Departamento de Estatística e Matemática Aplicada da UFC
CC0293- Análise Não Paramétrica
Testes de Posição Aplicáveis a duas Amostras Independentes- 21/ 03/2023
Professor: Maurício Mota

Introdução:

Vamos apresentar o capítulo 4 do livro Estatística Experimental nâo Paramétrica do professor Humberto de Campos. Vamos fazer uso do **R**.

4.1-Teste da Soma das Ordens-Wilcoxon

4.1.1-Generalidades

Comumente, ao confrontarmos dois tratamentos, o nosso interesse maior é o de averiguar se existe superioridade de um sobre outro quanto à natureza dos dados levantados.

Para este fim, são empregados os testes de posição, envolvendo duas populações X, grupo controle e Y o grupo tratamento.

No caso de populações independentes, destaca-se no campo não-paramétrico, pelo seu poder, o teste de **Wilcoxon**, introduzido por este autor em 1945, com a denominação "Teste da Soma das ordens"(Rank Sum Test).

4.1.2-Pressuposições

a) As duas amostras são casualizadas e independentes;

b) As variáveis (X e Y) são contínuas

4.1.3-Método

Consideramos as amostras $X_1, X_2, \ldots, X_n$ do nosso grupo controle e $Y_1, Y_2, \ldots, Y_m$ do nosso grupo tratamento e, segundo HOLLANDER e Wolf(1973), admitimos os modelos:

$$X_i = e_i \qquad\qquad\qquad\qquad (i = 1, 2, \ldots, n).$$

e

$$Y_j = \Delta + e_{n+j} \qquad\qquad\qquad\qquad (j = 1, 2, \ldots, m),$$

em que Δ representa o efeito do tratamento.

Procedemos à classificação conjunta das $N = n + m$ observações, em ordem crescente.

Definimos:

$$W = \sum_{j=1}^{m} O_j,$$

em que O_j representa a ordem de Y_j na classificação conjunta das $N = n + m$ observações.

As nossas hipóteses são

$$H_0 : \Delta = 0$$

contra uma das alternativas:

$$H_1 : \Delta > 0, \quad \text{ou } \; H_1 : \Delta > 0 \; \text{ ou } \; H_1 : \Delta \neq 0.$$

Para testarmos, ao nível α de significância:

Situação 1:

$$H_0 : \Delta = 0 \quad vs \quad H_1 : \Delta > 0.$$

Rejeitamos H_0 se $W \geq W_{1-\alpha}$
em que

$$P_0 \left(W \geq W_{1-\alpha} \right) = \alpha.$$

Situação 2:

$$H_0 : \Delta = 0 \quad vs \quad H_1 : \Delta < 0.$$

Rejeitamos H_0 se $W \leq W_{1-\alpha}$

em que

$$P_0\left(W \leq W_{1-\alpha}\right) = \alpha.$$

Situação 3:

$$H_0 : \Delta = 0 \quad vs \quad H_1 : \Delta \neq 0.$$

Rejeitamos H_0 se $W \geq W_{1-\alpha_1}$
ou se $W \leq W_{\alpha_2}$
em que

$$\alpha_1 + \alpha_2 = \alpha,$$

e geralmente , se considera

$$\alpha_1 = \alpha_2 = \frac{\alpha}{2}.$$

Os limites W_α e $W_{1-\alpha}$ são encontrados na tabela 8.

A aplicação dos testes unilaterais é recomendável quando já, a priori, esperamos um comportamento unidirecional de um dos tratamentos em relação ao outro. Na garnde maioria dos casos , não temos prévio conhecimento de qual dos tratamentos é esperado ser melhor ou pior,e, consequentemente, devemos aplicar o teste bilateral.

Embora empregando o teste bilateral para a comparação dos dois tratamentos A e B , decidimos:

a) A é superior a B, se $W_A \geq W_{1-\alpha_1}$.

b) A é inferior a B, se $W_A \leq W_{\alpha_2}$.

c) A não difere de B se $W_{\alpha_2} < W_A < W_{1-\alpha_1}$.

Alguns comentários se fazem necessários:

Comentário 1: Os valores máximo e mínimo de W são obtidos quando a variável Y_j ocupa as **m** primeiras e as últimas **m** posições na classificação conjunta das N observações.

Tais valores correspondem aos seguintes arranjos:

Para obter W_{min} teremos:

$$YY \ldots YXX \ldots XX.$$

Assim

$$W_{min} = \sum_{j=1}^{m} j = 1 + 2 + \ldots + m = \frac{m(m+1)}{2}.$$

Para obter W_{max} teremos:

$$XX \ldots XYY \ldots YY.$$

Assim

$$W_{max} = \sum_{j=n+1}^{N} j = (n+1) + (n+2) + \ldots + (m+n) = \frac{(n+1+N)m}{2},$$

ou

$$W_{max} = \frac{(2n+m+1)m}{2}.$$

Comentário 2: A média(mediana) dos possíveis valores de W, sob H_0, é:

$$W_{med} = \frac{m(m+n+1)}{2} = \frac{m(N+1)}{2}.$$

O livro não prova este resultado.

