8,5		2,5	
54,5	15,5	47,5	18,5	11		2,5	
6	8,5	31,5	23				
47,5	5	51	27,5				
27,5		41					
1		41					
T ₁ =317	216,5	414	277,5	105,5	122	71,5	69

Utilizando a equação (6.2), a correção (termo no denominador) será:

$$1 - \frac{960}{56^3 - 56} = 0,9945$$

assim o valor corrigido para empates será $H_{K-W}^{empates} = 18,566$ ($p = 0,009663$), cujo valor tabelado $\chi^2_{(5\%;7)} = 14,07$, nos faz concluir que H_0 será rejeitada a 5% de significância, logo o peso dos leitões, ao nascer, varia com o tamanho da cria.

```

c1=c(2,2.8,3.3,3.2,4.4,3.6,1.9,3.3,2.8,1.1)
c2=c(3.5,2.8,3.2,3.5,2.3,2.4,2,1.6)
c3=c(3.3,3.6,2.6,3.1,3.2,3.3,2.9,3.4,3.2,3.2)
c4=c(3.2,3.3,3.2,2.9,3.3,2.5,2.6,2.8)
c5=c(2.6,2.6,2.9,2,2,2.1)
c6=c(3.1,2.9,3.1,2.5)
c7=c(2.6,2.2,2.2,2.5,1.2,1.2)
c8=c(2.5,2.4,3,1.5)
kruskal.test(list(c1,c2,c3,c4,c5,c6,c7,c8))

```

Exercício 6.3. Com o objetivo de comparar três tratamentos distintos (*A*, *B* e *C*), 25 bovinos foram separados aleatoriamente e os níveis de ganho de peso (kg) foram anotados após o experimento. Utilize um nível de 5% para testar a hipótese de igualdade de tratamentos.

<i>A</i> :	15	15	25	25	31	36	50	50
<i>B</i> :	6	6	7	10	10	20	20	30
<i>C</i> :	10	10	17	25	25	28	35	-

R. H = 7,37; p = 0,025; A × B : p = 0,02

6.1.3 Comparações múltiplas

Evidentemente, quando rejeitamos H_0 , estamos admitindo que, pelo menos dois tratamentos diferem entre si. Assim, a finalidade do método é localizar, quando existem, as diferenças significativas entre pares de tratamentos.

Pequenas amostras ($n_i \leq 4$)

Para cada par $i \neq j$ tratamentos, determinamos a *d.m.s.*

$$d.m.s. = |T_i - T_j|; \quad i \neq j \quad (6.3)$$

será comparada ao valor tabelado Δ_1 (Anexo - Tabela dms para comparações múltiplas de Kruskal-Wallis), onde $n_1 = n_2 = \dots = n_k = n$, ou seja,

$$P(|T_i - T_j| \geq \Delta_1) = \alpha.$$

Para o caso de diferentes números de repetições entre os k tratamentos, tem-se:

$$d.m.s. = \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} h \quad (6.4)$$

onde h : é o limite tabelado do teste de Kruskal-Wallis. Neste caso consideramos as diferenças

$$|\bar{T}_i - \bar{T}_j|$$

onde

$$\bar{T}_i = \frac{T_i}{n_i} \quad \text{e} \quad \bar{T}_j = \frac{T_j}{n_j}.$$

Para melhor ilustração dos dois casos, vejamos o seguinte exemplo:

Exemplo 6.4. *Qualidade de vinho.*

Em uma pesquisa sobre qualidade de vinho, foram provados três tipos, por 4 degustadores. Cada degustador provou doze amostras (quatro de cada tipo) e atribuiu a cada uma delas uma nota de zero a dez. As médias das notas atribuídas pelos 4 degustadores a cada uma das amostras foram:

Tabela 6.7: *Notas de degustação.*

Tipo		
1	2	3
5,0 (1)	8,3 (7)	9,2 (11)
6,7 (2)	9,3 (12)	8,7 (9)
7,0 (3)	8,6 (8)	7,3 (5)
6,8 (4)	9,0 (10)	8,2 (6)
$T_1=10$	$T_2=37$	$T_3=31$

Verificar se há preferência dos degustadores por algum tipo de vinho, ao nível de 5% de significância.

Se o teste de Kruskal-Wallis é utilizado, teremos para $n = 4$ e $k = 3$:

$$T_1 = 10; T_2 = 37 \text{ e } T_3 = 31,$$

consequentemente: $H = 7,731$ já que $\alpha^* = 0,007$ (valor tabelado).

Pela valor tabelado, o *d.m.s.*, no qual rejeitaríamos $H_0 : T_1 = T_2 = T_3$, é $\alpha = 0,007 = 0,7\%$. Podemos então afirmar que pelo menos dois tratamentos diferem entre si.

Por outro lado, organizamos as diferenças:

$$\begin{aligned} |T_1 - T_2| &= 27^*, \\ |T_1 - T_3| &= 21, \\ |T_2 - T_3| &= 6. \end{aligned}$$

O valor tabelado nos dá, à taxa $\alpha = 0,011 = 1,1\%$, $\Delta_1 = 27$ ou ainda

$$P(|T_i - T_j| \geq 27) = 0,011 = 1,1\%.$$

Observe que para $\alpha = 5\%$, $\Delta_1 \simeq 24$. E, em nível de 5% $T_1 \neq T_3$ também !

Concluimos ao nível de 1,1% que $T_1 \neq T_2$, isto é, os degustadores preferem o tipo “2” ao tipo “1” já que $T_2 = 37$ e $T_1 = 10$.

Exemplo 6.5. *Considerando que $T_1 = 54$, $T_2 = 33$ e $T_3 = 18$, e que $n_1 = 5$, $n_2 = 5$ e $n_3 = 4$ podemos realizar as comparações múltiplas do seguinte modo:*

Como $\bar{T}_1 = 10,8$, $\bar{T}_2 = 6,6$ e $\bar{T}_3 = 4,5$, então, neste caso, sendo que os tratamentos não são igualmente repetidos, estruturamos:

$$\begin{aligned} d.m.s._{ij} &= \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right) h} \\ &= \sqrt{\frac{14(15)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} 5,643 \\ &= 9,94 \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \end{aligned}$$

onde $h = 5,643$ para um $\alpha = 5\%$, da seguinte forma:

(i, j)	$ \bar{T}_i - \bar{T}_j $		$d.m.s.$
(1,2)	4,2	<	6,29
(1,3)	6,3	<	6,67
(2,3)	2,1	<	6,67

Como nenhuma das diferenças de ranks médios é maior do que a d.m.s., concluímos que:

$$T_1 = T_2 = T_3.$$

Grandes amostras

Determinamos as diferenças $|\bar{T}_i - \bar{T}_j|$ e, ao nível de significância α , as d.m.s. segundo as quais $T_i \neq T_j$ são:

- para $n_1 = n_2 = \dots = n_k = n$ (k tratamentos)

$$d.m.s. = Q \sqrt{\frac{k(kn+1)}{12}} = Q \sqrt{\frac{k(N+1)}{12}} \quad (6.5)$$

e a Tabela da Amplitude Q nos fornece os valores de Q .

