

OBTENÇÃO DOS NÍVEIS DE SIGNIFICÂNCIA PARA OS
TESTES DE KRUSKAL-WALLIS, FRIEDMAN E COMPARAÇÕES
MÚLTIPLAS NÃO-PARAMÉTRICAS

ANTONIO CARLOS FONSECA PONTES

Dissertação apresentada à Escola Superior de
Agricultura "Luiz de Queiroz", Universidade
de São Paulo, para obtenção do título de
Mestre em Agronomia, Área de Concentração:
Estatística e Experimentação Agronômica.

P I R A C I C A B A
Estado de São Paulo - Brasil
Maio - 2000

**OBTENÇÃO DOS NÍVEIS DE SIGNIFICÂNCIA PARA OS
TESTES DE KRUSKAL-WALLIS, FRIEDMAN E COMPARAÇÕES
MÚLTIPLAS NÃO-PARAMÉTRICAS**

ANTONIO CARLOS FONSECA PONTES

Bacharel em Estatística

Orientador: Prof. Dr. **JOSÉ EDUARDO CORRENTE**

Dissertação apresentada à Escola Superior de
Agricultura "Luiz de Queiroz", Universidade
de São Paulo, para obtenção do título de
Mestre em Agronomia, Área de Concentração:
Estatística e Experimentação Agronômica.

P I R A C I C A B A
Estado de São Paulo - Brasil
Maio - 2000

Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - Campus "Luiz de Queiroz"/USP

Pontes, Antonio Carlos Fonseca

Obtenção dos níveis de significância para os testes de Kruskal-Wallis,
Friedman e comparações múltiplas não-paramétricas / Antonio Carlos Fonseca
Pontes. - - Piracicaba, 2000.

140 p.

Dissertação (mestrado) - - Escola Superior de Agricultura Luiz de Queiroz,
2000.

Bibliografia.

1. Análise de variância 2. Estatística não-paramétrica 3. Inferência estatística
4. Linguagem C 5. Linguagem de programação 6. Método estatístico I. Título

CDD 519.54

“Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor”

Agradecimentos

Aos professores do Departamento de Matemática e Estatística da Universidade Federal do Acre, pela confiança.

Aos professores e funcionários do Departamento de Ciências Exatas (ex- Matemática e Estatística) da ESALQ/USP pela colaboração nos momentos importantes desta caminhada.

Ao Prof. Dr. José Eduardo Corrente, meu orientador, cuja paciência, cooperação e apoio foram fundamentais.

Ao meu pai (*in memoriam*) e à minha mãe, eu ofereço.

Aos meus filhos, Junior, Eduardo e Ana, eu dedico.

À minha esposa, Loide, por tudo e um pouco mais.

SUMÁRIO

LISTA DE TABELAS	vii
RESUMO	xi
SUMMARY	xiii
1 INTRODUÇÃO	1
2 REVISÃO DA LITERATURA	10
2.1 Introdução	10
2.2 Modelo de Aleatorização	12
2.3 Estatística Linear de Postos	14
2.4 Teste de posição para comparação de dois tratamentos	15
2.4.1 Teste da Soma das Ordens de Wilcoxon	16
2.4.2 Teste de Mann-Whitney	17
2.4.3 Testes de Dispersão aplicáveis a duas amostras	18
2.5 Análise de Variância Não-Paramétrica: Classificação Simples (k amostras independentes)	18
2.5.1 Teste de Kruskal-Wallis	19
2.5.2 Teste de Jonckheere-Terpstra	22
2.5.3 Teste de Bhapkar	22
2.5.4 Teste de Fligner-Wolfe para tratamentos versus controle	23
2.5.5 Teste de Shirley (doses crescentes de substâncias <i>versus</i> controle)	24
2.5.6 Teste de Mack & Wolf (<i>umbrella alternatives</i>)	24
2.6 Análise de Variância Não-Paramétrica: Classificação Dupla (k amostras dependentes)	27
2.6.1 Teste de Friedman	28

2.6.2	Teste de Quade	29
2.6.3	Teste de Page	30
2.6.4	Outros testes não-paramétricos aplicáveis aos Delineamentos em Blocos Casualizados	31
2.7	Comparações Múltiplas	33
2.7.1	Introdução	33
2.7.2	Construção teórica do Método	38
2.7.3	Alguns Testes de Comparações Múltiplas Não-Paramétricos aplicáveis aos Delineamentos Inteiramente Casualizados	40
2.7.4	Alguns Testes de Comparações Múltiplas não-paramétricos aplicáveis aos Delineamentos em Blocos Casualizados	47
2.8	Testes não paramétricos para outros delineamentos experimentais	52
2.9	Outras Discussões em Estatística Experimental Não Paramétrica	53
3	METODOLOGIA	60
3.1	Testes de Permutações Sistemáticas e Testes de Permutações Aleatórias	60
3.1.1	Delineamentos Inteiramente Casualizados - Atribuição de postos conjunta	63
3.1.2	Delineamentos em Blocos Casualizados - atribuição de postos aos tratamentos dentro dos blocos	66
3.2	Teste de Kruskal-Wallis	68
3.2.1	Aspectos Gerais	68
3.2.2	Construção de Tabelas para o Teste de Kruskal-Wallis	72
3.3	Comparações Múltiplas para o delineamento inteiramente casualizado	73
3.3.1	Comparações múltiplas envolvendo todos os pares de tratamentos	74
3.3.2	Comparações Múltiplas envolvendo Tratamento <i>versus</i> Controle	76
3.3.3	Construção de Tabelas de Comparações Múltiplas	78
3.4	Teste de Friedman	79

3.4.1	Aspectos Gerais	79
3.4.2	Construção de Tabelas para o Teste de Friedman	81
3.5	Comparações Múltiplas para o delineamento em blocos casualizados	82
3.5.1	Comparações múltiplas envolvendo todos os pares de tratamentos	83
3.5.2	Comparações múltiplas entre os tratamentos e o controle	84
4	RESULTADOS E DISCUSSÃO	86
4.1	Construção dos Programas	86
4.1.1	Delineamentos Inteiramente Casualizados	86
4.1.2	Delineamentos em Blocos Casualizados	89
4.2	Construção de tabelas: delineamentos inteiramente casualizados	90
4.2.1	Tabelas para o Teste de Kruskal-Wallis	90
4.2.2	Construção de Tabelas para as Comparações Múltiplas em delineamentos Inteiramente Casualizados	96
4.3	Construção de tabelas: delineamentos em blocos casualizados	105
4.3.1	Construção de tabelas para o Teste de Friedman	106
4.3.2	Construção de tabelas para as comparações múltiplas em delineamentos em blocos casualizados	111
4.4	Exemplos utilizando o programa	115
4.4.1	Caso 1: delineamento inteiramente casualizado com número de repetições iguais e poucos tratamentos	115
4.4.2	Caso 2: delineamento inteiramente casualizado com número de tratamentos e de repetições grande	120
4.4.3	Caso 3: delineamentos inteiramente casualizados com número de repetições iguais - casos extremos	123
4.4.4	Caso 4: Delineamentos em Blocos Casualizados com pequeno número de repetições	124

4.4.5	Caso 5: Delineamentos em Blocos Casualizados utilizando as permutações aleatórias	127
4.4.6	Caso 6: Delineamentos em Blocos Casualizados - um caso com empates	129
5	CONCLUSÕES	131
	REFERÊNCIAS BIBLIOGRÁFICAS	136

LISTA DE TABELAS

1	Configurações para $k = 3$ tratamentos e $n_1 = n_2 = n_3 = 2$, sem empates.	91
2	Distribuição nula da estatística de Kruskal-Wallis (H) para $k = 3$ tratamentos e $n_1 = n_2 = n_3 = 2$	92
3	Configurações para $k = 3$ tratamentos e $n_1 = 1, n_2 = 2$ e $n_3 = 3$	93
4	Distribuição nula da estatística de Kruskal-Wallis (H) para $k = 3$ tratamentos e $n_1 = 1, n_2 = 2$ e $n_3 = 3$	94
5	Exemplo de uma configuração de postos para o caso com $k = 3$ tratamentos e $n_1 = 2, n_2 = 1$ e $n_3 = 2$, com empates ocorrendo entre os postos 2 e 3	94
6	Configurações para $k = 3$ tratamentos e $n_1 = 2, n_2 = 1$ e $n_3 = 2$, considerando a presença de empates.	95
7	Distribuição nula da estatística de Kruskal-Wallis (H) para $k = 3$ tratamentos e $n_1 = 2, n_2 = 1$ e $n_3 = 2$, com empate ocorrendo entre os postos 2 e 3.	96
8	Configurações para $k = 4$ tratamentos e 3 repetições por tratamento, com as diferenças entre totais de tratamentos iguais a 27 e 26.	97
9	Configurações para $k = 5$ tratamentos e duas repetições na qual as diferenças absolutas entre as somas dos postos são iguais a 15 ou 16 (máxima).	98
10	Valores das somas de postos de tratamentos ($R_i, i = 1, 2, 3$), diferenças absolutas entre estas somas e máximo destas diferenças ($\max_{i < j} R_i - R_j $).	100
11	Valores possíveis das diferenças entre as somas de postos de tratamentos (d) e sua distribuição de frequências.	100
12	Valores das somas dos postos ($R_i, i = 1, 2, 3$), das diferenças d_{i1} ($i = 2, 3$) e das diferenças máximas d_M	101

13	Valores possíveis das diferenças entre as somas de postos de tratamentos (d) e sua distribuição de frequências.	102
14	Médias dos postos dos tratamentos (\bar{R}_i), diferenças entre estas médias e valores das diferenças padronizadas (r_i^*).	103
15	Valores possíveis das diferenças padronizadas entre as somas de postos de tratamentos (r^*) e sua distribuição de frequências.	104
16	Esquema de distribuição de postos para um delineamento em blocos casualizados	105
17	Configurações para $k = 3$ tratamentos e $b = 2$ blocos	106
18	Configurações necessárias para o cálculo da estatística S no caso de $k = 3$ tratamentos e $b = 2$ blocos e somas dos postos ($R_i, i = 1, 2, 3$).	107
19	Distribuição nula exata da estatística S ($k = 3$ tratamentos e $b = 2$ blocos)	107
20	Configurações para $k = 3$ tratamentos e $b = 3$ blocos	108
21	Distribuição nula exata da estatística S ($k = 3$ e $b = 3$)	108
22	Alguns valores de S e as respectivas probabilidades de significância para $k = 4$ tratamentos e $b = 3$ blocos	109
23	Configurações para $k = 4$ tratamentos e $b = 3$ blocos	110
24	Somas dos postos ($R_i, i = 1, 2, 3$), diferenças entre as somas (d_{ij}) e diferença máxima (d_M) para $k = 3$ tratamentos e $b = 3$ blocos	111
25	Valores de diferenças entre somas de postos (d) e sua distribuição nula exata para as diferenças máximas e todas as diferenças ($k = 3$ tratamentos e $b = 3$ blocos)	112
26	Somas dos postos ($R_i, i = 1, 2, 3$), diferenças entre as somas de tratamentos e controle (d_{i1}) e diferença máxima (d_M) ($k = 3$ tratamentos e $b = 3$ blocos)	113

27	Diferenças entre somas de postos (d) e sua distribuição nula exata para as diferenças máximas e todas as diferenças ($k = 3$ e $b = 3$ - tratamentos <i>vs</i> controle)	113
28	Algumas configurações para $k = 3$ tratamentos e $b = 3$ blocos, diferenças (d_{i1}) entre somas de postos dos tratamentos (2 e 3) e o controle (1) e a diferença máxima para cada configuração considerada	114
29	Diferenças entre somas de postos (d) e sua distribuição nula exata para as diferenças máximas e todas as diferenças ($k = 3$ e $b = 3$ - tratamentos <i>vs</i> controle)	114
30	Dados relativos a quatro tratamentos e quatro repetições (McDonald & Thompson, 1967)	116
31	Diferenças entre os totais dos postos de tratamentos (Dif.), níveis de significância para as comparações entre todos os tratamentos (D-W e Geral) e resultados obtidos por McDonald & Thompson (1967).	117
32	Níveis de significância exatos (DW) para as comparações entre todos os tratamentos e obtidos através das execuções aleatórias 1 e 2 (DW Aleat1 e DW Aleat2), com seus respectivos erros percentuais.	119
33	Níveis de significância exatos (Geral) para as comparações entre todos os tratamentos e obtidos através das execuções aleatórias 1 e 2 (Geral Aleat1 e Geral Aleat2), com seus respectivos erros percentuais.	119
34	Níveis de significância aproximados para as diferenças entre todos os tratamentos obtidos pelos métodos dados em Campos (1983) e em Conover (1999).	121
35	Níveis de significância aproximados DW e Geral para as diferenças entre todos os tratamentos obtidos em duas execuções do programa desenvolvido (Aleat1 e Aleat2).	121

36	Algumas configurações extremas considerando $k = 4$ tratamentos e $n_i = 3$, ($i = 1, 2, 3, 4$) e as respectivas somas de postos de tratamentos (R_i).	123
37	Níveis de significância exatos para as diferenças entre todos os tratamentos obtidos através do programa para as configurações 1 e 2 da Tabela 36.	124
38	Níveis de significância exatos para as diferenças entre todos os tratamentos obtidos através do programa para as configurações 3 e 4 da Tabela 36.	125
39	Níveis de significância exatos para as diferenças entre os tratamentos e a testemunha obtidos através do programa para a configuração 3 da Tabela 36.	125
40	Postos e somas dos postos (R_i) referentes a um ensaio com $k = 4$ tratamentos e $b = 4$ blocos.	126
41	Níveis de significância exatos para as diferenças entre os tratamentos e a testemunha obtidos através do programa.	127
42	Níveis de significância exatos para as diferenças entre todos tratamentos obtidos através do programa.	127
43	Postos atribuídos a $k = 5$ tratamentos e $b = 6$ blocos.	128
44	Níveis de significância aproximados para as diferenças entre todos os tratamentos obtidos através do programa pelo teste de permutações aleatórias ($k = 5$ e $b = 6$).	128
45	Postos atribuídos a $k = 4$ tratamentos e $b = 3$ blocos com ocorrência de empates.	129

OBTENÇÃO DOS NÍVEIS DE SIGNIFICÂNCIA PARA OS TESTES DE KRUSKAL-WALLIS, FRIEDMAN E COMPARAÇÕES MÚLTIPLAS NÃO-PARAMÉTRICAS

Autor: ANTONIO CARLOS FONSECA PONTES

Orientador: Prof. Dr. JOSÉ EDUARDO CORRENTE

RESUMO

Uma das principais dificuldades encontradas pelos pesquisadores na utilização da Estatística Experimental Não-Paramétrica é a obtenção de resultados confiáveis. Os testes mais utilizados para os delineamentos com um fator de classificação simples inteiramente casualizados e blocos casualizados são o de Kruskal-Wallis e o de Friedman, respectivamente. As tabelas disponíveis para estes testes são pouco abrangentes, fazendo com que o pesquisador seja obrigado a recorrer a aproximações. Estas aproximações diferem dependendo do autor a ser consultado, podendo levar a resultados contraditórios. Além disso, tais tabelas não consideram empates, mesmo no caso de pequenas amostras. No caso de comparações múltiplas isto é mais evidente ainda, em especial quando ocorrem empates ou ainda, nos delineamentos inteiramente casualizados onde se tem número diferente de repetições entre tratamentos. Nota-se ainda que os *softwares* mais utilizados em geral recorrem a aproximações para fornecer os níveis de significância, além de não apresentarem resultados para as comparações múltiplas.

Assim, o objetivo deste trabalho é apresentar um programa, em linguagem C, que realiza os testes de Kruskal-Wallis, de Friedman e de comparações

múltiplas entre todos os tratamentos (bilateral) e entre os tratamentos e o controle (uni e bilateral) considerando todas as configurações sistemáticas dos postos ou com 1,000.000 de configurações aleatórias, dependendo do número total de permutações possíveis. Dois níveis de significância são apresentados: o *DW* ou *MaxDif*, baseado na comparação com a diferença máxima dentro de cada configuração e o *Geral*, baseado na comparação com todas as diferenças em cada configuração. Os valores do nível de significância *Geral* assemelham-se aos fornecidos pela aproximação normal.

Os resultados obtidos através da utilização do programa mostram, ainda, que os testes utilizando as permutações aleatórias podem ser bons substitutos nos casos em que o número de permutações sistemáticas é muito grande, já que os níveis de probabilidade são bastante próximos.

**OBTAINING SIGNIFICANCE LEVELS FOR KRUSKAL-WALLIS,
FRIEDMAN AND NONPARAMETRIC MULTIPLE
COMPARISON TESTS**

Author: ANTONIO CARLOS FONSECA PONTES

Adviser: Prof. Dr. JOSÉ EDUARDO CORRENTE

SUMMARY

One of the most difficulties for the researchers in using Nonparametric Methods is to obtain reliable results. Kruskal-Wallis and Friedman tests are the most used for one-way layout and for randomized blocks, respectively. Tables available for these tests are not too wild, so the research must use approximate values. These approximations are different, depending on the author and the results can be not similar. Furthermore, these tables do not taking account tied observations, even in the case of small sample. For multiple comparisons, this is more evident, specially when tied observations occur or the number of replications is different. Many softwares like SAS, STATISTICA, S-Plus, MINITAB, etc., use approximations in order to get the significance level and they do not present results for multiple comparison.

Thus, the aim of this work is to present a routine in C language that runs Kruskal-Wallis, Friedman and multiple comparison among all treatments (bi-tailed) and between treatment and control (uni and bi-tailed), considering all the systematic configurations of the ranks or with more than 1000000 random ones, depending on the total number of possible permutations. Two levels of significance

are presented: *DW* or *MaxDif*, based on the comparison of the maximum difference within each configuration and the *Geral*, based on the comparison of all differences for each configuration. The *Geral* values of the significance level are very similar for the normal approximation.

The obtaining results through this routine show that, the tests using random permutations can be nice substitutes for the case of the number of systematic permutations is too large, once the levels of probability are very near.

1 INTRODUÇÃO

A Estatística Não-Paramétrica pode ser vista como uma coleção de métodos estatísticos que, segundo a definição ampla, não é relacionada com parâmetros específicos ou ainda, mantém suas propriedades independente da distribuição subjacente aos dados (métodos de distribuição livre). Tais procedimentos têm sido utilizados há muitos anos, apesar de nem sempre conhecidos pelo nome não-paramétrico.

Os métodos em geral utilizados para comparações de medidas em populações, ou ainda, para verificar se os dados de uma determinada amostra provêm de uma população com uma distribuição especificada, baseiam-se na pressuposição de que a distribuição dos dados é conhecida, ou seja, que tem-se ou pode-se estimar os parâmetros da possível distribuição teórica que se adapta aos dados.

Inferências sobre os valores dos parâmetros de uma especificada família de distribuições, ou outras características destas que implicam especificar valores destes parâmetros, são denominadas inferências paramétricas, pois os parâmetros é que caracterizam a forma dos dados e sua distribuição. Em geral, é razoável assumir que as observações vêm de uma particular família de distribuições. Além disso, a experiência, respaldada pela teoria, sugere que, para muitas medidas, inferências baseadas na suposição de que as observações formam uma amostra aleatória de alguma distribuição normal podem ser utilizadas mesmo que a suposição de normalidade esteja incorreta. Mas, isto nem sempre é verdadeiro. Supor que amostras provêm de uma família de distribuições especificada pode, em muitos casos, não ser apropriada. Além disso, existem casos em que as medidas obtidas podem não ser precisas, ou ainda, o número de elementos examinados é pequeno. Nestes casos a

análise não-paramétrica é mais adequada do que a tradicional análise paramétrica. Até mesmo quando as medidas são precisas pode não ser adequado supor que os dados provêm de uma distribuição normal já que a normalidade implica em certas propriedades de simetria e dispersão. Pode estar claro que a distribuição não pode ser normal simplesmente observando a média, o desvio padrão e a amplitude dos valores como no caso em que todas as observações são nulas ou positivas e o desvio padrão é maior que a média. Há várias distribuições conhecidas que são mais achatadas que a normal, como é o caso da distribuição uniforme, ou não simétricas como, por exemplo, as distribuições gama e exponencial. Na prática, geralmente pode-se afirmar apenas que a amostra é originária de uma distribuição que é não simétrica ou muito alongada ou muito achatada, etc. Nestes casos, métodos não-paramétricos (ou como eles são algumas vezes denominados, métodos de distribuição livre) são novamente apropriados.

É comum encontrar situações em que a aplicação de testes paramétricos não é possível ou torna-se muito complicada, seja pela falta de informações a respeito da forma da distribuição da população, seja pela dificuldade de obtenção de estimativas confiáveis dos parâmetros populacionais. Existem, ainda, casos em que os dados não são quantitativos, e sim, qualitativos, podendo-se apenas estabelecer uma ordem entre eles. Os métodos não-paramétricos são relacionados ao desenvolvimento de procedimentos de inferência estatística que não fazem qualquer suposição explícita sobre a forma da distribuição dos dados. Portanto, para se trabalhar com os métodos não-paramétricos as exigências são menores. O fato de as pressuposições para a realização dos testes não-paramétricos serem fracas não significa que não haja pressuposições para a utilização dos mesmos.

A Estatística Não-Paramétrica é ainda uma ferramenta muito útil na análise de dados quando as pressuposições do modelo não se verificam, ou seja, quando os dados provenientes de um experimento não possuem normalidade ou homogeneidade de variâncias.

A base dos testes não-paramétricos está na ordenação (*ranks*) dos dados e não no seu valor intrínseco. Se, por um lado, perde-se em precisão na troca dos valores da variável por seus respectivos postos, ganha-se em eficiência e facilidade no entendimento dos resultados.

Paradoxalmente, muitos testes que são universalmente considerados como não-paramétricos ou de distribuição livre, envolvem parâmetros e distribuições (em geral as distribuições normal e binomial). Isto ocorre porque o termo não-paramétrico ou distribuição livre não se aplica às distribuições dos testes estatísticos e sim ao fato de que os métodos podem ser aplicados a amostras originadas de populações tendo qualquer das classes amplas de distribuições que necessitam apenas ser especificadas em termos gerais, ou seja, como sendo contínuas, simétricas, idênticas, diferindo apenas na mediana ou média, etc. Não é necessário especificar famílias tais como a normal, uniforme, exponencial, etc. Logicamente, o termo distribuição livre pode, então, ser mais apropriado mas o termo não-paramétrico é de uso mais difundido. Kendall & Stuart (1952) utilizam os termos distribuição livre (*distribution-free*) e não-paramétrico de formas distintas: não-paramétrico é a descrição do problema e distribuição livre é o método usado para resolver o problema. Entretanto, muitos procedimentos de distribuição livre foram desenvolvidos para problemas não-paramétricos e há então uma livre intercambialidade de sentido nos termos não-paramétrico e distribuição livre como utilizados na literatura em geral. É claro que se o problema paramétrico (por exemplo, testar uma hipótese paramétrica ou estimar um parâmetro) o método a ser utilizado pode ou não ser de distribuição livre.

Atualmente, os termos não-paramétrico ou distribuição livre utilizados para designar um teste são usados indistintamente. Um teste é denominado não-paramétrico se nenhuma suposição é feita com respeito à forma específica da distribuição das medidas. Por outro lado, um teste é denominado de distribuição livre se a distribuição de sua estatística de teste sob a hipótese nula não depende da forma particular da distribuição dos dados. Assim, ambos os termos referem-se

a testes que são válidos sem a suposição de uma distribuição artificial (por exemplo, normalidade). Estes testes são baseados em postos dos dados e na aleatorização, onde se consideram todas as possíveis permutações (rearranjos).

Métodos não-paramétricos são geralmente os únicos possíveis em experimentos cujos resultados especificam ordem simplesmente, postos ou contagens em várias categorias. Deve-se fazer uma distinção entre testes de aleatorização e testes não-paramétricos. Apesar dos testes não-paramétricos utilizarem, na sua forma exata, da aleatorização para a obtenção dos níveis de significância para a realização dos testes de hipóteses, os testes de aleatorização são mais abrangentes, compreendendo uma gama ampla de testes, paramétricos ou não. Pode-se, por exemplo, executar testes de aleatorização utilizando, dentre outros, os testes t ou F.

Testes aleatórios (*randomization tests*) ou testes de permutação são testes de significância que são admissíveis para medidas intervalares ou de escala. Testes de postos são testes de significância que são admissíveis para dados com ao menos uma escala ordinal assim denominados porque as medidas são substituídas por seus postos. Em geral, o posto 1 é assinalado à menor medida, o posto 2 à segunda menor medida e assim por diante, mas outras formas de atribuição de postos são admissíveis. Medidas de igual valor são consideradas empatadas e, em geral, é utilizado o posto médio destas medidas. Empates têm probabilidade nula de ocorrer se as medidas tomadas são contínuas mas, devido à limitação dos instrumentos de medida, quando o número de dados é grande prevê-se a ocorrência de empates, com a utilização de termos visando a correção destes. Além disso, muitas vezes, ao utilizar aproximações dos testes não-paramétricos (exatos) para distribuições conhecidas (aproximada, em geral, para a distribuição normal ou derivadas dela), uma correção para continuidade é usada.

Sprenst (1993) creditou ao Rei Nabucodonosor da Babilônia uma aplicação informal de um teste de permutação em cerca de 600 A.C., enquanto outros autores atribuem, de acordo com o texto Bíblico, ao Profeta Daniel. Arbuth-

nott é citado freqüentemente como um dos primeiros a utilizar o teste dos Sinais em 1710. De acordo com Hollander & Wolfe (1999), o marco inicial da estatística não-paramétrica é 1936, com a publicação do trabalho de Hotelling e Pabst sobre correlação de postos. O campo da estatística não-paramétrica propriamente dito vem a se estabelecer através dos trabalhos de Friedman em 1940, Mathisen em 1943, Wilcoxon em 1945 (que apresentou o teste das somas dos postos para duas amostras considerando amostras iguais e o testes de postos pareados), Mann & Whitney em 1947 (que expandem o teste de soma dos postos de Wilcoxon para amostras desiguais), dentre outros. Desenvolvimentos isolados em estatística não-paramétrica ocorreram anteriormente em trabalhos de Pearson em 1900 e 1911, Spearman em 1904, Kolmogorov em 1933, Kendall em 1938 e Smirnov em 1939. O primeiro autor a utilizar o termo não-paramétrico para designar esta sub-área da Estatística foi Wolfowitz em 1942. Outras contribuições importantes foram dadas por Pitman, Hodges & Lehmann e Chernoff & Savage, em 1948, 1956 e 1958, respectivamente, mostrando a eficiência dos testes não-paramétricos em relação aos seus competidores paramétricos. Outra técnica importante, introduzida por Quenouille em 1949, foi o *Jackknife*, que possibilitou a utilização de métodos não-paramétricos numa variedade de situações. Os artigos de Kruskal & Wallis (1952) e Jonckheere (1954) são decisivos para o desenvolvimento da Estatística Experimental Não-Paramétrica. Em 1963, Hodges & Lehmann mostraram como obter estimadores, com propriedades desejáveis, dos testes de postos, abrindo caminho para a utilização de estimadores em delineamentos experimentais e para testar e estimar parâmetros de regressão não-paramétrica. Hollander & Wolfe (1999) consideram ainda que o principal avanço ocorrido em estatística não-paramétrica nas últimas duas décadas foi o método *bootstrap*, que é um método computacionalmente intensivo para obtenção de erros padrões e intervalos de confiança em muitos casos, incluindo situações nas quais é difícil, quando não impossível, o uso de aproximação paramétrica.

Rank Correlation Methods de M.G. Kendall, publicado em 1948, foi

o primeiro livro de métodos relacionados a postos mas o primeiro livro específico de métodos não-paramétricos foi o de Siegel (1956). Hoje o campo cresceu, incluindo tópicos modernos como estimação robusta, regressão não-paramétrica e semi-paramétrica. Aplicações estão presentes em quase todas as áreas e como reflexo deste crescimento, foi lançado em 1991 um periódico, *Journal of Nonparametric Statistics* com trabalhos relacionados à esta área.

Como os métodos utilizados para a obtenção de estatística de teste se baseiam em estatísticas de ordem, isso significa que, dada uma amostra aleatória, uma estatística de ordem dessa amostra aleatória são os valores ordenados e, portanto, dependentes, o que torna as deduções muito mais difíceis. Por conta disso, o que existe disponível para análise de experimentos não é muito e torna-se inviável quando o tipo de experimento é complicado. Em vista disso, outros métodos têm sido propostos como o método *bootstrap*, as quase-verossimilhanças, que são semi-paramétricos, dentre outros.

Por sua natureza característica, as estatísticas não-paramétricas diferem das estatísticas clássicas, e estas diferenças pedem consideração especial no estudo do assunto. Concentrações de drogas, substâncias químicas, ou substâncias tóxicas que são variáveis não negativas que freqüentemente têm distribuições assimétricas são fontes de dados apropriados. Conjuntos de dados ideais para a utilização de métodos não-paramétricos são aqueles claramente não normais ou com distribuições assimétricas. Nestes, a mediana populacional é mais relevante que a média populacional. Consideram-se também aqueles para os quais a normalidade ou simetria é, no mínimo, suspeita ou ainda conjuntos de dados com dados discrepantes (*outliers*), para os quais os procedimentos não-paramétricos tendem a ser mais robustos. Se o procedimento não-paramétrico tem um concorrente clássico, então os resultados de ambos devem ser comparados. Muitas vezes as duas aproximações conduzirão a mesma conclusão, mas as suposições só podem ser válidas para um dos procedimentos. Quando os procedimentos dão resultados diferentes, normalmente o

método não-paramétrico é considerado mais apropriado. O procedimento correto para esta determinação, no entanto, deve ser escolhido baseado no delineamento experimental, nos parâmetros de interesse, nas hipóteses a serem testadas, e no poder das suposições que podem ser feitas. Outra vantagem dos procedimentos da estatística não-paramétrica é que, em geral, eles são dados sob a ótica do desenvolvimento inicial da distribuição exata, o que leva a um maior conhecimento das vantagens e desvantagens do teste que está sendo utilizado.

Em termos de *softwares*, nos pacotes mais conhecidos e utilizados, existem apenas os procedimentos básicos para uma análise de um delineamento inteiramente casualizado ou em blocos casualizados, sem qualquer menção sobre comparações múltiplas, apesar de tais resultados, mesmo sendo assintóticos, estarem disponíveis na literatura. Conjuntos de macros para pacotes não específicos foram desenvolvidos, como por exemplo para o Minitab, incluindo o teste de Kolmogorov-Smirnov para uma e duas amostras, teste de dispersão de Moses e Jackknife para duas amostras, o tau de Kendall e *kernel regression*. No SAS, os procedimentos não-paramétricos mais comuns não são encontrados dentro de um procedimento específico. Nele o teste de sinal, o teste de Kruskal-Wallis, a Correlação de postos de Spearman e o teste de Friedman são executados pelos procedimentos *PROC UNIVARIATE*, *PROC NPAR1WAY*, *PROC CORR* e *PROC RANK*, respectivamente. Já o S-Plus possui alguns procedimentos não-paramétricos em sua biblioteca, tais como o de Wilcoxon, Kruskal-Wallis, o de Friedman, dentre outros. *Softwares* estatísticos como o SAS, S-Plus, STATISTICA trabalham com aproximações normais para a obtenção dos níveis mínimos de significância para os testes não-paramétricos mais comuns apesar dos cálculos envolvidos nestes métodos serem relativamente simples.

Foram desenvolvidos, especificamente para métodos não-paramétricos, vários *softwares* estatísticos, incluindo NPSTAT em 1978, NPSP em 1983, NONPAR em 1989, StatXact em 1991, o SENP em 1992 e o XploRe em 1995, mas eles são pouco conhecidos e utilizados. Alguns deles trabalham com os testes na sua forma exata ou

ainda com testes aleatórios. No que se refere à Análise de Variância, a maioria deles se limita aos testes gerais, não trabalhando com Métodos de Comparações Múltiplas na sua forma exata.

Um dos grandes problemas ao se trabalhar métodos não-paramétricos é a obtenção dos níveis ou probabilidades de significância exatos para os testes utilizados. A construção de tabelas exatas abrangentes é trabalhosa, exigindo atenção para cada caso em particular. Conseqüentemente, a publicação de tabelas razoavelmente abrangentes requer um grande número de páginas somente para estas, como, por exemplo, em Hollander & Wolfe (1999) (180 páginas ou cerca de 23% do total de páginas do livro) ou Campos (1983)(87 páginas ou cerca de 25% do total). Isto faz com que, para o pesquisador, a consulta a estas tabelas se torne exaustiva. Mesmo em livros com grande espaço dedicado à publicação de tabelas, casos simples não podem ser consultados. Por exemplo, a tabela para o teste de Kruskal-Wallis mais abrangente considera até cinco tratamentos, com, no máximo, três repetições ou quatro tratamentos e até quatro repetições. Com isto, o pesquisador é levado a utilizar a aproximação normal em delineamentos com um número razoavelmente pequeno de tratamentos e poucas observações, o que pode resultar em erros nas conclusões. Esta lacuna nas tabelas é sentida especialmente quando o número de repetições dos tratamentos não são iguais. No caso do teste de Friedman, as tabelas são mais abrangentes, considerando até seis tratamentos e seis repetições. Quando pretende-se efetuar comparações múltiplas a dificuldade é maior ainda já que as tabelas são menos abrangentes.

Em qualquer dos casos mencionados os empates não são contemplados. Assim, na presença de empates, a tendência é a utilização de aproximações, utilizando a distribuição de χ^2 , para os testes de Kruskal-Wallis ou de Friedman, ou a distribuição normal, normal multivariada, dentre outras, no caso das comparações múltiplas. Em alguns casos é utilizada a tabela exata do teste correspondente para dar uma idéia aproximada do nível de significância da estatística de teste, desprezando-se

os empates e atribuindo-se aleatoriamente os postos àquelas observações empatadas. Outro método às vezes utilizado, quando o número de empates não é muito grande, é construir as diferentes configurações que seriam obtidas sem empates e, a partir delas, obter um *nível de significância médio* utilizando-se as tabelas exatas que existem na literatura.

Em qualquer desses casos mencionados, pode-se notar as dificuldades para a obtenção dos níveis de significância com uma precisão razoável. Some-se a isto, no caso das comparações múltiplas, a dificuldade em decidir qual tipo de erro a ser utilizado, se *comparisonwise* ou *experimentalwise*.

Assim, o objetivo deste trabalho é propor um *software* em linguagem C que obtenha tabelas e níveis de significância do teste de Kruskal-Wallis (ensaios inteiramente casualizados), do Teste de Friedman (ensaios em blocos casualizados) e para as comparações múltiplas entre todos os tratamentos e entre tratamentos e controle em ambos os tipos de delineamentos. Estes níveis de significância podem ser exatos, obtidos através de permutações sistemáticas ou aproximados, calculados por meio de testes de permutação aleatória.

2 REVISÃO DA LITERATURA

2.1 Introdução

A Estatística Não-Paramétrica não deve ser olhada apenas como um caso especial de análise de dados, e sim em conjunto com as técnicas paramétricas visando confrontar ambas, buscando, em cada caso, a melhor maneira de se trabalhar com os dados experimentais.

Conover & Iman (1981) fazem um apanhado geral sobre os procedimentos não-paramétricos, tratando-os como procedimentos de transformação de postos de forma que os estudos de testes paramétricos e não-paramétricos possam ser realizados de uma forma unificada, tornando possível aos iniciantes na estatística entender as similaridades e diferenças entre os dois tipos de análise. Segundo estes autores, existem várias formas pelas quais os postos podem ser assinalados às observações, sugerindo os seguintes tipos:

RT-1 (*rank transformation 1*): são atribuídos postos ao conjunto completo de observações, do menor ao maior, com a menor observação tendo posto 1, a segunda menor posto 2 e assim por diante; postos médios são assinalados em caso de empates (p.ex., na atribuição de postos para efetuar o teste de Kruskal-Wallis).

RT-2 (*rank transformation 2*): as observações são particionadas em subconjuntos e em cada um deles atribui-se os postos internamente, independentemente dos outros subconjuntos (é o caso, por exemplo, da atribuição de postos para efetuar o teste de Friedman).

RT-3 (*rank transformation 3*): esta transformação é a RT-1 aplicada após uma reexpressão apropriada dos dados (por exemplo, médias apropriadas podem ser

subtraídas dos dados antes de se proceder a atribuição dos postos).

RT-4 (*rank transformation 4*): idem ao anterior, somente que em relação a RT-2.

Outros tipos de atribuição de postos podem ser utilizados para casos específicos, como, por exemplo, atribuir postos dos extremos para o centro, fazendo com que os elementos com postos mais afastados do valor central tenham menores postos ou vice-versa.

Dentre os testes não-paramétricos destacam-se:

1. Teste Binomial, aplicado a dados dicotômicos;
2. Teste de χ^2 para dados nominais;
3. Teste de Kolmogorov-Smirnov, para verificar se os dados podem ser ajustados a uma distribuição qualquer, e Teste de Lilliefors, específico para o caso do ajuste à distribuição normal;
4. Teste do Sinal, Teste de McNemar e o Teste das Ordens Assinaladas (*Signed Rank Test*), apropriados para se efetuar comparações entre duas amostras pareadas;
5. Teste da Soma das Ordens (Wilcoxon), o Teste de Mann-Whithney, o Teste de Kolmogorov-Smirnov e o Teste Exato de Fisher, aplicáveis a duas amostras independentes;
6. Testes de Ansary-Bradley e o de Moses, que são relacionados à dispersão para duas amostras independentes;
7. Testes de Kruskal-Wallis, de Jonckheere-Tempstra, de Mack-Wolfe e de Fligner-Wolfe para $k \geq 3$ amostras independentes (delineamento inteiramente casualizado), cada um deles relativo a uma hipótese alternativa específica;

8. Testes de Friedman, de Page, de Quade, de Mack-Skillings e de Skillings-Mack para delineamentos em blocos casualizados.

2.2 Modelo de Aleatorização

Considere k populações sob k diferentes tratamentos e das quais retiram-se amostras de tamanhos n_1, n_2, \dots, n_k . Sejam X_{ij} a variável aleatória representando a j -ésima observação na i -ésima amostra, a_{ij} , $j = 1, \dots, n_i$ e $i = 1, \dots, k$ as observações obtidas no experimento. De acordo com Manoukian (1986), o método de aleatorização de Pitman consiste em, sob a hipótese nula, ou seja, todos os tratamentos têm efeitos equivalentes, considerar-se as observações como obtidas de uma única população. Então, existem $\binom{N}{n_1 \dots n_k}$ maneiras de formar k amostras de tamanhos n_1, \dots, n_k da amostra original combinada $(a_{11}, \dots, a_{1n_1}, a_{21}, \dots, a_{2n_2}, \dots, a_{k1}, \dots, a_{kn_k})$ que, sob a hipótese nula, são todas equiprováveis.

Seja Ω o conjunto destas k amostras. Neste caso, Ω tem $N!/[n_1! \dots n_k!]$ elementos, $N = \sum_{i=1}^k n_i$, que podem ser denotados pelas k -uplas (z_1, \dots, z_k) onde z_i tem n_i componentes, de modo que todas as permutações de seus componentes são consideradas equiprováveis. Denote ainda (Z_1, \dots, Z_k) um vetor aleatório tomando valores de Ω com probabilidades iguais a $P[(Z_1, \dots, Z_k) = (z_1, \dots, z_k)] = [n_1! \dots n_k!]/N!$ para todos (z_1, \dots, z_k) em Ω .