Comentário 3: A amplitude do intervalo, onde varia W é:

$$A_W = W_{max} - W_{min} = \frac{(2n+1+m)m}{2} - \frac{m(m+1)}{2} = \frac{2mn + m + m^2 - m^2 - m}{2} = mn.$$

Comentário 4: A estatística W é uma varável discreta.

Comentário 5: Consideramos n, como o tamanho menor da amostra.

Comentário 6: A distribuição de W, sob H_0, é simétrica em relação a sua média. Esta propriedade permite-nos concluir que :

$$W_\alpha = m(n+m+1) - W_{1-\alpha},$$

ou seja

$$P_0\left(W \leq W_\alpha\right) = P_0\left(W \leq m(n+m+1) - W_{1-\alpha}\right).$$

4.1.4-Aproximação Normal

Sabemos que

$$\mu = E(W) = \frac{m(N+1)}{2} \quad \sigma^2 = \frac{nm(N+1)}{12}.$$

Quando m e n tendem a infinito temos:

$$W* = \frac{W-\mu}{\sigma} \sim N(0,1,),$$

aproximadamente.

Portanto, para grandes amostras (m e n grandes) utilizamos a aproximação Normal , através da estatística $W*$.

Assim, para as hipóteses:

$$H_0 : \Delta = 0 \quad vs \quad H_1 : \Delta > 0.$$

Rejeitamos H_0 se $W^* \geq z_\alpha$,

em que z_α é o limite superior da distribuição normal ao nível α de significância.

Em casos de maior precisão é recomendável aplicarmos a correção de continuidade, através da fórmula:

$$W* = \frac{(W \pm 0,5) - \mu}{\sigma} \sim N(0,1)$$

aproximadamente.

O sinal positivo se aplica aos limites inferiores e sinal negativo

aos limites superiores. Isto se justifica admitindo que cada valor w de W assumido pela variável discreta seja o ponto médio do intervalo $(w - 0,5, w + 0,5)$.

A título de ilustração , admitamos $m = 4$, $n = 8$ e $w = 35$.

Queremos calcular a probabilidade exata $P(W \geq 35)$ e pela aproximação normal com e sem fator de correção.

Temos que $N = m + n = 12$.

A média de W é dada por:

$$\mu = \frac{m(N+1)}{2} = \frac{4(13)}{2} = 26,$$

A variância de W é dada por:

$$\sigma^2 = \frac{mn(N+1)}{122} = \frac{4 \times 8(13)}{12} = \frac{104}{3},$$

Vamos calcular a probabilidade de:

O Valor mínimo de W é dado por:

$$w_{min} = \frac{m(m+1)}{2} = \frac{4 \times 5}{2} = 10.$$

$$p = P(W \geq 35) = P(U \geq 35 - 10) = P(U \geq 25) = 1 - P(U \leq 24).$$

Vamos utilizar pacote R:

```
> m=4;n=8;N=m+n;N
[1] 12
>
>
> mu=m*(N+1)/2;mu
[1] 26
>
> sigma2=m*n*(N+1)/12;sigma2
[1] 34.66667
> require(MASS)
> fractions(sigma2)
[1] 104/3
> sigma=sqrt(sigma2);sigma
[1] 5.887841
>
>
>
> ### probabilidade exata
>
> w_min=m*(m+1)/2;w_min
[1] 10
>
> ##P(W>=35)=P(U>=35-10)=P(U>=25)=1- P(U <=24).
>
>
>
>
> pex=1-pwilcox(24,8,4);pex
[1] 0.07676768
>
> round(pex,3)
```

```
[1] 0.077
>
>
> z=(35-mu)/sigma;z;round(z,2)
[1] 1.528574
[1] 1.53
>
>
> pasc=1-pnorm(1.53);pasc;round(pasc,4)
[1] 0.06300836
[1] 0.063
>
> z1=(34.5-mu)/sigma;z1;round(z1,2)
[1] 1.443653
[1] 1.44
>
>
> pacc=1-pnorm(1.44);pacc;round(pacc,3)
[1] 0.0749337
[1] 0.075
>
```

Para altos valores de m e n a correção é dispensável.

4.1.5-Empates

Quando ocorrem empates entre os valores de X e Y, utilizamos, par a obtenção de W, a média das ordens dos valores empatados e, como no caso usual, tomamos

$$W = \sum_{j=1}^{m} O_j.$$

Se tivéssemos por exemplo:

X(Controle)	Y(Tratamento)
2,3	1,8
3,2	2,3
3,8	2,3
4,5	3,2

Obteríamos o arranjo:

Amostra	1,8	2,3	2,3	2,3	3,2	3,2	3,8	4,5
Grupo	Y	X	Y	Y	XY	Y	X	X
Posto	1	3	3	3	5,5	5,5	7	8

A soma dos postos do grupo tratamento

$$W = 1 + 3 + 3 + 5,5 = 12,5.$$

Observamos que empates entre valores de X ou entre valores de Y , não afetam o cálculo da estatística W , embora afete sua distribuição nula.

A média é a mesma mas a variância é afetada pelos empates:

A variância é dada por:

$$Var(W) = \frac{mn}{12N(N-1)} \left[N(N^2-1) - \sum_{i=1}^{k} t_i(t_i-1)(t_i+1) \right],$$

em que: N=m+n;

k = número de grupos com empates;

t_i =número de observações no grupo i.

