TESTES NÃO PARAMÉTRICOS

Estatística – Aulas 19, 20 e 21

Teste não paramétrico

- Um teste de hipótese que não requer condições específicas sobre o formato das populações ou o valor de nenhum parâmetro de população
- Mais fácil de fazer do que testes paramétricos
- Menos eficiente do que testes paramétricos (é necessário uma evidência maior para rejeitar a hipótese nula)

Testes Não Paramétricos

- Teste dos Sinais
- Teste de Wilcoxon
- Teste de Kruskal-Wallis
- Correlação de postos
- Testes de Corrida

TESTES NÃO PARAMÉTRICOS

Teste dos sinais

Objetivos

- Usar teste dos sinais para testar uma mediana de população.
- Usar teste dos sinais de amostra emparelhada para testar a diferença entre duas medianas de população (amostras dependentes)

Teste dos sinais para uma mediana de população

- Teste não paramétrico que pode ser usado para testar uma mediana de população contra um valor hipotético k

  - Teste unicaudal à esquerda:
    H_0: mediana $\geq k$ e H_a: mediana $< k$

  - Teste unicaudal à direita:
    H_0: mediana $\leq k$ e H_a: mediana $> k$

  - Teste bicaudal:
    H_0: mediana $= k$ e H_a: mediana $\neq k$

- Para usar o teste de sinais, primeiro compare cada entrada na amostra com a mediana hipotética *k*:
  - se a entrada está abaixo da mediana, designe um sinal –
  - se a entrada está acima da mediana, designe um sinal +
  - se a entrada for igual à mediana, designe um valor 0 (zero)
- Compare o número dos sinais – e + (ignore os zeros)

Estatística do teste para o teste de sinais

- Quando $n \leq 25$, a estatística do teste x é o menor número dos sinais + ou –
- Quando $n > 25$, a estatística do teste é:

$$z = \frac{(x + 0.5) - 0.5n}{\frac{\sqrt{n}}{2}}$$

Onde:

x é o menor número dos sinais + ou –

n é o tamanho da amostra (número total dos sinais + e –)

Desempenhando um teste de sinais para uma mediana de população

Em palavras	*Em símbolos*
1. Faça a afirmação. Identifique as hipóteses nula e alternativa.	Afirme H_0 e H_a.
2. Especifique o nível de significância.	Identifique α.
3. Determine o tamanho da amostra n designando sinais + e – para a informação de amostra.	n = número total de sinais + e –
4. Determine o valor crítico.	Se $n \leq 25$, use a Tabela valores críticos para o teste de sinais. Se $n > 25$, use a Tabela z.

Em palavras	*Em símbolos*
5. Calcule a estatística do teste.	Se $n \leq 25$, use x Se $n > 25$, use $$z = \frac{(x+0.5)-0.5n}{\frac{\sqrt{n}}{2}}$$
6. Tome a decisão de rejeitar ou não rejeitar a hipótese nula 7. Interprete a decisão no contexto da afirmação original.	Se a estatística de teste é menor ou igual ao valor crítico, rejeite H_0. Do contrário, não rejeite H_0.

Exemplo: usando o teste de sinais

Um gerente de banco afirma que o número mediano de clientes por dia não é maior que 750. Um atendente duvida da precisão dessa afirmação. O número de clientes do banco para 16 dias selecionados aleatoriamente está listado a seguir. Em $\alpha = 0{,}05$, o atendente pode rejeitar a afirmação do gerente do banco?

775	765	801	742
754	753	739	751
745	750	777	769
756	760	782	789

Solução: usando o teste de sinais

- H_0: mediana ≤ 750
- H_a: mediana > 750
- Compare cada valor com a mediana hipotética 750:

775	765	801	742
754	753	739	751
745	750	777	769
756	760	782	789

+	+	+	–
+	+	–	+
–	0	+	+
+	+	+	+

- Há 3 sinais – e 12 sinais +
  $n = 12 + 3 = 15$

- **α = 0,05** Use a Tabela $(n \leq 25)$

	Unicaudal, α = 0,005	α = 0,01	α = 0,025	α = 0,05
n	Bicaudal, α = 0,01	α = 0,02	α = 0,05	α = 0,10
8	0	0	0	1
9	0	0	1	1
10	0	0	1	1
11	0	1	1	2
12	1	1	2	2
13	1	1	2	3
14	1	2	3	3
15	2	2	3	3
16	2	2	3	4
17	2	3	4	4
18	3	3	4	5
19	3	4	4	5
20	3	4	5	5
21	4	4	5	6
22	4	5	5	6
23	4	5	6	7
24	5	5	6	7
25	5	6	6	7

O valor crítico é 3

- **Estatística do teste:**

  $\boldsymbol{x} = \mathbf{3}$ ($n \leq 25$; utilize menores números para + or –)

- **Decisão:** **Rejeite H_0**

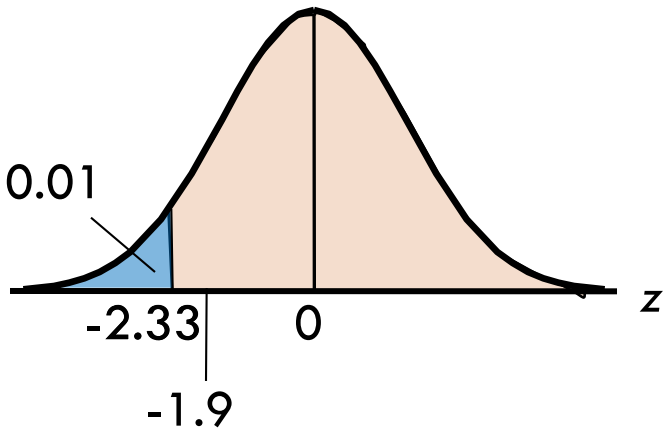
No nível de significância de 5%, o atendente pode rejeitar a afirmação do gerente de banco de que o número mediano de clientes por dia não é maior que 750.

