

MATD49-Estatística não paramétrica

Testes para duas amostras

Teste de Kolmogorov-Smirnov para duas amostras

Introdução I

- É utilizado na **comparação de dois grupos independentes**;
- Verifica a **concordância entre duas distribuições acumuladas**;
- Diferente do teste de Kolmogorov-Smirnov para uma amostra, o teste para duas amostras não especifica a forma exata das duas distribuições, apenas verifica a discordância máxima entre as duas distribuições acumuladas a partir dos dados de duas amostras.

Objetivo:

Comparar duas amostras independentes e determinar se elas são provenientes da mesma população, ou de populações com a mesma distribuição.

Introdução II

Suposições:

- As amostras são aleatórias e independentes;
- Os grupos devem ser **mutuamente exclusivos**;
- A escala de **mensuração deve ser no mínimo ordinal**. Pode-se também utilizar os **postos** das variáveis;
- Se a variável de interesse é **contínua**, o teste é **exato**. No caso de variáveis **discretas**, o teste é apenas **aproximado**.
- Dados:
 - Considere n_x observações de uma amostra aleatória X_1, X_2, \dots, X_{n_x} da população 1;
 - Considere n_y observações de uma amostra aleatória Y_1, Y_2, \dots, Y_{n_y} da população 2;
 - Supomos que F_X e G_Y sejam as funções de distribuição correspondentes as populações 1 e 2;
 - As amostras poderão ser de tamanhos diferentes.

Hipóteses do teste

O Teste pode ser bilateral ou unilateral:

- Hipótese nula:

$$H_0 : F_X(t) = G_Y(t), \text{ para todo } t \in \mathbb{R}.$$

- Hipótese alternativa:

- Bilateral: $H_1 : F_X(t) \neq G_Y(t)$, para pelo menos um valor de $t \in \mathbb{R}$;
- Unilateral à direita: $H_1 : F_X(t) > G_Y(t)$, para pelo menos um valor de $t \in \mathbb{R}$;
- Unilateral à esquerda: $H_1 : F_X(t) < G_Y(t)$, para pelo menos um valor de $t \in \mathbb{R}$.

Estatística de teste

Sejam $S_X(t)$ e $S_Y(t)$ as funções distribuições empíricas baseadas nas observações de (X_1, \dots, X_{n_x}) e (Y_1, \dots, Y_{n_y}) , respectivamente.

- **Bilateral:** Defina T_1 como a maior distância vertical absoluta entre $S_X(t)$ e $S_Y(t)$:

$$T_1 = \sup_t |S_X(t) - S_Y(t)|;$$

- **Unilateral à direita:** T_1^+ , a maior distância vertical entre $S_X(t)$ e $S_Y(t)$:

$$T_1^+ = \sup_t [S_X(t) - S_Y(t)];$$

- **Unilateral à esquerda:** T_1^- , a maior distância vertical entre $S_Y(t)$ e $S_X(t)$:

$$T_1^- = \sup_t [S_Y(t) - S_X(t)].$$

Decisão do teste

- Rejeitamos H_0 ao nível de significância α se a estatística teste associada à hipótese escolhida, T_1 , T_1^+ ou T_1^- , exceder o quantil $1 - \alpha$ encontrado na Tabela do teste de Kolmogorov-Smirnov;
- Os valores dos quantis da distribuição da estatística do teste são tabelados e podem ser encontrados por exemplo, em [Conover, 1996], Tabelas A19 (tamanhos de amostras iguais) e A20 (tamanhos de amostras diferentes);
- Para amostras grandes, aproximações são dadas no final de cada tabela.

Aspecto computacional

No R:

```
ks.test(x,y)
```

No SAS:

```
PROC NPAR1WAY DATA=dados;  
CLASS tipo_amostra;  
VAR var_analisada;  
EXACT;  
RUN;
```


Exemplo 1

O coordenador de um curso deseja saber se os alunos que ingressam via vestibular têm desempenho diferente dos que entram por outros meios. Os resultados estão apresentados na tabela abaixo.