No exemplo temos:

$$k = 2; t_1 = 3; t_2 = 2; m = 4; n = 4; N = 8.$$

$$\mu = \frac{4}{2} = 18.$$

$$\sigma^2 = \frac{16}{12 \times 8 \times 7} \left[8(64-1) - \sum_{i=1}^{2} t_i(t_i-1)(t_i+1) \right],$$

$$\sigma^2 = \frac{1}{42} [504 - 3 \times 2 \times 4 - 2 \times 1 \times 3] = \frac{1}{42} [504 - 24 - 6]$$

$$\sigma^2 = \frac{474}{42} = \frac{79}{7} = 11,29$$

Note que

$$w* = \frac{12,5-18}{\sqrt{11,29}} = -1,64.$$

$$P(W* \leq 12,5) = P(Z < -1,64) = 0,05.$$

```
> m=4;n=4;N=m+n;N
[1] 8
>
>
> mu=m*(N+1)/2;mu
[1] 18
>
>
> t=c(3,2)
>
>
> aux=sum(t*(t-1)*(t+1));aux
[1] 30
>
>
> sigma2=((m*n)/(12*N*(N-1)))*(N*(N^2-1)-aux);sigma2
[1] 11.28571
>
> round(sigma2,2)
[1] 11.29
>
> sigma=sqrt(sigma2);sigma
[1] 3.359422
>
> z=(12.5-mu)/sigma;z
[1] -1.637187
>
> round(z,2)
[1] -1.64
> pnorm(-1.64)
[1] 0.05050258
```

```
>
```

4.1.6- Distribuição nula de W.

A fim de ilustrar a distribuição nula de W, consideremos $m = 2$ e $n = 4$. Assim temos $\binom{6}{2} = 15$ possíveis grupamentos, conforme se verifica a seguir:

Grupamentos	W	Grupamentos	W
TTCCCC	3	CTCCCT	8
TCTCCC	4	CCTTCC	7
TCCTCC	5	CCTCTC	8
TCCCTC	6	CCTCCT	9
TCCCCT	7	CCCTTC	9
CTTCCC	5	CCCCTCT	10
CTCTCC	6	CCCCTT	11
CTCCTC	7		

Podemos obter a tabela:

$W = w$	$P_0(W = w)$	$P_0(W \geq w)$	$P_0(W \leq w)$
3	0,067	1,000	0,067
4	0,067	0,933	0,133
5	0,133	0,867	0,267
6	0,133	0,733	0,400
7	0,200	0,600	0,600
8	0,133	0,400	0,733
9	0,133	0,267	0,867
10	0,067	0,133	0,999
11	0,067	0,067	1,000

Vamos gerar esta tabela usando o pacote R:

```
\en>
> m=2;n=4;N=m+n;m;n;N
[1] 2
[1] 4
[1] 6
>
```

```
> ##Número de grupamentos:
>
>
> choose(N,m)
[1] 15
>
> ###a probabilidade de cada grupamento
>
> p=1/choose(N,m)
>
> p;round(p,3)
[1] 0.06666667
[1] 0.067
>
> w_min=m*(m+1)/2;w_min
[1] 3
>
>
>
> u=0:(m*n);u
[1] 0 1 2 3 4 5 6 7 8
>
> w=u+3;w
[1]  3  4  5  6  7  8  9 10 11
>
> pu=dwilcox(u,m,n)
>
> ###A acumulada de U é dada por:
>
>
> Pu=pwilcox(u,m,n)
>
>
> pw=pu
>
> Pw=Pu
>
>
```

```
> ###Note que
>
>
> ###P(U >=U)=P(U=u)+P(U >u)=P(U=u)+1 -F(u)
>
> Su=dwilcox(u,m,n)+1-pwilcox(u,m,n)
> Sw=Su
>
>
>
> tab=cbind(w,pw,Sw,Pw);tab; round(tab,3)
w          pw         Sw         Pw
[1,]  3 0.06666667 1.00000000 0.06666667
[2,]  4 0.06666667 0.93333333 0.13333333
[3,]  5 0.13333333 0.86666667 0.26666667
[4,]  6 0.13333333 0.73333333 0.40000000
[5,]  7 0.20000000 0.60000000 0.60000000
[6,]  8 0.13333333 0.40000000 0.73333333
[7,]  9 0.13333333 0.26666667 0.86666667
[8,] 10 0.06666667 0.13333333 0.93333333
[9,] 11 0.06666667 0.06666667 1.00000000
w    pw    Sw    Pw
[1,]  3 0.067 1.000 0.067
[2,]  4 0.067 0.933 0.133
[3,]  5 0.133 0.867 0.267
[4,]  6 0.133 0.733 0.400
[5,]  7 0.200 0.600 0.600
[6,]  8 0.133 0.400 0.733
[7,]  9 0.133 0.267 0.867
[8,] 10 0.067 0.133 0.933
[9,] 11 0.067 0.067 1.000
>
```

