

**UNIVERSIDADE FEDERAL DO PARANÁ
DEPARTAMENTO DE ESTATÍSTICA**

MATERIAL DIDÁTICO

ESTATÍSTICA NÃO PARAMÉTRICA
Elaborado a partir de notas de aula do prof. Paulo Ricardo B. Guimarães

2016

**UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE CIENCIAS EXATAS
DEPARTAMENTO DE ESTATISTICA**

Índice

Introdução.....	pag. 5
I - Métodos Estatísticos Não-Paramétricos.....	pag. 6
1.1 - Suas vantagens e desvantagens	
1.2 – Conceitos básicos	
1.3 - Discussão geral dos testes estatísticos	
1.4 - Escolha do teste estatístico adequado	
1.5 - Níveis de mensuração.	
II - Testes para o Caso de Uma Amostra (Testes de aderência).....	pag. 11
2.1 - Teste binomial	
2.2 - Teste do Qui-quadrado	
2.3 - Teste de Kolmogorov-Srnirnov	
2.4 - Teste de Lilliefors	
2.5 - Teste de Aleatorização	
III - Testes para Duas Amostras Relacionadas (dependentes).....	pag. 25
3.1 - Teste de McNemar	
3.2 - Teste dos Sinais	
3.3 - Teste de Walsh	
3.4 - Teste de Wilcoxon	
3.5 - Teste de Aleatorização	
IV- Testes para duas Amostras Independentes.....	pag. 44
4.1 - Teste Exato de Fischer	
4.2 - Teste do X^2	
4.3 - Teste da Mediana	
4.4 - Teste U de Mann-Whitney	
4.5 - Teste Kolmogorov-Srnirnov	
4.6 - Testes das Iterações de Wald-Wolfowitz	
4.7 - Teste de Aleatorização	

V - Testes para K - Amostras Relacionadas.....	pag. 62
5.1 - Teste Q de Cochran	
5.2 - Teste de Friedman	
5.3 - Comparações Múltiplas	
VI - Testes para K amostras Independentes.....	pag. 70
6.1 - Teste da mediana	
6.2 - Teste do X^2	
6.3 - Teste de Kruskal-Wallis	
6.4 - Comparações Múltiplas	
VII - Medidas de Associação.....	pag. 80
7.1 - Coeficiente de Contingência C	
7.2 - Coeficiente de Spearman	
ANEXOS.....	pag. 82

Referências Bibliográficas

- 1 - SIEGEL, S.; CASTELLAN, J. (2006). *Estatística Não - Paramétrica para as Ciências do Comportamento*, 2. Ed. Editora Penso.
- 2 - CAMPOS, H. (1983). *Estatística Experimental Não - Paramétrica*, ESALQ.
- 3 - HOLLANDER, M.; WOLFG, D.A. (1999). *Nonparametric Statistical Method*, John Wiley & Sons.
- 4 - CONOVER, W.J. (1998). *Practical Nonpararnetric Statistics*, John Wiley & Sons.
- 5 - NEGRILLO, B. G. (1992) *Métodos não-paramétricos uni e multivariados*. Piracicaba, CIAGRI.
- 6 - SPRENT, P., SMEETON, N.C. (2001) *Applied Nonparametric Statistical Methods*. New York: Chapman Hall.

INTRODUÇÃO

A Estatística Não paramétrica pode ser definida como uma coleção alternativa de métodos estatísticos aplicada a conjuntos de dados onde as condições necessárias para aplicação de uma técnica clássica inferencial (p. ex., Intervalo de Confiança, Teste de Hipótese) não são satisfatoriamente atendidas. É também bastante útil no tratamento de dados onde o nível de mensuração das observações não é dos melhores.

Tais procedimentos são usados há muitos anos atrás, informalmente, embora não com o nome atual:

- O rei Nabucodonossor aplicou o teste da permutação, 600 anos AC;
- Cálculos da probabilidade binomial foram feitos em 1710 pelo médico inglês Arbuthnott;
- O primeiro livro-texto denotado aos métodos Não-paramétricos foi escrito por Siegel (1956) e é usado até hoje pela leitura acessível a leigos.

No entanto, Savage designa o ano de 1936 como o verdadeiro início da Estatística Não-Paramétrica, marcado pela publicação do artigo de Hotelling e Pabst sobre correlação por postos.

O seu maior crescimento ocorreu nos últimos 40 anos.

Atualmente esta área da Estatística está bastante desenvolvida e os tópicos modernos são Estimação de Densidade, Regressão Não-Paramétrica e Semi-Paramétrica.

Em 1991 foi lançado o Journal of Nonparametric Statistics.

O tema central em Estatística é a chamada Inferência Estatística que aborda dois tipos de problemas fundamentais: a estimação de parâmetros de uma população, e o teste de hipóteses. Este último é ao que será tratado nesse curso. Na Inferência Estatística procuramos tirar conclusões sobre um grande número de eventos com base na observação de apenas parte deles. Os testes relacionados à Inferência Estatística nos dizem qual a margem de diferença que deve ser encontrada na amostra para que possamos afirmar que elas representam realmente diferenças nos tratamentos (grupos). Como nesses procedimentos, na verdade são testadas hipóteses a respeito dos parâmetros populacionais, esses são chamados de “Paramétricos”.

Algumas técnicas não são tão rigorosas na especificação de condições acerca dos parâmetros da população da qual a amostra foi obtida. Consequentemente as conclusões não são tão poderosas quanto às obtidas por técnicas paramétricas. Essas técnicas são chamadas de “distribuição livre” ou “não paramétricas”.

I - MÉTODOS ESTATÍSTICOS NÃO PARAMÉTRICOS

1.1 – SUAS VANTAGENS E DESVANTAGENS

Vantagens

- ✓ Dispensam Normalidade dos dados;
- ✓ O p – valor é exato (no caso paramétrico o cálculo do p-valor se baseia na distribuição Normal);
- ✓ São testes mais simples;
- ✓ São úteis quando é difícil estabelecer uma escala de valores quantitativa para os dados;
- ✓ São mais eficientes que os paramétricos quando não existe normalidade.

Desvantagens

- ✓ Proporcionam um desperdício de informações, já que em geral não consideram a magnitude dos dados;
- ✓ Quando as suposições do modelo estatístico são atendidas são menos eficientes que os paramétricos;
- ✓ A utilização das tabelas dos testes.

1.2 – CONCEITOS BÁSICOS

Amostra Aleatória

X_1, X_2, \dots, X_n constituem uma amostra aleatória se cada X_i apresenta a mesma distribuição e independência entre si.

Parâmetro

É um valor desconhecido, associado à uma população, que se deseja estimar.

Estimador

É uma função da amostra aleatória que estima o valor do parâmetro.

Nível de Confiança

É a probabilidade de abrangência do intervalo de confiança para certo parâmetro.

Hipótese Nula (H_0)

É uma afirmação acerca do parâmetro (ou parâmetros) em questão que expressa sempre uma posição conservadora, em forma de igualdade, que geralmente se deseja rejeitar. No caso não-paramétrico estas afirmações nem sempre se referirão a parâmetros, estando associadas ao interesse do pesquisador.

Hipótese Alternativa (H_a)

É a hipótese que se deseja testar, expressa sempre em oposição à hipótese nula. As evidências amostrais auxiliam a sua definição.

Teste de Hipótese

É uma regra de decisão que, com base na amostra, irá rejeitar ou não H_0 .

Região Crítica

É um subconjunto de valores tão extremos, que nos leva a rejeição da hipótese nula.

Erro Tipo I

Ocorre quando se rejeita H_0 quando na verdade esta hipótese é verdadeira. A probabilidade do Erro Tipo I (α) é chamada nível de significância.

Erro Tipo II

Ocorre quando H_0 não é rejeitada quando na verdade deveria ter sido. A probabilidade do Erro Tipo II é representada por β .

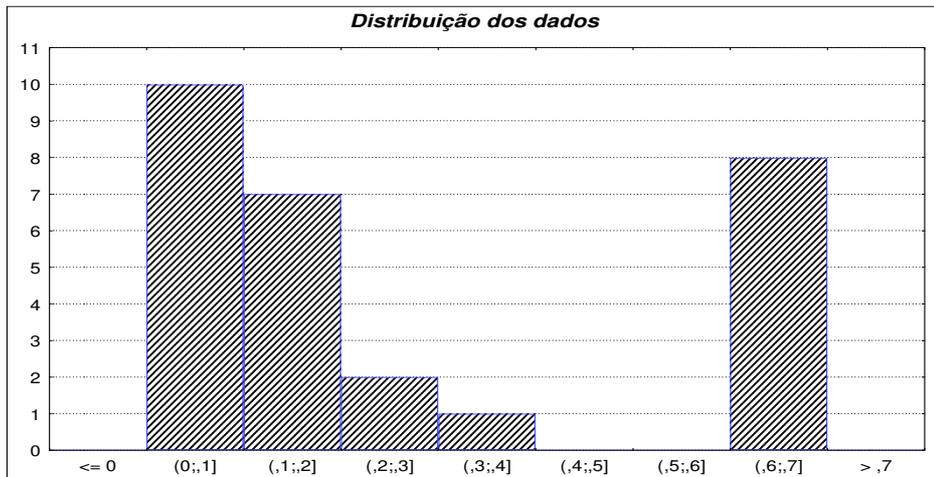
Posto (Rank)

Quando se atribui às observações originais, números correspondentes às suas posições na classificação (ranking em ordem crescente). Cada número é chamado de posto.

1.3 – DISCUSSÃO GERAL DOS TESTES ESTATÍSTICOS

Em um exemplo, serão apresentados os problemas que podem surgir ao se utilizar testes paramétricos quando as suas suposições não são atendidas.

Considerar um experimento onde foram observadas concentrações de partículas em 28 amostras de solo após aplicação de um pesticida. O histograma para o conjunto de dados observado é dado abaixo:



$$\bar{X} = 0,314 \quad s^2 = 0,07 \quad s = 0,264$$

O fabricante afirma que a concentração média de pesticida no solo se dará na ordem de $0,35\text{g/cm}^3$. Testando essa afirmação:

$$H_0 : \mu = 0,35 \quad H_a : \mu < 0,35 \quad t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{0,314 - 0,35}{0,264/\sqrt{28}} = -0,72$$

Não existem evidencias para rejeitar a hipótese nula ao usar o teste de hipótese paramétrico. Assim a afirmação do fabricante estará comprovada. Porém;

- ✓ Pode-se notar uma grande assimetria na distribuição dos dados, o que causou um aumento do valor da média em favor da hipótese nula;
- ✓ Observa-se que 8 amostras apresentaram alto teor de concentração de inseticida ($>0,60$) e a grande maioria (71%) não passou de 0,40;
- ✓ Essas amostras de solo foram obtidas perto de um rio (parte baixa da área) para onde provavelmente “desceu” o pesticida levado pela ação das chuvas;

- ✓ O que se nota na verdade é que o produto parece não ter se espalhado homogeneamente;
- ✓ Neste caso talvez se tivesse sido testada a mediana (0,19) ao invés da média teriam-se obtido resultados diferentes;
- ✓ Essas 8 amostras com alto teor de concentração não podem simplesmente ser eliminadas do estudo, pois constituem um número considerável de informações além de apontar o fato do pesticida se espalhar de forma desigual na área considerada.

É interessante, no momento oportuno, que se retorne a este exemplo, aplicando a técnica estatística adequada e se compare com o resultado obtido no teste paramétrico.

1.4 – ESCOLHA DO TESTE ESTATÍSTICO ADEQUADO

É importante a definição de critérios que nos ajudem a decidir qual o teste ideal para determinado problema.

Um desses critérios, sem dúvida, é o Poder do Teste ($1 - \beta$). O teste que apresenta uma maior probabilidade de rejeitar H_0 quando H_0 é falsa, entre todos os testes de nível α deve ser escolhido. Mas só isto não basta e nem sempre é simples de ser obtido, portanto precisamos de outras informações para escolher o teste mais adequado:

- ✓ Como foi obtida a amostra, ou seja, o plano experimental;
- ✓ Natureza e características da população (pessoas, objetos, áreas, animais, etc.);
- ✓ Tipo de Mensuração dos dados (escala de mensuração).

Quando se usa um teste paramétrico existem pressupostos a serem verificados, além do nível mínimo de mensuração exigido ser a escala intervalar.

Quando essas suposições não são verificadas é possível que o teste nos leve a resultados errôneos.

No caso não-paramétrico, o primeiro critério a ser verificado deve ser o nível de mensuração dos dados.

1.5 – NÍVEIS DE MENSURAÇÃO

a) Escala Nominal

É o mais baixo nível de mensuração. Utiliza símbolos ou números simplesmente para distinguir elementos em diferentes categorias (como um nome), não havendo entre eles, geralmente, possibilidade de comparação do tipo maior-menor, melhor-pior).

Exemplos:

- Masculino (M), Feminino (F)
- Perfeito (1), Defeituosa (0)
- Europeu (1), Americano(2), Africano (3), Asiático(4)

b) Escala Ordinal

Utiliza números apenas para classificarmos elementos numa ordem crescente ou decrescente. Existe assim algum tipo de relação entre as categorias embora a diferença entre elas seja de difícil quantificação.

Exemplos:

- Classes sócio - econômicas: (A, B, C, D, E)
- Patentes do Exército (soldado, cabo, sargento, etc)
- Opinião de um determinado produto (Ruim, Regular, Bom, Muito bom, Excelente)

c) Escala Intervalar (Intervalo de medida)

Ocorre quando a escala tem as características da escala ordinal e ainda é possível quantificar a diferença entre dois números desta escala.

Exemplo:

Temperatura, Peso, Altura, Rendimentos

obs: Alguns autores apontam ainda a existência de outra escala: a Escala de Razão, equivalente a escala intervalar porém, o valor zero é o verdadeiro ponto de origem.

II – TESTES PARA O CASO DE UMA AMOSTRA (Testes de Aderência)

Estes testes são úteis para verificar se determinada amostra pode provir de uma população especificada (se ajusta a uma distribuição de probabilidade). São usualmente conhecidos como testes de aderência ou bondade do ajuste. Neste caso, retira-se uma amostra aleatória e compara-se a distribuição amostral observada com a distribuição teórica de interesse.

2.1 – TESTE BINOMIAL

É aplicado em amostras provenientes de populações que se constituem de apenas 2 categorias (variáveis dicotômicas). Os experimentos binomiais se caracterizam por n repetições independentes com probabilidade constante de sucesso.

Exemplo: (masculino, feminino), (negativo, positivo), (defeituoso, perfeita)

É útil para verificar se a proporção de sucesso p observada na amostra indica que ela pode pertencer a uma população com uma determinada probabilidade de sucesso P.

2.1.1 – Pressupostos

- ✓ Cada observação é classificada como sucesso ou fracasso;
- ✓ A probabilidade p de sucesso não se altera com a repetição do experimento;
- ✓ As n tentativas são independentes.

2.1.2 – O Método

Este método irá calcular a probabilidade da variável de interesse apresentar resultados mais extremos do que os observados.

- ✓ Inicia-se com a formulação das hipóteses:

$$H_0 : p = p_0$$

$$H_a : p > p_0 \text{ ou } H_a : p < p_0 \text{ (unilateral)} \text{ ou } H_a : p \neq p_0 \text{ (bilateral)}$$

- ✓ Determinar n, número de repetições;
- ✓ Determinar a frequência observada (x);
- ✓ Calcula-se a probabilidade de ocorrência de valores mais extremos do que o observado (número de sucessos):

$$P(X \leq x) = \sum_{i=0}^x \binom{n}{i} p^i q^{n-i} \text{ ou } P(X \geq x) = \sum_{i=x}^n \binom{n}{i} p^i q^{n-i}$$

*dependendo da hipótese alternativa

Esta probabilidade equivale ao p-valor. Caso $p < \alpha$, rejeita-se H_0 .

Exemplo 1 - $H_0 : p = 0,5$ versus $H_a : p < 0,5$; $x = 5$ (número de sucessos em $n=11$ tentativas).
Considera-se x como sendo o número de sucessos observados.

$$\hat{p} = \frac{5}{11} = 0,45 \qquad P(X \leq 5) = \sum_{i=0}^5 \binom{n}{i} p^i q^{n-i} = \sum_{i=0}^5 \binom{11}{i} p^i q^{11-i} = 0,5$$

Conclusão: Não se rejeita a hipótese nula. O p-valor (0,5) é considerado alto, muito superior aos níveis usuais de significância.