Notas	Vestibular (X)	Outros meios (Y)
[0 – 4.0]	1	0
(4.0 – 6.0]	1	0
(6.0 – 7.0]	2	6
(7.0 – 8.5]	28	55
(8.5 – 10.0]	10	6
Total	42	67

Vamos verificar se as duas populações têm o mesmo comportamento ao nível $\alpha = 5\%$.

Exemplo 2 - [Korosteleva, 2014]

Uma droga contra câncer será testada em pacientes com leucemia. A tabela abaixo apresenta os dados de 14 pacientes, 8 tratamento e 6 controle, de acordo com sua contagem de glóbulos brancos (10^9 células por litro)

Tratamento	12.33	10.44	12.72	13.13	13.50	16.82	17.60	14.37
Controle	16.45	18.63	13.12	18.94	19.34	22.50		

Teste:

- H_0 : a droga não tem efeito *versus*
- H_1 : O tratamento é eficaz ($F_C(t) < F_T(t)$)

e decida ao nível de significância 5%.

Testes para duas amostras relacionadas

Introdução I

- Os testes não-paramétricos para duas amostras relacionadas são úteis quando se deseja estabelecer comparações entre dois tipos de procedimentos e outras variáveis podem interferir nos resultados;
- Para evitar que um grupo de indivíduos seja naturalmente superior ao outro, é comum proceder algum tipo de pareamento entre os indivíduos;
- O tipo mais comum de pareamento é utilizando cada indivíduo como seu próprio controle, submetendo-o aos dois tratamentos em ocasiões diferentes;
- Outro tipo de pareamento é tentar seleccionar, para cada par, indivíduos que sejam tão semelhantes quanto possível;

Introdução II

- Os testes paramétricos para dados pareados (o teste t pareado, por exemplo), apresentam algumas condições que muitas vezes não são satisfeitas na amostra (normalidade e independência);
- Casos comuns onde o teste t não é aplicável:
 - Dados qualitativos e não quantitativos, isto é, dados em escala de mensuração nominal ou ordinal;
 - As diferenças entre os dados pareados são apenas ordinais, isto é, podemos apenas estabelecer qual membro do par é maior do que o outro mas não é possível especificar em quanto é maior;
 - Amostras pequenas com escores fortemente assimétricos.

Teste de McNemar

Introdução

- O Teste de McNemar é baseado na distribuição Binomial, pois a variável de teste é dicotômica;
- É particularmente aplicável aos experimentos do tipo *antes e depois* (*pré-teste e pós-teste*) em que cada sujeito é utilizado como seu próprio controle e a medida é efetuada em escala nominal ou ordinal;
- Cada indivíduo se encaixa em uma das duas categorias de resposta, antes e depois do tratamento;
- Este teste avalia a mudança de uma categoria para outra, após a exposição da unidade amostral ao tratamento. Com isso, deseja-se testar a eficiência do tratamento.

Pressupostos

- Dados em escala nominal ou ordinal (dicotômicos);
- Dados pareados;
- Indicado para amostras com mais de 20 unidades;
- Cada elemento da amostra é tomado como seu próprio controle;
- Os pares de observações (X_i, Y_i) são mutuamente independentes, em que X_i representa a situação "pré" e Y_i representa a situação "pós" do i -ésimo indivíduo.

Dados

- Os dados podem ser resumidos em uma tabela de contingência 2×2 da seguinte forma:

Variável X: antes (Pré-teste) ($X_i = 0$ ou 1)	Variável Y: depois (Pós-teste)		Total
	Ausência da categoria (fracasso = 0)	Presença da categoria (sucesso = 1)	
Ausência da categoria (fracasso = 0)	a	b	a+b
Presença da categoria (sucesso = 1)	c	d	c+d
Total	a+c	b+d	a+b+c+d

em que:

$b + c$ = Número de observações discordantes;

$a + d$ = Número de observações em concordância;

$a + b + c + d$ = tamanho da amostra pareada.

