Profª. Maigan Alcântara

Monitora: Iris Souza

09/11/2021

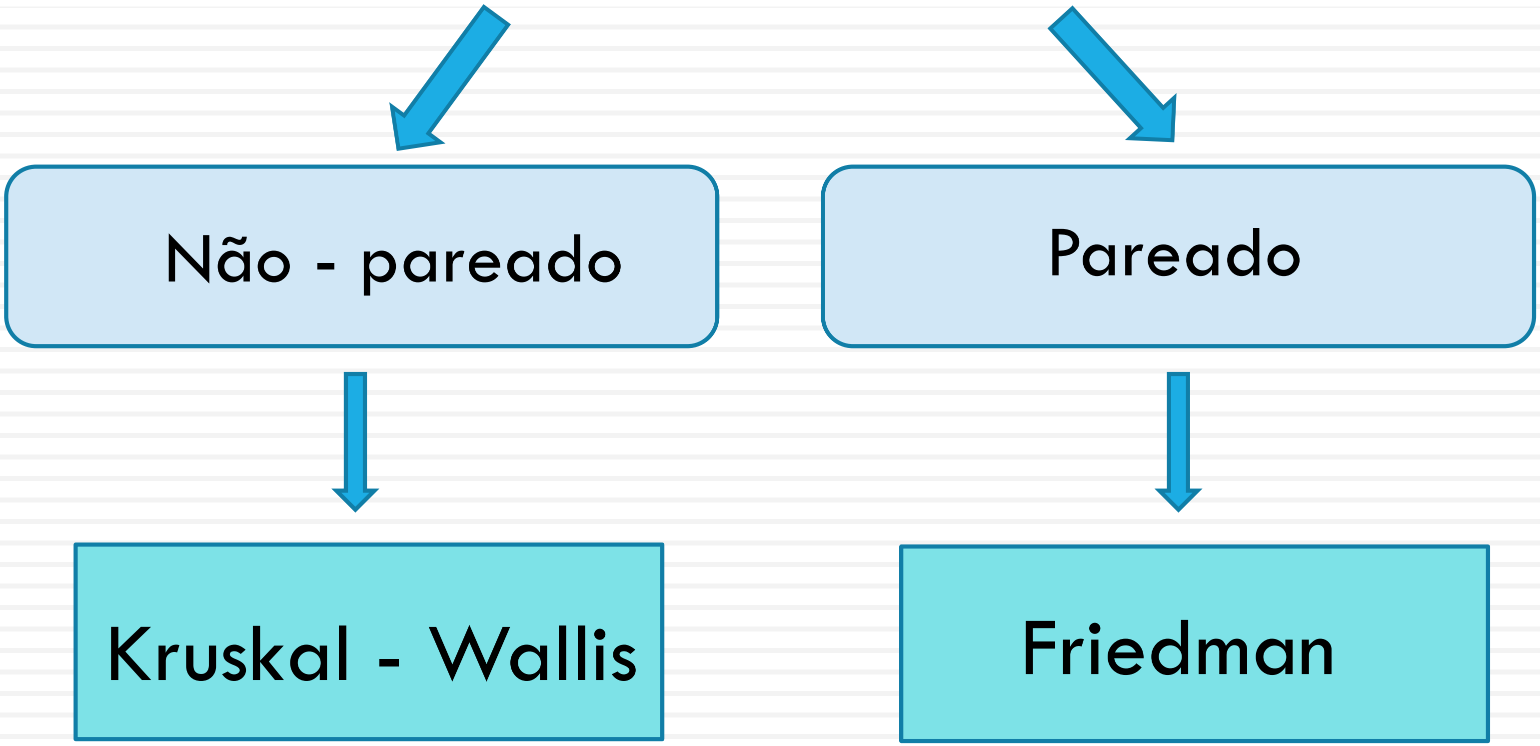
Estatística Paramétrica e Não-Paramétrica com uso do software R