Na distribuição evidenciamos :

a)

$$P_0(W = w) = P(W = m(N + 1) - w)$$

$$m(N + 1) = 2 \times 7 = 14$$

Assim

$$P_0(W = w) = P(W = 14 - w)$$

$$P_0(W = 3) = P(W = 11) = 0,067.$$

$$P_0(W = 4) = P(W = 10) = 0,067.$$

$$P_0(W = 5) = P(W = 9) = 0,133.$$

$$P_0(W = 6) = P(W = 8) = 0,133.$$

b)

$$P(W \geq w) = P(m(N + 1) - w) = P(W \leq 14 - w).$$

$$P(W \geq 3) = P(W \leq 11) = 1.$$

$$P(W \geq 4) = P(W \leq 10) = 0,933.$$

$$P(W \geq 5) = P(W \leq 9) = 0,867.$$

$$P(W \geq 6) = P(W \leq 8) = 0,733.$$

$$P(W \geq 7) = P(W \leq 7) = 0,600.$$

$$P(W \geq 8) = P(W \leq 6) = 0,400.$$

$$P(W \geq 9) = P(W \leq 5) = 0,267.$$

$$P(W \geq 10) = P(W \leq 4) = 0,133.$$

$$P(W \geq 11) = P(W \leq 3) = 0,167.$$

c) a distribuição de W é simétrica em torno se sua média:

$$\mu = \frac{m * (N+1)}{2} = 7.$$

$$P(W=8) = P(W=7+1) = P(W=7-1) = P(W=6) = 0,133 = \frac{2}{15}.$$

$$P(W=9) = P(W=7+2) = P(W=7-2) = P(W=5) = 0,133 = \frac{2}{15}.$$

$$P(W=10) = P(W=7+3) = P(W=7-3) = P(W=4) = 0,067 = \frac{1}{15}.$$

$$P(W=11) = P(W=7+4) = P(W=7-4) = P(W=3) = 0,067 = \frac{1}{15}.$$

No caso de observações empatadas a distribuição nula de W se altera e, consequentemente, os níveis de significância das tabelas usuais, sem empates são apenas aproximados. A título de ilustração, admitamos $m = 2$ e $n = 3$ com a terceira e a quarta estatísticas de ordem empatadas.

Assim os postos valem valem:

$$1; 2; 3,5; 35; 5.$$

Os arranjos e os valores assumidos por W são dados a seguir:

ARRANJOS	W	ARRANJOS	W
TTCCC	1+2=3	CTCTC	2+3,5=5,5
TCTCC	1+3,5=4,5	CTCCT	2+5=7
TCCTC	1+3,5=4,5	CCTTC	3,5+3,5=7
TCCCT	1+5=6	CCTCT	3,5+5=8,5
CTTCC	2+3,5=5,5	CCCTT	3,5+5=8,5

Cada arranjo tem probabilidade $0,1$. A distribuição nula de W é dada por:

$W = w$	$P(W = w)$	$P(W \geq w)$	$P(W \leq w)$
3	0,1	1,0	0,1
4,5	0,2	0,9	0,3
5,5	0,2	0,7	0,5
6	0,1	0,5	0,6
7	0,2	0,40	0,8
8,5	0,2	0,2	1,0

Note que a simetria foi perdida no caso de empates.

Assim por exemplo, se $w = 8,5$

$$P(W \geq 8,5) = 0,2$$

Para olhar na tabela temos:

Pela tabela do livro do Humberto temos:

$$P(W \geq 9) = 0,1 \quad e \quad P(W \geq 8) = 0,2$$

Olhando no R temos:

```
> 1-pwilcox(5,2,3)
[1] 0.1
>
```

$$P(W \geq 8,5) = P(W \geq 9) = P(W - 3 \geq 9 - 3) = P(U \geq 6) = 1 - P(U \leq 5) = 1 - 0,9 = 0,1.$$

A tabela de Wilcoxon do R não leva em conta os empates.

Observamos que , se fossem ouras duas estatísticas empatadas, a distribuição nula já sofreria alterações, demonstrando claramente, a complexidade do problema.

4.1.7- Estimativa de Δ.

Baseados nos modelos:

$$X_i = e_i, \; i = 1, 2, \ldots, n$$

e

$$Y_j = \Delta + e_{n+j}, \; j = 1, 2, \ldots, m$$

referidos anteriormente, HODGES e LEHMANN (1950) apresentaram a seguinte metodologia para estimar Δ, ou seja:

a) Determinamos as mn diferenças do tipo:

$$U_{ij} = Y_j - X_i,$$

classificando-as em ordem crescente

b) Obtemos assim as estatísticas de ordem:

$$U^{(1)} \leq U^2 \leq ldots \leq U^{mn};$$

c) A estimativa de Δ é:

$$\hat{\Delta} = \text{mediana dos } U^(i's).$$

4.1.8- intervalo de confiança para Δ.