Hipóteses do teste

Hipótese nula:

$H_0 : P(X_i = 0; Y_i = 1) = P(X_i = 1; Y_i = 0), \forall i$, isto é, a probabilidade de um elemento da amostra mudar da categoria "0" para a categoria "1" é igual a probabilidade de mudar de "1" para "0".

Hipótese alternativa:

- **Bilateral:** $H_1 : P(X_i = 0; Y_i = 1) \neq P(X_i = 1; Y_i = 0), \forall i$;
- **Unilateral à direita:** $H_1 : P(X_i = 0; Y_i = 1) < P(X_i = 1; Y_i = 0), \forall i$;
- **Unilateral à esquerda:** $H_1 : P(X_i = 0; Y_i = 1) > P(X_i = 1; Y_i = 0), \forall i$.

Estatística de teste I

- A hipótese nula admite que o número de elementos que mudaram de categoria seja o mesmo nas duas direções ($0 \rightarrow 1$ e de $1 \rightarrow 0$);
- O número de mudanças na amostra é igual ao número de observações discordantes ($b + c$). Assim, sob H_0 , o esperado é que se tenha $\frac{b+c}{2}$ mudanças em cada direção;
- O teste de McNemar está apenas interessado na significância da mudança, assim o interesse é avaliar o que ocorreu nas células com frequências b e c da tabela de contigência apresentada anteriormente;
- Deseja-se avaliar se as frequências observadas são significativamente diferentes das frequências esperadas nas duas células onde as categorias não concordam;
- Portanto, as células com frequências a e d são desconsideradas neste teste.

Estatística de teste II

- A estatística de teste é definida por:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} = \frac{(b - \frac{b+c}{2})^2}{\frac{b+c}{2}} + \frac{(c - \frac{b+c}{2})^2}{\frac{b+c}{2}} = \frac{(b - c)^2}{b + c},$$

que tem distribuição aproximadamente Qui-Quadrado com 1 grau de liberdade.

- Como a estatística de teste é aproximadamente χ_1^2 , este teste é recomendado para amostra com tamanho maior do que 20.
- Quando todas as frequências esperadas são pequenas, a aproximação da estatística do teste pela qui-quadrado não pode ser boa.

Estatística de teste III

Correção de continuidade de Yates:

$$\chi^2 = \frac{(|b - c| - 1)^2}{b + c}.$$

Regra de Decisão:

Para um nível de significância α , a regra de decisão do teste é:

$$\text{Rejeita-se } H_0 \text{ se } \chi_{calc}^2 > \chi_{\alpha;1}^2.$$

Aspecto computacional

No R:

```
mcnemar.test()
```

No SAS:

```
PROC FREQ data = tabela_dados ;  
TABLES var_antes*var_depois /AGREE;  
WEIGHT cont;  
RUN;
```

Exemplo

Numa campanha política, após uma série de fatos que supostamente denegririam a imagem do candidato B em favor do candidato A, investigou-se as mudanças ocorridas com relação à preferência do eleitorado. Os resultados são os seguintes.

Antes	Depois		Total
	A	B	
A	83	47	130
B	18	52	70
Total	101	99	200

Teste de Wilcoxon dos sinais(*signed ranks*)

Introdução

Este teste é utilizado com dados pareados e considera o valor das diferenças das observações e seus postos para avaliar a eficiência de uma intervenção.

Finalidade:

Testar a existência de mudanças na classificação das observações após a exposição da unidade amostral a uma intervenção.

Pressupostos:

- Dados quantitativos e pareados;
- Os pares de observações são independentes entre si;
- Escala de mensuração no mínimo ordinal;
- Teste especialmente eficiente para pequenas amostras(ideal quando $n \leq 50$ e não há empates de postos).

