ESTATÍSTICA NÃO PARAMÉTRICA

São extremamente interessantes para análises de dados qualitativos além dos quantitativos

- A aplicação dessas técnicas não exige suposições quanto à distribuição da população da qual se tenha retirado amostras para análises.

- Podem ser aplicadas a dados que se disponham simplesmente em ordem, ou mesmo para estudo de variáveis nominais.Contrariamente à estatística paramétrica, onde as variáveis são, na maioria das vezes, intervalares.

- Exigem poucos cálculos e são aplicáveis para análise de pequenas amostras.

- Independe dos parâmetros populacionais e amostrais (média, variância, desvio padrão).

TIPOS DE TESTE

- Qui-Quadrado
- Teste dos sinais
- Teste de Wilcoxon
- Teste de Mann-Whitney
- Teste da Mediana
- Teste de Kruskal-Wallis
- Teste Kolmogorov Smirnov
- Shapiro Wilk
- Teste Levene

QUI-QUADRADO (c^2)

Restrições ao uso:

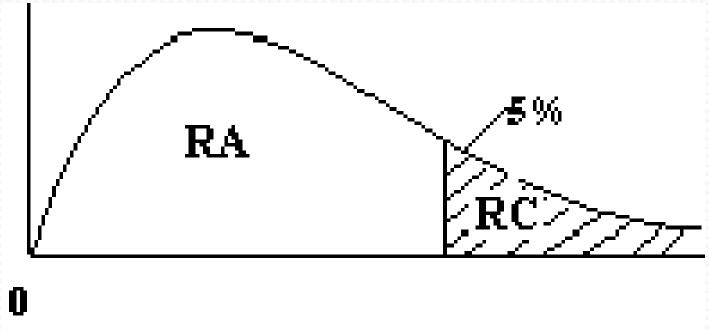
Se o número de classes é k=2, a freqüência esperada mínima deve ser ^{3} 5;

Se k >2, o teste não deve ser usado se mais de 20% das freqüências esperadas forem abaixo de 5 ou se qualquer uma delas for inferior a 1.

ADEQUAÇÃO DOS DADOS

PROCEDIMENTO

1. Enunciar as hipóteses (Ho e H1);
2. Fixar a; escolher a variável c^2 com gl= (k-1). k é o número de eventos;
3. Com auxílio da tabela de c^2, determinar RA (região de aceitação de Ho) e RC (região de rejeição de Ho)



INDEPENDÊNCIA DE VARIÁVEIS

A representação das freqüências observadas é dada por uma tabela de dupla entrada ou tabela de contingência.

PROCEDIMENTO

1. **Ho: as variáveis são independentes;**

   H1: as variáveis são dependentes;

2. **Fixar a.** Escolher a variável qui-quadrado com **gl** = (L-1) x (C-1), onde L = número de linhas da tabela de contingência e C+ número de colunas.

3. **Com auxílio da tabela calculam-se RA e RC**

Exemplo de variável qualitativa nominal

- Um pesquisador está interessado em saber se existe uma predisposição racial para a ocorrência de sarna sarcóptica em cães. Foram selecionados 300 animais aleatoriamente num município sendo procedidos exames para verificar a presença de sarna. Nos animais selecionados identificou-se a presença de animais da raça poodle, cocker e SRD.
- H_0: A proporção de animais com sarna é igual nos diferentes raças
  H_1: A proporção de animais com sarna não é igual para as diferentes raças

	Sarna +	Sarna -	Total
Poodle	2	48	50
Coacker	3	67	70
SRD	40	140	180
Total	45	255	300

Teste de χ^2

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{k} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

- Sob H_0, pode-se mostrar que a distribuição amostral de qui-quadrado tem distribuição aproximadamente qui-quadrada com gl=(r-1)(k-1) graus de liberdade

Teste de χ^2

Observado	Sarna +	Sarna –	Total
Poodle	2	48	50
Cocker	3	67	70
SRD	40	140	180
Total	45	255	300