Exemplo: usando o teste de sinais

Uma revendedora de carros afirma dar aos clientes uma oferta mediana de troca de pelo menos $ 6.000. Uma amostra aleatória de 103 transações revelou que a oferta de troca para 60 automóveis era menor que $ 6.000 e a oferta de troca para 40 automóveis era maior que $ 6.000. Em $\alpha = 0{,}01$, você pode rejeitar a afirmação da revendedora?

Solução: usando o teste de sinais

- H_0: mediana $\geq$ 6000
- H_a: mediana < 6000
- α = 0.01
- **Valor crítico:** $n > 25$



- **Estatística do teste:**

Há 60 sinais – e 40 sinais +

n = 60 + 40 = 100

x = 40

$$z = \frac{(40+0.5)-0.5(100)}{\sqrt{100}/2} = -1.9$$

- **Decisão:** **Não rejeitar a H_0**

No nível de significância 1%, você não pode rejeitar a afirmação da revendedora.

Exercício

Durante um anúncio do tempo, um meteorologista declara que a mediana da temperatura máxima para o mês de janeiro em San Diego é de 66° F. As temperaturas máximas para 18 dias em janeiro selecionadas aleatoriamente estão listadas a seguir. Em $\alpha=0{,}01$, você pode rejeitar a afirmação do metereologista?

78	74	72	72	70	70	72	78	74
71	72	72	72	74	77	79	75	73

Exercício

Uma instituição de serviços financeiros relata que a quantidade mediana de débito de cartão de crédito para as famílias é de no mínimo $ 2.200. Em uma amostra aleatória de 104 famílias que possuem débito, você vê que o débito para 60 famílias é menor que $ 2.200 e o débito para 44 famílias é maior que $ 2.200. Em α=0,02 você pode rejeitar a afirmação da instituição?

Teste de sinais de amostras emparelhadas

- Usado para testar a diferença entre duas medianas de população quando as populações não são normalmente distribuídas
- Para o teste de sinais de amostras emparelhadas ser usado, as seguintes condições devem ser atendidas:
  1. Uma amostra deve ser selecionada aleatoriamente de cada população.
  2. As amostras devem ser dependentes (emparelhadas)
- A diferença entre os dados é encontrada e o sinal da diferença é registrado.

Desempenhando um teste de sinais de amostra emparelhada

Em palavras	*Em símbolos*
1. Faça a afirmação. Identifique as hipóteses nula e alternativa.	Afirme H_0 e H_a.
2. Especifique o nível de significância.	Identifique α.
3. Determine o tamanho da amostra n encontrando a diferença para cada par de informação. Designe um sinal + para uma diferença positiva, um sinal – para uma diferença negativa, e um 0 para nenhuma diferença.	n = número total de sinais + e –

Em palavras	*Em símbolos*
4. Determine o valor crítico.	
5. Encontre a estatística de teste.	x = menor número de sinais + e –
6. Tome a decisão de rejeitar ou falhar em rejeitar a hipótese nula. 7. Interprete a decisão no contexto da afirmação original.	Se o teste estatístico é menor ou igual ao valor crítico, rejeite H_0. Caso contrário, não rejeitar H_0.

Exemplo: usando o teste de sinais de amostra emparelhada

Um psicólogo afirma que o número de infratores reincidentes diminuirá se infratores primários completarem um curso de reabilitação em especial. Você seleciona aleatoriamente 10 presídios e registra o número de infratores reincidentes durante um período de dois anos. Então, depois de infratores primários completarem o curso, você registra o número de infratores reincidentes em cada presídio por um outro período de dois anos. Os resultados são mostrados na tabela a seguir. Em $\alpha = 0{,}025$, você pode apoiar a afirmação do psicólogo?

Presídio	1	2	3	4	5	6	7	8	9	10
Antes	21	34	9	45	30	54	37	36	33	40
Depois	19	22	16	31	21	30	22	18	17	21

Solução:

- H_0: Número de infratores reincidentes não diminuirá.

- H_a: Número de infratores reincidentes diminuirá. (afirm.)

  - Determine o sinal da diferença entre o “antes” e o “depois”.

Solução: teste de sinais de amostra emparelhada

Presídio	1	2	3	4	5	6	7	8	9	10
Antes	21	34	9	45	30	54	37	36	33	40
Depois	19	22	16	31	21	30	22	18	17	21
Sinal	+	+	–	+	+	+	+	+	+	+

α = 0.025 (unicaudal)

- ***n*** **= 1+ 9 = 10**
- **Valor crítico:**

	Unicaudal, α = 0,005	α = 0,01	α = 0,025	α = 0,05
n	Bicaudal, α = 0,01	α = 0,02	α = 0,05	α = 0,10
8	0	0	0	1
9	0	0	1	1
10	0	0	1	1
11	0	1	1	2
12	1	1	2	2
13	1	1	2	3

O valor crítico é 1

Sinal	+	+	–	+	+	+	+	+	+	+

- Estatística do teste:

  $x = 1$ (o menor número de sinais + ou –)

- Decisão: Rejeite H_0.

No nível de significância 2,5%, você pode apoiar a afirmação do psicólogo de que o número de infratores reincidentes diminuirá.

Exercício

Um médico pesquisador afirma que uma nova vacina diminuirá o número de resfriados em adultos. Você seleciona aleatoriamente 14 adultos e registra o número de resfriados que cada um teve em um período de um ano. Após dar a vacina a cada adulto, você registra novamente o número de resfriados que cada um teve em um período de um ano. Os resultados são mostrados na tabela. Em $\alpha=0,05$, você apoia a afirmação do pesquisador?

Adulto	Antes da vacina	Depois da vacina
1	3	2
2	4	1
3	2	0
4	1	1
5	3	1
6	6	3
7	4	3
8	5	2
9	2	2
10	0	2
11	2	3
12	5	4
13	3	3
14	3	2

Exercícios

Uma organização trabalhista estima que a mediana de ganhos por hora paga a trabalhadores do sexo masculino é $12,16. Em uma amostra aleatória de 41 trabalhadores, 16 receberam menos que $12,16 por hora, 23 receberam mais que $12,16 por hora e 2 receberam $ 12,16 por hora. Em alfa = 0,01, você pode rejeitar a afirmação da organização?