A estatística da análise de variância clássica para um experimento em classificação simples ou delineamento inteiramente casualizado (*one-way layout*) é dada por

$$\Psi = \frac{\frac{\sum_{i=1}^k n_i (\bar{X}_i - \bar{X}_{..})^2}{(k-1)}}{\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2}{(N-k)}}$$

onde $\bar{X}_{..} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}}{N}$ e $\bar{X}_i = \frac{\sum_{j=1}^{n_i} X_{ij}}{n_i}$

Baseado nesta expressão, pode-se utilizar a seguinte estatística para trabalhar com o modelo aleatorizado:

$$S' = \frac{\frac{\sum_{i=1}^k (\bar{Z}_{i.} - \bar{Z}_{..})^2}{(k-1)}}{\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_{i.})^2}{(N-k)}}$$

onde $Z_i = (Z_{i1}, \dots, Z_{in_i})$ com i variando de 1 a k .

Usando o fato que $\sum_{i=1}^k \sum_{j=1}^{n_i} Z_{ij}$ e $\sum_{i=1}^k \sum_{j=1}^{n_i} Z_{ij}^2$ são constantes, pode-se reescrever S' obtendo:

$$S' = \frac{(N-k) \left[\sum_{i=1}^k n_i \bar{Z}_{i.}^2 - C_1 \right]}{(k-1) \left[C_2 - \sum_{i=1}^k n_i \bar{Z}_{i.}^2 \right]}$$

onde $C_1 = N \bar{Z}_{..}^2 > 0$, $C_2 = \sum_{i=1}^k \sum_{j=1}^{n_i} Z_{ij}^2 > 0$ e $C_2 \geq \sum_{i=1}^k n_i \bar{Z}_{i.}^2$.

Como S' é uma função monotonamente crescente de $S = \sum_{i=1}^k n_i \bar{Z}_{i.}^2$, pode-se trabalhar com esta última estatística sem perda de generalidade. Seja $\sum_{i=1}^k n_i \bar{a}_i^2 = s_0$ e considere q como o número de valores S , que podem ser obtidos de Ω e que são maiores ou iguais a s_0 . Se, para um dado nível de significância α , ocorre que $\alpha \geq \frac{q(n_1! \dots n_k!)}{N!}$, a hipótese nula é rejeitada. Substituindo Z_{ij} em S pelo posto de X_{ij} na amostra combinada, recaí-se, a menos de constantes, na estatística de Kruskal-Wallis.

A estatística

$$\Psi' = \frac{\frac{\sum_{i=1}^k n_i (\bar{Z}_{i.} - \bar{Z}_{..})^2}{(k-1)}}{\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_{i.})^2}{(N-k)}}$$

é equivalente à S já que seu denominador é uma constante positiva. Os momentos da estatística Ψ' podem ser comparados com seus respectivos clássicos paramétricos (obtidos através de Ψ para grandes amostras).

Para este fim, seja $(a_{11}, \dots, a_{kn_k}) = (c_1, \dots, c_N)$, $\bar{c} = \frac{\sum_{i=1}^N c_i}{N}$, ou seja, $\bar{Z}_{..} = \bar{c}$. Pode-se verificar que $E(\Psi') = 1$ e, para amostras de tamanhos n_1, \dots, n_k

grandes

$$E[(\Psi' - 1)^2] = \frac{2}{k-1} \left[1 + \frac{1}{N} \left(3 - \frac{\sum_{i=1}^N (c_i - \bar{c})^4}{\left[\frac{\sum_{i=1}^N (c_i - \bar{c})^2}{N} \right]^2} \right) \right].$$

Estas esperanças podem ser comparadas com a média e a variância da estatística F de Fisher ($E[\Psi] = 1$ e $Var[\Psi] = \frac{2}{k-1}$), sob a hipótese nula, sugerindo assim a aproximação da distribuição de Ψ' pela Distribuição de Fisher com $(k-1)$ e $(N-k)$ graus de liberdade para grandes amostras.

2.3 Estatística Linear de Postos

A principal ferramenta na aplicação de testes não-paramétricos é a estatística linear de postos. De acordo com Negrillo (1992), se X_1, X_2, \dots, X_N é uma amostra aleatória com função distribuição F absolutamente contínua e se r_1, r_2, \dots, r_N são os seus postos respectivos, então, a estatística $T_N = \sum_{i=1}^N c_i a(r_i)$ é denominada estatística linear de postos, utilizada para fazer inferências relativas a uma sequência de observações X_i , obtidas, em geral, sob diferentes condições experimentais. Na fórmula acima, c_i são as constantes de regressão que servem para identificar as variações experimentais ou amostrais e $a(r_i)$ é a função escore avaliada nos postos r_i . Se os postos r_i são uniformemente distribuídos sobre \mathfrak{R} , a estatística T_N tem as seguintes características:

i) $E(T_N) = N\bar{c}\bar{a}$ onde $\bar{a} = \frac{1}{N} \sum_{i=1}^N a(r_i)$ e $\bar{c} = \frac{1}{N} \sum_{i=1}^N c_i$;

ii) $Var(T_N) = \frac{1}{N-1} \sum_{i=1}^N (c_i - \bar{c})^2 \sum_{i=1}^N [a(r_i) - \bar{a}]^2$

iii) T_N é simétrica em torno de $E(T_N)$;

iv) T_N é assintoticamente normal.

2.4 Teste de posição para comparação de dois tratamentos

Considere, como em Campos (1983), que N elementos experimentais sejam avaliados para um estudo comparativo, dos quais n deles, selecionados ao acaso, receberão o novo tratamento, e os restantes $m = N - n$ servirão como controle. Seja Y a variável que representa a resposta ao tratamento e X a variável que representa o controle. Suponha ainda que X e Y tenham distribuições acumuladas F e G , respectivamente. Desta forma, para se fazer inferências sobre os resultados amostrais deve-se supor que a hipótese de não efeito do tratamento estabelece que X e Y têm a mesma distribuição, ou seja, $F = G$ e que a população deve ser suficientemente grande para poder considerar X e Y independentes. As duas pressuposições levam ao modelo populacional no qual:

- i) X_1, \dots, X_m e Y_1, \dots, Y_n são variáveis aleatórias independentes;
- ii) X_i tem distribuição F para todo $i = 1, \dots, m$;
- iii) Y_j tem distribuição G para todo $j = 1, \dots, n$;
- iv) a hipótese a ser testada é a de não efeito do tratamento, ou seja $H_0 : G = F$.

A introdução do modelo populacional vem da necessidade de se estabelecer a legitimidade para a extensão dos resultados e do poder do teste a ser utilizado, ou seja, do poder de rejeição da hipótese quando o tratamento traz resultados efetivos.

Desta forma, todas as possíveis escolhas dos $N = n + m$ elementos são igualmente prováveis. Ao final do estudo, os elementos são ordenados, de acordo com o interesse do pesquisador, de forma crescente ou decrescente.

Sob a hipótese nula de igualdade dos tratamentos (ou seja, efeitos de tratamentos iguais) o posto não é afetado por qual elemento recebe qual tratamento, ou seja, considerando s_1, s_2, \dots, s_n os postos dos elementos que receberam o tratamento e r_1, r_2, \dots, r_m os postos dos elementos do controle ($m + n = N$) tem-se, sob H_0 , $P_H(S_1 = s_1, \dots, S_n = s_n) = 1/\binom{N}{n}$ para cada uma das n -uplas (s_1, s_2, \dots, s_n) ,

lembrando que os postos referentes ao tratamento ($r'_i s$) são determinados uma vez conhecidos os postos dos elementos que receberam o controle, e vice-versa.

Os testes não paramétricos (ou de distribuição livre) têm a vantagem da simplicidade por supor unicamente a aleatorização dos indivíduos ao controle e ao tratamento. Como nenhuma pressuposição é feita com relação à natureza ou à origem dos dados, isto pode limitar o alcance dos resultados.

2.4.1 Teste da Soma das Ordens de Wilcoxon

Wilcoxon propôs, em 1945, o Teste da Soma das Ordens como uma forma de decidir se um determinado tratamento é superior a um tratamento padrão (testemunha). Neste caso, a estatística a ser utilizada é $T_N = \sum_{i=1}^N c_i R_i$ e a hipótese nula, de não efeito do tratamento, será rejeitada se a soma dos postos dos n elementos que o receberam for suficientemente grande, ou seja, H_0 será rejeitada se $W_S \geq c$, onde $W_S = S_1 + S_2 + \dots + S_n$ onde S_i é o posto do elemento i e c é o valor crítico, determinado de tal forma que, sob H_0 a probabilidade de obtenção da desigualdade acima é igual ao nível de significância α , ou seja, $P_H(W_S \geq c) = \alpha$. Como nestes casos as distribuições são discretas, o valor de α não é pré-determinado como quando se trabalha com as distribuições normais, F , t de *Student*, etc.

Sob a hipótese nula, a distribuição de W_S pode ser obtida como segue: considere $\#(w; n, m)$ o número de casos em que a soma dos n postos de tratamentos seja igual a w , quando se tem os postos $1, 2, \dots, N$ divididos entre os n elementos que receberam o tratamento e os $m = N - n$ elementos do controle; neste caso temos que $P_H(W_S = w) = \frac{\#(w; n, m)}{\binom{N}{n}}$. Pode-se então obter, pela soma das probabilidades individuais, $P_H(W_S \geq c) = \alpha$ tal que α seja o mais próximo possível de um valor pré-fixado. Com o advento dos micro-computadores de alta capacidade, as dificuldades para aplicação deste método praticamente desapareceram. Porém, as tabelas confeccionadas para este teste trabalham com outras estatísticas com o intuito de facilitar a construção e a procura das mesmas. Sejam k_1 e k_2 , $k_1 \leq k_2$. Então a

distribuição de $W_S - \frac{1}{2}n(n+1)$ é a mesma quando temos $m = k_1$ e $n = k_2$ ou $m = k_2$ e $n = k_1$. Além disso, o valor mínimo de W_s é $\frac{1}{2}n(n+1)$ e então o valor tabelado é de $W_{XY} = W_S - \frac{1}{2}n(n+1)$ ou ainda $W_{YX} = W_R - \frac{1}{2}m(m+1)$, com $W_R = R_1 + R_2 + \dots + R_m$, onde R_j é o posto do elemento j do controle. Sejam X_1, X_2, \dots, X_m as observações do controle e Y_1, Y_2, \dots, Y_n as observações do tratamento. Considerando todos os mn pares (X_i, Y_j) , tem-se que W_{XY} é o número de pares tais que $X_i < Y_j$, e que W_{YX} é o número de pares tais que $X_i > Y_j$. Para valores grandes de n e m , pode-se utilizar a aproximação da distribuição de W_S ou W_R para uma distribuição normal, baseada no Teorema Central do Limite, onde $\mu = E(W_S) = [n(N+1)]/2$ e $\sigma^2 = Var(W_S) = [mn(N+1)]/12$.

A ocorrência de empates de elementos do controle e do tratamento é tratada convenientemente tomando-se para cada termo a média dos postos, levando, porém, a distribuições aproximadas. O teste bilateral pode também ser aplicado nestes casos.

Existem casos onde o interesse está em tratamentos cujos postos mantenham uma relativa concentração em torno de um valor central, ou seja, apenas tratamentos que tenham pequena dispersão é de interesse. Devido a sua construção, o método de Wilcoxon não é adequado para se trabalhar nesses casos já que ocorrência de postos extremos no controle (ou mesmo no tratamento) tende a ser anulada se ocorre em ambos os extremos. Um teste alternativo é o de Siegel-Tukey, onde os postos são atribuídos dos valores centrais aos valores extremos.

2.4.2 Teste de Mann-Whitney

Considerando as amostras X_1, \dots, X_m e Y_1, \dots, Y_n , com $m \geq n$ e classificando estes dados como no caso do Teste de Wilcoxon, Mann & Whitney definiram a estatística U que representa o número de pares (X_i, Y_j) tais que $X_i < Y_j$, ou seja,

$$U = \sum_{i=1}^m \sum_{j=1}^n a_{ij} \quad \text{onde} \quad a_{ij} = \begin{cases} 1 & \text{se } X_i < Y_j \quad i = 1, 2, \dots, m \\ 0 & \text{se } X_i > Y_j \quad j = 1, 2, \dots, n \end{cases}$$

Muitos autores denominam este teste por *Wilcoxon-Mann-Whitney* devido a relação existente entre a estatística W_S (de Wilcoxon) e a estatística U , ou seja, $W_S = \sum_{j=1}^n \{\#(X < Y_j)\} + \sum_{j=1}^n \{\#(Y \leq Y_j)\} = U + \sum_{j=1}^n j = U + [n(n+1)]/2$.

2.4.3 Testes de Dispersão aplicáveis a duas amostras

Em relação aos testes de dispersão para duas amostras independentes podemos citar o Teste de Ansari-Bradley e o Teste de Moses. Segundo Campos (1983) este último nada mais é do que a aplicação do Teste de Wilcoxon às variâncias dos dados originais.

O teste de Ansari-Bradley é um teste muito versátil e de fácil aplicação, que concorre com o teste F, apesar de ser menos eficiente. A sua estatística é baseada na soma dos postos que classificam as variáveis da seguinte forma: ordem 1 (um) aos dois valores extremos (maior e menor valor), ordem 2 (dois) aos dois valores extremos seguintes, e assim por diante. Dessa forma, a amostra com maior dispersão tenderá a ter menor soma das ordens de seus valores. Obviamente, se a ordenação for feita de modo inverso, a amostra com maior dispersão será aquela que tem maior soma das ordens. Para a utilização deste teste de dispersão pressupõe-se que as medianas das amostras sejam aproximadamente iguais. Caso isto não ocorra, propõe-se utilizar um ajuste aos dados de um dos grupos subtraindo destes a diferença entre a mediana dos grupos (amostras).

2.5 Análise de Variância Não-Paramétrica: Classificação Simples (k amostras independentes)

Nesta seção alguns testes utilizados na comparação de k tratamentos, ou seja, para verificar se k amostras independentes provêm de uma mesma população, são apresentados. O mais conhecido destes testes é o Teste de Kruskal-Wallis, cuja alternativa é a mais geral possível. Alguns testes com características específicas na hipótese alternativa (hipótese ordenada, tratamentos *versus* controle, etc.) também

foram desenvolvidos, visando atender casos com alguma utilização na prática.

2.5.1 Teste de Kruskal-Wallis

Comparações envolvendo mais do que dois tratamentos sem a utilização de controle local (característicos dos delineamentos inteiramente casualizados) são bastante utilizados em todas as áreas da ciência. O teste não-paramétrico mais utilizado para este tipo de comparação é o teste proposto por Kruskal & Wallis (1952). Em paralelo com o que é feito para dois tratamentos, atribui-se postos às variáveis, independentemente dos tratamentos (ou seja, utiliza-se o RT-1, conforme descrição dada anteriormente).

Seja X_{ij} ($i = 1, \dots, k; j = 1, \dots, n_i$), um conjunto de variáveis aleatórias independentes e $F_i(x)$ a distribuição contínua de X_{ij} . O modelo considerado neste caso é $X_{ij} = \mu + \tau_i + \epsilon_{ij}$, onde μ é a média geral e os τ_i 's são os efeitos de tratamentos com a restrição $\sum_{i=1}^k n_i \tau_i = 0$. Os parâmetros $\mu + \tau_i$ são os parâmetros de locação e os erros ϵ_{ij} são considerados independentes e identicamente distribuídos. Suponha ainda que $F_i(x) = G(x - \tau_i)$, $i = 1, \dots, k$. A hipótese a ser testada é a de nenhuma diferença entre os efeitos de tratamentos, ou seja, $H_0 : \tau_1 = \dots = \tau_k$ contra a hipótese alternativa $H_a : \tau_i \neq \tau_j$ para pelo menos um par $i, j \in [1, \dots, k]$ ($i \neq j$)

Considerando N elementos submetidos a k tratamentos, cada um deles com n_i repetições, ou seja, $N = \sum_{i=1}^k n_i$ e denotando o posto de X_{ij} na amostra combinada por R_{ij} , assume-se que dentro de cada grupo eles são numerados em ordem crescente, ou seja, $R_{11} < \dots < R_{1n_1}, R_{21} < \dots < R_{2n_2}, \dots, R_{k1} < \dots < R_{kn_k}$.

Sob a hipótese nula de não diferença entre tratamentos, as $\binom{N}{n_1 \dots n_k}$ possíveis divisões dos inteiros $1, \dots, N$ nos k grupos de tamanhos

n_1, \dots, n_k têm, cada uma delas, a probabilidade

$$P_H(R_{11} = r_{11}, \dots, R_{1n_1} = r_{1n_1}; \dots; R_{k1} = r_{k1}, \dots, R_{kn_k} = r_{kn_k}) = \frac{1}{\binom{N}{n_1 \dots n_k}}$$

ou seja,

$$P_H(R_{11} = r_{11}, \dots, R_{1n_1} = r_{1n_1}; \dots; R_{k1} = r_{k1}, \dots, R_{kn_k} = r_{kn_k}) = \frac{n_1! \dots n_k!}{N!}$$

e esta é a distribuição nula básica dos postos, que é uma generalização da distribuição para dois tratamentos, ou seja, sob a hipótese nula, $(X_{11}, \dots, X_{kn_k})$ pode ser considerada como uma amostra de tamanho N vindo de uma única população.

Sejam

$$R_{i.} = \sum_{j=1}^{n_i} R_{ij} ; \quad R_{..} = \sum_{i=1}^k R_{i.} ; \quad \bar{R}_{i.} = \frac{R_{i.}}{n_i} ; \quad \bar{R}_{..} = \frac{R_{..}}{N} = \frac{N+1}{2}.$$

A estatística de Kruskal-Wallis é definida por

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \left(n_i \bar{R}_{i.} - \frac{N+1}{2} \right)^2$$

onde o coeficiente $\frac{12}{N(N+1)}$ é um fator de normalização adequado.

Através de P_H pode-se determinar a distribuição nula da estatística H . Se a hipótese nula não é verdadeira, então H deverá tomar valores grandes e portanto, neste caso, rejeitamos H_0 em favor de H_a , isto é, com um nível de significância α , rejeitamos H_0 se $H \geq c_N$, onde $P_{H_0}[H \geq c_N] = \alpha$. Sob a hipótese nula, pode-se considerar a amostra combinada $(X_{11}, \dots, X_{1n_1}, \dots, X_{k1}, \dots, X_{kn_k})$ como sendo proveniente de uma mesma população. Portanto, se $c(n_1, \dots, n_k; h)$ denota o número de partições de N observações dentro de todas os k subgrupos possíveis de tamanhos n_1, n_2, \dots, n_k tais que $H = h$, então

$$P[H = h] = c(n_1, \dots, n_k; h) \frac{n_1! \dots n_k!}{N!}.$$

A estatística H pode ser reescrita de muitas outras formas equivalentes:

$$H = \frac{12}{N(N+1)} \left[\sum_{i=1}^k \frac{R_{i.}^2}{n_i} - \frac{N(N+1)^2}{4} \right] = \frac{12}{N(N+1)} \left(\sum_{i=1}^k \frac{R_{i.}^2}{n_i} \right) - 3(N+1) = \sum_{i=1}^k T_i^2$$

onde

$$T_i = \frac{R_{i.} - \frac{n_i(N+1)}{2}}{\frac{N^{3/2}}{\sqrt{12}} \sqrt{\frac{n_i}{N} \left(1 + \frac{1}{N}\right)}}$$

Nota-se que $R_{i.}$ e, por conseqüência, $\bar{R}_{i.}$ são invariantes sob as permutações dos postos r_{i1}, \dots, r_{in_i} , isto é, pode-se supor, como anteriormente, a ordem crescente para os postos dentro de cada tratamento. Sob a hipótese nula, as seguintes relações ocorrem:

$$E[R_{i.}] = \sum_{j=1}^{n_i} E[R_{ij}] = \frac{n_i(N+1)}{2}$$

$$E[R_{i.}R_{j.}] = \sum_{\ell=1}^{n_i} \sum_{m=1}^{n_j} E[R_{i\ell}R_{jm}] = \frac{n_i n_j (N+1)(3N+2)}{2} \quad \text{para } i \neq j$$

$$Cov[R_{i.}, R_{j.}] = - \left(\frac{N^3}{12} \right) \left(\frac{n_i}{N} \right) \left(\frac{n_j}{N} \right) \left(1 + \frac{1}{N} \right) = \frac{n_i n_j (N+1)}{12} \quad \text{para } i \neq j$$

$$\sigma^2(R_{i.}) = \left(\frac{N^3}{12} \right) \left(\frac{n_i}{N} \right) \left(1 - \frac{n_i}{N} \right) \left(1 + \frac{1}{N} \right) = \frac{n_i(N-n_i)(N+1)}{12}$$

Considerando a estatística T_i tem-se:

$$E[T_i] = 0 \quad ; \quad Cov[T_i, T_j] = \delta_{ij} - \sqrt{\frac{n_i n_j}{N N}}$$

onde δ_{ij} é o delta de Kronecker, ou seja, $\delta_{ii} = 1$ e $\delta_{ij} = 0$ para $i \neq j$. Dessa forma, tem-se que

$$E[H] = \sum_{i=1}^k E[T_i^2] = k - 1$$

e ainda

$$\sigma^2(H) = 2(k-1) - \frac{2}{5N(N+1)} \left[3k^2 - 6k + N(2k^2 - 6k + 1) \right] - \frac{6}{5} \sum_{i=1}^k \frac{1}{n_i}$$

Segundo Manoukian (1986), a distribuição nula assintótica da estatística H é $\chi_{k-1}^2(\eta)$ quando $(N \rightarrow \infty)$, onde η é o parâmetro de não-centralidade.

2.5.2 Teste de Jonckheere-Terpstra

Outro teste bastante utilizado é o Teste de Jonckheere-Terpstra, apresentado inicialmente por Jonckheere (1954) onde a hipótese alternativa é do tipo $H_1 : F_1(x) \leq F_2(x) \leq \dots \leq F_k(x)$, ou seja, por um prévio conhecimento dos dados, sabe-se que existe uma ordenação dos tratamentos. Este teste pode ser considerado como uma extensão do Teste unilateral de Wilcoxon-Mann-Whitney quando o número de amostras (tratamentos) independentes é maior que dois.

Considere o modelo $X_{ij} = \mu + \tau_i + \epsilon_{ij}$ e seja $F_i(x)$ a função distribuição das variáveis aleatórias independentes X_{ij} . Vê-se que $\mu + \tau_i$ são parâmetros de locação. Suponha ainda que os erros sejam independentes e identicamente distribuídos e que $F_i(x) = G(x - \tau_i)$. O Teste é baseado então na estatística

$$J = \sum_{j < i} \left[U_{ij} - \frac{n_i n_j}{2} \right] \quad \text{onde} \quad U_{ij} = \sum_{a=1}^{n_i} \sum_{b=1}^{n_j} \psi(X_{ia} - X_{ib}).$$

Devido ao fato da estatística de Mann-Whitney ter uma estreita relação com a estatística de Wilcoxon e, ao mesmo tempo, ser a base para o teste de Jonckheere, pode-se obter o poder do teste de Jonckheere para algumas alternativas simples, utilizando-se a tabela de distribuição de W_G do teste de Wilcoxon.

Büning (1999) apresenta dois testes adaptativos do tipo Jonckheere para alternativas ordenadas, considerando estatísticas JT (*Jonckheere-type*) e medidas de assimetria e curtose, definindo quatro categorias e a estatística adequada em cada caso. De acordo com o autor, a estatística de Jonckheere baseada nos escores de Mann-Whitney só deve ser utilizada quando houver pouca assimetria e grau de curtose médio. Para outros casos, a utilização de tipos de escores diferentes é recomendada.

2.5.3 Teste de Bhapkar

Bhapkar (1961) apresenta um teste com hipóteses nula e alternativas equivalentes ao de Kruskal-Wallis, mas diferente na sua construção. Considere $v^{(i)}$

o número de conjuntos de k elementos que podem ser formados escolhendo uma observação de cada amostra, tal que a observação da i -ésima amostra seja menor, ou seja $v_i = \sum_{s=1}^{n_i} \prod_{r \neq i} X_{rs} > X_{ij}, s = 1, 2, \dots, n_r$.

A estatística de teste proposta por Bhapkar é

$$V = N(2k - 1) \left[\sum_{i=1}^k p_i (u^{(i)} - k^{-1}) - \left\{ \sum_{i=1}^k p_i (u^{(i)} - k^{-1}) \right\}^2 \right]$$

onde

$$N = \sum_{i=1}^k n_i \quad p_i = \frac{n_i}{N} \quad u^{(i)} = \frac{v^{(i)}}{n_1 \cdot n_2 \dots n_k}.$$

Para H_0 verdadeira temos $u^{(i)} = k^{-1}$ o que implica em $V = 0$, ou seja, V é a medida do desvio de H_0 e V converge para uma distribuição qui-quadrado com $k - 1$ graus de liberdade.

2.5.4 Teste de Fligner-Wolfe para tratamentos versus controle

Hollander & Wolfe (1999) apresentam um teste, proposto em 1982, cuja finalidade é verificar qual tratamento (se ele existe) difere de um tratamento pré-qualificado como controle. Considerando, sem perda de generalidade, o tratamento 1 como controle, a hipótese a ser testada é $H_0 : t_i = t_1$ para $i = 2, \dots, k$, contra a hipótese alternativa de que pelo menos um dos tratamentos difere unilateralmente do controle. Para calcular a estatística para o Teste de Fligner-Wolfe (FW), ordena-se os dados da mesma maneira utilizada para o teste de Kruskal-Wallis (RT-1) incluindo todos os tratamentos e o controle, e calcula-se $FW = \sum_{j=2}^k \sum_{i=1}^{n_j} r_{ij}$, ou seja, a estatística FW é a soma dos postos atribuídos aos tratamentos (exceto o controle). O teste a ser realizado é unilateral, podendo a hipótese alternativa ser $t_i \geq t_1$ com ao menos uma desigualdade ocorrendo, e neste caso, rejeitamos H_0 ao nível de significância α se $FW \geq f_\alpha$, onde f_α é um valor tabelado de tal forma que a probabilidade do erro tipo I seja igual a α . Por outro lado, caso a hipótese alternativa seja $t_i \leq t_1$, a rejeição de H_0 , ao nível de significância α , ocorrerá se $FW \leq N_*(N + 1) - f_\alpha$ onde $N_* = \sum_{j=2}^k n_j$.

São dadas também aproximações para grandes amostras e correção para o caso de ocorrerem empates.

2.5.5 Teste de Shirley (doses crescentes de substâncias *versus* controle)

Shirley (1977) propôs um teste não paramétrico para grandes amostras, baseado no Teste de Williams (paramétrico) para comparação de doses crescentes de uma substância com um dose nula (controle) visando determinar o menor nível de dose para a qual há evidência de um efeito tóxico. Esta estatística é baseada na diferença entre a média das doses envolvidas e a dose zero, num procedimento do tipo *stepwise*, iniciando pela dose maior. Caso não seja detectada significância em alguma das doses, não é dada continuidade ao procedimento. Williams (1986) propôs uma modificação à estatística de Shirley visando torná-la mais efetiva. Não foram encontradas porém, na bibliografia pesquisada, distribuição nula e tabelas para pequenas amostras para este tipo de teste.

2.5.6 Teste de Mack & Wolf (*umbrella alternatives*)

Existem casos em que temos alguma informação a priori relativa às possíveis alternativas à hipótese nula de k distribuições idênticas e, para isto, testes específicos para detectar estas alternativas devem ser utilizados. Mack & Wolf (1981) propuseram procedimentos não-paramétricos para a classe de alternativas da forma $H_1 : F_1(x) \geq \dots \geq F_{g-1}(x) \geq F_g(x) \leq F_{g+1}(x) \leq \dots \leq F_k(x)$ com ao menos uma desigualdade estrita para ao menos um valor de x , contra a hipótese nula usual $H_0 : F_1(x) = \dots = F_{g-1}(x) = F_g(x) = F_{g+1}(x) = \dots = F_k(x)$.

Estas são denominadas alternativas *guarda-chuva* (*umbrella alternatives*) devido à configuração da correspondente mediana populacional e g é denominado o pico do *guarda-chuva*. Tais alternativas são análogas à regressão quadrática, e o caso especial em que o ponto g é conhecido a priori como da k -ésima população, corresponde à alternativa monotona ordenada (Teste de Jonckherre).

Alternativas deste tipo são apropriadas, por exemplo, na avaliação de ganho marginal na eficiência da performance como uma função do grau de treinamento, efetividade de drogas em função do tempo, produção como função da quantidade de fertilizante aplicado, reação a níveis crescentes de certas drogas aplicadas, efeito da idade nas respostas a certos estímulos, dentre outras. No artigo mencionado são considerados testes para o caso em que o ponto de pico (g) é conhecido e para o caso em que o ponto é desconhecido a priori (este último caso ocorre com mais frequência na prática).

Para o primeiro caso (pico conhecido), os autores propõem rejeitar H_0 para grandes valores da estatística $A_g = \sum \sum_{1 \leq i < j \leq g} U_{ij} + \sum \sum_{g \leq i < j \leq k} U_{ij}$ onde U_{uv} corresponde ao número das u -ésimas observações amostrais que precedem a v -ésima observação amostral, ou seja, é a estatística de Mann-Whitney entre a u -ésima e a v -ésima amostra. Nota-se que as direções das contagens de Mann-Whitney diferem em ambos os lados do ponto de *guarda-chuva* e que nenhuma comparação é feita entre as amostras que precedem o ponto e aquelas que o seguem (desde que as relações de ordem entre as correspondentes distribuições são consideradas desconhecidas). Para obter a distribuição nula exata para pequenas amostras da estatística de teste A_g , coloca-se A_g como a soma de várias variáveis aleatórias independentes (sob H_0), cada uma das quais tem uma distribuição nula de Mann-Whitney apropriada. Isto faz com que se possa usar as propriedades conhecidas de convoluções de variáveis independentes e obter a distribuição nula de A_g para uma variedade de valores de tamanhos de amostras sem que seja necessário obter todas as configurações dos postos apropriados. A distribuição nula exata completa de A_g foi tabulada para alguns valores de k e g e algumas combinações dos tamanhos de amostras. Nota-se ainda que a distribuição nula de A_g pode ser tabulada para quaisquer valores de k , g e combinações de tamanhos de amostras, desde que as distribuições de Mann-Whitney necessárias sejam facilmente obtidas.

Para o caso mais comum na prática, em que o ponto de pico (*umbrella*

point) é desconhecido os autores propõem inicialmente a utilização dos dados para estimar o pico desconhecido (g) e então usar a estatística correspondente ao valor conhecido de g como dada no caso anterior. Assim, a rejeição de H_0 em favor da hipótese alternativa do tipo *guarda-chuva*, para g desconhecido, ocorre para valores grandes da estatística $A_{\hat{g}} = \frac{\sum_{i=1}^k \chi_i [A_t - \mu_0(A_t)]}{\sigma_0(A_t)}$ onde A_t é a estatística do pico conhecido (no t -ésimo grupo) dada anteriormente, $\mu_0(A_t)$ e $\sigma_0^2(A_t)$ são, respectivamente, a média e a variância sob a hipótese nula com $g = t$ e as variáveis aleatórias χ_1, \dots, χ_k indica(m) qual ou quais grupos foram estimados pelas amostras para ser(em) o(s) grupo(s) de pico. Em geral há apenas um grupo de pico e, neste caso, o valor de $A_{\hat{g}}$ é igual ao da estatística padronizada conhecida com pico no t -ésimo grupo, ou seja, $g = t$. A definição anterior é necessária pois, para certas combinações de tamanhos de amostras, existe uma probabilidade positiva de que existam dois ou mais grupos de pico e em tais situações o valor de $A_{\hat{g}}$ é igual a média das r estatísticas padronizadas g conhecidas correspondendo aos picos para cada uma das r amostras. Estes empates não causam problemas pois, no caso em que se calcula a distribuição nula exata, eles estão sendo levados em conta e, no caso assintótico (todos os tamanhos de amostras tendendo ao infinito), a probabilidade de um empate entre estas variáveis é nula. Para obter valores críticos aproximados para o método com ponto de pico desconhecido, foram utilizadas simulações através do método de Monte Carlo.

Pan (1996), propôs testes com distribuição livre para alternativas do tipo *guarda-chuva*, baseados na decomposição das alternativas, unindo as de ordem simples, ou seja, considerando-se a diferença e a associação entre um cone convexo de ordenação do tipo *guarda-chuva* com pico diferente de 1 ou k , e um cone convexo de alternativas simples (ou seja, com pico 1 ou k). Por exemplo, se temos $k = 3$ tratamentos com pico no tratamento 2, de tal forma que $\theta_1 \leq \theta_2 \geq \theta_3$, consideramos a união de dois cones de ordenamento simples, ou seja, com $\theta_1 \leq \theta_3 \leq \theta_2$ e com $\theta_3 \leq \theta_1 \leq \theta_2$. Generalizando, para k tratamentos com pico no tratamento g , este será decomposto em $s = \binom{k-1}{g-1}$ cones distintos de ordenamentos simples. Para cada

um deles calculamos a estatística de Jonckheere-Tempstra J_r , para $r = 1, \dots, s$, dada por

$$J_r = \sum_{1 \leq i \leq j \leq k} U_{ij} \quad \text{onde} \quad U_{ij} = \sum_{u=1}^{n_i} \sum_{v=1}^{n_j} \psi(X_{iv} - X_{ju}) \quad \text{e} \quad \begin{cases} \psi(x) > 0 & \text{se } x > 0 \\ \psi(x) = 0 & \text{se } x \leq 0 \end{cases}$$

Para testar a hipótese nula de igualdade dos tratamentos contra a hipótese alternativa $H_1 : \theta_1 \leq \dots \leq \theta_{g-1} \leq \theta_g \geq \theta_{g+1} \geq \dots \geq \theta_k$ com pico conhecido g rejeitamos H_0 para valores grandes da estatística de teste $B_p = \max_{1 \leq r \leq s} J_r$. Trabalhou-se com testes tanto para o pico do *guarda-chuva* conhecido como para desconhecido. Através de estudos simulados de poder, o autor mostra que estes testes são mais balanceados e, em geral, mais poderosos que o proposto por Mack & Wolfe (1981).

2.6 Análise de Variância Não-Paramétrica: Classificação Dupla (k amostras dependentes)

Quando compara-se k tratamentos, pode existir um outra fonte de variação detectável pelo experimentador que pode ser isolada. Pode-se então dividir as unidades experimentais a serem testadas em subgrupos mais homogêneos, denominados blocos. Então, se em cada bloco todos os tratamentos forem alocados de forma aleatória, tem-se o delineamento em blocos casualizados completos de tamanho k . Delineamentos em blocos casualizados podem ser considerados como uma generalização dos testes para dados pareados. Neste tipo de delineamento tem-se dois fatores (tratamentos e blocos), cada um dos quais com dois ou mais níveis e considera-se que inexistem interações entre os níveis dos fatores.

O teste mais conhecido para a comparação de k tratamentos para o caso do delineamento em blocos casualizados é o Teste de Friedman, cuja alternativa é a mais geral possível. Outros testes visando constituir uma alternativa mais poderosa ao Teste de Friedman foram construídos, como, por exemplo, o teste de Quade.

Outros ainda, com características específicas na hipótese alternativa como o teste de Page (hipótese alternativa ordenada), também foram desenvolvidos.

2.6.1 Teste de Friedman

Seja X_{ij} , $i = 1, \dots, k$ e $j = 1, \dots, b$, um conjunto de variáveis aleatórias independentes e $F_i(x)$ a distribuição contínua de X_{ij} . Considere, como em Manoukian (1986), o modelo $X_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$ onde μ é a média geral, τ_i ($i = 1, \dots, k$) o efeito do tratamento i , β_j ($j = 1, \dots, b$) referente ao efeito do bloco j , com as restrições $\sum_{i=1}^k \tau_i = 0$ e $\sum_{j=1}^b \beta_j = 0$. Os parâmetros $\mu + \tau_i + \beta_j$ são os parâmetros de locação e os erros ϵ_{ij} são considerados independentes e identicamente distribuídos. Suponha ainda que $F_{ij}(x) = F_j(x - \tau_i)$, $F_j(x) = G(x - \beta_j - \mu)$ e que $G(x)$ é contínua e estritamente decrescente. A hipótese a ser testada é a de que não há nenhuma diferença entre os efeitos de tratamentos contra a hipótese alternativa de que pelo menos um par de tratamentos diferem entre si, ou seja

$$H_0 : \tau_1 = \dots = \tau_k$$

vs

$$H_a : \tau_i \neq \tau_j \text{ para pelo menos um par } i, j \in [1, \dots, k] \text{ (} i \neq j \text{)}$$

Substituindo os valores X_{ij} por seus postos, atribuídos de acordo com a transformação de postos $RT-2$ dada em Conover & Iman (1981), tem-se que R_{ij} corresponde ao postos do i -ésimo tratamento no j -ésimo bloco. Considerando a não ocorrência de empates dentro do mesmo bloco tem-se que o vetor de postos atribuídos aos k tratamentos dentro do bloco j , $\{R_{1j}, R_{2j}, \dots, R_{kj}\}$ é alguma permutação dos inteiros $\{1, 2, \dots, k\}$. Considera-se os blocos independentes e sob a hipótese nula, tem-se que $P(R_{ij} = r) = \frac{1}{k}$, $P(R_{ij} = r, R_{i'j} = s) = \frac{1}{k(k-1)}$, ($i \neq i'$) e ainda, $E[R_{ij}] = \frac{k+1}{2}$, $\sigma^2(R_{ij}) = \frac{k^2-1}{12}$ e $Cov(R_{ij}, R_{i'j}) = -\frac{k+1}{12}$, ($i \neq i'$).

Ainda sob H_0 , tem-se que todas as $k!$ possíveis ordenações são igualmente prováveis e ainda, as atribuições de postos nos diferentes blocos são indepen-

dentos. Assim, para b blocos tem-se que

$$P_H(R_{11} = r_{11}, \dots, R_{k1} = r_{k1}; \dots; R_{1b} = r_{1b}, \dots, R_{kb} = r_{kb}) = \frac{1}{(k!)^b}$$

Para testar a hipótese nula, utiliza-se a seguinte estatística

$$S = \frac{12b}{k(k+1)} \sum_{i=1}^k (\bar{R}_{i.} - \bar{R}_{..})^2 = \frac{12}{k(k+1)} \left[\sum_{i=1}^k \frac{R_{i.}^2}{b} - \frac{R_{..}^2}{bk} \right]$$

onde $R_{i.} = \sum_{j=1}^b R_{ij}$, $\bar{R}_{..i} = \frac{R_{i.}}{b}$, $R_{..} = \sum_{j=1}^b \sum_{i=1}^k R_{ij} = \frac{bk(k+1)}{2}$ e $\bar{R}_{..} = \frac{R_{..}}{b} = \frac{k+1}{2}$.

A hipótese nula será rejeitada para grandes valores de S e, considerando-se $c(k, b, s)$ como o número de S igual a s dentre as $(k!)^b$ permutações tem-se

$$P[S = s] = \frac{c(k, b, s)}{(k!)^b}$$

Reescrevendo a estatística S como $S = \sum_{i=1}^k Z_i^2$ onde

$$Z_i = \sqrt{\frac{12}{bk(k+1)}} \sum_{j=1}^b (R_{ij} - \bar{R}_{..}) = \sqrt{\frac{12}{bk(k+1)}} \sum_{j=1}^b \left(R_{ij} - \frac{k+1}{2} \right)$$

tem-se, sob H_0 que $E[Z_i] = 0$ e $Cov[Z_i, Z_{i'}] = \delta_{ii'} - \left(\frac{1}{k}\right)^{1/2} \left(\frac{1}{k}\right)^{1/2}$ e, para $b \rightarrow \infty$, Z_i tem distribuição assintótica $N\left(0, 1 - \frac{1}{k}\right)$.