Esperado	Sarna +	Sarna –	Total
Poodle	7,5	42,5	50
Cocker	10,5	59,5	70
SRD	27	153	180
Total	45	255	300

Teste de χ^2

$$\chi^2 = \frac{(2-7{,}5)^2}{7{,}5} + \frac{(48-42{,}5)^2}{42{,}5} + \frac{(3-10{,}5)^2}{10{,}5} + \frac{(67-59{,}5)^2}{59{,}5} + \frac{(40-27)^2}{27} + \frac{(140-153)^2}{153}$$

$$\chi^2 = 18{,}411$$

```
              C9       C10     Total
    1          2        48        50
            7,50     42,50

    2          3        67        70
           10,50     59,50

    3         40       140       180
           27,00    153,00

Total         45       255       300

Chi-Sq =   4,033 +  0,712 +
           5,357 +  0,945 +
           6,259 +  1,105 = 18,411
DF = 2, P-Value = 0,000
```

Teste de Kruskal-Wallis

Análise de variância por postos

• É uma alternativa não-paramétrica à análise que se faz por recorrência à **estatística F.**

• Pode ser usado para comparar **várias amostras** independentes O teste de Kruskal-Wallis ***não*** pode ser usada para testar diferenças em amostras **pareadas**

• É uma análise da variância que emprega posições (**soma de filas**) em lugar de mensurações como critério de avaliação.

•*Dados devem ser ordinais*, onde seja possível atribuir posições

• Exige amostras aleatórias independentes.

• *O tamanho mínimo de cada amostra deve ser 6.*

TESTE DE KRUSKALL-WALLIS

O teste de Kruskall-Wallis é uma generalização para **$k > 2$** amostras, do teste de Mann-Whitney.

A estatística de teste baseia-se nos postos das observações e como tal, a variável em estudo (nos diferentes grupos) é uma variável ordinal.

Suponha-se então a existência de **k** populações $X_1, X_2, \ldots, X_k$ das quais foram retiradas **k** amostras aleatórias

$X_{11}, X_{12}, \ldots, X_{1n_1}$	da população X_1
$X_{21}, X_{22}, \ldots, X_{1n_2}$	da população X_2
... ...	
$X_{k1}, X_{k2}, \ldots, X_{1n_k}$	da população X_k

e que existe independência, não só entre os elementos de cada amostra mas também entre os elementos de amostras distintas.

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_k$$

$$H_1: \exists\ i, j: \mu_i \neq \mu_j, \quad \text{sendo } \mu_i = E(X_i),\ i = 1,2,\ldots,k$$

A estatística de teste baseia-se nos postos das observações:

1. Ordenem-se as k amostras conjuntamente. A observação de mais baixo valor tomará o posto 1, a segunda o posto 2 e assim sucessivamente.

2. Caso existam empates, será atribuído o mesmo posto às observações empatadas. Este é a média aritmética dos postos que lhe corresponderiam se tais empates não existissem.

Seja $R(X_{ij})$ o posto atribuído a X_{ij} e

$$R_i = \sum_{j=1}^{n_i} R(X_{ij})$$

a soma dos pontos das observações da i-ésima amostra (i=1,2,...,k).

Seja

$$N = \sum_{i=1}^{k} n_i$$

o número total de observações.

Seqüência do teste

• Converter cada observação em posições crescentes em uma única fila.

• Observar os empates e considerar a posição média.

• Contabilizar a soma de fila de cada amostra.

• Calcular a Estatística H e comparar.

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

Estatística Teste

N = número total de observações

K = número de amostras

n_j = número de observações na j-ésima amostra

R_j = soma dos postos da j-ésima amostra

- Comparar com os valores críticos

$$\chi^2_{crítico}(\alpha, gl = k-1)$$

Se a hipótese nula de igualdade de médias, **é verdadeira**, os postos devem ficar **bem dispersos** entre as amostras. Os quadrados das somas de postos divididos pelos respectivos tamanhos amostrais devem ser aproximadamente **iguais**.

