

Testes de Hipóteses Não Paramétricos:

Teste à igualdade de duas distribuições

- Vamos estudar os testes não paramétricos que habitualmente são usados como alternativa aos testes paramétricos da diferença de médias.
- A vantagem destes testes não paramétricos deve-se ao facto de não ser necessário impor qual a forma da distribuição de cada população (nos testes paramétricos foi sempre imposto que a distribuição era Normal ou pelo menos aproximadamente Normal). Aqui apenas interessa testar se a distribuição pode ser considerada a mesma.

Princípios Básicos na Realização dos Testes à igualdade de duas distribuições

- 1 São definidas duas **hipóteses**:
 - ▶ **Hipótese Nula** = H_0 - é a hipótese que indica que as duas amostras são provenientes de populações com a mesma distribuição.
 - ▶ **Hipótese Alternativa** = H_1 - é a hipótese que se contrapõe à hipótese nula, ou seja, que indica que o que foi colocado na hipótese nula não se verifica (por ser diferente, maior ou menor).
- 2 É definida uma **Estatística Teste**, que é a base da realização do teste e consiste analisar posições.
- 3 São construídas duas regiões:
 - ▶ **Região de Aceitação** = RA - conjunto de valores para os quais H_0 é admissível.
 - ▶ **Região de Rejeição ou Região Crítica** = RC - conjunto de valores para os quais H_0 não é admissível.
- 4 A **regra de decisão** define as condições de rejeição ou não rejeição da hipótese nula:
 - ▶ Se o Valor Observado da Estatística de Teste sob a hipótese H_0 pertencer à Região de Aceitação, então Não se Rejeita H_0
 - ▶ Se o Valor Observado da Estatística de Teste sob a hipótese H_0 pertencer à Região Crítica, então Rejeita-se H_0
- 5 **Erros de decisão** - um teste de hipóteses nem sempre conduz a decisões corretas, a análise de uma amostra pode falsear as conclusões quanto à população. Como já vimos, um dos erros é o chamado **Erro de 1ª espécie** ou **Nível de significância do teste**:
$$\alpha = P[\text{rejeitar } H_0 \mid H_0 \text{ verdadeira}]$$
para minimizar este erro fixa-se o seu valor.
- 6 As regiões de aceitação e de rejeição (RA e RC) são definidas à custa do valor fixado para o nível de significância (α).

Na prática, em vez de calcular a região crítica (RC) e a região de aceitação (RA), é usual calcular-se o **Valor-p** (ou **p-value**).

Valor-p (ou p-value)

É a probabilidade associada ao valor da estatística de teste, considerando H_0 verdadeira.

- Se o valor-p for pequeno significa que, no caso de H_0 ser verdadeira, estamos perante um evento muito raro, pouco provável de ocorrer, então deve optar-se por rejeitar H_0 .

Portanto, o valor-p também permite tomar decisões:

- ▶ se valor-p $\leq \alpha$, então rejeita-se H_0
- ▶ se valor-p $> \alpha$, então não se rejeita H_0

Teste à igualdade de duas distribuições

- Suponha que tem duas amostras e pretende verificar se podem ser consideradas provenientes da mesma população, ou seja, pretende testar:

H_0 — As duas amostras são provenientes de populações com a mesma distribuição
contra

H_1 — As duas amostras são provenientes de populações com distribuição distinta

- Os testes não paramétricos que habitualmente são usados como alternativa aos testes paramétricos da diferença de médias são:
 - ▶ Teste de Wilcoxon - para amostras emparelhadas
 - ▶ Teste de Mann-Whitney - para amostras independentes

Teste de Wilcoxon

Objetivo

Testar se duas amostras aleatórias emparelhadas podem ser consideradas provenientes de populações com a mesma distribuição, para tal vamos testar se as duas amostras aleatórias emparelhadas são originárias de populações com igual **mediana**.