Construção do teste

- Para cada par (X_i, Y_i) , $i = 1, 2, \dots, N$, determinar a diferença $(D_i = Y_i - X_i)$, com sinal, entre os dois valores;
- Determinar n : total de diferenças D_i não nulas, $n < N$;
- Para atribuir postos r_i aos valores absolutos dos D_i 's (considerar $|D_i| = |Y_i - X_i|$);
- Em caso de empate, atribuir a média dos postos que as observações teriam se não empatassem;
- Atribuir a cada posto o sinal "+" ou "-" do D_i que ele representa.

Para $D_i = Y_i - X_i$ a diferença entre os escores do par i , temos que:

- A distribuição de cada D_i é simétrica;
- Os D_i 's são independentes e têm a mesma média;
- A escala de mensuração dos D_i 's deve ser, no mínimo, intervalar.

Hipóteses do teste

O teste pode ser bilateral ou unilateral:

Hipótese nula:

$H_0 : E(D) = 0$ (i.e. $E(Y) = E(X)$), ou equivalentemente, a média do pós-teste é igual a média do pré-teste.

Hipótese alternativa:

- **Bilateral:** $H_1 : E(D) \neq 0$ ($E(Y) \neq E(X)$);
- **Unilateral à direita:** $H_1 : E(D) > 0$ ($E(Y) > E(X)$);
- **Unilateral à esquerda:** $H_1 : E(D) < 0$ ($E(Y) < E(X)$).

Estatística de teste I

- Seja R_i o posto (ou *rank*) com o sinal de D_i :

$$R_i = r_i \cdot d_i, \quad d_i = \mathbb{I}(D_i > 0) - \mathbb{I}(D_i < 0) \quad (1)$$

- **positiva**, se a diferença original de D_i é "+";
- **negativa**, se a diferença original de D_i é "-".

Estatística de teste:

A estatística de teste é a soma dos postos com sinal positivo:

$$T^+ = \sum_i R_i \cdot \mathbb{I}(d_i = 1). \quad (2)$$

- Os quantis inferiores da distribuição exata de T^+ quando não existe empates nos postos e $n \leq 50$ são tabelados (sob a hipótese nula de que os D_i 's tem média 0);

Estatística de teste II

- Os quantis superiores são encontrados a partir da relação:

$$\omega_p = \frac{n(n+1)}{2} - \omega_{1-p};$$

- Ex.: Para $n = 8$ a 5% temos dos valores tabelados que $\omega_{5\%} = 6$. Logo:

$$\omega_{95\%} = \frac{n(n+1)}{2} - \omega_{5\%} = \frac{8 \cdot 9}{2} - 6 = 36 - 6 = 30;$$

- Quando n é grande (sem empates nos postos), a distribuição de T^+ , sob H_0 , é aproximadamente normal com média $\mu = \frac{n(n+1)}{4}$ e variância $\sigma^2 = \frac{n(n+1)(2n+1)}{24}$;
- Alguns autores consideram:
 - $T^+ = \sum_i R_i \mathbb{I}(d_i = 1)$;
 - $T^- = \sum_i R_i \mathbb{I}(d_i = -1)$;
 - E a estatística de teste bilateral é:

$$T_1 = \min\{T^-, T^+\}$$

Estatística de teste III

- Alternativamente, pode-se propor a estatística:

$$T = \frac{\sum_{i=1}^n R_i}{\sqrt{\sum_{i=1}^n R_i^2}},$$

que usa a soma de todos os postos R_i com seus sinais "+" ou "-".

- Caso não haja empates e $n > 50$, a estatística T se reduz a:

$$T = \frac{\sum_{i=1}^n R_i}{\sqrt{n(n+1)(2n+1)/6}};$$

- A distribuição de T é aproximadamente Normal-padrão;
- Se existem muitos empates entre os postos ou se $n > 50$, a estatística T deverá ser usada.