UNIVASF
UNIVERSIDADE FEDERAL DO VALE DO SÃO FRANCISCO

ProeX
PRÓ-REITORIA DE EXTENSÃO

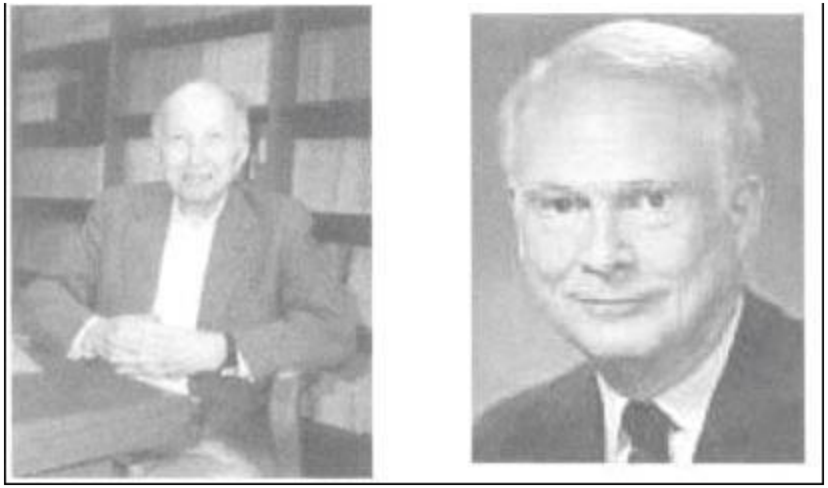
UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
CODAI
COLÉGIO AGRÍCOLA DOM AGOSTINHO IKAS DA UFRPE

Testes Não Paramétricos para mais de dois grupos:



Teste Kruskal-Wallis

É um teste não paramétrico utilizado para comparar três ou mais populações. Ele é usado para testar a hipótese nula de que todas as populações possuem funções de distribuição iguais contra a hipótese alternativa de que ao menos duas das populações possuem funções de distribuição diferentes.



William Henry Kruskal (1919 - 2005) e Wilson Allen Wallis (1912-1998).

Teste Kruskal-Wallis

- É uma extensão do teste de Mann-Whitney (soma dos postos de Wilcoxon), i.e. para k = 2 populações, este teste é equivalente ao teste de Wilcoxon da soma de postos;
- É equivalente ao teste de Mantel-Haenszel aplicado aos postos dos dados;
- Pode ser interpretado como a versão não-paramétrica do teste F da ANOVA com 1 fator;
- Não leva em consideração formas específicas de distribuição;
- Deseja-se testar se k amostras aleatórias(possivelmente tamanhos diferentes) de uma v.a. possuem a mesma distribuição.

Teste Kruskal-Wallis

PRESSUPOSTOS

- As amostras são aleatórias e independentes entre si;
- A escala de mensuração é no mínimo ordinal.

HIPÓTESES

- H0 :

Todas as k populações têm funções de distribuição idênticas.

- H1 :

Pelo menos uma das populações difere das demais.

Teste Kruskal-Wallis

CONSTRUÇÃO DO TESTE

Considere a i–ésima amostra aleatória de tamanho ni :

$$X_{i1}, X_{i2}, \ldots, X_{in_i}, \; i = 1, 2, \ldots, k.$$

Seja N o número total de observações

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

Teste Kruskal-Wallis

CONSTRUÇÃO DO TESTE

Os dados podem ser organizados em colunas:

Amostra 1	Amostra 2	...	Amostra k
X_{11}	X_{21}	$\dots$	X_{k1}
X_{12}	X_{22}	$\dots$	X_{k2}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
X_{1n_1}	X_{2n_2}	$\dots$	X_{kn_k}

Teste Kruskal-Wallis

CONSTRUÇÃO DO TESTE

Atribua posto 1 à menor observação do total de N observações, posto 2 a segunda menor, e assim por diante, até a maior de todas as N observações, que recebe posto N (**quando não há empates**).

Quando há observações iguais, calcular a média dos postos. Sejam :

- Rij o posto da observação de Xij
- $R_{i\circ}$ a soma dos postos da i–ésima amostra:

$$R_{i.} = \sum_{j=1}^{n_i} R_{ij}, \quad i = 1, 2, \ldots, k,$$

CONSTRUÇÃO DO TESTE

e R• • a média geral dos postos:

$$R_{\cdot\cdot} = \frac{\sum_{i=1}^{k} R_{i\cdot}}{N}.$$

Se não há empates, teremos que a soma total de postos(numerador de (1)) é a soma de uma PA e portanto R• • = (N + 1)/2.