Obs: Caso $n > 25$ e $p \approx \frac{1}{2}$ ou $npq \geq 9$, deve-se utilizar a aproximação pela distribuição Normal:

$$Z = \frac{x - np}{\sqrt{npq}} \text{ e procede-se como num teste de hipótese convencional}$$

Obs: neste caso, os procedimentos para pequenas e grandes amostras, devem fornecer resultados muito próximos. Deve-se enfatizar, no entanto, que o p-valor exato será aquele obtido utilizando a metodologia para pequenas amostras, sem utilizar aproximação para uma distribuição de probabilidade teórica. Com o desenvolvimento de recursos computacionais cada vez mais precisos, muitas vezes perde o sentido utilizar a aproximação pela Normal. Isto vale para todas as demais situações apresentadas neste material, onde se apresenta o procedimento para grandes amostras.

Exemplo 2: Uma empresa lançou recentemente no mercado uma dieta de emagrecimento. E pretende-se averiguar se a porcentagem de usuários acima de 60 anos é significativamente maior que 30%. Para tal houve escolha aleatória de 16 indivíduos entre os adeptos da nova dieta, e apurou-se que apenas 5 tinham idade superior a 60 anos.

$$H_0 : p = 0,30 \quad \text{versus} \quad H_a : p > 0,30$$

x = número de sucessos = 5 e considerando $n = 16$ repetições.

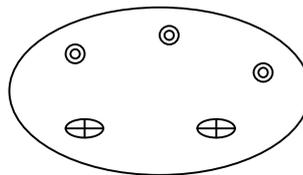
Obs: note que a hipótese alternativa foi definida pelos dados amostrais, pois $\hat{p} = \frac{5}{16} = 0,3125$.

$$P(X \geq 5) = \sum_{i=5}^n \binom{n}{i} p^i q^{n-i} = 1 - \sum_{i=0}^4 \binom{16}{i} 0,3^i 0,7^{16-i} = 0,551$$

Conclusão: Com um nível de significância de 5% não rejeitamos a hipótese nula, pois o p-valor é aproximadamente igual a 0,55. Não podemos afirmar que a proporção de seguidores com idades superiores a 60 anos é maior que 30%.

Exemplo 3: Num ensaio de degustação de café, cada mesa era constituída de 5 amostras, sendo duas delas de café “mole” e as 3 restantes, de café “comum”. Dos 8 degustadores que foram utilizados, 3 classificaram corretamente os tipos de café.

Teste a hipótese de que os degustadores conseguem distinguir o café “mole” dos demais.



Primeiramente precisamos saber qual a probabilidade de um degustador distinguir por acaso os dois cafés “moles” dentre as 5 amostras. $p_0 = \frac{1}{C_5^2} = \frac{1}{10} = 0,10$ ou $\frac{2}{5} \times \frac{1}{4}$

$$H_0 : p = 0,10 \quad \text{versus} \quad H_a : p > 0,10$$

$$P(X \geq 3) = \sum_{i=0}^N \binom{N}{i} p^i q^{N-i} = \sum_{i=3}^8 \binom{8}{i} 0,10^i 0,90^{8-i} = 0,0381$$

Conclusão: A probabilidade de 3 ou mais degustadores distinguirem corretamente os tipos de café aleatoriamente é de apenas 0,0381 ou 3,81%. Ao nível de significância de 5% rejeitamos a hipótese nula. Isto indica que o resultado observado não deve ser devido ao acaso.

Exercício - A Lei de Mendel afirma que para 4 genes, 3 são dominantes e 1 é recessivo. Em certo experimento observou-se 54 genes dominantes e 26 recessivos. Verifique, pelo teste Binomial, se a lei se aplica ao experimento realizado.

2.1.3 – Comentários

É a técnica mais poderosa aplicável a dados medidos em escala nominal e ainda por cima dicotômicos.

Quando a variável aleatória em estudo é contínua e procede-se uma dicotomização haverá certamente perda de eficiência.

2.2 – TESTE DE QUI-QUADRADO (χ^2)

É um teste amplamente utilizado em análise de dados provenientes de experimentos onde o interesse está em observar frequências em diversas categorias (pelo menos duas).

É uma prova de aderência útil para comprovar se a frequência observada difere significativamente da frequência esperada, esta geralmente especificada por uma distribuição de probabilidade.

2.2.1 – Exigências do Teste

- ✓ Quando o número de categorias é igual a 2 ($k=2$) as frequências esperadas devem ser superiores a 5;
- ✓ Quando $k > 2$, não deve haver mais de 20% das frequências esperadas abaixo de 5 e nenhuma frequência esperada igual a zero;
- ✓ Para evitar frequências esperadas pequenas deve-se combinar as categorias (juntar) até que as exigências sejam atendidas;
- ✓ Caso as categorias sejam combinadas em apenas duas e mesmo assim as exigências não tenham sido atendidas, deve-se utilizar o Teste Binomial;
- ✓ As observações devem ser independentes.

2.2.2 – O Método

Após se definir a hipótese nula como a proporção esperada definida pela distribuição de probabilidade em questão, testa-se se as frequências observadas diferem muito das frequências esperadas da seguinte forma:

$$\chi^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i} \quad \text{em que} \quad \begin{cases} k = \text{número de categorias (classes)} \\ o_i = \text{frequência observada na categoria } i \\ e_i = \text{frequência esperada na categoria } i \end{cases}$$

Quanto maior o valor de χ^2 maior é a probabilidade das frequências observadas divergirem das frequências esperadas.

A estatística do teste χ^2 tem distribuição Qui-Quadrado com ν graus de liberdade onde:

- $\nu = k - 1$ se as frequências esperadas puderem ser calculadas sem precisar estimar os parâmetros distribucionais;
- $\nu = k - m - 1$ se as frequências esperadas só puderem ser calculadas após a estimação dos m parâmetros populacionais.

Exemplo 1: Deseja-se testar se a posição de largada de um cavalo (por dentro ou por fora) influencia no resultado de uma corrida de cavalos.

Tabela 2.1 – Distribuição do número de vitórias dos cavalos segundo a posição de largada.

Posição	1	2	3	4	5	6	7	8
Número de Vitórias	29	19	18	25	17	10	15	11
	18*	18*	18*	18*	18*	18*	18*	18*

Fonte: Livro-Siegel

* Resultado esperado pela hipótese nula

$$H_0 : f_1 = f_2 = \Lambda = f_8 \quad \text{versus} \quad H_a : f_1 \neq f_2 \neq \Lambda \neq f_8$$

$$\chi^2 = \sum_{k=1}^8 \frac{(o_i - e_i)^2}{e_i} = \frac{(29-18)^2}{18} + \frac{(19-18)^2}{18} + \Lambda + \frac{(11-18)^2}{18} = 16,3$$

χ^2_7 indica que o valor 16,3 está associado à um p-valor = 0,02224, considerado um valor bilateral.

Conclusão: Ao nível de significância de 1% não rejeitamos H_0 , porém a um nível de 5% rejeitamos H_0 . São necessárias mais repetições do experimento para as conclusões serem mais confiáveis.

Exemplo 2: A tabela dada a seguir apresenta o número observado de falhas mecânicas, por hora, em uma linha de montagem a partir de um experimento com duração de 40 horas.

Tabela 2.2 – Distribuição do número de falhas mecânicas por hora em uma linha de montagem.

Falhas	0	1	2	3	4	5	6	7	+ de 7
Freq. Observada	0	6	8	11	7	4	3	1	0
Freq. Esperada	1,6	5,2	8,3	8,9	7,1	4,6	2,4	1,1	0,7

Fonte: Livro-Siegel

Um engenheiro afirma que o processo descrito acima segue uma distribuição de Poisson com média igual a 3,2.

H_0 : A distribuição das falhas mecânicas por hora se ajusta a um processo de Poisson com $\lambda = 3,2$

H_a : A distribuição das falhas mecânicas por hora não se ajusta a um processo de Poisson com $\lambda = 3,2$

Como a tabela dada apresenta mais de 20% das frequências com valores inferiores a 5 devemos unir as categorias. Assim:

Tabela 2.3 – Junção das Categorias em que as frequências são inferiores ao valor 5.

Falhas	0 e 1	2	3	4	5 ou mais
Freq. Observada	6	8	11	7	8
Freq. Esperada	6,8	8,3	8,9	7,1	8,8

Fonte: Livro-Siegel

$$\chi^2 = \sum_{k=1}^5 \frac{(o_i - e_i)^2}{e_i} = 0,675$$

$\nu = 5 - 1 = 4$ então χ_4^2 indica que o valor 0,675 acusa um p-valor = 0,954375.

Conclusão: Ao nível de significância de 5% não podemos rejeitar H_0 . A distribuição das falhas mecânicas parece se ajustar satisfatoriamente a um processo de Poisson com uma média de 3,2 falhas por hora.

Exercício - O recenseamento de 320 famílias com 5 filhos revelaram os dados abaixo:

Tabela 2.4 – Distribuição do número de famílias com 5 filhos de acordo com o sexo

<i>Número de filhos</i>	<i>5 mulheres 0 homens</i>	<i>4 mulheres 1 homem</i>	<i>3 mulheres 2 homens</i>	<i>2 mulheres 3 homens</i>	<i>1 mulher 4 homens</i>	<i>0 mulheres 5 homens</i>	<i>Total</i>
<i>Número de famílias</i>	18	56	110	88	40	8	320
<i>e_i</i>							320

Teste a hipótese de que o nascimento de homens e mulheres é igualmente provável. Caso H_0 seja rejeitada, estimar o parâmetro da distribuição e refazer o teste.

2.2.3 - Comentários

- ✓ Para variáveis nominais é a única técnica adequada de verificação da bondade do ajuste;
- ✓ Para variáveis ordinais não é sensível ao efeito da ordem. Quando a hipótese levar em conta a ordem, o teste de Qui-Quadrado deixa de ser a melhor opção.

2.3 – TESTE DE KOLMOGOROV – SMIRNOV

Este teste de aderência proposto em 1933 avalia a concordância entre a distribuição observada da amostra e uma determinada distribuição teórica. Avalia se os dados amostrais se aproximam razoavelmente de uma determinada distribuição. Para isso utiliza-se a função **distribuição acumulada** observada, compara-se com a teórica, determina-se o ponto em que estas distribuições mais divergem, e testa-se se essa divergência é significativa ou não.

2.3.1 – Exigência do Teste

- ✓ Os dados devem seguir ao menos uma escala ordinal.

2.3.2 – O Método

Seja $F_0(X)$ uma distribuição teórica acumulada e $S_n(X)$ uma distribuição acumulada observada em uma amostra de n observações.

Encontra-se a seguir o maior valor das diferenças entre $F_0(X)$ e $S_n(X)$, ou seja,

$$D = \max |F_0(X) - S_n(X)|$$

Compara-se o valor observado com o valor crítico (**Tábua E** – Siegel) bilateral.

Na tabela associa-se o valor observado com o seu p-valor.

Exemplo 1: Num experimento para calibrar a luminosidade adequada de uma nova máquina fotográfica, foram tiradas 5 fotos de cada uma das 10 pessoas que participaram do experimento. A cada pessoa perguntou-se qual das fotos apresentava uma qualidade maior, de 1 a 5, onde 1 representa um grau baixo e 5 um grau alto de luminosidade.

$H_0 : f_1 = f_2 = \Lambda = f_5$ X H_a : As frequências não são iguais

	1	2	3	4	5
Frequência	0	1	0	5	4
$F_0(X)$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	$\frac{5}{5}$
$S_{10}(X)$	$\frac{0}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{6}{10}$	$\frac{10}{10}$
$F_0(X) - S_{10}(X)$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{5}{10}$	$\frac{2}{10}$	0

Conclusão: $D = 0,5$. Para $n = 10$, $p\text{-valor} < 0,01$. Portanto rejeita-se H_0 , em favor de H_a . Os indivíduos apresentam preferência significativa em relação ao grau de luminosidade.

Exemplo 2: Verifique se os dados abaixo podem ser ajustados por uma distribuição de Poisson com média igual a 1,2.

X_i	f_i
0	15
1	25
2	10
3	5
4	4
5	1

$F_0(X_i)$	$S_n(X_i)$	$F_0(X_i) - S_n(X_i)$
0,3012	0,250	0,0512*
0,6626	0,666	0,0041
0,8794	0,833	0,0461
0,9661	0,917	0,0494
0,9921	0,983	0,0088
0,9983	1,000	0,0017

H_0 : Os dados seguem uma distribuição Poisson (1,2) ou $F = F_0(X)$

H_a : Os dados não seguem uma distribuição Poisson (1,2) ou $F \neq F_0(X)$

Conclusão: $D = \max|F_0(X_i) - S_n(X_i)| = 0,0512$ e com isto $p > 0,20$. Então não podemos rejeitar a hipótese nula.

Como vimos no exemplo acima os dados foram agrupados em classes (i), onde era associado a cada valor X_i a sua respectiva frequência f_i . Este teste também pode ser aplicado quando as classes são definidas por intervalos; no entanto desta forma ele perde em precisão.

Exercício: As produções médias (sacas) de um experimento envolvendo adubação em milho em 115 propriedades encontram-se tabuladas abaixo

Classes	f_i	x_i	$F(x_i)$	$S(x_i)$	$F(x_i) - S(x_i)$
2700-3000	13				
3000-3300	18				
3300-3600	24				
3600-3900	32				
3900-4200	17				
4200-4500	11				
	115				

Pode-se admitir que a produção média segue uma distribuição normal com média $\mu = 3.600$ e $\sigma^2 = 3.000.000$?

2.3.3 – Comentários

Este teste tem vantagens em relação ao χ^2 , são elas:

- ✓ Pode ser aplicado para pequenas amostras;
- ✓ É na maioria dos casos mais poderoso do que o χ^2 .

2.4 – TESTE DE LILLIEFORS

No caso em que se deseja testar normalidade e a média e a variância não são previamente especificadas, mas sim estimadas através dos dados da amostra, deve-se utilizar o teste de Lilliefors.

Este teste tem procedimento análogo ao Kolmogorov – Smirnov porém utiliza uma tabela própria e mais adequada a este tipo de situação. (**Tabela 6** – Campos)

Exercício: Resolva o exercício anterior estimando os parâmetros populacionais.

$$\bar{X} = 3593,478 \text{ e } S^2 = 191.601,8$$

Classes	f_i	x_i	$F(x_i)$	$S(x_i)$	$F(x_i) - S(x_i)$
2700-3000	13				
3000-3300	18				
3300-3600	24				
3600-3900	32				
3900-4200	17				
4200-4500	11				
	115				

2.5 – TESTE DE ALEATORIZAÇÃO DAS ITERAÇÕES

Para comprovar a propriedade de aleatoriedade de uma amostra utilizamos o teste de aleatorização, que faz uso da análise das iterações (sequência de símbolos idênticos).

Este teste, basicamente, verifica o número de iterações existentes na amostra; se o número de iterações é muito grande ou muito pequeno sugere-se falta de aleatoriedade.

2.5.1 - Exigência do Teste

Exige-se ao menos que os dados sigam uma escala nominal e que eles possam ser divididos em duas categorias.

2.5.2 – Método

- ✓ Seja n_1 o número de elementos da categoria 1, n_2 o número de elementos da categoria 2 e $N = n_1 + n_2$
- ✓ Se n_1 e $n_2 < 20$ verificamos o número r de iterações, obtemos os limites inferior e superior, que definem o número aceitável de iterações em caso de aleatoriedade (Tábua F1 e F2 – Siegel)

Se n_1 ou $n_2 > 20$ usamos a seguinte aproximação:

$$\mu_r = \frac{2n_1n_2}{n_1 + n_2} + 1 \qquad \sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}$$

$$Z = \frac{r - \mu_r}{\sigma_r}$$

E fazemos uso da tabela da distribuição normal padrão.

Isto pode ser necessário na falta de recurso computacional para este teste, embora o p-valor exato será sempre obtido fazendo uso do método para pequenas amostras.