Regra de decisão

- Para $n < 50$: Utilizar os quantis da distribuição de T^+ disponíveis na tabela A12 em [Conover, 1996]. Para o nível de significância α :
 - **Bilateral**: Rejeita-se H_0 se $T^+ < \omega_\alpha$ ou $T^+ > \omega_{1-\alpha}$;
 - **Unilateral à direita**: Rejeita-se H_0 se $T^+ > \omega_{1-\alpha}$;
 - **Unilateral à esquerda**: Rejeita-se H_0 se $T^+ < \omega_\alpha$;
- Para grandes amostras, utilizar a decisão usual do teste para v.a. Normal-padrão;

Comentários:

- O teste do sinal de Wilcoxon pode também funcionar como um teste de mediana, i.e. quando temos apenas uma amostra, Y_1, \dots, Y_n e temos interesse em testar as hipóteses:
 H_0 : A mediana de Y é igual a m **versus** H_1 : A mediana de Y é não igual a m ;
- Para realizar o teste, forma-se os pares: $(m, Y_1), \dots, (m, Y_n)$;
- É possível simular a função poder para este teste. Veja [Lehmann and D'abrera, 2006] e [Gibbons and Chakraborti, 2011](Seção 5.7.3);
- Para cálculo de tamanho de amostra e cálculo de intervalos de confiança, veja [Gibbons and Chakraborti, 2011](Seções 5.7.4 e 5.7.5);

Aspecto computacional

No R:

```
wilcox.test(x, # grupo 1
y, # grupo 2
paired=TRUE, # faz o teste pareado
exact=TRUE, # nao fazer a aprox. normal
alternative = c('two.sided', 'less', 'greater'))
```

No SAS:

```
DATA database;
INPUT var1 var2 @@;
diff=var1-var2;
DATALINES;
;
PROC UNIVARIATE DATA=database;
VAR diff;
RUN;
```

O proc univariate vai calcular automaticamente as estatísticas:

- do teste t,
- do teste dos sinais: **Sign (M)**,
- e do teste dos postos com sinais: **Signed Rank (S)**.

O SAS utiliza uma tabela diferente das tabelas A19 e A20 de [Conover, 1996], pois sua estatística de teste é centralizada na média dos postos: $n(n+1)/4$. Enquanto [Conover, 1996] utiliza a estatística (2), O SAS centraliza T^+ na sua média e utiliza a estatística:

$$S = T^+ - n(n+1)/4.$$

Os níveis descritivos e conclusões, no entanto, são equivalentes.

Exemplo 1

Avaliar se o nível de ansiedade de 10 crianças que serão submetidas a alguma intervenção cirúrgica se reduz após serem submetidas a uma orientação profissional sobre o que é uma cirurgia. O nível de ansiedade, no pré-teste e no pós-teste, nas crianças foi medido em uma escala de 1 a 7. Considere $\alpha = 5\%$.

Antes da orientação (Pré-Teste)	5	6	6	7	6	5	4	4	3	3
Depois da orientação (Pós-Teste)	5	6	3	3	4	2	2	5	6	4

Exemplo 2 - [Korosteleva, 2014]

Deseja-se avaliar se o preço de imóveis mudou ao longo de dois anos. Os dados são apresentados na Tabela 34 a seguir

mês	Ano 1	Ano 2	Sinal da diferença	Posto(r_i)
Jan	572	593	-21	4
Fev	572	588	-16	3
Mar	578	586	-8	1
Abr	591	581	+10	2
Mai	601	576	+25	5
Jun	606	568	+38	8
Jul	602	560	+42	9
Ago	600	555	+45	11
Set	602	553	+49	12
Out	604	560	+44	10
Nov	602	566	+36	7
Dez	598	571	+27	6

Tabela 1: Índice de preços de imóveis em dois anos

Teste de Mann-Whitney (*Wilcoxon's rank-sum test*)