Assim, de acordo com Manoukian (1986) a distribuição nula assintótica da estatística S é $\chi_{k-1}^2(\eta)$, onde η é o parâmetro de não-centralidade.

2.6.2 Teste de Quade

De acordo com Conover (1999), o Teste de Quade é uma extensão do teste dos postos assinalados de Wilcoxon, tendo mais poder do que o teste de Friedman nos casos em que temos apenas três tratamentos, mesmo poder quando tem-se entre três e cinco tratamentos e menor poder nos demais casos. Para o cálculo da estatística de teste no Teste de Quade, calcula-se a amplitude de cada bloco,

utilizando-se os valores originais, ou seja, obtém-se

$$Amplitude_j = \max_i \{X_{ij}\} - \min_i \{X_{ij}\}, \quad i = 1, \dots, k; \quad j = 1, \dots, b.$$

Assinala-se então postos aos blocos de acordo com as amplitudes, obtendo-se Q_1, Q_2, \dots, Q_b e em seguida calcula-se $S_{ij} = Q_j \left[R_{ij} - \frac{k+1}{2} \right]$ onde R_{ij} são os postos assinalados às observações dentro dos blocos, como no teste de Friedman. Calcula-se ainda $S_i = \sum_{j=1}^b S_{ij}$ ($i = 1, 2, \dots, k$) e então a estatística para o Teste de Quade é dada por

$$T = \frac{(b-1)B}{A-B} \quad \text{onde} \quad A = \sum_{i=1}^k \sum_{j=1}^b S_{ij}^2 \quad \text{e} \quad B = \frac{1}{b} \sum_{i=1}^k S_i^2.$$

Verifica-se que $A = b(b+1)(2b+1)k(k+1)(k-1)/72$ se não ocorrem empates. Se $A = B$ considera-se o ponto na região crítica e o p -valor é dado por $\alpha = \left(\frac{1}{k!}\right)^{b-1}$.

Conover (1999) destaca que a distribuição exata de T é difícil de ser encontrada e então a distribuição F é utilizada como uma aproximação, ou seja, rejeita-se a hipótese nula de igualdade dos tratamentos ao nível de significância α se o valor calculado T excede o quantil $1 - \alpha$ da distribuição F com $k-1$ e $(b-1)(k-1)$ graus de liberdade. Entretanto, para pequenas amostras, pode-se obter a distribuição exata da estatística T pela listagem de todas as possíveis $(k!)^b$ configurações de postos às unidades experimentais e $k!$ atribuições de postos às amplitudes dos blocos, totalizando $k!(k!)^b = (k!)^{b+1}$ configurações possíveis.

2.6.3 Teste de Page

O Teste de Page é um teste não-paramétrico onde, considerando-se a hipótese nula usual de igualdade entre os tratamentos $H_0 : \tau_1 = \tau_2 = \dots = \tau_k$, tem-se como hipótese alternativa $H_1 : \tau_1 \leq \tau_2 \leq \dots \leq \tau_k$ ou seja, por informações anteriores sabe-se que, no caso da hipótese nula ser rejeitada, existe uma ordenação entre os tratamentos. Este teste é baseado na correlação de postos entre a ordenação

hipotética e as somas dos postos dos tratamentos, onde as atribuições de postos às unidades experimentais é feita através do *RT-2* como no Teste de Friedman. A estatística de Page (L) é então dada por

$$L = \sum_{i=1}^k iR_i = R_1 + 2R_2 + \dots + kR_k \quad \text{onde} \quad R_i = \sum_{j=1}^b R_{ij}$$

Para testar H_0 contra a alternativa ordenada H_1 ao nível de significância α , rejeitamos H_0 se $L \leq l_\alpha$ onde a constante l_α é escolhida de forma que a probabilidade de se cometer o erro tipo I, ou seja, rejeitar H_0 quando ela é verdadeira, seja α . Tabelas são encontradas dando valores de l_α . A aproximação, no caso de grandes amostras, é $L^* = \{L - E_0(L)\} / [\sqrt{Var_0(L)}]$ onde $E_0(L) = [bk(k+1)^2]/4$ e $Var_0(L) = [bk^2(k+1)(k^2-1)]/144$ e, quando H_0 é verdadeira, L^* tem (para $b \rightarrow \infty$), distribuição assintoticamente normal de padrão (média 0 e variância 1) e rejeitamos a hipótese nula se $L^* \geq z_\alpha$. Hettmansperger (1975) propõe uma extensão ao Teste de Page para o caso geral de números desiguais de observações por célula.

2.6.4 Outros testes não-paramétricos aplicáveis aos Delineamentos em Blocos Casualizados

Scheirer *et alli* (1976) apresentam um método para a análise de dados em postos para experimentos fatoriais em delineamentos em blocos casualizados, que é uma extensão do teste de Kruskal-Wallis, permitindo o cálculo dos efeitos de interação e contrastes lineares. Os autores partem da idéia do particionamento da estatística de Kruskal-Wallis em componentes pelo uso de contrastes lineares ou em efeitos principais e interações, da mesma maneira como é feito na Análise de Variância Paramétrica.

Skillings & Mack (1981) apresentam uma estatística baseada no Teste de Friedman que pode ser utilizada tanto para delineamentos em blocos balanceados como desbalanceados (com perda de dados), apresentando tabelas exatas para o caso em que houve perda de parcelas em apenas um dos tratamentos, considerando

ainda comparações múltiplas para este caso. Para o caso de delineamentos em blocos incompletos balanceados são apresentados procedimentos de comparações múltiplas com tabelas exatas.

Fawcett & Salter (1984) apresentam um estudo, utilizando o método de Monte Carlo, comparando o teste F e três outros testes baseados em postos: o teste de Quade, a transformação de postos para delineamentos em blocos casualizados e a estatística AR , que aplica a transformação de postos aos dados após o alinhamento destes. Nesta última, o modelo clássico para blocos casualizados $y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$ é substituído pelo modelo $y'_{ij} = \mu + \tau_i + \epsilon_{ij}$ onde $y'_{ij} = y_{ij} - (\bar{y}_{.j} - \bar{y}_{..})$. Os autores afirmam não ser possível obter conclusões gerais relativas ao poder e robustez das estatísticas consideradas no artigo. Entretanto, concluem que, mesmo quando as suposições clássicas são satisfeitas, o teste F pode ser menos robusto e de menor poder que os testes baseados em postos. Evidencia-se ainda que o método de transformação de postos pode ser um meio de lidar com a perda de poder resultante da atribuição de postos aos blocos separadamente e da suscetibilidade aos efeitos dos parâmetros *nuisance* resultante da atribuição de postos sem o alinhamento.

Kepner & Robinson (1984) propõem um teste de distribuição livre com atribuição de postos para detectar alternativas ordenadas nos delineamentos em blocos casualizados e delineamentos em medidas repetidas para quatro ou menos tratamentos. Dado o modelo $X_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$, $i = 1, \dots, k$ e $j = 1, \dots, b$, considere

$$Y_{i1} = X_{i2} - X_{i1} \quad \text{se } k = 2$$

$$Y_{i1} = X_{i2} - X_{i1} \quad \text{e} \quad Y_{i2} = 2X_{i3} - X_{i1} - X_{i2} \quad \text{se } k = 3$$

$$Y_{i1} = X_{i3} - X_{i1}; \quad Y_{i2} = X_{i4} - X_{i2} \quad \text{e} \quad Y_{i3} = X_{i4} - X_{i3} + X_{i2} - X_{i1} \quad \text{se } k = 4,$$

rejeita-se a hipótese de igualdade dos tratamentos em favor da hipótese alternativa ordenada $H_1 : \tau_1 \leq \tau_2 \leq \dots \leq \tau_k$ para grandes valores da estatística $W = \sum_{i=1}^{k-1} \sum_{j=1}^b \psi_{ij}^* R_{ij}^+$, onde R_{ij}^+ o posto de $|Y_{ij}|$ no conjunto dos $n(k-1)$ valores absolutos (aos quais foram atribuídos postos de forma conjunta) e ψ_{ij}^* é o sinal de Y_{ij} .

House (1986) propõe uma versão não-paramétrica para grandes amostras do teste de Williams para comparar doses crescentes de uma substância com a dose zero (testemunha ou controle) aplicável ao delineamento em blocos casualizados. Supondo k tratamentos e o controle (totalizando $k + 1$ tratamentos), a estatística de teste utilizada é $M = \max_{1 \leq u \leq k} \frac{\sum_{j=u}^k \bar{R}_j}{k - u + 1} - \bar{R}_0 = \max_{1 \leq u \leq k} \sum_{j=u}^k \frac{\bar{R}_j - \bar{R}_0}{k - u + 1}$ que, multiplicada pela constante $\left\{ \left[\frac{(k+1)(k+2)}{12} \right] \frac{2}{n} \right\}^{-1/2}$, tem distribuição t de Student com infinitos graus de liberdade. A modificação sugerida por Williams (1986) ao teste proposto por Shirley (1977), ou seja, excluir todas as observações das doses para as quais a significância foi estabelecida e atribuir postos às observações restantes, pode ser utilizada neste caso também.

Shirley (1987) apresenta testes não-paramétricos para experimentos fatoriais em delineamentos inteiramente casualizados, propondo a utilização de reordenamento dos dados dentro de subconjuntos, podendo tratar cada nível do fator como um experimento independente. Experimentos multifatoriais também são tratados com as atribuições de postos sendo feitas às observações dentro dos subgrupos definidos por cada combinação de fatores cujos efeitos não estariam sendo investigados.

2.7 Comparações Múltiplas

2.7.1 Introdução

Em geral, estamos interessados não apenas em saber se existem diferenças entre tratamentos mas, principalmente, saber, dentre os tratamentos testados, qual ou quais aqueles que são mais interessantes para o nosso propósito. Os testes citados anteriormente servem apenas para testar a hipótese de não diferença entre as médias dos tratamentos, ou seja, ao rejeitar a hipótese nula, sabe-se que existe pelo menos um contraste entre médias que difere de zero, mas a forma como isto

ocorre não é dada. No campo paramétrico, existem vários testes para comparações múltiplas envolvendo pares de médias de tratamentos ou ainda envolvendo contrastes de médias de tratamentos. Dentre eles, são bastante utilizados o Teste de Tukey para pares de médias, o teste F para contrastes entre médias, o Teste de Duncan e o de Scheffé, dentre outros. Como são baseados nas pressuposições básicas da análise de variância, estes testes falham quando as mesmas são violadas. Quando estas violações são pequenas, estes testes podem ser usados com razoável precisão.

Entretanto, quando violações sérias nas pressuposições do modelo ocorrem ou ainda, quando os dados não são adequados para que seja realizada uma análise paramétrica, devemos recorrer a outros métodos. A adoção de modelos não-paramétricos é uma boa alternativa já que suas pressuposições são mais fracas que as dos paramétricos e são mais práticos para serem efetuados, especialmente se dispomos de ferramentas computacionais.

Os processos não-paramétricos empregados nas comparações múltiplas são geralmente menos eficientes que seus concorrentes do campo paramétrico, em especial quando os dados provêm de distribuições normais. Além disso, as tabelas, na maioria dos casos, são deficientes devida sua complexidade de estruturação. Apesar disto, muitas vezes os testes de comparações múltiplas não-paramétricos podem se apresentar como única alternativa na busca de um resultado fidedigno, em especial quando as violações ao modelo são severas (falta de normalidade e de homogeneidade de variâncias, principalmente). Assim, o emprego de comparações múltiplas não-paramétricas pode ser encarado como uma complementação aos testes apresentados anteriormente. Dentre os testes não-paramétricos mais utilizados para verificação de diferença entre médias em delineamentos inteiramente casualizados estão aqueles baseados no teste de Kruskal-Wallis, comparações com um controle, comparações pareadas baseadas no uso de atribuição de postos separadamente (Steel, 1960), comparações pareadas baseadas na atribuição de postos de forma conjunta (Dunn, 1964) e procedimentos *stepwise* (Campbell & Skillings, 1985).

A dificuldade de se estabelecer quando utilizar métodos de comparações múltiplas e qual o método adequado para uma situação específica, é mostrada, por exemplo, em Santos *et alli* (1997) onde foi feita uma avaliação do emprego dos testes de comparações múltiplas de médias em artigos publicadas na revista Pesquisa Agropecuária Brasileira (*PAB*) de 1980 a 1994. Constatou-se que, em 57% dos trabalhos levantados, houve aplicação inadequada do método, em cerca de 11,5%, aplicação parcialmente adequada e nos 37,5% restantes, a aplicação foi considerada adequada pelos autores. Considerou-se o procedimento inadequado quando os tratamentos eram de natureza quantitativa, e parcialmente adequados quando os tratamentos eram relacionados, combinações fatoriais ou ainda quando foram aplicados, simultaneamente, os métodos de comparações múltiplas e regressão a tratamentos quantitativos. Não houve a preocupação de verificar se os testes utilizados eram adequados do ponto de vista de normalidade dos dados, homogeneidade de variâncias, etc.

Outro fator a ser considerado nas comparações múltiplas é o tipo de erro utilizado. Vários procedimentos foram desenvolvidos e têm sido objeto de pesquisas devido especialmente, entre outros motivos, à divergência de opinião (filosófica) sobre como os níveis dos erros devem ser controlados, o que afeta o poder comparativo dos procedimentos. De acordo com Steel (1961a) existem várias definições de taxas de erros e a decisão sobre qual delas é a mais apropriada e qual nível de significância deve ser usado para um determinado experimento deve ser feita pelo experimentador. Taxas de erro são definidas baseadas no fato da hipótese nula ser verdadeira.

1. *Per-comparison error rate* ou *comparisonwise error rate* é definida como o número de comparações falsamente declaradas significativas dividida pelo número total de comparações efetuadas. Aqui a comparação é a unidade conceitual. Quando comparações são escolhidas sem referência aos resultados experimentais, um teste dá a taxa de erro *comparisonwise* como o valor tabulado não importante se as comparações são independentes ou não. Quando todas as possíveis comparações são feitas, o uso de testes baseados nesta taxa de erro

(*comparisonwise error rate*) quase certamente leva a falsas declarações de significância, especialmente quando o experimento inclui muitos tratamentos e se os níveis de significância usuais são utilizados. Ou seja, se muitas comparações são feitas, a probabilidade de encontrar ao menos uma diferença significativa falsa é alta.

2. *Experimentwise error rate* ou *experimentalwise error rate* definida como o número de experimentos com ao menos uma diferença falsamente declarada significativa dividido pelo número Total de Experimentos. Neste caso, o experimento é a unidade conceitual. Para comparações pareadas, testa-se a maior diferença quando ela ocorre. Se esta diferença é significativa, então este experimento é um para o qual ao menos uma diferença foi declarada significante. Esta taxa de erro é encontrada nos testes de Tukey (baseado na amplitude estudantizada) e de Dunnett (teste de tratamentos contra o controle). A crítica a estes testes é que são conservativos, ou seja, requerem grandes diferenças entre tratamentos para declará-los significativamente diferentes e esta diferença cresce com o número de tratamentos.

As taxas de erros aqui apresentadas não são tão diferentes como podem parecer à primeira vista. Se α_C e α_E representam, respectivamente, as taxas de erros *comparisonwise* e *experimentwise*, então para k comparações independentes temos $(1 - \alpha_E) = (1 - \alpha_C)^k$.

Como em Hsu (1996), pode-se considerar quatro tipos de comparações múltiplas, de acordo com o tipo de inferência para o qual cada um deles é delineado:

1. Comparações de todos os contrastes (*All-Contrasts Comparisons*): os parâmetros de interesse associados (sob os quais obtém-se inferência direta), são todos os contrastes do tipo $\sum_{i=1}^k c_i \cdot \mu_i$, onde $\sum_{i=1}^k c_i = 0$;
2. Comparações pareadas (*All-pairwise Comparisons*): neste caso, os parâmetros

de interesse são as $\frac{k(k-1)}{2}$ diferenças pareadas das médias de tratamentos, ou seja, $\mu_i - \mu_j$, para todo $i < j$;

3. Comparações Múltiplas com o Melhor (*Multiple Comparisons with the Best*): aqui os parâmetros de interesse são as diferenças entre cada tratamento e o melhor dentre os outros tratamentos. Tem-se então dois casos a considerar

(a) supondo que o maior efeito de tratamento implica no melhor tratamento, então os parâmetros de interesse são $\mu_i - \max_{j \neq i} \mu_j$ para $i = 1, \dots, k$. Assim, quando $\mu_i - \max_{j \neq i} \mu_j < 0$, então o tratamento i não é o melhor, pois há outro tratamento melhor que ele; se $\mu_i - \max_{j \neq i} \mu_j > 0$ o tratamento i é o melhor;

(b) supondo que o menor efeito de tratamento implica no melhor tratamento, então os parâmetros de interesse são $\mu_i - \min_{j \neq i} \mu_j$, para $i = 1, \dots, k$.

4. Comparações Múltiplas com o Controle (*Multiple Comparisons with the Control*): supondo que o tratamento k seja um controle, então os parâmetros de interesse são $\mu_i - \mu_k$ para $i = 1, \dots, k - 1$.

Outras situações podem ser consideradas, como por exemplo, os parâmetros de interesse primário como sendo a diferença entre a média de cada tratamento e a média de todos os tratamentos (*multiple comparisons with the mean*), dentre outros.

Além desta divisão entre os procedimentos de comparações múltiplas, são encontrados na literatura modos diferentes de se proceder: atribuição de postos de forma conjunta a todos os dados (*joint rank procedure*, atribuição de postos apenas aos tratamentos de interesse na comparação especificada (*pairwise procedure*) e procedimentos do tipo passo a passo (*stepwise procedure*).

Dentre os livros que tratam de comparações múltiplas destacam-se Miller (1981), Hochberg & Tamhane (1987), Hoppe (1993) e Hsu (1996).

2.7.2 Construção teórica do Método

De acordo com Manoukian (1986) considere o modelo

$$X_{ij} = \mu + \tau_i + \epsilon_{ij}$$

e a estatística de Kruskal-Wallis,

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k n_i (\bar{R}_{i.} - \bar{R}_{..})^2 = \frac{12}{N(N+1)} \sum_{i=1}^k \left(\frac{R_{i.}^2}{n_i} \right) - 3(N+1)$$

onde $R_{i.} = \sum_{j=1}^{n_i} R_{ij}$, $\bar{R}_{i.} = \frac{\sum_{j=1}^{n_i} R_{ij}}{n_i}$ e $\bar{R}_{..} = \frac{\sum_{i=1}^k R_{i.}}{N} = \frac{N+1}{2}$.

A hipótese $H_0 : \tau_1 = \tau_2 = \dots = \tau_k$ (contra H_a : nem todos os τ_i 's são iguais) é rejeitada, com um nível de significância α se $H \geq c_N(\alpha)$ onde

$$P_{H_0} [H \geq c_N(\alpha)] = \alpha$$

Mas,

$$H \geq c_N(\alpha) \quad \Rightarrow \quad \frac{12}{N(N+1)} \sum_{i=1}^k n_i (\bar{R}_{i.} - \bar{R})^2 \geq c_N(\alpha)$$

ou ainda

$$\sum_{i=1}^k n_i (\bar{R}_{i.} - \bar{R})^2 \geq \frac{N(N+1)}{12} c_N(\alpha)$$

Sejam $i \neq j$ fixados em $\{1, 2, \dots, k\}$ e $a_i = \sqrt{n_i}$, $a_j = -\sqrt{n_j}$. Se $u = (x_1, x_2)$ e $v = (y_1, y_2)$ com produto interno dado por $\langle u, v \rangle = x_1 y_1 + x_2 y_2$ e considerando $x_1 = \frac{1}{\sqrt{n_i}}$, $x_2 = \frac{1}{\sqrt{n_j}}$, $y_1 = \sqrt{n_i} (\bar{R}_{i.} - \bar{R})$ e $y_2 = \sqrt{n_j} (\bar{R}_{j.} - \bar{R})$. Então, aplicando a desigualdade de Cauchy-Schwarz, ou seja, $|\langle u, v \rangle| \leq \|u\| \cdot \|v\|$, onde

$$\|u\| = \sqrt{x_1^2 + x_2^2} = \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

e

$$\|v\| = \sqrt{y_1^2 + y_2^2} = \sqrt{n_i (\bar{R}_{i.} - \bar{R})^2 + n_j (\bar{R}_{j.} - \bar{R})^2}$$

obtém-se:

$$|\bar{R}_i - \bar{R}_j| = \left| \sum_{\ell=i,j} \frac{a_\ell (\bar{R}_\ell - \bar{R})}{\sqrt{n_\ell}} \right| \leq \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \sqrt{\sum_{\ell=i,j} n_\ell (\bar{R}_\ell - \bar{R})^2}$$

ou

$$P \left[|\bar{R}_i - \bar{R}_j| \leq \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \sqrt{\frac{N(N+1)}{12}} (c_N^{1/2}) \right] \geq P [H \leq c_N(\alpha)] \geq 1 - \alpha$$

que são as $\frac{k(k-1)}{2}$ desigualdades

$$|\bar{R}_i - \bar{R}_j| \geq \sqrt{\frac{N(N+1)}{12}} \left(\frac{1}{n_i} + \frac{1}{n_j} \right) (c_N^{1/2}(\alpha))$$

ocorrendo simultaneamente com uma probabilidade não menor que $1 - \alpha$. Se para qualquer par i, j ($i \neq j$), com valores em $\{1, 2, \dots, k\}$, $|\bar{R}_i - \bar{R}_j|$ não satisfaz a última desigualdade, então infere-se que $\tau_i \neq \tau_j$.

Verifica-se que:

$$P \left[|\bar{R}_i - \bar{R}_j| \leq c \text{ para todo } i, j \right] = P \left[\max_{i,j} |\bar{R}_i - \bar{R}_j| \leq c \right]$$

Para o caso em que $N \rightarrow \infty$, $\sqrt{\frac{12}{k(N+1)}} \max_{i,j} |\bar{R}_i - \bar{R}_j|$ tem a mesma distribuição que $\max_{i,j} |Y_i - Y_j| = Y_{(k)} - Y_{(1)}$ que é a amplitude de variáveis normais padrões independentes Y_1, \dots, Y_k .

A mesma análise pode ser aplicada ao teste de Friedman, fazendo-se as substituições adequadas. Assim, para este teste tem-se:

$$|\bar{R}_i - \bar{R}_j| \leq \sqrt{\frac{k(k+1)}{6b}} s_b^{1/2}$$

com uma probabilidade não menor que $1 - \alpha$, onde $P_{H_0} [S \geq s_b] = \alpha$. A distribuição assintótica de

$$\sqrt{\frac{12b}{k(k+1)}} \max_{i,j} |\bar{R}_i - \bar{R}_j|$$

para $b \rightarrow \infty$ coincide com a amplitude $Y_{(k)} - Y_{(1)}$ de variáveis normais padrão independentes Y_1, \dots, Y_k .

2.7.3 Alguns Testes de Comparações Múltiplas Não-Paramétricos aplicáveis aos Delineamentos Inteiramente Casualizados

Steel (1960) apresenta um teste de comparações múltiplas para a comparação simultânea de todos os pares de tratamentos num delineamento inteiramente casualizado balanceado. Para efetuar este teste deve-se atribuir os postos de forma conjunta aos tratamentos pareados (i e j , $1 \leq i \neq j \leq k$), assinalando posto 1 à menor observação e posto $2n$ à maior. Obtém-se, então, os valores R_{ij} , $i < j$ e calcula-se os conjugados $R'_{ij} = (2n + 1)n - R_{ij}$. O mínimo entre os valores de R_{ij} e R'_{ij} é usado como estatística do teste. Para a construção das tabelas o autor obteve a distribuição exata do mínimo de R_{ij} . É possível também a comparação de contrastes de médias fazendo a atribuição de postos apenas àqueles tratamentos que estão envolvidos no contraste a ser testado.

Dunn (1964) apresenta testes de comparações múltiplas utilizando atribuição de postos conjunta como uma alternativa à atribuição de postos pareados apresentada em Steel (1960). O procedimento proposto utiliza a idéia de postos conjuntos (RT-1) e, a partir desta atribuição, calcular os contrastes de médias dos postos de interesse e utilizar a aproximação normal adequada. Enfatiza-se ainda que, ambos os procedimentos são conservativos e comparando-os, conclui-se que o procedimento de Steel é mais adequado para pequenas amostras, devido à possibilidade de obtenção de tabelas, e também nos casos em que os tratamentos têm postos médios muito próximos; já o procedimento de postos conjuntos é considerado mais adequado para grandes amostras e nos casos em que as somas dos postos dos tratamentos são razoavelmente distanciadas. Sherman (1965) faz a comparação dos procedimentos de Steel (1960) e Dunn (1964) do ponto de vista das propriedades assintóticas, mostrando que ambos têm a mesma eficiência relativa assintótica de Pittman em relação aos testes clássicos paramétricos.

Odeh (1967) apresenta um teste referente à soma máxima dos postos, que constitui do seguinte: são retiradas amostras aleatórias de tamanho n de cada

uma de k populações e denotamos por X_{ij} a j -ésima observação da amostra da i -ésima população ($i = 1, \dots, k; j = 1, \dots, n$). Considerando r_{iu} o posto de X_{iu} na amostra combinada de tamanho kn , $R_i = \sum_{j=1}^n r_{ij}$, $R_M = \max_{1 \leq i \leq k} R_i$ e $R_m = \min_{1 \leq i \leq k} R_i$, rejeita-se a hipótese nula de que as k populações são idênticas se R_M excede um valor crítico λ tabelado (referente à soma máxima) ou ainda se R_m é menor que um valor crítico λ' (referente à soma mínima). Para a obtenção das tabelas da distribuição da estatística de teste e dos valores críticos, considera-se a Função distribuição exata de R_M dada por $P(R_M \leq \lambda) = P\left[\bigcap_{i=1}^k (R_i \leq \lambda)\right] = N(n_1, \dots, n_k | \lambda, \dots, \lambda) \frac{n_1! \dots n_k!}{N!}$ onde $N(n_1, \dots, n_k | \lambda, \dots, \lambda)$ é o número de seqüências para as quais ocorrem as desigualdades $R_1 \leq \lambda, \dots, R_k \leq \lambda$ e $N = \sum_{i=1}^k n_i$. Se $n_i = n$, para todo i , tem-se $P(R_M \leq \lambda) = N(n_1, \dots, n_k | \lambda, \dots, \lambda) \frac{(n!)^k}{N!}$ e ainda, da simetria tem-se, para o caso de amostras de tamanhos iguais, que $P[R_m \leq \lambda'] = P[R_M \leq \lambda]$ onde $\lambda + \lambda' = n^2k + n$.

Tobach *et alli* (1967) apresentam uma tabela para a realização de testes de comparações múltiplas através da estatística G para tratamentos com o mesmo número de repetições. A estatística G é definida como sendo a diferença absoluta entre as somas dos postos dos tratamentos a serem testados, dada a atribuição de postos feitas de forma conjunta a todos as r parcelas dos k tratamentos (método RT-1 de atribuição de postos). A tabela obtida fornece valores exatos do lado direito da função distribuição acumulada de G para um pequeno número de tratamentos e repetições sob a hipótese nula de que as $kr = N$ observações provêm de uma amostra aleatória da mesma população. Observa-se que G pode variar de 0 a $\sum_{i=0}^{r-1} (kr - i) - \sum_{i=1}^r i = (k-1)r^2$ e, considerando $\#(G \geq g) = \sum_{i=g}^{(k-1)r^2} \#(G = i)$ como o número de partições favoráveis ao evento $[G \geq g]$, então $P[G \geq g] = \frac{\#G \geq g}{\frac{1}{2} \binom{kr}{k} \binom{kr-r}{r}}$.

A diferença entre as somas dos postos quando r tende ao infinito é, de acordo com

Dunn (1964) assintoticamente normal com média 0 e variância $\sigma^2 = \frac{kr^2(kr+1)}{6}$. Consequentemente tem-se que $P[G \geq g] = 2P\left[Z \geq \frac{g}{\sigma}\right]$ onde $Z \sim N(0, 1)$.

Fairley & Pearl (1984) observam que os dois tipos mais comumente utilizados de comparações de dois tratamentos na Análise de Variância de delineamentos inteiramente casualizados com k populações são o procedimento de Kruskal-Wallis (ou *K-tipo*), baseado nas diferenças pareadas das somas dos postos, quando são atribuídos postos a todas as observações de todas as k populações conjuntamente (Nemenyi, 1963) e o procedimento *W-tipo*, baseado na estatística da soma dos postos de Wilcoxon, calculada para todos os pares de tratamentos (Steel, 1960). Estes autores investigaram a performance dos procedimentos de comparações com atribuição de postos pareados e o de atribuição conjunta, utilizando a eficiência de Bahadur aproximada com a finalidade de comparar pares de populações particulares, quando as alternativas eram distantes da hipótese nula. Concluíram que quando as populações a serem comparadas são bastante separadas, ou ainda, quando os dados provêm de um modelo de locação (*shift model*) com uma densidade unimodal subjacente, o procedimento de atribuição de postos pareados é melhor para se comparar duas populações adjacentes, enquanto que o procedimento de atribuição de postos conjunta é mais eficiente para se comparar duas populações mais distantes.

Fligner (1985) defende a utilização de testes pareados como substituto do teste de Kruskal-Wallis, pois, com um único teste verificamos se há diferenças entre tratamentos e onde estas estão localizadas. Além disso, num teste deste tipo não ocorrem paradoxos possíveis quando utiliza-se o teste de Kruskal-Wallis seguido de um teste de comparações múltipla, ou seja, um dos testes detectar diferenças e outro não. A estatística para utilização de testes pareados para k amostras é construída pela combinação de todas as estatísticas para duas amostras de Wilcoxon. O autor mostra que esta estatística é análoga de Kruskal-Wallis, tendo ambas eficiência relativa assintótica de Pittman entre um e dois. Porém, a eficiência aproximada de Bahadur

desta nova estatística em relação à estatística de Kruskal-Wallis é maior ou igual a um para todas as alternativas. O autor sugere ainda que tal estatística pode ser mais apropriada do que a generalização da estatística de Kruskal-Wallis para outros problemas envolvendo k amostras devido a sua versatilidade. Escrevendo a estatística de Kruskal-Wallis na forma

$$V_N = \frac{12}{N+1} \sum_{i < j} n_i n_j V_{ij}^2 \quad \text{onde} \quad V_{ij} = \frac{(\bar{R}_i - \bar{R}_j)}{N} \quad \text{e} \quad N = \sum_{i=1}^k n_i$$

e considerando W_{ij}^* a estatística de Wilcoxon calculada para a i -ésima e a j -ésima amostras, um valor análogo a V_N para postos pareados é dado por

$$W_N = \frac{12}{N+1} \sum_{i < j} n_i n_j W_{ij}^2 \quad \text{onde} \quad W_{ij} = \frac{W_{ij}^* - n_i(n_i + n_j + 1)}{n_i n_j}$$

que é a estatística considerada neste teste, lembrando que os vetores

$$V = (V_{12}, V_{13}, \dots, V_{1k}, V_{23}, \dots, V_{2k}, \dots, V_{(k-1)k})$$

$$W = (W_{12}, W_{13}, \dots, W_{1k}, W_{23}, \dots, W_{2k}, \dots, W_{(k-1)k})$$

padronizados de forma adequada, têm a mesma distribuição normal assintótica sob H_0 . Apesar de seu comportamento assintótico adequado, tal estatística não tem sido utilizada na prática nem havendo tabelas específicas para ela.

Damico & Wolfe (1987) e Damico & Wolfe (1989) apresentam a distribuição nula exata das estatísticas de postos utilizada para a realização de comparações múltiplas de todos os tratamentos e entre tratamentos e o controle, respectivamente, nos delineamentos inteiramente casualizados. Busca-se fazer inferências estatísticas sobre os efeitos dos parâmetros populacionais τ_1, \dots, τ_k baseadas nas k amostras aleatórias independentes destas populações nas situações onde o teste de Kruskal-Wallis levou à rejeição de H_0 , com o objetivo de descobrir quais são as razões desta rejeição. Um procedimento de comparações múltiplas baseado nos postos conjuntos para tomar decisões sobre todos os pares de efeitos de tratamentos é descrito a seguir para o caso geral, ou seja, com populações de tamanhos desiguais. Definindo:

$$R^* = N^* \max \left[\left| \bar{R}_u - \bar{R}_v \right|, u = 1, \dots, k-1; v = u+1, \dots, k \right]$$

onde N^* = mínimo múltiplo comum de $\{n_1, \dots, n_k\}$ e $\bar{R}_i = \frac{1}{n_j} \sum_{j=1}^{n_i} R_{ij}$ para $j = 1, \dots, k$ são os postos médios amostrais usados na estatística de Kruskal-Wallis. O procedimento de comparações múltiplas de todos os tratamentos associados é (para todo para $u \neq v = 1, \dots, k$): $\tau_u \neq \tau_v$ se e só se $N^* |\bar{R}_u - \bar{R}_v| < y(\alpha, k, n_1, \dots, n_k)$, onde $y(\alpha, k, n_1, \dots, n_k)$ é escolhido de forma a satisfazer

$$\begin{aligned} 1 - \alpha &= P_0(\text{decide-se que } \tau_u = \tau_v \text{ para todo par } u \neq v = 1, \dots, k) \\ &= P_0(N^* |\bar{R}_u - \bar{R}_v| < y(\alpha, k, n_1, \dots, n_k) \text{ para todo par } u \neq v = 1, \dots, k) \\ 1 - \alpha &= P_0(R^* < y(\alpha, k, n_1, \dots, n_k)) \end{aligned}$$

com P_0 indicando que a probabilidade é calculada sob H_0 , ou seja, sob a hipótese de que todos os efeitos de tratamentos são iguais. Portanto, α é a taxa de erro *experimentalwise* e $1 - \alpha$ é a probabilidade de tomar todas as $[k(k - 1)]/2$ decisões de igualdade corretas quando H_0 é verdadeira. Os valores críticos $y(\alpha, k, n_1, \dots, n_k)$ foram obtidos em McDonald & Thompson (1967) somente para o caso especial de tamanhos de amostras iguais ($n_1 = n_2 = \dots = n_k$) e algumas combinações de n e k . Damico & Wolfe (1987) obtêm tabelas para tamanhos de amostras diferentes, incluindo nas tabelas apresentadas, para todas as combinações de k, n_1, \dots, n_k consideradas ($k = 3$ e $1 \leq n_i \leq 6$; $k = 4$ e $\sum_i n_i \leq 16$) e algumas probabilidades da cauda superior para R^* de interesse. As tabelas obtidas servem de complemento ao Teste de Kruskal-Wallis podendo, entretanto, serem utilizadas no procedimento de comparações múltiplas de todos os tratamentos omitindo o teste de Kruskal-Wallis como descrito por Skillings (1983) e Fairley & Pearl (1984) ou ainda serem utilizadas na implementação dos procedimentos exatos do método proposto por Campbell & Skillings (1985), que apresenta uma aproximação *stepwise* para a condução de procedimentos em múltiplas comparações para delineamentos inteiramente casualizados.

Para a obtenção das tabelas para comparações múltiplas entre os tratamentos e o controle, considera-se, sem perda de generalidade, que o primeiro tratamento é o controle. Assim, X_{i1} , $i = 1, \dots, n_1$ é a i -ésima observação da população

controle e X_{ij} , $i = 1, \dots, n_j$ e $j = 2, \dots, k$ é a i -ésima observação do j -ésimo tratamento (temos aqui o controle e $k - 1$ tratamentos). A estatística de teste utilizada neste caso é

$$R^{**} = N^* \max \left| \bar{R}_u - \bar{R}_1 \right|, u = 2, \dots, k$$

Existem $(k - 1)$ comparações a serem realizadas e testadas. Decide-se que $\tau_u > \tau_1$ se e só se $N^* \left| \bar{R}_u - \bar{R}_1 \right| \geq y^*(\alpha, k - 1, n_1, \dots, n_k)$, onde $y(\alpha, k, n_1, \dots, n_k)$ é escolhido de forma a satisfazer

$$\begin{aligned} 1 - \alpha &= P_0(\tau_u = \tau_1 \quad \text{para todo } u = 2, \dots, k) \\ &= P_0(N^* \left| \bar{R}_u - \bar{R}_1 \right| < y^*(\alpha, k - 1, n_1, \dots, n_k) \quad \text{para todo } u = 2, \dots, k) \\ 1 - \alpha &= P_0(R_{**} < y^*(\alpha, k - 1, n_1, \dots, n_k)) \end{aligned}$$

com P_0 indicando que a probabilidade é calculada sob a hipótese nula (H_0) de que todos os efeitos de tratamentos e o do controle são iguais. Portanto $(1 - \alpha)$ é a probabilidade de tomar todas as $(k - 1)$ decisões de igualdade corretas quando H_0 é verdadeira. Como no caso anterior, os valores críticos $y(\alpha, k, n_1, \dots, n_k)$ foram previamente calculados somente para o caso especial de tamanhos de amostras iguais e algumas combinações de n e k . Neste artigo os autores estendem essas tabelas para tamanhos de amostras diferentes para probabilidades da cauda superior para R^{**} , para valores adequados de α , considerando $k = 3$ e $n_1 = 1(1)6$ e $1 \leq n_2 \leq n_3 \leq 6$; $k = 4$ e $n_1 = 1(1)61 \leq n_2 \leq n_3 \leq 6$.

Critchlow & Fligner (1991) descrevem dois métodos para se efetuar comparações múltiplas com distribuição livre para k ($k > 2$) populações. O primeiro método, baseado na atribuição conjunta de postos a todas as $N = \sum_{i=1}^k n_i$ observações foi inicialmente utilizada por Nemenyi em 1963 e depois por Dunn (1964) em conjunção com a desigualdade de Bonferroni. Sejam \bar{R}_i e \bar{R}_j as médias dos postos assinalados à i -ésima e j -ésima amostras nesta ordenação conjunta. Dois tratamentos são então declarados diferentes se $|\bar{R}_i - \bar{R}_j|$ excede um valor crítico apropriado.

A segunda aproximação, utilizando o método de Wilcoxon para postos pareados, foi proposta independentemente por Steel e Dwass em 1960 e depois inves-

tigado por Steel (1961). Seja S_{ij} a soma dos postos assinalados à i -ésima amostra na atribuição e postos conjunto das $n_i + n_j$ observações das i -ésima e j -ésima amostras. Os tratamentos i e j são declarados diferentes por este procedimento se S_{ij} cai numa região crítica apropriada.

Considere quatro propriedades desejáveis para os testes de comparações múltiplas:

- A) Controle da taxa de erro *experimentwise*: a probabilidade de declarar erroneamente que qualquer par de tratamentos diferem, quando de fato todos os tratamentos são iguais, deve ser menor ou igual a um valor pré definido α ;
- B) Controle da taxa de erro tipo I máxima: a probabilidade de declarar que algum par de tratamentos difere, quando isto é falso, deve ser menor ou igual a α . Esta é uma condição mais restritiva do que (A), desde que requer o controle da probabilidade de fazer qualquer declaração falsa sobre diferenças, mesmo em situações onde nem todos os tratamentos são idênticos;
- C) Intervalos de Confiança Simultâneos para as diferenças entre tratamentos: o procedimento de comparações múltiplas deve revelar não somente quais pares de tratamentos diferem mas também produzir intervalos de confiança simultâneos para os tamanhos destas diferenças;
- D) Métodos para a determinação do ponto crítico: estes incluem tabelas exatas para pequenas amostras, aproximações para grandes amostras e simulação das probabilidades dos limites superiores. Tais métodos são necessários para implementar o procedimento de fato; se eles não forem possíveis, um procedimento não será utilizado, apesar de suas outras propriedades apropriadas.