Vamos apresentar somente o processo analítico:

Para determinar o intervalo de confiança para Δ, com coeficiente de confiança $\gamma 1 - \alpha$, MOSES(1956) apresenta a seguinte marcha:

1) Determinamos

$$C_\alpha = \frac{m(N+n+1)}{2} - W_{1-\frac{\alpha}{2}} + 1 = W_{\frac{\alpha}{2}} - \frac{m(m+1)}{2} + 1.$$

2) Os extremos do intervalo de confiança são, então, dados por:

$$\Delta_I = U^{(C_\alpha)} \ \ e \ \ \Delta_S = U^{(mn+1-C_\alpha)}.$$

Quando utilizamos a a aproximação normal tomamos:

$$C_\alpha = \frac{mn}{2} - z_{\frac{\alpha}{2}} \sqrt{\frac{mn(N+1)}{12}}.$$

A título de ilustração tomemos as amostras:

$X_1 = 3,0$	$Y_1 = 2,5$
$X_2 = 3,8$	$Y_2 = 3,5$
$X_3 = 5,0$	$Y_3 = 5,4$
$X_4 = 5,6$	$Y_4 = 7,8$

Determine o intervalo de confiança para Δ ao nível $\gamma = 1 - 0,058 = 1 - \alpha$.
Temos que

$$\frac{\alpha}{2} = 0,029.$$

Além disso

$$m = n = 4 \ \ e \quad N = 8.$$

O livro diz que :

$$W_{0,971} = 25.$$

Vamos mostrar que essa afirmação é verdadeira:

```
>
> u=0:(m*n);u
[1]  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
>
> pu=dwilcox(u,m,n)
> Pu=pwilcox(u,m,n)
> Su=pu+1-Pu
> tab=cbind(u,pu,Pu,Su)
> round(tab,3)
u     pu     Pu     Su
[1,]    0 0.014 0.014 1.000
[2,]    1 0.014 0.029 0.986
[3,]    2 0.029 0.057 0.971
[4,]    3 0.043 0.100 0.943
[5,]    4 0.071 0.171 0.900
[6,]    5 0.071 0.243 0.829
[7,]    6 0.100 0.343 0.757
[8,]    7 0.100 0.443 0.657
[9,]    8 0.114 0.557 0.557
[10,]   9 0.100 0.657 0.443
[11,] 10 0.100 0.757 0.343
[12,] 11 0.071 0.829 0.243
[13,] 12 0.071 0.900 0.171
[14,] 13 0.043 0.943 0.100
[15,] 14 0.029 0.971 0.057
[16,] 15 0.014 0.986 0.029
[17,] 16 0.014 1.000 0.014
>
>
> w=u+10
```

```
> pw=pu;Pw=Pu;Sw=Su
>
> tab1=cbind(w,pw,Pw,Sw)
> round(tab1,3)
w    pw    Pw    Sw
[1,] 10 0.014 0.014 1.000
[2,] 11 0.014 0.029 0.986
[3,] 12 0.029 0.057 0.971
[4,] 13 0.043 0.100 0.943
[5,] 14 0.071 0.171 0.900
[6,] 15 0.071 0.243 0.829
[7,] 16 0.100 0.343 0.757
[8,] 17 0.100 0.443 0.657
[9,] 18 0.114 0.557 0.557
[10,] 19 0.100 0.657 0.443
[11,] 20 0.100 0.757 0.343
[12,] 21 0.071 0.829 0.243
[13,] 22 0.071 0.900 0.171
[14,] 23 0.043 0.943 0.100
[15,] 24 0.029 0.971 0.057
[16,] 25 0.014 0.986 0.029
[17,] 26 0.014 1.000 0.014
>
>
>
```

Note que na tabela1 temos:

$$P(W_s \geq 25) = 0,029.$$

$$C_{0,058} = \frac{m(N+n+1)}{2} - W_{1-\frac{\alpha}{2}} + 1 = \frac{4 \times 13}{2} - 25 + 1 = 26 - 25 + 1 = 2.$$

$$mn + 1 - C_{0,058} = 16 + 1 - 2 = 15.$$

A estimativa pontual de Δ é dada por:
Como $mn = 16$ temos:

$$\hat{\Delta} = \frac{U^{(8)} + U^{(9)}}{2} = \frac{-0,2+0,4}{2}\frac{0,2}{2} = 0,1.$$

$$IC(\Delta, 94, 42\%) = [U^{(2)}\ ,\ U^{(15)}] = [-2,5; 4].$$

A solução geral pelo R:

```
wilcox.test(Y,X,conf.level=1-0.058, conf.int = TRUE)

Wilcoxon rank sum exact test

data:  Y and X
W = 8, p-value = 1
alternative hypothesis: true location shift is not equal to 0
94.2 percent confidence interval:
-2.5  4.0
sample estimates:
difference in location
0.1
```