Introdução

- Este teste é equivalente ao teste de Wilcoxon da soma de postos (*rank-sum test*).
- Considere duas amostras aleatórias $X = \{X_1, X_2, \dots, X_n\}$ e $Y = \{Y_1, \dots, Y_m\}$.
- Se unirmos estas duas amostras em uma só, teremos um conjunto com $m + n = N$ observações, as quais podemos atribuir postos R_1, R_2, \dots, R_N .
- Defina agora

$$Z = (Z_1, Z_2, \dots, Z_N),$$

com $Z_i = 1$ se a i -ésima amostra vem do conjunto X e $Z_i = 0$ caso contrário, para $i = 1, 2, \dots, N = m + n$.

Exemplo: $(X_1, X_2, X_3, X_4) = (2, 9, 3, 4)$ e $(Y_1, Y_2, Y_3) = (1, 6, 10)$;

Vetor combinado ordenado: $(1, 2, 3, 4, 6, 9, 10) = (Y_1, X_1, X_3, X_4, Y_2, X_2, Y_3)$;

Logo, $Z = (0, 1, 1, 1, 0, 1, 0)$.

Suposições:

- As amostras são aleatórias e independentes entre si;
- A mensuração é ao menos ordinal.

Construção do teste I

A estatística do teste é dada por :

$$T = \sum_{i=1}^N R_i \cdot Z_i,$$

que é equivalente à soma dos postos de X.

- Para $n \leq 20$ e $m \leq 20$, alguns quantis inferiores (ω_{1-p}) da distribuição de T podem ser obtidos, por exemplo, em [Conover, 1996], Tabela A7;
- Os quantis superiores, simétricos aos da tabela (ω_p), podem ser obtidos através da seguinte relação:

$$\omega_p = n(m + n + 1) - \omega_{1-p};$$

Construção do teste II

- Para n, m maiores do que 20, uma aproximação à normal pode ser calculada:

$$\omega_p \approx \frac{n(N+1)}{2} + z_p \sqrt{\frac{mn(N+1)}{12}},$$

em que z_p é o p-quantil da normal padrão.

- Se houver muitos empates, uma correção é aplicada à estatística T (padronização pela média e desvio padrão):

$$T_1 = \frac{T - n(N+1)/2}{\sqrt{\frac{nm}{N(N-1)} \sum_{i=1}^N R_i^2 \frac{nm(N+1)^2}{4(N-1)}}}, \quad (3)$$

que tem distribuição normal-padrão.

Hipóteses do teste

O Teste pode ser bilateral ou unilateral:

- Hipótese nula:

$$H_0 : F_X(t) = G_Y(t), \text{ para todo } t \in \mathbb{R}.$$

- Hipótese alternativa:

- Bilateral:

$$H_1 : F_X(t) \neq G_Y(t), \text{ para pelo menos um valor de } t \in \mathbb{R};$$

- Unilateral à esquerda:

$$H_1 : F_X(t) < G_Y(t), \text{ para pelo menos um valor de } t \in \mathbb{R}.$$

- Unilateral à direita:

$$H_1 : F_X(t) > G_Y(t), \text{ para pelo menos um valor de } t \in \mathbb{R};$$

Regra de decisão

Rejeitamos H_0 ao nível α se:

- T for menor do que o $(\alpha/2)$ -quantil ou maior do que o $(1 - \alpha/2)$ -quantil no teste bilateral com $n, m \leq 20$.
- T for menor do que o α -quantil no teste unilateral à esquerda;
- T for maior do que o $(1 - \alpha)$ -quantil no teste unilateral à direita.

Se $n > 20$ ou $m > 20$, a aproximação normal em (3) deve ser utilizada e o teste é o teste normal usual. O p -valor pode ser calculado pela aproximação normal:

$$p = P \left(Z \leq \frac{T + 1/2 - n(N + 1)/2}{\sqrt{nm(N + 1)/12}} \right),$$

para o caso unilateral. No caso bilateral o p -valor será dado por $2p$.