Exemplo 1: 24 crianças foram avaliadas com relação a um índice de agressividade e em seguida converteram-se os dados em sinais positivos (+) e negativos (-), dependendo se o índice estava acima ou abaixo da mediana do grupo. Deseja-se verificar a aleatoriedade dos escores de agressividade com relação à ordem em que foram obtidos.

H_0 : Os escores de agressividade ocorrem de forma aleatória

H_a : Os escores de agressividade não ocorrem de forma aleatória

Sendo $N=24$, $n_1 = 12$ e $n_2 = 12$ tem-se a seguinte sequência de sinais:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
+	-	+	+	+	+	-	+	+	+	-	-	-	-	+	-	-	+	+	+	-	-	-	-
1	2	3	3	3	3	4	5	5	5	6	6	6	6	7	8	8	9	9	9	10	10	10	10

Conclusão: A **Tábua F** (Siegel) indica para $n_1 = n_2 = 12$ os limites 7 e 19, portanto $r = 10$ iterações não se encontra na região de rejeição. Os dados parecem ter sido gerados de forma aleatória.

Exemplo 2: Deseja-se verificar se a disposição de homens e mulheres numa fila de cinema se dá de forma aleatória. Foram observados 30 homens e 20 mulheres, que forneceram os seguintes resultados:

$$\begin{aligned}
 N &= 50 \\
 n_1 &= 30 \\
 n_2 &= 20 \\
 r &= 35
 \end{aligned}
 \quad
 \begin{aligned}
 \mu_r &= \frac{2 \cdot 30 \cdot 20}{50} + 1 = 25 \\
 \sigma_r &= \sqrt{\frac{2 \cdot 30 \cdot 20 (2 \cdot 30 \cdot 20 - 30 - 20)}{(30 + 20)^2 (30 + 20 - 1)}} = 3,356 \\
 Z &= \frac{35 - 25}{3,356} = 2,98
 \end{aligned}$$

Conclusão: Pela Tabela da Normal Padrão (teste bilateral) obtém-se um p-valor igual a 0,0028. Assim, rejeita-se a hipótese nula. Na fila observada, a ordem de homens e mulheres não foi aleatória.

Exercício: Suponha que um teste de QI tenha sido aplicado a 8 homens e 8 mulheres candidatos a um certo cargo de uma empresa. Os resultados foram os seguintes:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Sexo	H	H	M	H	M	M	M	H	M	M	M	H	H	H	H	M
QI	125	127	128	132	133	135	136	137	140	141	142	143	144	145	146	147

Existem evidências para acreditar que a ordem dos quocientes não seja aleatória?

2.5.3 - Comentários

Observe neste exercício que havia uma variável em escala intervalar e que esta foi transformada em nominal, levando a uma perda considerável de observações. Neste caso, é mais adequada a verificação da suposição de normalidade e se atendida aplicar um teste paramétrico.

Se não atendido o pressuposto, deve-se tentar a aplicação de uma técnica não-paramétrica para comparação de duas amostras, apresentada mais adiante.

III – TESTES PARA DUAS AMOSTRAS RELACIONADAS (dependentes)

Muitas vezes se tem interesse na avaliação do efeito de um tratamento. No entanto, utilizando dois grupos independentes para isso, é possível que ocorram disparidades entre os elementos dos grupos. Para evitar isso, é comum proceder algum tipo de pareamento entre os indivíduos. O tipo mais comum de pareamento é utilizando cada indivíduo como seu próprio controle, submetendo-o aos dois tratamentos em momentos diferentes. Outro tipo de pareamento é tentar selecionar, para cada par, indivíduos que sejam tão semelhantes quanto possível. Por exemplo: Gêmeos, órgãos (ouvidos, braços, pés etc..). Os testes não paramétricos apropriados a experimentos com amostras relacionadas são os seguintes:

3.1 – TESTE DE MCNEMAR

É um teste aplicável a situações do tipo “antes e depois” onde cada indivíduo é observado duas vezes: antes e depois de certo tratamento e deseja-se testar a sua eficiência. Cada indivíduo se encaixa em uma das duas categorias de resposta, antes e depois do tratamento. Este teste avalia se as mudanças de uma categoria para outra são significativas.

3.1.1. – Exigências do Teste

- ✓ Dados em escala nominal ou ordinal (dicotômicos);
- ✓ Os pares (X_i, Y_i) são mutuamente independentes onde X_i representa a situação “pré” e Y_i representa a situação “pós” do i -ésimo indivíduo.

3.1.2 – Método

As frequências observadas no experimento devem ser classificadas no quadro abaixo:

		Depois		
Antes		A	B	
A	a	b	a + b	
B	c	d	c + d	
	a + c	b + d		

A hipótese nula afirma que as mudanças ocorridas nos dois sentidos não são significativamente diferentes.

Devem-se considerar dois casos:

- ✓ $b + c \leq 20$, neste caso aplica-se o teste binomial considerando $p_0 = \frac{1}{2}$, $n = b + c$ e x a menor frequência entre b e c .
- ✓ $b + c > 20$, utiliza-se o teste de χ^2 com 1 grau de liberdade adaptado da seguinte forma:

$$\chi^2 = \frac{(b - c)^2}{b + c}$$

Mais uma vez aqui se enfatiza que neste caso os dois métodos devem fornecer resultados muito próximos e que a aplicação do método para grandes amostras só terá sentido caso não exista recurso computacional disponível para utilizar o procedimento para pequenas amostras.

Exemplo 1: Dois supermercados A e B disputam a preferência dos consumidores de uma cidade. O supermercado A realiza então uma campanha com distribuição de prêmios para aumentar a sua clientela. 100 consumidores foram acompanhados antes e depois da campanha, onde se perguntou a cada um deles sobre a sua preferência. Os resultados seguem abaixo:

		Depois	
Antes	A	B	
A	37	3	40
B	13	47	60
	50	50	

Teste agora a eficiência da campanha.

Como $b + c \leq 20$ aplicamos o teste Binomial:

$$N = b + c = 16$$

$$x = 3$$

$$\hat{p} = \frac{3}{16} = 0,1875$$

$$H_0 : P(A, B) = P(B, A)$$

$$H_a : P(A, B) < P(B, A)$$

$$P(B \leq 3) = 0,0106 \text{ (corresponde ao p-valor)}$$

Conclusão: Existem evidências para afirmar que a campanha foi eficiente.

Exemplo 2: Numa campanha política, após uma série de fatos divulgados com o objetivo de denegrir a imagem do candidato B em favor do candidato A, investigou-se as mudanças ocorridas com relação à preferência do eleitorado. Os resultados foram os seguintes.

Depois			
Antes	A	B	
A	83	47	130
B	18	52	70
	101	99	

$$H_0 : P(A, B) = P(B, A) \quad \text{versus} \quad H_a : P(A, B) > P(B, A)$$

$$\chi^2 = \frac{(47 - 18)^2}{47 + 18} = 12,93 \quad \text{cujo p-valor} < 0,001$$

Conclusão: Os fatos, na verdade, foram prejudiciais ao candidato A. Rejeita-se a hipótese de nulidade.

Exercício: Antes da implantação do divórcio no Brasil, 400 funcionários de uma empresa foram entrevistados, posicionando-os contra ou a favor. Depois de implantado, os mesmos funcionários foram novamente entrevistados. Houve mudança significativa de opinião dos entrevistados?

Depois			
Antes	Contra	Favor	
Contra	116	24	140
Favor	48	212	260
	164	236	

3.1.3 – Comentários

Quando os dados possuem escala nominal, a prova de McNemar é a única opção. Quando os dados permitem a aplicação de uma técnica paramétrica (como teste t para amostras pareadas), a eficiência desse teste será inferior.

Ex: para $b + c = 6 \Rightarrow \text{eficiência} = 95\%$ e $b + c > 30 \Rightarrow \text{eficiência} = 63\%$.

3.2 – TESTE DO SINAL

Este teste é aplicado a situações onde se deseja estabelecer comparação entre dois tratamentos sendo um deles o controle. Visando a redução da heterogeneidade entre os grupos é feito um pareamento das unidades amostrais. Este processo se faz necessário para evitar qualquer influência de fatores colaterais nos resultados da pesquisa. Este teste é na verdade uma Prova Binomial com $p_0 = \frac{1}{2}$.

3.2.1 – Exigência do Teste

- ✓ Os pares (X_i, Y_i) são mutuamente independentes para $i = 1, 2, \dots, n$;
- ✓ A escala de mensuração é ao menos ordinal.

3.2.2 – O Método

Consiste na atribuição de sinais + ou - para cada situação dependendo da diferença observada. Daí o nome Teste dos Sinais.

□ Pequenas Amostras

Caso $n \leq 25$, faz-se uso da Prova Binomial considerando $p_0 = \frac{1}{2}$, n sendo o número de pares, e x o número de sinais que corresponde à menor frequência.

Deve-se calcular $P(X \leq x)$.

□ **Grandes Amostras***

Quando $n > 25$ utilizamos a Aproximação Normal fazendo

$$z = \frac{x - \frac{n}{2}}{\sqrt{\frac{n}{4}}} = \frac{2x - n}{\sqrt{n}}$$

* mais uma vez, só há sentido na utilização da aproximação no caso de indisponibilidade de recurso computacional.

□ **Empates**

Algumas vezes pode ocorrer $X_i = Y_i$, ou seja, não há diferença entre os escores do par i . Neste caso, o indivíduo i deve ser eliminado da análise.

Exemplo 1: Para se testar a eficiência de um novo herbicida, foram analisadas 10 áreas. Em cada área, observou-se se a quantidade de ervas daninhas (em gramas) aumentou (+) ou diminuiu(-). Os dados foram os seguintes:

Tabela 3.1 – Resultado da aplicação do herbicida nas 10 áreas.

Área	Antes	Depois	Sinal
1	115,4	98,4	+
2	121	73,6	+
3	112,3	65,9	+
4	78,7	42,1	+
5	65,6	77,2	-
6	213,5	104,0	+
7	157,5	82,8	+
8	80,7	59,4	+
9	142,8	102,6	+
10	100,3	53,7	+

Verifique se o herbicida foi eficiente.

H_0 :O herbicida não é eficiente

H_a :O herbicida é eficiente

Como a amostra é pequena, usa-se a prova Binomial onde $n = 10$, $p = \frac{1}{2}$ e $x = 1$ ($\hat{p} = 0,10$)

$$p\text{-valor} = P(X \leq 1) = \binom{10}{1} \frac{1}{2}^{10} + \binom{10}{0} \frac{1}{2}^{10} = 0,0107.$$

Conclusão: Rejeita-se a hipótese nula, existem evidências para afirmar que o herbicida é eficiente, reduz significativamente a quantidade de ervas daninhas.

Exemplo 2: Extraíu-se uma amostra de 100 adultos de uma comunidade e perguntou-se a cada um a opinião sobre o tipo de punição a ser aplicado em casos de delinquência juvenil (se mais forte ou mais fraca). Em seguida exibiu-se um filme sobre instituições de reabilitação, depois se repetiu a pergunta. Os resultados seguem no quadro abaixo.

Antes	Depois	
	-	+
+	59	7
-	8	26

H_0 : O filme não produz efeito

H_a : O filme produz efeito

A escala de mensuração é ordinal e a amostra pode ser considerada grande. Como houve 15 empates, estes são excluídos da análise. Assim, sob H_0 , é de se esperar que metade dos demais 85 entrevistados mudem sua opinião de - para + e a outra metade, de + para -.

Assim,

$$z = \frac{59 - (42,5)}{\sqrt{\frac{85}{4}}} = 3,58 \text{ e o p-valor} = 0,00034 \text{ (bilateral).}$$

Conclusão: Rejeita-se H_0 , o filme teve efeito significativo sobre a atitude dos adultos.

Exercício: Uma firma submeteu 8 de seus empregados a um treinamento intensivo sobre um novo método a ser implantado, visando maior rendimento na produção. O resultado individual, informado pelo gerente segue abaixo:

Tabela 3.2 – Rendimento dos empregados após o treinamento, segundo o gerente.

Empregado	Antes	Depois	Sinal
1	6	9	
2	7	6	
3	8	11	
4	3	7	
5	6	9	
6	4	5	
7	7	4	
8	7	7	

Compare a eficiência do método.

3.2.3 – Comentários

A prova dos sinais exige nível de mensuração pelo menos em escala ordinal, ao contrário do Teste de McNemar que aceita um nível de mensuração em escala nominal.

3.2.4 – Poder – Eficiência

$$n = 6 \Rightarrow 95\%$$

$$n > 6 \Rightarrow 63\%$$

3.3 – TESTE DE WALSH

Este teste se aplica a dados pareados onde se supõe que diferenças entre os escores (d_i 's) provém de distribuição simétrica, não necessariamente Normal, ou seja, a média é aproximadamente igual à mediana.

O objetivo, mais uma vez, é verificar o efeito de um tratamento aplicado a indivíduos (antes e depois)

3.3.1 – Exigências do Teste

- ✓ Nível de Mensuração em escala intervalar

3.3.2 – O Método

Considere a tabela abaixo, onde são ilustrados os dados observados e os escores de diferenças.

Tabela 3.3 – Dados observados e escores de diferença

	<i>X</i>	<i>Y</i>	<i>D</i>
1	x_1	y_1	d_1
2	x_2	y_2	d_2
3	x_3	y_3	d_3
⋮	⋮	⋮	⋮
<i>J</i>	x_j	y_j	d_j
⋮	⋮	⋮	⋮
<i>N</i>	x_n	y_n	d_n

Devemos ordenar os d_i 's em ordem crescente, levando em conta o sinal (+ ou -). As hipóteses a serem testadas estão definidas a seguir:

$$H_0 : \text{Mediana} = 0 \quad \text{versus} \quad H_a : \text{Mediana} \begin{cases} \neq 0 \\ > 0 \\ < 0 \end{cases}$$

Após ordenar os escores d_i 's tais que $d_1 \leq d_2 \leq \dots \leq d_j \leq \dots \leq d_n$ consulta-se a Tabela H – Siegel que indicará a regra de decisão do teste. Esta regra depende do tamanho da amostra n , se o teste é unilateral ou bilateral, e dos valores d_i 's.

Exemplos de utilização da Tabela

$$\begin{aligned}
 \text{a) } n=5, H_a : \mu > 0 \text{ Se } & \begin{cases} \frac{1}{2}(d_1 + d_2) > 0 & p = 0,062 \\ d_1 > 0 & p = 0,031 \end{cases} \\
 \text{b) } n=6, H_a : \mu < 0 \text{ Se } & \left\{ \max \left[d_5 \text{ ou } \frac{1}{2}(d_4 + d_6) \right] \right\} < 0
 \end{aligned}$$

Exemplo 1 – Quinze sujeitos participaram de um experimento que consistia em aprender 10 sílabas sem sentido e associar a 5 delas uma reação negativa (choque elétrico). Após 40 horas foi solicitado aos indivíduos que evocassem a lista com as 10 sílabas. O pesquisador suspeita que as sílabas não associadas aos choques serão mais lembradas:

H_0 : Os indivíduos lembraram igualmente dos dois grupos de sílabas.

H_a : As sílabas não associadas a choques serão mais lembradas.

<i>Indivíduo</i>	<i>Sem Choque</i>	<i>Com Choque</i>	<i>D</i>
A	5	2	3
B	4	2	2
C	3	0	3
D	5	3	2
E	2	3	-1
F	4	2	2
G	2	3	-1
H	2	1	1
I	4	1	3
J	4	3	1
K	3	4	-1
L	1	2	-1
M	5	2	3
N	3	4	-1
O	1	0	1

Pela tabela H, considerando $n = 15$, teste unilateral e:

$$d_1 = d_2 = d_3 = d_4 = d_5 = -1$$

$$d_6 = d_7 = d_8 = 1$$

$$d_9 = d_{10} = d_{11} = 2$$

$$d_{12} = d_{13} = d_{14} = d_{15} = 3$$

$$\min\left\{\frac{1}{2}(d_1 + d_{12}), \frac{1}{2}(d_2 + d_{11})\right\} > 0 \quad \text{Verdadeira}$$

$$\min\left\{\frac{1}{2}(d_1 + d_{11}), \frac{1}{2}(d_2 + d_{10})\right\} > 0 \quad \text{Verdadeira}$$

$$\min\left\{\frac{1}{2}(d_1 + d_{10}), \frac{1}{2}(d_5 + d_6)\right\} > 0 \quad \text{Falsa}$$

$$p(\text{unilateral}) = 0,023 \quad \text{e} \quad p(\text{bilateral}) = 0,047$$

Conclusão: Rejeita-se a hipótese nula. O número de sílabas lembradas “sem choque” é significativamente superior ao número de sílabas lembradas “com choque”.