- A importância do teste de Wilcoxon advém do facto de ser geralmente considerado como alternativa não paramétrica ao teste t para a diferença de médias (teste de hipóteses paramétrico) quando são consideradas **amostras emparelhadas**.
- Este é um teste à igualdade de distribuições para duas amostras emparelhadas e baseia-se na posição dos valores observados da variável em estudo, incorporando a amplitude das diferenças existentes entre as duas variáveis.
- Tal como no caso dos testes paramétricos, para construir a estatística de teste é necessário passar para a amostra das diferenças:

$$D_i = Y_i - X_i, \quad i = 1, \dots, N$$

onde X_i e Y_i representam os elementos das amostras emparelhadas.



H_0 — As duas amostras emparelhadas são provenientes de populações com a mesma distribuição
vs

Seja M_D a mediana de $D \equiv Y - X$ onde X e Y representam as populações onde

<u>Teste bilateral</u>	<u>Teste unilateral direito</u>	<u>Teste unilateral esquerdo</u>
$H_0 : M_D = 0$	$H_0 : M_D = 0$	$H_0 : M_D = 0$
<i>vs</i>	<i>vs</i>	<i>vs</i>
$H_1 : M_D \neq 0$	$H_1 : M_D > 0$	$H_1 : M_D < 0$

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ↺ 🔍 ↻

Considerando que H_0 é verdadeira, o valor-p indica a probabilidade do valor observado da estatística de teste ocorrer e, tal como aconteceu na definição da região crítica, independentemente do tipo de teste (bilateral ou unilateral), pode-se considerar

- Se $\text{valor-p} > \alpha$, então, ao nível de significância α , a hipótese H_0 não é rejeitada, isto é, com base na amostra há evidências estatísticas que os dados são provenientes de populações com a mesma distribuição.
- Se $\text{valor-p} \leq \alpha$, então, ao nível de significância α , a hipótese H_0 é rejeitada, isto é, com base na amostra há evidências estatísticas que os dados não são provenientes de populações com a mesma distribuição.

Mediu-se a capacidade torácica de 8 indivíduos selecionados aleatoriamente. Esse grupo de indivíduos submeteu-se voluntariamente, durante um mês, a um treino especial que tinha por objetivo o aumento da capacidade torácica. No final do mês de treino, foi medida, de novo, a capacidade torácica. Os resultados de ambas as medições encontram-se na tabela seguinte:

Com base nos dados apresentados, poder-se-á concluir, com um nível de significância de 5% que o treino é eficaz?

Sendo amostras emparelhadas, poderíamos pensar num teste paramétrico para diferença de médias com dados emparelhados (ver exemplo 4, página 46 slides Cap. 5), onde a nossa amostra seria as diferenças *Depois – Antes*, ou seja, passaríamos a ter apenas uma amostra constituída por essas diferenças. Porém, não se sabe a distribuição da população, logo temos uma população qualquer:

População Qualquer σ conhecido $n \geq 30$	$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$	$\left] \bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}; \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right[$	<code>z.test()</code> <code>library(BSDA)</code>
População Qualquer σ desconhecido $n \geq 30$	$Z = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim N(0, 1)$	$\left] \bar{x} - z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}; \bar{x} + z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right[$	<code>z.test()</code> <code>library(BSDA)</code>

Por outro lado, $n = 8$, inferior a 30. Desta forma, não se pode usar o teste paramétrico, teremos que recorrer ao **teste não paramétrico de Wilcoxon**.

Observação: Poderíamos ir verificar a normalidade dos dados, através dos Testes de Ajustamento. Vamos assumir que não são normais.