Aspecto computacional

No R:

```
wilcox.test(x, # grupo 1  
y, # grupo 2  
paired=FALSE, # faz o teste da soma dos postos  
exact=TRUE, # nao fazer a aprox. normal  
alternative = c('two.sided', 'less', 'greater'))
```

No SAS:

```
PROC NPAR1WAY DATA=dados WILCOXON;  
CLASS tipo_amostra;  
VAR var_analisada;  
EXACT;  
RUN;
```

Exemplo 1 - [Conover, 1996], adaptado

Nove peças, 4 do tipo A e 5 do tipo B, foram testadas de acordo com seu grau de dureza e ordenadas da mais mole à mais dura, como mostra a tabela a seguir:

Tipo de peça	A	A	A	B	A	B	B	B	B
Posto	1	2	3	4	5	6	7	8	9

Deseja-se testar as seguintes hipóteses:

H_0 : Os tipos tem igual dureza *versus* H_1 : Os tipos tem dureza diferente.

O teste de Mann-Whitney para $n=4$ e $m=5$, nos dá, ao nível 5%, nos dá $\omega_{.025} = 12$ e portanto $\omega_{.975} = 4 * 10 - 12 = 28$.

Por outro lado, a estatística do teste é a soma dos postos do grupo A: $T = 1 + 2 + 3 + 5 = 11$. Logo, rejeitamos a hipótese nula a 5%.

O *p-valor* pode ser calculado de forma aproximada a partir da aproximação normal:

$$p = 2 * P(Z \leq (11 + 1/2 - 4 * 10/2) / (\sqrt{4 * 5 * 10/12})) = 2 * P(Z \leq -2.0821) = 0.038$$

Exemplo 2 - [Korosteleva, 2014]

Deseja-se avaliar se a participação em um programa de reforço de aprendizagem é eficaz e para isto 20 estudantes foram avaliados, 10 participantes e 10 não participantes do programa, de acordo com seu desempenho. Os dados são apresentados na Tabela 43 a seguir:

Participou Desempenho	Posto (rank)	Não participou Desempenho	Posto (rank)
3.98	18	3.42	7
3.45	9.5	2.56	2
3.66	13	2.00	1
3.78	14.5	3.19	6
3.90	16	3.00	4
4.00	19.5	3.56	11.5
3.78	14.5	3.56	11.5
3.12	5	4.00	19.5
3.45	9.5	2.78	3
3.97	17	3.44	8

Tabela 2: Desempenho de 20 estudantes de acordo com a participação em um programa de reforço de aprendizagem

Conclua se o programa foi eficaz ao nível 1%.

Exemplo 3, adaptado

Considere a tabela a seguir, contendo pesos de 18 indivíduos, 9 homens e 9 mulheres, em quilos.

Mulheres	68.9	61.2	73.3	51.8	63.4	64.6	48.4	48.8	48.5
Homens	67.8	60	63.4	76	89.4	73.3	67.3	61.3	62.4

Fonte: <http://www.sthda.com/english/wiki/unpaired-two-samples-wilcoxon-test-in-r>.

Queremos avaliar se o peso de mulheres e homens é igual ou não. faça o teste de Mann-Whitney e conclua ao nível 5%.

Acknowledgements

Agradecemos ao prof. Anderson Ara pela disponibilização de seu material didático, no qual nos baseamos para a elaboração destes slides. Alguns trechos desta apresentação são replicados de seu material.

Referências



Conover, W. J. (1996).
Practical nonparametric statistics.
John Wiley and sons, 3 ed. edition.



Gibbons, J. D. and Chakraborti, S. (2011).
Nonparametric statistical inference.
Crc Press, Cop.



Korosteleva, O. (2014).
Nonparametric methods in statistics with SAS applications.
Crc Press.



Lehmann, E. L. and D'abrera, H. J. M. (2006).
Nonparametrics : statistical methods based on ranks.
Springer.