



Universidade de Brasília
IE- Departamento de Estatística
Métodos Estatísticos 2

Juliana Magalhães Rosa

Teste de Mann-Whitney/ Wilcoxon para Duas Amostras

Brasília

2020

Sumário

1 Introdução	2
2 Estatística Ws	3
3 Estatística U	5
4 Considerações Finais	7

1 Introdução

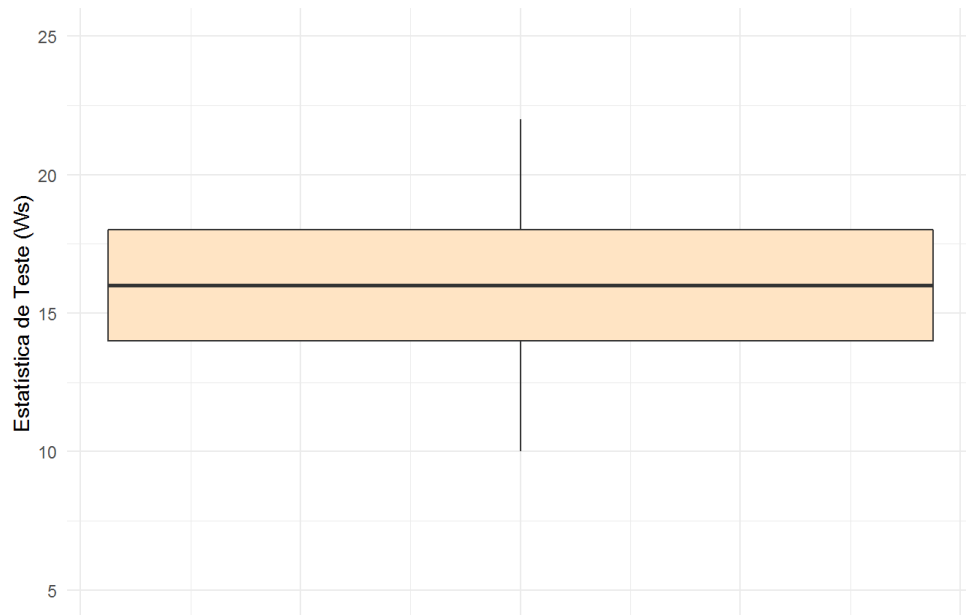
O Teste de Mann-Whitney/ Wilcoxon para Duas Amostras é um teste não paramétrico que compara as funções de distribuição de duas populações. Sua hipótese nula geralmente consiste na igualdade das distribuições, e sua hipótese alternativa na diferença das mesmas. Existem duas possibilidades de estatísticas a serem usadas para o teste. A primeira é a estatística W_s que é dada pela soma dos postos da segunda amostra. A segunda é a estatística U , que resulta da primeira: $U = W_s - m(m+1)/2$, sendo m o tamanho da segunda amostra.

Neste trabalho serão analisadas as distribuições dessas duas estatísticas para os casos $n=3, m=4$ e $n=4, m=3$. Para isso, serão gerados mil pares de amostras para o primeiro caso e mais mil pares de amostras para o segundo caso. Como o intuito é investigar as distribuições quando a hipótese nula é verdadeira, basta que as amostras geradas sigam uma mesma distribuição de probabilidade. Será utilizada, então, a distribuição uniforme com parâmetros 0 e 1, mas poderia ser escolhida qualquer outra distribuição sem alterar os resultados.