Uma organização afirma que a mediana de ganhos por hora paga a trabalhadores do sexo feminino é no máximo de $ 10,31. Em uma amostra aleatória de 23 trabalhadoras, 9 recebem menos que $ 10,31 por hora, 11 recebem mais que $ 10,31 por hora e 3 recebem $ 10,31 por hora. Em alfa=0,05, você pode rejeitar a afirmação da organização?

Resumo

- Usamos o teste de sinal para testar uma mediana de população.
- Usamos o teste dos sinais de amostra emparelhada para testar a diferença entre duas medianas de população (amostras dependentes)

TESTE DE WILCOXON

Testes Não Paramétricos

Objetivos

- Usar o teste de postos de sinais de Wilcoxon para determinar se duas amostras dependentes são selecionadas de populações que possuem a mesma distribuição.
- Usar o teste de soma de postos de Wilcoxon para determinar se duas amostras independentes são selecionadas de populações que possuem a mesma distribuição.

Teste de postos de sinais de Wilcoxon

- É um teste não paramétrico que pode ser usado para determinar se duas amostras *dependentes* foram selecionadas de populações que possuem a mesma distribuição.

- Ao contrário do teste de sinais, ele considera o tamanho das entradas de informação.

Desempenhando um teste de postos de sinais de Wilcoxon

Em palavras	*Em símbolos*
1. Escreva a afirmação. Identifique as hipóteses nula e alternativa.	Afirme H_0 e H_a.
2. Especifique o nível de significância.	Identifique α.
3. Determine o tamanho da amostra n, que é o número de pares de informações para os quais a diferença não é 0.	Use a Tabela
4. Determine o valor crítico.	

Em palavras	*Em símbolos*
5. Calcule a estatística de teste w_s. a. Complete a tabela usando os títulos usados à direita. b. Encontre a soma dos postos positivos e a soma dos postos negativos. c. Selecione o menor dos valores absolutos das somas.	Títulos: Amostra 1, Amostra 2, Diferença, Valor absoluto, Posto, e Posto com sinais. Posto com sinais leva o mesmo sinal da sua diferença correspondente.

Em palavras	*Em símbolos*
6. Tome a decisão de rejeitar ou não rejeitar a hipótese nula.	Se w_s é menor ou igual ao valor crítico, rejeite H_0. Do contrário, não rejeite H_0.
7. Interprete a decisão no contexto da afirmação original.	

Exemplo: teste de postos de sinais de Wilcoxon

Um psicólogo do esporte acredita que ouvir música afeta a duração das sessões de exercício de atletas. A duração do tempo (em minutos) de 10 sessões de exercício de atletas, enquanto ouviam música e enquanto não ouviam música, são mostrados na tabela. Em $\alpha = 0{,}05$, você pode apoiar a afirmação do psicólogo do esporte?

Com música	45	38	28	39	41	47	62	54	33	44
Sem música	38	40	33	36	42	41	54	47	28	35

Solução: teste de postos de sinais de Wilcoxon

- H_0: Não há diferença na duração das sessões de exercício de atletas.
- H_a: Há diferença na duração das sessões de exercício de atletas.
- $\alpha =$ 0.05 (teste bicaudal)
- $n =$ 10 (a diferença entre cada par de dados não é 0)

- Valor crítico: Tabela 9

	Unicaudal, $\alpha = 0{,}05$	$\alpha = 0{,}025$	$\alpha = 0{,}01$	$\alpha = 0{,}005$
n	Bicaudal, $\alpha = 0{,}10$	$\alpha = 0{,}05$	$\alpha = 0{,}02$	$\alpha = 0{,}01$
5	1	—	—	—
6	2	1	—	—
7	4	2	0	—
8	6	4	2	0
9	8	6	3	2
10	11	8	5	3
11	14	11	7	5
12	17	14	10	7
13	21	17	13	10
14	26	21	16	13
15	30	25	20	16

O valor crítico é 8.

Estatística do teste:

Com música	Sem música	Diferença	Valor absoluto	Postos	Postos de sinais
45	38	7	7	7,5	7,5
38	40	-2	2	2	-2
28	33	-5	5	4,5	-4,5
39	36	3	3	3	3
41	42	-1	1	1	-1
47	41	6	6	6	6
62	54	8	8	9	9
54	47	7	7	7,5	7,5
33	28	5	5	4,5	4,5
44	35	9	9	10	10

- Estatística do teste:

A soma dos postos negativos é:

$-1 + (-2) + (-4.5) = -7.5$

A soma dos postos positivos é:

$(+3) + (+4.5) + (+6) + (+7.5) + (+7.5) + (+9) + (+10) = 47.5$

$w_s = 7.5$ (O menor do valor absoluto dessas duas somas: $|-7.5| < |47.5|$)

Decisão: Rejeitar H_0

No nível de significância 5%, há evidência suficiente para apoiar a afirmação de que música faz a diferença na duração das sessões de exercício de atletas.

Exercício

Um pesquisador quer determinar se um novo remédio afeta o número de horas de dor de cabeça experimentadas por pessoas que sofrem deste mal. Para tanto, o pesquisador seleciona 7 pacientes e pergunta o número de horas de dor de cabeça por dia que cada um sente antes e depois de tomar o remédio. Os resultados são mostrados na tabela. Em $\alpha=0,05$, o pesquisador pode concluir que o novo remédio afeta o número de horas de dor de cabeça?

Paciente	1	2	3	4	5	6	7
Horas de dor de cabeça (antes)	0,8	2,4	2,8	2,6	2,7	0,9	1,2
Horas de dor de cabeça (depois)	1,6	1,3	1,6	1,4	1,5	1,6	1,7

Teste de soma de postos de Wilcoxon

- É um teste não paramétrico que pode ser usado para determinar se duas amostras *independentes* foram selecionadas de populações que possuem a mesma distribuição.
- Um requisito é que o tamanho de ambas as amostras deve ser pelo menos 10.
- n_1 representa o tamanho da menor amostra e n_2 representa o tamanho da maior amostra.
- Quando calcular a soma dos postos R, use os postos para a menor das duas amostras.