A aproximação de postos conjuntos satisfaz a propriedade (A), mas não satisfaz (B). Das duas primeiras propriedades, (B) é a mais desejável já que inclui a propriedade (A) e também controla as probabilidades de tipos de erros adicionais.

Por exemplo, tendo a configuração de tratamentos dada por $\tau_1 = \tau_2 < \tau_3$, a probabilidade de declarar erroneamente que $\tau_1 \neq \tau_2$ é necessariamente controlada por um procedimento de comparações múltiplas satisfazendo (B) mas nem sempre será por um procedimento que satisfaz (A) somente.

Com relação à propriedade (C) para a aproximação de postos conjuntos, Miller (1981) observa que é virtualmente impossível realizar os cálculos necessários para obter intervalos de confiança simultâneos. Esta é talvez a maior desvantagem da aproximação de postos conjuntos. Tabelas para amostras pequenas para $k = 3$ e $k = 4$ foram obtidas por Damico & Wolfe (1987) mas a padronização para amostras desiguais diferem do procedimento conservativo de Dunn (1964). Para os casos em que os tamanhos de amostras são desiguais e escapam do escopo das tabelas, podem ser feitas aproximações utilizando permutações aleatórias já que não há estudos sobre distribuições assintóticas. Finalmente, a aproximação de postos conjuntos tem a característica indesejável de que a comparação entre duas populações quaisquer é afetada pelas observações das populações restantes.

A aproximação de postos pareados satisfaz as propriedades (A), (B) e (C), sendo recomendada por Miller (1981). Apesar das óbvias vantagens desta aproximação, muitos livros de estatística não-paramétrica descrevem somente a aproximação de postos conjuntos. Isto ocorre talvez devido à falha da aproximação de postos pareados em satisfazer (D), ou seja, na obtenção de tabelas exatas para pequenas amostras e aproximações confiáveis. Critchlow & Fligner (1991) tentam sanar parte deste problema, apresentando uma tabela para três tratamentos e até sete repetições.

2.7.4 Alguns Testes de Comparações Múltiplas não-paramétricos aplicáveis aos Delineamentos em Blocos Casualizados

Steel (1959) apresenta um teste dos sinais visando as comparações entre tratamentos e o controle. Considerando o controle como o tratamento 1 e dado

X_{ij} , $i = 1, \dots, k$ e $j = 1, \dots, b$ a medida resposta no i -ésimo tratamento no j -ésimo bloco, o teste proposto considera o número de sinais positivos e negativos em cada um dos $k - 1$ conjuntos das b diferenças entre controle e tratamento. Para realizar o teste, considerando-se a hipótese alternativa de que os tratamentos dão respostas significativamente maiores que o controle, calcula-se as diferenças $X_{ij} - X_{1j}$, $i = 2, \dots, k$ e $j = 1, \dots, b$ e verifica-se o número de sinais negativos para cada um dos $k - 1$ conjuntos de b sinais, anotando-os r_i , $i = 1, \dots, k - 1$. Para julgar se há ou não significância, compara-se cada r_i com o valor crítico tabulado para o nível de significância conjunto fixado, para cada uma das $k - 1$ comparações, obtendo-se assim uma região crítica unilateral. A aplicação deste método no caso em que os tratamentos dão valores menores que o controle é óbvia. No caso de teste bilateral (mediana diferente de zero), toma-se o valor de r_i como o número de vezes que o sinal menos freqüente ocorre para cada um dos $k - 1$ conjuntos de b sinais. Valores de r_i menores que aqueles tabelados para um nível α fixado são considerados significantes. Como a taxa de erro aqui utilizada é *experimentalwise*, deve-se ter cuidado com o nível de significância escolhido, especialmente em experimentos com grande número de tratamentos. Empates entre tratamentos podem ocorrer na prática mas não foram levados em conta no procedimento, utilizando-se nestes casos um procedimento de atribuição de sinais. Rhyne & Steel (1965) apresentam tabelas mais completas para este teste.

Rhyne & Steel (1967) apresentam um teste do sinal para comparações entre todos os tratamentos. Considerando, como em Steel (1959), X_{ij} a medida resposta no i -ésimo tratamento no j -ésimo bloco, tem-se $c = \binom{k}{2}$ comparações. Considere os c conjuntos de b sinais relativos às diferenças $(X_{ij} - X_{i'j})$, $i, i' = 1, \dots, k$ para $i \neq i'$ e $j = 1, \dots, b$, o teste do sinal proposto considera o número de diferenças positivas e/ou negativas em cada um dos c conjuntos de b sinais de diferenças. Denotando o número de diferenças negativas nos c conjuntos por (r_1, \dots, r_c) , a distribuição do $\min(r_1, \dots, r_c)$ dá os valores críticos para testar simultaneamente todos os pares

de tratamentos contra alternativas unilaterais usando uma taxa de erro adequada. A distribuição do $\min_i[\min(r_i, n - r_i)]$ dá os valores críticos para a alternativa bilateral. Cada r_i na estatística de teste é comparada com o valor crítico através da tabela apropriada para a alternativa desejada, levando-se em conta cada uma das c comparações. O nível de significância se aplica ao conjunto das declarações de significância. Para obtenção das tabelas, verifica-se que há $s = k!$ permutações de k observações num bloco, cada uma produzindo um vetor de 0 's e 1 's que, sob a hipótese nula, são igualmente prováveis; assim, a probabilidade de que um dado vetor ocorra num simples ensaio é $1/k!$. Sejam os vetores v_1, v_2, \dots, v_s dos ensaios simples e suas probabilidades p_1, p_2, \dots, p_s onde $p_i = \frac{1}{k!}$ para todo i . A soma dos coeficientes de um ou mais termos apropriados na expansão da equação $(p_1 y_1 + p_2 y_2 + \dots + p_s y_s)^b = \frac{1}{k!} (y_1 + \dots + y_s)^b$ dá a probabilidade de obtenção de (r_1, \dots, r_c) em b ensaios. Três métodos são descritos para encontrar tais valores e, como o número de vetores $s = k!$ cresce rapidamente quando k cresce, considera-se uma aproximação normal multivariada para a distribuição conjunta dos (r_1, \dots, r_c) .

Shuster & Boyett (1979) apresentam métodos baseados em testes de permutação válidos para delineamentos gerais aplicados aos postos estratificados. Considerando k tratamentos e b blocos, sejam n_{ij} o número de elementos no bloco j que receberam o tratamento i , $n_{.j} = \sum_{i=1}^k n_{ij}$ o número de elementos no bloco j , y_{ijr} a medida associada com o r -ésimo elemento do bloco j que recebeu o tratamento i , $\bar{y}_{.j} = \frac{1}{n_{.j}} \sum_{i=1}^k \sum_{r=1}^{n_{ij}} y_{ijr}$ a média amostral do bloco j e

$$d_{ij} = \begin{cases} \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijl} & \text{se } n_{ij} > 0. \\ \bar{y}_{.j} & \text{se } n_{ij} = 0 \end{cases}$$

que é a definição da média amostral dos elementos que receberam o tratamento i no bloco j , supondo $n_{ij} > 0$. Considera-se ainda $Q_i = \sum_{j=1}^b n_{.j} d_{ij}$ $i = 1, 2, \dots, k$, a amplitude estratificada $R = \max_{1 \leq i \leq k} Q_i - \min_{1 \leq i \leq k} Q_i$ e a amplitude estratificada controle-tratamento $R_1 = Q_1 - \min_{2 \leq i \leq k} Q_i$, onde Q_1 corresponde ao controle e Q_i , $2 \leq i \leq k$

aos tratamentos. Supondo, sob H_0 , que os tratamentos são todos equivalentes, a distribuição das variáveis aleatórias dentro de um dado bloco é invariante sob as permutações das observações ou mudanças nas atribuições de tratamentos e os b vetores aleatórios dos y 's (um para cada bloco) são mutuamente independentes. Sob H_0 , Q_i é a estimativa clássica estratificada (os blocos são os estratos) do total dos y 's no experimento, baseado somente nos elementos que recebem o tratamento i . Nos blocos onde o tratamento i é ausente, projetamos o total do j -ésimo bloco como $n_{.j}\bar{y}_{.j}$.

Para cada uma das $M = \prod_{j=1}^b \binom{n_{.j}}{n_{1j}, n_{2j}, \dots, n_{kj}}$ simulações independentes escolhidas por amostragem aleatória simples com reposição de todas as possíveis alocações compatíveis com o delineamento, atribui-se aleatoriamente n_{ij} elementos ao tratamento i , independentemente, para cada bloco j e calcula-se a estatística desejada (R ou R_1) para esta alocação de elementos. Denotamos estes valores por r_ℓ ou $r_{1\ell}$ respectivamente, $\ell = 1, 2, \dots, M$. Seja R_c^α o menor número real x tal que a fração dos números R, r_1, \dots, r_M que são menores que x é ao menos $(1 - \alpha)$, onde R é a amplitude observada e r_ℓ é a ℓ -ésima amplitude simulada. Seja também R_{1c}^α definida de modo similar para $R_1, r_{11}, \dots, r_{1M}$. Assim, para comparações múltiplas entre todos os tratamentos, consideramos os tratamentos i e i' significativamente diferentes se e só se $|Q_i - Q_{i'}| \geq R_c^\alpha$. No caso de comparações unilaterais dos tratamentos $2, 3, \dots, k$ com o tratamento 1 (controle), declara-se o tratamento i superior ao controle se e só se $Q_1 - Q_i \geq R_{1c}^\alpha$. Os resultados são imediatos devido ao fato que, sob H_0 os vetores aleatórios $\{R, r_1, \dots, r_M\}$ e $\{R_1, r_{11}, \dots, r_{1M}\}$ são ambos intercambiáveis (com distribuições invariantes sob permutações). As simulações, no lugar da enumeração completa, são necessárias em muitas situações práticas devido ao grande número de configurações. O procedimento proposto é não-paramétrico, podendo ser aplicado a todo delineamento, balanceado ou não com efeitos mixtos e tamanhos de amostras arbitrários. Pode ser aplicado a grandes e pequenas amostras. Além disso, o procedimento é conceitualmente simples e pode ser aplicado a qualquer conjunto

de contrastes de interesse do usuário.

Como complemento ao teste de Quade, Conover (1999) apresenta um teste para a comparação de todos os tratamentos, baseado nas diferenças entre os valores de S_u e S_v , para

$$S_i = \sum_{j=1}^b Q_j \left[R_{ij} - \frac{k+1}{2} \right] \quad \text{onde} \quad Q_j = \max_i \{X_{ij}\} - \min_i \{X_{ij}\}$$

Assim, no caso da estatística de Quade ser rejeitada, considera-se os tratamentos u e v diferentes se a desigualdade

$$|S_u - S_v| > t_{(1-\alpha)/2} \left[\frac{2b}{(b-1)(k-1)} \left(\sum_{i=1}^k \sum_{j=1}^b S_{ij}^2 - \frac{1}{b} \sum_{i=1}^k S_i^2 \right) \right]$$

é satisfeita e onde $t_{(1-\alpha)/2}$ é o valor tabelado t de Student com $(b-1)(k-1)$ graus de liberdade e α é o mesmo nível de significância utilizado no teste de Quade.

Campos (1983) e Hollander & Wolfe (1999) apresentam procedimentos de comparações múltiplas entre todos os tratamentos e entre os tratamentos e o controle a serem utilizados nos casos em que a hipótese nula de igualdade de tratamentos é rejeitada no Teste de Friedman. No caso de comparações entre todos os tratamentos (bilateral), calcula-se as $\frac{k(k-1)}{2}$ diferenças absolutas das somas dos postos de tratamentos $|R_u - R_v|$, $1 \leq u < v \leq k$ e decide-se que $\tau_u \neq \tau_v$ se $|R_u - R_v| \geq r_\alpha$ onde a constante r_α é escolhida de forma que a taxa de erro *experimentalwise* seja igual a α , ou seja, r_α satisfaz a restrição $P_0(|R_u - R_v| < r_\alpha) = 1 - \alpha$, para $u = 1, \dots, k-1$; $v = u+1, \dots, k$. Os valores de r_α foram apresentados por McDonald & Thompson (1967) e posteriormente ampliados, abrangendo valores de $k \leq 15$ e $b \leq 15$. Nos casos em que temos valores de k ou de b não constantes da tabela, pode ser utilizada a aproximação normal utilizando a distribuição da amplitude de k variáveis normais padrões independentes (q_α), ou seja, decide-se que $\tau_u \neq \tau_v$ se $|R_u - R_v| \geq q_\alpha \left[\frac{bk(k+1)}{12} \right]^{1/2}$.

Supondo agora o tratamento 1 como controle, calcula-se as $k-1$ diferenças $(R_u - R_1)$, $u = 2, \dots, k$ e, a uma taxa de erro *experimentalwise* α , decide-se que $\tau_u > \tau_1$ se $(R_u - R_1) \leq r_\alpha^*$, onde r_α^* é tal que satisfaz a restrição

$P_0[(R_u. - R_{1.}) < r_\alpha^*] = 1 - \alpha$, para $u = 2, \dots, k$. Valores de r_α^* foram tabelados e, nos casos em que os valores de k ou b não estão disponíveis em tabelas, pode-se utilizar uma aproximação através da distribuição do máximo de $(k - 1)$ variáveis normais padrões com coeficiente de correlação igual a $1/2$ ($m_{\alpha,1/2}^*$), ou seja, decide-se que $\tau_u > \tau_1$ se $(R_u. - R_{1.}) \geq m_{\alpha,1/2}^* \left[\frac{bk(k+1)}{12} \right]^{1/2}$. Tais procedimentos podem ser facilmente adaptados para o caso em que tem-se valores dos tratamentos menores que o do controle.

2.8 Testes não paramétricos para outros delineamentos experimentais

Silva (1979) faz um estudo do método de análise univariada não-paramétrica, de experimentos em parcelas subdivididas e sub-subdivididas. Recomenda a utilização de testes não paramétricos quando o modelo é drasticamente violado e a análise paramétrica caso as exigências do modelo sejam plenamente satisfeitas. Nos casos intermediários, ou seja, onde o modelo é levemente violado, qualquer das alternativas é válida, sendo preferível a análise não-paramétrica pela sua simplicidade de aplicação. Nas comparações múltiplas em delineamentos inteiramente casualizados ou em blocos ao acaso para experimentos em parcelas subdivididas e sub-subdivididas o autor optou pela aplicação dos testes não-paramétricos citados quando havia mais que dois tratamentos ou ainda os testes de Wilcoxon e das ordens assinaladas, quando havia a presença de apenas dois tratamentos dentro dos efeitos a serem estudados (tratamentos principais, secundários, terciários ou desdobramentos).

Groggel & Skillings (1986) descrevem um teste de distribuição livre baseado nos postos para os efeitos de tratamentos num delineamento com $M \geq 1$ fatores, com fatores cruzados, sem interação e sem células vazias. A estatística de teste é $T = R'\Sigma^{-1}R$ onde $R' = [R_1^* - E_0(R_1^*), \dots, R_{L_1}^* - E_0(R_{L_1}^*)]$ e $\Sigma = (\sigma_{ij})$ para

$$\sigma_{ab} = \begin{cases} \sum_{j=1}^{L_c} \frac{(N_j - n_{ij})(N_j + 1)}{12n_{ij}} & \text{se } a = b \\ \frac{-(N + L_c)}{12} & \text{se } a \neq b \end{cases}$$

que recai na estatística de Kruskal-Wallis se $M = 1$ e na estatística de Friedman se $M = 2$. Esta estatística foi anteriormente proposta por Mack & Skillings (1980) para o caso com apenas dois fatores.

Salter & Fawcett (1993) utilizam simulações para mostrar que as transformações de postos alinhados (*aligned ranks transformation - ART*) podem ser utilizados para o teste de interações em modelos fatoriais, sendo quase tão poderosos quanto o teste F no caso das suposições da análise de variância serem satisfeitas e mais poderosos nos casos em que estas suposições são violadas. O teste de interações utilizado é uma combinação de um teste de postos alinhados modificado e um teste de transformação de postos modificado, ou seja, primeiramente é feito o alinhamento de todas as observações subtraindo-se os parâmetros *nuisance*, em seguida são atribuídos postos às observações alinhadas de forma conjunta e finalmente calcula-se o teste F para os postos alinhados.

Negrillo (1997) apresenta um modelo não-paramétrico para a análise de experimentos em parcelas subdivididas, determinando testes assintóticos para a significância dos efeitos das parcelas, dos efeitos das subparcelas e das interações de parcelas e subparcelas, baseados na distribuição de qui-quadrado. Apresenta ainda teste para comparações múltiplas, tanto para as parcelas como para as subparcelas, baseados na aproximação normal.

2.9 Outras Discussões em Estatística Experimental Não Paramétrica

Conover (1973) desenvolve a teoria de testes de postos, sem qualquer suposição em relação à forma da função de distribuição subjacente aos dados (se discreta ou contínua), apresentando a distribuição assintótica da estatística linear de postos sob a hipótese nula de simetria e aleatorização.

Still & White (1981) discutem a utilização de testes aleatórios aproximados como uma alternativa ao teste F na análise de variância. Discussões relativas ao procedimento utilizado pelos autores bem como erros apontados naquele artigo

são feitas por Bradbury (1987a, 1987b) e debatidos por White & Still (1987) e White (1987a, 1987b). Com relação aos experimentos inteiramente casualizados não foram detectadas discrepâncias significativas entre o poder do teste F e do teste de aleatorização aproximado ou do teste de aleatorização exato para dados provenientes de uma população normal. O teste de aleatorização aproximado não é sensível a p -valores muito pequenos, mas Bradbury (1987a) considera ser este um preço pequeno a ser pago em função da falta de suposições da distribuição do teste de aleatorização aproximado e sua conseqüente robustez. Still & White (1981) concordam entretanto, que tanto o teste de aleatorização aproximado como o teste de aleatorização exato, apesar da esperada perda de poder devido à ausência de suposições da distribuição subjacente, são ferramentas poderosas quando há falha das suposições e também quando temos delineamentos pequenos.

Maritz (1981) indica que, num sentido amplo, um método estatístico de distribuição livre é um método que não necessita, para sua validação ou utilização, de nenhuma pressuposição a respeito da forma da distribuição que é tomada para gerar os valores amostrais nos quais serão baseadas as inferências sobre a distribuição da população. Um método não pode ser utilizável, a menos que ele seja válido, mas o contrário não é verdade. Os termos *validade* e *utilidade* são usados num sentido semi-técnico e relacionados com as noções estatísticas usuais de consistência e eficiência, respectivamente. Os grandes atrativos destes métodos são que eles são válidos sob mínimas suposições referentes às formas das distribuições subjacentes, são baseados em grande parte em permutações simples ou idéias de aleatorização e têm propriedades de eficiência e robustez bastante satisfatórias. Entretanto, os principais impedimentos para a utilização destes métodos são que os resultados dos testes de distribuição livre não são geralmente fáceis de interpretar em termos de quantidades físicas, como o são os resultados da análise paramétrica, e, em algumas situações mais complexas, dificuldades computacionais grandes podem aparecer.

Maritz (1981) ainda destaca que poucos testes denominados de dis-

tribuição livre realmente o são. Muitos dos argumentos são simplificados se a distribuição subjacente pode ser tomada como contínua (e isto é feito em geral). Outras suposições são necessárias, dependendo do problema. Por exemplo, em problemas de locação com uma amostra, a suposição de simetria é importante. Portanto, o termo distribuição livre deve ser interpretado com alguma qualificação, ou seja, os métodos são desenvolvidos sem a especificação paramétrica detalhada das distribuições. O termo não-paramétrico é preferido a distribuição livre por alguns, mas quando se trata de fazer inferências sobre parâmetros, o termo distribuição livre parece ser mais apropriado.

Skillings (1983) discute duas questões de interesse envolvendo comparações múltiplas não-paramétricas. A primeira delas é saber se podemos utilizar um procedimento de comparações múltiplas como um teste de igualdade de k tratamentos e, caso afirmativo, qual procedimento tem melhor performance como um teste. Foram utilizados no trabalho os procedimentos de comparações múltiplas de Nemenyi (tipo Scheffé, que utiliza a estatística de Kruskal-Wallis e é extremamente conservativo) e o teste de amplitude (McDonald & Thompson (1967)) para serem comparados com o teste de Kruskal-Wallis. Os resultados do trabalho destes autores mostram que para valores pequenos de k alguns procedimentos de comparações múltiplas têm boa performance como teste. As vantagens de utilizar um teste de comparações múltiplas como teste é que, num único procedimento, podemos determinar se e de que forma os tratamentos diferem e, além disso, o resultado deste procedimento não é condicionado ao resultado do teste de Kruskal-Wallis. A segunda questão é saber qual dos procedimentos de comparações múltiplas (com atribuição conjunta de postos ou atribuição de postos separadamente) tem melhor performance como um teste e como um dispositivo de separação de tratamentos. O procedimento que utiliza usando atribuição de postos separadamente é baseado na atribuição de postos às observações de apenas dois tratamentos a cada vez. Estes procedimentos foram comparados em termos de suas funções poder assintoticamente (usando eficiência relativa assintótica), e para

grandes amostras (usando um estudo de simulação). O procedimento de postos separados e o de postos conjuntos são assintoticamente equivalentes e, portanto, têm as mesmas eficiências assintóticas e, conseqüentemente, para grandes amostras, os dois procedimentos devem fornecer aproximadamente os mesmos resultados tanto na forma de um teste como um procedimento de separação de tratamentos. Uma comparação assintótica do teste de Kruskal-Wallis com Comparações Múltiplas é mais complicada pois os procedimentos têm distribuições assintóticas diferentes e, por este motivo, foi utilizado um estudo simulado para efetuar a comparação dos poderes. A primeira conclusão foi a de que o teste do tipo Scheffé é muito conservativo e como conseqüência tem menor poder que os outros três procedimentos. O teste de Steel (teste de postos separados) parece ter poder ligeiramente menor que os testes de amplitude e o de Kruskal-Wallis na maioria dos casos. Os poderes empíricos do teste de Kruskal-Wallis e o de amplitude são aproximadamente os mesmos em muitos casos. O teste de Kruskal-Wallis mostrou ser um pouco melhor que o de amplitude para grande valores de k , especialmente quando há vários tratamentos com postos próximos ao maior posto.

Apesar de alguns autores (por exemplo, Miller (1981)) indicarem preferência por postos separados baseados numa razão intuitiva, argumentando contra os postos conjuntos, com base no fato que neste a comparação dos tratamentos 1 e 3 dependem do tratamento 2 e que esta dependência não ocorre no procedimento de postos separados, pode-se, por outro lado, refutar que atribuindo-se postos a todos os tratamentos utilizam-se mais informação e a tendência é a obtenção de respostas mais fidedignas. Em termos de comparações de grandes amostras, o autor verificou que ambos os procedimentos são igualmente poderosos. Portanto, não parece haver um procedimento melhor do que outro para todos os casos em termos de poder nesta comparação.

Supondo que as somas dos postos dos k tratamentos R_1, R_2, \dots, R_k estão relacionadas em ordem crescente, Skillings (1983) relata os seguintes casos es-

peciais: procedimentos de postos conjuntos são superiores quando comparam-se os tratamentos extremos R_1 e R_k ; o procedimento de postos conjuntos é pior, em geral, quando comparam-se R_i e R_j , $i \neq j = 2, \dots, k - 1$; quando somente um tratamento difere dos outros tratamentos, ambos os procedimentos se assemelham, com uma leve vantagem do procedimento de postos separados, se a diferença é grande; o poder do procedimento de postos separados para uma dada composição de tratamentos é o mesmo, não importando os tamanhos dos outros tratamentos, não ocorrendo para o de postos conjuntos; ambos os procedimentos parecem ser bastante conservativos na categoria de comparação de erros. Um complicador relacionado ao teste de postos separados é a indisponibilidade de tabelas para pequenas amostras, já que a aproximação para grandes amostras é boa apenas para tamanhos de amostra razoavelmente grandes e, nestes casos, há pequena diferença entre os procedimentos. Por outro lado, não é recomendável a utilização dos testes de atribuição conjunta de postos para separação de tratamentos, já que eles têm algumas deficiências em termos de poder. Portanto, parece ser desejável a obtenção de um novo procedimento que tenha tabelas disponíveis e combine as vantagens de ambos os procedimentos para a separação de tratamentos.

Shirley (1987) observa que a utilização de testes de comparações múltiplas baseados na estatística de Kruskal-Wallis para detectar diferenças entre tratamentos na análise de experimentos inteiramente casualizados, quando a hipótese nula é falsa, não deve ser feita pois a utilização de $\frac{(N - 1)(N + 1)}{12}$ como a variância de um posto individual quando um grupo tem valores muito maiores que o de outro, é uma superestimativa que propiciará uma perda de eficiência nos procedimentos de comparações múltiplas. Sugere, também, a comparação de pares de grupos mais homogêneos através do ordenamento dos elementos destes grupos, excluindo as outras observações.

Noether (1987) discute o problema da determinação do tamanho da amostra adequada para testar uma hipótese, de tal forma que os testes tenham poder

maior ou igual a $1 - \beta$, para os seguintes testes: teste do sinal, teste de Wilcoxon, teste de Independência de Kendall e a estatística de Mann-Whitney para o teste de Wilcoxon para duas amostras. Neste artigo, o autor faz também uma comparação entre o tamanho de amostra mínima obtida para o teste de Wilcoxon e o do Sinal. Estes resultados são importantes para a discussão da eficiência destes testes em relação aos testes paramétricos equivalentes.

Ludbrook & Dudley (1998) consideram os testes de permutação mais adequados do que os testes clássicos (testes t e F) em parte dos trabalhos biomédicos, devido ao pequeno tamanho das amostras, bem como a utilização de aleatorização ao invés de amostras aleatórias. Além disso, sugerem que existem poucas publicações na área que utilizam testes de permutação devido ao pouco conhecimento dos editores dos periódicos bem como a falta de *softwares* estatísticos mais amplamente difundidos que executem testes de permutação. Alertam ainda que, numa amostra de artigos obtidos destes periódicos, em 96% dos casos utilizou-se aleatorização para a construção de grupos experimentais e em 4% apenas amostragem aleatória; enquanto isso, em 84% dos casos, o teste t ou F foi utilizado para a obtenção de conclusões e em 16% apenas foram utilizados testes não-paramétricos. Desta forma, erros grosseiros advindos da utilização dos testes t e F em dados não provenientes de populações normais e com tamanhos de grupos pequenos foram detectados, verificando-se a diferença dos resultados obtidos quando da utilização dos testes de permutação em relação aos resultados obtidos através dos testes clássicos.

Hollander & Wolfe (1999) consideram as seguintes vantagens dos testes não paramétricos fundamentais para o rápido crescimento e desenvolvimento destes testes:

1. Métodos não-paramétricos requerem poucas suposições sobre as populações de onde provêm os dados; em especial, não é necessária a suposição de normalidade;
2. Procedimentos não-paramétricos possibilitam a obtenção dos valores exatos (p -

- valores) para a realização dos testes, probabilidades exatas relacionadas aos intervalos de confiança, taxas de erros experimentais exatas para os testes de comparações múltiplas;
3. Em geral, as técnicas não-paramétricas são mais fáceis de serem aplicadas do que seus concorrentes paramétricos;
 4. Procedimentos não-paramétricos são mais fáceis de serem entendidos e aceitos;
 5. Apesar dos procedimentos não-paramétricos deixarem transparecer uma certa perda de informação, investigações mostraram que estes são um pouco menos eficientes que seus competidores paramétricos quando as suposições de normalidade são satisfeitas, e que podem ser significativamente mais eficientes do que seus competidores quando as populações de onde provêm os dados não são normais;
 6. Métodos não-paramétricos são relativamente menos sensíveis a observações discrepantes (*outliers*);
 7. Procedimentos não-paramétricos podem ser aplicáveis a situações onde os procedimentos baseados na teoria normal não são utilizáveis;
 8. O desenvolvimento dos computadores e dos *softwares* têm facilitado muito a obtenção dos p -valores exatos para os testes não-paramétricos.

3 METODOLOGIA

3.1 Testes de Permutações Sistemáticas e Testes de Permutações Aleatórias

Testes de permutação (*permutation test*) são testes que se baseiam em resultados obtidos a partir das permutações dos dados de um experimento. A forma como esta permutação é realizada depende do tipo de delineamento que está sendo realizado e do teste que se deseja aplicar. O teste de permutações sistemáticas é realizado quando todas as permutações possíveis para um dado conjunto de observações experimentais sob um determinado modelo é utilizado para obter os níveis ou as probabilidades de significância. Essas permutações são obtidas de forma sistemática, de maneira que seja assegurado que todas as permutações estarão contempladas e, ao mesmo tempo, não haver redundância. Como o número de permutações possíveis cresce muito rapidamente, muitas vezes torna-se inviável a obtenção de todas elas. Assim, pode-se utilizar um número fixo de permutações, obtidas através da aleatorização (atribuição aleatória) de acordo com o tipo de experimento que está sendo analisado e as hipóteses que estão sendo testadas.

O teste de permutação é realizado da seguinte maneira: uma estatística de teste é calculada para os dados experimentais, então os dados são permutados (divididos ou rearranjados) repetidamente e a estatística de teste é calculada para uma das permutações resultantes. Os dados permutados, que podem incluir ou não aqueles que representam os resultados obtidos, constituem o conjunto de referência para determinar a significância. A proporção de dados permutados no conjunto de referência que têm a estatística de teste com valores maiores ou iguais (ou, em certos casos, menor ou igual) ao valor para os resultados obtidos experimentalmente é o

p -valor (significância ou valor de probabilidade). A significância calculada com base na distribuição das estatísticas de teste geradas pela permutação dos dados é característica de todos os testes de permutação; quando a base para a permutação dos dados é a atribuição aleatória então um teste de permutação é denominado Teste de Permutações Aleatórias.

Esta definição é suficientemente abrangente para incluir procedimentos denominados testes aleatórios que dependem não só de amostras aleatórias mas também da aleatorização. A concepção moderna de Testes Aleatórios é um teste de permutação que é baseado somente na aleatorização, não importando como a amostra é selecionada.

A hipótese nula para um teste aleatório tradicional é que as medidas para cada unidade experimental que é atribuída aleatoriamente será a mesma, tanto sob uma atribuição escolhida como sob qualquer atribuição que possa ser obtida através do procedimento de escolha aleatória. Portanto, quando a hipótese nula para o teste de permutação tradicional (hipótese de nenhuma diferença entre tratamentos) é verdadeira, atribuições aleatórias de elementos aos tratamentos divide aleatoriamente as medidas entre os tratamentos. Cada conjunto de dados permutados no conjunto de referência, que funciona como uma *tabela de significância* para o teste aleatório, representa os resultados que teriam sido obtidos para uma particular atribuição no caso da hipótese nula ser verdadeira. Um teste aleatório é válido para qualquer tipo de amostra, independente da maneira como ela é selecionada. Esta é uma propriedade extremamente importante pois, o uso de amostras não aleatórias é comum na experimentação e as estatísticas paramétricas não são válidas para tais amostras, sendo as tabelas de estatísticas paramétricas válidas somente para os casos em que todas as amostras que têm a mesma probabilidade de serem retiradas de uma população. Violações da pressuposição de amostras aleatórias podem tomar várias formas. Uma violação é a utilização de todos os elementos; outra se refere à escolha sistemática de certos tipos de elementos; outra ainda, quando faz-se inferências estatísticas sobre

uma população baseadas em amostras aleatórias de outra população (por exemplo, uma sub-população). Além disso, é muito difundida a utilização de amostras não aleatórias em experimentos tanto por escolha do pesquisador como pelas dificuldades intrínsecas na obtenção de amostras realmente aleatórias.

Um teste estatístico para o qual a significância dos resultados experimentais é determinada pela permutação dos dados repetidamente para calcular t , F ou alguma outra estatística de teste é denominado teste aleatório. Testes aleatórios, quando as estatísticas de teste convencionais são calculadas, não são alternativas aos testes convencionais e sim testes com a significância determinada por um procedimento especial. Portanto, teste aleatório não é um tipo especial de teste estatístico, mas sim, um método de determinar a significância estatística que pode ser utilizada em conjunção com vários tipos de testes.

Sabe-se que o número de permutações necessárias para um teste de permutação cresce rapidamente com o aumento do tamanho da amostra. Mesmo para pequenas amostras, o tempo necessário para a execução de tais testes é proibitivo. O advento de computadores, entretanto, tornou a utilização de testes aleatórios possível com a utilização de 10.000, 100.000 e até 1.000.000 de casos. Outro fator que leva os testes de permutação e os testes de aleatorização a serem uma boa opção é a intelegibilidade destes testes, ou seja, a correspondência entre o modo que os dados são permutados para obter a significância e o modo como atribuições aleatórias são feitas num experimento que permite ao pesquisador verificar se o método de permutar os dados é apropriado para aquele experimento. Essa intelegibilidade se torna mais importante ainda nos casos em que as pessoas para as quais o experimentador deve comunicar seus resultados, não tem, ou têm pequeno conhecimento de estatística, já que seus resultados são extremamente claros e precisos, além de serem muito versáteis. Por causa desta versatilidade, eles podem ser utilizados tanto para assegurar a validade de testes estatísticos existentes como para desenvolver novos testes para propósitos específicos, dando portanto grande flexibilidade no planejamento e análise de dados

experimentais.

Testes paramétricos de todos os tipos, incluindo testes relativamente complexos como análise de variância de um esquema fatorial, análise de covariância e análise de variância multivariada, tornam-se de distribuição livre quando a significância é determinada por um procedimento de teste aleatório.

Enquanto que a utilização de testes de postos ordenados para assegurar a validade dos resultados levam, em geral, a uma perda do poder em relação aos testes paramétricos, testes aleatórios podem ser utilizados com a mesma validade sem provocar esta perda em relação aos testes baseados na distribuição normal.

Quando a questão de violações das pressuposições paramétricas ocorre, a determinação de significância por meio de um teste aleatório é uma boa resposta. Para delineamentos inteiramente casualizados e teste t , a atribuição aleatória aos tratamentos é feita como se segue: tipicamente, a unidade experimental é um indivíduo e o número de unidades experimentais a serem assinaladas a cada tratamento é fixada pelo pesquisador; dos indivíduos a serem utilizados no experimento, o pesquisador seleciona aleatoriamente n_1 para o primeiro tratamento, n_2 dos indivíduos restantes para o segundo tratamento e assim por diante, e qualquer indivíduo pode ser atribuído a qualquer dos tratamentos, sendo determinado de forma aleatória (com tamanho de amostra constante) qual tratamento um indivíduo tomará.

Pode-se considerar testes aleatório baseado em permutações sistemáticas para determinar significância dos resultados para uma comparação de dois ou mais tratamentos, com o número de elementos por tratamento iguais ou diferentes. Para a aplicação deste método é necessária a obtenção de um modo sistemático de listar as permutações de dados, assegurando assim que todas elas serão consideradas.

3.1.1 Delineamentos Inteiramente Casualizados - Atribuição de postos conjunta

Supondo k tratamentos com n_1, n_2, \dots, n_k elementos respectivamente,

e ainda $N = \sum_{i=1}^k n_i$, tem-se que o número total de permutações é $\frac{N!}{n_1!n_2!\dots n_k!}$. Se este número for adequado à capacidade de processamento do programa e do computador que se está trabalhando, pode-se indexar os resultados obtidos de 1 a N e permutar os números indexados ao invés dos resultados. A N -upla $(1, 2, \dots, N)$ conduz à primeira permutação, que pode ser representada por um número de N dígitos $123\dots N$. As permutações dos números indexados são listadas do menor (que representa o resultado obtido) para o maior número de N dígitos, colocando os valores indexados em ordem crescente dentro de um tratamento. Esta última imposição assegura que não haverá redundância na listagem resultante de uma mesma combinação aparecendo em ordem diferente dentro de um tratamento. Diante do exposto e afim de elaborar um programa computacional para obter tabelas para estes números, dois caminhos são considerados:

- i) Para cada permutação obtida, utilizando os valores originais associados aos valores indexados, obter a estatística $\sum_{i=1}^k \frac{T_i^2}{n_i}$ onde T_i é o total do tratamento i . Observa-se aqui que esta estatística tem o mesmo p -valor daquela que seria obtida utilizando a estatística F tradicional da Análise de Variância. Assim, calculamos a probabilidade de F como a proporção das $\frac{N!}{n_1!n_2!\dots n_k!}$ permutações que fornecerão valores maiores que o valor obtido com os dados do experimento, que é aquele referente à primeira permutação.
- ii) Utilizando-se os valores indexados atribuindo-se o valor 1 à observação de menor valor, 2 à segunda menor observação, ..., N à maior observação (RT-1, de acordo com Conover & Iman, 1981) será obtida a estatística H de Kruskal-Wallis, que é equivalente à F para dados transformados em postos. Efetuando-se as permutações e calculando-se, em cada uma delas, o valor da estatística de Kruskal-Wallis, pode-se obter a distribuição exata de H . Para calcular p -valor exato para os dados do experimento pode-se utilizar a estatística $\sum_{i=1}^k \bar{R}_i^2 = \sum_{i=1}^k \frac{R_i^2}{n_i}$.

para efetuar as comparações pois ela é equivalente à estatística H , diferindo apenas por constantes.

Observa-se que, quando as amostras têm tamanhos iguais, o cálculo pode ser simplificado tomando-se a estatística $\sum_{i=1}^k T_i^2$ que é equivalente à $\sum_{i=1}^k \frac{T_i^2}{n_i}$ e, conseqüentemente, equivalente à estatística F . Tal redução é importante para simplificar o programa e reduzir consideravelmente o seu tempo de execução. Similarmente, no caso de postos e de amostras de tamanhos iguais pode-se utilizar a estatística $\sum_{i=1}^k R_i^2$. Além disso, quando o teste a ser efetuado for bilateral, como é no caso do teste de Kruskal-Wallis, pode-se utilizar um subconjunto do conjunto completo de permutações, ou seja, somente $\frac{1}{k!}$ do número total de permutações para k tratamentos, já que para cada divisão dos dados em k grupos de tamanhos iguais existem $k!$ maneiras de assinalar os k tratamentos àqueles grupos, todos eles fornecendo o mesmo valor da estatística de teste.

Para a determinação da significância através de permutação para amostras relativamente grandes, é necessário um método alternativo. Um método possível é o Método de Permutação Aleatória, no qual um número fixo de permutações são selecionadas aleatoriamente e a probabilidade ou o valor de significância é baseado naquelas permutações. Esse teste pode ser efetuado para tratamentos com amostras de tamanhos iguais ou não e o procedimento para sua execução é o mesmo do caso anterior no que se refere à atribuição dos valores indexados. A modificação ocorre onde, ao invés de obter todas as permutações de forma sequencial, será utilizado um algoritmo de geração de números aleatórios que selecionará n_1 números indexados, sem reposição, dentre os N , para assinalar as medidas correspondentes ao tratamento A . Do restante, seleciona-se aleatoriamente, sem reposição, n_2 números para assinalar as medidas correspondentes ao tratamento B , continuando assim até que todos os números indexados sejam atribuídos aos tratamentos. Cada uma dessas atribuições corresponde a uma permutação dos dados. Seja $NPERM$ o número de permutações escolhidas. Faz-se $NPERM - 1$ permutações e para cada uma delas calcula-se o valor

da estatística de teste referente a F ou a H , como anteriormente. Calcula-se então a probabilidade de F ou de H como a proporção das $NPERM$ permutações que fornecem um valor maior que o valor da estatística de teste resultante dos dados originais. A inclusão ou não da configuração resultante dos dados originais é irrelevante quando o número de permutações geradas é grande. No caso de não inclusão, faz-se uma permutação a mais.

