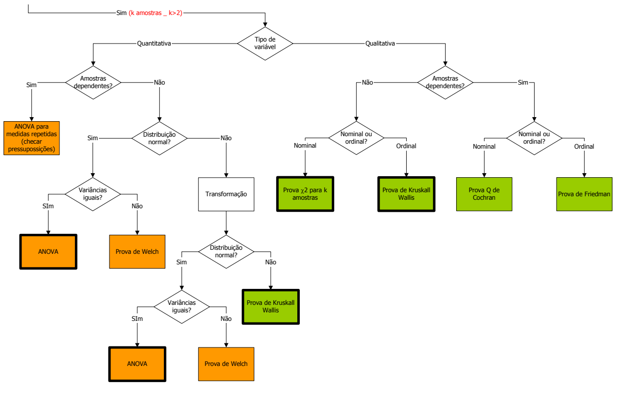
Comparando k amostras (k>2)

Abordagem não-paramétrica

Aula de hoje

Testes de Tendência Central (média, mediana, proporção)	Classificação	Variável 1	Variável 2	Número Grupos	Dependência	Premissas
Teste Z	Paramétrico	Quantitativa	-	1	-	Variância pop. * conhecida
Teste t	Paramétrico	Quantitativa	-	1	-	Distribuição normal
Wilcoxon (teste dos sinais, Wilcoxon p/ 1 amostra)	Não-Paramétrico	Quantitativa / Qualitativa ordinal	-	1	-	
Teste t p/ 2 amostras	Paramétrico	Quantitativa	Nominal	2	Independentes	Distribuição normal; variâncias iguais
Teste t p/ 2 amostras com variâncias diferentes	Paramétrico	Quantitativa	Nominal	2	Independentes	Distribuição normal; variâncias diferentes
Teste t pareado	Paramétrico	Quantitativa	Nominal	2	Dependentes	Distribuição normal
ANOVA	Paramétrico	Quantitativa	Nominal	3 ou mais	Independentes	Distribuição normal; variâncias iguais
ANOVA p/ medidas repetidas	Paramétrico	Quantitativa	Nominal	3 ou mais	Dependentes	Distribuição normal; Esfericidade
Mann-Whitney (Wilcoxon não-pareado)	Não-Paramétrico	Quantitativa / Qualitativa ordinal	Nominal	2	Independentes	
Wilcoxon (Wilcoxon pareado, teste dos sinais)	Não-Paramétrico	Quantitativa / Qualitativa ordinal	Nominal	2	Dependentes	
Kruskal-Wallis	Não-Paramétrico	Quantitativa / Qualitativa ordinal	Nominal	3 ou mais	Independentes	
Friedman	Não-Paramétrico	Quantitativa / Qualitativa ordinal	Nominal	3 ou mais	Dependentes	
Teste p/ 1 proporção	Paramétrico	Nominal	-	1	-	
Teste p/ 2 proporções	Paramétrico	Nominal	Nominal	2	Independentes	

k Amostras (k>2)



Quando utilizar a comparação de k amostras na abordagem não-paramétrica?

- ANOVA paramétrica não pode ser utilizada
  - Dados não tem distribuição normal e/ou as variâncias não são iguais (Teste de Kruskal Wallis)
- Quando as variáveis medidas são do tipo qualitativo nominal (Teste de χ^2) ou ordinal (Teste de Kruskal Wallis)

Princípio do Teste de Kruskal Wallis

- A prova de Kruskal-Wallis é útil para decidir se as k amostras provêm da mesma população
- Realização da prova:
  - Cada uma das observações é substituída pelo seu posto, isto é, as observações são dispostas em ordem crescente atribuindo-se ao menor valor o posto 1 e ao maior o posto N

Princípio do Teste de Kruskal Wallis

- Feito isto, realiza-se a soma dos postos para cada uma das amostras. O teste de Kruskal-Wallis determina se estas somas são tão dispares que não seja provável que elas se originem da mesma população

Estatística de Kruskal-Wallis

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

R_i é a soma dos postos da amostra i $(i = 1,2,\ldots,k)$

n_i é o número de observações na amostra i

$N = n_1 + n_2 + \ldots + n_k$

- Desde que os tamanhos das k amostras não sejam muito pequenos, a estatística H tem distribuição χ^2 com $gl=k-1$ graus de liberdade

Exemplo I (FSP – Profa. Maria do Rosário D O Latorre)