- Dadas duas amostras independentes, o teste estatístico *z* para o teste de soma de postos de Wilcoxon é:

$$z = \frac{R - \mu_R}{\sigma_R}$$

Onde: *R* = soma dos postos para a menor amostra.

$$\mu_R = \frac{n_1(n_1 + n_2 + 1)}{2}$$

e

$$\sigma_R = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

Desempenhando um teste de soma de testes de Wilcoxon

Em palavras	*Em símbolos*
1. Escreva a afirmação. Identifique as hipóteses nula e alternativa.	Afirme H_0 e H_a.
2. Especifique o nível de significância.	Identifique α.
3. Determine o(s) valor(es) crítico(s)	Use a Tabela Normal
4. Determine o tamanho das amostras.	$n_1 \leq n_2$

Em palavras	*Em símbolos*
5. Encontre a soma dos postos para a menor amostra.	R
a. Liste a informação combinada em ordem ascendente direita.	
b. Posicione a informação combinada.	
c. Adicione a soma dos postos para a menor amostra.	

Em palavras	*Em símbolos*
6. Calcule a estatística do teste.	$\boxed{z = \frac{R - \mu_R}{\sigma_R}}$
7. Tome a decisão de rejeitar ou falhar em rejeitar a hipótese nula.	Se z está na área de rejeição, rejeite H_0. Do contrário, não rejeite H_0.
8. Interprete a decisão no contexto da afirmação original.	

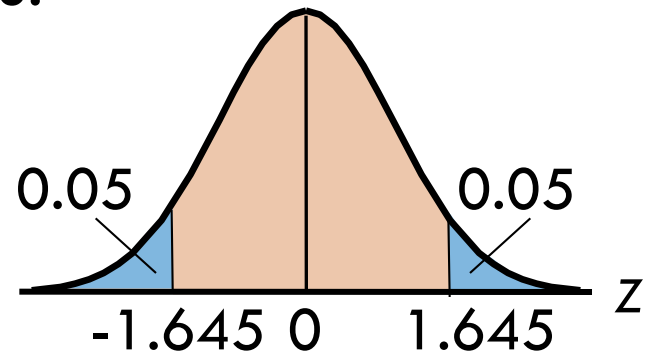
Exemplo: teste de soma de postos de Wilcoxon

A tabela mostra os ganhos (em milhares de dólares) de uma amostra aleatória de representantes de vendas farmacêuticas de 10 homens e 12 mulheres. Em $\alpha = 0{,}10$, você pode concluir que há uma diferença entre os ganhos de homens e mulheres?

Ganho dos homens	58	73	94	81	78	74	66	75	97	79		
Ganho das mulheres	66	57	81	73	65	78	71	67	64	77	80	70

Solução: teste de soma de postos de Wilcoxon

- H_0: Não há diferença entre os ganhos de homens e mulheres.
- H_a: Há uma diferença entre os ganhos de homens e mulheres.
- $\alpha =$ **0.10 (Teste bicaudal)**
- **Região de rejeição:**



Para encontrar os valores de R, μ_R, e σ_R, construa uma tabela que mostre os dados combinados em ordem ascendente e seus postos correspondentes.

Informação ordenada	Amostra	Posto
57	F	1
58	M	2
64	F	3
65	F	4
66	M	5.5
66	F	5.5
67	F	7
70	F	8
71	F	9
73	M	10.5
73	F	10.5

Informação ordenada	Amostra	Posto
74	M	12
75	M	13
77	F	14
78	M	15.5
78	F	15.5
79	M	17
80	F	18
81	M	19.5
81	F	19.5
94	M	21
97	M	22

- Já que a amostra menor é a amostra de homens, R é a soma dos postos dos homens.

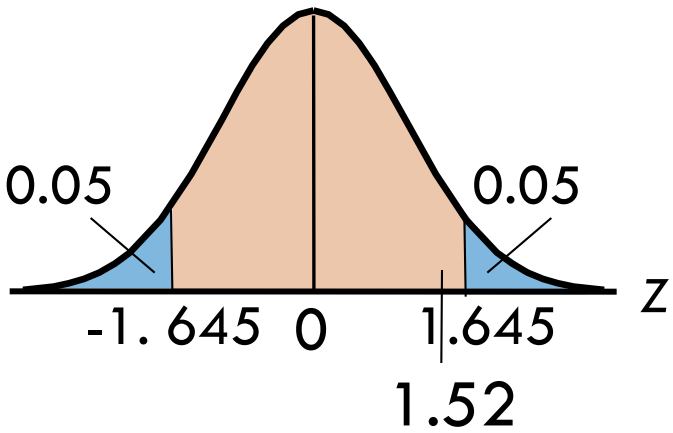
$R = 2 + 5.5 + 10.5 + 12 + 13 + 15.5 + 17 + 19.5 + 21 + 22 = 138$

- Usando $n_1 = 10$ e $n_2 = 12$, podemos encontrar μ_R, e σ_R.

$$\mu_R = \frac{n_1(n_1+n_2+1)}{2} = \frac{10(10+12+1)}{2} = 115$$

$$\sigma_R = \sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}} = \sqrt{\frac{(10)(12)(10+12+1)}{12}} \approx 15.17$$

- H_0: **Não há diferença nos ganhos.**
- H_a: **Há uma diferença nos ganhos.**
- α = **0.10**
- **Região de rejeição:**



- **Estatística do teste**

$$z = \frac{R - \mu_R}{\sigma_R} = \frac{138 - 115}{15.17}$$

$$\approx 1.52$$

- **Decisão:** **Não rejeitar H_0**

No nível de significância 10%, você não pode concluir que há uma diferença entre os ganhos de homens e mulheres.

Exercício

Um pesquisador de cuidados neonatais conduz um estudo de uma seleção aleatória de mães com menos de 20 anos de idade e mães entre 20 e 24 anos de idade. Um número de semanas que cada mãe alimentou seu bebê com leite materno é mostrado na tabela. Em $\alpha=0{,}01$, você pode rejeitar a afirmação do pesquisador de que não há diferença no número de meses em que as mães nesses dois grupos etários alimentam seus bebês com leite materno

Idade da mãe	Duração (em semanas)											
Menos de 20 anos	21	24	24	30	11	17	10	28	21	12	15	
De 20 a 24 anos	25	28	26	23	22	13	28	21	27	32	19	27

Objetivos

- Usamos o teste de postos de sinais de Wilcoxon para determinar se duas amostras dependentes são selecionadas de populações que possuem a mesma distribuição.
- Usamos o teste de soma de postos de Wilcoxon para determinar se duas amostras independentes são selecionadas de populações que possuem a mesma distribuição.