Exemplo 6.6. Estabeleça as comparações múltiplas considerando os seguintes dados provenientes de um estudo, onde se deseja verificar se existe diferença entre quatro rações quanto ao ganho de peso (em kg) de suínos.

$$\begin{array}{cc} \bar{T}_1 = 8,8; & \bar{T}_2 = 16,5 \\ \bar{T}_3 = 11,1 & \bar{T}_4 = 5,6 \\ \hline n_1 = n_2 = n_3 = n_4 = 5 \end{array}$$

Deste modo pode-se formar as seguintes diferenças entre os ranks médios:

(i, j)	$ \bar{T}_i - \bar{T}_j $
(1,2)	7,7
(1,3)	2,3
(1,4)	3,2
(2,3)	5,4
(2,4)	10,9*
(3,4)	5,5

Considerando $\alpha = 5\%$, para $k = 4$, da Tabela da Amplitude Q temos $Q = 3,633$, logo

$$d.m.s. = 3,633 \sqrt{\frac{4[4(5)+1]}{12}} = 9,6.$$

Concluí-se que as rações 2 e 4 diferem quanto ao ganho de peso, ao nível de 5%, pois $|\bar{T}_2 - \bar{T}_4| = 10,9 > 9,6 = d.m.s.$

- no caso de tratamentos não igualmente repetidos:

$$d.m.s.ij = \left| z_{\frac{\alpha}{k(k-1)}} \right| \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \quad (6.6)$$

onde $z_{\frac{\alpha}{k(k-1)}}$ é um limite superior da distribuição normal.

Exemplo 6.7. Considere que no exemplo anterior $n_1 = n_2 = 6$, $n_3 = 5$ e $n_4 = 8$ ($N = 25$), onde $T_1 = 50$; $T_2 = 97$; $T_3 = 45$ e $T_4 = 33$. Estabeleça as comparações múltiplas ao nível de 5%.

$$\begin{aligned} \bar{T}_1 &= 8,33; & \bar{T}_2 &= 16,17 \\ \bar{T}_3 &= 9,0 & \bar{T}_4 &= 4,13 \end{aligned}$$

Deste modo pode-se formar as seguintes diferenças, considerando

$$z_{\frac{\alpha}{k(k-1)}} = z_{\frac{0,05}{4(4-1)}} = z_{0,0041} = -2,64$$

(i, j)	$ \bar{T}_i - \bar{T}_j $		$d.m.s.$	$= \left z_{\frac{\alpha}{k(k-1)}} \right \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$
(1,2)	7,84	<	11,22	$= 2,64 \sqrt{54,17 \left(\frac{1}{6} + \frac{1}{6} \right)}$
(1,3)	0,67	<	11,77	$= 2,64 \sqrt{54,17 \left(\frac{1}{6} + \frac{1}{5} \right)}$
(1,4)	4,2	<	10,49	$= 2,64 \sqrt{54,17 \left(\frac{1}{6} + \frac{1}{8} \right)}$
(2,3)	7,17	<	11,77	$= 2,64 \sqrt{54,17 \left(\frac{1}{6} + \frac{1}{5} \right)}$
(2,4)	12,04*	>	10,49	$= 2,64 \sqrt{54,17 \left(\frac{1}{6} + \frac{1}{8} \right)}$
(3,4)	4,87	<	11,08	$= 2,64 \sqrt{54,17 \left(\frac{1}{5} + \frac{1}{8} \right)}$

onde as $d.m.s.$ foram obtidas por meio de (6.6).

Concluí-se que as rações 2 e 4 diferem quanto ao ganho de peso, ao nível de 5%.

Para realizar comparações múltiplas para o Teste de Kruskal-Wallis, a função `kruskalmc` está implementada na livreria `pgirmess` do R.

```
resp=c(0.44,0.44,0.54,0.32,0.21,0.28,0.7,0.77,  
0.48,0.64,0.71,0.75,0.8,0.76,0.34,0.80,0.73,0.8)  
categ=as.factor(rep(c('A','B','C'),times=1,each=6))  
boxplot(resp~categ)  
kruskal.test(resp,categ)  
library(pgirmess)  
kruskalmc(resp, categ)  
kruskalmc(resp, categ, probs=0.01)  
kruskalmc(resp, categ, cont='one-tailed')  
kruskalmc(resp, categ, cont='two-tailed')
```

Considere os dados da Tabela (6.2)

```
kruskal.test(tempo,tratamento)  
kruskalmc(tempo,tratamento,probs=0.05,cont=NULL)
```

Exercício 6.4. Num experimento de competição de variedades de cana-de-açúcar, foram registrados os rendimentos, em kg, de açúcar por tonelada de cana, obtendo-se os seguintes resultados para três tipos:

Co 419	Co 421	Co 678
105,3	121,6	123,8
99,9	122,0	122,9
104,8	124,9	118,5
105,1	123,3	-
100,2	-	-

Estabeleça as comparações múltiplas entre os tratamentos, ao nível de 5% de significância.

R. $H_{K-W} = 8,12$ ($p < 0,01723$); $T_1 \neq T_2$.

```
x=c(105.3,99.9,104.8,105.1,100.2)
y=c(121.6,122,124.9,123.3)
z=c(123.8,122.9,118.5)
rend=c(x,y,z)
variedade=factor(rep(1:3,c(length(x),length(y),length(z))),
labels=c('Co419','Co421','Co678'))
boxplot(rend~variedade)
kruskal.test(rend,variedade)
kruskalmc(rend,variedade,probs=0.05,cont=NULL)
```

Exercício 6.5. Dados de um delineamento completamente aleatorizado: três métodos diferentes de cultivo de milho foram aleatoriamente atribuídos a um grande número de lotes de terra, e a produção por pedaço de terra (campo) foi calculada para cada lote. A hipótese nula pode ser formulada como segue:

H_0 : Os métodos são equivalentes.

As observações são apresentadas na Tabela (6.8). Teste H_0 ao nível de 5% de significância. **R.**

Tabela 6.8: Métodos para análise de cultivo de milho.

Método		
1	2	3
83	101	78
91	100	82
94	91	81
89	93	77
89	96	79
96	95	81
91	94	80
92	-	81
90	-	-

$H_{K-W} = 18,02 > 5,99 = X_{5\%,2}^2$ ($p = 0,00012$) "Rejeita-se H_0 " - $M_1 \neq M_2$.

Exercício 6.6. Num experimento sobre a eficiência de inseticidas no feijoeiro, foram obtidos os seguintes resultados de produção (em kg/ha): Verifique, ao nível de 5% de significância se existe

Tabela 6.9: Tratamento na produção de feijão.

Tratamento		
Testemunha	Ekatin	EPN
1451	1451	1534
1534	1575	1658
1493	1534	1700
1285	1493	1700
5763	6053	6592
1441	1513	1648

diferença entre os tratamentos.