Hipótese a ser testada

- X – medida da capacidade torácica antes do treino
- Y – medida da capacidade torácica depois do treino
- amostras aleatórias emparelhadas
- $D = Y - X$

$$H_1: M_D > 0$$

(o treino é eficaz)

Observação: ser eficaz significa que houve aumento da capacidade torácica, ou seja, a diferença depois-antes será superior a zero.

$$H_0: M_D = 0 \quad vs \quad H_1: M_D > 0$$

Dados

- teste unilateral direito
- nível de significância $= \alpha = 0.05$

No R:

```
# Exemplo 1

# X - medida da capacidade torácica antes do treino
# Y - medida da capacidade torácica depois do treino
# amostras emparelhadas -> teste de Wilcoxon
# D=Y-X

# H0: as duas amostras emparelhadas são provenientes de populações com a mesma distribuição
# a mediana de D = 0
# contra
# H1: as duas amostras emparelhadas são provenientes de populações com distribuição distinta
# a mediana de D > 0

# Exemplo 1
# amostras
amostra1.antes <- c(3.5,3.6,4.1,2.9,3.4,4.2,3.9,4.1)
amostra1.depois <- c(3.4,3.9,4.5,3.1,3.9,4.4,3.8,4.1)

# Teste de Wilcoxon
wilcox.test(x=amostra1.depois, y=amostra1.antes, alternative = "greater", mu=0, paired=TRUE)

#ou
exemplo.1 <- wilcox.test(x=amostra1.depois, y=amostra1.antes, alternative = "greater", mu=0, paired=TRUE)
exemplo.1$statistic # T+
exemplo.1$p.value # valor-p
exemplo.1$null.value # H0: MD=0
exemplo.1$alternative # H1: MD>0
```

Valor que está em H_0

$$p - value = 0.03744$$

DECISÃO: Como $p - value$ é 0.03744 e é $\leq \alpha$, pois $\alpha = 0.05$, então rejeita-se H_0 .

Conclusão: Com base nas amostras e ao nível de significância de 5%, conclui-se que o treino é eficaz, na medida em que contribui para o aumento da capacidade torácica.

Exemplo 2

Num estudo sobre nutrição pretende-se avaliar uma determinada dieta com base na perda de peso. Num grupo de 10 pessoas analisou-se o peso antes e depois do plano de dieta. Os pesos (em kg) foram os seguintes:

Individuo	1	2	3	4	5	6	7	8	9	10
Antes da dieta	82.7	73.2	84.1	84.1	81.6	78.9	85.6	80.2	84.5	73.8
Depois da dieta	74.5	73.2	79.1	85.6	81.6	79.6	81.5	80.2	86.9	73.8

Com base nos dados apresentados, poder-se-á concluir, com um nível de significância de 5% que dieta é eficaz?

Hipótese a ser testada

- X – peso, em kg, antes da dieta
- Y – peso, em kg, depois da dieta
- amostras aleatórias emparelhadas
- $D = Y - X$

$$H_1: M_D < 0$$

(a dieta é eficaz)

$$H_0: M_D = 0 \quad vs \quad H_1: M_D < 0$$

Dados

- teste unilateral esquerdo
- nível de significância = $\alpha = 0.05$

No R:

```
# Exemplo 2

# X - antes da dieta
# Y - depois da dieta
# amostras emparelhadas -> teste de Wilcoxon
# D=Y-X

# H0: as duas amostras emparelhadas são provenientes de populações com a mesma distribuição
# a mediana de D = 0
# contra
# H1: as duas amostras emparelhadas são provenientes de populações com distribuição distinta
# a mediana de D < 0

# amostras
amostra2.antes <- c(82.7,73.2,84.1,84.1,81.6,78.9,85.6,80.2,84.5,73.8)
amostra2.depois <- c(74.5,73.2,79.1,85.6,81.6,79.6,81.5,80.2,86.9,73.8)

# Teste de Wilcoxon
wilcox.test(x=amostra2.depois, y=amostra2.antes, alternative = "less", mu=0, paired=TRUE)

#ou
exemplo.2 <- wilcox.test(x=amostra2.depois, y=amostra2.antes, alternative = "less", mu=0, paired=TRUE)
exemplo.2$statistic # T+
exemplo.2$p.value # valor-p
exemplo.2$null.value # H0: MD=0
exemplo.2$alternative # H1: MD<0
```

Valor que está em H_0

$$p - value = 0.20084$$

DECISÃO: Como $p - value$ é 0.20084, não é $\leq \alpha$, pois $\alpha = 0.05$, então não se rejeita H_0 .