Teste de Kruskal-Wallis

Testes Não Paramétricos

Objetivos

- Usar o teste de Kruskal-Wallis para determinar se três ou mais amostras foram selecionadas de populações que apresentam a mesma distribuição.

Teste de Kruskal-Wallis

- Um teste não paramétrico que pode ser usado para determinar se três ou mais amostras independentes foram selecionadas de populações que possuem a mesma distribuição.

- As hipóteses nula e alternativa para o teste de Kruskal-Wallis são as seguintes:

H_0: Não há diferença na distribuição das populações.
H_a: Há diferença na distribuição das populações.

- Duas condições para usar o teste de Kruskal-Wallis:
  - Cada amostra deve ser selecionada aleatoriamente.
  - O tamanho de cada amostra deve ser no mínimo 5.
- Se essas condições são alcançadas, a distribuição de amostragem para o teste Kruskal-Wallis é aproximada por uma distribuição quiquadrado com graus de liberdade $k - 1$, onde k é o número de amostras.

Estatística de teste para o teste de Kruskal-Wallis

- Dadas três ou mais amostras independentes, o teste estatístico *H* para o teste de Kruskal-Wallis é:

$$H = \frac{12}{N(N+1)}\left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \ldots + \frac{R_k^2}{n_k}\right) - 3(N+1)$$

onde :

k representa o número de amostras;

n_i é o tamanho da i-ésima amostra,;

N é a soma dos tamanhos das amostras;

R_i é a soma dos postos da i-ésima amostra.

Desempenhando um teste Kruskal-Wallis

Em palavras	*Em símbolos*
1. Faça a afirmação. Identfique as hipóteses nula e alternativa.	Afirme H_0 e H_a.
2. Especifique o nível de significância.	Identifique α.
3. Identifique os graus de liberdade.	g.l. $= k - 1$
4. Determine o valor crítico e a região de rejeição.	Use a Tabela Qui-quadrado.

Em palavras	Em símbolos

5. Encontre a soma dos postos para a menor amostra.

   a. Liste a informação combinada em ordem ascendente.

   b. Classifique a informação combinada.

6. Calcule a estatística de teste.

$$H = \frac{12}{N(N+1)} \cdot \left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \ldots + \frac{R_k^2}{n_k} \right) - 3(N+1)$$

Em palavras	*Em símbolos*
7. Tome a decisão de rejeitar ou falhar em rejeitar a hipótese nula.	Se H está na área de rejeição, rejeite H_0. Do contrário, falhe em rejeitar H_0.
8. Interprete a decisão no contexto da afirmação original.	

Exemplo: o teste Kruskal-Wallis

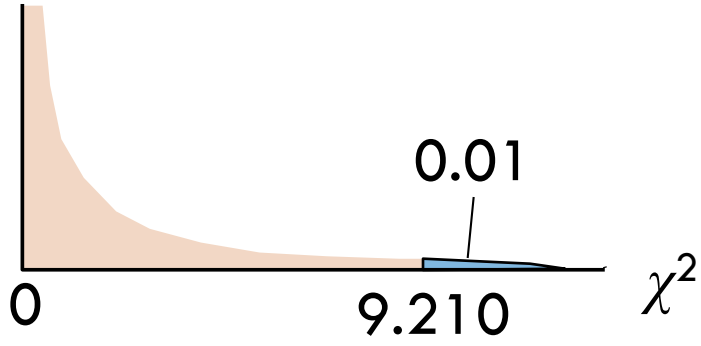
Você quer comparar as taxas de pagamento por hora de atuários que trabalham na Califórnia, Indiana e Maryland. Para tal, você seleciona aleatoriamente vários atuários em cada Estado e registra suas taxas de pagamento por hora. As taxas de pagamento por hora são mostradas na tabela. Em $\alpha = 0{,}01$, você pode concluir que as distribuições das taxas de pagamento por hora dos atuários nesses três estados são diferentes? (*Adaptado de U.S. Bureau of Labor Statistics.*)

Amostra das taxas de pagamento por hora		
CA (Amostra 1)	IN (Amostra 2)	MD (Amostra 3)
40.50	33.45	49.68
44.98	40.12	44.94
47.78	38.65	48.80
43.20	35.98	49.20
37.10	35.97	40.37
49.88	4570	48.79
42.05	42.05	53.82
52.94	35.97	45.35
41.70	38.25	53.25
43.85		43.57

Solução: o teste de Kruskal-Wallis

- H_0: **Não há diferença nas taxas de pagamento por hora nos três estados.**
- H_a: **Há diferença nas taxas de pagamento por hora nos três estados.**

- $\alpha = 0.01$
- **g.l.** $= k - 1 = 3 - 1 = 2$
- **Região de rejeição:**



A tabela mostra as informações combinadas em ordem ascendente e os postos correspondentes.