R. $H_{K-W} = 5,84$ ($p = 0,0584$). "Não rejeita-se H_0 "

Exercício 6.7. Aplique um teste estatístico específico (considerando não normalidade nos dados) para verificar se há diferença significativa entre as seguintes amostras de dados independentes ao nível de 5% significância.

X_1	X_2	X_3
125	100	95
100	99	90
122	105	86
127	103	96
115	116	88
129	98	89
130		

R. $H_{K-W} = 14,37$ ($p = 0,0008$). "Rejeita-se H_0 " - $X_1 \neq X_3$

Exercício 6.8. Numa pesquisa realizada pelo Dr. Aldo Duvilio Ometto, do Departamento de engenharia Rural da ESALQ, sobre **Influência do tamanho do disco na razão**. Foram provados três diâmetros distintos (em cm). Os resultados parciais, em kgf, para solo argiloso foram:

D_1	D_2	D_3
2.212,8 (13)	2.195,2 (12)	1.770,3 (4)
2.025,3 (9)	2.031,5 (11)	1.800,0 (5)
1.989,0 (8)	1.876,5 (7)	1.852,8 (6)
2.232,8 (14)	1.750,3 (2)	1.769,0 (3)
2.027,8 (10)	1.060,3 (1)	

Verificar se há diferença significativa de comportamento dos tamanhos dos discos ao nível de 5% significância.

R. $H_{K-W} = 5,4$ ($p = 0,0672$). "Não rejeita-se H_0 "

Capítulo 7

Análise de Correlação

A correlação entre duas variáveis pode ser verificada utilizando o teste de Pearson, caso há a suposição de normalidade nos dados.

No caso em que a suposição de normalidade não é considerada ou a característica dos próprios dados pressupõe isto (escala no mínimo ordinal), então pode-se fazer uso de estatísticas alternativas, como: Teste de correlação por postos de Spearman ou de Kendall.

7.1 Teste de Spearman

De todas as estatísticas baseadas em postos, o coeficiente de correlação por postos de Spearman, ρ (rho), foi a que surgiu primeiro e é talvez a mais conhecida hoje. Esta estatística será aqui representada por r_s . É uma medida de associação que exige que ambas as variáveis se apresentem em escala de mensuração pelo menos ordinal, de modo que os objetivos ou indivíduos em estudo possam ser dispostos por postos em duas séries ordenadas.

Para calcular o coeficiente de correlação por postos para um conjunto de pares de dados, inicialmente ordena-se os dados da amostra X do mais baixo para o mais alto, ou vice-versa; em seguida, os de Y da mesma maneira; determina-se a soma dos quadrados das diferenças $d_i = x_i - y_i$, entre os postos de X e de Y, e utiliza-se a seguinte regra:

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n} \quad (7.1)$$

onde n: é o número de pares.

Quando há empates, procede-se o cálculo pela média dos mesmos.

Exemplo 7.1. *Correlação entre nº de horas de estudo (X) e nota no exame (Y).*

Dispondo os postos, tem-se a Tabela 7.1, em que a correlação de Spearman é dada por:

Tabela 7.1: <i>Horas de estudo versus nota no exame.</i>					
X	Y	X' = Posto de X	Y' = Posto de Y	d = X' - Y'	d ²
9	56	5	4	1,0	1,00
5	44	2,5	2	0,5	0,25
11	79	7	8	-1,0	1,00
13	72	8	7	1,0	1,00
10	70	6	6	0	0,00
5	54	2,5	3	-0,5	0,25
18	94	10	10	0,0	0,00
15	85	9	9	0,0	0,00
2	33	1	1	0,0	0,00
8	65	4	5	-1,0	1,00

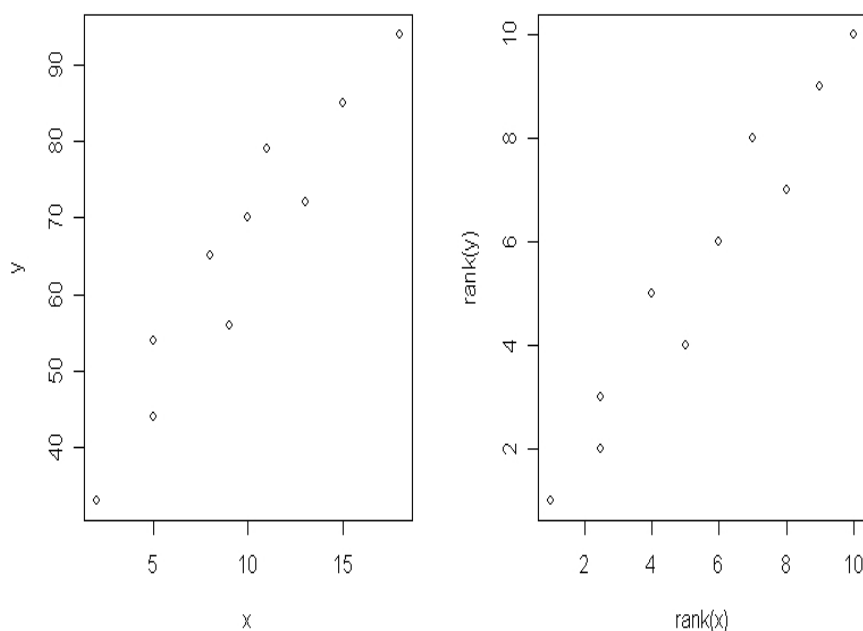


Figura 7.1: *Correlação entre X e Y.*

Então, pela expressão (7.1) tem-se que:

$$r_s = 1 - \frac{6(4, 50)}{10^3 - 10} \simeq 0,9727.$$

A análise é similar ao de Pearson (paramétrico), isto é, como $-1 \leq r_s \leq +1$, então quanto mais próximo de 1 o valor de r_s estiver, mais correlacionadas as variáveis X e Y estarão.

A principal vantagem da utilização de r_s , é que pode-se testar a hipótese nula de não haver qualquer correlação, sem ter de fazer quaisquer suposições sobre as populações das quais provem

as amostras.

$$H_0 : \rho = 0 \text{ "Não há correlação entre X e Y"}$$

$$H_a : \rho \neq 0.$$

H_0 será rejeitada, ao nível $\alpha\%$ de significância, se $r_s > r_{s\alpha\%;n}$ (para $n \leq 30$ podemos utilizar a Tabela de valores críticos da correlação de Spearman do Anexo). Considerando o exemplo, como $r_s = 0,97 > 0,564 = r_{s5\%;10}$, rejeita-se H_0 .

Sob hipótese nula de não haver qualquer correlação - na realidade, a hipótese nula é que os x 's e os y 's sejam aleatoriamente emparelhados - a distribuição amostral de r_s tem média 0 e desvio-padrão $\sigma_{r_s} = \frac{1}{\sqrt{n-1}}$.