Conclusão: Com base nas amostras e ao nível de significância de 5%, conclui-se que a dieta não parece ser eficaz, na medida em que não parece contribuir para a perda de peso.

Exemplo 3

Com o objetivo de avaliar uma dada disciplina que está dividida em teórica e prática, um professor pediu a um grupo de alunos que realizassem dois testes, um dos testes apenas com a componente teórica e outro teste só com a componente prática. Os resultados encontram-se na tabela seguinte:

Aluno	1	2	3	4	5	6	7	8	9	10
teste teórico	10	12	13	14	11	12.4	15	9.8	12.9	12.9
teste prático	9.8	11.6	12	14	11	13	16	12	13	13.4

Será possível concluir, para um nível de significância de 5%, que não há diferenças nos resultados dos testes?

Hipótese a ser testada

- X – nota no teste teórico
- Y – nota no teste prático
- amostras aleatórias emparelhadas
- $D = Y - X$

$$H_0: M_D = 0$$

(não há diferenças nos resultados do teste)

$$H_0: M_D = 0 \quad vs \quad H_1: M_D \neq 0$$

Dados

- teste bilateral
- nível de significância = $\alpha = 0.05$

```
# Exemplo 3

# T - teste teórico
# P - teste prático
# amostras emparelhadas -> teste de Wilcoxon
# D=P-T

# H0: as duas amostras emparelhadas são provenientes de populações com a mesma distribuição
# a mediana de D = 0
# contra
# H1: as duas amostras emparelhadas são provenientes de populações com distribuição distinta
# a mediana de D diferente 0

# amostras
amostra3.testeT <- c(10,12,13,14,11,12.4,15,9.8,12.9,12.9)
amostra3.testeP <- c(9.8,11.6,12,14,11,13,16,12,13,13.4)

# Teste de Wilcoxon
wilcox.test(x=amostra3.testeP, y=amostra3.testeT, alternative = "two.sided", mu=0, paired=TRUE)

#ou
exemplo.3 <- wilcox.test(x=amostra3.P, y=amostra3.T, alternative = "two.sided", mu=0, paired=TRUE)
exemplo.3$statistic # T+
exemplo.3$p.value # valor-p
exemplo.3$null.value # H0: MD=0
exemplo.3$alternative # H1: MD<>0
```

$$p - value = 0.40024$$

DECISÃO: Como $p - value$ é 0.40024, não é $\leq \alpha$, pois $\alpha = 0.05$, então não se rejeita H_0 .

Conclusão: Com base nas amostras e ao nível de significância de 5%, conclui-se que não há diferenças nos resultados dos testes.

Teste de Mann-Whitney

- O Teste de Mann-Whitney (também chamado Teste de Mann-Whitney-Wilcoxon ou Teste de Wilcoxon-Mann-Whitney), é um teste não paramétrico aplicado para duas amostras independentes.
- A importância do teste de Mann-Whitney advém do facto de ser geralmente considerado como alternativa não paramétrica ao teste t para a diferença de médias (teste de hipóteses paramétrico para $\mu_1 - \mu_2$) quando são consideradas amostras independentes.
- Este é um teste à igualdade de distribuições para duas amostras independentes e baseia-se na posição dos valores observados da variável em estudo.
- A posição de uma observação é o número de ordem que lhe corresponde considerando a ordenação indistinta das duas amostras independentes envolvidas.

Objetivo

Testar se duas amostras aleatórias independentes podem ser consideradas provenientes de populações com a mesma distribuição, para tal vamos testar se as duas amostras aleatórias independentes são originárias de populações com igual mediana.