Exercício: Trabalhando-se com duas variedades de cana-de-açúcar, constatou-se que a variedade B produzia acima de 15% a mais do que a variedade A. Para a comprovação de tal resultado, foram consideradas as unidades de produção (t/he) das duas variedades, em 10 locais distintos, segundo a tabela abaixo:

Tabela 3.4 – Distribuição da produção das variedades de cana-de-açúcar em 10 locais diferentes

Locais	X	Y	$X^* = 1,15 * X$	Y-X*	d	D
	Variedade A	Variedade B	1,15 * Var.A			
1	75,8	93,5				
2	84,3	100,2				
3	78,4	95,4				
4	81,5	98,6				
5	85,6	102,5				
6	68,7	80,5				
7	70,4	73,6				
8	79,7	93,5				
9	78,3	94,6				
10	83,5	105,1				

Assumindo que as diferenças se distribuem de forma simétrica, construa as hipóteses apropriadas comparando as duas variedades de cana-de-açúcar.

3.3.3 – Comentário

Comparada à prova paramétrica t , a prova de Walsh acusa poder eficiência de 95% para a maioria dos valores de n e α .

3.4 – TESTE DE WILCOXON

Este teste para dados pareados, ao invés de considerar apenas o sinal das diferenças entre os pares, considera o valor dessas diferenças, sendo assim um teste não paramétrico dos mais poderosos e “populares”. Este teste é o primeiro a ser apresentado que utiliza ranks, pois atribui postos ao ordenar as diferenças entre os pares.

3.4.1 – Exigências do Teste

- ✓ Os pares (X_i, Y_i) são mutuamente independentes;
- ✓ As diferenças d_i são variáveis contínuas, com distribuição simétrica;
- ✓ Nível de mensuração em escala intervalar.

3.4.2 – O Método

Considere, as diferenças d_i 's onde $d_i = Y_i - X_i$. Devemos ordenar os d_i 's, porém agora sem considerar o sinal da diferença (em módulo).

□ Pequenas Amostras ($n \leq 25$)

Considerar T como a menor soma de postos de mesmo sinal, ou seja $T = \min(T_-, T_+)$. Em seguida, consultam-se os valores de T tabelados e respectivos p-valores.

Compara-se então o valor de T calculado com aqueles tabelados. Caso o valor calculado seja inferior ao valor tabelado, deve-se rejeitar a hipótese nula.

□ Grandes Amostras ($n > 25$)

Neste caso T tem distribuição aproximadamente Normal e pode-se usar a aproximação considerando:

$$\hat{\mu}_T = \frac{n(n+1)}{4} \quad e \quad \hat{\sigma}_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

Calcula-se assim a estatística $z = \frac{T - \hat{\mu}_T}{\hat{\sigma}_T}$ e compara-se com os valores tabelados da distribuição de Z (Normal Padrão). **Para ser usada em último caso, quando não existe recurso computacional adequado.**

□ **Empates**

Considerar duas situações:

- a) Quando $X_i = Y_i$, ou seja, a informação pré equivale a informação pós para um mesmo indivíduo, descarta-se este par da análise e redefine-se n como sendo o número de pares tais que $X_i \neq Y_i$.
- b) Quando duas ou mais d_i 's tem o mesmo valor atribui-se como posto a média dos postos que seriam atribuídos a eles caso não ocorresse empate.

Exemplo

D_i	 d_i 	Postos	Cálculo para Empates
-5	5	2	$\Rightarrow \frac{1+2+3}{3} = 2$
5	5	2	
5	5	2	
7	7	4	$\Rightarrow \frac{6+7}{2} = 6,5$
10	10	5	
-13	13	6,5	
13	13	6,5	
15	15	8	

Exemplo 1: Num experimento estudando a prevalência de oclusão normal em crianças do sexo masculino e feminino de 11 escolas, obtiveram-se os seguintes resultados:

Escola	% de oclusão Normal		d_i	$ d_i $	p
	Masculino (X)	Feminino (Y)			
1	8,7	7,7	1,0	1,0	4
2	18,6	9,6	9,0	9,0	9
3	8,0	16,0	-8,0	8,0	6
4	12,9	13,4	-0,5	0,5	2
5	10,9	9,6	1,3	1,3	5
6	13,4	13,0	0,4	0,4	1
7	11,9	23,7	-11,8	11,8	11
8	14,3	6,2	8,1	8,1	7
9	20,0	9,6	10,4	10,4	10
10	14,4	13,8	0,6	0,6	3
11	6,6	15,1	-8,5	8,5	8

Verifique se há influência do sexo na % de ocorrência de oclusão normal.

Aplicando o teste de Wilcoxon testam-se as seguintes hipóteses:

$$H_0 : \text{Masculino} = \text{Feminino} \quad \text{versus} \quad H_a : \text{Masculino} \neq \text{Feminino}$$

Considerando $T=27$, $\alpha = 0,05$ (bilateral) e $n = 11$, a tabela de valores críticos fornece um valor limite de 11. Assim p-valor deverá ser maior que 0,05.

Conclusão: Não se pode rejeitar H_0 , ou seja, a porcentagem de oclusão normal não difere entre os sexos.

Pelo teste de Wilcoxon: é recomendável quando é possível quantificar as diferenças.

Pelo Teste dos Sinais: é recomendável quando se conhece somente os sinais das diferenças.

Exemplo 2: No exercício resolvido utilizando o teste de Walsh não se rejeitou H_0 . Aplicando o teste de Wilcoxon:

$T=9$ e $n = 10$ e o p-valor = 0,032 (teste unilateral)

Conclusão: A decisão aqui seria de rejeitar a hipótese nula.

Exemplo 3: Considere o quadro de notas abaixo referente a estudo para comparar a eficiência de um novo método de aprendizagem.

<i>Aluno</i>	<i>Nota 1</i>	<i>Nota 2</i>	<i>di</i>	<i>Postos</i>
1	5.1	5.8	0.7	10.5
2	4.2	4.7	0.5	7.5
3	3.3	6.1	2.8	18.5
4	6	7.4	1.4	14
5	7.1	6.5	-0.6	9
6	3.2	3.2	0	-
7	2.8	4.5	1.7	15.5
8	1.4	3.5	2.1	17
9	0.8	4	3.2	20
10	5.6	5.5	-0.1	1.5
11	3.8	3.9	0.1	1.5
12	4.7	4.5	-0.2	3
13	6.2	5	-1.2	13
14	5.4	5.4	0	-
15	1.2	6	4.8	23
16	2.2	5	2.8	18.5
17	1.9	2.2	0.3	5
18	5.4	1.8	-3.6	22
19	5.5	5.2	-0.3	5
20	5.2	6	0.8	12
21	3.8	4.5	0.7	10.5
22	4.2	4.7	0.5	7.5
23	1.8	5.1	3.3	21
24	2.6	2.9	0.3	5
25	4.3	2.6	-1.7	15.5
26	1.4	6.5	5.1	24

Utiliza-se a aproximação para a normal dado que a amostra aqui considerada é grande.

H_0 :O novo método não é eficiente

H_a :O novo método foi eficiente

$$\mu_T = \frac{24(24+1)}{4} = 150 \quad \sigma = \sqrt{\frac{24 \cdot 25 \cdot 49}{24}} = 35 \quad \Rightarrow T = 69 \quad Z = -2,31$$

Conclusão: Como o valor de p é igual a 0,0104 existem evidências para se acreditar que o novo método é eficiente.

Exercício: Um grupo de 8 indivíduos se submete a um estímulo. A tabela abaixo apresenta as medidas de pressão sanguínea (mm/Hg) antes e depois do estímulo. O pesquisador desconfia que os estímulos aumentem a pressão sanguínea. Teste esta afirmação.

<i>Paciente</i>	<i>Antes (Xi)</i>	<i>Depois (Yi)</i>	<i>di (Yi - Xi)</i>	<i>Posto</i>
1	118	127		
2	120	128		
3	128	136		
4	124	131		
5	130	135		
6	136	138		
7	128	125		
8	140	136		

Conclusão: Existem evidências de que existe diferença significativa entre a pressão sanguínea medida antes e depois do estímulo.

3.4.3 - Comentários

Quando as suposições paramétricas estão atendidas, a eficiência do teste de Wilcoxon é de aproximadamente 95% tanto para pequenas como para grande amostras.

Embora semelhante, é uma prova mais poderosa que a de Walsh.

3.5 – TESTE DE ALEATORIZAÇÃO

Este teste, bastante poderoso, é utilizado para se obter a probabilidade exata, sob H_0 , da ocorrência de um resultado qualquer relacionado a pares de dados. Não se exige aqui verificação da suposição de Normalidade. Este teste usa basicamente todas as combinações possíveis de resultados (2^n), levando em conta o sinal, e verifica quantas somas dessas combinações se encontram acima de certo valor crítico, valor este definido pelo nível de significância.

Se o resultado observado estiver dentro da região de rejeição, rejeita-se H_0 .

3.5.1 - Exigência do Teste

- ✓ Escala de mensuração em nível intervalar.

3.5.2 – O Método

□ Pequenas Amostras (n <12)

Primeiramente calcula-se o número de combinações possíveis de resultados variando-se o sinal das diferenças encontradas (2^n). Define-se, pelo nível de significância, o número de combinações que irão definir a região de rejeição. Compara-se então a soma de escores observada com os valores críticos das regiões de rejeição.

□ Grandes Amostras

- Se $12 \leq n < 25$, não é prático utilizar o teste de aleatorização. Sugere-se aplicação da prova de Wilcoxon;
- Se $n \geq 25$ e as diferenças não variam muito, tal que $\frac{(d_{\max})^2}{\sum d_i^2} \leq \frac{5}{2n}$ aplica-se a aproximação normal, utilizando-se:

$$\text{Média} = 0 \quad \text{desvio padrão} = \sqrt{\sum d_i^2} \quad \text{e} \quad z = \frac{\sum d_i}{\sqrt{\sum d_i^2}}$$

Caso $\frac{(d_{\max})^2}{\sum d_i^2} > \frac{5}{2n}$ deve-se aplicar o teste de Wilcoxon.

Exemplo1: Considere a comparação de dois momentos relativos à um tratamento aplicado a indivíduos de forma pareada, segundo tabela abaixo.

<i>Indivíduo</i>	<i>Antes</i>	<i>Depois</i>	<i>Diferença</i>
1	38	19	+19
2	46	19	+27
3	61	62	-1
4	18	12	+6
5	35	28	+7
6	72	59	+13
7	24	28	-4
8	47	44	+3
Soma das diferenças			+70

Assim, existem $2^8 = 256$ resultados diferentes se forem alterados os sinais das diferenças observadas. Exemplo: +19, +27, -1, +6, +7, +13, -4, -3; é um resultado diferente do observado e assim por diante.

Se for considerado o nível de significância de $\alpha = 0,05$, tem-se $256 \times 0,05 = 12,8$ que é o número de resultados que deverá pertencer a região de rejeição. Se forem considerados apenas os 12 resultados mais extremos, tem-se $\alpha = 0,047$, ou seja, $\left(\frac{12}{256}\right)$. Definem-se então as hipóteses:

H_0 : O tratamento não surtiu efeito;

H_a : O tratamento surtiu efeito.

A região de rejeição é bilateral, ficando dessa forma, 6 resultados para as somas de escores positivos e os outros 6 para somas de escores negativos.

Dado que se obteve uma soma de diferenças igual a +70, esta soma cairá na região de rejeição caso esteja entre as 6 somas extremas positivas, são elas:

	Resultados								$\sum d_i$
1	+19	+27	+1	+6	+7	+13	+4	+3	80
2	+19	+27	-1	+6	+7	+13	+4	+3	78
3	+19	+27	+1	+6	+7	+13	+4	-3	74
4	+19	+27	+1	+6	+7	+13	-4	+3	72
5	+19	+27	-1	+6	+7	+13	+4	-3	72
6	+19	+27	-1	+6	+7	+13	-4	+3	70

O resultado observado se encontra, portanto na região de rejeição $\sum d_i = 70$

Conclusão: Rejeita-se H_0 , portanto existe efeito significativo do tratamento.

Exercício: Um grupo de 6 pilotos da Aeronáutica testou dois novos altímetros digitais quanto à quantidade de erros de leitura. Os escores para cada piloto, deste teste de indução ao erro, foram os seguintes :

	Altímetro A	Altímetro B	Diferença
1	73	78	
2	78	81	
3	82	83	
4	71	74	
5	81	80	
6	90	81	

Teste a diferença entre os altímetros.

3.5.3 – Comentários

Dado que o teste de aleatorização utiliza todas as informações da amostra, o seu poder-eficiência é de 100%. Recomenda-se o uso desta prova sempre que n não afete a praticidade do teste.

Na presença de empates, considera-se n como sendo o número de casos em que $d_i \neq 0$.

IV – TESTES PARA DUAS AMOSTRAS INDEPENDENTES

Estes testes se aplicam a planos amostrais onde se deseja comparar dois grupos independentes. Esses grupos podem ter sido formados de duas maneiras diferentes:

- Extraíu-se uma amostra da população A e outra amostra da população B;
- Indivíduos da mesma população foram alocados aleatoriamente a um dos dois tratamentos em estudo.

Diferente do caso pareado, não se exige que as amostras tenham o mesmo tamanho.

Quando as suposições paramétricas são atendidas, o teste t de Student para amostras independentes é mais adequado para comparação das médias dos grupos.

Quando estas não são atendidas, deve-se aplicar alguma das provas não-paramétricas apresentadas a seguir:

4.1 – TESTE EXATO DE FISHER

Este teste faz uso de tabelas de contingência 2 X 2, para se comparar 2 grupos. É indicado quando o tamanho das duas amostras independentes é pequeno e consiste em determinar a probabilidade exata de ocorrência de uma frequência observada, ou de valores mais extremos.

4.1.1 – Exigências do Teste

- ✓ Amostras aleatórias e independentes;
- ✓ Duas classes mutuamente exclusivas;
- ✓ Nível de Mensuração em escala nominal ao menos.

4.1.2 – O Método

Considere a definição de duas amostras I e II, agrupadas em duas classes – e +.

	-	+	
I	A	B	A+B
II	C	D	C+D
	A+C	B+D	N

Calcula-se, em seguida, a probabilidade de interesse. A probabilidade de ocorrência das frequências observadas nas caselas acima, se obtém com o uso da distribuição hipergeométrica, ou seja:

$$P = \frac{\binom{A+C}{A} \binom{B+D}{B}}{\binom{N}{A+B}}$$

Ou da mesma forma:

$$P = \frac{(A+B)!(C+D)!(A+C)!(B+D)!}{N!A!B!C!D!}$$

Como a hipótese deseja testar a probabilidade de ocorrência de uma situação mais extrema, devem-se calcular as probabilidades referentes às frequências observadas e das demais situações extremas.

Exemplo 1: Numa classe de 24 alunos, comparou-se o rendimento de estudantes provenientes de escolas particulares e escolas públicas. Os resultados seguem abaixo:

	Acima da Média	Abaixo da Média	Total
A (Particular)	5	7	12
B (Pública)	10	2	12
Total	15	9	24

$$H_0 : P(A) = P(B)$$

$$H_a : P(A) < P(B)$$

$$p_2 = \frac{12!12!15!9!}{24!5!7!10!2!} = 0,04$$

$$p_1 = \frac{12!12!15!9!}{24!4!8!11!1!} = 0,0045$$

$$p_0 = \frac{12!12!15!9!}{24!3!9!12!0!} = 0,0002$$

Assim $p = p_0 + p_1 + p_2 = 0,0447$ (teste unilateral) e $2p = 0,0894$ (teste bilateral)

Considerando o caso unilateral, rejeita-se H_0 . Os alunos de escola pública aparentemente têm um rendimento superior aos alunos da escola particular.