Essa distribuição pode ser aproximada por uma distribuição normal mesmo para valores relativamente pequenos de n , baseamos o teste da hipótese nula na estatística que tem aproximadamente a distribuição normal padronizada.

```
x=c(9,5,11,13,10,5,18,15,2,8)
y=c(56,44,79,72,70,54,94,85,33,65)
par(mfrow=c(1,2))
plot(x,y)
plot(rank(x),rank(y))
summary(lm(y~x))
?cor.test
cor.test(x,y,method='pearson')
cor.test(x,y,method='spearman')
```

7.1.1 Grandes amostras ($n \geq 10$)

Pode-se testar a significância de r_s por:

$$z_{r_s} = \frac{r_s - 0}{\frac{1}{\sqrt{n-1}}} = r_s \sqrt{n-1} \sim N(0, 1). \quad (7.2)$$

Exemplo 7.2. Do exemplo anterior, sabe-se que $n = 10$ e $r_s = 0,97$, portanto ao nível de 1%, por exemplo, pode-se testar

$$H_0 : \rho = 0 \text{ versus } \rho \neq 0.$$

Utilizando (7.2), segue que:

$$z_{r_s} = r_s \sqrt{n-1} = 0,97 \sqrt{10-1} = 2,91,$$

e como $z_{0,5\%} = \pm 2,575$, z_{r_s} está na região de rejeição de H_0 ($z_{r_s} > z_{\alpha/2}$). Conclui-se que $\rho \neq 0$ (correlação significativa), isto é, que há uma relação entre o tempo de estudo e as notas obtidas, para a população que originou a amostra.

7.1.2 Pequenas amostras ($n < 10$)

Pode-se testar a significância de r_s por meio da equação a seguir, que segue a distribuição t -Student (Tabela de valores percentuais da distribuição t -Student do Anexo):

$$t_{r_s} = r_s \sqrt{\frac{n-2}{1-r_s^2}} \sim t_{\alpha\%,(n-2)}. \quad (7.3)$$

Exemplo 7.3. *Correlação de Spearman.*

Considerando uma amostra de tamanho $n = 8$, e que o coeficiente de correlação de Spearman calculado foi $r_s = 0,62$, utilizando (7.3), ao nível de 5% segue que:

$$t_{r_s} = r_s \sqrt{\frac{n-2}{1-r_s^2}} = 0,62 \sqrt{\frac{8-2}{1-0,62^2}} = +1,94,$$

e como $t_{2,5\%;6} = \pm 2,447$, então t_{r_s} pertence a região de não rejeição de H_0 . Conclui-se que a correlação não é significativa, isto é, $\rho = 0$.

Observação: A eficiência do coeficiente de correlação por postos de Spearman, quando comparado com a mais poderosa prova paramétrica de correlação r , o coeficiente de Pearson, é de cerca de 91%. Isto é, quando se utiliza r_s com uma amostra para comprovar a existência de uma associação na população, e quando as suposições e exigências inerentes ao uso adequado do coeficiente r de Pearson são satisfeitas, isto é, quando a população tem distribuição normal bivariada e a mensuração é feita no mínimo em escala de intervalo, então r_s tem 91% de eficiência de r na rejeição de H_0 .

Exemplo 7.4. *Vendas por cidades.*

As vendas de 12 semanas (em unidades de R\$ 10.000,00) de uma loja de departamento no centro de uma cidade X e em uma sua filial em um subúrbio Y, são:

X	71	64	67	58	80	63	69	59	76	60	66	55
Y	49	31	45	24	68	30	40	37	62	22	35	19

- Construa o gráfico de X vs Y e ajuste um modelo linear simples; **R.** $\hat{Y} = -89,26 + 1,95X$;
- Calcule r_s e teste à 5% a significância de ρ ; **R.** $r_s = 0,91$;
- Calcule r_p , compare com r_s obtido no item (b) e comente. **R.** $r_p = 0,948$.

$$r_p = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

```
Cidade=c(71,64,67,58,80,63,69,59,76,60,66,55)
Vendas=c(49,31,45,24,68,30,40,37,62,22,35,19)
pred=lm(Vendas~Cidade)
summary(pred)
plot(Cidade,Vendas)
abline(pred)
text(75,40,'Reta estimada: Vendas=-89.2+1.95Cidade',cex=0.6)
text(75,38,'R²=0,89',cex=0.6)
text(75,35,'r_p=0,95; r_s=0,91',cex=0.6)
cor.test(Cidade,Vendas,method='pearson')
cor.test(Cidade,Vendas,method='spearman')
```

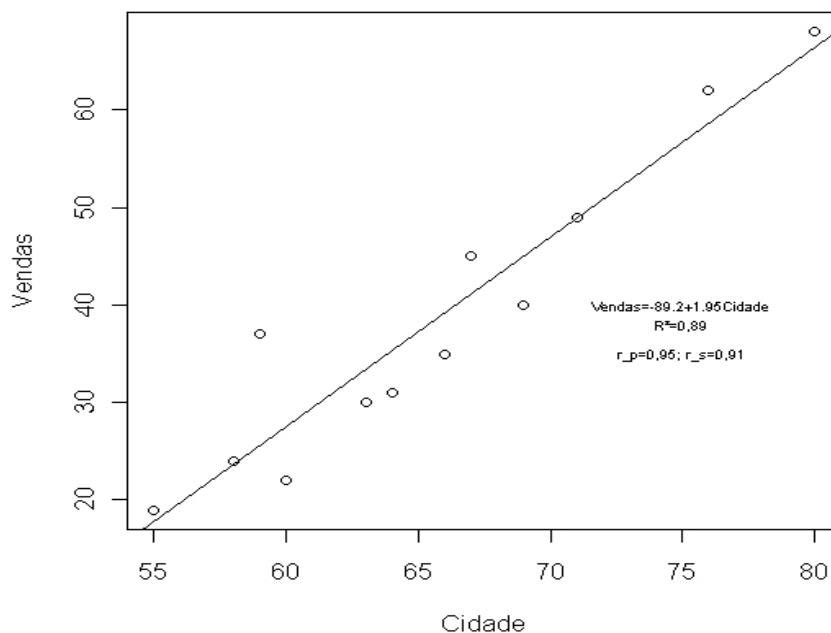


Figura 7.2: Correlação entre Cidade e Vendas.