Pressupostos do Teste

- As duas amostras de dimensões n e m foram retiradas de forma independente e aleatória das respectivas populações.
- A variável em análise é uma variável aleatória contínua.
- Se do teste resultar que as populações diferem, isso acontece somente em relação às respectivas medianas.

Observação: Para populações simétricas as conclusões que se tiram para as medianas são igualmente válidas para as médias.

Formulação das Hipóteses a Testar:

H_0 — As duas amostras independentes são provenientes de populações com a mesma distribuição
 vs

H_1 — As duas amostras independentes são provenientes de populações com distribuição distinta

Seja M_X a mediana da população X e M_Y a mediana da população Y , então é possível testar as seguintes hipóteses:

<u>Teste bilateral</u>	<u>Teste unilateral direito</u>	<u>Teste unilateral esquerdo</u>
$H_0 : M_X = M_Y$ vs	$H_0 : M_X = M_Y$ vs	$H_0 : M_X = M_Y$ vs
$H_1 : M_X \neq M_Y$	$H_1 : M_X > M_Y$ □	$H_1 : M_X < M_Y$

Cálculo do valor-p

O cálculo do valor-p depende do tipo de teste considerado:

- Teste bilateral: $\text{valor-p} = 2 \times \text{mínimo} \{P(U \leq U_{obs}), P(U \geq U_{obs})\}$
- Teste unilateral direito: $\text{valor-p} = P(U \geq U_{obs})$
- Teste unilateral esquerdo: $\text{valor-p} = P(U \leq U_{obs})$

Regra de Decisão com base no valor-p

- Se $\text{valor-p} > \alpha$, então, ao nível de significância α , a hipótese H_0 não é rejeitada, isto é, com base na amostra há evidências estatísticas que os dados são provenientes de populações com a mesma distribuição.
- Se $\text{valor-p} \leq \alpha$, então, ao nível de significância α , a hipótese H_0 é rejeitada, isto é, com base na amostra há evidências estatísticas que os dados não são provenientes de populações com a mesma distribuição.

O valor-p pode ser visto como o menor valor de α (nível de significância) para o qual os dados observados indicam que H_0 deve ser rejeitada.

Exemplo 4

Um investigador pretende conhecer o efeito da inalação prolongada de óxido de cádmio. Para o efeito sujeita um grupo de 15 animais de laboratório às inalações e confronta os resultados dos níveis de hemoglobina com os do grupo de controlo (que não foram sujeitos às inalações) constituído por 10 animais. Os resultados apresentam-se na tabela seguinte:

X	14.4	14.2	13.8	16.5	14.1	16.6	15.9	15.6	14.1	15.3
	15.7	16.7	13.7	15.3	14.0					
Y	17.4	16.2	17.1	17.5	15.0	16.0	16.9	15.0	16.3	16.8

Será possível concluir, para um nível de significância de 5%, que a inalação prolongada de óxido de cádmio reduz os níveis de hemoglobina?

logo poderíamos pensar num teste para $\mu_1 - \mu_2$ com população qualquer. Recorrendo ao formulário vem que:

Condições			R
Populações Quaisquer σ_1 e σ_2 conhecidos Amostras Independentes $n_1 \geq 30$ e $n_2 \geq 30$	D. A.	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$	z.test() library(BSDA)
	I. C.	$\left[(\bar{X}_1 - \bar{X}_2) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}; (\bar{X}_1 - \bar{X}_2) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$	
Populações Quaisquer σ_1 e σ_2 desconhecidos Amostras Independentes $n_1 \geq 30$ e $n_2 \geq 30$	D. A.	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0, 1)$	z.test() library(BSDA)
	I. C.	$\left[(\bar{X}_1 - \bar{X}_2) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}; (\bar{X}_1 - \bar{X}_2) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right]$	

Porém, $n_X = 15$ e $n_Y = 10$, ambos inferiores a 30...