Teste Kruskal-Wallis

ESTATÍSTICA DO TESTE

A estatística de teste T é definida como:

$$T = \frac{1}{S^2}\left(\sum_{i=1}^{k}\frac{R_{i\cdot}^2}{n_i} - \frac{N(N+1)^2}{4}\right),$$

em que

$$S^2 = \frac{1}{N-1}\left(\sum_{i,j}R_{ij}^2 - \frac{N(N+1)^2}{4}\right)$$

Teste Kruskal-Wallis

ESTATÍSTICA DO TESTE

Esta estatística mede a razão entre :

- A soma diferenças quadráticas das médias dos tratamentos para a média geral (soma de quadrados entre tratamentos) e;
- As diferenças quadráticas dos postos em relação à média geral (quadrado médio total, variância);

Teste Kruskal-Wallis

ESTATÍSTICA DO TESTE

Se **não há empates**, teremos que:

$$S^2 = N(N+1)/12$$

e a estatística de teste reduz a:

$$T = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_{i\cdot}^2}{n_i} - 3(N+1);$$

Teste Kruskal-Wallis

ESTATÍSTICA DO TESTE

- A distribuição exata de T é tabelada (veja [Conover, 1996], Tabela A8) para $k = 3$ e $n_i \leq 5$ (no caso sem empates). No entanto, a distribuição exata é complexa e nos demais casos utiliza-se uma aproximação de T por a qui-quadrado com $k - 1$ graus de liberdade;

- Rejeita H_0 ao nível α se T for maior que o quantil $1 - \alpha$ (dos valores tabelados ou da χ^2_{k-1})

Teste Kruskal-Wallis

ASPECTO COMPUTACIONAL

No R, Pode ser realizado pelo pacote kruskal.test ou pelo pacote PMCMRplus

```
stats::kruskal.test(resposta,tratamento)
    kruskal.test(resposta~tratamento)
PMCMRplus::kruskalTest(resposta, tratamento, dist = "Chisquare")
      ::kruskalTest(resposta, tratamento)
# dist =  "KruskalWallis" p/ teste com a distr. exata.
### comparacoes multiplas ###
      kwAllPairsConoverTest(resposta~tratamento)
```

Teste Kruskal-Wallis – Exemplo 1

Os dados a seguir são de uma experiência clássica agrícola para avaliar o rendimento de culturas divididas em quatro grupos diferentes. Para manter a simplicidade, identificamos os tratamentos usando os números inteiros {1,2,3,4}. Queremos avaliar se os dados provém de distribuições igualmente distribuídas.

1 83	1 91	1 94	1 89	1 89	1 96	1 91	1 92	1 90	1 84	2 91	2 90	2 81	2 83	2 84
2 91	2 89	3 101	3 100	3 91	3 93	3 96	3 95	3 94	3 81	4 78	4 82	4 81	4 77	4 79

Teste Kruskal-Wallis – Exemplo 1

Os dados a seguir [illegible]periência clássica agrícola para avaliar [illegible]ividida[s] em quatro grupos dif[illegible] manter a simpli[illegible], identificamos os trat[illegible] usando os números [illegible]ros {1,2,3,4}. Querem[illegible]ar se [illegible] de distribuições igualm[illegible]tribuídas.

1	1	[illegible]	[illegible]	1	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2	2	2	2
83	91	[illegible]	[illegible]	[illegible]	96	[illegible]	[illegible]	[illegible]	[illegible]	91	[illegible]	81	83	84
2	2	3	3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4	4	4	4
91	89	101	100	91	93	[illegible]	[illegible]	94	[illegible]	[illegible]	[illegible]	81	77	79

Teste Kruskal-Wallis

EXERCÍCIO 1

Verificar a influência do fator Idade sobre a variável tempo (em dias) para conseguir um emprego, considerando as seguintes amostras:

Acima de 40 anos	Entre 25 e 40	Abaixo de 25
63	33	25
20	42	31
43	27	6
58	28	14
	51	18

Ao nível de 5% de significância, é possível afirmar que o fator idade tem influência sobre o tempo para encontrar trabalho?

Teste Kruskal-Wallis

EXERCÍCIO 2 [Korosteleva, 2014]

Em uma competição de arco e flecha, 3 competidores disputam o primeiro lugar, que será definido estatisticamente. A pontuação por acerto pode ser 10 para o círculo menor, 5 a 9 para os círculos intermediários e 1 a 4 para os círculos mais externos. Cada competidor tem direito a 10 flechas. Decida se houve ganhador, ou se todos obtiveram o mesmo desempenho, com base nos dados abaixo.

Monica		Bob		Jeff	
Pontuação	Posto	Pontuação	Posto	Pontuação	Posto
3	6.5	2	4	1	1.5
4	10.5	2	4	1	1.5
4	10.5	3	6.5	2	4
5	16	4	10.5	4	10.5
5	16	4	10.5	4	10.5
5	16	5	16	5	16
10	22	10	22	10	22
10	22	10	22	10	22
				10	22

Teste de Friedman

É um teste não paramétrico utilizado para comparar três ou mais populações. Ele é usado para testar a hipótese nula de que todas as populações possuem funções de distribuição iguais contra a hipótese alternativa de que ao menos duas das populações possuem funções de distribuição diferentes.