Exemplo 2: Num estudo sobre fecundidade de duas raças bovinas foram feitos ensaios de fecundidade obtendo-se os seguintes resultados:

	<i>Fecundos</i>	Não - Fecundos	<i>Total</i>
Raça A	3	7	10
Raça B	4	1	5
Total	7	8	15

Verifique se as duas raças diferem quanto à fecundidade.

$$H_0 : P(A) = P(B)$$

$$H_a : P(A) \neq P(B)$$

$$p_1 = \frac{10!5!7!8!}{15!3!7!4!1!} = 0,093$$

$$p_0 = \frac{10!5!7!8!}{15!2!8!5!0!} = 0,007$$

Assim $p = p_0 + p_1 = 0,1$ (teste unilateral) e $p = 0,2$ (teste bilateral)

Não se rejeita H_0 . Não há diferença significativa entre as duas raças com relação à fecundidade.

Exercício: Numa pesquisa sobre desquites, realizada entre as classes média e alta, foram obtidos os seguintes resultados.

<i>Classe</i>	Amigável	Não - Amigável	<i>Total</i>
Alta	6	4	10
Média	2	8	10
Total	8	12	20

A proporção de desquites amigáveis é maior na classe alta?

4.1.3 – Comentário

O Teste Exato de Fisher é a alternativa ao caso de duas amostras independentes, quando o tamanho da amostra é pequeno, pois nesse caso o teste χ^2 não se aplica.

4.2 – TESTE DE χ^2

Aplicável a dados representados em forma de frequência para detectar significância estatística da diferença entre dois grupos independentes. Tem as mesmas características do teste para uma amostra, com os mesmos procedimentos e restrição com relação ao tamanho das frequências.

4.2.1 – Exigência do Teste

- ✓ Nível de mensuração em escala nominal (ao menos);
- ✓ $N > 20$ e frequências esperadas superiores a 5 quando ocorre o caso 2x2;
- ✓ Se $k > 2$, o número de células com frequência esperada inferior a 5 deve ser menos de 20% do total de células.

4.2.2 – O Método

Testamos a hipótese nula obtendo a estatística

$$\chi^2 = \sum \frac{(O - E)^2}{E} \quad \text{onde} \quad \begin{cases} O \Rightarrow \text{Frequência Observada} \\ E \Rightarrow \text{Frequência Esperada} \end{cases}$$

Os valores críticos de χ^2 dependem do nível de significância e dos graus de liberdade:
 $g.l. = (k - 1)$

Exemplo 1: Antes da descoberta da teoria da doença causada por germes a taxa de mortalidade após cirurgias era muito alta devido à infecção. Louis Pasteur e Joseph Lister foram os grandes responsáveis pela teoria dos germes. Lister achava que se ácido carbólico fosse usado como um desinfetante, a chance de sobrevivência do paciente poderia ser melhorada. Ele o utilizou para desinfetar tudo na sala de operação que pudesse ter contato com o paciente. As roupas do paciente também foram desinfetadas. Lister comparou 40 operações (amputações) nas quais este procedimento foi utilizado, com 35 amputações nas quais ele não tinha sido usado. Os resultados são sumarizados na tabela abaixo:

	Paciente viveu	Paciente morreu	Total
Ácido usado	34	6	40
Ácido não usado	19	16	35
Total	53	22	75

Ao nível de significância de 1%, teste se o uso ou não do ácido carbólico interfere no resultado da cirurgia (paciente viveu ou morreu), ou seja, se a proporção de pacientes sobreviventes com o uso do ácido é maior do que a proporção de pacientes sobreviventes sem o uso do ácido.

H_0 : A taxa de mortalidade não se altera com o uso do ácido carbólico;

H_a : A taxa de mortalidade se altera com o uso do ácido carbólico.

Observados	+	-
Usou	34	6
Não usou	19	16

Esperados	+	-
Usou	28,27	11,73
Não usou	24,73	10,27

$$\chi^2 = \sum \frac{(O - E)^2}{E} = 1,16 + 2,8 + 1,33 + 3,2 = 8,49$$

com 1 grau de liberdade sendo que $0,001 < p < 0,01$

Conclusão: Rejeitamos a hipótese nula, pois o p-valor é baixo. A proporção de pacientes sobreviventes com uso do ácido carbólico na desinfecção é significativamente superior à proporção de sobreviventes sem uso do ácido.

Exemplo 2: Considere um estudo em que dois grupos de indivíduos foram formados: um de pessoas baixas e outro de pessoas altas. Essas pessoas foram avaliadas com relação ao seu grau de liderança. O objetivo é saber, se a altura da pessoa definirá se uma pessoa será líder ou não. Os dados seguem abaixo:

	Baixo	Alto
Líder	12 (19,9)	32 (24,1)
Liderado	22 (16,3)	14 (19,7)
Não-Classificado	9 (6,8)	6 (8,2)

H_0 : grau de liderança (baixo) = alto

H_a : grau de liderança (baixo) \neq alto

$$\chi^2 = 10,67$$

$$g.l. = (3 - 1)(2 - 1) = 2$$

$$0,001 < p - \text{valor} < 0,01$$

Conclusão: Rejeita-se a hipótese nula pois o p-valor é baixo. Pode-se afirmar que, estatisticamente, a altura das pessoas define a liderança. O fato de a pessoa ser alta aumenta sua probabilidade de se tornar líder.

Exercício

Em dois hospitais foi coletada a quantidade de pacientes em 5 diagnósticos diferentes conforme tabela abaixo. Teste se há diferença entre os hospitais.

	D 1	D 2	D 3	D 4	D 5	Total
Hospital 1	34	16	12	4	2	68
Hospital 2	12	6	6	4	0	28
Total	46	22	18	8	2	96

	d1	d2	d3	d4	d5
h1					
h2					

4.2.3 – Comentários

Deve-se destacar, sobre este teste, a sua aplicabilidade no caso em que as frequências são suficientemente grandes e o nível de mensuração é pelo menos nominal.

No caso em que certas frequências são pequenas, utiliza-se o critério de agrupar classes, tomando-se o cuidado de preservar o sentido das classes. Por exemplo, em uma situação com mensuração nominal, pode não ser correto agrupar certas classes:

Negro	Amarelo	Branco	Mestiço	Pardo
5	3	12	8	3

O Teste Qui-quadrado é sempre bilateral.

Sugere-se que, quando $n < 20$, se utilizar a prova exata de Fisher. Quando n é grande o poder do teste tende para 1.

4.3 – TESTE DA MEDIANA

Este teste, como o nome sugere, se utiliza da mediana para efetuar a comparação de dois grupos independentes. Assim, o teste se baseia no cálculo da probabilidade de que os dois grupos provenham de populações com a mesma mediana.

4.3.1 – Exigência do Teste

- ✓ O nível de mensuração em escala ordinal

4.3.2 – O Método

Primeiramente deve-se obter o valor mediano referente aos dados fornecidos pelos dois grupos, conjuntamente (mediana combinada).

Em seguida deve-se construir a tabela de contingência 2 x 2 abaixo.

	Grupo I	Grupo II
Acima da mediana	A	B
Abaixo ou igual à mediana	C	D

Com o auxílio da distribuição hipergeométrica calcula-se a probabilidade de ocorrência dessa disposição de frequências.

$$P(A, B) = \frac{\binom{A+C}{A} \binom{B+D}{B}}{\binom{A+B+C+D}{A+B}} e \begin{cases} n_1 = A+C \\ n_2 = B+D \end{cases}$$

Como a hipótese deseja testar a probabilidade de ocorrência de uma situação mais extrema, devemos calcular as probabilidades referentes às frequências observadas e das demais situações extremas, assim como no Teste exato de Fisher.

Se $P(A, B)$ é inferior ao nível de significância, rejeita-se H_0 .

Exemplo 1: Os dados apresentam escores de um teste de inteligência emocional aplicado a funcionários de dois setores de uma empresa: Recursos Humanos e Controle de Qualidade. Verifique a hipótese de que as amostras provêm de populações com a mesma mediana.

H_0 : Os escores de inteligência emocional dos dois setores provêm de populações com a mesma mediana.

H_a : Os escores de inteligência emocional dos dois setores não provêm de populações com a mesma mediana.

RH	CQ
15	8
12	7
13	14
18	10
9	8
11	6
12	7
13	6
14	8
14	4
12	
12	
11	
10	
n₁=14	n₂=10

A mediana obtida neste caso é igual a 11

	RH	CQ	
Acima	10	1	11
Abaixo ou igual	4	9	13
	14	10	

$$p_1 = \frac{\binom{14}{10} \binom{10}{1}}{\left[\begin{matrix} 24 \\ 11 \end{matrix} \right]} = 0,004$$

$$p_0 = \frac{\binom{14}{11} \binom{10}{0}}{\left[\begin{matrix} 24 \\ 11 \end{matrix} \right]} \approx 0$$

$P(A,B) = 0,004$

Rejeita-se H_0 . Existe diferença significativa entre os escores emocionais dos dois setores.

Exercício: Nos mesmos dois setores da empresa citada anteriormente, registrou-se o salário de alguns funcionários. Verifique agora se os salários provêm de populações com mesma mediana.

H_0 : Os salários dos dois setores provêm de populações com a mesma mediana.

H_a : Os salários dos dois setores não provêm de populações com a mesma mediana.

RH	CQ
(em salários mínimos)	
4	11
3	10
8	7
3	6
5	5
7	8
2	9
	10
$n_1=7$	$n_2=8$

A Mediana neste exemplo será igual a

	RH	CQ	Total
Acima	1	5	6
Abaixo ou igual	6	3	9
Total	7	8	

4.3.3 – Comentários

Na verdade este teste recai na aplicação do teste exato de Fisher quando o tamanho da amostra é pequena ($n < 20$) e o teste χ^2 se $n \geq 20$.

Para $n_1 + n_2$ pequeno o poder eficiência é de 95%

Para $n_1 + n_2$ grande o poder eficiência é de 63%

4.4 – TESTE U DE MANN – WHITNEY

Esta prova se aplica na comparação de dois grupos independentes, para verificar se eles pertencem ou não à mesma população.

4.4.1 – Exigência do Teste

- ✓ Nível de Mensuração em escala ordinal (pelo menos);
- ✓ Amostras independentes.

4.4.2 – O Método

Primeiramente ordenam-se os valores misturados dos dois grupos, em ordem crescente. Considera-se como 1 o grupo com menos observações e como 2 o grupo com mais observações. Em seguida, obtêm-se as somas de postos de cada grupo, respectivamente (R_1 e R_2).

Calcula-se

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 \quad \text{e} \quad U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

O valor da estatística do teste (U) será o menor entre U_1 e U_2 .

- Para $n_2 < 9$, utiliza-se a Tábua J – Siegel*.
- Quando $9 \leq n_2 \leq 20$ utiliza-se a Tábua K – Siegel*.
- Quando $n_2 > 20$, utiliza-se a aproximação Normal* dada por:

$$\mu_U = \frac{n_1 n_2}{2} \quad \sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} \quad z = \frac{U - \mu_U}{\sigma_U}$$

* o uso de tabelas e aproximações pela normal são recomendados apenas quando não houver recursos computacionais adequados. O ideal é sempre o uso de um software que forneça o p-valor exato.

Se $U < U$ crítico (tabelado), rejeita-se H_0 .

Exemplo 1: Num experimento onde se comparavam ratos treinados e ratos sem treinamento (controle), com relação a um teste de aprendizado (tempo em segundos para sair de um labirinto), registraram-se os seguintes valores:

Ratos Treinados (E)	78	64	75	45	82
Postos (E)	7	4	6	1	8
Ratos Controle (C)	110	70	53	51	
Postos (C)	9	5	3	2	

H_0 : Não há diferença entre os ratos com e sem treinamento

H_a : Os ratos treinados levam menos tempo para realizar a tarefa

$$U_1 = (4 \times 5) + \frac{(4 \times 5)}{2} - 19 = 11 \text{ e } U_2 = (4 \times 5) + \frac{(5 \times 6)}{2} - 26 = 9. \text{ Então } U = 9.$$

Pela Tábua J – Siegel, considerando $n_2 = 5$ e $U = 9$, obtêm-se um p-valor = 0,452 (unilateral), portanto não se pode rejeitar a hipótese de que os ratos treinados são iguais aos controles.

Exemplo 2: Dois tipos de solução química, A e B, foram ensaiadas para a determinação do Ph. As análises de 10 amostras de cada solução estão apresentadas na tabela que segue.

	A	Posto (A)	B	Posto (B)
$H_0: P_{HA} = P_{HB}$	7,49	13	7,28	2
	7,35	4,5	7,35	4,5
$H_a: P_{HA} > P_{HB}$	7,54	19	7,52	17,5
	7,48	11	7,50	14,5
	7,48	11	7,38	7
	7,37	6	7,48	11
	7,51	16	7,31	3
	7,50	14,5	7,22	1
	7,52	17,5	7,41	8
	7,56	20	7,45	9
	$R_A = 132,5$		$R_B = 77,5$	

$$U = 100 + \frac{110}{2} - 132,5$$

$$U = 22,5$$

$$U' = 100 + \frac{110}{2} - 77,5$$

$$U' = 77,5$$

O valor crítico para $n_1 = n_2 = 10$ em que $\alpha = 0,05$ (teste unilateral) será $U_c = 27$.

Conclusão: Com um p-valor $\approx 0,025$ existem evidências de que o Ph da solução A é maior que o Ph da solução B.

Exercício: Num ensaio sobre competição de variedades de tomates foram consideradas as produções individuais (em kg) de 15 plantas da variedade A e dez da variedade B, obtendo-se os seguintes resultados:

Var. A		Var. B	
4.3	2.2	4.2	6.2
3.8	1.8	4.8	6.8
5.2	4.5	4.7	
2.5	1.7	6.5	
3.5	3.6	6.3	
4.1	4.5	5.9	
5.1	5.0	7.2	
4.0		5.1	

Verifique se as variedades A e B diferem em produtividade. Os valores entre parênteses são os postos correspondentes.

4.4.3 – Comentário

Poder-eficiência em torno de 95%.

4.5 – TESTE DE KOLMOGOROV – SMIRNOV

Este teste se aplica na comparação de dois grupos independentes, quando os dados observados se posicionam de maneira dispersa, quando ordenados em ordem crescente, ou seja, não é possível notar uma distinção de um dos tratamentos em relação ao outro. Esta prova verifica a concordância entre duas distribuições acumuladas.

4.5.1 – Exigência do Teste

- ✓ Nível de mensuração em escala ordinal.

4.5.2 – O Método

Deve-se construir uma distribuição de probabilidade acumulada para cada uma das amostras, considerando sempre intervalos iguais.

Em seguida, para cada intervalo subtraí-se uma função da outra. Seja $S_{n_1}(x) = \frac{k}{n_1}$ onde k é o número de escores não superiores a x . Da mesma forma define-se $S_{n_2}(x)$ para o outro grupo. Assim, deve-se obter:

$$D = \text{máx}|S_{n_1}(x) - S_{n_2}(x)|$$

Sugere-se a utilização de tantos intervalos quanto possível. Para verificação da significância estatística, deve-se considerar:

□ Pequenas Amostras (n_1 e $n_2 \leq 40$)

- $n_1 = n_2$: utiliza-se a Tábua L – Siegel;
- $n_1 \neq n_2$: utiliza-se a Tabela 9 – Campos.

□ Grandes Amostras (n_1 ou $n_2 > 40$)

- bilateral: utiliza-se a Tábua M – Siegel;
- unilateral: calcula-se a estatística

$$\chi^2 = 4D^2 \frac{n_1 n_2}{n_1 + n_2}$$

que possui distribuição Qui-quadrado com 2 graus de liberdade.

Os procedimentos acima são recomendados sempre que recursos computacionais não estejam disponíveis para obtenção do p-valor.

Exemplo 1: No exemplo anterior que se tratava da comparação de duas soluções químicas com relação ao grau de P.H. obteve-se:

A		B	
7.49	7.37	7.28	7.48
7.35	7.51	7.35	7.31
7.54	7.50	7.52	7.22
7.48	7.52	7.50	7.41
7.48	7.46	7.38	7.45

Para verificar se há diferença significativa entre as soluções usando o teste KS, serão considerados 7 intervalos.