O passo a passo:

```
X
[1] 3.0 3.8 5.0 5.6
>
> Y
[1] 2.5 3.5 5.4 7.8
>
> L1=Y[1]-X;L1
[1] -0.5 -1.3 -2.5 -3.1
> L2=Y[2]-X;L2
[1]  0.5 -0.3 -1.5 -2.1
> L3=Y[3]-X;L3
[1]  2.4  1.6  0.4 -0.2
> L4=Y[4]-X;L4
[1] 4.8 4.0 2.8 2.2
>
> Aux=c(L1,L2,L3,L4);Aux
[1] -0.5 -1.3 -2.5 -3.1  0.5 -0.3 -1.5 -2.1  2.4  1.6  0.4 -0.2  4.8  4.0  2.8
[16]  2.2
```

```
>
> U=sort(Aux);U
[1] -3.1 -2.5 -2.1 -1.5 -1.3 -0.5 -0.3 -0.2  0.4  0.5  1.6  2.2  2.4  2.8  4.0
[16]  4.8
>
> matrix(U,nrow=4,ncol=4)
[,1] [,2] [,3] [,4]
[1,] -3.1 -1.3  0.4  2.4
[2,] -2.5 -0.5  0.5  2.8
[3,] -2.1 -0.3  1.6  4.0
[4,] -1.5 -0.2  2.2  4.8
>
>
> Delta_est=median(U);Delta_est
[1] 0.1
>
> C_alfa=m*(N+n+1)/2 -25 +1;C_alfa
[1] 2
>
>
> LS=(m*n) +1-C_alfa;LS
[1] 15
>
> U_I=U[2];U_I
[1] -2.5
> U_S=U[15];U_S
[1] 4
>
```

4.1.9- Exemplos.

Vamos fazer o Exemplo 1:

Exemplo 1 Um lote de sementes de milho foi tratado com um determinado produto químico com o objetivo de aumentar o vigor dos "seedlings". Após o tratamento foram semeados oito canteiros com sementes não tratados e cinco com sementes tratadas. Duas semanas após a germinação foram tomadas, ao acaso, 50 plantas de cada canteiro, que foram posteriormente pesadas. Os pesos(em gramas) obtidos foram o que se seguem:

Não Tratadas (X)	Tratadas (Y)
103,7 ; 93,2	98,7
88,5;81,4	112,4
75,4;78,1	117,3
97,8;105,4	102,5
	114,3

a) Teste a eficiência do tratamento.

b) Estime o efeito do tratamento e determine seu intervalo de confiança de 95%.

Solução: Vamos começar direto no R:

Nosso controle $C = X$ e nosso tratamento $T = Y$.

Vamos fazer uma diagrama box-plot para os dois grupos. Percebe-se claramente que a mediana do grupo Tratado é amaior que a medina do grupo controle.

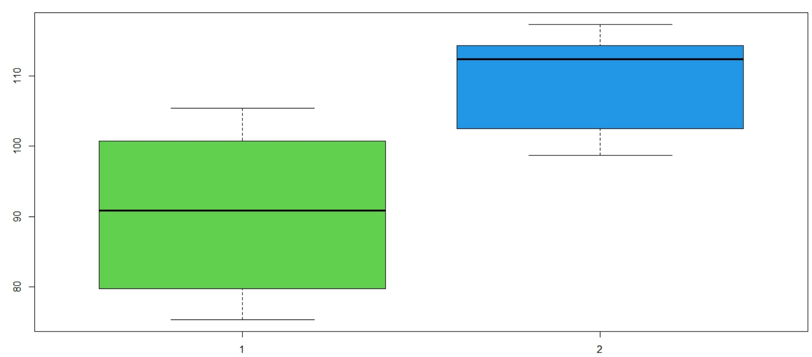

Figura 1:

A solução direta no R é sem emoção. Veja:

```
>
> Con=c(103.7,93.2,88.5,81.4,75.4,78.1,97.8,105.4)
>
> n=length(Con);n
[1] 8
>
> Trat=c(98.7,112.4,117.3,102.5,114.3)
>
```

```
> m=length(Trat);m
[1] 5
>
> boxplot(Con,Trat,col=c(3,4))
>
>
> #####  H_0: Delta=0 vs H_1: Delta /=0
>
> wilcox.test(Trat,Con)

Wilcoxon rank sum exact test

data:  Trat and Con
W = 36, p-value = 0.01865
alternative hypothesis: true location shift is not equal to 0

>
>
> wilcox.test(Con,Trat)

Wilcoxon rank sum exact test

data:  Con and Trat
W = 4, p-value = 0.01865
alternative hypothesis: true location shift is not equal to 0

>
>
>
> ####Analise as duas saídas!!!!!
>
>
> #####  H_0: Delta=0 vs H_1: Delta >0
>
>
> wilcox.test(Trat,Con, alternative="greater")

Wilcoxon rank sum exact test
```

```
data:  Trat and Con
W = 36, p-value = 0.009324
alternative hypothesis: true location shift is greater than 0

>
>
> wilcox.test(Con,Trat, alternative="greater")  #####epa!!!!!!!!!!

Wilcoxon rank sum exact test

data:  Con and Trat
W = 4, p-value = 0.9946
alternative hypothesis: true location shift is greater than 0

>
>
> wilcox.test(Con,Trat, alternative="less")

Wilcoxon rank sum exact test

data:  Con and Trat
W = 4, p-value = 0.009324
alternative hypothesis: true location shift is less than 0

>
>
>
> ####Estimar Delta:
>
>
>
> wilcox.test(Trat,Con, conf.int = TRUE)

Wilcoxon rank sum exact test

data:  Trat and Con
W = 36, p-value = 0.01865
```

```
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
5.5 34.3
sample estimates:
difference in location
19.35

>
>
```

Vamos explicar detalhadamente cada saída do R:

Temos $m = 5$, $n = 8$ e $N = m + n = 13$.