Ambos os procedimentos vistos acima, testes de permutação e testes de permutação aleatória, podem ser adaptados para a utilização em comparações múltiplas.

3.1.2 Delineamentos em Blocos Casualizados - atribuição de postos aos tratamentos dentro dos blocos

Para a realização de testes de permutação ou de testes de permutação aleatória nos casos em que o delineamento utilizado é em blocos casualizados, tem-se duas opções iniciais:

1. Obter as permutações e, utilizando os dados do experimento, calcular a estatística F_ℓ de Fisher para cada configuração ℓ ; a partir destas estatísticas, calcular a probabilidade de significância da estatística F das observações originais, comparando-a com as obtidas através das permutações.
2. Realizar a transformação dos dados do experimento em postos, utilizando a RT-2 como dado em Conover & Iman (1981), considerando como subgrupos os blocos; a partir dos dados do experimento calcular a estatística de S de Friedman que será comparada, para a obtenção da probabilidade de significância, com as estatísticas S_ℓ geradas a partir das permutações dos postos intra-blocos.

Caso o número de permutações possíveis ultrapasse a capacidade de armazenamento (do *hardware* ou do *software*) ou ainda faa com que o tempo de processamento seja proibitivo, o teste de permutações aleatórias pode ser utilizado,

ou seja, constrói-se um número fixo de configurações aleatórias que serão utilizadas para o cálculo da probabilidade de significância da estatística de teste. Obviamente que esta probabilidade será aproximada, tendo, em geral, um valor mais próximo do valor exato quanto maior for o número de configurações aleatórias utilizadas.

O número total de configurações possíveis para um dado experimento depende do número de tratamentos k e da quantidade de repetições ou blocos b . Como no caso dos delineamentos em blocos casualizados atribui-se postos aos tratamentos dentro de cada bloco, tem-se, para cada bloco, $k!$ permutações. Assim, considerando o conjunto dos b blocos, existem $np = (k!)^b$ configurações possíveis. No caso do testes de Friedman, calcula-se, para cada configuração obtida a estatística

$$S_\ell = \frac{12}{bk(k+1)} \sum_{i=1}^k R_i^2 - 3b(k+1) \quad \ell = 1, \dots, np$$

Duas reduções importantes podem ser feitas visando diminuir o tempo de processamento das configurações. A primeira é utilizar a estatística $Q_\ell = \sum_{i=1}^k R_i^2$ para realizar as comparações, tendo em vista que Q_ℓ difere de S_ℓ apenas por constantes. A outra redução refere-se ao número total de configurações a serem consideradas. Pode-se fixar uma configuração referente ao primeiro bloco, variando as dos blocos restantes. Com isto, o número de permutações cairá para $np = (k!)^{(b-1)}$. Esta diminuição tem sua origem na seguinte concepção: considere k tratamentos cujas somas de postos são $R_{1.} = t_1, R_{2.} = t_2, \dots, R_{k.} = t_k$ de tal forma que a estatística $S_\ell = s$; tomando qualquer configuração tal que, $R_i = t_i, i = 1, \dots, k$ onde a estrutura interna de postos dentro de cada bloco não se altere, então as estatísticas $S_{\ell'}$ obtidas a partir destas configurações terão o mesmo valor s . Desta forma, quando tem-se comparações bilaterais, como é no caso da estatística de Friedman, o número de permutações necessárias para efetuar o teste é $(k!)^{(b-1)}$.

3.2 Teste de Kruskal-Wallis

3.2.1 Aspectos Gerais

A estatística de Kruskal-Wallis é definida por

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \left(n_i \bar{R}_i - \frac{N+1}{2} \right)^2$$

ou ainda,

$$H = \frac{12}{N(N+1)} \left[\sum_{i=1}^k \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4} \right] = \frac{12}{N(N+1)} \left(\sum_{i=1}^k \frac{R_i^2}{n_i} \right) - 3(N+1) = \sum_{i=1}^k T_i^2$$

onde

$$T_i = \frac{R_i - \frac{n_i(N+1)}{2}}{\frac{N^{3/2}}{\sqrt{12}} \sqrt{\frac{n_i}{N} \left(1 + \frac{1}{N} \right)}}$$

Através de

$$P_H (R_{11} = r_{11}, \dots, R_{1n_1} = r_{1n_1}; \dots; R_{k1} = r_{k1}, \dots, R_{kn_k} = r_{kn_k}) = \frac{1}{\binom{N}{n_1 \dots n_k}}$$

$$P_H (R_{11} = r_{11}, \dots, R_{1n_1} = r_{1n_1}; \dots; R_{k1} = r_{k1}, \dots, R_{kn_k} = r_{kn_k}) = \frac{n_1! \dots n_k!}{N!}$$

pode-se determinar a distribuição nula da estatística H . Nos casos em que são verificadas grandes disparidades entre os valores R_i (somadas dos postos dos tratamentos), H deverá tomar valores grandes e, portanto, a hipótese nula de igualdade dos efeitos dos tratamentos não é verdadeira. Nesse caso, rejeita-se H_0 em favor de H_a , isto é, com um nível de significância α rejeita-se H_0 se $H \geq c_N$, onde $P_{H_0} [H \geq c_N] = \alpha$. Sob a hipótese nula, pode-se considerar a amostra combinada $(X_{11}, \dots, X_{1n_1}, \dots, X_{k1}, \dots, X_{kn_k})$ como sendo proveniente de uma mesma população. Portanto, se $c(n_1, \dots, n_k; h)$ denota o número de H 's obtido de $\frac{N!}{n_1! \dots n_k!}$ partições de N observações dentro de todas as k subgrupos possíveis de tamanhos n_1, n_2, \dots, n_k , tais que $H = h$, então

$$P[H = h] = \frac{c(n_1, \dots, n_k; h) n_1! \dots n_k!}{N!}.$$

A aproximação normal através do χ^2 com $k - 1$ graus de liberdade utilizada para o teste de Kruskal-Wallis quando temos grupos de tamanhos iguais, segundo Lehmann & D'abrera (1974) é adequada, exceto para níveis de significância baixos, mesmo para $k = 3$, cada um deles com 5 ou mais repetições ou ainda para $k > 3$ e grupos de tamanhos maiores ou iguais a 4.

De acordo com Conover (1999), outra forma que a estatística de teste para o Teste de Kruskal-Wallis pode tomar é

$$H = \frac{1}{E^2} \left(\sum_{i=1}^k \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4} \right) \quad \text{onde} \quad E^2 = \frac{1}{N-1} \left(\sum_{i=1}^k \sum_{j=1}^{n_i} R_{ij}^2 - \frac{N(N+1)^2}{4} \right)$$

e, quando não há empates, essa estatística reduz-se a

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1).$$

Considerando os dois casos, com e sem empates, pode-se verificar o que ocorre com a estatística H :

i) Primeiro Caso: sem empates

É fácil verificar que quando não ocorrem empates temos que

$$\sum_{i=1}^k \sum_{j=1}^{n_i} R_{ij}^2 = \frac{N(N+1)(2n+1)}{6}$$

e assim E^2 pode ser escrita como

$$\begin{aligned} E^2 &= \frac{1}{N-1} \left(\sum_{i=1}^k \sum_{j=1}^{n_i} R_{ij}^2 - \frac{N(N+1)^2}{4} \right) \\ &= \frac{1}{N-1} \left[\frac{N(N+1)(2N+1)}{6} - \frac{N(N+1)^2}{4} \right] \end{aligned}$$

ou ainda

$$E^2 = \frac{N(N+1)}{N-1} \left[\frac{2(2N+1) - 3(N+1)}{12} \right] = \frac{N(N+1)}{12}.$$

Com este valor de E^2 obtemos para H a seguinte expressão

$$H = \frac{12}{N(N+1)} \left[\sum_{i=1}^k \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4} \right] = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1).$$

ii) Segundo Caso: com empates

Consideremos a fórmula de H

$$H = \frac{1}{E^2} \left(\sum_{i=1}^k \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4} \right).$$

Multiplicando e dividindo o valor de H por $\frac{12}{N(N+1)}$ temos

$$H = \frac{\frac{12}{N(N+1)} \left[\sum_{i=1}^k \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4} \right]}{\frac{12}{N(N+1)} E^2}.$$

Utilizando a fórmula de E^2 temos:

$$\frac{12}{N(N+1)} S^2 = \frac{12}{N(N+1)} \frac{1}{(N-1)} \left[\sum_{i=1}^k \sum_{j=1}^{n_i} R_{ij}^2 - \frac{N(N+1)^2}{4} \right]$$

e portanto,

$$\begin{aligned} \frac{12}{N(N+1)} S^2 &= \frac{12}{N^3 - N} \sum_{i=1}^k \sum_{j=1}^{n_i} R_{ij}^2 - \frac{3(N+1)}{N-1} \\ &= \frac{1}{N^3 - N} \left[12 \sum_{i=1}^k \sum_{j=1}^{n_i} R_{ij}^2 - 3(N+1)^2 \right]. \end{aligned}$$

Considerando g grupos empatados, temos, em cada um deles, a substituição de seus valores pelos respectivos valores médios. Seja t_i o número de observações empatadas. Vejamos alguns casos particulares:

- (a) Para $t_i = 2$: considerando empatadas as observações de postos j e $j-1$, estas serão substituídas por duas observações com postos iguais a $j+1/2$; como trabalha-se com os postos elevados ao quadrado, a diferença entre o valor sem e com empates será:

$$\text{sem empate: } j^2 + (j+1)^2 = 2j^2 + 2j + 1$$

$$\text{com empate: } 2(j+1/2)^2 = 2j^2 + 2j + 1/2$$

$$\text{diferença: } 1/2.$$

- (b) Para $t_i = 3$: as observações de postos $j - 1$, j e $j + 1$ serão substituídas por três observações com postos iguais a j , ou seja:

$$\text{sem empate: } (j - 1)^2 + j^2 + (j + 1)^2 = 3j^2 + 2$$

$$\text{com empate: } 3j^2 = 3j^2$$

$$\text{diferença: } 2.$$

- (c) Para $t_i = 4$: as observações de postos $j - 1$, j , $j + 1$ e $j + 2$ serão substituídas por quatro observações com postos iguais a $j + 1/2$, ou seja:

$$\text{sem empate: } (j - 1)^2 + j^2 + (j + 1)^2 + (j + 2)^2 = 4j^2 + 4j + 6$$

$$\text{com empate: } 4(j + 1/2)^2 = 4j^2 + 4j + 1$$

$$\text{diferença: } 5.$$

- (d) Para $t_i = 5$: as observações de postos $j - 2$, $j - 1$, j , $j + 1$ e $j + 2$ serão substituídas por cinco observações com postos iguais a j , ou seja:

$$\text{sem empate: } (j - 2)^2 + (j - 1)^2 + j^2 + (j + 1)^2 + (j + 2)^2 = 5j^2 + 10$$

$$\text{com empate: } 5j^2 = 5j^2$$

$$\text{diferença: } 10.$$

e assim por diante.

A seqüência $(1/2, 2, 5, 10, \dots)$ tem como elemento gerador $\frac{t^3 - t}{12}$ para $t \geq 2$.

Generalizando, a diferença quando ocorrem empates em g grupos é dada por $\frac{1}{12} \sum_{i=1}^g (t_i^3 - t_i)$

Assim, no caso em que ocorrem empates tem-se que

$$\frac{12}{N(N + 1)} E^2 = \frac{1}{N^3 - N} \left[12 \sum_{i=1}^k \sum_{j=1}^{n_i} R_{ij}^2 - 3(N + 1)^2 \right]$$

$$\begin{aligned}
&= \frac{1}{N^3 - N} \left\{ 12 \left[\frac{N(N+1)(2N+1)}{6} - \frac{1}{12} \sum_{i=1}^g (t_i^3 - t_i) \right] \right\} - \\
&\quad - \frac{1}{N^3 - N} 3N(N+1)^2 \\
&= \frac{1}{N^3 - N} \left[N(N+1)(N-1) - \sum_{i=1}^g (t_i^3 - t_i) \right] \\
&= \frac{(N^3 - N) - \sum_{i=1}^g (t_i^3 - t_i)}{N^3 - N}.
\end{aligned}$$

Assim

$$\frac{12}{N(N+1)} E^2 = 1 - \frac{\sum_{i=1}^g (t_i^3 - t_i)}{N^3 - N}$$

e portanto,

$$H = \frac{1}{1 - \frac{\sum_{i=1}^g (t_i^3 - t_i)}{N^3 - N}} \left\{ \frac{12}{N(N+1)} \left[\sum_{i=1}^k \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4} \right] \right\}.$$

3.2.2 Construção de Tabelas para o Teste de Kruskal-Wallis

Considerando como anteriormente, k tratamentos com n_1, n_2, \dots, n_k repetições, respectivamente, e ainda $N = \sum_{i=1}^k n_i$ tem-se, no caso geral, um total de

$$\begin{aligned}
\# \text{Total de Permutações} &= \binom{N}{n_1 \dots n_k} = \frac{N!}{n_1! n_2! \dots n_k!} \\
&= C_N^{n_1} \cdot C_{N-n_1}^{n_2} \dots C_{N-(n_1+n_2+\dots+n_{k-2})}^{n_{k-1}}
\end{aligned}$$

permutações possíveis.

Caso se tenha todos os tratamentos com o mesmo número de repetições, ou seja, $n_1 = n_2 = \dots = n_k = n$, e sem a ocorrência de empates, tem-se $N = kn$ e então o número de casos cai drasticamente pois não importa a ordem em que as somas dos postos dos tratamentos ocorre e, portanto, o número total de casos fica dividido por $k!$, ou seja, neste caso tem-se

$$\# \text{ Total de Permutações} = \frac{N!}{n_1! n_2! \dots n_k! k!} = \frac{(nk)!}{(n!)^k k!} = \frac{(nk-1)!}{(n!)^{(k-1)} (k-1)!}.$$

Quando ocorrem empates, procede-se da mesma maneira, apenas substituindo os valores das observações empatadas pela média (ou pela mediana ou por outro valor qualquer designado) dos postos destas medidas. A ocorrência de empates não altera o número total de permutações e para o teste exato, empates dentro de um mesmo tratamento devem ser levados em conta na obtenção de todos os possíveis valores da estatística e, portanto, na construção do nível de significância, apesar de não alterar o valor da estatística de Kruskal-Wallis para aquela determinada configuração.

Lehman & D'Abbrera (1974) consideram a Distribuição Nula da estatística de Kruskal-Wallis na presença de empates (H_c) através de um caso em que tem-se três tratamentos (A, B e C) e a seguinte composição de postos: A : 3;3 B : 3;5;6,5 C : 1;6,5, totalizando sete repetições. Assim tem-se um total de $\frac{7!}{2!3!2!} = 210$ modos igualmente prováveis de dividir os sete elementos nos três tratamentos. Mas existem agora $C_3^2 = 3$ modos de escolher dois dos três elementos com posto médio 3 para receber o tratamento A e $C_1^2 = 2$ maneiras de escolher um dos dois elementos com posto médio 6,5 para receber o tratamento B. Feitas estas $3 \times 2 = 6$ escolhas, resta apenas uma forma de atribuir os elementos restantes ao tratamento C. Assim, a probabilidade da composição inicial é dada por $\frac{6}{210} = \frac{1}{35}$, se não for considerada a possibilidade de troca entre os tratamentos 1 e 3 que têm o mesmo tamanho (levando em conta isto, teríamos a probabilidade $\frac{12}{210}$). Já no caso de uma composição do tipo A : 1;3 B : 3;5;6,5 C : 3;6,5 tem probabilidade $\frac{12}{210}$ (ou $\frac{24}{210}$, quando se inclui a possibilidade de intercambialidade dos tratamentos). Também no caso de empates a estatística de Kruskal-Wallis pode ser aproximada por χ_{k-1}^2 mas aqui a precisão da aproximação é menor.

3.3 Comparações Múltiplas para o delineamento inteiramente casualizado

Serão apresentados os métodos de comparações múltiplas não-paramétricas para os delineamentos inteiramente casualizados, para pequenas e

grandes amostras e ainda considerando-se, de acordo com Hsu (1996), dois casos importantes: comparações entre todos os tratamentos e comparações entre os tratamentos e um tratamento designado como controle. Nos casos de grandes amostras ou ainda quando tem-se número de repetições diferentes em cada tratamentos, as aproximações utilizadas podem diferir de autor para autor.

3.3.1 Comparações múltiplas envolvendo todos os pares de tratamentos

Nas comparações múltiplas envolvendo todos os pares de tratamentos tem-se $\frac{k(k-1)}{2}$ comparações a serem efetuadas. Quando o número de tratamentos é pequeno e são feitas poucas repetições, pode-se utilizar um teste exato. Nos casos em que tem-se grandes amostras (número de tratamentos e/ou repetições grande), na maioria dos textos é utilizada uma aproximação normal mas o uso de testes de permutação aleatória é uma boa alternativa.

a) Pequenas Amostras de Tamanhos Iguais ($n_1 = \dots = n_k = n$)

Para cada par de tratamentos i, j ($1 \leq i < j \leq k$) determina-se a diferença $|R_i - R_j|$ onde R_i e R_j representam as somas dos postos atribuídos aos tratamentos i e j , respectivamente, na classificação conjunta de todas as $N = kn$ observações referentes aos k tratamentos. Havendo empates, utiliza-se os postos médios. A diferença mínima significativa (d.m.s.) a uma taxa de erro experimental α , segundo a qual admite-se que $\tau_i \neq \tau_j$ é:

$$d.m.s. = \Delta \quad \text{onde} \quad P(|R_i - R_j| \geq \Delta) = \alpha$$

onde Δ é o limite tabelado para α, n e k e é obtido a partir de todas as configurações de postos possíveis. Para a obtenção do nível mínimo de significância tem-se que

$$P(|R_i - R_j| = r) = \frac{n!n!(N-2n)!A_k^2}{N!} (\#\{(i, j); |R_i - R_j| = r\}).$$

b) Pequenas Amostras (Tamanhos Desiguais)

Dados n_i e n_j os tamanhos das amostras dos tratamentos i e j , respectivamente, calcula-se as diferenças entre as médias de tratamentos, ou seja, $|\bar{R}_i - \bar{R}_j|$ onde $\bar{R}_i = \frac{R_i}{n_i}$ e $\bar{R}_j = \frac{R_j}{n_j}$ e utiliza-se como diferença mínima significativa

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

onde h = limite de tabela do Teste de Kruskal-Wallis. Este procedimento, proposto por Nemenyi em 1963 é baseado na adaptação do teste de Scheffé aplicado às comparações não-paramétricas mas, por ser extremamente conservativo, não é muito recomendada sua utilização. Outra alternativa é utilizar o procedimento de Damico & Wolfe (1987), ou seja, tomar

$$R^* = N^* \max [|R_u - R_v|, u = 1, \dots, k-1; v = u+1, \dots, k]$$

onde $N^* =$ mínimo múltiplo comum de $\{n_1, \dots, n_k\}$ e $R_j = \frac{1}{n_j} \sum_{i=1}^{n_j} R_{ij}$ para $j = 1, \dots, k$ são os postos médios amostrais usados na estatística de Kruskal-Wallis. Para efetuar este teste os autores apresentam tabelas razoavelmente abrangentes, mas pode-se construir todas as configurações de atribuição de postos às unidades amostrais e, conseqüentemente, obter resultados mais abrangentes.

c) Amostras Grandes (tamanhos de amostras que não constam das tabelas usuais)

Uma alternativa para o caso em que temos grandes amostras é calcular as diferenças $|\bar{R}_i - \bar{R}_j|$ e, a uma taxa de erro experimental α , as diferenças mínimas significativas segundo as quais $\tau_i \neq \tau_j$ são:

i) Para $n_1 = n_2 = \dots = n_k = n$ (amostras de tamanhos iguais)

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

onde os valores de Q são tabelados e referem-se à amplitude de variáveis independentes com distribuição normal padrão.

- ii) Caso em que os tratamentos não são igualmente repetidos: Supondo n_i e n_j como os tamanhos dos tratamentos a serem comparados, a d.m.s. é dada por

$$d.m.s. = z_{\alpha/[k(k-1)]} \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

onde $z_{\alpha/[k(k-1)]}$ é o limite superior da distribuição normal. Este procedimento foi proposto por Dunn (1964) e é baseado na distribuição normal multivariada assintótica de $(\bar{R}_1, \bar{R}_2, \dots, \bar{R}_k)$ quando $n_i \rightarrow \infty$, $\frac{n_i}{N} \rightarrow \lambda_i$, $0 < \lambda_i < 1$. Sob a hipótese nula, ou seja, de igualdade das distribuições, ocorre que $E(\bar{R}_i) = \frac{N+1}{2}$; $Var(\bar{R}_i) = \frac{(N+1)(N-n_i)}{12n_i}$ e $Cov(\bar{R}_i, \bar{R}_{i'}) = \frac{-(N+1)}{12}$ ($1 \leq i \leq k$ e $1 \leq i' < i \leq k$). Sob H_0 as diferenças entre as médias dos postos dos tratamentos têm distribuições marginais normais com média 0 e variância $\left[\frac{N(N+1)}{12} \right] \left(\frac{1}{n_i} + \frac{1}{n_j} \right)$.

Outra alternativa para o caso em que temos grandes amostras (de tamanhos iguais ou não) é utilizar o processo de construção de testes de permutação aleatória aliado ao teste proposto por Damico & Wolfe (1987).

3.3.2 Comparações Múltiplas envolvendo Tratamento *versus* Controle

Designando a testemunha (ou tratamento controle) por t_1 e admitindo k tratamentos (incluindo a testemunha), pode-se fazer $(k-1)$ comparações referentes às diferenças do tipo $R_i - R_1$, ou $R_1 - R_i$ (unilaterais) ou ainda $|R_i - R_1|$ (bilaterais). Tem-se aqui, como no caso das comparações entre todos os tratamentos, a possibilidade de trabalhar com o procedimento exato ou com aproximações.

1. Procedimento Exato: Este procedimento é utilizado para o caso em que tem-se pequenas amostras e mesmo número de tratamentos. São tabelados os valores

da diferença mínima significativa (*d.m.s.* = δ) a uma taxa α para comparações unilaterais, ou seja,

$$\tau_i > \tau_1 \quad \text{se} \quad R_{i.} - R_{1.} \geq \delta \quad \text{ou} \quad \tau_i < \tau_1 \quad \text{se} \quad R_{1.} - R_{i.} \leq \delta.$$

Para comparações bilaterais utiliza-se a estatística $|R_{i.} - R_{1.}|$ que será comparada com o valor tabelado δ_1 , a uma taxa α e

$$\tau_i \neq \tau_1 \quad \text{se} \quad |R_{i.} - R_{1.}| \geq \delta_1.$$

Na construção de tabelas para a obtenção do nível de significância no caso em que tem-se controle *versus* tratamento, verifica-se que o número total de configurações é $\frac{N!}{n_1!n_2!\dots,n_k!}$ e para cada uma delas são realizadas $k - 1$ comparações.

2. Procedimento Aproximado (para grandes amostras): para situações que fogem ao escopo das tabelas apresentadas nos livros, pode-se utilizar o mesmo procedimento adotado para comparações entre todos os tratamentos, apenas lembrando que no caso de tratamentos *versus* testemunha são realizadas $k - 1$ comparações. Assim, tomando-se $\bar{R}_{i.}$ como a média dos postos do tratamento i e $\bar{R}_{1.}$ a média dos postos do controle (ou testemunha), tem-se, supondo a hipótese nula (H_0) de igualdade entre tratamentos e testemunha, para cada uma das possíveis hipóteses alternativas $H_{11} : \tau_i > \tau_1$, $H_{12} : \tau_i < \tau_1$ e $H_{13} : \tau_i \neq \tau_1$:

$$H_{11} : \text{rejeita-se } H_0 \quad \text{se} \quad \bar{R}_{i.} - \bar{R}_{1.} \geq z_{\alpha/(k-1)} \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_1} \right)}$$

$$H_{12} : \text{rejeita-se } H_0 \quad \text{se} \quad \bar{R}_{1.} - \bar{R}_{i.} \geq z_{\alpha/(k-1)} \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_1} + \frac{1}{n_i} \right)}$$

$$H_{13} : \text{rejeita-se } H_0 \quad \text{se} \quad |\bar{R}_{i.} - \bar{R}_{1.}| \geq z_{\alpha/[2(k-1)]} \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_1} \right)}$$

onde $z_{\alpha/(k-1)}$ e $z_{\alpha/[2(k-1)]}$ são os limites da distribuição normal padrão. Este procedimento é utilizado também nos casos em que tem-se amostras de tamanhos

desiguais, que não são cobertos pelas tabelas em geral, mesmo para um pequeno número de repetições em cada amostra. Um caso comum que ocorre na prática é aquele em que os tratamentos têm todos o mesmo tamanho ($n_2 = \dots = n_k = n$) e para a testemunha tem-se um número diferente (em geral maior) de repetições ($n_1 = b$). Quando isto acontece, utiliza-se, para os mesmos casos vistos anteriormente, substitui-se $z_{\alpha/(k-1)}$ pelo valor tabelado de Q_α (para o caso unilateral) e $z_{\alpha/[2(k-1)]}$ por $Q_{1\alpha}$ (no caso bilateral), sendo Q e Q_1 referentes à amplitude de variáveis independentes com distribuição normal padrão. Novamente aqui o procedimento de testes de permutação aleatória pode ser utilizado em conjunto com o teste de Damico & Wolfe (1989).

3.3.3 Construção de Tabelas de Comparações Múltiplas

Na construção de tabelas para efetuar as comparações entre tratamentos, tanto no caso de comparações entre todos os tratamentos como nas comparações entre tratamentos e testemunha, constroem-se a distribuição exata sob H_0 . A idéia inicial para a construção do modelo para ambos os casos é a mesma, coincidindo com aquela que foi utilizada para o Teste de Kruskal-Wallis, ou seja, todos os casos possíveis são obtidos através da permutação dos postos (no caso de pequenas amostras) ou obtém-se um conjunto relativamente grande de permutações aleatórias dos postos. Para cada uma das configurações obtidas calcula-se a média das somas dos postos de cada tratamento e utiliza-se o método de Damico & Wolfe (1987) (quando temos comparações entre todos os tratamentos) ou de Damico & Wolfe (1989) (comparações entre tratamentos e o controle).

Pode-se ainda utilizar um método geral, que compara todas as diferenças em todas as configurações obtidas, relacionadas ao tipo de teste que está sendo efetuado, ou seja, o número total de comparações a serem efetuadas é

$$\begin{aligned} (\# \text{ de configurações}) * \frac{k(k-1)}{2} & \quad \text{comparações entre todos os tratamentos} \\ (\# \text{ de configurações}) * (k-1) & \quad \text{comparações entre tratamentos e o controle} \end{aligned}$$

Nos casos em que tem-se o número de repetições iguais para todos os tratamentos e ainda testes de comparações múltiplas bilaterais, pode-se utilizar a mesma redução do número de casos vista para o Teste de Kruskal-Wallis, ou seja, trabalhar com as $\frac{N!}{(n!)^k k!}$ permutações.

Quando se têm amostras de tamanhos iguais e comparações entre tratamentos e testemunha (ou controle) ou ainda quando se têm hipóteses unilaterais para comparações entre todos os tratamentos, a direção da diferença torna-se importante (ou seja, considera-se $R_i - R_j \neq R_j - R_i$ para $i \neq j$) e, portanto, utiliza-se, como no caso mais geral de tratamentos com número de repetições diferentes, as $\frac{N!}{n_1! n_2! \dots n_k!}$ configurações.

Os empates serão tratados de forma semelhante ao que foi feito no Teste de Kruskal-Wallis, ou seja, calcula-se os postos médios para cada grupo de observações empatadas e a partir deste conjunto de postos, contrói-se as todas as configurações ou um número fixado de configurações aleatórias, dependendo do número de permutações totais possíveis.

3.4 Teste de Friedman

3.4.1 Aspectos Gerais

No Teste de Friedman a estatística S é dada pela soma das diferenças quadráticas entre as médias dos postos dos tratamentos ($R_{i.}$, $i = 1, \dots, k$) e o posto médio geral ($R_{..}$), ou seja,

$$S = \frac{12b}{k(k+1)} \sum_{i=1}^k (\bar{R}_{i.} - \bar{R}_{..})^2 = \frac{12b}{k(k+1)} \sum_{i=1}^k \left(\bar{R}_{i.} - \frac{k+1}{2} \right)^2$$

ou ainda, na sua forma mais utilizada,

$$S = \frac{12}{bk(k+1)} \sum_{i=1}^k R_{i.}^2 - 3b(k+1)$$

As estatísticas obtidas por Kruskal-Wallis e por Friedman são parecidas no que se refere à motivação para a construção das mesmas. O que diferencia uma

da outra é a forma como os postos são atribuídos s variáveis. É fácil observar que, se as somas dos postos de todos os tratamentos são iguais ($R_i = R_{i'}$ para $i \neq i'$), então S é nula e se os valores de R_i são muito díspares, o valor de S torna-se alto. Os valores críticos para a estatística de Friedman são tabelados para até seis tratamentos ($3 \leq k \leq 6$) com o número máximo de repetições ou blocos variando (seis para $k = 6$, oito para $k = 4$ ou $k = 5$ e treze, para $k = 3$).

A hipótese nula de igualdade de tratamentos é rejeitada quando $S \geq c_\alpha$, onde c_α é o valor tabelado para uma valor fixado de k e b . Para grandes valores de k e b , a probabilidade $P_H(S \geq c_\alpha)$ pode ser aproximada pela área à direita de c_α sob a curva do χ^2 com $k - 1$ graus de liberdade, ou seja, $P_H(S \geq c_\alpha) \simeq \chi_{k-1}^2(c_\alpha)$.

Empates entre observações entre blocos diferentes não são levados em conta devido à forma de atribuição de postos às variáveis. Quando ocorrem empates entre observações num mesmo bloco, deve-se tomar o posto médio destas e utilizar a correção para empates dada em Lehmann & D'Abbrera (1974). Assim, denotando e_j como o número de grupos empatados no bloco j e $d_{\ell j}$ como o número de observações empatadas no grupo ℓ do bloco j , tem-se que a estatística de Friedman, corrigida por C para empates, é dada por

$$S_E = \frac{S}{E} \quad \text{onde} \quad E = 1 - \frac{1}{bk(k^2 - 1)} \sum_{j=1}^b \sum_{\ell=1}^{e_j} (d_{\ell j}^3 - d_{\ell j})$$

Pode-se utilizar, ainda neste caso, a aproximação da estatística S_E através da distribuição de χ^2 com $k - 1$ graus de liberdade, mas a precisão desta depende não só do número de observações do experimento ($kb \geq 30$) mas também da configuração de empates.

Conover (1999) recomenda, na presença de empates, a utilização da estatística

$$T_F = \frac{(b - 1)T}{b(k - 1) - T}$$

onde

$$T = \frac{(k - 1) \left[\sum_{i=1}^k R_i^2 - bC_1 \right]}{A_1 - C_1} = \frac{(k - 1) \sum_{i=1}^k \left[R_i - \frac{b(k+1)}{2} \right]^2}{A_1 - C_1}$$

e ainda, $R_{.j} = \sum_{i=1}^k R_{ij}$, ($j = 1, \dots, b$) é a soma dos postos referentes ao bloco j ,
 $A_1 = \sum_{i=1}^k \sum_{j=1}^b R_{ij}^2$ é a soma dos quadrados de todos os postos, empatados ou não,
e $C_1 = \frac{bk(k+1)^2}{4}$ é denominado *fator de correção*. Estes resultados são utilizados apenas sob a forma aproximada devido à dificuldade de obtenção de sua distribuição nula. A estatística T pode ser aproximada por uma distribuição de χ^2 com $k-1$ graus de liberdade. Conover (1999) recomenda, entretanto, o uso de T_F , cuja aproximação, quando H_0 é verdadeira, é feita através da distribuição F com $(k-1)$ e $(b-1)(k-1)$ graus de liberdade.

Quando o número de tratamentos é pequeno, Lehmann & D'Abbrera (1974) propõem uma modificação ao teste de Friedman, visando torná-lo mais sensível. Este método, denominado alinhamento das observações, busca eliminar, de certa forma, o efeito de bloco. Para este fim, dentro de cada bloco subtrai-se de cada observação deste uma estimativa de sua locação, utilizando para este fim a média ou a mediana das observações do bloco. Considerando \hat{R}_{ij} o posto alinhado, ou seja, o posto de observação alinhada do i -ésimo tratamento no j -ésimo bloco, a estatística de teste será dada por

$$\hat{Q} = \frac{1}{\sum_{i=1}^k \sum_{j=1}^b (\hat{R}_{ij} - \hat{R}_{.j})^2} \sum_{i=1}^k \left[b\hat{R}_{.i} - \frac{1}{2}b(bk+1) \right]^2$$

onde $\hat{R}_{.i} = \frac{1}{b} \sum_{j=1}^b \hat{R}_{ij}$ é a média dos postos alinhados do i -ésimo tratamento nos b blocos e $\hat{R}_{.j} = \frac{1}{k} \sum_{i=1}^k \hat{R}_{ij}$ é a média dos postos alinhados dos k tratamentos no j -ésimo bloco.

3.4.2 Construção de Tabelas para o Teste de Friedman

Considerando como anteriormente, k tratamentos e b blocos ou

repetições tem-se, sob H_0 ,

$$\# \text{ de combinações possíveis para } b \text{ blocos} = k!k! \dots k! = (k!)^b$$

Este número de combinações ocorre pois são de interesse apenas as $k!$ permutações entre elementos do mesmo bloco, tomadas de forma independente para cada bloco, que serão combinadas posteriormente. Assim, para cada permutação dos tratamentos no bloco 1 tem-se $k!$ permutações do bloco 2, e assim por diante, até considerarmos as $k!$ permutações do bloco b . Sob a hipótese nula, todas estas combinações são equiprováveis. Pode-se reduzir ainda mais o número de configurações de interesse se considerarmos que, fixados os valores das somas dos postos dos k tratamentos, a troca entre estes valores não interferem no valor da estatística S , ou seja, o número de configurações distintas de interesse é

$$\# \text{ de combinações de interesse} = \frac{(k!)^b}{k!} = (k!)^{(b-1)}$$

Com este número de configurações é que obtém-se as tabelas de probabilidades de significância exatas para pequenos valores de k e b .

Como a atribuição de postos é feita às unidades experimentais dentro de cada bloco, a ocorrência de valores iguais em blocos diferentes não precisa ser considerada. Quando ocorrem empates entre observações dentro do mesmo bloco substitui-se os valores destas observações por uma medida adequada (em geral, a média dos postos destas medidas) e realiza-se as permutações com estes valores. A ocorrência de empates não altera o número total de permutações e para o teste exato.

3.5 Comparações Múltiplas para o delineamento em blocos casualizados

Como no caso das comparações múltiplas nos delineamentos inteiramente casualizados, dois tipos de comparações entre tratamentos serão consideradas, agora para o caso em que temos um delineamento em blocos casualizados: as comparações entre todos os tratamentos e comparações entre os tratamentos e um tratamento designado como controle.

Em qualquer um desses casos o procedimento de testes de permutações sistemáticas e de permutações aleatórias podem ser utilizados para os casos de pequenas e grandes amostras, respectivamente. Através do teste de permutação são obtidos resultados equivalentes aos das tabelas em uso, além de fornecer a alternativa de um teste em que o valor obtido é comparado com todas as possíveis diferenças entre tratamentos e não apenas a diferença máxima de uma determinada configuração. Além disso, no caso de grandes amostras, os resultados obtidos pelo teste de permutações aleatórias são bem consistentes devido à utilização de um grande número de configurações aleatórias. Não existe uma unanimidade quanto ao tipo de teste assintótico a ser realizado nestes casos gerando diferenças nas conclusões, para o mesmo conjunto de dados.

Como no caso dos delineamentos inteiramente casualizados e do Teste de Friedman, a ocorrência de empates não é contemplada nas tabelas, levando a resultados aproximados, mesmo no caso de pequenas amostras. Além disso, nos casos de grandes amostras, não existe concordância entre os autores no que se refere a qual aproximação utilizar. As aproximações aqui apresentadas são as mais comumente encontradas na literatura.

3.5.1 Comparações múltiplas envolvendo todos os pares de tratamentos

Considera-se dois casos complementares para as $\binom{k}{2} = \frac{k(k-1)}{2}$ comparações entre todos os pares de tratamentos:

1. Pequenas Amostras Para o teste bilateral são obtidas as diferenças absolutas entre as somas dos postos dos tratamentos $|R_u - R_v|$, $u = 1, \dots, (k-1)$ e $v = (u+1), \dots, k$ e, a uma taxa de erro *experimentalwise* α considera-se $\tau_u \neq \tau_v$ se $|R_u - R_v| \geq \delta_1$. Assim, δ_1 é a diferença mínima significativa (d.m.s.) a uma taxa α , ou seja

$$P_H(|R_u - R_v| \geq \delta_1) = \alpha$$

Os valores de δ_1 são tabelados para valores pequenos de k e b . Esta tabela foi obtida por McDonald & Thompson (1967), obtidos através da distribuição nula destas diferenças, ou seja, montando-se todas as configurações possíveis e, a partir delas, obtendo-se os valores das probabilidades de significância.

2. Grandes Amostras Nos casos em que tem-se grandes amostras, ou seja, valores de k e de b que fogem ao escopo das tabelas exatas, Campos (1983) e Hollander & Wolfed (1999), dentre outros, apresentam uma diferença mínima significativa (d.m.s.) aproximada dada por

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

onde Q é a amplitude estudantizada, tabelada para valores de k e para α fixados.

Conover (1999) apresenta um método de comparações múltiplas aproximado baseado na distribuição t de *Student*, a ser utilizado apenas no caso de rejeição da hipótese nula pelo teste de Friedman, onde os tratamentos i e j são considerados diferentes se ocorre a desigualdade

$$|R_i. - R_j.| > t_{1-\frac{\alpha}{2}} \left[\frac{2(b \sum_{i=1}^k \sum_{j=1}^b R_{ij}^2 - \sum_{i=1}^k R_i^2)}{(b-1)(k-1)} \right]^{1/2}$$

onde $t_{1-\frac{\alpha}{2}}$ é o valor tabelado t de Student com $(b-1)(k-1)$ graus de liberdade.

3.5.2 Comparações múltiplas entre os tratamentos e o controle

Considere, sem perda de generalidade, o tratamento 1 como sendo a testemunha ou tratamento controle. Podem ser efetuadas $k-1$ comparações entre o controle os $k-1$ tratamentos restantes e os testes podem ser unilaterais e bilaterais. Como anteriormente, serão considerados dois casos complementares

1. Pequenas Amostras A diferença mínima significativa ($d.m.s. = \lambda$) para as comparações unilaterais considerando os tratamentos com efeitos maiores que o do controle decide-se que $\tau_i > \tau_1$ se $R_i. - R_{1.} \geq \lambda$ ($i = 2, \dots, k$), onde λ é escolhida

de tal forma que a taxa *experimentalwise* seja igual a α , ou seja, λ satisfaz a restrição

$$P_H [(R_i. - R_{1.}) < \lambda, i = 1, \dots, k] = 1 - \alpha$$

Se os efeitos de tratamentos são menores que o do controle, decide-se que $\tau_i > \tau_1$ se $R_{1.} - R_i. \geq \lambda$ para $i = 2, \dots, k$, com λ , neste caso, satisfazendo a restrição

$$P_H [(R_i. - R_{1.}) < \lambda, i = 1, \dots, k] = 1 - \alpha$$

Ainda, no caso de comparações bilaterais, $\tau_i \neq \tau_1$ se $|R_i. - R_{1.}| \geq \lambda_1$, com λ_1 satisfazendo

$$P_H [|R_i. - R_{1.}| < \lambda_1, i = 1, \dots, k] = 1 - \alpha$$

A obtenção de valores tabelados para λ e λ_1 foi feita a partir da construção de todas as configurações possíveis e obtidas então as diferenças entre tratamentos e controle.