	7,21-7,25	7,26-7,30	7,31-7,35	7,36-7,40	7,41-7,45	7,46-7,5	7,51-7,55
$S_{10} A(x)$	0/10	0/10	1/10	2/10	2/10	7/10	10/10
$S_{10} B(x)$	1/10	2/10	4/10	5/10	7/10	9/10	10/10
$ S_{10} A - S_{10} B $	1/10	2/10	3/10	3/10	5/10	2/10	0

$D = \max |S_{10} A(x) - S_{10} B(x)| = 5/10$ e consequentemente $K_C = 5$ (numerador da maior diferença entre as duas distribuições acumuladas). Pela tabela, K_D crítico = 7. Desta forma, não podemos rejeitar H_0 .

Conclusão: Pelo teste de Mann-Whitney obteve-se p-valor de aproximadamente 0,05 e pelo teste de KS, p-valor = 0,168. Assim não existem evidências para afirmar que existem diferenças significativas entre as soluções químicas.

Exemplo 2: Um coordenador de curso deseja saber se os alunos que ingressam via vestibular têm desempenho diferente dos que entram por outros meios. Os resultados do índice de rendimento, após um ano de curso, estão apresentados abaixo:

Ingresso	Aprovados por média (8,5 -10)	Aprovados por média (7 -8,4)	Aprovados em final (6 -6,9)	Aprovados em final (5 - 5,9)	Reprovados (0 - 4,9)
Vestibular	10 (0.238)	28 (0.905)	2 (0.952)	1 (0.976)	1 (1,0)
Outros	6 (0.089)	55 (0.910)	6 (1,0)	0 (1,0)	0 (1,0)

$$D = 0,238 - 0,089 = 0,149$$

$$D_{\text{crítico}} = 1,36 \sqrt{\frac{n_1 + n_2}{n_1 * n_2}} = 0,268 \quad \text{considerando } \alpha=0,05$$

Conclusão: Não existem evidências suficientes para rejeitar H_0 . Não se pode afirmar que existe diferença significativa entre o desempenho dos dois grupos de estudantes.

Exercício: Aplique o teste de Kolmogorov – Smirnov aos dados referentes à comparação de variedades de tomates.

	1.6-2.1	2.2-2.7	2.8-3.3	3.4-3.9	4.0-4.5	4.6-5.1	5.2-5.7	5.8-6.3	6.4-6.9	7.0-7.5
Var(A)S ₁₅ A	1/15	3/15	3/15	7/15	12/15	14/15	15/15	1	1	1
Var(B)S ₁₀ A	0	0	0	0	1/10	4/10	4/10	7/10	9/10	1
S ₁₅ A-S ₁₀ A										

4.6 - TESTE DAS ITERAÇÕES DE WALD-WOLFOWITZ

Este teste se aplica a comparação de dois tratamentos (grupos), para verificar se existe diferença entre eles, com relação a diversos aspectos:

- Tendência central;
- Variabilidade;
- Assimetria;
- etc;

4.6.1 - Exigência do Teste

- ✓ Nível de mensuração em escala ordinal (pelo menos).

4.6.2 - Método

Primeiramente, deve-se ordenar os $n_1 + n_2$ escores em ordem crescente. Determina-se, em seguida, o número de iterações desta série ordenada.

A idéia do teste consiste em verificar se os valores desta série estarão bem misturados, o que indica então, que os dois grupos devem pertencer à mesma população. Se isso se verificar, r , o número de iterações, será bastante alto.

- **Pequenas amostras** (n_1 e $n_2 \leq 20$)

A tábua F_1 - Siegel fornece os valores críticos de r para um nível de significância de 5%. Se r é menor que o valor crítico fornecido pela tabela, deve-se rejeitar H_0 .

□ **Grandes amostras** (n_1 ou $n_2 > 20$)

Neste caso, utiliza-se a aproximação Normal, fazendo:

$$\mu_r = \frac{2n_1n_2}{n_1 + n_2} + 1$$

$$\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}$$

e obtendo-se $z = \frac{r - \mu_r}{\sigma_r}$. Determina-se, então, o p-valor associado a z, consultando-se a tabela.

Obs: no caso de empates entre escores de grupos diferentes, deve-se obter os diferentes números r de iterações para cada possível série, considerando todas as ordenações possíveis. Se alguns r's fornecerem resultados significativos e outros não, deve-se tirar a média de todos os p-valores obtidos.

□ **Poder Eficiência**

Cerca de 75% para amostras de tamanho 20.

4.7 - TESTE DE ALEATORIZAÇÃO PARA DUAS AMOSTRAS INDEPENDENTES

Esta prova é bastante útil na comprovação da hipótese de igualdade entre as características de duas amostras independentes pequenas. Consiste em obter a probabilidade exata associada aos valores observados, supondo que eles pertencem à mesma população.

4.7.1 - Exigências do Teste

- ✓ Nível de mensuração em escala intervalar.

4.7.2 – O Método

□ Pequenas amostras

Inicialmente, calcula-se o número de maneiras diferentes que os escores poderiam estar dispostos, considerando os tamanhos de amostra n_A e n_B , fazendo $\binom{n_A + n_B}{n_A}$ e assim, o número de resultados que irão compor a região de rejeição será $\binom{n_A + n_B}{n_A} \times \alpha$ (onde α é o nível de significância). Esta região será então, formada pelos resultados mais extremos.

Verifica-se então se o resultado observado se encontra entre esses resultados extremos.

□ Grandes amostras

Neste caso, aplica-se a aproximação pelo teste t fazendo-se

$$t = \frac{\bar{A} - \bar{B}}{\sqrt{\frac{\Sigma(B - \bar{B})^2 + \Sigma(A - \bar{A})^2}{n_A + n_B - 2} \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}}$$

Esta estatística tem aproximadamente a distribuição t de Student com g.l.= $n_A + n_B - 2$.

- Poder Eficiência = 100%

V – TESTES PARA COMPARAÇÃO DE K AMOSTRAS RELACIONADAS

Neste capítulo serão abordadas técnicas não-paramétricas de comparação de 3 ou mais grupos relacionados entre si. Imagine então a seguinte estrutura de dados:

	<i>TRATAMENTOS</i>				
	1	2	3	...	k
Unidade 1	X_{11}	X_{21}	X_{31}	...	X_{k1}
Unidade 2	X_{12}	X_{22}	X_{32}	...	X_{k2}
Unidade 3	X_{13}	X_{23}	X_{33}	...	X_{k3}
...
Unidade n	X_{1n}	X_{2n}	X_{3n}	...	X_{kn}

As unidades 1 a n representam as unidades amostrais utilizadas no experimento e os tratamentos são as k condições de avaliação (tempo, dietas, distância, etc...).

5.1 - TESTE Q DE COCHRAN

Este teste é uma extensão do teste de McNemar para duas amostras e se aplica na verificação de diferença significativa entre 3 ou mais grupos de frequências ou proporções.

5.1.1 - Exigência Do Teste

- ✓ Nível de mensuração em escala Nominal ou Ordinal dicotomizada

5.1.2 - Método

Se for considerada a disposição de dados ilustrada acima, com n linhas e k colunas, podemos testar a diferença entre os k tratamentos fazendo uso da expressão:

$$Q = \frac{(k - 1) \left[k \sum_{j=1}^k G_j^2 - \left(\sum_{j=1}^k G_j \right)^2 \right]}{k \sum_{i=1}^n L_i - \sum_{j=1}^n L_i^2}$$

onde G_j é o total da coluna j e L_i é o total da linha i. A estatística Q acima tem distribuição aproximadamente Qui-quadrado com g.l. = k-1.

Exemplo 1: Deseja-se analisar a influência da atitude de um entrevistador sobre as respostas a uma pesquisa sobre opinião pública, dadas por donas de casa. Os entrevistadores são treinados de acordo com os seguintes métodos:

- I – Demonstrando interesse, cordialidade, entusiasmo;*
- II – Demonstrando formalismo, reserva e cortesia;*
- III – Demonstrando modo abrupto, formalismo áspero.*

Cada entrevistador visita 18 casas (uma casa em cada bairro) e aplica um dos métodos de abordagem. Os resultados estão demonstrados a seguir. Obs: 0: não ; 1: sim

Bairro	Ent.1	Ent.2	Ent.3	L_i
1.	0	0	0	0
2.	1	1	0	2
3.	0	1	0	1
4.	0	0	0	0
5.	1	0	0	1
6.	1	1	0	2
7.	1	1	0	2
8.	0	1	0	1
9.	1	0	0	1
10.	0	0	0	0
11.	1	1	1	3
12.	1	1	1	3
13.	1	1	0	2
14.	1	1	0	2
15.	1	1	0	2
16.	1	1	1	3
17.	1	1	0	2
18.	1	1	0	2
G_j	13	13	3	29

H_0 : A probabilidade de um “sim” é a mesma nos três tipos de entrevistas.

H_a : A probabilidade de um “sim” difere entre pelo menos dois tipos de entrevistas.

$$Q = \frac{(3-1)[3(13^2 + 13^2 + 3^2) - (29)^2]}{3 \cdot 29 - 63} = \frac{400}{24} = 16.7$$

Com 2 graus de liberdade, obtém-se p-valor < 0,001.

Conclusão: Existe diferença significativa entre as formas de abordagem.

Exemplo 2: Um fabricante de sapatos para senhoras mostra 4 de seus últimos lançamentos (A, B, C e D) a 7 compradores de lojas de calçados. As encomendas estão resumidas abaixo. Estabelecer e testar a hipótese adequada.

Comprador	Modelo			
	A	B	C	D
1	X	X		
2		X	X	
3		X	X	X
4				X
5			X	
6			X	
7			X	

H_0 : A probabilidade de compra é a mesma nos quatro modelos.

H_a : A probabilidade de compra é diferente entre pelo menos 2 modelos.

$$Q = \frac{3[4(1^2 + 3^2 + 5^2 + 2^2) - (11)^2]}{4 \cdot 11 - 21} = \frac{3 \cdot 35}{23} = 4,565$$

Com 3 graus de liberdade, obtém-se p-valor igual a 0,20.

Conclusão: Não existem diferenças significativas entre os modelos de sapatos.

Exercício: Um professor deseja descobrir a melhor maneira de demonstrar uma fórmula matemática a seus alunos. Existem 3 maneiras possíveis de demonstrar a fórmula, mas talvez seus alunos não as considerem igualmente fáceis (ou difíceis) de entender. Ele explica cada uma das 3 demonstrações a seus 18 alunos e depois pergunta a cada estudante qual das 3 demonstrações ele havia entendido. Os resultados estão apresentados abaixo. Existe diferença entre os três métodos?

Estudante	A	B	C	L
1	X	X	X	
2		X		
3		X		
4		X		
5		X	X	
6				
7		X		
8	X	X		
9	X	X	X	
10				
11				
12		X		
13		X		
14		X	X	
15		X		
16	X	X		
17	X	X	X	
18		X		
G				

5.2 - TESTE DE FRIEDMAN

Este teste é útil quando se deseja comprovar a hipótese de que as k amostras relacionadas provêm da mesma população. Neste tipo de estudo observa-se o mesmo grupo de indivíduos sob cada uma das k condições, ou então se formam conjuntos de indivíduos homogêneos entre si, e estes são alocados aleatoriamente a cada uma das condições.

Podemos considerar este teste uma extensão da comparação de duas amostras pareadas.

5.2.1 – Exigência do Teste

- ✓ Nível de mensuração em escala ordinal (pelo menos)

5.2.2 – O Método

Considera-se aqui a mesma disposição dos dados observada anteriormente, com n linhas e k colunas. Para cada linha da tabela atribui-se postos de 1 a k .

Calcula-se em seguida a estatística χ_r^2 de Friedman dada por:

$$\chi_r^2 = \frac{12}{nk(k+1)} \sum_{i=1}^k R_i^2 - 3n(k+1)$$

onde R_i é a soma dos postos na coluna i .

□ Pequenas Amostras

- $k = 3$ e n variando de 2 a 9: tábua N - Siegel (pag. 311);
- $k = 4$ e n variando de 2 a 4: tábua N - Siegel (pág. 312).

□ Grandes Amostras

Valores críticos: Tabela Qui-quadrado com g.l. = $k - 1$.

Exemplo 1: Um teste de consumo de combustível envolvendo carros produzidos por três fabricantes foi realizado e os resultados, em quilômetros por litro de combustível estão apresentados na tabela abaixo. Estabelecer e testar a hipótese adequada.

Modelo	Fabricante		
	G	F	C
Pequeno	9,0	11,3	10,6
Médio- 6 cil.	9,4	10,9	10,2
Médio- 8 cil.	8,1	8,6	9,1
Grande-8 cil.	8,3	8,6	8,8
Esporte	8,2	9,2	9,5

$$\chi^2_r = \frac{12}{5.3(3+1)}(5^2 + 12^2 + 13^2) - 3.5.4 = 7,6$$

Pela tabela N (Siegel) p – valor = 0,024 . Existem diferenças significativas entre os fabricantes, com relação ao consumo de combustível.

Exemplo 2: Uma pesquisa foi delineada para avaliar a eficácia de um programa para a redução do estresse baseado em meditação, para pacientes com problemas de ansiedade. Foi medido o escore de Ansiedade da Escala Hamilton em quatro momentos diferentes: recrutamento (IR), pré-tratamento (Pré), pós-tratamento (Pós) e três meses após o pós-tratamento (3-M). O tratamento aparenta ter surtido efeito?

Sujeito	IR	Pré	Pós	3-M
1	21	21	16	19
2	30	38	10	21
3	38	19	15	6
4	43	33	30	24
5	35	34	25	10
6	40	40	31	30
7	27	15	11	6
8	18	11	4	7
9	31	42	23	27
10	21	23	21	17
11	18	24	16	13
12	28	8	5	2
13	40	37	31	19
14	35	32	12	21

$$\chi_r^2 = \frac{12}{14.4.5} (50,5^2 + 47^2 + 23,5^2 + 19^2) - 3.14.5 = 33,107$$

Com graus de liberdade igual a 3, o p-valor é inferior a 0,001.

Exercício: Num ensaio sobre adubação nitrogenada de alface, compararam-se 4 tratamentos:

Tratamento 1: controle

Tratamento 2: 5 g de salitre/10 litros de água

Tratamento 3: 10g de salitre/10 litros de água

Tratamento 4: 20g de salitre/10 litros de água

Os resultados de produção se encontram abaixo:

Área	Trat.1	Trat.2	Trat.3	Trat.4
I	3640	4200	4700	5300
II	4890	4550	6020	5900
III	4800	5320	5250	5150
IV	4460	5500	5580	5560

5.3 – COMPARAÇÕES MÚLTIPLAS

Com a finalidade de complementar a análise do teste de Friedman, quando a hipótese H_0 é rejeitada, deve-se proceder ao teste de comparações múltiplas para identificar as diferenças entre pares de tratamentos.

5.3.1 – O Método

□ Pequenas Amostras (k e n \leq 15)

Primeiramente, calcula-se para cada par de tratamentos a diferença $| R_i - R_j |$ onde R_i é a soma de postos atribuída ao tratamento i.

Em seguida obtém-se a diferença mínima significativa (d.m.s.) na tabela 24 (Campos).

As diferenças observadas entre as somas de postos, superiores a d.m.s. indicam diferenças significativas entre aqueles tratamentos.

□ Grandes Amostras

Para k e/ou n maiores que 15 calcula-se a d.m.s através da expressão

$$d.m.s. = Q \sqrt{\frac{nk(k+1)}{12}}$$

onde os valores de Q podem ser obtidos na tabela 17 (Campos).

Exercício: Aplicar o teste de comparações múltiplas aos exemplos e exercícios do teste de Friedman.