Milton Friedman(1912 - 2006).

Teste de Friedman

- É uma extensão do teste dos sinais de Wilcoxon, logo, também utiliza os postos das observações. É também um caso especial do teste de Mantel-Haenszel geral, o qual vimos anteriormente o caso $k \times 2$
- É útil para estudos de medidas repetidas ou delineamento em blocos;
- Neste tipo de estudo, observa-se o mesmo grupo de indivíduos sob cada um das k tratamentos, ou então formam-se conjuntos de indivíduos homogêneos que são alocados aleatoriamente a cada um dos tratamentos;
- É uma alternativa não paramétrica para a ANOVA com blocos casualizados;
- O teste examina os postos (ranks) dos dados em cada tratamento para determinar se as distribuições das variáveis são provenientes da mesma população.

Teste de Friedman

SUPOSIÇÕES:

- A variável de interesse é medida no mínimo em escala ordinal;
- Os blocos são independentes, i.e. a variabilidade dentro de um bloco não influencia os resultados de outro bloco.
- Em cada bloco (amostra), as observações podem ser ordenadas de acordo com algum critério de interesse.

Teste de Friedman

DADOS

- Consistem de b vetores aleatórios independentes k–variados (X_{i1}, X_{i2}, . . . , X_{ik}), chamados blocos (ou amostras), i = 1, . . . , b.

- A variável aleatória X_{ij} representa a observação associada ao bloco i e ao tratamento j.

- Os dados podem ser organizados em forma da Tabela que será apresentada.

Teste de Friedman

CONSTRUÇÃO DO TESTE

Atribuir postos de 1 a k para as observações do bloco i, i = 1, 2, . . . , b, ditos Rij ; Em caso de empates, atribuir a média dos postos; A soma dos postos para cada tratamento, R•j , é definida por:

$$R_{\cdot j} = \sum_{i=1}^{b} R_{ij}, \text{ para todo } j = 1, 2, \ldots, k$$

Teremos assim uma nova tabela de postos:

Teste de Friedman

Amostras (Blocos)	Tratamentos 1	2	$\ldots$	k
1	X_{11}	X_{12}	$\ldots$	X_{1k}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
b	X_{b1}	X_{b2}	$\ldots$	X_{bk}
Total	-	-	-	-

Postos 1	2	$\ldots$	k
R_{11}	R_{12}	$\ldots$	R_{1k}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
R_{b1}	R_{b2}	$\ldots$	R_{bk}
$R_{.1}$	$R_{.2}$	$\ldots$	$R_{.k}$

Teste Kruskal-Wallis

ESTATÍSTICA DO TESTE

Diferente do teste de Kruskal-Wallis, aqui temos postos 1, . . . , k para cada bloco. Logo, a soma total dos postos, se não há empates, será b vezes a soma de uma PA de k termos. A média geral dos postos será então:

$$R_{..} = \frac{1}{k} b \left(\frac{k(k+1)}{2} \right) = \frac{b(k+1)}{2};$$

Teste de Friedman

ESTATÍSTICA DO TESTE

No caso **sem empates**, Friedman propôs a estatística do teste como:

$$
\begin{aligned}
T_1 &= \frac{12}{bk(k+1)} \sum_{j=1}^{k} \left(R_j - \frac{b(k+1)}{2} \right)^2 \\
&= \frac{12}{bk(k+1)} \sum_{j=1}^{k} R_j^2 - 3b(k+1),
\end{aligned}
$$

Note que, assim como no teste de Kruskal-Wallis, esta estatística também está relacionada com a soma das diferenças quadráticas das médias dos tratamentos para a média geral

Teste de Friedman

ESTATÍSTICA DO TESTE

<u>**Se há empates**</u>, um ajustamento na estatística T1 precisa ser feito:

- Seja A1 a soma dos quadrados dos postos

$$A_1 = \sum_{i=1}^{b} \sum_{j=1}^{k} [R_{ij}]^2;$$

- Calcule o fator de correção dado por:

$$C_1 = bk(k+1)^2/4;$$

Teste de Friedman

ESTATÍSTICA DO TESTE

Então, a estatística T1, modificada na **presença de empates**, é:

$$T_1^* = \frac{(k-1)\left(\sum_{j=1}^{k} R_j^2 - bC_1\right)}{A_1 - C_1} = \frac{(k-1)\sum_{j=1}^{k}\left(R_j - \frac{b(k+1)}{2}\right)^2}{A_1 - C_1};$$