Vamos ordenar amostra conjunta:

$$75,4 < 78,1 < 81,4 < 88,5 < 93,2 < 97,8 < 98,7 < 102,5 < 103,7 < 105,4 < 112,4 < 114,3 < 117,3$$

Vamos corresponder:

$$C(1)C(2)C(3)C(4)C(5)C(6)T(7)T(8)C(9)C(10)T(11)T(12)T(13)$$

A soma de postos do grupo tratamento vale

$$W_S = 7 + 8 + 11 + 12 + 13 = 51.$$

A soma de postos do grupo controle vale

$$W_R = 1 + 2 + 3 + 4 + 5 + 6 + 9 + 10 = 40.$$

$$W_S + W_r = 51 + 40 = 91 = \frac{13 \times 14}{2} = 91.$$

Note que tanto os valores 51 e 40 não aparecem na saída do R.

O valor mínimo de W_S é:

$$w_{min} = \frac{m(m+1)}{2} = \frac{5 \times 6}{2} = 15.$$

O valor da estatística de Mann-Whitney é dado por:

$$U_s = W_s - w_{min} = 51 - 15 = 36,$$

valor que aparece como resultado do comando $wilcox.test(Trat, Con)$.

A soma de postos do grupo tratamento vale

$$W_R = 7 + 8 + 11 + 12 + 13 = 40.$$

O valor mínimo de W_R é:

$$w_{min} = \frac{n(n+1)}{2} = \frac{8 \times 9}{2} = 36$$

O valor da estatística de Mann-Whitney é dado por:

$$U_r = W_r - w_{min} = 40 - 36 = 4,$$

valor que aparece como resultado do comando $wilcox.test(Con, Trat)$.

O nível descritivo do teste Unilateral será:

$$nd = P(W_s \geq 51) = P(W_s - 15 \geq 51 - 15) = P(U_s \geq 36) = 1 - P(U_s \leq 35) =$$

```
>
> m=5;n=8
> nd=1-pwilcox(35,m,n);nd;round(nd,3)
[1] 0.009324009
[1] 0.009
>
>
```

A fim de estimar o efeito do tratamento e determinar seu intervalo de confiança, organizamos a tabela das diferenças

$$U_{ij} = Y_j - X_i$$

que , já em ordem crescente, são:

-6,7	4,7	9,3	14,0	19,5	23,3	28,8	36,2
-5,0	5,5	10,2	14,6	20,6	23,9	31,0	37,0
-2,9	7,0	10,6	16,5	21,1	24,4	32,9	38,9
-1,2	8,7	11,9	17,3	21,1	25,8	34,3	39,2
0,9	8,9	13,6	19,2	23,1	27,1	35,9	41,9

Desde que $mn = 5 \times 8 = 40$ obtemos a seguinte estimativa do efeito de tratamento :

$$\hat{\Delta} = \frac{U^{(20)} + U^{(21)}}{2} = \frac{19,2 + 19,5}{2} = \frac{38,7}{2} = 19,35,$$

isto é, o tratamento produz um aumento de 19,35 g no peso de 50 plantas .

Essa tabela das diferenças pode ser obtida como:

```
>
> Con;Trat
[1] 103.7  93.2  88.5  81.4  75.4  78.1  97.8 105.4
[1]  98.7 112.4 117.3 102.5 114.3
>
>
>
> L1=Trat[1] -Con;L1
[1] -5.0  5.5 10.2 17.3 23.3 20.6  0.9 -6.7
> L2=Trat[2] -Con;L2
[1]  8.7 19.2 23.9 31.0 37.0 34.3 14.6  7.0
>
> L3=Trat[3] -Con;L3
[1] 13.6 24.1 28.8 35.9 41.9 39.2 19.5 11.9
>
> L4=Trat[4] -Con;L4
[1] -1.2  9.3 14.0 21.1 27.1 24.4  4.7 -2.9
>
> L5=Trat[5] -Con;L5
[1] 10.6 21.1 25.8 32.9 38.9 36.2 16.5  8.9
>
>
> Aux=c(L1,L2,L3,L4,L5);Aux
[1] -5.0  5.5 10.2 17.3 23.3 20.6  0.9 -6.7  8.7 19.2 23.9 31.0 37.0 34.3 14.6
[16]  7.0 13.6 24.1 28.8 35.9 41.9 39.2 19.5 11.9 -1.2  9.3 14.0 21.1 27.1 24.4
[31]  4.7 -2.9 10.6 21.1 25.8 32.9 38.9 36.2 16.5  8.9
>
> U=sort(Aux);U
[1] -6.7 -5.0 -2.9 -1.2  0.9  4.7  5.5  7.0  8.7  8.9  9.3 10.2 10.6 11.9 13.6
[16] 14.0 14.6 16.5 17.3 19.2 19.5 20.6 21.1 21.1 23.3 23.9 24.1 24.4 25.8 27.1
[31] 28.8 31.0 32.9 34.3 35.9 36.2 37.0 38.9 39.2 41.9
>
```

```
>
>
> matrix(U,nrow=5,ncol=8)
[,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,] -6.7  4.7  9.3 14.0 19.5 23.9 28.8 36.2
[2,] -5.0  5.5 10.2 14.6 20.6 24.1 31.0 37.0
[3,] -2.9  7.0 10.6 16.5 21.1 24.4 32.9 38.9
[4,] -1.2  8.7 11.9 17.3 21.1 25.8 34.3 39.2
[5,]  0.9  8.9 13.6 19.2 23.3 27.1 35.9 41.9
>
>
>
> delta_est=median(U);delta_est
[1] 19.35
>
>
```

Agora programar com mais eficiência.