VI – TESTES PARA COMPARAÇÃO DE K – AMOSTRAS INDEPENDENTES

Neste capítulo serão abordados testes não-paramétricos de comparação de 3 ou mais grupos independentes entre si, úteis quando as suposições paramétricas não forem atendidas. Considere-se então, a seguinte estrutura de dados:

TRATAMENTOS				
1	2	3	...	K
X_{11}	X_{21}	X_{31}	...	X_{k1}
X_{12}	X_{22}	X_{32}	...	X_{k2}
X_{13}	X_{23}	X_{33}	...	X_{k3}
...
X_{1l}	X_{2m}	X_{3n}	...	X_{kp}

Onde as informações do tratamento i não estão relacionadas às informações do tratamento j, e os tamanhos das amostras em cada tratamento não necessariamente são iguais.

As provas que serão vistas aqui são competidores do teste F (Análise de Variância).

6.1 - TESTE DA MEDIANA

A função deste teste é verificar se k grupos independentes pertencem à mesma população ou populações com a mesma mediana.

6.1.1 - Exigências do Teste

- ✓ Nível de mensuração em escala ordinal (pelo menos)

6.1.2 – O Método

Inicialmente deve-se calcular a mediana geral, considerando todos os valores observados. Então, cada score é substituído por um sinal + (se for superior à mediana) ou - (se não for superior à mediana) e estas informações são representadas pela tabela k x 2 a seguir:

TRATAMENTOS					
	1	2	3	...	k
+	O_{1+}	O_{2+}	O_{3+}	...	O_{k+}
-	O_{1-}	O_{2-}	O_{3-}	...	O_{k-}

Calcula-se, em seguida a estatística χ^2 dada por

$$\chi^2 = \sum_{j=1}^k \frac{(O_j - E_j)^2}{E_j}$$

onde E_j são as frequências esperadas relacionadas a cada casela na tabela de dados.

Compara-se então, o valor da estatística com os valores críticos da distribuição Qui-quadrado considerando-se g.l. = k-1.

Exemplo1: Suponha que se deseja estudar a influência do nível de instrução das mães sobre o grau de interesse das mesmas pela educação de seus filhos. O pesquisador toma o número de visitas espontâneas que cada mãe faz à escola durante o ano letivo. Ele obtém também o grau educacional de 44 mães, que irão constituir a amostra.

Grau educacional atingido pela mãe					
8º grau	10º grau	12º grau	Universidade (parcial)	Universidade (diplomada)	Pós- graduação
4	2	2	9	2	2
3	4	0	4	4	6
0	1	4	2	5	
7	6	3	3	2	
1	3	8			
2	0	0			
0	2	5			
3	5	2			
5	1	1			
1	2	7			
	1	6			
		5			
		1			

H_0 : Não há diferença na mediana de visitas à escola nos diferentes níveis educacionais da mãe.

H_a : Existe diferença na mediana de visitas à escola nos diferentes níveis educacionais da mãe.

Mediana = 2,5 visitas.

	8º grau	10º grau	12º grau	Univ(parcial)	Univ (comp)	Pós
> 2,5	5	4	7	3	2	1
e_j	5	5,5	6,5	2	2	1
≤ 2,5	5	7	6	1	2	1
e_j	5	5,5	6,5	2	2	1

$$\chi^2 = \frac{(5-5)^2}{5} + \frac{(4-5,5)^2}{5,5} + K + \frac{(4-5)^2}{5} = 1,295 \quad p\text{-valor} = 0,73$$

Considerando $\alpha = 0,05$, $n=44$ e $g.l = k-1 = 4-1=3$ (agregando as 3 últimas classes), o valor tabelado de Qui-quadrado = 7,81.

Conclusão: Não temos evidências suficientes para rejeitar H_0 . Portanto não existe diferença no nível educacional das mães com relação ao número de visitas à escola.

Exemplo 2: 186 indivíduos com excesso de peso foram submetidos a quatro dietas diferentes. Abaixo estão apresentados os resultados, do peso perdido após um período de quatro semanas. Verificar se as dietas provêm de populações com a mesma mediana.

		<i>Dietas</i>			
		A	B	C	D
+		23 (18)	10 (20)	10 (25)	50 (30)
-		13 (18)	30 (20)	40 (25)	10 (30)
Total		36	40	50	60

$$\chi^2 = 1,39 + 5 + 9 + 13,33 + 1,39 + 5 + 9 + 13,33 = 57,44$$

$p\text{-valor} < 0,001$

Conclusão: Temos evidências suficientes para rejeitar H_0 , isto significa que as dietas provêm de populações com diferentes medianas.

Exercício: Quatro grupos de crianças de escolas públicas urbanas (A), particulares urbanas (B), públicas rurais (C) e particulares rurais (D) são submetidos a um teste físico. Os resultados estão apresentados abaixo. Verificar se há diferença no desempenho mediano.

A	B	C	D
16	4	12	15
27	9	16	26
10	13	20	28
17	19	24	29
12	22	18	30
23	18		33
21	13		21
12			
22			

<i>Escolas</i>				
	A	B	C	D
+				
-				
Total				

	A/B	C/D		A/C	B/D
+					
-					
Total					

6.2 - TESTE DE KRUSKAL-WALLIS

É um teste bastante eficaz na comparação de k tratamentos independentes. Ele nos indica se há diferença entre pelo menos dois deles. É na verdade uma extensão do teste de Wilcoxon para duas amostras independentes e se utiliza dos postos atribuídos aos valores observados.

6.2.1 – Exigências do Teste

- ✓ Nível de mensuração em escala ordinal (pelo menos)

6.2.2 - Método

Primeiramente deve-se atribuir a cada valor observado, um posto, sempre atribuindo o menor posto ao menor valor e o maior posto ao maior valor. Após se efetuar a soma dos postos para cada tratamento (R_j) calcula-se a estatística H, onde

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1)$$

- **Quando K = 3 e os n_j 's ≤ 6 :**

Compara-se este valor com o valor crítico da tabela 14 - Campos, onde $n_1 \leq n_2 \leq n_3$;

- **Quando K > 3 ou pelo menos um $n_j > 6$**

Utiliza-se a aproximação Qui-quadrado com g.l. = k - 1.

- **Empates**

Atribui-se aos empates a média dos postos.

Exemplo 1: Suponha-se que um pesquisador resolva investigar o grau de autoritarismo do pessoal de escolas, usando a Escala de Autoritarismo de Adorno. Os escores estão apresentados abaixo. Verificar se há diferença entre os grupos.

Professores de E.F.	Outros professores	Administradores
96	82	115
128	124	149
83	132	166
61	135	147
101	109	

	4	2	7
	9	8	13
POSTOS	3	10	14
	1	11	12
	5	6	
TOTAL	22	37	46

$$H = \frac{12}{14.15} (96,8 + 273,8 + 529) - 45 = 6,406$$

$$h = 5,643$$

$$0,025 < p < 0,05$$

Rejeita-se H_0 . É necessário realizar o teste de comparações múltiplas entre os grupos:

Exemplo 2: Considere um experimento de alimentação de suínos, em que se usaram 4 rações (A, B, C, D), cada uma fornecida a 5 animais. Os aumentos de peso observados, em Kg, após 1 anos foram os seguintes:

	A	B	C	D
	35 (14,5)	40 (17)	39 (16)	27 (6,5)
	19 (4)	35 (14,5)	27 (6,5)	12 (1)
	31 (12)	46 (20)	20 (5)	13 (2)
	15 (3)	41 (18)	29 (9)	28 (8)
	30 (10,5)	33 (13)	45 (19)	30 (10,5)
R	44	82,5	55,5	28
Posto médio	8,8	16,5	11,1	5,6

$H = 9,037$ ($0,02 < p < 0,05$). Existem diferenças significativas entre as rações.

Exercício: Numa pesquisa sobre qualidade de vinho, foram provados três tipos por cinco degustadores. Cada degustador provou 12 amostras (4 de cada tipo) e atribuiu a cada uma delas uma nota de zero a dez. As médias das notas atribuídas pelos 5 degustadores a cada uma das amostras foram:

Tipo 1	Tipo 2	Tipo 3
5,0	8,3	9,2
6,7	9,3	8,7
7,0	8,6	7,3
6,8	9,0	8,2

Verifique se há preferência dos degustadores por algum dos tipos de vinho.

6.3 - COMPARAÇÕES MÚLTIPLAS

Este teste é utilizado para complementar a análise do teste de Kruskal-Wallis e verificar entre que pares de tratamentos existe diferença significativa.

6.3.1 - Método

□ Pequenas amostras

⇒ para n's iguais

Primeiramente, calcula-se para cada par de tratamentos a diferença

$$| R_i - R_j |$$

onde R_i é a soma de postos atribuída ao tratamento i .

Em seguida obtém-se a diferença mínima significativa (d.m.s.) na tabela 16 - Campos. As diferenças observadas entre as somas de postos, superiores a d.m.s. indicam diferenças significativas entre aqueles tratamentos.

⇒ para n's diferentes

$$d.m.s. = \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right) h}$$

onde h é obtido pela tabela 14 - Campos e deve-se utilizar

$$|\overline{R}_i - \overline{R}_j|$$

onde $\overline{R}_i = \frac{R_i}{n_i}$

□ Grandes amostras

⇒ para n's iguais

Determinam-se as diferenças $|\overline{R}_i - \overline{R}_j|$ e então calcula-se $d.m.s. = Q \sqrt{\frac{k(N+1)}{12}}$
 a tabela 17 - Campos fornece os valores de Q .

⇒ para n's diferentes

$$d.m.s. = Z \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

onde $Z = Z(\alpha/[k(k-1)])$ obtido na tabela Normal Padrão Z.

Exemplos: Verificar diferenças entre tratamentos nos exercícios anteriores do teste de Kruskal-Wallis.

Exercício: Três marcas de lâmpadas elétricas são testadas para verificar se elas possuem a mesma vida média. Os dados estão apresentados abaixo. Estabelecer e testar a hipótese adequada.

Marcas		
A	B	C
73	84	82
64	80	79
67	81	71
62	77	75
70	82	72
65	79	73
68	78	75
71	83	76
	82	

VII – MEDIDAS DE ASSOCIAÇÃO

Frequentemente há interesse em verificar a existência de associação entre dois conjuntos de valores e também o grau desta associação. É necessário também testar a significância estatística deste coeficiente.

No caso paramétrico, a medida usual é o coeficiente de correlação r de Pearson que exige mensuração dos escores no mínimo ao nível intervalar. Ainda, se estivermos interessados em comprovar a significância de um valor observado de r de Pearson deveremos supor que os escores provenham de uma distribuição normal. Quando estas suposições não são atendidas podemos utilizar um dos coeficientes de correlação não-paramétricos e suas respectivas provas de significância. Sempre importante lembrar que este coeficiente avalia o grau de relação linear.

7.1 - COEFICIENTE DE CONTINGÊNCIA C

Este coeficiente mede a associação entre dois conjuntos de atributos quando um ou ambos os conjuntos são medidos em escala nominal.

Considere uma tabela de contingência $k \times r$, que representa as frequências cruzadas dos escores A (divididos em k categorias) e escores B (divididos em r categorias). O grau de associação entre dois conjuntos de atributos é calculado por:

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}} \text{ onde } \chi^2 \text{ é a estatística Qui-quadrado calculada conforme já visto anteriormente.}$$

O p-valor associado à estatística Qui-quadrado com $(r-1) \times (k-1)$ graus de liberdade é a prova de significância do coeficiente de contingência C.

O coeficiente C assumirá valor zero quando não houver associação, porém nunca será igual à

1. O limite superior do coeficiente é dado por $\sqrt{\frac{k-1}{k}}$ (quando $k = r$). Note que para calcular o coeficiente C, a tabela de contingência deve satisfazer as restrições do teste Qui-quadrado. (Ver exemplo Siegel, pág. 223)

7.2 - COEFICIENTE DE CORRELAÇÃO DE SPEARMAN

É uma medida de associação que exige que ambas as variáveis se apresentem em escala de mensuração pelo menos ordinal. Basicamente equivale ao coeficiente de correlação de Pearson aplicado à dados ordenados. Assim,

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} = r_s$$

ou seja, o coeficiente de correlação de Spearman se utiliza da expressão do coeficiente de Pearson porém calculado com postos, ao invés dos dados originais. Esta expressão equivale à

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n} \quad \text{onde } d_i = x_i - y_i \text{ a diferença de postos dos escores X e Y.}$$

Para verificar a significância do valor observado de r_s deve-se consultar a tábua P (Siegel) para n entre 4 e 30. Para n maior ou igual a 10 pode-se usar a expressão de t de Student

$$t = r_s \sqrt{\frac{n-2}{1-r_s^2}} \quad \text{onde t tem n-2 graus de liberdade. (Ver exemplo Siegel, pág. 231)}$$

ANEXOS

PROVA BINOMIAL

Exemplo: Testes de Diferenças Sensoriais são comumente utilizados em Controle de Qualidade e Avaliação da qualidade. O teste triangular (Bradley – 1963), teste de diferenças sensoriais que fornece uma aplicação útil do modelo Binomial, é dado a seguir. Para cada um de n painelistas três amostras-teste são apresentados em ordem aleatória. É sabido que duas das amostras são idênticas; a terceira é diferente. Ao painalista é então solicitado que selecione a amostra diferente, com base num atributo sensorial específico. Se os painelistas são juizes treinados, homogêneos, o experimento pode ser visto como n tentativas de Bernoulli repetidas e independentes, onde um sucesso corresponde a uma identificação correta da amostra. (Se os painelistas não são juizes treinados, homogêneos, pode-se questionar a validade da suposição de probabilidade constante ao longo das repetições). Sob a hipótese de que não há nenhuma base para discriminação, a probabilidade p de sucesso é $1/3$. A suspeita de que existe discriminação corresponde a valores de p maiores que $1/3$.

Byer e Abrams (1953) consideraram um teste triangular de amargor no qual cada degustador recebeu três copos, dois contêm a mesma solução de quinino e o terceiro uma solução de quinino diferente. No primeiro teste de amargura, as soluções contiveram 0,0075% e 0,005%, respectivamente, de sulfato de quinina. As seis ordens de apresentação são: LHH, HLH, HHL, HLL, LHL, LLH (L - denota a mais baixa concentração e H - a concentração mais alta), foram distribuídos aleatoriamente entre os degustadores. Depois de 50 tentativas, houve 25 seleções corretas e 25 seleções incorretas.

Considere B como a estatística do teste Binomial (B é o numero de identificações corretas).

Considere ainda $H_0: p = 1/3$ contra a hipótese alternativa $p > 1/3$, e usa-se a aproximação para grandes amostras. Fixamos $\alpha = 0,05$. Na Tabela da Normal padrão encontra-se $z = 1,645$ e assim a aproximação ao nível de 0,05 de significância estabelece a seguinte regra de decisão: Rejeita-se H_0 se $B \geq 1,645$; caso contrário não se rejeita.

Assim, $n = 50$ e B (o número correto de identificações) = 25. Assim considerando $p_0 = 1/3$:

$$B^* = \frac{25 - 50(1/3)}{[50(1/3)(2/3)]^{1/2}} = 2,5$$

Sendo $B = 2,5 > 1,645$, rejeita-se $H_0: p = 1/3$ a favor de $p > 1/3$ ao nível de 0,05 de significância. Assim, existem evidências de discriminação do sabor amargo (Na Tabela da Normal padronizada nota-se também que o menor nível de significância ao qual rejeitamos H_0 em favor $p > 1/3$ seria 0,0062 – p-valor.)

TRABALHO APLICADO: DESCOBRINDO ERROS

Um médico me ligou outro dia com algumas dúvidas a respeito de uma análise estatística, realizada por um amigo, na sua dissertação de mestrado. Este médico estava comparando os resultados de dois exames realizados em um grupo de 40 pacientes.

Ele me passou poucas informações por telefone. Disse apenas que no primeiro exame 10 pacientes apresentaram resultado positivo enquanto que no segundo exame foram apenas 5 pacientes. Disse ainda que no relatório constava:

Os resultados apresentaram nível de concordância de 82,5%

Qui-quadrado = 35,7 $p > 0,0001$

Tente reproduzir a análise estatística apontando os erros cometidos. Explique os resultados ao médico, com linguagem própria, pois ele está bastante confuso às vésperas da defesa de sua dissertação.