2. Grandes amostras Em casos que fogem do escopo das tabelas existentes, Campos (1983) e Hollander & Wolfe (1999) apresentam aproximações para o caso unilateral baseadas na distribuição do máximo de $k - 1$ distribuições normais padrão com correlação comum igual a $1/2$. Para comparações bilaterais, Campos (1983) apresenta uma aproximação baseada naquela obtida para comparações entre todos os tratamentos. Nestas aproximações o valor tabelado tem como multiplicador a quantidade $\sqrt{\frac{bk(k+1)}{6}}$.

4 RESULTADOS E DISCUSSÃO

4.1 Construção dos Programas

Para obtenção dos níveis de significância dos testes não-paramétricos dois programas foram construídos, um para delineamentos inteiramente casualizados e o outro para delineamentos em Blocos Casualizados, denominados *Programa IC* e *Programa BC*, respectivamente. O primeiro trabalha com o Teste de Kruskal-Wallis e o segundo com o Teste de Friedman e ambos fazem comparações múltiplas entre todos os tratamentos (bilateral) e entre os tratamentos e o controle (uni e bilateral), fornecendo taxas de erro *experimentalwise*, além de obter níveis de significância para contrastes escolhidos pelo usuário.

A idéia central para a construção dos programas é simples: obtém-se todas as configurações referentes à distribuição dos postos às unidades experimentais e a partir destas configurações calcula-se o nível de significância para o teste desejado. No caso de delineamentos inteiramente casualizados, pode-se ter números de repetições em cada tratamento iguais ou diferentes. Casos com empates podem ser tratados normalmente nestes programas. Nos casos em que o número de configurações totais a serem montadas é muito grande, o programa recorre ao método das permutações aleatórias, ou seja, um grande número de permutações aleatórias é gerado de acordo com o delineamento escolhido e a partir destas configurações calcula-se um nível de significância aproximado.

4.1.1 Delineamentos Inteiramente Casualizados

O *Programa IC* foi projetado com o intuito de fornecer as probabili-

dades de significância tanto para o teste de Kruskal-Wallis como para as comparações múltiplas de todos os tratamentos e dos tratamentos com o controle.

A entrada de informações e valores no *Programa IC* é feita da seguinte forma:

1. Número de tratamentos (que deve ser menor ou igual a 20);
2. Número de repetições para cada tratamento, observando que o número total de unidades experimentais deve ser menor ou igual a 100;
3. Entrada dos dados, que deve ser feita com os postos atribuídos às unidades experimentais;
4. Verificação se existe interesse em proceder às comparações múltiplas e o tipo de comparação a ser efetuada, se de todos os tratamentos, de tratamentos com o controle ou contrastes. No caso de comparações entre tratamentos e o controle, pede-se informar qual tratamento será utilizado como controle e qual tipo de comparação a ser efetuada (bilateral, unilateral para tratamentos maiores que o controle ou unilateral para tratamentos menores que o controle). Se a opção escolhida forem os contrastes, o programa solicita os coeficientes de cada um dos $k - 1$ contrastes a serem feitos.

Nos casos em que se tem grandes amostras e a obtenção do número total de configurações se torna impraticável, utiliza-se um teste de permutações aleatórias, ou seja, constrói-se 1.000.000 de configurações aleatórias e, a partir delas, obtém-se a probabilidade de significância desejada. Uma alternativa seria fornecer ao usuário o tempo aproximado necessário para a obtenção de todas as configurações para, a partir desta informação, ele poder escolher qual tipo utilizar: configurações totais ou configurações obtidas de forma aleatória. Optou-se, entretanto, por estabelecer um limite para o número de configurações (10.000.000) e, caso o número total ultrapasse este limite, o programa realiza as configurações aleatórias. Nos testes realizados, o

tempo médio para a obtenção dos resultados com 1.000.000 de configurações aleatórias foi de cerca de dois minutos, mas este tempo cresce de acordo com o tamanho do experimento devido ao processamento da aleatorização.

O programa pode ser facilmente adaptado para obtenção de tabelas, tanto para o teste de Kruskal-Wallis como para comparações múltiplas. Pela forma como foi construído, para as comparações múltiplas utilizou-se a atribuição de postos conjunta e a taxa de erro *experimentalwise*. Entretanto, o programa pode ser utilizado para a realização dos testes de comparações múltiplas *pairwise* ou mesmo para o *stepwise*. É possível ainda modificar a forma de entrada dos dados, permitindo que ela seja feita também com os dados brutos e, a partir de um subprograma, realizar a transformação destes dados em seus postos respectivos.

Os resultados apresentados pelos programa são:

1. Valor da Estatística de Kruskal-Wallis (H) para os dados de entrada;
2. Probabilidade de Significância de H , exata ou aproximada, através de permutações aleatórias, de acordo com o número total de configurações;
3. Comparações Múltiplas, de todos os tratamentos (bilateral) ou dos tratamentos com o controle (unilateral e bilateral), de acordo com a escolha feita pelo usuário; em ambos os casos, dois níveis de significância são obtidos:
 - (a) o primeiro, de acordo com Damico & Wolfe (1987) (para comparações entre todos os tratamentos) ou Damico & Wolfe (1989) (para comparações entre os tratamentos e o controle), denotado DW , em que cada módulo da diferença entre tratamentos é comparado com a diferença máxima (em módulo) de cada configuração;
 - (b) o segundo, denominado *Geral*, em que cada módulo da diferença entre tratamentos é comparado com todas as diferenças entre tratamentos ou entre os tratamentos e o controle (em módulo) de todas as configurações.

4. Se o item escolhido for contrastes, o programa fornece os níveis de significância dos contrastes escolhidos pelo usuário através dos dados de entrada.

4.1.2 Delineamentos em Blocos Casualizados

O *Programa BC*, destinado a delineamentos em blocos casualizados, foi projetado para fornecer as probabilidades de significância tanto para o teste de Friedman como para as comparações múltiplas de todos os tratamentos e dos tratamentos com o controle, além dos contrastes.

A entrada de informações e valores no programa é feita da seguinte forma:

1. Número de tratamentos (que deve ser menor ou igual a 20);
2. Número de blocos ou repetições para cada tratamento, menor ou igual a 20;
3. Entrada dos postos atribuídos às unidades experimentais, por blocos dentro dos tratamentos;
4. Verificação se existe interesse em proceder às comparações múltiplas e o tipo de comparação a ser efetuada, se de todos os tratamentos ou de tratamentos com o controle; neste caso, pede-se informar qual tratamento será utilizado como controle.
5. Caso a opção do usuário seja pelos contrastes, é solicitada a entrada dos coeficientes de $k - 1$ contrastes.

Como no caso anterior, quando tem-se grandes amostras e a obtenção do número total de configurações se torna impraticável, utiliza-se um teste de permutações aleatórias, ou seja, constrói-se 2.000.000 de configurações aleatórias e, a partir delas, obtém-se a probabilidade de significância desejada. Estas configurações

foram obtidas atribuindo-se postos aos tratamentos de forma aleatória dentro dos blocos. As observações feitas em relação ao programa para delineamentos inteiramente casualizados também são válidas aqui (adaptações possíveis, entrada de dados, etc).

Os resultados apresentados pelo programa são:

1. Valor da Estatística de Friedman (S) para os dados de entrada;
2. Probabilidade de Significância de S , exata (utilizando permutações sistemáticas) ou aproximada (através de permutações aleatórias), dependendo do número total de configurações;
3. Comparações Múltiplas, de todos os tratamentos (bilateral) ou dos tratamentos com o controle (unilateral e bilateral), de acordo com a escolha feita pelo usuário; em ambos os casos, dois níveis de significância são obtidos:
 - (a) o primeiro, denominado *Max* em que cada módulo da diferença entre tratamentos é comparado com a diferença máxima (em módulo) de cada configuração;
 - (b) o segundo, denominado *Geral*, em que cada módulo da diferença entre tratamentos é comparado com todas as diferenças entre tratamentos ou entre os tratamentos e o controle (em módulo) de todas as configurações.
4. Níveis de significância dos contrastes escolhidos pelo usuário através dos dados de entrada, caso tenha sido esta a opção.

4.2 Construção de tabelas: delineamentos inteiramente casualizados

4.2.1 Tabelas para o Teste de Kruskal-Wallis

Para a obtenção da distribuição nula da estatística de Kruskal-Wallis (H) para alguns casos específicos, com um pequeno número de tratamentos e poucas repetições, obtém-se todas as permutações possíveis e para cada uma delas calcula-se

o valor da estatística H , montando então uma tabela de distribuição de frequências para os valores de H . Tabelas contendo probabilidades de significância pequenas existem, sendo a mais abrangente encontrada em Hollander & Wolfe (1999), onde tem-se $k = 3$ tratamentos e tamanhos de amostras (n_i) menores ou iguais a seis (para $n_1 = n_2 = n_3 = n$ tem-se $n \leq 8$), $k = 4$ e $n_i \leq 4$ e para $k = 5$ e $n_i \leq 3$.

Serão apresentados alguns exemplos de delineamentos, variando o número de tratamentos e de repetições, com e sem empates.

Exemplo 1: Sejam $k = 3$ tratamentos e $n_1 = n_2 = n_3 = 2$, ou seja, número de repetições iguais, e ainda sem empates. Neste caso tem-se $N = 6$ e, portanto, $\frac{6!}{2!2!2!3!} = 15$ permutações possíveis que são mostradas na Tabela 1, com os respectivos valores das somas dos postos R_i , $i = 1, 2, 3$ e dos valores h_s , $s = 1, \dots, 15$ da estatística H , para cada combinação.

Tabela 1: Configurações para $k = 3$ tratamentos e $n_1 = n_2 = n_3 = 2$, sem empates.

Tratamentos	1	2	3	h_ℓ	1	2	3	h_ℓ	1	2	3	h_ℓ
	1	3	5		1	3	4		1	3	4	
	2	4	6		2	5	6		2	6	5	
$R'_i s$	3	7	11	4,571	3	8	10	3,714	3	9	9	3,429
	1	2	5		1	2	4		1	2	4	
	3	4	6		3	5	6		3	6	5	
$R'_i s$	4	6	11	3,714	4	7	10	2,571	4	8	9	2,000
	1	2	5		1	2	3		1	2	3	
	4	3	6		4	5	6		4	6	5	
$R'_i s$	5	5	11	3,429	5	7	9	1,143	5	8	8	0,857
	1	2	4		1	2	3		1	2	3	
	5	3	6		5	4	6		5	6	4	
$R'_i s$	6	5	10	2,000	6	6	9	0,857	6	8	7	0,286
	1	2	4		1	2	3		1	2	3	
	6	3	5		6	5	4		6	4	5	
$R'_i s$	7	5	9	1,143	7	7	7	0,000	7	6	8	0,286

A partir das configurações da Tabela 1, pode-se obter a tabela de dis-

Tabela 2: Distribuição nula da estatística de Kruskal-Wallis (H) para $k = 3$ tratamentos e $n_1 = n_2 = n_3 = 2$.

h_s	$\#(H = h_s)$	$P(H = h_s)$	$P(H \geq h_s)$
4,571	1	1/15	1/15 = 0,0667
3,714	2	2/15	3/15 = 0,2000
3,429	2	2/15	5/15 = 0,3333
2,571	1	1/15	6/15 = 0,4000
2,000	2	2/15	8/15 = 0,5333
1,143	2	2/15	10/15 = 0,6667
0,857	2	2/15	12/15 = 0,8000
0,286	2	2/15	14/15 = 0,9333
0,000	1	1/15	15/15 = 1,0000

tribuição de frequências para os valores possíveis h_s de H (Tabela 2) e, com estes resultados, obter a significância de qualquer das possíveis configurações de atribuição de postos aos tratamentos no caso considerado.

Exemplo 2: Considere o caso em que tem-se $k = 3$ tratamentos e $N = 6$, com $n_1 = 1$, $n_2 = 2$ e $n_3 = 3$. Este é o caso mais geral possível sem a ocorrência de empates, ou seja, tem-se tratamentos com números diferentes de repetições. Existem $\frac{N!}{n_1!n_2!n_3!} = \frac{6!}{1!2!3!} = 60$ permutações ou configurações possíveis, dadas na Tabela 3. Para cada configuração foi obtida a estatística de Kruskal-Wallis.

Organizando a tabela de distribuição de frequência para os valores da estatística de Kruskal-Wallis obtém-se a tabela para a distribuição exata da estatística de Kruskal-Wallis neste caso específico (Tabela 4).

Exemplo 3: Supondo agora um delineamento em que tem-se alguns tratamentos com número de repetições iguais e que haja ocorrência de empates. É necessário lembrar que, quando ocorrem empates, deve-se tratar caso a caso, pois cada tipo de empate pode levar a uma configuração diferente e que as tabelas disponíveis não contemplam empates.

Tomando como exemplo um delineamento com três tratamentos onde $n_1 = 2$, $n_2 = 1$ e $n_3 = 2$ e considerando ainda que a configuração obtida com os

Tabela 3: Configurações para $k = 3$ tratamentos e $n_1 = 1$, $n_2 = 2$ e $n_3 = 3$.

Trat.	1	2	3	h_s	1	2	3	h_s	1	2	3	h_s	1	2	3	h_s	1	2	3	h_s
	1	2	4		1	2	3		1	2	3		1	2	3		1	3	2	
		3	5			4	5			5	4			6	4			4	5	
			6				6				6				5				6	
R_i	1	5	15	4,286	1	6	14	3,095	1	7	13	2,381	1	8	12	2,143	1	7	13	2,381
	1	3	2		1	3	2		1	4	2		1	4	2		1	5	2	
		5	4			6	4			5	3			6	3			6	3	
			6				5				6			5					4	
R_i	1	8	12	2,143	1	9	11	2,381	1	9	11	2,381	1	10	10	3,095	1	11	9	4,286
	2	1	4		2	1	3		2	1	3		2	1	3		2	3	1	
		3	5			4	5			5	4			6	4			4	5	
			6				6				6			5					6	
R_i	2	4	15	3,857	2	5	14	2,381	2	6	13	1,381	2	7	12	0,857	2	7	12	0,857
	2	3	1		2	3	1		2	4	1		2	4	1		2	5	1	
		5	4			6	4			5	3			6	3			6	3	
			6				5				6			5					4	
R_i	2	8	11	0,810	2	9	10	1,238	2	9	10	1,238	2	10	9	2,143	2	11	8	3,524
	3	1	4		3	1	2		3	1	2		3	1	2		3	2	1	
		2	5			4	5			5	4			6	4			4	5	
			6				6				6			5					6	
R_i	3	3	15	4,286	3	5	13	1,238	3	6	12	0,429	3	7	11	0,095	3	6	12	0,429
	3	2	1		3	2	1		3	4	1		3	4	1		3	5	1	
		5	4			6	4			5	2			6	2			6	2	
			6				5				6			5					4	
R_i	3	7	11	0,095	3	8	10	0,238	3	9	9	0,857	3	10	8	1,952	3	11	7	3,524
	4	1	3		4	1	2		4	1	2		4	1	2		4	2	1	
		2	5			3	5			5	3			6	3			3	5	
			6				6				6			5					6	
R_i	4	3	14	3,524	4	4	13	1,952	4	6	11	0,238	4	7	10	0,095	4	5	12	0,857
	4	2	1		4	2	1		4	3	1		4	3	1		4	5	1	
		5	3			6	3			5	2			6	2			6	2	
			6				5				6			5					3	
R_i	4	7	10	0,095	4	8	9	0,429	4	8	9	0,429	4	9	8	1,238	4	11	6	4,286
	5	1	3		5	1	2		5	1	2		5	1	2		5	2	1	
		2	4			3	4			4	3			6	3			3	4	
			6				6				6			4					6	
R_i	5	3	13	3,524	5	4	12	2,143	5	5	11	1,238	5	7	9	0,857	5	5	11	1,238
	5	2	1		5	2	1		5	3	1		5	3	1		5	4	1	
		4	3			6	3			4	2			6	2			6	2	
			6				4				6			4					3	
R_i	5	6	10	0,810	5	8	8	1,381	5	7	9	0,857	5	9	7	2,381	5	10	6	3,857
	6	1	3		6	1	2		6	1	2		6	1	2		6	2	1	
		2	4			3	4			4	3			5	3			3	4	
			5				5				5			4					5	
R_i	6	3	12	4,286	6	4	11	3,095	6	5	10	2,381	6	6	9	2,143	6	5	10	2,381
	6	2	1		6	2	1		6	3	1		6	3	1		6	4	1	
		4	3			5	3			4	2			5	2			5	2	
			5				4				5			4					3	
R_i	6	6	9	2,143	6	7	8	2,381	6	7	8	2,381	6	8	7	3,095	6	9	6	4,286

Tabela 4: Distribuição nula da estatística de Kruskal-Wallis (H) para $k = 3$ tratamentos e $n_1 = 1$, $n_2 = 2$ e $n_3 = 3$.

h_s	$\#(H = h_s)$	$P(H = h_s)$	$P(H \geq h_s)$
4,286	6	3/30	3/30 = 0,1000
3,857	2	1/30	4/30 = 0,1333
3,524	4	2/30	6/30 = 0,2000
3,095	4	2/30	8/30 = 0,2667
2,381	10	5/30	13/30 = 0,4333
2,143	6	3/30	16/30 = 0,5333
1,952	2	1/30	17/30 = 0,5667
1,381	2	1/30	18/30 = 0,6000
1,238	6	3/30	21/30 = 0,7000
0,857	6	3/30	24/30 = 0,8000
0,810	2	1/30	25/30 = 0,8333
0,429	4	2/30	27/30 = 0,9000
0,238	2	1/30	28/30 = 0,9333
0,095	4	2/30	30/30 = 1,0000

dados do experimento seja aquela dada na Tabela 5, ou seja, existe empate entre as unidades amostrais com postos 2 e 3.

Tabela 5: Exemplo de uma configuração de postos para o caso com $k = 3$ tratamentos e $n_1 = 2$, $n_2 = 1$ e $n_3 = 2$, com empates ocorrendo entre os postos 2 e 3

Tratamentos (i)	1	2	3	
	1,0	2,5	4,0	
		2,5	5,0	
R_i	3,5	2,5	9,0	H=3,15

Tem-se então $\frac{5!}{2!1!2!} \frac{1}{2!} = 15$ configurações possíveis. Aqui, o último termo (ou seja, $\frac{1}{2!}$) refere-se aos dois tratamentos (1 e 3) que têm o mesmo número de repetições. Estas configurações estão listadas na Tabela 6.

É possível, a partir dos dados obtidos, organizar a tabela de distribuição de frequência para os valores da estatística de Kruskal-Wallis (Tabela 7). O valor he_s equivale ao valor de h_s com a correção para empates, ou seja, $he_s = h_s \left(\frac{20}{19} \right)$, alterando

Tabela 6: Configurações para $k = 3$ tratamentos e $n_1 = 2$, $n_2 = 1$ e $n_3 = 2$, considerando a presença de empates.

Tratamentos	1	2	3	h_s	1	2	3	h_s	1	2	3	h_s
	1,0	2,5	4,0		1,0	4,0	2,5		1,0	5,0	4,0	
	2,5		5,0		2,5		5,0		2,5		2,5	
R_i	3,5	2,5	9,0	3,15	3,5	4,0	7,5	2,10	3,5	5,0	6,5	2,90
	1,0	2,5	2,5		1,0	5,0	2,5		1,0	2,5	4,0	
	4,0		5,0		4,0		2,5		5,0		2,5	
R_i	5,0	2,5	7,5	0,75	5,0	5,0	5,0	2,00	6,0	2,5	6,5	0,15
	1,0	4,0	2,5		2,5	1,0	4,0		2,5	1,0	2,5	
	5,0		2,5		2,5		5,0		4,0		5,0	
R_i	6,0	4,0	5,0	0,60	5,0	1,0	9,0	3,60	6,5	1,0	7,5	2,10
	2,5	2,5	1,0	h	2,5	5,0	1,0	h	2,5	1,0	2,5	
	4,0		5,0		4,0		2,5		5,0		4,0	
R_i	6,5	2,5	6,0	0,15	6,5	5,0	3,0	2,25	7,5	1,0	6,5	2,10
	2,5	2,5	1,0		2,5	4,0	1,0		4,0	2,5	1,0	
	5,0		4,0		5,0		2,5		5,0		2,5	
R_i	7,5	2,5	5,0	0,75	7,5	4,0	3,5	2,10	9,0	2,5	3,5	3,15

apenas o valor da estatística mas não a distribuição em si. Para os dados do experimento obteve-se $H = 3,15$ e, portanto, $He = 3,316$, com um nível de significância $p = 0,2000$, não havendo rejeição da hipótese de igualdade dos tratamentos.

Caso o empate ocorresse entre outras unidades amostrais, a tabela seria modificada e esta é uma das dificuldades quando ocorrem empates, ou seja, não existe meios de se construir uma tabela para o caso geral. Assim, considerando-se o mesmo número de tratamentos e repetições, para cada configuração de empates uma tabela diferente é obtida.

O cálculo da estatística com correção para empates e a utilização da tabela de H sem a correção é um paliativo, mas pode levar a resultados incorretos. O uso da aproximação através da distribuição de χ^2 para pequenas amostras é inviável devido ao erro que ocorre nestas situações. Mesmo nos casos em que temos amostras razoavelmente grandes, mas com muitas observações empatadas, a aproximação não tem se mostrado muito boa.

Tabela 7: Distribuição nula da estatística de Kruskal-Wallis (H) para $k = 3$ tratamentos e $n_1 = 2$, $n_2 = 1$ e $n_3 = 2$, com empate ocorrendo entre os postos 2 e 3.

h_s	he_s	$\#(H = h_s)$	$P(H = h_s)$	$P(H \geq h_s)$
3,60	3,789	1	1/15	1/15 = 0,1333
3,15	3,316	2	2/15	3/15 = 0,2000
2,90	3,053	1	1/15	4/15 = 0,2667
2,25	2,368	1	1/15	5/15 = 0,3333
2,10	2,211	4	4/15	9/15 = 0,6000
2,00	2,105	1	1/15	10/15 = 0,6667
0,75	0,789	2	2/15	12/15 = 0,8000
0,60	0,632	1	1/15	13/15 = 0,8667
0,15	0,158	2	2/15	15/15 = 1,0000

4.2.2 Construção de Tabelas para as Comparações Múltiplas em delineamentos Inteiramente Casualizados

Para a obtenção de tabelas para efetuar as comparações múltiplas trabalhou-se com alguns casos específicos com um número pequeno de tratamentos e poucas repetições. Em alguns deles, obteve-se a probabilidade de significância apenas para valores de interesse. Quando possível, utilizou-se os quadros com todas as permutações e calculou-se as estatísticas de teste para as comparações de todos os tratamentos ou para as comparações entre tratamentos e o controle (ou testemunha), obtendo-se então a tabela completa de interesse.

Exemplo 1: Suponha que tem-se $k = 4$ tratamentos, cada um deles com $n = 3$ repetições. Neste caso o número de configurações possíveis é grande (15.400 se considerarmos comparações bilaterais) e não existe um método simples de obtenção deste valor, dada uma determinada diferença entre os totais dos postos dos tratamentos. Obteve-se, então as probabilidades de significância para alguns casos com o objetivo de mostrar as dificuldades para a realização de testes de comparações múltiplas. Considere que pretende-se calcular a probabilidade de que a diferença (absoluta) entre as somas dos postos dos tratamentos seja maior ou igual a 24, ou

seja, $P(|R_i - R_j| \geq 24)$. Como a diferença máxima entre os postos dos tratamentos é igual a 27, tem-se que:

$$P(|R_i - R_j| \geq 24) = \sum_{d=24}^{27} P(|R_i - R_j| = d)$$

As configurações para as quais obtém-se a diferença absoluta entre as somas dos postos dos tratamentos máxima ($27 = |33 - 6|$) e a diferença absoluta igual a 26 ($26 = |33 - 7| = |32 - 6|$) são mostradas na Tabela 8.

Tabela 8: Configurações para $k = 4$ tratamentos e 3 repetições por tratamento, com as diferenças entre totais de tratamentos iguais a 27 e 26.

T_1	T_2	T_3	T_4	T_1	T_2	T_3	T_4	T_1	T_2	T_3	T_4
1	r_{21}	r_{31}	10	1	r_{21}	r_{31}	10	1	r_{21}	r_{31}	9
2	r_{22}	r_{32}	11	2	r_{22}	r_{32}	11	2	r_{22}	r_{32}	11
3	r_{23}	r_{33}	12	4	r_{23}	r_{33}	12	3	r_{23}	r_{33}	11
6	R_2	R_3	33	7	R_2	R_3	33	6	R_2	R_3	32

Tem-se ainda que a diferença igual a 25 pode ser obtida quando tem-se uma soma de postos máxima igual a 33 e mínima igual a 8 (existem duas formas desta última, ou seja, tomando os postos 1, 2 e 5 ou 1, 3 e 4), ou máxima igual a 32 e mínima igual a 7 (uma maneira apenas) ou ainda máxima igual a 31 e mínima igual a 6 (duas maneiras de se obter a soma dos postos igual a 31, ou seja, 9, 10 e 12 ou 8, 11 e 12). Assim, temos que $(\#|R_i - R_j| = 25) = 5$. Para obter a diferença igual a 24 existem 10 maneiras, ou seja, 33-9 (três configurações), 32-8 (duas configurações), 31-7 (duas configurações) ou ainda 30-6 (três configurações). Desta forma, denotando $D_{ij} = |R_i - R_j|$ tem-se que

$$[\#(D_{ij} \geq 24)] = [\#(D_{ij} = 27)] + [\#(D_{ij} = 26)] + [\#(D_{ij} = 25)] + [\#(D_{ij} = 24)] = 18$$

Como o número total de configurações é igual a $N! = (nk)! = 12!$ e o número de configurações para uma determinada diferença é $3!3!6!A_4^2 = 311040$, a probabilidade desejada é dada por

$$P(D_{ij} \geq 24) = \frac{(18)(311040)}{4790001600} = 0,011688311$$

Este resultado coincide com o obtido por Damico & Wolfe (1987), ou seja, obtém-se o número de configurações em que a diferença máxima (em módulo) é igual a um valor pré-definido d e divide-se pelo número total de configurações. Para a obtenção da probabilidade de significância denominada *Geral* do programa seria necessário o desdobramento de cada configuração, fazendo o cálculo de todas as diferenças (em módulo) e verificando-se o número de diferenças iguais ao valor pré-definido. Este número seria então dividido pelo número total de comparações, ou seja, número total de configurações multiplicado por $\binom{k}{2} = \binom{4}{2} = 6$ e o resultado obtido seria $P_{Geral}[D_{ij} \geq 24] = 0,001948$.

Exemplo 2: Supondo $k = 5$ tratamentos e $n = 2$ ($n_i = 2, i = 1, \dots, 5$) repetições para cada tratamento, e que pretende-se calcular a probabilidade de que a diferença entre as somas dos postos dos tratamentos seja maior ou igual a 15, ou seja, $P(|R_i - R_j| \geq 15)$. Como o diferença máxima entre os postos dos tratamentos é igual a 16, tem-se que $P(|R_i - R_j| \geq 15) = P(|R_i - R_j| = 16) + P(|R_i - R_j| = 15)$. Existem $N! = 10!$ configurações possíveis e a Tabela 9 mostra a configuração em que o módulo da diferença entre as somas de postos de dois tratamentos atinge o máximo (16) e os casos em que esta diferença absoluta é igual a 15.

Tabela 9: Configurações para $k = 5$ tratamentos e duas repetições na qual as diferenças absolutas entre as somas dos postos são iguas a 15 ou 16 (máxima).

T_1	T_2	T_3	T_4	T_5	T_1	T_2	T_3	T_4	T_5	T_1	T_2	T_3	T_4	T_5
1	r_{21}	r_{31}	r_{41}	9	1	r_{21}	r_{31}	r_{41}	9	1	r_{21}	r_{31}	r_{41}	8
2	r_{22}	r_{32}	r_{42}	10	3	r_{22}	r_{32}	r_{42}	10	2	r_{22}	r_{32}	r_{42}	10
3	R_2	R_3	R_4	19	4	R_2	R_3	R_4	19	3	R_2	R_3	R_4	18

Assim, $\max_{i < j} |R_i - R_j| = 19 - 3 = 16$ e o número de casos em que a configuração com diferença máxima ocorre é $(2!2!6!A_5^2)(1)$ já que esta configuração ocorre de apenas uma maneira e portanto

$$P(|R_i - R_j| \geq 16) = P(|R_i - R_j| = 16) = \frac{2!2!6!A_5^2}{10!} = \frac{1}{63} = 0,01587$$

Para obter a probabilidade de ocorrência de diferenças (em módulo) maiores do que 15, deve-se lembrar que $|R_i - R_j| = 15$ pode ocorrer de duas maneiras, ou seja, $(19 - 4)$ ou $(18 - 5)$. Assim, tem-se

$$P(|R_i - R_j| = 15) = \frac{(2!2!6!A_5^2)(2)}{10!} = \frac{2}{63} = 0,03174$$

e então $P(|R_i - R_j| \geq 15) = 3/63 = 0,047619$ que coincide com os resultados de Damico & Wolfe (1987).

Utilizando-se todas as $\binom{5}{2} = 10$ comparações entre tratamentos para cada configuração obtém-se $P_{Geral}(|R_i - R_j| \geq 15) = 0,004762$. A relação entre o resultado obtido utilizando-se todas as comparações entre pares de tratamentos de todas as configurações possíveis (denominado *Geral*) e o resultado comparando-se o máximo de cada configuração (*DW*) é $P_{Geral}[D_{ij} \geq 15] = P_{DW}(D_{ij} \geq 24) / \binom{k}{2}$, como no exemplo anterior, podendo induzir à falsa conclusão de que esta relação é sempre verdadeira. Os exemplos posteriores mostram que isto nem sempre ocorre.

Exemplo 3: com $k = 3$ tratamentos e $n_1 = n_2 = n_3 = 2$ e portanto, $N = 6$ e supondo comparações bilaterais a serem feitas entre todos os tratamentos, tem-se $\frac{6!}{(2!2!2!3!)} = 15$ permutações possíveis. As configurações são as mesmas obtidas no Exemplo 1 de 4.3.

Os resultados dos $\frac{k(k-1)}{2} = 3$ módulos das diferenças entre as somas dos postos dos tratamentos e o máximo destas diferenças para cada configuração são mostrados na Tabela 10.

Utilizando a notação $D_M = \max_{i < j} |R_i - R_j|$ e $D = |R_i - R_j|$ obtém-se os resultados apresentados na Tabela 11. Os valores de $P(D_M \geq d)$ coincidem com os obtidos por Damico & Wolfe (1987) e os de $P(D \geq d)$ são aqueles denominados *Geral* obtidos através do programa. Verifica-se que a relação entre as probabilidades não é direta, ou seja, notando que $\binom{k}{2} = 3$, nem sempre $P(D \geq d) = \frac{P(D_M \geq d)}{3}$. Esta relação ocorre apenas para valores em que a diferença é próxima da máxima ($d = 7$ e $d = 8$).

Tabela 10: Valores das somas de postos de tratamentos (R_i , $i = 1, 2, 3$), diferenças absolutas entre estas somas e máximo destas diferenças ($\max_{i < j} |R_i - R_j|$).

R_1	R_2	R_3	$ R_1 - R_2 $	$ R_1 - R_3 $	$ R_2 - R_3 $	$\max_{i < j} R_i - R_j $
3	7	11	4	8	4	8
3	8	10	5	7	2	7
3	9	9	6	6	0	6
4	6	11	2	7	5	7
4	7	10	3	6	3	6
4	8	9	4	5	1	5
5	5	11	0	6	6	6
5	7	9	2	4	2	4
5	8	8	3	3	0	3
6	5	10	1	4	5	5
6	6	9	0	3	3	3
6	8	7	2	1	1	2
7	5	9	2	2	4	4
7	7	7	0	0	0	0
7	6	8	1	1	2	2

Tabela 11: Valores possíveis das diferenças entre as somas de postos de tratamentos (d) e sua distribuição de frequências.

d	$\#(D_M = d)$	$P(D_M \geq d)$	$\#(D = d)$	$P(D \geq d)$
8	1	0,0667	1	0,0222
7	2	0,2000	2	0,0667
6	3	0,4000	5	0,1778
5	2	0,5333	4	0,2667
4	2	0,6667	6	0,4000
3	2	0,8000	6	0,5333
2	2	0,9333	8	0,7111
1	0	0,9333	6	0,8444
0	1	1,0000	7	1,0000

Exemplo 4: Utilizando, como no exemplo anterior, $k = 3$ tratamentos e $n_1 = n_2 = n_3 = 2$ e portanto, $N = 6$ e supondo agora comparações dos tratamentos com o controle (considerando o primeiro tratamento como controle) tem-se $6!/(2!2!2!) = 90$ permutações possíveis.

Pretende-se realizar o teste unilateral, para a hipótese alternativa de que $\tau_j > \tau_1$ para $j = 2, \dots, k$ e os resultados das $k - 1 = 2$ diferenças entre os tratamentos $d_{21} = R_2 - R_1$ e $d_{31} = R_3 - R_1$, e o $d_M = \max(d_{21}, d_{31})$ referentes a cada configuração são mostrados na Tabela 12. Este último valor é necessário para o cálculo da probabilidade de significância como em Damico & Wolfe (1989). Para obter os valores relativos à hipótese alternativa $\tau_j < \tau_1$ basta trocar os sinais fazendo $d_{12} = -d_{21}$ e $d_{13} = -d_{31}$ e tomar $d_m = \max(d_{12}, d_{13}) = \min(d_{21}, d_{31})$.

Tabela 12: Valores das somas dos postos (R_i , $i = 1, 2, 3$), das diferenças d_{i1} ($i = 2, 3$) e das diferenças máximas d_M .

R_1	R_2	R_3	d_{21}	d_{31}	d_M	R_1	R_2	R_3	d_{21}	d_{31}	d_M	R_1	R_2	R_3	d_{21}	d_{31}	d_M
3	7	11	4	8	8	3	11	7	8	4	8	7	3	11	-4	4	4
7	11	3	4	-4	4	11	3	7	-8	-4	-4	11	7	3	-4	-8	-4
3	8	10	5	7	7	3	10	8	7	5	7	8	3	10	-5	2	2
8	10	3	2	-5	2	10	3	8	-7	-2	-2	10	8	3	-2	-7	-2
3	9	9	6	6	6	3	9	9	6	6	6	9	3	9	-6	0	0
9	9	3	0	-6	0	9	3	9	-6	0	0	9	9	3	0	-6	0
4	6	11	2	7	7	4	11	6	7	2	7	6	4	11	-2	5	5
6	11	4	5	-2	5	11	4	6	-7	-5	-5	11	6	4	-5	-7	-5
4	7	10	3	6	6	4	10	7	6	3	6	7	4	10	-3	3	3
7	10	4	3	-3	3	10	4	7	-6	-3	-3	10	7	4	-3	-6	-3
4	8	9	4	5	5	4	9	8	5	4	5	8	4	9	-4	1	1
8	9	4	1	-4	1	9	4	8	-5	-4	-4	9	8	4	-4	-5	-4
5	5	11	0	6	6	5	11	5	6	0	6	5	5	11	0	6	6
5	5	11	0	6	6	11	5	5	-6	-6	-6	11	5	5	-6	-6	-6
5	7	9	2	4	4	5	9	7	4	2	4	7	5	9	-2	2	2
7	9	5	2	-2	2	9	5	7	-4	-2	-2	9	7	5	-2	-4	-2
5	8	8	3	3	3	5	8	8	3	3	3	8	5	8	-3	0	0
8	8	5	0	-3	0	8	5	8	-3	0	0	8	8	5	0	-3	0
6	5	10	-1	4	4	6	10	5	4	-1	4	5	6	10	1	5	5
5	10	6	5	1	5	10	5	6	-5	-4	-4	10	6	5	-4	-5	-4
6	6	9	0	3	3	6	9	6	3	0	3	6	6	9	0	3	3
6	9	6	3	0	3	9	6	6	-3	-3	-3	9	6	6	-3	-3	-3
6	8	7	2	1	2	6	7	8	1	2	2	7	6	8	-1	1	1
7	8	6	1	-1	1	8	6	7	-2	-1	-1	8	7	6	-1	-2	-1
7	5	9	-2	2	2	7	9	5	2	-2	2	5	7	9	2	4	4
5	9	7	4	2	4	9	5	7	-4	-2	-2	9	7	5	-2	-4	-2
7	7	7	0	0	0	7	7	7	0	0	0	7	7	7	0	0	0
7	7	7	0	0	0	7	7	7	0	0	0	7	7	7	0	0	0
7	6	8	-1	1	1	7	8	6	1	-1	1	6	7	8	1	2	2
6	8	7	2	1	2	8	6	7	-2	-1	-1	8	7	6	-1	-2	-1

Com os valores da Tabela 12 pode-se construir, como no exemplo anterior, a Tabela 13 de valores das diferenças entre tratamentos (d), as frequências absolutas, as frequências acumuladas e as probabilidades de significância a serem uti-

Tabela 13: Valores possíveis das diferenças entre as somas de postos de tratamentos (d) e sua distribuição de frequências.

d	$\#(D_M = d)$	$\#(D_M \geq d)$	$P(D_M \geq d)$	$\#(D = d)$	$\#(D \geq d)$	$P(D \geq d)$
8	2	2	0,0222	2	2	0,0111
7	4	6	0,0667	4	6	0,0333
6	8	14	0,1556	10	16	0,0889
5	6	20	0,2222	8	24	0,1333
4	8	28	0,3111	12	36	0,2000
3	8	36	0,4000	12	48	0,2667
2	10	46	0,5111	16	64	0,3556
1	6	52	0,5778	12	76	0,4222
0	14	66	0,7333	18	104	0,5778
< 0	24	90	1,0000	76	180	1,0000

lizadas nos testes de comparações múltiplas de acordo com Damico & Wolfe (1989) e o caso geral obtido pelo programa. A notação utilizada foi a mesma do exemplo anterior, ou seja, $\#(D_M = d)$, $\#(D_M \geq d)$, $P(D_M \geq d)$ e $\#(D = d)$, $\#(D \geq d)$, $P(D \geq d)$ para as frequências absolutas, frequências acumuladas e as probabilidades de significância referentes aos testes de Damico & Wolfe (1987) e ao caso *Geral* obtido pelo programa, respectivamente.

Exemplo 5: Para mostrar a construção de tabelas para comparações múltiplas supondo tratamentos com números de repetições diferentes, tomou-se, como no Exemplo 2 da Seção 4.2, $k = 3$ tratamentos com $n_1 = 1$, $n_2 = 2$ e $n_3 = 3$ repetições, totalizando $N = 6$ unidades experimentais. Serão aqui utilizadas aqui as $\frac{6!}{1!2!3!} = 60$ configurações construídas no exemplo citado e as médias $\bar{R}_i = R_i/n_i$, $i = 1, 2, 3$ dos tratamentos. Considere as $\binom{k}{2} = \binom{3}{2} = 3$ diferenças entre médias de tratamentos $d_{ij} = |\bar{R}_i - \bar{R}_j|$ e as *diferenças padronizadas* $r_1^* = N^*d_{12}$, $r_2^* = N^*d_{13}$ e $r_3^* = N^*d_{23}$, onde N^* o mínimo múltiplo comum entre n_1 , n_2 e n_3 (nesse caso, $N^* = 6$). Para cada configuração obtém-se ainda, $r^* = N^* \max d_{ij} = \max(r_t^*)$ para $i < j$ e $1 \leq t \leq [k(k-1)]/2$ e o quadro de valores para as comparações múltiplas entre todos os tratamentos, mostrado na Tabela 14, é obtido.