- A distribuição exata de T1 e T ∗ 1 é difícil ser encontrada e uma aproximação é comumente usada;
- T1 e T ∗ 1 têm, sob H0 distr. aprox. Qui-quadrado com (k – 1) graus de liberdade;

Teste de Friedman

ESTATÍSTICA DO TESTE

Uma estatística alternativa é a estatística dos experimentos em blocos ao acaso na ANOVA calculada sobre os postos R(Xij):

$$T_2 = \frac{(b-1)\,T_1}{b(k-1) - T_1};$$

A estatística T2 tem, aproximadamente, distribuição F com k1 = (k – 1) e k2 = (b – 1)(k – 1) graus de liberdade.

Teste de Friedman

DECISÃO DO TESTE

- Fixado α, rejeitamos H_0 se T_1 exceder o quantil $(1 - \alpha)$ da distribuição Qui-quadrado com $(k - 1)$ graus de liberdade.

- De maneira similar, rejeitamos H_0 ao nível de significância α se T_2 exceder o quantil $(1 - \alpha)$ da distribuição F com

  $k_1 = (k - 1)$ e $k_2 = (b - 1)(k - 1)$ graus de liberdade.

Teste de Friedman

ASPECTO COMPUTACIONAL

No R, pode ser realizado

```
friedman.test(resp, trat, bloco)
friedman.test(resp ~ trat | bloco, data)
```

Teste de Friedman – Exemplo

- Um teste de consumo de combustível envolvendo carros produzidos por três fabricantes foi realizado e os resultados, em quilômetros por litro de combustível estão apresentados na tabela abaixo. Estabelecer e testar a hipótese adequada. Considere $\alpha = 5\%$

Modelo	Fabricante		
	G	F	C
Pequeno	9.0	11.3	10.6
Médio- 6 cil.	9.4	10.9	10.2
Médio- 8 cil.	8.1	8.6	9.1
Grande-8 cil.	8.3	8.6	8.8
Esporte	8.2	9.2	9.5

- Aqui podemos testar se existe diferença do consumo do carro de acordo com o modelo.

Teste de Friedman – Exemplo

- Um teste de cons mbustível envolvendo carros produzidos po do e os resultados, em quilômetr apresentados na tabela abelecer e testar a tese adequada. Consi

Model	F
Pequer	11.3
Médio- 6	
Médio- 8	
de-8	

- Aqui podemos testar se diferer onsumo do carro de acordo com o modelo o fabri

Teste de Friedman

EXERCÍCIO 1 [Lehmann and D'abrera, 2006]

Num estudo sobre hipnose, 8 sujeitos tiveram a tensão elétrica na superfície da pele medida (em milivolts) em 4 situações emocionais distintas: medo, alegria, tristeza e calma. Avalie se existe diferença na tensão entre os diferentes estados emocionais.

Sujeito	Medo	Alegria	Tristeza	Calma
1	23.1	22.7	22.5	22.6
2	57.6	53.2	53.7	53.1
3	10.5	9.7	10.8	8.3
4	23.6	19.6	21.1	21.6
5	11.9	13.8	13.7	13.3
6	54.6	47.1	39.2	37.0
7	21.0	13.6	13.7	14.8
8	20.3	23.6	16.3	14.8

Teste de Friedman

EXERCÍCIO 2 Velocidades de atletas

6 atletas de ciclismo tiveram suas velocidades médias calculadas ao longo de 4 trechos de uma prova. Avalie se algum atleta se destacou dos demais.

	Trecho			
Atleta	A	B	C	D
1	32.60	36.40	29.50	29.40
2	42.70	47.10	32.90	40.00
3	35.30	40.10	33.60	35.00
4	35.20	40.30	35.70	40.00
5	33.20	34.30	33.20	34.00
6	33.10	34.40	33.10	34.10

Teste de Friedman

- Beall, G. (1942). The transformation of data from entomological field experiments so that the analysis of variance becomes applicable. Biometrika, 32:243.
- Conover, W. J. (1996). Practical nonparametric statistics. John Wiley and sons, 3 ed. edition.
- Korosteleva, O. (2014). Nonparametric methods in statistics with SAS applications
- Lehmann, E. L. and D'abrera, H. J. M. (2006). Nonparametrics : statistical methods based on ranks.
- Notas de aula do prof Anderson Ara