Exercício 7.1. Considere os dados a seguir:

- a) Calcule r_s para os seguintes dados amostrais, os quais representam o número de minutos que 12 mecânicos levaram para montar uma máquina de manhã X e à tarde Y ; **R. $r_s = 0,65$**
- b) Ao nível de 5% de significância, teste se o valor de r_s é significativo; **R. $p = 0,02493$**
- c) Calcule o valor de r (Pearson) utilizando o método paramétrico e o compare com o obtido no item (a). **R. $r_p = 0,67$**

X	10,8	16,6	11,1	10,3	12,0	15,1	13,7	18,5	17,3	14,2	14,8	15,3
Y	15,1	16,8	10,9	14,2	13,8	21,5	13,2	21,1	16,4	19,3	17,4	19,0

Exercício 7.2. São apresentadas, a seguir, as notas atribuídas por três juízes aos trabalhos de dez artistas:

Juiz A:	5	8	4	2	3	1	10	7	9	6
Juiz B:	3	10	1	4	2	5	6	7	8	9
Juiz C:	8	5	6	4	10	2	3	1	7	9

- a) Calcule r_s para cada par de classificações e decida;
- R. $r_s(A \times B) = 0,61$, $r_s(A \times C) = -0,05$ e $r_s(B \times C) = -0,18$,**
- b) quais os dois juízes mais semelhantes em suas opiniões sobre os artistas;
- R. A e B**
- c) quais os dois juízes que mais diferem em suas opiniões sobre os artistas.
- R. B e C**

Exercício 7.3. *Doze pares de gêmeos idênticos fizeram testes psicológicos para medir suas agressividades. A ênfase está no exame do grau de similaridade entre os gêmeos dentro do mesmo conjunto. Teste ao nível de 1% e de 5% se a correlação é significativa.*

Primogênito	86	71	77	68	91	72	77	91	70	71	88	87
Segundo	88	77	76	64	96	72	65	90	65	80	81	72

R. $r_s = 0,43$, $p = 0,1634$, “Não rejeita-se H_0 ”

7.2 Teste de Kendall

O coeficiente de correlação por postos de Kendall, τ (tau) é uma medida de correlação útil para o mesmo tipo de dados ao qual se aplica o coeficiente de Spearman (ρ). Isto é, se ambas as variáveis X e Y acusam mensuração no mínimo ao nível ordinal, de forma que se possa atribuir a cada indivíduo um posto tanto em X como em Y, então τ dará uma medida do grau de associação entre os dois conjuntos de postos. Conhece-se a distribuição amostral de τ sob a hipótese de nulidade, portanto τ , como ρ , está sujeito a provas de significância.

Uma vantagem de τ sobre ρ é que τ pode ser generalizado para um coeficiente de correlação parcial (Siegel, 1975) o que não será apresentado neste texto.

7.2.1 Pequenas amostras ($n \leq 10$)

Considere a seguinte situação em que peçamos a dois juizes, X e Y, para atribuir ordens a quatro objetos. Por exemplo, poderíamos solicitar que classificassem quatro ensaios por ordem de qualidade de estilo. Representemos os quatro ensaios por **a**, **b**, **c** e **d**. As ordens obtidas foram:

Ensaio	a	b	c	d
Juiz X	3	4	2	1
Juiz Y	3	1	4	2

Reordenando os ensaios, de forma que as ordens atribuídas pelo Juiz X apareçam na ordem natural (1,2,...,n), temos:

Ensaio	d	c	a	b
Juiz X	1	2	3	4
Juiz Y	2	4	3	1

Estamos agora em condições de determinar o grau de correspondência entre os julgamentos de X e de Y. As ordens atribuídas pelo Juiz já estando em sua ordem natural, passamos a determinar quantos pares de ordens atribuídas pelo Juiz Y se acham em sua ordem correta (natural) em relação um ao outro.

Consideremos primeiro todos os pares de ordens em que figura a ordem “2” do Juiz Y - a ordem mais à esquerda em seu conjunto. O primeiro par, 2 e 4, está na ordem correta: 2 precede 4. Como a ordem é natural, atribuímos o escore +1 a esse par. As ordens 2 e 3 constituem o segundo par, que também está na ordem correta (2 precede 3), recebendo, assim, também o escore +1. O terceiro par consiste das ordens 2 e 1. Esses escores não estão na ordem natural: 2 não precede 1. Atribuímos então ao par o escore -1. O total de escores de todos os pares de ordens que incluem a ordem “2” é

$$S_1 = (+1) + (+1) + (-1) = +1$$

Consideramos em seguida todos os pares possíveis de ordens que incluem a ordem “4” (segunda ordem do Juiz Y a contar da esquerda) e uma outra ordem que lhe segue. Um par é 4 e 3; os elementos não estão na ordem natural; atribui-se ao par o escore -1. Outro par é 4 e 1, que também não está na ordem natural, recebendo, por isso, o escore -1. O total desses escores é

$$S_2 : (-1) + (-1) = -2.$$

Considerando agora a ordem “3” e as ordens que lhe seguem, obtemos um único par: 3 e 1, cujos elementos não estão na ordem natural; o par recebe o escore

$$S_3 = -1.$$

O total de todos os escores assim atribuídos é

$$S : (+1) + (-2) + (-1) = -2.$$

Agora, qual é o total *máximo possível* que poderíamos ter obtido para os escores atribuídos a todos os pares de ordens do Juiz Y? Obteríamos esse total máximo se as ordens dos juízes X e Y tivessem apresentado perfeita concordância, porque então, colocadas as ordens de X em sua ordem natural, cada para de ordens do Juiz Y se apresentaria também na ordem natural, recebendo assim, o escore +1. O total máximo possível - que ocorreria no caso de concordância perfeita entre X e Y seria $\binom{4}{2} = 6$. Observamos que o número total de pares possíveis

$$\binom{\text{nº de ensaios}}{2} = \binom{n}{2} = \frac{n!}{2!(n-2)!} = \frac{n(n-1)}{2},$$

ou seja, as n observações podem ser pareadas de $\frac{n(n-1)}{2}$ maneiras diferentes.

O grau de relacionamento entre os dois conjuntos de ordens é dado pela razão do total efetivo de escores +1 e -1, para o total máximo possível. O coeficiente de correlação por ordens de Kendall é a razão:

$$\tau_k = \frac{\text{total efetivo}}{\text{total máximo possível}} = \frac{S}{\binom{n}{2}} = \frac{-2}{\binom{4}{2}} = \frac{-2}{6} = -0,33.$$

Pode-se considerar τ como função do número mínimo de inversões ou permutas entre elementos vizinhos, necessário para transformar uma ordem em outra, isto é, τ é uma espécie de coeficiente de desordenamento.

De forma geral temos:

$$\tau_k = \frac{\text{total efetivo}}{\text{total máximo possível}}. \quad (7.4)$$

O escore máximo possível será $\binom{n}{2}$, que se pode exprimir como $\frac{n(n-1)}{2}$ (e será o denominador da expressão). Quanto ao numerados, denotemos por S a soma dos escores +1 e -1 para todos os pares, então, para amostras pequenas ($n \leq 10$) o coeficiente de Kendall será dado por:

$$\tau_k = \frac{S}{\frac{n(n-1)}{2}} = \frac{2S}{n(n-1)}. \quad (7.5)$$

O cálculo de S pode ser abreviado considerando da seguinte forma:

Colocadas em sua ordem natural as ordens do Juiz X, as ordens correspondentes de Y se apresentam nesta ordem:

$$\text{Juiz Y: } 2 \quad | \quad \begin{matrix} \sup & \sup & \inf \\ 4 & 3 & 1 \end{matrix}$$

Podemos determinar o valor de S partindo do primeiro número à esquerda e contando o número de ordens à sua direita que lhe são *superiores*. Deste número subtraímos o número de ordens à direita do primeiro que lhe são *inferiores*. Procedendo desta forma para todas as ordens e somando os resultados, obtemos S .