2 Estatística Ws

Distribuição da Estatística do Teste de Mann-Whitney/ Wilcoxon

Amostras: $n=3$ e $m=4$



Distribuição da Estatística do Teste de Mann-Whitney/ Wilcoxon

Amostras: $n=4$ e $m=3$

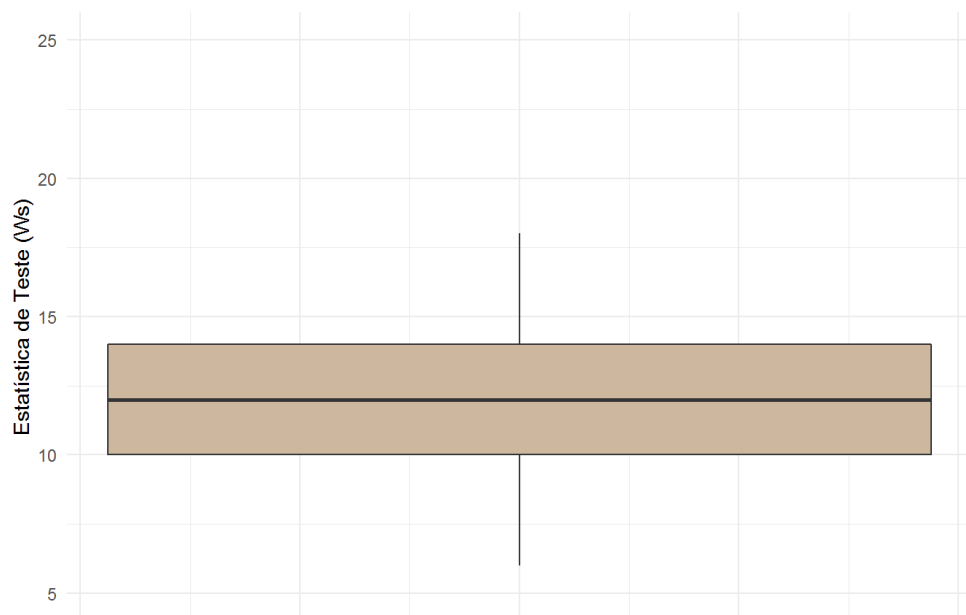


Tabela 1- Quantis Inferiores e Superiores da Estatística W_s para $n=3$, $m=4$ e $n=4$, $m=3$, Software R, 2020.

Quantis/ Amostras	$n=3$ e $m=4$	$n=4$ e $m=3$
0.1%	10	6
0.5%	10	6
1%	10	6
2.5%	10	6
5%	11	7
10%	12	8
90%	20	16
95%	21	17
97.5%	22	18
99%	22	18
99.5%	22	18
99.9%	22	18

Fonte: função `sumarize`, pacote `dplyr`, Software R.

Pelos gráficos acima, é possível perceber que as distribuições de W_s são idênticas em todos os aspectos menos um: a distribuição se desloca para baixo no eixo das ordenadas quando são invertidos os tamanhos da amostra, passando de $n=3$, $m=4$ para $n=4$, $m=3$. Isso ocorre porque a estatística W_s é justamente a soma dos postos da segunda amostra. Logo, ao diminuir o tamanho dessa amostra, são somados menos elementos para se obter a estatística, e, portanto, seu valor diminui. Também se nota, tanto pelos gráficos como pelos quantis tabelados, que a distribuição dessa estatística é simétrica.