Informa-ções ordenadas	Amostra	Posto
33.45	IN	1
35.97	IN	2.5
35.97	IN	2.5
35.98	IN	4
37.10	CA	5
38.25	IN	6
38.65	IN	7
40.12	IN	8
40.37	MD	9
40.50	CA	10

Informa-ções ordenadas	Amostra	Posto
41.70	CA	11
42.05	IN	12.5
42.05	CA	12.5
43.20	CA	14
43.57	MD	15
43,85	CA	16
44.94	MD	17
44.98	CA	18
45.35	MD	19
45.70	IN	20

Informa-ções ordenadas	Amostra	Posto
47.78	CA	21
48.79	MD	22
48.80	MD	23
49.20	MD	24
49.68	MD	25
49.88	CA	26
52.94	CA	27
53.25	MD	28
53.82	MD	29

A soma dos postos para cada amostra é:

$R_1 = 5 + 10 + 11 + 12.5 + 14 + 16 + 18 + 21 + 26 + 27$
$= 160.5$

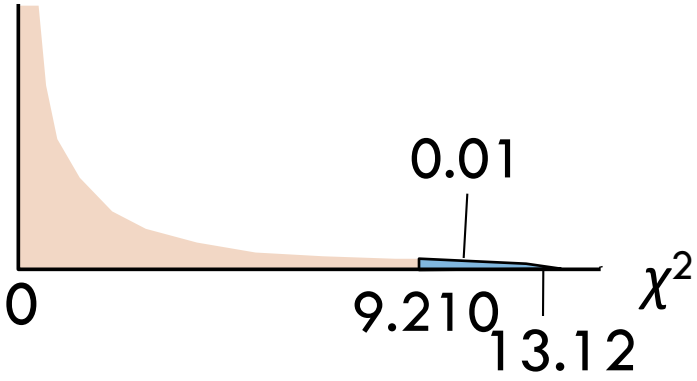
$R_2 = 1 + 2.5 + 2.5 + 4 + 6 + 7 + 8 + 12.5 + 20$
$= 63.5$

$R_3 = 9 + 15 + 17 + 19 + 22 + 23 + 24 + 25 + 28 + 29$
$= 211$

$$H = \frac{12}{N(N+1)}\left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \ldots + \frac{R_k^2}{n_k}\right) - 3(N+1)$$

$$= \frac{12}{29(29+1)}\left(\frac{160.5^2}{10} + \frac{63.5^2}{9} + \frac{211^2}{10}\right) - 3(29+1) \approx 13.12$$

- H_0: **Não há diferença.**
- H_a: **Há diferença.**
- $\alpha =$ **0.01**
- **g.l. = 3 – 1 = 2**
- **Região de rejeição:**



- **Estatística do teste:**

$$H \approx 13.12$$

- **Decisão:** **Rejeitar H_0.**

Em um nível de 1% de significância, você pode concluir que há uma diferença nas taxas de pagamento por hora nos pagamentos dos atuários na Califórnia, Indiana e Maryland.

Exercício

Um representante de seguro independente quer determinar se há uma diferença nos prêmios anuais para seguro de casa em três cidades: São Paulo, Rio de Janeiro e Porto Alegre. Para tal, o representante seleciona aleatoriamente várias casas em cada cidade e registra os prêmios anuais na tabela abaixo. Em $\alpha=0,05$ o representante pode concluir que as distribuições dos prêmios anuais nessas cidades são diferentes?

Cidade	Prêmio Anual (em reais)						
São Paulo	639	588	710	799	669	617	625
Rio de Janeiro	785	890	807	693	759	975	691
Porto Alegre	593	545	650	718	479	590	

Objetivo

- Usamos o teste de Kruskal-Wallis para determinar se três ou mais amostras foram selecionadas de populações que apresentam a mesma distribuição.

CORRELAÇÃO DE POSTOS

Testes Não Paramétricos

Objetivos

- Usar o coeficiente de correlação de postos de Spearman para determinar se a correlação entre duas variáveis é significante.

O coeficiente de correlação de postos de Spearman

- É uma medida da força da relação entre duas variáveis.
- Equivalente não paramétrico para o coeficiente de correlação de Pearson.
- É calculado usando os postos de entradas de amostras de dados emparelhados.
- Denotado como $\boldsymbol{r}_{\mathbf{s}}$.

O coeficiente de correlação de postos de Spearman r_s

- A fórmula para o coeficiente de correlação de postos de Spearman é:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

Onde:

n é o número de entradas de informações emparelhadas;

d é a diferença entre os postos de uma entrada de informação emparelhada.

- Os valores de r_s variam de -1 a 1, inclusive.
  - Se os postos das entradas de informação são idênticas, r_s é igual a +1.
  - Se os postos estão em ordem "reversa", r_s é igual a -1.
  - Se não há relação, r_s é igual a 0.
- Para determinar se a correlação entre as variáveis é significante, você pode desempenhar um teste de hipótese para o coeficiente de correlação da população ρ_s.

- As hipóteses nula e alternativa para esse teste são as seguintes.

  H_0: $\rho_s = 0$ (Não há correlação entre as variáveis.)
  H_a: $\rho_s \neq 0$ (Há uma correlação significativa entre as variáveis.)

Testando a significância do coeficiente de correlação de postos de Spearman

Em palavras	*Em símbolos*
1. Faça a afirmação. Identifique as hipóteses nula e alternativa	Afirme H_0 e H_a.
2. Especifique o nível de significãncia.	Identifique α.
3. Determine o valor crítico.	Use a tabela 10 no Apêndice B.
4. Encontre a estatística do teste.	$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$

Em palavras	*Em símbolos*
5. Tome a decisão de rejeitar ou falhar em rejeitar a hipótese nula.	Se $\lvert r_s \rvert$ é maior que o valor crítico, rejeite H_0. Do contrário, não rejeite H_0.
6. Interprete a decisão dentro do contexto da afirmação original.	

Exemplo: coeficiente de correlação de postos de Spearman

A tabela mostra os preços (em dólares por 100 libras) recebidos por fazendeiros nos Estados Unidos para carne bovina e de carneiro de 1999 a 2005. Em $\alpha = 0{,}05$, você pode concluir que há uma correlação entre os preços das carnes bovina e de carneiro?

(*Fonte: U.S. Department of Agriculture.*)

Ano	Carne bovina	Carne de carneiro
1999	63.4	74.5
2000	68.6	79.8
2001	71.3	66.9
2002	66.5	74.1
2003	79.7	94.4
2004	85.8	101.0
2005	89.7	110.0

Solução: coeficiente de correlação de postos de Spearman

- H_0: $\rho_s = 0$ (Não há correlação entre os preços da carne bovina e de carneiro)
- H_a: $\rho_s \neq 0$ (Há uma correlação entre os preços da bovina e de carneiro)
  - $\alpha =$ 0.05
  - $n =$ 7
  - **Valor crítico:**

n	$\alpha = 0{,}10$	$\alpha = 0{,}05$	$\alpha = 0{,}01$
5	0,900	—	—
6	0,829	0,886	—
7	0,714	0,786	0,929
8	0,643	0,738	0,881
9	0,600	0,700	0,833
10	0,564	0,648	0,794
11	0,536	0,618	0,818

O valor crítico é 0.786.