- Um pesquisador deseja comparar o valor do índice de massa corporal entre homens casados (grupo 1), solteiros (grupo 2) e viúvos ou separados (grupo 3). Para tanto analisou uma amostra de 19 indivíduos descritos abaixo.
- H_0: IMC são iguais para os três grupos
  H_1: IMC para os três grupos não são iguais

Exemplo

Indivíduo	IMC	Grupo	Posto	R1	R2	R3
1	26,5	1	8	8		
2	32,7	2	16		16	
3	20,4	3	2			2
4	31,6	2	14		14	
5	22,5	1	3	3		
6	25	1	5	5		
7	30,2	3	13			13
8	26,4	1	7	7		
9	19,3	2	1		1	
10	27,6	1	9	9		
11	31,7	3	15			15
12	28,1	1	10	10		
13	22,7	2	4		4	
14	36,5	3	18			18
15	36,9	3	19			19
16	25,1	2	6		6	
17	33,2	3	17			17
18	30,1	2	12		12	
19	28,7	3	11			11
			Soma	42	53	95

Saída do Minitab

13/11/2002 16:02:01

```
Welcome to Minitab, press F1 for help.
```

Kruskal-Wallis Test: IMC versus Grupo

```
Kruskal-Wallis Test on IMC

Grupo        N    Median    Ave Rank        Z
1            6     26,45         7,0    -1,58
2            6     27,60         8,8    -0,61
3            7     31,70        13,6     2,11
Overall     19                  10,0

H = 4,78  DF = 2  P = 0,092
```

Interpretação

- Para um nível de significância α=5%, como valor de p=0,092=9,2% é superior a 5%, não temos evidência para rejeitar a hipótese nula. Assim, concluímos que não há diferença no IMC dos três grupos.

Curiosidade

Sob a hipótese nula, z_i tem distribuição aproximadamente Normal com média $\mu = 0$ e desvio padrão $\sigma = 1$.

Os valores de z_i indicam como a média dos postos $\overline{R_i}$ para o grupo i difere da média dos postos $\overline{R}$ para a totalidade das observações N.

$$z_i = \frac{\overline{R_i} - (N+1)/2}{\sqrt{(N+1)(N/n_i - 1)/12}}$$

Comparações Múltiplas

1) Utilizar teste semelhante ao de Tukey, *teste de Nemenyi*, ou o *teste de Dunn*. Minitab não realiza essa comparação. Veja como realizar o teste no livro "Biostatistical Analysis" de J. H. Zar, 3.ed.,p. 226.
2) Fazer as comparações 2 a 2 utilizando o teste de Mann-Whitney, com o critério de Bonferroni para o nível de significância de cada comparação (α/c, onde c é o número de comparações a serem feitas). Veja o livro "Statistics for Health Professionals", de Susan Shott, Saunders, 1990.

Variável qualitativa ordinal

- O teste de Kruskal-Wallis também pode ser utilizado para comparar k amostras (k>2) para variável qualitativa ordinal

Exemplo 2: variável qualitativa ordinal

- Comparação do nível de conhecimento sobre febre aftosa em proprietários de fazendas de diferentes tamanhos. O nível de conhecimento foi aferido com base em questionário aplicado aos fazendeiros e que permitiu categorizá-lo em: nenhum (0), baixo (1), médio (2) e elevado (3). Existe diferença no grau de conhecimento sobre febre aftosa entre proprietários de fazendas de pequeno, médio e grande porte?

Hipóteses do teste

- H_0: Não existe diferença no grau de conhecimento
  H_1: Existe diferença no grau de conhecimento

Teste de Kruskal-Wallis

```
Kruskal-Wallis Test: Conhecimento versus Tamanho

Kruskal-Wallis Test on Conhecimento

Tamanho    N  Median  Ave Rank      Z
Grande    27   2.000      39.1   3.46
Média     18   1.000      25.2  -1.53
Pequena   15   1.000      21.3  -2.36
Overall   60              30.5

H = 12.40  DF = 2  P = 0.002
H = 13.38  DF = 2  P = 0.001  (adjusted for ties)
```

Vejam os dados no roteiro da aula prática (Aftosa.mtw)

Interpretação

- O valor de p=0,001=0,1% obtido é inferior ao nível de significância de 5%. Portanto, rejeitamos a hipótese nula e concluímos que há diferença no nível de conhecimento dos proprietários de fazendas de diferentes tamanhos (pequena, média e grande).

Referências

- Noether GE. Introdução à Estatística: Uma abordagem não-paramétrica. 2.ed., Guanabara Dois, 1976.
- Shott S. Statistics for Health Professionals. Saunders, 1990.
- Zar JH. Biostatistical Analysis. 3.ed. Prentice-Hall, 1996.