Verificar se o **número de empates é grande**, pois isto afetará o valor de H. Consequentemente, pode ser necessário ajustar o valor de H dividindo-o pela quantidade

$$1 - \frac{\sum \left(t^3 - t\right)}{N^3 - N}$$

onde t é o número de empates num grupo de empates

Exemplo (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
```

Uma companhia deseja comparar cinco máquinas diferentes (A, B, C, D e E), em um experimento projetado para determinar se existe diferença de desempenho entre as elas.

Cada um de cinco operários experientes trabalharam com as máquinas por períodos de tempo iguais. A tabela abaixo apresenta o número de unidades produzidas por cada máquina. Testar a hipótese de que não existe diferença entre as máquinas aos níveis de significância (a)0,05 e (b)0,01.

A	68	72	77	42	53
B	72	53	63	53	48
C	60	82	64	75	72
D	48	61	57	64	50
E	64	65	70	68	53

Etapa 1: .

H0: Não existe diferença entre as máquinas

H1: Existe diferença entre as máquinas

Etapa 2: Estabelecendo o nível de significância: $\alpha = 0{,}05$ **e** $\alpha = 0{,}01$

Etapa 3: Estabelecendo a estatística de teste: **H**

Etapa 4: Estabelecendo os valores críticos

- Como existem cinco amostras (*A, B, C, D e E*), $k = 5$
- Como cada amostra consiste de cinco valores temos

$N_1=N_2=N_3=N_4=N_5=5$, resulta que $N = N_1+N_2+N_3+N_4+N_5=25$.

$$\chi^2 \text{crítico}(\alpha = 0{,}05 \text{ e } gl = k-1 = 4) = 9{,}49$$

Etapa 5: O valor da Estatística Teste

Ordenando-se todos os valores crescentemente e atribuindo-se postos apropriados aos empates.

	POSIÇÕES					Soma dos Postos
A	17,5	21	24	1	6,5	70
B	21	6,5	12	6,5	2,5	48,5
C	10	25	14	23	21	93
D	2,5	11	9	14	4	40,5
E	14	16	19	17,5	6,5	73

R_1 = 70, R_2 = 48,5, R_3 = 93, R_4 = 40,5 e R_5 = 73.

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

$$H = \frac{12}{(25)(26)} \left[\frac{(70)^2}{5} + \frac{(48{,}5)^2}{5} + \frac{(93)^2}{5} + \frac{(40{,}5)^2}{5} + \frac{(73)^2}{5} \right] - 3(26) = 6{,}44$$

Etapa 6: Como **H** < χ^2crítico (6,44 < 9,49), não podemos rejeitar a hipótese da não existência de diferença entre as máquinas ao nível 0,05, e por esta razão, certamente também não podemos rejeitá-la ao nível 0,01.

Agora vamos resolver este problema fazendo uma correção para os empates

Observação	48	53	64	68	72	
Número de empates (t)	2	4	3	2	3	
t^3-t	6	60	24	6	24	$\sum(t^3 - t) = 120$

$$\sum(t^3 - t) = 6 + 60 + 24 + 6 + 24 = 120$$

$$1 - \frac{\sum(t^3 - t)}{N^3 - N} = 1 - \frac{120}{(25)^3 - 25} = 0{,}9923$$

$$\boxed{H_c = \frac{6{,}44}{0{,}9923} = 6{,}49}$$

Esta correção não é suficiente para alterar a decisão adotada anteriormente

http://www.portalaction.com.br/content/teste-de-kruskal-wallis

Tempo	Lâmpada
73	A
64	A
67	A
62	A
70	A
84	B
80	B
81	B
77	B
82	C
79	C
71	C
75	C

TESTE DE KRUSKAL-WALLIS

DADOS DO PROCESSO

Informação	*Valor*
Kruskal-Wallis qui-quadrado	8,403296703
Graus de Liberdade	2
P-valor	0,014970879

Fatores	*Limite Inferior*	*Efeito*	*Limite Superior*
A	63,06081601	67,2	71,33918399
B	75,87225162	80,5	85,12774838
C	72,12225162	76,75	81,37774838