Carne bovina	Posto	Carne de carnei-ro	Posto	d	d^2
63.4	1	74.5	3	-2	4
68.6	3	79.8	4	-1	1
71.3	4	66.9	1	3	9
66.5	2	74.1	2	0	0
79.7	5	94.4	5	0	0
85.8	6	101.0	6	0	0
89.7	7	110.0	7	0	0

$\Sigma d^2 = 14$

- H_0: $\rho_s = 0$
- H_a: $\rho_s \neq 0$
- $\alpha = 0.05$
- $n = 7$
- **Valor crítico:**

  O valor crítico é 0.786.

- **Estatística do teste:**

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$$= 1 - \frac{6(14)}{7(7^2-1)} = 0.75$$

- **Decisão: Falhar em rejeitar H_0.**

  No nível de significância 5%, você não pode concluir que há uma correlação significante entre os preços de carne bovina e de carneiro entre 1999 e 2005.

Objetivo

- Usamos o coeficiente de correlação de postos de Spearman para determinar se a correlação entre duas variáveis é significante.

TESTES DE CORRIDAS

Testes Não Paramétricos

Objetivos

- Usar o teste de corridas para determinar se um conjunto de dados é aleatório.

Teste de corrida para aleatoriedade

- Uma corrida é uma sequência de dados que possuem a mesma característica.
- Cada corrida é precedida por e seguida de dados com uma característica diferente ou por nenhum dado.
- O número de dados em uma corrida é chamado de extensão da corrida.

Exemplo: encontrando o número de corridas

Uma máquina que distribui líquidos foi projetada para encher garrafas de um litro. Um inspetor de controle de qualidade decide se cada garrafa foi enchida até um nível aceitável e passa na inspeção (*P*) ou se falha na inspeção (*F*). Determine o número de corridas para cada sequência e encontre a extensão de cada corrida.

P P F F F F P F F F P P P P P P

Solução: Há 5 corridas.

Extensão da corrida:	*P P*	*F F F F*	*P*	*F F F*	*P P P P P P*
	2	4	1	3	6

Exemplo: encontrando o número de corridas

P P P P P P P P F F F F F F F F

P F P F P F P F P F P F P F P F

Exemplo: encontrando o número de corridas

P P P P P P P P ***F F F F F F F F***

2 corridas de extensão 8

P F P F P F P F P F P F P F P F

16 corridas de extensão 1

Teste de corridas para aleatoriedade

- Um teste não paramétrico que pode ser usado para determinar se uma sequência de dados de amostra é aleatória.
- Só pode ser usado quando cada valor em um conjunto de dados pode ser categorizado em uma ou duas categorias separadas
- Se tem poucas corridas ou corridas demais, geralmente não é aleatório
- As hipóteses nula e alternativa para esse teste são as seguintes:

H_0: A sequência de dados é aleatória.

H_a: A sequência de dados não é aleatória.

n_1 = número de dados com uma característica

n_2 = número de dados com outra característica

G = número de corridas

Quando $n_1 \leq 20$ e $n_2 \leq 20$, a estatística do teste para o teste de corridas é **G**, o número de corridas.

Quando $n_1 > 20$ ou $n_2 > 20$, a estatística do teste para o teste de corridas é:

$$z = \frac{G - \mu_G}{\sigma_G}$$

onde

$$\mu_G = \frac{2n_1n_2}{n_1 + n_2} + 1 \quad \text{e} \quad \sigma_G = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}$$

Desempenhando um teste de corridas para aleatoriedade

Em palavras	*Em símbolos*
1. Determine a afirmação. Identifique as hipóteses nula e alternativa.	Afirme H_0 e H_a.
2. Especifique o nível de significância. (Use $\alpha = 0.05$ para o teste de corridas.)	Identifique α.
3. Determine o número de dados que possui cada característica e o número de corridas.	Determine n_1, n_2, e G.

Em palavras	*Em símbolos*
4. Determine os valores críticos.	Se $n_1 \leq 20$ e $n_2 \leq 20$, use a Tabela "Valores Críticos para o número de corridas" Se $n_1 > 20$ ou $n_2 > 20$, use a Tabela Z
5. Calcule a estatística do teste.	Se $n_1 \leq 20$ e $n_2 \leq 20$, use G. Se $n_1 > 20$ ou $n_2 > 20$, use $$z = \frac{G - \mu_G}{\sigma_G}.$$

Em palavras	*Em símbolos*
6. Tome a decisão de rejeitar ou falhar em rejeitar a hipótese nula.	Se $G \leq$ ao menor valor crítico ou $G \geq$ ao maior valor crítico, rejeite H_0. Do contrário, não rejeite H_0.
7. Interprete a decisão dentro do contexto da afirmação original.	

Exemplo: usando o teste de corridas

Um administrador de pessoal de uma companhia de construção registra acidentes relatados por trabalhadores durante o seu período de trabalho. A sequência a seguir mostra se alguns acidentes foram relatados durante cada mês em um ano recente. *I* representa um mês no qual ao menos um acidente foi relatado e *N* representa um mês no qual nenhum acidente foi relatado. Em $\alpha = 0{,}05$, você pode concluir que a ocorrência de acidentes em cada mês não é aleatória?

I I N N N I N I I N N N

Solução: usando o teste de corridas

- H_0: A ocorrência de acidentes é aleatória.
- H_a: A ocorrência de acidentes não é aleatória.