Tabela 14: Médias dos postos dos tratamentos (\bar{R}_i), diferenças entre estas médias e valores das diferenças padronizadas (r_i^*).

R_1	R_2	R_3	d_{12}	d_{13}	d_{23}	r_1^*	r_2^*	r_3^*	r^*	R_1	R_2	R_3	d_{12}	d_{13}	d_{23}	r_1^*	r_2^*	r_3^*	r^*
1,0	2,5	5,0	1,5	4,0	2,5	9	24	15	24	1,0	3,0	4,7	2,0	3,7	1,7	12	22	10	22
1,0	3,5	4,3	2,5	3,3	0,8	15	20	5	20	1,0	4,0	3,0	3,0	2,0	1,0	18	12	6	18
1,0	3,5	4,3	2,5	3,3	0,8	15	20	5	20	1,0	4,0	3,0	3,0	2,0	1,0	18	12	6	18
1,0	4,5	3,7	3,5	2,7	0,8	21	16	5	21	1,0	4,5	3,7	3,5	2,7	0,8	21	16	5	21
1,0	5,0	3,3	4,0	2,3	1,7	24	14	10	24	1,0	5,5	3,0	4,5	2,0	2,5	27	12	15	27
2,0	2,0	5,0	0,0	3,0	3,0	0	18	18	18	2,0	2,5	4,7	0,5	2,7	2,2	3	16	13	16
2,0	3,0	4,3	1,0	2,3	1,3	6	14	8	14	2,0	3,5	4,0	1,5	2,0	0,5	9	12	3	12
2,0	3,5	4,0	1,5	2,0	0,5	9	12	3	12	2,0	4,0	3,7	2,0	1,7	0,3	12	10	2	12
2,0	4,5	3,3	2,5	1,3	1,2	15	8	7	15	2,0	4,5	3,3	2,5	1,3	1,2	15	8	7	15
2,0	5,0	3,0	3,0	1,0	2,0	18	6	12	18	2,0	5,5	2,7	3,5	0,7	2,8	21	4	17	21
3,0	1,5	5,0	1,5	2,0	3,5	9	12	21	21	3,0	2,5	4,3	0,5	1,3	1,8	3	8	11	11
3,0	3,0	4,0	0,0	1,0	1,0	0	6	6	6	3,0	3,5	3,7	0,5	0,7	0,2	3	4	1	4
3,0	3,0	4,0	0,0	1,0	1,0	0	6	6	6	3,0	3,5	3,7	0,5	0,7	0,2	3	4	1	4
3,0	4,0	3,3	1,0	0,3	0,7	6	2	4	6	3,0	4,5	3,0	1,5	0,0	1,5	9	0	9	9
3,0	5,0	2,7	2,0	0,3	2,3	12	2	14	14	3,0	5,5	2,3	2,5	0,7	3,2	15	4	19	19
4,0	1,5	4,7	2,5	0,7	3,3	15	4	19	19	4,0	2,0	4,3	2,0	0,3	2,3	12	2	14	14
4,0	3,0	3,7	1,0	0,3	0,7	6	2	4	6	4,0	3,5	3,3	0,5	0,7	0,2	3	4	1	4
4,0	2,5	4,0	1,5	0,0	1,5	9	0	9	9	4,0	3,5	3,3	0,5	0,7	0,2	3	4	1	4
4,0	4,0	3,0	0,0	1,0	1,0	0	6	6	6	4,0	4,0	3,0	0,0	1,0	1,0	0	6	6	6
4,0	4,5	2,7	0,5	1,3	1,8	3	8	11	11	4,0	5,5	2,0	1,5	2,0	3,5	9	12	21	21
5,0	1,5	4,3	3,5	0,7	2,8	21	4	17	21	5,0	2,0	4,0	3,0	1,0	2,0	18	6	12	18
5,0	2,5	3,7	2,5	1,3	1,2	15	8	7	15	5,0	3,5	3,0	1,5	2,0	0,5	9	12	3	12
5,0	2,5	3,7	2,5	1,3	1,2	15	8	7	15	5,0	3,0	3,3	2,0	1,7	0,3	12	10	2	12
5,0	4,0	2,7	1,0	2,3	1,3	6	14	8	14	5,0	3,5	3,0	1,5	2,0	0,5	9	12	3	12
5,0	4,5	2,3	0,5	2,7	2,2	3	16	13	16	5,0	5,0	2,0	0,0	3,0	3,0	0	18	18	18
6,0	1,5	4,0	4,5	2,0	2,5	27	12	15	27	6,0	2,0	3,7	4,0	2,3	1,7	24	14	10	24
6,0	2,5	3,3	3,5	2,7	0,8	21	16	5	21	6,0	3,0	3,0	3,0	3,0	0,0	18	18	0	18
6,0	2,5	3,3	3,5	2,7	0,8	21	16	5	21	6,0	3,0	3,0	3,0	3,0	0,0	18	18	0	18
6,0	3,5	2,7	2,5	3,3	0,8	15	20	5	20	6,0	3,5	2,7	2,5	3,3	0,8	15	20	5	20
6,0	4,0	2,3	2,0	3,7	1,7	12	22	10	22	6,0	4,5	2,0	1,5	4,0	2,5	9	24	15	24

Pode-se, então, obter as distribuições nulas de R_{DW}^* e de R_{Geral}^* , calculando as probabilidades $p_{DW} = P[R^* \geq N^* \max d_{ij}] = P[R_{DW}^* \geq \max(r_t^*)]$ como em Damico & Wolfe (1987) e $p_{Geral} = P[R^* \geq N^* d_{ij}] = P[R_{Geral}^* \geq r_t^*]$, para $t = 1, 2, 3$, a partir dos dados da Tabela 14.

Assim, a distribuição nula para as comparações múltiplas entre todos os $k = 3$ tratamentos, considerando $n_1 = 1$, $n_2 = 2$ e $n_3 = 3$ repetições e a não ocorrência de empates, é dada na Tabela 15.

Os exemplos apresentados mostram que os procedimentos para comparações múltiplas apresentados em Damico & Wolfe (1987) (para comparações entre todos os tratamentos) e Damico & Wolfe (1989) (para comparações entre os tratamentos e o controle) são, em geral, mais conservativos do que o procedimento de comparação denominado *Geral*, pois este último utiliza-se de todas as diferenças entre tratamentos (ou entre tratamentos e testemunha) enquanto que aqueles utilizam como elemento de comparação a diferença máxima dentro de cada configuração.

Tabela 15: Valores possíveis das diferenças padronizadas entre as somas de postos de tratamentos (r^*) e sua distribuição de frequências.

r^*	$\#[R^* = \max(r_t^*)]$	$\#[R^* \geq \max(r_t^*)]$	p_{DW}	$\#(R^* = r_t^*)$	$\#(R^* \geq r_t^*)$	p_{Geral}
27	2	2	0,0333	2	2	0,0111
24	4	6	0,1000	4	6	0,0333
22	2	8	0,1333	2	8	0,0444
21	8	16	0,2667	8	16	0,0889
20	4	20	0,3333	4	20	0,1111
19	2	22	0,3667	2	22	0,1222
18	8	30	0,5000	12	34	0,1889
17	0	30	0,5000	2	36	0,2000
16	2	32	0,5333	6	42	0,2333
15	4	36	0,6000	14	56	0,3111
14	4	40	0,6667	6	62	0,3444
13	0	40	0,6667	2	64	0,3556
12	6	46	0,7667	18	82	0,4556
11	2	48	0,8000	2	84	0,4667
10	0	48	0,8000	6	90	0,5000
9	2	50	0,8333	12	102	0,5667
8	0	50	0,8333	8	110	0,6111
7	0	50	0,8333	4	114	0,6333
6	6	56	0,9333	16	130	0,7222
5	0	56	0,9333	8	138	0,7667
4	4	60	1,0000	10	148	0,8222
3	0	60	1,0000	12	160	0,8889
2	0	60	1,0000	6	166	0,9222
1	0	60	1,0000	4	170	0,9444
0	0	60	1,0000	10	180	1,0000

4.3 Construção de tabelas: delineamentos em blocos casualizados

Considere k tratamentos com b repetições ou blocos, em que cada tratamento ocorre apenas uma vez dentro de cada bloco. Suponha que os dados originais X_{ij} já estejam transformados em seus respectivos postos R_{ij} , com a atribuição feita da seguinte maneira: considerando o bloco j , atribui-se o posto 1 à variável de menor valor, posto 2 à variável de segundo menor valor, e assim por diante, até a atribuição do posto k à variável de maior valor dentro do bloco; utiliza-se o mesmo procedimento para todos os blocos e obtém-se então o esquema da Tabela 16.

Tabela 16: Esquema de distribuição de postos para um delineamento em blocos casualizados

Tratamentos	Bloco 1	Bloco 2	...	Bloco b	$R_{i.}$
1	R_{11}	R_{12}	...	R_{1b}	$R_{1.}$
2	R_{21}	R_{22}	...	R_{2b}	$R_{2.}$
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
k	R_{k1}	R_{k2}	...	R_{kb}	$R_{k.}$
$R_{.j}$	$R_{.1}$	$R_{.2}$...	$R_{.b}$	$R_{..}$

A hipótese a ser testada é a usual de igualdade dos k tratamentos, $H_0 : \tau_1 = \dots = \tau_k$, contra a hipótese alternativa de que existe pelo menos um par de tratamentos cujos efeitos diferem entre si, ou seja, $H_1 : \tau_i \neq \tau_j$ para $1 \leq i < j \leq k$.

A estatística de teste para o teste de Friedman, quando não ocorrem empates, é dada por

$$S = \frac{12}{bk(k+1)} \sum_{i=1}^k R_{i.} - 3b(k+1)$$

e, caso ocorram empates, considerando-se e_j grupos de empates dentro do bloco j e $d_{\ell j}$ observações empatadas do grupo ℓ no bloco j , utiliza-se a correção C , ou seja,

$$S_E = \frac{S}{C} \quad \text{onde} \quad C = 1 - \frac{1}{bk(k^2-1)} \sum_{j=1}^b \sum_{\ell=1}^{e_j} (d_{\ell j}^3 - d_{\ell j})$$

O procedimento consiste em obter todas as configurações e, em cada

uma delas calcular o valor da estatística desejada, seja para o teste de Friedman ou para realizar as comparações múltiplas.

4.3.1 Construção de tabelas para o Teste de Friedman

Exemplo 1: Considerando-se três tratamentos ($k = 3$) e dois blocos ($b = 2$), como foi visto anteriormente, existem $(3!)^2 = 36$ configurações possíveis de distribuição dos postos, ou seja, considerando a tripla (r_1, r_2, r_3) como sendo a dos postos dos tratamentos 1, 2 e 3, respectivamente, tem-se, para os dois blocos, as possibilidades delineadas na Tabela 17.

Tabela 17: Configurações para $k = 3$ tratamentos e $b = 2$ blocos

Bloco 1	Bloco 2	Bloco 1	Bloco 2	Bloco 1	Bloco 2
(1,2,3)	(1,2,3)	(2,1,3)	(1,2,3)	(3,1,2)	(1,2,3)
(1,2,3)	(1,3,2)	(2,1,3)	(1,3,2)	(3,1,2)	(1,3,2)
(1,2,3)	(2,1,3)	(2,1,3)	(2,1,3)	(3,1,2)	(2,1,3)
(1,2,3)	(2,3,1)	(2,1,3)	(2,3,1)	(3,1,2)	(2,3,1)
(1,2,3)	(3,1,2)	(2,1,3)	(3,1,2)	(3,1,2)	(3,1,2)
(1,2,3)	(3,2,1)	(2,1,3)	(3,2,1)	(3,1,2)	(3,2,1)
(1,3,2)	(1,2,3)	(2,3,1)	(1,2,3)	(3,2,1)	(1,2,3)
(1,3,2)	(1,3,2)	(2,3,1)	(1,3,2)	(3,2,1)	(1,3,2)
(1,3,2)	(2,1,3)	(2,3,1)	(2,1,3)	(3,2,1)	(2,1,3)
(1,3,2)	(2,3,1)	(2,3,1)	(2,3,1)	(3,2,1)	(2,3,1)
(1,3,2)	(3,1,2)	(2,3,1)	(3,1,2)	(3,2,1)	(3,1,2)
(1,3,2)	(3,2,1)	(2,3,1)	(3,2,1)	(3,2,1)	(3,2,1)

Mas, destas $(3!)^2 = 36$ configurações, apenas $(3!)^1 = 6$ são distintas já que, por exemplo, os pares de triplas $[(1, 2, 3); (1, 2, 3)]$, $[(1, 3, 2); (1, 3, 2)]$, $[(2, 1, 3); (2, 1, 3)]$, $[(2, 3, 1); (2, 3, 1)]$, $[(3, 1, 2); (3, 1, 2)]$ e $[(3, 2, 1); (3, 2, 1)]$ levam à mesma soma de postos ($R_u = 2, R_v = 4, R_w = 6$) e ainda com os mesmos valores geradores, ou seja, (1, 1) para $R_u = 2$, (2, 2) para $R_v = 4$ e (3, 3) para $R_w = 6$. Fixando então o bloco $j = 2$ e fazendo variar os elementos do bloco $j = 1$, tem-se as somas de postos dadas na Tabela 18.

Tabela 18: Configurações necessárias para o cálculo da estatística S no caso de $k = 3$ tratamentos e $b = 2$ blocos e somas dos postos (R_i , $i = 1, 2, 3$)

ℓ	Bloco 1	Bloco 2	$R_1.$	$R_2.$	$R_3.$
1	(1,2,3)	(1,2,3)	$R_1. = 2$,	$R_2. = 4$,	$R_3. = 6$
2	(1,3,2)	(1,2,3)	$R_1. = 2$,	$R_2. = 5$,	$R_3. = 5$
3	(2,1,3)	(1,2,3)	$R_1. = 3$,	$R_2. = 3$,	$R_3. = 6$
4	(2,3,1)	(1,2,3)	$R_1. = 3$,	$R_2. = 5$,	$R_3. = 4$
5	(3,1,2)	(1,2,3)	$R_1. = 4$,	$R_2. = 3$,	$R_3. = 5$
6	(3,2,1)	(1,2,3)	$R_1. = 4$,	$R_2. = 4$,	$R_3. = 4$

Para cada uma destas configurações calcula-se a estatística S de Friedman e com estes valores obtém-se a Tabela 19, com a distribuição nula de S para este caso específico, ou seja, $k = 3$ tratamentos e $b = 2$ blocos, sem empates.

Tabela 19: Distribuição nula exata da estatística S ($k = 3$ tratamentos e $b = 2$ blocos)

s	$\#(S = s)$	$\#(S \geq s)$	$P(S \geq s)$
4	1	1	$1/6 = 0,1667$
3	2	3	$3/6 = 0,5000$
1	2	5	$5/6 = 0,8333$
0	1	6	$6/6 = 1,0000$

Exemplo 2: Seja $k = 3$ tratamentos e $b = 3$ blocos ou repetições, tem-se $(3!)^3 = 216$ configurações, das quais $(3!)^2 = 36$ são consideradas para a obtenção da distribuição nula de S e, por consequência, para construção das tabelas. Fixando-se o bloco 3, tem-se as 36 configurações mostradas na Tabela 20.

Calcula-se então as estatísticas S de Friedman para cada uma das 36 configurações. Com estes valores, pode-se obter a tabela com a distribuição nula de S (Tabela 21) considerando $k = 3$ tratamentos e $b = 3$ blocos, sem empates entre elementos do mesmo tratamento. Estes valores são os mesmos obtidos nas tabelas clássicas para o Teste de Friedman.

Tabela 20: Configurações para $k = 3$ tratamentos e $b = 3$ blocos

ℓ	Bl.1	Bl.2	Bl.3	R_1	R_2	R_3	ℓ	Bl.1	Bl.2	Bl.3	R_1	R_2	R_3
1	1,2,3	1,2,3	1,2,3	3	6	9	19	2,3,1	1,2,3	1,2,3	4	7	7
2	1,2,3	1,3,2	1,2,3	3	7	8	20	2,3,1	1,3,2	1,2,3	4	8	6
3	1,2,3	2,1,3	1,2,3	4	5	9	21	2,3,1	2,1,3	1,2,3	5	6	7
4	1,2,3	2,3,1	1,2,3	4	7	7	22	2,3,1	2,3,1	1,2,3	5	8	5
5	1,2,3	3,1,2	1,2,3	5	5	8	23	2,3,1	3,1,2	1,2,3	6	6	6
6	1,2,3	3,2,1	1,2,3	5	6	7	24	2,3,1	3,2,1	1,2,3	6	7	5
7	1,3,2	1,2,3	1,2,3	3	7	8	25	3,1,2	1,2,3	1,2,3	5	5	8
8	1,3,2	1,3,2	1,2,3	3	8	7	26	3,1,2	1,3,2	1,2,3	5	6	7
9	1,3,2	2,1,3	1,2,3	4	6	8	27	3,1,2	2,1,3	1,2,3	6	4	8
10	1,3,2	2,3,1	1,2,3	4	8	6	28	3,1,2	2,3,1	1,2,3	6	6	6
11	1,3,2	3,1,2	1,2,3	5	6	7	29	3,1,2	3,1,2	1,2,3	7	4	7
12	1,3,2	3,2,1	1,2,3	5	7	6	30	3,1,2	3,2,1	1,2,3	7	5	6
13	2,1,3	1,2,3	1,2,3	4	5	9	31	3,2,1	1,2,3	1,2,3	5	6	7
14	2,1,3	1,3,2	1,2,3	4	6	8	32	3,2,1	1,3,2	1,2,3	5	7	6
15	2,1,3	2,1,3	1,2,3	5	4	9	33	3,2,1	2,1,3	1,2,3	6	5	7
16	2,1,3	2,3,1	1,2,3	5	6	7	34	3,2,1	2,3,1	1,2,3	6	7	5
17	2,1,3	3,1,2	1,2,3	6	4	8	35	3,2,1	3,1,2	1,2,3	7	5	6
18	2,1,3	3,2,1	1,2,3	6	5	7	36	3,2,1	3,2,1	1,2,3	7	6	5

Tabela 21: Distribuição nula exata da estatística S ($k = 3$ e $b = 3$)

s	$\#(S = s)$	$\#(S \geq s)$	$P(S \geq s)$
6,0000	1	1	$1/36 = 0,02778$
4,6667	6	7	$7/36 = 0,19444$
2,6667	6	13	$13/36 = 0,36111$
2,0000	6	19	$19/36 = 0,52778$
0,6667	15	34	$34/36 = 0,94444$
0,0000	2	36	$36/36 = 1,00000$

Exemplo 3: Considere agora $k = 4$ tratamentos e $b = 3$ blocos ou repetições, tem-se $(4!)^3 = 331.776$ configurações, das quais $(4!)^2 = 576$ são consideradas para a obtenção da distribuição nula de S .

Pode-se montar uma tabela de significância para alguns valores de S (Tabela 22), obtidas a partir das configurações apenas para valores grandes da estatística S (Tabela 23).

Percebe-se que, mesmo com um número pequeno de tratamentos e repetições torna-se complicada a obtenção de todas as configurações necessárias para a construção das tabelas. Mesmo se for feita apenas a verificação das configurações que levam a valores grandes de S como neste caso, não é tão fácil, pois inexistente uma regra geral.

Estes valores podem ser encontrados tanto nas tabelas usuais como através do *Programa BC*. A maneira como tais níveis de significância são encontrados é a mesma em ambos os casos, ou seja, através da obtenção de todas as permutações geradoras dos valores de interesse da estatística S . Como o intuito do programa é também a obtenção dos níveis de significância para as comparações múltiplas, ele gera todas as configurações possíveis.

Tabela 22: Alguns valores de S e as respectivas probabilidades de significância para $k = 4$ tratamentos e $b = 3$ blocos

s	$\#(S = s)$	$\#(S \geq s)$	$P(S \geq s)$
9,0	1	1	0,001736
8,2	9	10	0,017361
7,4	9	19	0,032986
7,0	12	31	0,053819
6,6	12	43	0,074653

Tabela 23: Configurações para $k = 4$ tratamentos e $b = 3$ blocos

Blocos	Tratamentos					h_ℓ	Tratamentos					h_ℓ	Tratamentos					h_ℓ	Tratamentos					h_ℓ
	1	2	3	4			1	2	3	4			1	2	3	4			1	2	3	4		
B_1	1	2	3	4			1	2	3	4			1	2	3	4			1	3	2	4		
B_2	1	2	3	4			1	2	3	4			1	3	2	4			1	2	3	4		
B_3	1	2	3	4			1	3	2	4			1	2	3	4			1	2	3	4		
$R_{i.}$	3	6	9	12		9,0	3	7	8	12		8,2	3	7	8	12		8,2	3	7	8	12		8,2
B_1	1	2	3	4			1	2	3	4			1	2	4	3			1	2	3	4		
B_2	1	2	3	4			1	2	4	3			1	2	3	4			1	2	3	4		
B_3	1	2	4	3			1	2	3	4			1	2	3	4			2	1	3	4		
$R_{i.}$	3	6	10	11		8,2	3	6	10	11		8,2	3	6	10	11		8,2	4	5	9	12		8,2
B_1	1	2	3	4			2	1	3	4			1	2	3	4			1	2	3	4		
B_2	2	1	3	4			1	2	3	4			1	2	3	4			2	1	3	4		
B_3	1	2	3	4			1	2	3	4			2	1	4	3			1	2	4	3		
$R_{i.}$	4	5	9	12		8,2	4	5	9	12		8,2	4	5	10	11		7,4	4	5	10	11		7,4
B_1	2	1	3	4			1	2	3	4			1	2	3	4			2	1	3	4		
B_2	1	2	3	4			1	2	4	3			2	1	4	3			1	2	4	3		
B_3	1	2	4	3			2	1	3	4			1	2	3	4			1	2	3	4		
$R_{i.}$	4	5	10	11		7,4	4	5	10	11		7,4	4	5	10	11		7,4	4	5	10	11		7,4
B_1	1	2	4	3			1	2	4	3			2	1	4	3			1	3	2	4		
B_2	1	2	3	4			2	1	3	4			1	2	3	4			1	2	3	4		
B_3	2	1	3	4			1	2	3	4			1	2	3	4			1	2	4	3		
$R_{i.}$	4	5	10	11		7,4	4	5	10	11		7,4	4	5	10	11		7,4	3	7	9	11		7,0
B_1	1	3	2	4			1	2	4	3			1	2	3	4			1	2	3	4		
B_2	1	2	4	3			1	2	3	4			1	3	2	4			1	2	4	3		
B_3	1	2	3	4			1	3	2	4			1	2	4	3			1	3	2	4		
$R_{i.}$	3	7	9	11		7,0	3	7	9	11		7,0	3	7	9	11		7,0	3	7	9	11		7,0
B_1	1	2	4	3			1	2	3	4			1	3	2	4			1	2	3	4		
B_2	1	3	2	4			1	3	2	4			1	2	3	4			2	1	3	4		
B_3	1	2	3	4			2	1	3	4			2	1	3	4			1	3	2	4		
$R_{i.}$	3	7	9	11		7,0	4	6	8	12		7,0	4	6	8	12		7,0	4	6	8	12		7,0
B_1	1	3	2	4			2	1	3	4			2	1	3	4			1	3	2	4		
B_2	2	1	3	4			1	2	3	4			1	3	2	4			1	3	2	4		
B_3	1	2	3	4			1	3	2	4			1	2	3	4			1	2	4	3		
$R_{i.}$	4	6	8	12		7,0	4	6	8	12		7,0	4	6	8	12		7,0	3	8	8	11		6,6
B_1	1	3	2	4			1	2	4	3			1	3	2	4			1	2	4	3		
B_2	1	2	4	3			1	3	2	4			1	2	4	3			1	3	2	4		
B_3	1	3	2	4			1	3	2	4			1	2	4	3			1	2	4	3		
$R_{i.}$	3	8	8	11		6,6	3	8	8	11		6,6	3	7	10	10		6,6	3	7	10	10		6,6
B_1	1	2	4	3			1	3	2	4			1	3	2	4			2	1	3	4		
B_2	1	2	4	3			1	3	2	4			2	1	3	4			1	3	2	4		
B_3	1	3	2	4			2	1	3	4			1	3	2	4			1	3	2	4		
$R_{i.}$	3	7	10	10		6,6	3	7	10	10		6,6	3	7	10	10		6,6	4	7	7	12		6,6
B_1	1	3	2	4			2	1	3	4			2	1	3	4								
B_2	2	1	3	4			1	3	2	4			2	1	3	4								
B_3	2	1	3	4			2	1	3	4			1	3	2	4								
$R_{i.}$	3	7	10	10		6,6	3	7	10	10		6,6	3	7	10	10		6,6						

4.3.2 Construção de tabelas para as comparações múltiplas em delineamentos em blocos casualizados

Exemplo 1: Considere três tratamentos ($k = 3$) e três blocos ($b = 3$), sem a presença de empates, como no Exemplo 2 da Seção 4.3.1.

Tem-se $(3!)^2 = 216$ configurações que, no caso de comparações entre todos os tratamentos (bilateral) reduz-se a $(3!)^2 = 36$ configurações distintas, apresentadas na Tabela 24.

A partir dos resultados da Tabela 24, obtém-se as três possíveis diferenças absolutas entre totais de tratamentos $d_{12} = |R_{1.} - R_{2.}|$, $d_{13} = |R_{3.} - R_{1.}|$ e $d_{23} = |R_{3.} - R_{2.}|$, o máximo destas diferenças (d_M) e finalmente a distribuição de frequências das diferenças e do máximo (mostradas na Tabela 25). Tais resultados coincidem com os obtidos através do *Programa BC* e com as tabelas usuais.

Tabela 24: Somas dos postos ($R_{i.}$, $i = 1, 2, 3$), diferenças entre as somas (d_{ij}) e diferença máxima (d_M) para $k = 3$ tratamentos e $b = 3$ blocos

ℓ	$R_{1.}$	$R_{2.}$	$R_{3.}$	d_{12}	d_{13}	d_{23}	d_M	ℓ	$R_{1.}$	$R_{2.}$	$R_{3.}$	d_{12}	d_{13}	d_{23}	d_M
1	3	6	9	3	6	3	6	19	4	7	7	3	3	0	3
2	3	7	8	4	5	1	5	20	4	8	6	4	2	2	4
3	4	5	9	1	5	4	5	21	5	6	7	1	2	1	2
4	4	7	7	3	3	0	3	22	5	8	5	3	0	3	3
5	5	5	8	0	3	3	3	23	6	6	6	0	0	0	0
6	5	6	7	1	2	1	2	24	6	7	5	1	1	2	2
7	3	7	8	4	5	1	5	25	5	5	8	0	3	3	3
8	3	8	7	5	4	1	5	26	5	6	7	1	2	1	2
9	4	6	8	2	4	2	4	27	6	4	8	2	2	4	4
10	4	8	6	4	2	2	4	28	6	6	6	0	0	0	0
11	5	6	7	1	2	1	2	29	7	4	7	3	0	3	3
12	5	7	6	2	1	1	2	30	7	5	6	2	1	1	2
13	4	5	9	1	5	4	5	31	5	6	7	1	2	1	2
14	4	6	8	2	4	2	4	32	5	7	6	2	1	1	2
15	5	4	9	1	4	5	5	33	6	5	7	1	1	2	2
16	5	6	7	1	2	1	2	34	6	7	5	1	1	2	2
17	6	4	8	2	2	4	4	35	7	5	6	2	1	1	2
18	6	5	7	1	1	2	2	36	7	6	5	1	2	1	2

Tabela 25: Valores de diferenças entre somas de postos (d) e sua distribuição nula exata para as diferenças máximas e todas as diferenças ($k = 3$ tratamentos e $b = 3$ blocos)

d	$\#(d_M = d)$	$\#(d_M \geq d)$	$P(d_M \geq d)$	$\#(d_{ij} = d)$	$\#(d_{ij} \geq d)$	$P(d_{ij} \geq d)$
6	1	1	0,02778	1	1	0,00926
5	6	7	0,19444	6	7	0,06481
4	6	13	0,36111	12	19	0,17593
3	6	19	0,52778	14	33	0,30556
2	15	34	0,94444	27	60	0,55556
1	0	34	0,94444	36	96	0,88889
0	2	36	1,00000	12	108	1,00000

Exemplo 2: Considere como no Exemplo 1 da Seção 4.3.1, que tem-se três tratamentos ($k = 3$) e dois blocos ($b = 2$) e que pretende-se fazer comparações unilaterais entre os tratamentos (2 e 3) e o controle (1) ($\tau_i > \tau_1, i = 2, 3$). Considere-se as $(3!)^2 = 36$ configurações, as diferenças $d_{21} = R_2. - R_1.$ e $d_{31} = R_3. - R_1.$ e o máximo destas diferenças (d_M), obtendo-se os resultados encontrados na Tabela 26.

Com estes valores pode-se construir a tabela com a distribuição de frequências das diferenças e do máximo e as probabilidades necessárias para obter a distribuição nula para o caso considerado (Tabela 27).

Exemplo 3: Considere o mesmo caso do exemplo 1 desta seção, ou seja, três tratamentos ($k = 3$) e três blocos ($b = 3$) e que pretende-se obter comparações entre os tratamentos e o controle (tratamento 1) e considere comparações unilaterais. Tem-se $(3!)^2 = 216$ configurações distintas, de onde obtém-se as diferenças entre totais de tratamentos e o controle ($d_{21} = R_2. - R_1.$ e $d_{31} = R_3. - R_1.$) e o máximo destas diferenças (d_M), para apenas algumas configurações (Tabela 28).

Nota-se que neste caso seria necessária a construção das 216 configurações para a obtenção da tabela completa dos níveis de significância pois a comparação a ser feita não é bilateral, tornando a obtenção de todas elas um trabalho extremamente cansativo e tedioso.

Tabela 26: Somas dos postos (R_i , $i = 1, 2, 3$), diferenças entre as somas de tratamentos e controle (d_{i1}) e diferença máxima (d_M) ($k = 3$ tratamentos e $b = 3$ blocos)

ℓ	R_1	R_2	R_3	d_{21}	d_{31}	d_M	ℓ	R_1	R_2	R_3	d_{21}	d_{31}	d_M
1	2	4	6	2	4	4	19	3	5	4	2	1	2
2	2	5	5	3	3	3	20	3	6	3	3	0	3
3	3	3	6	0	3	3	21	4	4	4	0	0	0
4	3	5	4	2	1	2	22	4	6	2	2	-2	2
5	4	3	5	-1	1	1	23	5	4	3	-1	-2	-1
6	4	4	4	0	0	0	24	5	5	2	0	-3	0
7	2	5	5	3	3	3	25	4	3	5	-1	1	1
8	2	6	4	4	2	4	26	4	4	4	0	0	0
9	3	4	5	1	2	2	27	5	2	5	-3	0	0
10	3	6	3	3	0	3	28	5	4	3	-1	-2	-1
11	4	4	4	0	0	0	29	6	2	4	-4	-2	-2
12	4	5	3	1	-1	1	30	6	3	3	-3	-3	-3
13	3	3	6	0	3	3	31	4	4	4	0	0	0
14	3	4	5	1	2	2	32	4	5	3	1	-1	1
15	4	2	6	-2	2	2	33	5	3	4	-2	-1	-1
16	4	4	4	0	0	0	34	5	5	2	0	-3	0
17	5	2	5	-3	0	0	35	6	3	3	-3	-3	-3
18	5	3	4	-2	-1	-1	36	6	4	2	-2	-4	-2

Tabela 27: Diferenças entre somas de postos (d) e sua distribuição nula exata para as diferenças máximas e todas as diferenças ($k = 3$ e $b = 3$ - tratamentos *vs* controle)

d	$\#(d_M = d)$	$\#(d_M \geq d)$	$P(d_M \geq d)$	$\#(d_{ij} = d)$	$\#(d_{ij} \geq d)$	$P(d_{ij} \geq d)$
4	2	2	0,0556	2	2	0,0278
3	6	8	0,2222	8	10	0,1389
2	6	14	0,3889	8	18	0,2500
1	4	18	0,5000	8	26	0,3611
0	10	28	0,7778	20	46	0,6389
< 0	8	36	1,0000	24	72	1,0000

Tabela 28: Algumas configurações para $k = 3$ tratamentos e $b = 3$ blocos, diferenças (d_{i1}) entre somas de postos dos tratamentos (2 e 3) e o controle (1) e a diferença máxima para cada configuração considerada

Bloco1	Bloco2	Bloco3	d_{21}	d_{31}	d_M	Bloco1	Bloco2	Bloco3	d_{21}	d_{31}	d_M
123	123	123	3	6	6	213	123	213	-1	4	4
132	132	132	6	3	6	213	213	123	-1	4	4
123	123	132	4	5	5	231	132	231	4	-1	4
123	132	123	4	5	5	231	231	132	4	-1	4
123	132	132	5	4	5	123	231	132	4	2	4
132	123	123	4	5	5	123	132	231	4	2	4
132	123	132	5	4	5	123	132	213	2	4	4
132	132	123	5	4	5	123	213	132	4	2	4
123	213	123	1	5	5	132	213	123	2	4	4
123	123	213	1	5	5	132	123	213	2	4	4
132	132	231	5	1	5	132	231	123	4	2	4
132	231	132	5	1	5	132	123	231	4	2	4
213	123	123	1	5	5	213	123	132	2	4	4
213	132	132	1	5	5	213	132	123	2	4	4
123	213	213	-1	4	4	213	132	123	4	2	4
132	231	231	4	-1	4	213	123	132	4	2	4

Para os valores de diferença iguais a 6, 5 e 4 obteve-se, então, a distribuição de frequências das diferenças e do máximo e as probabilidades de significância que são mostradas na Tabela 29. Observa-se que para a comparação com o máximo de cada configuração existe significância apenas para a diferença igual a 6 (a máxima possível nesta configuração).

Tabela 29: Diferenças entre somas de postos (d) e sua distribuição nula exata para as diferenças máximas e todas as diferenças ($k = 3$ e $b = 3$ - tratamentos *vs* controle)

d	$\#(d_M = d)$	$\#(d_M \geq d)$	$P(d_M \geq d)$	$\#(d_{ij} = d)$	$\#(d_{ij} \geq d)$	$P(d_{ij} \geq d)$
6	2	2	0,00926	2	2	0,00463
5	12	14	0,06481	12	14	0,03241
4	18	32	0,14815	24	36	0,08333

4.4 Exemplos utilizando o programa

Foram utilizados exemplos encontrados na literatura para a aplicação e obtenção dos resultados referentes ao teste de Kruskal-Wallis e seu nível de significância exato (quando possível) ou seu nível de significância aproximado baseado em permutações aleatórias. Também são executados os testes de comparações múltiplas de todos os tratamentos ou dos tratamentos *versus* o controle, de acordo com a escolha do pesquisador. Os resultados obtidos foram comparados com os encontrados na literatura e nas tabelas, quando disponíveis. Nos casos em que não havia disponibilidade de tabelas exatas, os resultados obtidos através do programa foram comparados com as aproximações para grandes amostras utilizadas na literatura. Em alguns casos aqui exemplificados foram feitas todas as permutações e, portanto, calculada a probabilidade de significância exata, apenas com o objetivo de verificar a precisão do teste de permutação aleatória.

4.4.1 Caso 1: delineamento inteiramente casualizado com número de repetições iguais e poucos tratamentos

McDonald & Thompson (1967) apresentam dados referentes à percentagem de quebra, em quebras por unidade de comprimento, de quatro tipos de estames (*yarn*), com o objetivo de verificar se estes diferem em relação à referida taxa. Foram utilizadas quatro variedades (quatro tratamentos), com quatro repetições cada uma ($n_i = 4$, $i = 1, 2, 3, 4$). Os dados da Tabela 30 apresentam a taxa de quebra (x_{ij} , $i = 1, 2, 3, 4$ e $j = 1, 2, 3, 4$) e os respectivos postos (r_{ij}) para cada um dos $N = 16$ resultados obtidos do experimentos.

Os autores concluem, inicialmente, que existem diferenças significativas entre os tratamentos, baseados na estatística de Kruskal-Wallis pois, para $H = 9,8162$, o nível de significância, de acordo com Hollander & Wolfe (1999) está entre 0,0051 (para $H = 9,949$) e 0,0100 (para $H = 9,287$). Se a aproximação normal for utilizada, verifica-se que, para χ^2 com 3 graus de liberdade, este nível é de

Tabela 30: Dados relativos a quatro tratamentos e quatro repetições (McDonald & Thompson, 1967)

Repetições (j)	Variedades de Estames (Tratamentos)							
	x_{1j}	r_{1j}	x_{2j}	r_{2j}	x_{3j}	r_{3j}	x_{4j}	r_{4j}
1	70	16	12	2	10	1	29	6
2	52	14	18	3	43	11	31	7
3	51	13	35	8	28	5	41	10
4	67	15	36	9	26	4	44	12
$R_{i.}$	58		22		21		35	
$\bar{R}_{i.}$	14,50		5,50		5,25		8,75	

0,020195. Concluem também, baseados nos postos atribuídos aos dados, que a variedade 1 difere significativamente (ao nível de 5%) das variedades 2 e 3, mas que entre os pares de variedades (1,4), (2,3), (2,4) e (3,4) não existem diferenças significativas.

Com o objetivo de comparar resultados foram feitas três execuções do programa: na primeira, todas as permutações foram obtidas, calculando-se o nível de significância exato; nas duas execuções seguintes, utilizou-se o procedimento de permutações aleatórias com o objetivo de verificar a precisão de tal procedimento, bem como a comparação com os resultados de McDonald & Thompson (1967).

1) *Execução com todas as combinações:*

O programa construiu $\frac{16!}{4!4!4!4} = 15.765.750$ configurações em cerca de 5 minutos (aproximadamente 3.100.000 combinações por minuto) obtendo o valor da estatística de Kruskal-Wallis ($H = 9,816176$) e seu nível de significância exato (0,005780). Foram realizadas também as comparações entre todos os tratamentos e a Tabela 31 fornece: a diferença entre os tratamentos, a probabilidade de significância denominada DW e a probabilidade de significância *Geral*, conforme visto em (4.2). Foram acrescentados também os resultados obtidos por McDonald & Thompson (1967) para comparações.

Verifica-se, inicialmente, que o resultado obtido através do programa para a estatística de Kruskal-Wallis é consistente, estando dentro do intervalo obtido

Tabela 31: Diferenças entre os totais dos postos de tratamentos (Dif.), níveis de significância para as comparações entre todos os tratamentos (D-W e Geral) e resultados obtidos por McDonald & Thompson (1967).

Tratamentos	Dif.	D-W	Geral	McDonald-Thompson
1-2	(36)	0,024856	0,004256	significativo (0,026)
1-3	(37)	0,017480	0,002957	significativo ($0,012 < n.s. < 0,026$)
1-4	(23)	0,359032	0,095618	não significativo
2-3	(1)	0,999851	0,971031	não significativo
2-4	(13)	0,814598	0,367579	não significativo
3-4	(14)	0,776327	0,329861	não significativo

através da tabela de Hollander & Wolfe (1999) (entre 0,0051 e 0,0100). É perceptível também a diferença entre o valor da probabilidade de significância exata (0,005780) e o da aproximada através do χ^2 com 3 g.l. (0,020195). Caso o nível de significância adotado fôsse de 1%, esta diferença faria com que houvesse rejeição da hipótese nula se o valor exato fosse utilizado, e não rejeição se a aproximação normal fosse adotada. Isto mostra a fragilidade da aproximação, devendo a mesma ser utilizada com muita cautela, em especial para pequenas amostras.