Assim, para o nosso exemplo, as ordens à direita de 2 e superiores à 2 são 3 e 4, enquanto que 1 é inferior a 2. A ordem 2 contribui, assim, com $(2-1) = +1$ para S . Para a ordem 4, não há à sua direita ordens que lhe sejam superiores; 3 e 1 são as duas ordens, que são inferiores a 4. A contribuição de 4 para S , então, $(0-2) = -2$. Quanto a ordem 3, sua contribuição para S é $(0-1) = -1$. Essas contribuições totalizam

$$S : (+1) + (-2) + (-1) = -2.$$

Conhecido S , podemos aplicar a fórmula (7.5) para calcular o valor de τ para as ordens atribuídas pelos juízes:

$$\tau_k = \frac{2S}{n(n-1)} = \frac{2(-2)}{4(4-1)} = -0,33.$$

Podemos testar, ao nível $\alpha\%$, se a correlação entre as variáveis é significativa:

$$H_0 : \tau = 0$$

$$H_a : \tau \neq 0.$$

O valor crítico da Estatística de Kendall está tabulado no Anexo e rejeitaremos H_0 se p -valor for menor do que α , nível de significância fixado.

Considerando o exemplo corrente, $|S| = 2$ e $n = 4$ fornecem um $p = 0,375$, logo conclui-se que $\tau_k = -0,33$ não é significativo ao nível de 5%.

```
x=c(3,4,2,1)
y=c(3,1,4,2)
cor.test(x,y,method='kendall')
```

7.2.2 Grandes amostras ($n > 10$)

Quando n é suficientemente grande ($n > 10$), a expressão 7.5 pode ser redefinida por:

$$z_{\tau_k} = \frac{\tau_k}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}} \sim N(0,1). \quad (7.6)$$

Observação: Os testes de Spearman e Kendall não são comparáveis, apesar de serem igualmente poderosos na rejeição de H_0 , pois utilizam equivalentemente as informações contidas nos dados. Quando utilizados em dados aos quais o coeficiente de Pearson é aplicável, τ têm eficiência de 91%.

Exemplo 7.5. *Correlação entre Autoritarismo e Aspirações de Status Social.*

A Tabela (7.2) apresenta os escores de autoritarismo e aspirações de status social de 12 estudantes. Determine o coeficiente de correlação de Kendall.

Tabela 7.2: *Escores de Autoritarismo e Aspirações de Status Social de 12 estudantes.*

Estudante	1	2	3	4	5	6	7	8	9	10	11	12
Autoritarismo	82	98	87	40	116	113	111	83	85	126	106	117
Aspirações de Status Social	42	46	39	37	65	88	86	56	62	92	54	81

Tabela 7.3: *Ordens de Autoritarismo e Aspirações de Status Social de 12 estudantes.*

Autoritarismo (X)	2	6	5	1	10	9	8	3	4	12	7	11
Aspirações de Status Social (Y)	3	4	2	1	8	11	10	6	7	12	5	9

Tabela 7.4: Ordens de Autoritarismo na forma natural

Autoritarismo (X)	1	2	3	4	5	6	7	8	9	10	11	12
Aspirações de Status Social (Y)	1	3	6	7	2	4	5	10	11	8	9	12

Dispostos em sua ordem natural as ordens de X, determinamos o valor de S para a correspondente ordem das ordens de Y:

$$\begin{aligned}
 S &= (11 - 0) + (9 - 1) + (6 - 3) + (5 - 3) + (7 - 0) + \\
 &\quad + (6 - 0) + (5 - 0) + (2 - 2) + (1 - 2) + (2 - 0) + (1 - 0) \\
 S &= +44.
 \end{aligned}$$

A ordem relativa a autoritarismo mais à esquerda é 1. Esta ordem tem 11 ordens à sua direita, que lhe são superiores, e nenhuma que lhe seja inferior. Sua contribuição para S é $(11 - 0) = 11$. A seguir, a ordem 3 contribui com $(9 - 0)$ para S, e assim por diante, determinamos $S = +44$. Daí por 7.5:

$$\tau_k = \frac{2S}{n(n-1)} = \frac{2(+44)}{12(12-1)} = +0,67,$$

o que representa o grau de relacionamento entre o autoritarismo e as aspirações de status social acusado pelos 12 estudantes.

Exemplo 7.6. *Teste para a correlação entre Autoritarismo e Aspirações de Status Social.*

Considerando o Exemplo anterior, que a correlação entre autoritarismo e grau de aspirações de status social é de 0,67 entre os 12 estudantes, segundo Kendall, se considerarmos essa amostra como uma amostra aleatória de uma população, podemos testar se essas duas variáveis apresentam associação na população ao nível de 5%. Utilizando (7.6), temos:

$$z_{\tau_k} = \frac{\tau_c}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}} = \frac{+0,67}{\sqrt{\frac{2[2(12)+5]}{9(12)(12-1)}}} = +3,03.$$

As hipóteses a serem testadas são:

H_0 : $\tau = 0$: Não existe correlação entre autoritarismo e o grau de aspirações de status social;

H_a : $\tau \neq 0$: Existe correlação entre autoritarismo e o grau de aspirações de status social.

Para $\alpha = 5\%$, sabemos que $z_{2,5\%} = \pm 1,96$, e como $z_{\tau_k} = +3,03$ [$p = 2(0,0005) = 0,001$] pertence a região de rejeição de H_0 , concluímos que existe correlação entre autoritarismo e o grau de aspirações de status social.

```

Autoritarismo=c(82,98,87,40,116,113,111,83,85,126,106,117)
ASS=c(42,46,39,37,65,88,86,56,62,92,54,81)
cor.test(x,y,method='kendall')
cor.test(x,y,method='sperman')
cor.test(x,y,method='pearson')
plot(Autoritarismo,ASS)

```

Exercício 7.4. Um marido e sua esposa que foram jogar boliche juntos registraram os seus pontos em 10 jogadas para ver se existia correlação entre seus pontos (escores), nos quais foram:

Jogo	1	2	3	4	5	6	7	8	9	10
Marido	147	158	131	142	183	151	196	129	155	158
Esposa	122	128	125	123	115	120	108	143	124	123

Através do teste de Kendall, verifique ao nível de 5% se os escores são correlacionados.

R. $S = -22$; $\tau_k = -0,48$ ($Z_c = -1,97$; $p = 0,048$).

Exercício 7.5. Ferreira (1970), estudando a densidade básica média de *Eucalyptus grandis* considerou:

X : Densidade básica média ao nível do D.A.P (amostras Pressler);

Y : Densidade básica média ao nível do D.A.P. (secções transversais do tronco).