I I N N N I N I I N N N

- n_1 = número de *I*s = 5
- n_2 = número de *N*s = 7
- G = número de corridas = 6
- α = 0.05
- Valor crítico: Porque $n_1 \leq 20$, $n_2 \leq 20$, use a Tabela

- **Valor crítico:**
- ***n_1 = 5 n_2 = 7 G = 6***

	Valor de n_2										
	2	3	4	5	6	7	8	9	10	11	12
2	1 6	1 6	1 6	1 6	1 6	1 6	1 6	1 6	1 6	1 6	2 6
3	1 6	1 8	1 8	2 8	2 8	2 8	2 8	2 8	2 8	2 8	2 8
4	1 6	1 8	1 9	2 9	2 9	2 10	3 10	3 10	3 10	3 10	3 10
5	1 6	1 8	2 9	2 10	3 10	3 11	3 11	3 12	3 12	4 12	4 12
6	1 6	2 8	2 9	3 10	3 11	3 12	3 12	4 13	4 13	4 13	4 13
7	1 6	2 8	2 10	3 11	3 12	3 13	4 13	4 14	5 14	5 14	5 14
8	1 6	2 8	3 10	3 11	3 12	4 13	4 14	5 14	5 15	5 15	6 16

O menor valor crítico é 3 e o maior valor crítico é 11.

- H_0: **aleatório**
- H_a: **não aleatório**
- $n_1 = 5 \quad n_2 = 7$
- $G = 6$
- $\alpha = 0.05$
- **Valor crítico:**

Menor valor crítico= 3
Maior valor crítico =11

- **Estatística do teste:**

$$G = 6$$

- **Decisão: Não rejeitar H_0**

No nível de significância 5%, você não possui evidência suficiente para apoiar a afirmação de que a ocorrência de acidentes não é aleatória. Então, parece que os acidentes relatados por trabalhadores durante o período de trabalho do administrador ocorrem aleatoriamente.

Exemplo: usando o teste de corridas

Você quer determinar se a seleção de empregados recentemente contratados em uma grande companhia é aleatória em relação ao gênero. Os gêneros de 36 empregados recentemente contratados são mostrados a seguir. Em $\alpha = 0{,}05$, você pode concluir que a seleção não é aleatória?

M M F F F F M M M M M M

F F F F F M M M M M M M

F F F M M M M F M M F M

Solução: usando o teste de corridas

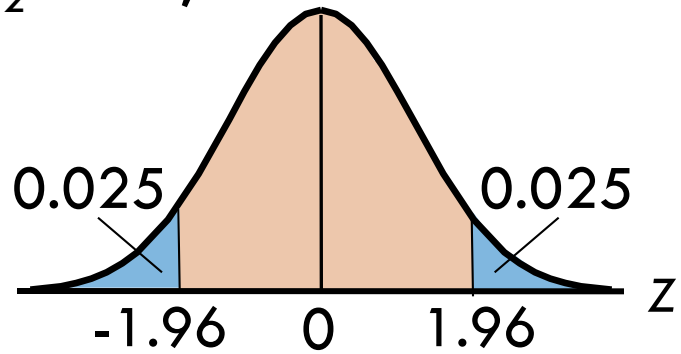
- H_0: A seleção de empregados é aleatória.
- H_a: A seleção de empregados não é aleatória.

$$M\,M \quad F\,F\,F\,F \quad M\,M\,M\,M\,M\,M$$
$$F\,F\,F\,F\,F \quad M\,M\,M\,M\,M\,M\,M$$
$$F\,F\,F \quad M\,M\,M\,M \quad F \quad M\,M \quad F \quad M$$

- **n_1 = número de *F*s = 14**
- **n_2 = número de *M*s = 22**
- **G = número de corridas = 11**

- H_0: aleatória
- H_a: não aleatória
- $n_1 = 14$ $n_2 = 22$
- $G = 11$
- $\alpha = 0{,}05$
- Valor crítico:

$n_2 > 20$, use a Tabela Z



- Estatística do teste:

- Decisão:

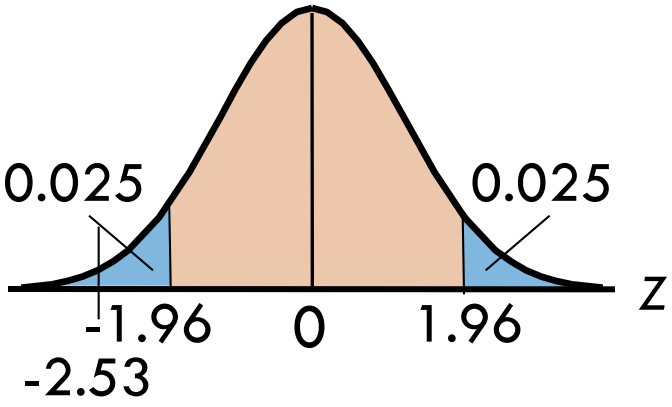
$$\mu_G = \frac{2n_1n_2}{n_1+n_2} + 1 = \frac{2(14)(22)}{14+22} + 1 \approx 18.11$$

$$\sigma_G = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}\frac{1}{n}}$$

$$= \sqrt{\frac{2(14)(22)[2(14)(22) - 14 - 22)}{(14+22)^2(14+22-1)}} \approx 2.81$$

$$z = \frac{G - \mu_G}{\sigma_G} = \frac{11 - 18.11}{2.81} \approx -2.53$$

- H_0: aleatório
- H_a: não aleatório
- $n_1 = 14$ $n_2 = 22$
- $G = 11$
- $\alpha = 0{,}05$
- Valor crítico:



- Estatística do teste:

  $z = -2.53$

- Decisão: Rejeite H_0

Você tem evidência suficiente, no nível de significância 5%, para apoiar a afirmação de que a seleção de empregados em relação ao gênero não é aleatória.

Exercício

Um treinador registra os resultados de jogadas de moeda no início de cada jogo de futebol em uma temporada. Os resultados estão a seguir, onde H representa cara e T representa coroa. O técnico afirmou que as jogadas de moeda não foram aleatórias. Use um teste de corridas para testar a afirmação do técnico.

H T T T H T H H T T T T H T H H

Exercício

Um gerador de números produz a sequencia de dígitos mostrada a seguir, onde O representa o dígito impar e E o dígito par. Teste a afirmação de que estes dígitos não foram gerados aleatoriamente.

O O O E E E E O O O O O E E E E O O E E E E O O O O E E E E O O

Objetivo

- Usamos o teste de corridas para determinar se um conjunto de dados é aleatório.