Agora vamos construir o intervalo de confiança para Δ.

Exemplo 2: Em um estudo sobre a determinação do teor de fósforo em duas regiões, A,B, foram coletadas em cada uma delas, dez amostras de solo e procedida a determinação daquele elemento. Os resultados em (e.mg/100 g de solo) foram:

Região A	Região B
0,24;0,29	0,21;0,31
0,37;0,42	0,35; 0,21
0,33;0,37	0,41;0,31
0,35;0,39	0,28;0,28
0,18;0,38	0,21;0,33

Verifique, ao nível de significância $\alpha = 0,052$ se as duas regiões diferem quanto ao teor de fósforo.

Solução:

Vamos testar se

$$H_0 : med(A) = Med(B)\ vs\ \ H_1 : \Delta = med(A) - Med(B) = 0\ \ vs\ H_1 : \Delta = med(A) - Med(B) \neq 0.$$

Vamos inicialmente fazer direto no R:

```
\end{
> A=c(24,29,37,42,33,37,35,19,18,38)/100;A;m=length(A);m
[1] 0.24 0.29 0.37 0.42 0.33 0.37 0.35 0.19 0.18 0.38
[1] 10
>
>
> B=c(21,31,35,21,41,31,28,28,21,33)/100;B;n=length(B);m
[1] 0.21 0.31 0.35 0.21 0.41 0.31 0.28 0.28 0.21 0.33
[1] 10
>
> C=c(A,B)
> Co=sort(C);Co
[1] 0.18 0.19 0.21 0.21 0.21 0.24 0.28 0.28 0.29 0.31 0.31 0.33 0.33 0.35 0.35
[16] 0.37 0.37 0.38 0.41 0.42
>
> table(C)####Note os empates!!!!!!!!
C
0.18 0.19 0.21 0.24 0.28 0.29 0.31 0.33 0.35 0.37 0.38 0.41 0.42
1    1    3    1    2    1    2    2    2    2    1    1    1
>
> PostoC=rank(Co); PostoC
[1]  1.0  2.0  4.0  4.0  4.0  6.0  7.5  7.5  9.0 10.5 10.5 12.5 12.5 14.5 14.5
[16] 16.5 16.5 18.0 19.0 20.0
>
>
> Ao=sort(A);Ao
[1] 0.18 0.19 0.24 0.29 0.33 0.35 0.37 0.37 0.38 0.42
>
> PostoA=c(1,4,5,11,12,13,16,17,18,19)
> W_s=sum(PostoA);W_s
[1] 116
> w_min=m*(m+1)/2;w_min
[1] 55
>
> U_s=W_s-w_min;U_s
[1] 61
```

```
>
>
> wilcox.test(A,B,conf.int=TRUE)

Wilcoxon rank sum test with continuity correction

data:  A and B
W = 61, p-value = 0.4258
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
-0.04001978  0.09995229
sample estimates:
difference in location
0.03008498

Warning messages:
1: In wilcox.test.default(A, B, conf.int = TRUE) :
não é possível computar o valor de p exato com o de desempate
2: In wilcox.test.default(A, B, conf.int = TRUE) :
impossível computar os intervalos de confiança exatos com empate
>
>
```

4.1.10- Exercícios Propostos.

1) Estruture a distribuição nula de W , para $n = 4$ e $m = 3$, com a quarta, quinta e sexta estatísticas de ordem empatadas. Confronte com a tabela usual.

2) Foram feitas determinações de Brix para duas variedades (A,B) de cana-de-açúcar, tomando-se, para cada uma delas ,oito colmos distintos. Os resultados permitiram o seguinte arranjo:

$$X < X < X < Y < X < X < Y < Y < X < X < X < Y < Y < Y < Y < Y$$

não havendo empates.

Verifique se as duas variedades diferem quanto ao Brix.

3) Dois tipos de cirurgiões discutiam sobre o grau de dificuldade das operações de apêndice e das cesarianas.

   Um dos grupos afirmava que as cesarianas eram mais demoradas e o outro afirmava 0 contrário. Foram então tomados os tempos, em minutos, gastos nos dois tipos de operação, com os seguintes resultados:

Apêndice	70;62;76
Cesariana	70;122;122;70;137;100

   a) Obtenha a distribuição de W com a configuração de empates apresentada.

   b) Baseado nos dados, a que grupo daria razão? Calcule o nível descritivo exato usando a tabela obtida no item **a**.

   c) Obtenha a estimativa de Δ e seu intervalo de confiança, ao nível $1 - \alpha = 0,834$.

4)

5)

6)

7)