Através do teste de Kendall, verifique ao nível de 5% se as densidades obtidas pelos dois processos são correlacionadas.

X :	0,602	0,6366	0,604	0,548	0,590	0,592	0,625	0,641	0,606	0,502	0,588	0,594	0,626
Y :	0,619	0,620	0,620	0,538	0,616	0,601	0,664	0,652	0,579	0,501	0,590	0,622	0,606

R. $\tau_k = 0,555$ ($Z_c = 2,63$; $p = 0,00852$).

Exercício 7.6. Custódio (1969), no seu estudo sobre seleção de matéria prima destinada à formação de populações básica para o melhoramento de galinhas, obteve, dentre outros, os seguintes dados:

Plantel	Peso do corpo com 200 dias (kg)	Peso do ovo com 210 dias (g)	Plantel	Peso do corpo com 200 dias (kg)	Peso do ovo com 210 dias (g)
1	1,615	23,13	9	2,411	49,43
2	1,809	52,90	10	2,272	51,49
3	1,677	52,85	11	2,513	51,26
4	1,684	53,58	12	2,394	51,20
5	1,651	52,40	13	2,043	53,25
6	1,784	52,46	14	2,309	50,11
7	1,569	49,60	15	2,201	47,82
8	1,457	52,27			

a) Determine o coeficiente de correlação de Kendall; **R.** $\tau_k = -0,086$

b) Verifique, ao nível de 5%, se o peso do ovo está correlacionado com o peso do corpo da ave.

R. $T = 48$; $p = 0,6972$.

Exercício 7.7. Numa faculdade tomaram-se as médias de aprovação no primeiro semestre dos doze primeiros alunos classificados nos exames vestibulares e cujos resultados foram:

Classificação no vestibular:	1	2	3	4	5	6	7	8	9	10	11	12
Média de aprovação:	9,5	8,7	9,0	8,6	9,3	9,3	8,4	8,8	8,5	7,0	7,3	7,0

Verifique, via Kendall, se as notas de aprovação na faculdade estão correlacionadas com a classificação no vestibular, ao nível de 5% de significância.

R. $\tau_k = -0,62$ ($Z_c = -2,76$; $p = 0,006$).

Referências Bibliográficas

- ARANGO, H.G.. *Bioestatística: Teórica e computacional*. Guanabara Koogan S.A. RJ, 2001.
- BYRT, T., BISHOP, J. and CARLIN, J.B.. Bias, prevalence and kappa. *Journal of Clinical Epidemiology*. v.46: 423, 1993.
- CALLEGARI-JACQUES, S. M.. *Bioestatística: Princípios e Aplicações*. Ed. Artmed. RS, 2003.
- CAMPOS, H.. *Estatística Experimental Não-Paramétrica*. 4ª Ed. USP/ESALQ, 1983.
- CARVAJAR, S. S. R.. *Elementos de Estatística* (com aplicações às ciências médicas e biológicas). Rio de Janeiro: UFRJ, 1970.
- CASTELLANOS, R. S e MERINO, A. P.. *Psicoestadística-Contrastes Paramétricos y no Paramétricos*. Ediciones Piramide S.A., Madrid, 1989.
- CENTENO, A. J.. *Curso de Estatística aplicada à Biologia*. Goiânia: Ed. Universidade Federal de Goiás, 1981.
- COSTA NETO, P. L. O.. *Estatística*. 5ª Edição. Ed. Edgard Blucher Ltda, SP, 1985.
- DANIEL, W. W.. *Biostatistics: a foundation for analysis in the health sciences*. 2 Ed. New York: John Wiley, 1999.
- DOWNING, D e CLARK, J.. *Estatística aplicada*. 2ª Ed. Saraiva. SP, 2002.
- FORANTTINI, O. P.. *Epidemiologia Geral*. 1ª Ed. Artes Médicas, Ltda., 1980.
- FREUND, J. E. & SIMON, G. A.. *Estatística Aplicada: Economia, Administração e Contabilidade*. 9ª Edição. Ed. Bookman, Porto Alegre, 2000.
- GUEDES, M. L. S. e GUEDES, J. S.. *Bioestatística - Para profissionais de saúde*. 1ª Ed. Ao livro técnico S.A. RJ, 1988.
- MEDRONHO, R. A.. *Epidemiologia*. Ed. Atheneu. SP, 2003.
- PEREIRA, M. G.. *Epidemiologia: Teoria e Prática*. Ed. Guanabara & Koogan, 1995.
- RODRIGUES, P. C.. *Bioestatística*. EdUFF, RJ, 2002.
- ROSNER, B. *Fundamentals of Biostatistics*. Duxbury Press, 4ª edição New York, 1994.
- SHESKIN, D. J.. *Handbook of Parametric and Nonparametric Statistical Procedures*. 3ª Ed. Chapman & Hall, 2004.

- SIEGEL, S.. *Estatística não-paramétrica*. 1ª Ed. RJ, Editora McGraw Hill do Brasil Ltda, 1981.
- SOARES, J. F. e SIQUEIRA, A.L.. *Introdução à estatística médica*. Ed. UFMG. 1999.
- SOARES, J. F e BARTMAN, F. C.. *Métodos estatísticos em medicina e biologia*. 14º Colóquio Brasileiro de Matemática. IMPA/CNPq. RJ, 1983.
- SOKAL, R.R.. *Biometry: The Principles and Practice of Statistics in Biological Research*. New York: W.F. Freeman, 1995.
- SOUNIS, E. *Bioestatística: Princípios fundamentais, metodologia, estatística, aplicação à ciências biológicas*. 2ª Ed. Revisada, McGraw-Hill, 1975.
- SPRENT, P. e SMEETON, N. C.. *Applied nonparametric statistical methods*, 3ª Ed. Chapman & Hall, 2000.
- STEEL, R.G.D.. *Principles and Procedures of Statistics a Biometrical Approach*. 2ª Ed. New York: Mc Graw-Hill, 1980.
- WAYNE, W. D.. *Biostatistics. A foundation for analysis in the health sciences*. 7ª Ed. J&S. NY, 1999.
- VIEIRA, S.. *Introdução à Bioestatística*. 5ª Edição Revista e Ampliada. Editora Campus, 1980.