Logo devemos usar teste não paramétrico (supondo que os dados não são normais)

Hipótese a ser testada

$H_1: M_X - M_Y < 0$
(a inalação prolongada de óxido de cádmio reduz os níveis de hemoglobina)

$$H_0: M_X = M_Y \quad vs \quad H_1: M_X < M_Y$$

- M_X – mediana dos valores da hemoglobina dos animais sujeitos à inalação de óxido de cádmio
- M_Y – mediana dos valores da hemoglobina dos animais do grupo de controlo

Dados

- amostras aleatórias independentes
- da população X foi retirada uma amostra de dimensão $n = 15$
- da população Y foi retirada uma amostra de dimensão $m = 10$
- Teste unilateral esquerdo
- nível de significância $= \alpha = 0.05$

No R:

```
# Exemplo 4

# I - valores da hemoglobina dos animais sujeitos à inalação de óxido de cádmio
# C - valores da hemoglobina dos animais do grupo de controlo
# amostras independentes -> Mann-Whitney

# H0: as duas amostras independentes são provenientes de populações com a mesma distribuição
# a mediana de X é igual à mediana de Y
# contra
# H1: as duas amostras independentes são provenientes de populações com distribuição distinta
# a mediana de X é inferior à mediana de Y

# amostras
amostra4.i <- c(14.4,14.2,13.8,16.5,14.1,16.6,15.9,15.6,14.1,15.3,15.7,16.7,13.7,15.3,14.0)
amostra4.c <- c(17.4,16.2,17.1,17.5,15.0,16.0,16.9,15.0,16.3,16.8)

# Teste de Mann-Whitney
wilcox.test(x=amostra4.i, y=amostra4.c, alternative = "less", mu=0, paired=FALSE)

#ou
exemplo.4 <- wilcox.test(x=amostra4.i, y=amostra4.c, alternative = "less", mu=0, paired=FALSE)
exemplo.4$statistic # Uobs
exemplo.4$p.value # valor-p
exemplo.4$null.value # H0: MX-MY=0
exemplo.4$alternative # H1: MX-MY<0
```

$$p - value = 0.0030$$

DECISÃO: Como $p - value$ é 0.0030, $\leq \alpha$, pois $\alpha = 0.05$, então rejeita-se H_0 .

Conclusão: Com base nas amostras e ao nível de significância de 5%, conclui-se que a inalação prolongada de óxido de cádmio reduz os níveis de hemoglobina.

Exemplo 5

Na tabela seguinte indicam-se os valores dos Triglicéridos, em g/L, em 10 doentes com enfarte do miocárdio e em 8 indivíduos escolhidos para controlo (que não sofreram enfarte do miocárdio):

Doentes	1.62	0.51	1.29	0.71	0.52	2.10	0.88	0.99	0.51	1.59
Controlo	0.92	1.29	2.81	0.82	4.48	0.71	1.10	0.41		

Será possível concluir, para um nível de significância de 5%, que os indivíduos que sofreram enfarte do miocárdio possuem valores dos Triglicéridos superiores?

Hipótese a ser testada

$$H_0 : M_X = M_Y \quad vs \quad H_1 : M_X > M_Y$$

- M_X – mediana dos valores dos Triglicéridos dos doentes com enfarte do miocárdio
- M_Y – mediana dos valores dos indivíduos do grupo de controlo

$H_1: M_X - M_Y > 0$
(indivíduos que sofreram enfarte possuem valores dos Triglicéridos superiores)

Dados

- amostras aleatórias independentes
- da população X foi retirada uma amostra de dimensão $n = 10$
- da população Y foi retirada uma amostra de dimensão $m = 8$
- Teste unilateral direito
- nível de significância $= \alpha = 0.05$

No R:

EXEMPLO 5

```
# X - valores dos triglicéridos dos doentes com enfarte do miocárdio
# Y - valores dos triglicéridos do grupo de controlo
# amostras independentes -> Mann-Whitney

# H0: as duas amostras independentes são provenientes de populações com a mesma distribuição
# a mediana de X é igual à mediana de Y
# contra
# H1: as duas amostras independentes são provenientes de populações com distribuição distinta
# a mediana de X é superior à mediana de Y

# amostras
amostra5.x <- c(1.62,0.51,1.29,0.71,0.52,2.10,0.88,0.99,0.51,1.59)
amostra5.y <- c(0.92,1.29,2.81,0.82,4.48,0.71,1.10,0.41)

# Teste de Mann-Whitney
wilcox.test(x=amostra5.x, y=amostra5.y, alternative = "greater", mu=0, paired=FALSE)

#ou
exemplo.5 <- wilcox.test(x=amostra5.x, y=amostra5.y, alternative = "greater", mu=0, paired=FALSE)
exemplo.5$statistic # Uobs
exemplo.5$p.value # valor-p
exemplo.5$null.value # H0: MX-MY=0
exemplo.5$alternative # H1: MX-MY>0
```

$$p - value = 0.68774$$

DECISÃO: Como $p - value$ é 0.68774, não é $\leq \alpha$, pois $\alpha = 0.05$, então não se rejeita H_0 .

Conclusão: Com base nas amostras e ao nível de significância de 5%, conclui-se que os indivíduos que sofreram enfarte do miocárdio não possuem valores dos Triglicéridos superiores.

Exemplo 6

Considere as seguintes amostras relativas à precipitação anual nos distritos de Beja e Évora:

Beja	607.4	809.1	488.8	481.1	592.8	345.4	620.0	407.7	513.3	527.4
Évora	694.5	629.6	676.9	430.3	727.2					

Será possível concluir, para um nível de significância de 5%, que não há diferenças na precipitação anual nestes dois distritos?

Hipótese a ser testada

$H_0: M_X - M_Y = 0$
(não há diferenças na precipitação anual nestes dois distritos)

$$H_0 : M_X = M_Y \quad vs \quad H_1 : M_X \neq M_Y$$

- M_X — mediana da precipitação anual em Beja
- M_Y — mediana da precipitação anual em Évora

Dados

- amostras aleatórias independentes
- da população X foi retirada uma amostra de dimensão $n = 10$
- da população Y foi retirada uma amostra de dimensão $m = 5$
- Teste bilateral
- nível de significância $= \alpha = 0.05$

No R:

```
# EXEMPLO 6

# X - precipitação anual no distrito de Beja
# Y - precipitação anual no distrito de Évora
# amostras independentes -> Mann-Whitney

# H0: as duas amostras independentes são provenientes de populações com a mesma distribuição
# a mediana de X é igual à mediana de Y
# contra
# H1: as duas amostras independentes são provenientes de populações com distribuição distinta
# a mediana de X é diferente da mediana de Y

# amostras
amostra6.x <- c(607.4, 809.1, 488.8, 481.1, 592.8, 345.4, 620.0, 407.7, 513.3, 527.4)
amostra6.y <- c(694.5, 629.6, 676.9, 430.3, 727.2)

# Teste de Mann-Whitney
wilcox.test(x=amostra6.x, y=amostra6.y, alternative = "two.sided", mu=0, paired=FALSE)

#ou
exemplo.6 <- wilcox.test(x=amostra6.x, y=amostra6.y, alternative = "two.sided", mu=0, paired=FALSE)
exemplo.6$statistic # Uobs
exemplo.6$p.value # valor-p
exemplo.6$null.value # H0: MX-MY=0
exemplo.6$alternative # H1: MX-MY<>0
```

$$p - value = 0.1292$$

DECISÃO: Como $p - value$ é 0.1292, não é $\leq \alpha$, pois $\alpha = 0.05$, então não se rejeita H_0 .

Conclusão: Com base nas amostras e ao nível de significância de 5%, conclui-se que não há diferenças na precipitação anual nestes dois distritos.