3 Estatística U

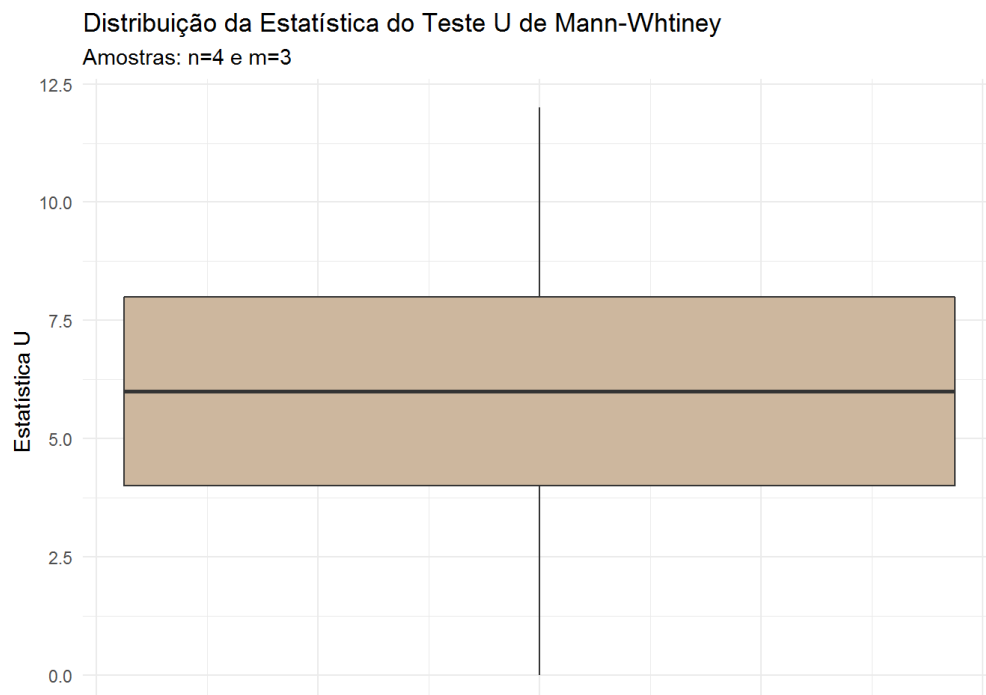
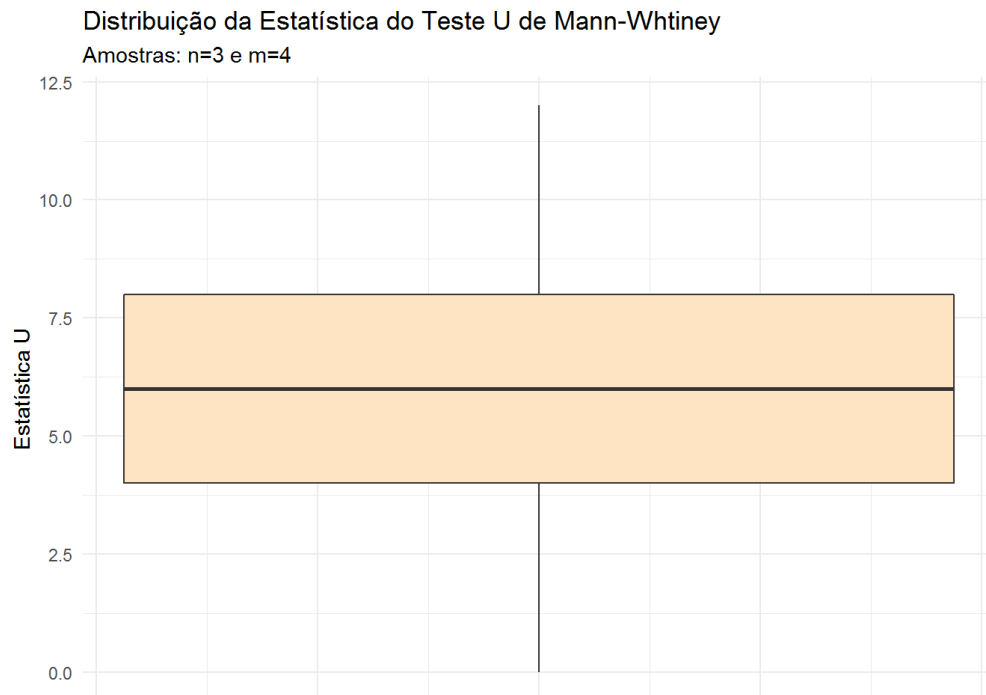


Tabela 2- Quantis Inferiores e Superiores da Estatística U para $n=3$, $m=4$ e $n=4$, $m=3$, Software R, 2020.

Quantis/ Amostras	$n=3$ e $m=4$	$n=4$ e $m=3$
0.1%	0	0
0.5%	0	0
1%	0	0
2.5%	0	0
5%	1	1
10%	2	2
90%	10	10
95%	11	11
97.5%	12	12
99%	12	12
99.5%	12	12
99.9%	12	12

Fonte: função `sumarize`, pacote `dplyr`, Software R.

Para a estatística U, as distribuições são realmente idênticas ao se inverter o tamanho das amostras. Isso ocorre porque, para se obter U, se parte de W_s e é retirado um valor fixo proporcional ao tamanho da segunda amostra. Logo, esse valor serve como um fator corretivo para impedir que o tamanho das amostras influencie na distribuição da estatística de teste. Outra consequência do uso da estatística U em lugar da W_s é que o valor mínimo dessa nova estatística é igual a zero, como fica evidente tanto pelo gráfico como pela tabela da distribuição de probabilidades. Apesar de todas essas mudanças, a distribuição de U parece ter as mesmas características da de W_s , inclusive a de simetria.

4 Considerações Finais

Em suma, tanto a estatística W_s como a estatística U funcionam para testar igualdade das funções de distribuição de duas populações. Porém, a segunda tem vantagem no sentido de manter sua distribuição idêntica quando os tamanhos das amostras são invertidos e de ter valor mínimo em zero.