Formulários

Formulário I - Testes de aderência¹

Qui-quadrado

$$Q^2 = \sum_{i=1}^k \frac{(n_i - np_i^0)^2}{np_i^0} \sim \chi_{\alpha\%, (k-1)}^2$$

$$Q^2 = \sum_{i=1}^k \frac{[n_i - n\pi_i(\hat{\theta})]^2}{n\pi_i(\hat{\theta})} \sim \chi_{\alpha\%, (k-1)}^2$$

$$\chi^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i} \sim \chi_{\alpha\%, (k-1-s)}^2$$

$$\chi_{cy}^2 = \sum_{i=1}^k \frac{(|o_i - e_i| - 0,5)^2}{e_i} \sim \chi_{\alpha\%, (k-1-s)}^2$$

e_i : valor esperado

Kolmogorov-Smirnov

$$d_c = \max |F(x) - G(x)| \sim \begin{cases} Se\ n \leq 40 \rightarrow d_{\alpha\%, n} \\ Se\ n > 40\ e\ \alpha = 5\% : 1,36/\sqrt{n} \\ Se\ n > 40\ e\ \alpha = 1\% : 1,63/\sqrt{n} \end{cases}$$

Shapiro-Wilk

$$T = \frac{\left[\sum_{i=1}^{k=n/2} a_i (X^{n-i+1} - X^i) \right]^2}{s^2(n-1)} \sim T_{\alpha\%, n}$$

$$G = \begin{cases} Se\ n \leq 6 \\ W = \ln \left[\frac{T - d_n}{1 - T} \right] \sim N(0, 1) \\ Se\ 7 \leq n \leq 50 \\ W = b_n + c_n \ln \left[\frac{T - d_n}{1 - T} \right] \sim N(0, 1) \\ b_n, c_n\ e\ d_n\ \text{valores tabelados.} \end{cases}$$

¹Para fazer uso deste formulário nas avaliações, não faça nenhum tipo de anotação.

Formulário II - Testes Aplicáveis a duas amostras²

Dados pareados

$$Z_{\text{sinais}} = \frac{x - 0.5n}{\sqrt{0.25n}} \sim N(0, 1)$$

$$\chi_{McN}^2 = \frac{(|b - c| - 1)^2}{b + c} \sim \chi_{\alpha\%;1}^2$$

$$Wilcoxon = \frac{\min\{|T_-|; |T_+|\} - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \sim N(0, 1)$$

	Kappa	Concordância
	$k \leq 0,2$	péssima
	$0,2 < k \leq 0,4$	razoável
	$0,4 < k \leq 0,6$	boa
	$0,6 < k \leq 0,8$	muito boa
	$0,8 < k \leq 1$	excelente

$$p_o = \frac{a + d}{n}; \quad p_c = \frac{n_1 m_1 + n_2 m_2}{n^2}; \quad Kappa = \frac{p_o - p_c}{1 - p_c}$$

Dados não-pareados

$$\text{Sinais: } Z_c = \frac{p_1 - p_2}{\sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim N(0, 1)$$

$$Z_c = \frac{2(p_1 - p_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim N(0, 1)$$

$$\chi_{cYates}^2 = \sum_{i=1}^k \frac{(|o_i - e_i| - \frac{1}{2})^2}{e_i} \sim \chi_{[L-1 \times C-1; \alpha\%]}^2$$

$$\chi_c^2 = \frac{n(ad - bc)^2}{n_1 n_2 m_1 m_2} \sim \chi_{\alpha\%;1}^2$$

$$\chi_{cYates}^2 = \frac{n(|ad - bc| - \frac{n}{2})^2}{n_1 n_2 m_1 m_2} \sim \chi_{\alpha\%;1}^2$$

$$p_{\text{Fisher}} = \frac{n_1! n_2! m_1! m_2!}{a! b! c! d! n!}$$

$$\text{Mann-Whitney: } U_i = n_1 n_2 + \frac{n_i(n_i + 1)}{2} - T_i, \quad i = 1, 2; \quad U_c = \min\{U_1; U_2\}$$

$$Z_c = \frac{\frac{n_1 n_2}{w} - \frac{2}{12}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \sim N(0, 1); \quad w = \min\{U_1; U_2\}$$

$$D_{\text{Kolmogorov-Smirnov}} = \max\{|S_{n_1}(X) - S_{n_2}(X)|\} \sim D_{\alpha\%}; \quad S_{n_i}(X) = \frac{k}{n_i}, \quad i = 1, 2.$$

$$\text{Hollander e Wolfe: } J_3 = \frac{n_1 n_2}{d} D_{K-S}; \quad d : \text{mdc}\{n_1; n_2\}$$

$$D_{5\%} = 1,36 \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

²Para fazer uso deste formulário nas avaliações, não faça nenhum tipo de anotação.

Formulário III - Testes Aplicáveis a k amostras³

Dados pareados

$$Q_{Cochran} = \frac{k(k-1) \sum_{j=1}^k (T_j)^2 - (k-1)N^2}{kN - \sum_{i=1}^n (S_i)^2} \sim \chi_{\alpha\%, k-1}^2$$

$$H_{Friedman} = \frac{12}{Nk(k+1)} \sum_{j=1}^k T_j^2 - 3N(k+1) \sim \chi_{\alpha\%, (k-1)}^2$$

$$d.m.s = |T_i - T_j|; \quad i \neq j \quad \text{ou} \quad d.m.s. = Q \sqrt{\frac{nk(k+1)}{12}}$$

Dados não-pareados

$$H_{Kruskal-Wallis} = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{T_j^2}{n_j} - 3(N+1) \sim \chi_{\alpha\%, (k-1)}^2$$

$$H_{K-W}^{empates} = \frac{12}{N(N+1)} \frac{\sum_{j=1}^k \frac{T_j^2}{n_j} - 3(N+1)}{\sum_{i=1}^g T_i} \chi_{\alpha\%, (k-1)}^2; \quad T = t^3 - t$$
$$1 - \frac{\sum_{i=1}^g T_i}{N^3 - N}$$

$$d.m.s = |T_i - T_j|; \quad i \neq j \quad \text{ou} \quad d.m.s_{ij} = \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right) h}$$

$$d.m.s. = Q \sqrt{\frac{k(N+1)}{12}} \quad \text{ou} \quad d.m.s_{ij} = \left| z_{\frac{\alpha}{k(k-1)}} \right| \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

Formulário IV - Testes de Correlação

Análise de Correlação

$$r_{Pearson} = \frac{\sum_{i=1}^n (x_i - \bar{x}) (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

$$r_{Spearman} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n}$$

$$t_{r_s} = r_s \sqrt{\frac{n-2}{1-r_s^2}} \sim t_{\alpha\%, (n-2)}$$

$$z_{r_s} = r_s \sqrt{n-1} \sim N(0, 1)$$

$$\tau_{Kendall} = \frac{2S}{n(n-1)}$$

$$z_{\tau_k} = \frac{\tau_k}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}} \sim N(0, 1)$$

³Para fazer uso deste formulário nas avaliações, não faça nenhum tipo de anotação.

Tabelas

- Normal padrão
- t-Student
- Qui-quadrado
- Binomial
- Poisson
- Kolmogorov-Smirnov - Uma amostra
- Coeficientes de Shapiro-Wilk
- Shapiro-Wilk
- Lilliefors
- Wilcoxon
- Mann-Whitney
- Kolmogorov-Smirnov - Duas amostras
- Kolmogorov-Smirnov - Hollander-Wolfe
- Friedman
- Kruskal-Wallis
- Comparações múltiplas de Friedman
- Comparações múltiplas de Kruskal-Wallis
- Amplitude Q
- Spearman
- Kendall