**Intervalo de Confiança
para a Probabilidade de Sucesso
(Clopper – Pearson)**

Para grandes amostras, utiliza-se a aproximação normal para construção do intervalo de confiança da proporção de sucesso:

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

No entanto, utilizando este procedimento obtêm-se um intervalo de confiança simétrico, o que não corresponde à realidade em geral. É proposto então um procedimento mais adequado:

Intervalo de Confiança de Clopper – Pearson

$$L.I. = \frac{B}{B + (n - B + 1) - f_{\alpha/2, 2(n-B+1), 2B}}$$

$$L.S. = 1 - \frac{n - B}{n - B + (B + 1)f_{\alpha/2, 2(B+1), 2(n-B)}} = \frac{(B + 1)f_{\alpha/2, 2(B+1), 2(n-B)}}{(n - B) + (B + 1)f_{\alpha/2, 2(B+1), 2(n-B)}}$$

em que $\begin{cases} B \Rightarrow \text{número de sucessos em } n \text{ repetições} \\ n \Rightarrow \text{número de repetições} \\ f \Rightarrow \text{valor da distribuição } F \text{ com g.l.} = n_1 \text{ e } n_2 \end{cases}$

ANÁLISE DE DADOS REAIS

Neste estudo foram acompanhados 21 pacientes obesos, que passaram por tratamentos para perda de PESO. O primeiro tratamento, seguido por dez pacientes, era composto simplesmente de uma dieta (grupo PLACEBO) e o segundo tratamento, seguido por outros 11 pacientes, era composto de uma dieta e mais a droga XENICAL (grupo XENICAL). No início e no final dos tratamentos (6 meses após), foram mensuradas nos pacientes as variáveis Peso, Índice de Massa Corpórea (IMC) e a taxa de Leptina. Deseja-se verificar, neste estudo, se há diferença significativa entre os tratamentos para as variáveis mencionadas acima.

METODOLOGIA ESTATÍSTICA

Inicialmente, para verificação da homogeneidade dos grupos analisados foram construídas tabelas de contingência e calculadas algumas medidas descritivas. Dependendo da variável em análise, foi aplicado o teste estatístico mais adequado (Teste exato de Fisher, Teste “t” de Student, Teste de Mann-Whitney).

Observada a homogeneidade dos grupos, verificou-se, em cada grupo, se no final do tratamento houve variação significativa para as variáveis PESO, IMC e LEPTINA aplicando-se o Teste estatístico não-paramétrico de Wilcoxon, dado que a suposição de Normalidade das observações não foi comprovada.

Finalmente, os dois tratamentos foram comparados, verificando-se as diferenças significativas entre as variações de PESO, IMC e LEPTINA através do Teste estatístico não-paramétrico de Mann-Whitney.

PLACEBO		IDADE	PESO		IMC		LEPTINA	
PACIENTE	SEXO		INICIAL	FINAL	INICIAL	FINAL	INICIAL	FINAL
733	2	35	92,6	90,5	39,07	38,16	39	38,83
736	1	69	86,2	79,5	28,16	25,98	2,8	1,95
738	2	36	74,9	74,4	35,66	35,38	22,06	15,41
741	2	50	82,5	81,2	34,8	34,26	36,52	37,41
747	2	47	91,5	77,7	36,16	30,73	16,26	7,05
748	1	46	84,8	81,5	31,64	30,3	5,7	4,07
750	2	57	78,2	80,9	32,55	33,7	17,1	18,94
752	1	52	87,4	84,7	31,43	30,37	8,34	13,6
755	2	61	69,7	68,7	33,15	32,67	29,86	22,14
758	2	29	79,5	73,7	30,69	28,45	7,67	4,92

XENICAL		IDADE	PESO		IMC		LEPTINA	
PACIENTE	SEXO		INICIAL	FINAL	INICIAL	FINAL	INICIAL	FINAL
734	2	61	104	97	41,14	38,37	14,72	9,25
735	2	45	87	82,3	33,9	32,14	13,69	21,1
737	1	43	117,1	117	42,58	42,54	28,85	37,23
740	2	56	68,8	65,5	29,4	27,99	13,37	12,27
742	2	54	74	71,3	31,62	30,47	28,04	22,36
745	2	44	87,6	86,1	35,6	35	11,04	26,39
751	2	49	71,3	70,1	31,68	31,15	18,73	17,62
753	1	49	104,5	99,4	35,78	33,99	11,33	8,66
756	2	64	86	82,6	35,39	33,99	36,98	27,95
757	2	59	94,3	90,9	44,27	42,67	65,41	78,39
759	2	45	86,8	85,5	32,38	31,79	18,91	22,48

Estadística Não paramétrica nos softwares (colaboração do aluno Ângelo Cabral)

Teste	Exercício	Resultados				
		Apostila	R	SPSS	Statistica	BioEstat
Binomial	Exemplo 1 página 12	p = 0,5	p = 0,5	p = 1 (bi)	-	p = 0,3818
Binomial	Exemplo 2 página 13	p = 0,551	p = 0,5501	p = 0,5501	-	p = 0,4569
Binomial	Exemplo 3 página 13	p = 0,0381	p = 0,0381	p = 0,0381	-	p = 0,0381
Binomial	Exercício página 14	p = 0,0848 (ex)	p = 0,0847	p = 0,078	-	p = 0,0607
		p = 0,0607 (ap)		Z = 1,42		Z = 1,5492
Qui-Quadrado	Exemplo 1 página 16	p = 0,0222	p = 0,0222	p = 0,022	p = 0,0222	p = 0,0222
		X ² = 16,3	X ² = 16,3333	X ² = 16,333	X ² = 16,3333	X ² = 16,333
Qui-Quadrado	Exemplo 2 página 16	p = 0,9544	p = 0,9547	p = 0,955	p = 0,9544	p = 0,9544
		X ² = 0,675	X ² = 0,6727	X ² = 0,673	X ² = 0,6746	X ² = 0,675
Qui-Quadrado	Exercício página 18	p = 0,0354	p = 0,0354	p = 0,035	p = 0,0353	p = 0,0353
		X ² = 11,96	X ² = 11,96	X ² = 11,96	X ² = 11,96	X ² = 11,96
KS	Exemplo 1 página 20	p < 0,01	p = 0,0004	p = 0	-	-
		D = 0,5	D = 0,65	D = 0,65		
KS	Exemplo 2 página 20	p > 0,20	p = 0	p = 0,998	-	-
		D = 0,0512	D = 0,4126	D = 0,051		
KS	Exercício página 21	p < 0,01	p = 0	p = 0	-	p < 0,01
		D = 0,333	D = 0,3325	D = 0,333		D = 0,3325
Lilliefors	Exercício página 22	p < 0,01	-	p = 0,005	-	p < 0,01
		D = 0,116		D = 0,161		D = 0,1614

Iterações	Exemplo 1 página 23	r = 10	-	p = 0,297 (bi)	-	r = 10
				r = 10		
Iterações	Exemplo 2 página 24	p = 0,0028	-	-	-	-
		Z = 2,98				
Iterações	Exercício página 24	r = 8	-	p = 0,796 (bi)	-	r = 8
				r = 8		
McNemar	Exemplo 1 página 26	p = 0,0106	p = 0,0124	p = 0,0213 (bi)	p = 0,0245	p = 0,0106
			X ² = 6,25		X ² = 5,06 (c)	
			X ² = 5,0625 (c)			
McNemar	Exemplo 2 página 27	p < 0,001	p = 0,0003	p = 0,001 (c)	p = 0,0005	p = 0,0005
		X ² = 12,93	X ² = 12,94	X ² = 12,062 (c)	X ² = 12,06 (c)	X ² = 12,0615 (c)
			X ² = 12,0615 (c)			
McNemar	Exercício página 27	p = 0,0047	p = 0,0047	p = 0,007 (c)	p = 0,0067 (c)	p = 0,0067 (c)
		X ² = 8	X ² = 8	X ² = 7,347 (c)	X ² = 7,35 (c)	X ² = 7,3472 (c)
			X ² = 7,3472 (c)			
Sinais	Exemplo 1 página 29	p = 0,0107	-	p = 0,0215 (bi)	p = 0,0269	p = 0,0107
					Z = 2,2136	
Sinais	Exemplo 2 página 30	p (bi) = 0,0003	-	p (bi) = 0,001	p (bi) = 0,0005	p (bi) = 0,0005
		Z = -3,58		Z = -3,471	Z = 3,4709	Z = 3,4709
Sinais	Exercício página 31	p = 0,2266	-	p = 0,4531 (bi)	p = 0,4497	p = 0,2266
					Z = 0,756	
Wilcoxon	Exemplo 1 página 38	p > 0,05 (bi)	p = 0,5937 (bi)	p = 0,594 (bi)	p = 0,5937 (bi)	p = 0,5937 (bi)
		T = 27	V = 27	T = 27	T = 27	T = 27
Wilcoxon	Exemplo 2 página 38	p = 0,032	p = 0,02967	p = 0,059 (bi)	p = 0,0593 (bi)	p = 0,0297
		T = 9	V = 9	T = 9	T = 9	T = 9
Wilcoxon	Exemplo 3 página 39	p = 0,0104	p = 0,0092	p = 0,021 (bi)	p = 0,0207 (bi)	p = 0,0092
		T = 69	V = 67,5	T = 69	T = 69	T = 68
Wilcoxon	Exercício página 40	p = 0,03	p = 0,0342	p = 0,068 (bi)	p = 0,0687 (bi)	p = 0,0344
		T = 5	V = 5	T = 5	T = 5	T = 5

Fisher	Exemplo 1 página 45	p (uni) = 0,0447	p (uni) = 0,0447	p (uni) = 0,045	p (uni) = 0,0447	p (uni) = 0,0447
		p (bi) = 0,0894	p (bi) = 0,0894	p (bi) = 0,089	p (bi) = 0,0894	p (bi) = 0,0894
Fisher	Exemplo 2 página 46	p (uni) = 0,1	p (uni) = 0,1002	p (uni) = 0,1	p (uni) = 0,1002	p (uni) = 0,1002
		p (bi) = 0,2	p (bi) = 0,1189	p (bi) = 0,119	p (bi) = 0,1189	p (bi) = 0,1189
Fisher	Exercício página 46	p (uni) = 0,085	p (uni) = 0,0849	p (uni) = 0,085	p (uni) = 0,0849	p (uni) = 0,0849
		p (bi) = 0,17	p (bi) = 0,1698	p (bi) = 0,17	p (bi) = 0,1698	p (bi) = 0,1698
Qui-Quadrado	Exemplo 1 página 48	p < 0,01	p = 0,003561	p = 0,008 (c)	p = 0,0036	p = 0,0078 (c)
		X ² = 8,49	X ² = 8,4952	X ² = 8,495	X ² = 8,5	X ² = 8,495
Qui-Quadrado	Exemplo 2 página 49	p < 0,01	p = 0,004719	p = 0,005	p = 0,00472	p = 0,0047
		X ² = 10,67	X ² = 10,7122	X ² = 10,712	X ² = 10,7122	X ² = 10,712
Qui-Quadrado	Exercício página 49	p > 0,5	p = 0,8088	p = 0,809	p = 0,8088	p = 0,8088
		X ² = 0,969	X ² = 0,9687	X ² = 0,969	X ² = 0,9687	X ² = 0,969
Mediana 2 amostras	Exemplo 1 página 52	p = 0,004	-	p (bi) = 0,0045	-	p = 0,0027
		Md = 11		Md = 11		X ² = 8,9833 Md = 11
Mediana 2 amostras	Exercício página 53	p = 0,0839	-	p (bi) = 0,1189	-	p (uni) = 0,0775
		Md = 7		Md = 7		p (bi) = 0,1026 Md = 7
Mann-Witney	Exemplo 1 página 55	p = 0,452 (ex)	p = 0,4032	p = 0,9048 (bi)(ex)	p = 0,8065 (bi)	p = 0,4032
		U = 9	W = 9	U = 9	U = 9	U = 9
Mann-Witney	Exemplo 2 página 55	p aprox 0,025	p = 0,01857	p = 0,0355 (bi)(ex)	p = 0,0376 (bi)	p = 0,0188
		U = 22,5	W = 22,5	U = 22,5	U = 22,5	U = 22,5
Mann-Witney	Exercício página 56	p < 0,002	p = 0,0003216	p = 0,0002 (bi)(ex)	p = 0,0006 (bi)	p = 0,0003
		U = 13,5	W = 13,5	U = 13,5	U = 13,5	U = 13,5
KS	Exemplo 1 página 58	p = 0,168	p = 0,1641	p = 0,164	p > 0,1	p > 0,05
		Kc = 5	D = 0,5	D = 0,5	D = 0,5	Kc = 5
		Kd = 7				Kd = 7
KS	Exemplo 2 página 58	-	-	-	-	p > 0,05
		D = 0,149 Dc = 0,268				D = 0,1485 Dc = 0,2677

KS	Exercício página 59	p aprox 0,005	-	-	-	p < 0,01
		D = 0,7				D = 0,7
						Dc = 0,6654
Cochran	Exemplo 1 página 63	p < 0,001	-	p = 0,0002	p = 0,0002	p = 0,0002
		Q = 16,7		Q = 16,667	Q = 16,6667	Q = 16,6667
Cochran	Exemplo 2 página 64	p = 0,20	-	p = 0,207	p = 0,2065	p = 0,2065
		Q = 4,565		Q = 4,565	Q = 4,5652	Q = 4,5652
Cochran	Exercício página 65	p = 0,0002	-	p = 0,0002	p = 0,0002	p = 0,0002
		Q = 16,6667		Q = 16,667	Q = 16,6667	Q = 16,6667
Friedman	Exemplo 1 página 67	p = 0,024	p = 0,02237	p = 0,022	p = 0,02237	p = 0,0224
		X ² = 7,6	X ² = 7,6	X ² = 7,6	X ² = 7,6	X ² = 7,6
Friedman	Exemplo 2 página 67	p < 0,001	p = 0	p = 0	p = 0	p < 0,0001
		X ² = 33,107	X ² = 33,8321	X ² = 33,832	X ² = 33,8321	X ² = 33,1071
Friedman	Exercício página 68	p = 0,068 (tab)	p = 0,07515	p = 0,075	p = 0,07516	p = 0,0752
		X ² = 6,9	X ² = 6,9	X ² = 6,9	X ² = 6,9	X ² = 6,9
Mediana k amostras	Exemplo 1 página 71	p = 0,74	-	p = 0,73	p = 0,7303	-
		X ² = 1,295		X ² = 1,295	X ² = 1,2951	
Mediana k amostras	Exemplo 2 página 72	p < 0,001	-	p = 0	p = 0	-
		X ² = 57,44		X ² = 57,444	X ² = 57,4444	
Mediana k amostras	Exercício página 73	p = 0,1266 (AC)	-	p = 0,127 (AC)	p = 0,1266 (AC)	-
		X ² = 2,333 (AC)		X ² = 2,333 (AC)	X ² = 2,3333 (AC)	
				p = 1 (AB)	p = 1 (AB)	
				X ² = 0 (AB)	X ² = 0 (AB)	
Kruskal-Wallis	Exemplo 1 página 75	0,25 < p < 0,05	p = 0,04065	p = 0,041	p = 0,0406	p = 0,0406
		H = 6,406	H = 6,4057	H = 6,406	H = 6,4057	H = 6,4057
Kruskal-Wallis	Exemplo 2 página 76	0,2 < p < 0,05	p = 0,02854	p = 0,029	p = 0,0285	p = 0,0285
		H = 9,037	H = 9,0576 (c)	H = 9,058 (c)	H = 9,0576 (c)	H = 9,0576 (c)
Kruskal-Wallis	Exercício página 77	p (tab) = 0,007	p = 0,02095	p = 0,021	p = 0,021	p = 0,021
		H = 7,731	H = 7,7308	H = 7,731	H = 7,7308	H = 7,7308

Softwares Utilizados

R - versão 3.0.1 com pacotes básicos (gratuito)

SPSS - versão 20

Statistica - versão 7

BioEstat - versão 5.3 (gratuito)

Legenda

(c): valores com correção

(tab): p-valor tabelado

(uni): p-valor unilateral

(bi): p-valor bilateral

(ex): p-valor exato

(ap): p-valor aproximado

cor amarela para diferença de metodologia identificada

cor vermelha para erros ou diferença de metodologia não identificada