Quanto aos testes de comparações múltiplas, no caso dos tratamentos 1 e 2, encontrou-se uma pequena diferença entre as probabilidades de significância obtidas através do programa (0,0248) e as da tabela de McDonald & Thompson (1967) (0,026). Nos outros casos os resultados puderam ser comparados apenas de forma geral devido à ausência de valores exatos na tabela.

Deve-se notar que os resultados obtidos pelo programa, foram consistentes nos casos em que nem todos os tratamentos têm o mesmo número de repetições e naqueles com número pequeno de repetições iguais, onde foram construídas todas as configurações. Além disso, os resultados obtidos ao se trabalhar com permutações aleatórias (como pode ser verificado nos itens seguintes desta seção) mostraram-se bastante estáveis, não havendo, pois, motivos para duvidar dos resultados gerados a

partir de amostras com o mesmo número de repetições.

2) *Execuções aleatórias:*

Foram feitas duas execuções do programa, aqui denominadas *Aleat1* e *Aleat2*, cada uma delas com 1.000.000 de combinações aleatórias. Foram calculadas, em cada execução, o nível de significância do Teste de Kruskal-Wallis ($n.s.(KW)$) aproximado e obteve-se então os erros em relação ao teste exato, dados por

$$Erro(\%) = 100 \frac{|n.s.exato(KW) - n.s.Aleati(KW)|}{n.s.exato(KW)}$$

O mesmo procedimento foi efetuado em relação às comparações múltiplas, obtendo-se o erro em relação ao teste proposto por Damico & Wolfe (1987) (*DW*) e ao teste, aqui denominado *Geral*, para cada uma das execuções. O tempo de execução foi de cerca de 45 segundos para cada uma das configurações aleatórias e obteve-se, como anteriormente, $H = 9,816176$.

Os níveis de significância para o teste de Kruskal-Wallis foram de 0,005701 e 0,005805, nas execuções aleatórias 1 e 2, respectivamente. Portanto, os erros, em relação ao nível de significância exato foram de 1,37% e 0,43%, respectivamente. Os resultados das execuções aleatórias 1 e 2 assim como os erros em relação ao teste exato nos casos de comparações múltiplas entre todos os tratamentos são mostrados nos Tabelas 32 e 33, respectivamente.

Nota-se que o teste de permutações aleatórias pode ser um bom substituto nos casos em que tem-se um número de configurações muito grande. Os erros máximos apresentados para os níveis de significância foram de 1,4% no caso da estatística de Kruskal-Wallis e 0,7% para as comparações múltiplas. Os erros maiores ocorrem nos valores em que as probabilidades de significância são menores, como era esperado, mas não são significativos.

Tabela 32: Níveis de significância exatos (DW) para as comparações entre todos os tratamentos e obtidos através das execuções aleatórias 1 e 2 (DW Aleat1 e DW Aleat2), com seus respectivos erros percentuais.

Tratamentos	DW Exato	DW Aleat1	Erro (%) Aleat1	DW Aleat2	Erro (%) Aleat2
1-2	0,024856	0,024811	0,1810	0,024898	0,1690
1-3	0,017480	0,017355	0,7151	0,017427	0,3032
1-4	0,359032	0,359004	0,0078	0,358916	0,0323
2-3	0,999851	0,999864	0,0013	0,999832	0,0019
2-4	0,814598	0,814953	0,0436	0,814822	0,0275
3-4	0,776327	0,776772	0,0689	0,776512	0,0354

Tabela 33: Níveis de significância exatos (Geral) para as comparações entre todos os tratamentos e obtidos através das execuções aleatórias 1 e 2 (Geral Aleat1 e Geral Aleat2), com seus respectivos erros percentuais.

Tratamentos	Geral Exato	Geral Aleat1	Erro Geral Aleat1 (%)	Geral Aleat2	Erro Geral Aleat2 (%)
1-2	0,004256	0,004252	0,0940	0,004258	0,0470
1-3	0,002957	0,002938	0,6425	0,002947	0,3381
1-4	0,095618	0,095479	0,1454	0,095432	0,1945
2-3	0,971031	0,970932	0,0102	0,971071	0,0041
2-4	0,367579	0,367807	0,0620	0,367724	0,0394
3-4	0,329861	0,330044	0,0555	0,329981	0,0364

4.4.2 Caso 2: delineamento inteiramente casualizado com número de tratamentos e de repetições grande

Conover (1999, p.291) apresenta dados de quatro tratamentos com números de repetições diferentes ($n_1 = 9; n_2 = 10; n_3 = 7; n_4 = 8$) totalizando $N = 34$ repetições o que proporcionaria $1,1033 \times 10^{18}$ permutações (demoraria aproximadamente 6361 séculos para serem efetuadas pelo programa criado). Foram construídas 1.000.000 (um milhão) de configurações aleatórias, com um tempo de execução de cerca de dois minutos.

Este exemplo se constitui no caso mais geral possível pois tem-se tratamentos com números de repetições diferentes e ainda empates entre elementos dentro do mesmo tratamento e entre elementos de tratamentos diferentes. Conover (1999) calcula o valor da estatística de Kruskal-Wallis $H = 25,464371$ (com a correção para empates tem-se $H_c = 25,6288$) e conclui que existem diferenças entre tratamentos a um nível de significância menor que 0,1% (este nível, utilizando a distribuição de χ^2 com 3 graus de liberdade é de $0,00001235 = 0,0012\%$).

Para o cálculo do nível de significância aproximado da estatística de Kruskal-Wallis através do teste de permutações aleatórias, foram feitas duas simulações, cada uma delas com 1.000.000 de permutações, resultando em ambas um nível de significância nulo (em nenhuma delas um valor maior ou igual a H foi obtido). Como as médias de tratamentos são $\bar{R}_1 = 21,8333$, $\bar{R}_2 = 15,3000$, $\bar{R}_3 = 29,5714$ e $\bar{R}_4 = 4,8125$, percebe-se, intuitivamente que o tratamento 4 tem média bem menor que a dos outros três tratamentos; mas será esta diferença significativa?

Para efetuar as comparações múltiplas entre todos os tratamentos, Conover (1999) ignorou os empates e calculou, para cada par de tratamentos i, j , a diferença absoluta entre as médias dos tratamentos e comparou com as diferenças mínima significativas dada por $dms_{ij} = t_{1-(\alpha/2)} \left(S^2 \frac{N-1-H}{N-k} \right)^{1/2} (1/n_i + 1/n_j)^{1/2}$ onde $S^2 = [N(N+1)]/12$, ou seja, $dms_{ij} = (2,041)(24,911)^{1/2} (1/n_i + 1/n_j)^{1/2}$ (ver Tabela 34) concluindo que todos os tratamentos diferem entre si ao nível de significância

Tabela 34: Níveis de significância aproximados para as diferenças entre todos os tratamentos obtidos pelos métodos dados em Campos (1983) e em Conover (1999).

Populações	$ \bar{R}_i - \bar{R}_j $	Conover (1999)		Campos (1983)	
		$dms_{ij}(5\%)$	n.m.s	$dms_{ij}(5\%)$	n.m.s
1-2	6,533	4,681	0,007263	11,810	0,1533
1-3	7,738	5,134	0,004073	12,949	0,1231
1-4	17,021	4,950	0,000000	12,484	0,0004
2-3	14,271	5,020	0,000002	12,661	0,0036
2-4	10,488	4,832	0,000101	12,187	0,0264
3-4	24,759	5,272	0,000000	13,297	0,0000

Tabela 35: Níveis de significância aproximados DW e Geral para as diferenças entre todos os tratamentos obtidos em duas execuções do programa desenvolvido (Aleat1 e Aleat2).

Populações	D-W	Geral	D-W	Geral
	Aleat1	Aleat1	Aleat2	Aleat2
1-2	0,547632	0,181914	0,547230	0,181583
1-3	0,392619	0,112362	0,391982	0,112014
1-4	0,001085	0,000182	0,001082	0,000183
2-3	0,012740	0,002290	0,012725	0,002283
2-4	0,132946	0,029198	0,132722	0,029221
3-4	0,000000	0,000000	0,000000	0,000000

de 5%. Entretanto, os resultados obtidos utilizando-se a aproximação para grandes amostras de acordo com Campos (1983) ou ainda com o programa desenvolvido neste trabalho, diferem sensivelmente dos apresentados em Conover (1999). Utilizando a aproximação dada em Campos (1983), ou seja, $d.m.s. = z_{\alpha/[k(k-1)]} \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$, tem-se, para $\alpha' = 0,06/12 = 0,005$ os resultados apresentados na Tabela 34.

Os níveis de significância obtidos em duas permutações aleatórias gerados pelo *Programa IC* são mostrados na Tabela 35. Os resultados apresentados nas Tabelas 34 e 35 mostram que a aproximação utilizada em Conover (1999) detecta diferenças imperceptíveis pelos outros métodos, levando a conclusões completamente

diferentes daquelas obtidas tanto pela aproximação normal como pelo programa desenvolvido. Os níveis de significância *Geral* obtidos através do teste de permutações aleatórias (Tabela 35) se assemelham com àqueles gerados pela aproximação normal para tratamentos com números de repetições diferentes utilizada em Campos (1983) (Tabela 34).

A aproximação pelas permutações aleatórias mostra-se novamente consistente, com os resultados obtidos por ambas aproximações (Aleat1 e Aleat2) bastante próximos, mesmo neste caso onde o número possível de configurações totais é extremamente elevado. Como no exemplo anterior, o método proposto por Damico & Wolfe (1987) para comparações de todos os tratamentos mostra-se extremamente conservativo em relação aos outros métodos.

4.4.3 Caso 3: delineamentos inteiramente casualizados com número de repetições iguais - casos extremos

- Delineamento inteiramente casualizado: $k = 4$ tratamentos e $n_i = 3$ ($i = 1, 2, 3, 4$)
- $\frac{12!}{(3!3!3!3!)}$ configurações possíveis. Considere as configurações (extremas) mostradas na Tabela 36.

Tabela 36: Algumas configurações extremas considerando $k = 4$ tratamentos e $n_i = 3$, ($i = 1, 2, 3, 4$) e as respectivas somas de postos de tratamentos (R_i).

	Configuração 1				Configuração 2				Configuração 3				Configuração 4			
	T_1	T_2	T_3	T_4	T_1	T_2	T_3	T_4	T_1	T_2	T_3	T_4	T_1	T_2	T_3	T_4
	1	4	7	10	1	2	7	8	1	4	5	6	1	2	7	8
	2	5	8	11	3	4	9	10	2	8	7	9	4	3	10	9
	3	6	9	12	5	6	11	12	3	12	11	10	5	6	11	12
R_i	6	15	24	33	9	12	27	30	6	24	23	25	10	11	28	29

- Resultados obtidos utilizando-se o programa
- Para a configuração 4 foram obtidos os níveis de significância para três contrastes ortogonais Y_1 , Y_2 e Y_3 :

$$Y_1 = -1T_1 - 1T_2 + 1T_3 + 1T_4 \quad n.s.(Y_1) = 0,002165$$

$$Y_2 = -1T_1 + 1T_2 + 0T_3 + 0T_4 \quad n.s.(Y_2) = 0,961255$$

$$Y_3 = 0T_1 + 0T_2 - 1T_3 + 1T_4 \quad n.s.(Y_3) = 0,961255$$

Pelos resultados obtidos percebe-se que as comparações realizadas com o máximo das diferenças entre tratamentos é extremamente conservativa, detectando diferenças apenas no caso extremo da

Tabela 37: Níveis de significância exatos para as diferenças entre todos os tratamentos obtidos através do programa para as configurações 1 e 2 da Tabela 36.

Pares de Tratamentos	Configuração 1			Configuração 2		
	Dif.	DW	Geral	Dif.	DW	Geral
1 vs 2	9	0,805519	0,355303	3	0,995325	0,789177
1 vs 3	18	0,203766	0,042533	18	0,203766	0,042533
1 vs 4	27	0,000649	0,000108	21	0,071818	0,012446
2 vs 3	9	0,805519	0,355303	15	0,391493	0,104221
2 vs 4	18	0,203766	0,042533	18	0,203766	0,042533
3 vs 4	9	0,805519	0,355303	3	0,995325	0,789177

configuração 1. Mesmo no caso da configuração 4, onde tem-se os tratamentos 1 e 2 com somas de postos pequenas e os tratamentos 3 e 4 com somas de postos grandes, não são detectadas diferenças significativas pelo método *DW* proposto por Damico & Wolfe (1987). O método *Geral*, que faz as comparações com todas as diferenças entre tratamentos detecta algumas destas diferenças, sendo portanto mais recomendável.

O uso de contrastes, visto aqui para a configuração 4 parece ser bastante interessante no intuito de mostrar diferenças entre grupos de tratamentos.

4.4.4 Caso 4: Delineamentos em Blocos Casualizados com pequeno número de repetições

Campos (1983, p.238 e 253) apresenta um exemplo de um ensaio de adubação com $k = 4$ tratamentos e $b = 4$ blocos, onde o primeiro tratamento é a testemunha. Os resultados, já transformados em postos, estão apresentados na

Tabela 38: Níveis de significância exatos para as diferenças entre todos os tratamentos obtidos através do programa para as configurações 3 e 4 da Tabela 36.

Pares de Tratamentos	Configuração 3			Configuração 4		
	Dif.	DW	Geral	Dif.	DW	Geral
1 vs 2	18	0,203766	0,042533	1	1,000000	0,961255
1 vs 3	17	0,261818	0,059307	18	0,203766	0,042532
1 vs 4	19	0,154156	0,029762	19	0,154156	0,029762
2 vs 3	1	1,000000	0,961255	17	0,261818	0,059307
2 vs 4	1	1,000000	0,961255	18	0,203766	0,042532
3 vs 4	2	0,997922	0,872511	1	1,000000	0,961255

Tabela 39: Níveis de significância exatos para as diferenças entre os tratamentos e a testemunha obtidos através do programa para a configuração 3 da Tabela 36.

Comparações	Dif.	DW	Geral
1 vs 2	18	0,0573701	0,021266
1 vs 3	17	0,0771266	0,029654
1 vs 4	19	0,0415909	0,014881

Tabela 40.

Os níveis de significância denominados *CompMax*, obtidos através do *Programa BC* coincidem com aqueles apresentados nas tabelas em Campos (1983) e Hollander & Wolfe (1999) e referem-se às comparações entre as diferenças obtidas com os dados do experimento e a diferença máxima em cada configuração. Os níveis de significância denominados *Geral* refere-se à comparação entre as diferenças dos dados experimentais com todas as diferenças em todas as configurações. Obviamente, o nível de significância *CompMax* é mais conservativo que o *Geral*, não detectando, por exemplo, diferenças entre o tratamento 4 e a testemunha.

Tabela 40: Postos e somas dos postos (R_i) referentes a um ensaio com $k = 4$ tratamentos e $b = 4$ blocos.

Blocos	Trat.1	Trat.2	Trat.3	Trat.4
I	1	2	3	4
II	2	1	4	3
III	1	4	3	2
IV	1	2	4	3
R_i	5	9	14	12

Utilizando o programa referente ao delineamento em blocos casualizados, obtém-se o valor da estatística de Friedman ($S = 6,9$) e seu nível de significância ($p(S) = 0,0677807$). Foi solicitado também os níveis de significância para as $k-1 = 3$ comparações unilaterais entre os tratamentos (2, 3 e 4) e a testemunha. Os resultados obtidos pelo programa são apresentados na Tabela 41.

Se considerássemos as comparações entre todos os tratamentos (bilateral), os resultados, apresentados na Tabela 42, mostram que, utilizando o nível de significância *CompMax* não foram detectadas diferenças entre os tratamentos, enquanto que o *Geral* constata diferenças significativas entre os tratamentos 1 e 3.

Tabela 41: Níveis de significância exatos para as diferenças entre os tratamentos e a testemunha obtidos através do programa.

Trat. vs Test.	CompMax	Geral
2 vs 1	0,351852	0,176408
3 vs 1	0,020490	0,007186
4 vs 1	0,091761	0,036990

Tabela 42: Níveis de significância exatos para as diferenças entre todos tratamentos obtidos através do programa.

Trat. vs Test.	CompMax	Geral
1 vs 2	0,799696	0,352811
1 vs 3	0,077691	0,014371
1 vs 4	0,292607	0,073978
2 vs 3	0,632595	0,224826
2 vs 4	0,900391	0,508681
3 vs 4	0,992405	0,691551

4.4.5 Caso 5: Delineamentos em Blocos Casualizados utilizando as permutações aleatórias

Os dados em postos apresentados na Tabela 43 são fictícios, onde tem-se $k = 5$ tratamentos em $b = 6$ blocos, totalizando $N = 30$ parcelas.

Foi utilizado o teste de permutações aleatórias, com o intuito de verificar a sua precisão em relação ao teste exato e os resultados das comparações múltiplas entre todos os tratamentos são apresentados na Tabela 44. O valor da estatística de Friedman foi $S = 17,4667$ com um nível de significância de 0,000108.

As tabelas existentes mostram que para $k = 5$ tratamentos e $b = 6$ blocos, diferenças de 15, 16 e 17 unidades têm níveis de significância iguais a 0,049, 0,028 e 0,013, respectivamente, que são bem próximos dos obtidos pelas configurações aleatórias.

Tabela 43: Postos atribuídos a $k = 5$ tratamentos e $b = 6$ blocos.

Blocos	Trat.1	Trat.2	Trat3	Trat.4	Trat.5
I	1	2	5	4	3
II	1	3	2	5	4
III	1	2	3	5	4
IV	2	1	3	4	5
V	2	1	3	4	5
VI	2	1	4	3	5
R_i	9	10	20	25	26

Tabela 44: Níveis de significância aproximados para as diferenças entre todos os tratamentos obtidos através do programa pelo teste de permutações aleatórias ($k = 5$ e $b = 6$).

Trat. vs Test.	Dif.	CompMax	Geral
1 vs 2	1	0,999768	0,928861
1 vs 3	11	0,314624	0,053460
1 vs 4	16	0,025660	0,002834
1 vs 5	17	0,012122	0,001276
2 vs 3	10	0,426296	0,082411
2 vs 4	15	0,049119	0,005753
2 vs 5	16	0,025660	0,002834
3 vs 4	5	0,931313	0,419560
3 vs 5	6	0,866283	0,322417
4 vs 5	1	0,999768	0,928861

4.4.6 Caso 6: Delineamentos em Blocos Casualizados - um caso com empates

Considere que existem $k = 4$ tratamentos e $b = 3$ blocos e que ocorram empates entre alguns elementos do delineamento, ou seja, que a configuração resultante dos dados experimentais seja a da Tabela 45.

Tabela 45: Postos atribuídos a $k = 4$ tratamentos e $b = 3$ blocos com ocorrência de empates.

Blocos	Trat.1	Trat.2	Trat3	Trat.4
I	1,0	2,5	2,5	4
II	1,0	2,5	2,5	4
III	1,5	1,5	3,5	3,5
R_i	3,5	6,5	8,5	11,5

Com estes postos obteve-se, através do programa, o valor da estatística de Friedman, sem a correção para empates ($S = 6,8$) com um nível de significância exato igual a 0,027778. Apesar do programa não fazer a correção, o nível de significância obtido considera todos os empates. Com a correção para empates tem-se $S_C = 7,8462$, com um nível de significância (aproximado) entre 0,017 (para $S = 8,2$) e 0,033 (para $S = 7,4$). Utilizando a aproximação através da distribuição qui-quadrado, o nível de significância seria de 0,0493.

Para as comparações múltiplas entre todos os tratamentos as tabelas apresentam os níveis de significância iguais a 0,049 e 0,007 para as diferenças iguais a 8 e 9, respectivamente. Os níveis para diferenças iguais a 8, exatos utilizando permutações sistemáticas e aproximados pelas permutações aleatórias são, respectivamente, 0,02778 e 0,028072. A diferença máxima que pode ser obtida com a configuração dada na Tabela 45 é igual 8 e portanto, uma diferença igual a 9 não pode ser comparada.

No caso das comparações múltiplas entre os tratamentos (2, 3 e 4) e o controle (tratamento 1), tem-se, na tabela para $k = 4$ e $b = 3$ blocos que os níveis de significância para as diferenças d (ns_d) são: $ns_6 = 0,1046$, $ns_7 = 0,0460$, $ns_8 = 0,0122$ e $n.s._9 = 0,0017$. Utilizando o programa obtém-se, para diferenças iguais a 8, os níveis de significância exato ($nse_8 = 0,006944$) e aproximado ($nsa_8 = 0,006863$). As diferenças iguais a 6, 7 e 9 não ocorrem nesta configuração de dados.

Pelos resultados apresentados, verifica-se que a aproximação utilizando a correção para empates, para a estatística de Friedman tem uma aproximação boa, mesmo num caso em que o número de empates é significativo. Já a aproximação pela distribuição de qui-quadrado mostrou-se ruim neste caso, onde tem-se poucas repetições e tratamentos.

Em relação às comparações múltiplas, não existem aproximações no caso de ocorrência de empates e o uso das tabelas existentes para a obtenção dos níveis de significância não é aconselhável, como mostram os resultados acima, devido às grandes discrepâncias entre os valores exatos e os valores das tabelas, que foram concebidas para casos em que não ocorrem empates. A aproximação através das permutações aleatórias mostrou-se uma boa alternativa nos casos em que a obtenção de todas as permutações sistemáticas não for possível, resultando em valores bem próximos dos exatos (erros próximos a 1%), tanto para comparações entre todos os tratamentos como para comparações entre tratamentos e a testemunha.

5 CONCLUSÕES

A abrangência das tabelas sempre foi um ponto fraco dos testes não-paramétricos. Mesmo nos livros-textos mais recentes e com grande número de páginas dedicadas às tabelas direcionadas aos diversos testes, não é possível obter resultados, muitas vezes simples, com poucas repetições ou tratamentos, quando ocorrem empates. Mesmo quando não ocorrem empates, as tabelas deixam a desejar pois elas são incompletas, não fornecendo os níveis de significância para a maioria dos casos. Outro ponto a destacar refere-se às aproximações sugeridas pelos autores. Nos casos dos testes de Kruskal-Wallis e de Friedman, em geral utilizam-se a aproximação através da distribuição de χ^2 , que mostra-se fraca nos casos em que o número de repetições é diferente para os tratamentos ou ainda quando ocorrem empates. Outros autores consideram mais adequada a aproximação através da distribuição F de Fisher, que também é falha nos casos citados.

A situação é mais crítica no caso de comparações múltiplas (entre todos os tratamentos ou entre os tratamentos e o controle). Aqui, além das dificuldades anteriores, como a falta de tabelas exatas nos casos em que ocorrem empates e sua pequena abrangência, várias aproximações sugeridas. A escolha de qual aproximação é mais adequada para cada caso não é tarefa muito fácil, não havendo, até o momento, consenso entre os pesquisadores.

Outro fator, associado às comparações múltiplas refere-se ao tipo de erro a ser utilizado (*comparisonwise* ou *experimentalwise*) e o método associado a cada tipo de erro (atribuição de postos de forma conjunta ou atribuição de pos-

tos apenas aos tratamentos que estão sendo comparados). O uso da taxa de erro *comparisonwise*, apesar de ser defendida por vários autores, pressupõe a quebra da estrutura montada para o delineamento experimental pois cada par de tratamentos é tratado isoladamente, com atribuição de postos àqueles que estão sendo comparados. Esta quebra de estrutura não parece adequada aos propósitos da análise de variância, onde busca-se fazer uma análise conjunta dos tratamentos envolvidos. O uso do método *stepwise*, onde inicialmente analisa-se os tratamentos mais discrepantes dentro do conjunto total de tratamentos, e, caso a diferença seja significativa, faz-se as comparações intermediárias, requer também a quebra da estrutura montada para a análise de variância. A utilização da taxa de erro *experimentalwise* aproveita toda a estrutura de análise mas, por outro lado, requer grandes diferenças para detectar significância entre tratamentos, em especial quando o número de tratamentos aumenta.

Os resultados obtidos por McDonald & Thompson (1967), complementados por Damico & Wolfe (1987) para as comparações múltiplas entre todos os tratamentos em delineamentos inteiramente casualizados consideram a taxa de erro *experimentalwise*, realizando as comparações com a diferença máxima (entre as $k(k - 1)/2$ diferenças) de cada configuração. Este tipo de comparação leva a resultados bem mais conservativos, dificultando sobremaneira a detecção de diferenças significativas entre tratamentos. O mesmo problema ocorre em relação às comparações entre tratamentos e o controle neste mesmo tipo de delineamento, cujas tabelas foram obtidas em Damico & Wolfe (1989), onde as comparações são feitas com o máximo das $k - 1$ comparações realizadas em cada configuração. Os testes de comparações múltiplas entre todos os tratamentos e entre os tratamentos e o controle para os delineamentos em blocos casualizados também são baseados nas comparações com o máximo das diferenças de cada configuração. Em todos os casos considerados, os níveis de significância exatos podem diferir sensivelmente das aproximações recomendadas em cada

caso, em especial se estes níveis são pequenos, levando a conclusões totalmente discrepantes.

Atenção especial deve ser dada aos dois níveis de significância obtidos através do programa para as comparações múltiplas (*DW* ou *CompMax* e *Geral*). Os níveis de significância obtidos através da comparação com a diferença máxima dentro de cada configuração (*DW* ou *CompMax*) são utilizados em Campos (1983) e Hollander & Wolfe (1999), dentre outros, mas são extremamente conservativos. O nível de significância obtido através da comparação com todas as diferenças dentro de cada configuração (*Geral*), além de ser menos conservativo, fornece resultados próximos daqueles obtidos com a aproximação normal, como pode ser verificado nos casos apresentados na Seção 4.4.

Os programas apresentados oferecem uma alternativa válida em todos os casos, tanto para os delineamentos inteiramente casualizados como para os em blocos casualizados, pois trabalham com todas as configurações de atribuição de postos aos tratamentos (nos casos em que o número total de configurações permite). Quando o número total de configurações torna-se muito grande para a capacidade computacional, trabalha-se com os testes de permutações aleatórias, onde cada configuração é obtida pela aleatorização dos postos aos tratamentos, adequado ao delineamento com que se está trabalhando. Assim, estes programas oferecem as seguintes vantagens:

1. Apresentam níveis de significância exatos para os casos de pequenas amostras, tanto para o teste de Kruskal-Wallis como para o teste de Friedman, podendo substituir, com vantagens, as exaustivas tabelas.
2. Oferecem os valores exatos dos níveis de significância para as comparações múltiplas entre todos os tratamentos e para os tratamentos *versus* o controle, nos delineamentos inteiramente casualizados e em blocos casualizados, nos casos de pequenas amostras.

3. Podem ser utilizados também nos casos em que ocorrem empates, fornecendo níveis de significância exatos também nestes casos, tanto para o teste adequado ao delineamento (Kruskal-Wallis ou Friedman) como para as comparações múltiplas e os contrastes.
4. Nos casos de comparações múltiplas fornecem dois níveis de significância distintos: o primeiro, denominado *DW* (no delineamento inteiramente casualizado) ou *CompMax* (no delineamento em blocos casualizados) refere-se à comparação entre a diferença obtida e o máximo das diferenças dentro de cada configuração; o segundo, denominado *Geral* em ambos os delineamentos, utiliza como fator de comparação todas as diferenças dentro de cada configuração.
5. Possibilita a obtenção dos níveis de significância para os contrastes escolhidos pelo usuário, entre médias ou totais de tratamentos, dando uma maior versatilidade para análise dos dados.
6. Fornece resultados aproximados para os casos de grandes amostras, tanto para o Teste de Kruskal-Wallis como para o teste de Friedman, utilizando o teste de permutações aleatórias, que mostraram-se mais confiáveis do que as aproximações em geral utilizadas nestes casos. Nos testes efetuados, os erros, em relação ao valor exato dos níveis de significância, foram menores do que 2%.
7. Utiliza-se do teste de permutações aleatórias para fornecer níveis de significância aproximados para comparações múltiplas entre todos os tratamentos, entre tratamentos e o controle e para os contrastes. Os erros cometidos ao utilizar tal aproximação foram baixos (próximos a 1,0%).

A metodologia utilizada para a obtenção de níveis exatos de significância adotada neste trabalho pode ser estendida a outros tipos de ensaios e para delineamentos mais complicados (delineamentos com mais de dois fatores, com

ou sem interação, delineamentos em parcelas subdivididas, dentre outros), bastando verificar como se processa a distribuição dos tratamentos às unidades experimentais e construir as configurações possíveis a partir desta atribuição inicial. Cuidados especiais devem ser tomados nos casos em que ocorrem interações, mas métodos de análise que abrangem tais casos estão disponíveis na literatura.

REFERÊNCIAS BIBLIOGRÁFICAS

- BHAPKAR, V.P. A nonparametric test for the problem of several samples. **Annals of Mathematical Statistics**, v.32, p.1109-1117, 1961.
- BRADBURY, I. Analysis of variance versus randomization tests - a comparison. **British Journal of Mathematical and Statistical Psychology**, v.40, p.177-187, 1987a.
- BRADBURY, I. Reply to White & Still **British Journal of Mathematical and Statistical Psychology**, v.40, p.192-193, 1987b.
- BÜNING, H. Adaptative Jonckheere-type tests for ordered alternatives. **Journal of Applied Statistics**, v.26, n.5, p.541-551, 1999.
- CAMPBELL, G.; SKILLINGS, J.H. Nonparametric stepwise multiple comparison procedures. **Journal of the American Statistical Association**, v.80, n.392, p.998-1003, 1985.
- CAMPOS, H. de. **Estatística experimental não-paramétrica**. 4.ed. Piracicaba: FEALQ, 1983. 349 p.
- CONOVER, W.J. Rank tests for one sample, two sample, and k samples without the assumption of a continuous distribution function. **The Annals of Statistics**, v.1, n.6, p.1105-1125, 1973.
- CONOVER, W.J. **Practical nonparametric statistics**. 3.ed. New York: John Wiley, 1999. 584 p.
- CONOVER, W.J.; IMAN, R.L. Rank transformations as a bridge between parametric and nonparametric statistics. **The American Statistician**, v.35, n.3, p.124-129, 1981.

- CRITCHLOW, D.E.; FLIGNER, M.A. On distribution-free multiple comparisons in the one-way analysis of variance. **Communications in Statistics - Theory and Methods**, v.20, n.1, p.127-139, 1991.
- DAMICO, J.A.; WOLFE, D.A. Extended tables of the exact distribution of a rank statistic for all treatments multiple comparisons in one-way layout designs. **Communications in Statistics - Theory and Methods**, v.16, n.8, p.2343-2360, 1987.
- DAMICO, J.A.; WOLFE, D.A. Extended tables of the exact distribution of a rank statistic for treatments versus control multiple comparisons in one-way layout designs. **Communications in Statistics - Theory and Methods**, v.18, n.9, p.3327-3353, 1989.
- DUNN, O.J. Multiple comparisons using rank sums. **Technometrics**, v.6, n.3, p.241-252, 1964.
- EDGINGTON, E.S. **Randomization tests**. 3.ed. New York: Marcel Dekker, 1995. 409 p.
- FAIRLEY, D.; PEARL, D.K. The Bahadur efficiency of paired versus joint ranking procedures for pairwise multiple comparisons. **Communications in Statistics - Theory and Methods**, v.13, n.12, p.1471-1481, 1984.
- FAWCETT, R.F.; SALTER, K.C. A Monte Carlo study of the F test and three tests based on ranks of treatment effects in randomized block designs. **Communications in Statistics - Simulation and Computation**, v.13, n.2, p.213-225, 1984.
- FLIGNER, M.A. Pairwise versus joint ranking: another look at the Kruskal-Wallis statistic. **Biometrika**, v.72, n.3, p.705-709, 1985.
- GROGGER, D.J.; SKILLINGS, J.H. Distribution-free tests for main effects in multifactor designs. **The American Statistician**, v.40, n.2, p.99-102,

1986.

HETTMANSPERGER, T.P. Non-parametric inference for ordered alternatives in a randomized block design. **Psychometrika**, v.40, n.1, p.53-62, 1975.

HOCHBERG, Y.; TAMHANE, A.C. **Multiple comparisons procedures**. New York: John Wiley, 1987. 450 p.

HOLLANDER, M.; WOLFE, D.A. **Nonparametric statistical methods**. 2.ed. New York: John Wiley, 1999. 787 p.

HOPPE, F.M. (Ed.) **Multiple comparisons, selection and applications in biometry**. New York: Marcel Dekker, 1993. 558 p.

HOUSE, D.E. A nonparametric version of Williams' Test for a randomized block design. **Biometrics**, v.42, n.1, p.187-190, 1986.

HSU, J.C. **Multiple Comparisons: theory and methods**. London: Chapman & Hall, 1996. 277p.

JONCKHEERE, A.R. A distribution-free k-sample test against ordered alternatives. **Biometrika**, v.41, n., p.133-145, 1954.

KENDALL, M.G.; STUART, A. **The advanced theory of statistics**. London: Charles Griffin & Company, 1967. v.2, 690p.

KEPNER, J.L.; ROBINSON, D.H. A distribution-free rank test for ordered alternatives in randomized complete block designs. **Journal of the American Statistical Association**, v.79, n.385, p.212-217, 1984.

KRUSKAL, W.H.; WALLIS, W.A. Use of ranks in one-criterion variance analysis. **Journal of the American Statistical Association**, v.47, n.260, p.583-621, 1952.

LEHMANN, E.L.; D'ABRERA, H.J.M. **Nonparametrics: statistical based on ranks**. San Francisco: Holden-Day, 1975. 457 p.

- LUDBROOCK, J.; DUDLEY, H. Why permutation tests are superior to t and F tests in Biomedical Research. **The American Statistician**, v.52, n.2, p.127-132, 1998.
- MACK, G.A.; SKILLINGS, J.H. A Friedman-type rank test for main effects in a two-factor ANOVA. **Journal of the American Statistical Association**, v.75, n.372, p.947-951, 1980.
- MACK, G.A.; WOLFE, D.A. K-sample rank tests for umbrella alternatives. **Journal of the American Statistical Association**, v.78, n.373, p.175-181, 1981.
- MANOUKIAN, E.B. **Mathematical nonparametric statistics**. Glasgow: Gordon and Breach Science, 1986. 326p.
- MARITZ, J.S. **Distribution-free statistical methods**. London: Chapman and Hall, 1981. 264p.
- McDONALD, B.J.; THOMPSON Jr, W.A. Rank sum multiple comparisons in one- and two-way classifications. **Biometrika**, v.54, n.3, p.487-497, 1967.
- MILLER, R.G. **Simultaneous statistical inference**. 2.ed. New York: McGraw-Hill, 1981. 299p.
- NEGRILLO, B.G. **Métodos não-paramétricos uni e multivariados**. Piracicaba: S.N., 1992. 228p.
- NEGRILLO, B.G. Testes de hipóteses não-paramétricos para experimentos em parcelas subdivididas (*Split Plot*). **Revista de Matemática e Estatística**, v.15, p.145-156, 1997.
- NOETHER, G.E. Sample size determination for some common nonparametric tests. **Journal of the American Statistical Association**, v.82, n.398, p.645-647, 1987.

- ODEH, R.E. The distribution of the maximum sum of ranks. **Technometrics**, v.9, n.2, p.271-278, 1967.
- PAN, G. Distribution-free tests for umbrella alternatives. **Communications in Statistics - Theory and Methods**, v.25, n.12, p.3185-3194, 1994.
- RHYNE, A.L.; STEEL, R.G.D. A multiple comparisons sign test: all pairs of treatments. **Biometrics**, v.23, n.4, p.539-549, 1967.
- RHYNE Jr., A.L.; STEEL, R.G.D. Tables for a treatments versus control multiple comparisons sign test. **Technometrics**, v.7, n.3, p.293-306, 1965.
- SALTER, K.C.; FAWCETT, R.F. The ART test of interaction: a robust and powerful rank test of interaction in factorial models. **Communications in Statistics - Simulation and Computation**, v.22, n.1, p.137-153, 1993.
- SANTOS, J.W.; MOREIRA, J.A.N.; BELTRÃO, N.E.M. Avaliação do emprego dos testes de comparação de médias na Revista Pesquisa Agropecuária Brasileira (PAB) de 1980 a 1994. **Pesquisa Agropecuária Brasileira**, v.33, n.3, p.225-230, 1998.
- SCHEIRER, C.J.; RAY, W.S.; HARE, N. The analysis of ranked data derived from completely randomized factorial designs. **Biometrics**, v.32, n.2, p.429-434, 1976.
- SHERMAN, E. A note on multiple comparisons using rank sums. **Technometrics**, v.7, n.2, 255-256, 1965.
- SHIRLEY, E. A non-parametric equivalent of Williams' Test for contrasting increasing dose levels of a treatment. **Biometrics**, v.33, n.2, p.387-389, 1977.
- SHIRLEY, E. Applications of ranking methods to multiple comparisons procedures and factorial experiments. **Applied Statistics**, v.36, p.205-213,

1987.

SHUSTER, J.J.; BOYETT, J.M. Nonparametric multiple comparisons procedures. **Journal of the American Statistical Association**, v.74, n.366, p.379-382, 1979.

SIEGEL, S. **Estatística não paramétrica (para as ciências do comportamento)**. Trad. de A.A. de Farias. São Paulo: Ed. McGraw-Hill do Brasil, 1977. 355p.

SILVA, H.M.G. da. Análise não-paramétrica de experimentos em parcelas subdivididas. Piracicaba, 1979. 135p. Dissertação (M.S.) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.

SKILLINGS, J.H. Nonparametric approaches to testing and multiple comparisons in a one-way ANOVA. **Communications in Statistics - Simulation and Computation**, v.12, n.4, p.373-387, 1983.

SKILLINGS, J.H.; MACK, G.A. On the use of a Friedman-Type Statistic in balanced and unbalanced block designs. **Technometrics**, v.23, n.2, 171-177, 1981.

SPRENT, P. **Applied nonparametric statistical methods**. 2.ed. New York: Chapman & Hall, 1993. 342p.

STEEL, R.G.D. A multiple comparisons sign test: treatments versus control. **American Statistical Association Journal**, v.54, p.379-382, 1959.

STEEL, R.G.D. A rank sum test for comparing all pairs of treatments. **Technometrics**, v.3, n.2, p.197-207, 1960.

STEEL, R.G.D. Query 163: Error rates in Multiple Comparisons. **Biometrics**, v.17, n.2, p.326-328, 1961a.

STEEL, R.G.D. Some rank sum multiple comparisons tests. **Biometrics**, v.17, n.4, p.539-551, 1961b.

- STILL, A.W.; WHITE, A.P. The approximate randomization test as an alternative to the F test in analysis of variance. **British Journal of Mathematical and Statistical Psychology**, v.34, p.243-252, 1981.
- TOBACH, E.; SMITH, M.; ROSE, G.; RICHTER, D. A table for making rank sum multiple paired comparisons. **Technometrics**, v.9, n.4, p.561-567, 1967.
- WHITE, A.P. Monte Carlo randomization tests: a clarification. **British Journal of Mathematical and Statistical Psychology**, v.40, p.194-195, 1987a.
- WHITE, A.P. Monte Carlo randomization tests: a replay to Bradbury. **British Journal of Mathematical and Statistical Psychology**, v.40, p.177-187, 1987b.
- WHITE, A.P.; STILL, A.W. Monte Carlo randomization tests: a replay to Bradbury. **British Journal of Mathematical and Statistical Psychology**, v.40, p.188-191, 1987.
- WILLIAMS, D.A. A note on Shirley's nonparametric test for comparing several dose levels with a zero-dose control. **Biometrics**, v.42, n.2, p.183-186, 1986.