Estatística Não Paramétrica

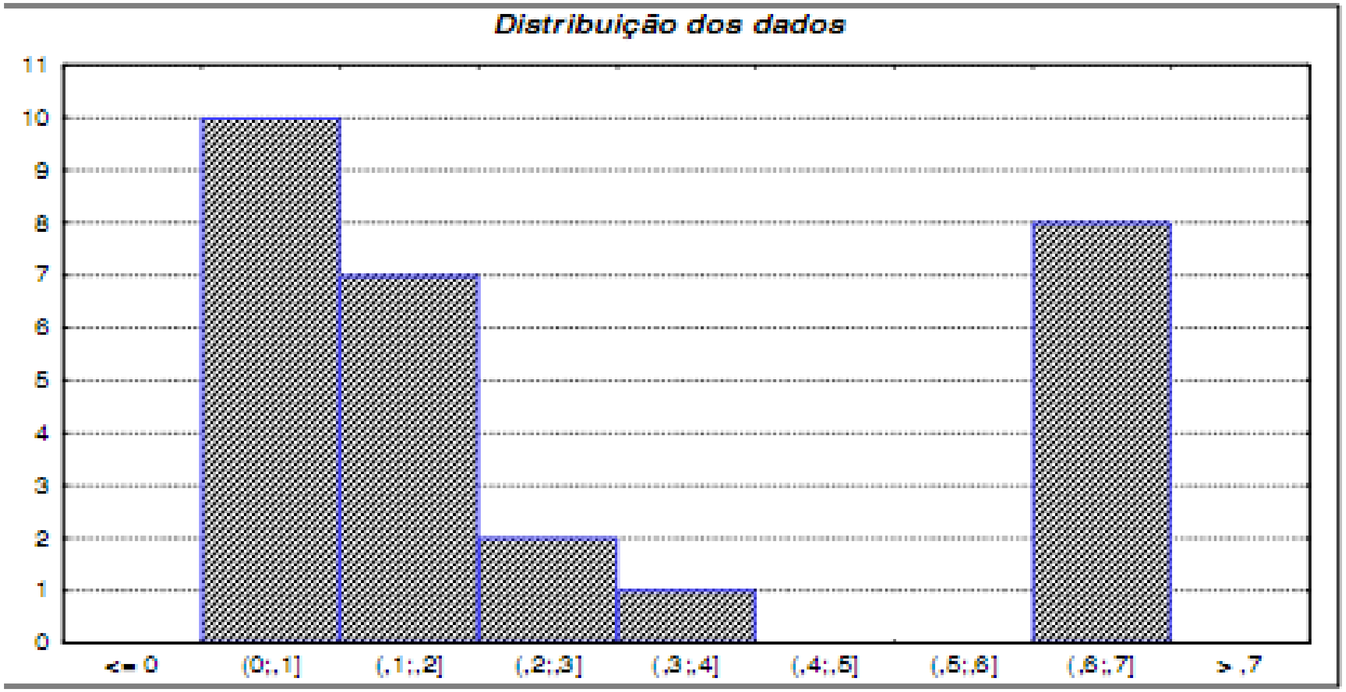
- Como construir testes de hipóteses para uma amostra
- Como construir testes de hipóteses para duas amostras dependentes
- Como construir testes de hipóteses para duas amostras independentes

Motivação

- Os testes t assumem que a população é aproximadamente normal.
- Situações que uma população difere muito de uma normal, é recomendável testes não paramétricos.
- Vantagens dos testes não paramétricos:
  - Envolvem cálculos mais simples.
  - Não exigem populações normais.
- Desvantagem:
  - Não são tão eficientes quanto os testes paramétricos. Precisa-se de evidência mais forte (aumentar o tamanho da amostra) para rejeitar hipótese nula.

Exemplo

Concentrações de partículas em 28 amostras de solo após aplicação de um pesticida.



$\overline{X} = 0{,}314 \qquad s^2 = 0{,}07 \qquad s = 0{,}264$

$$H_0 : \mu = 0,35$$
$$H_a : \mu < 0,35$$

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}} = \frac{0,314 - 0,35}{0,264/\sqrt{28}} = -0,72$$

- Não temos porque rejeitar a hipótese nula ao usarmos o teste de hipótese paramétrico (valor da tabela -1.0733). Assim a afirmação do fabricante estará comprovada. Porém: podemos notar uma grande assimetria na distribuição dos dados, o que causou um aumento do valor da média em favor da hipótese nula;

Teste dos Sinais para mediana de uma amostra

- Foi introduzido em 1710 por Arbuthnott. Um dos testes mais fáceis.
- Os dados podem ser convertidos para sinais de mais ou de menos. A idéia é analisar a freqüência dos sinais de mais ou de menos.
- Suposição: Amostra aleatória.

Exemplo 1

- Considere uma amostra de 12 pessoas saudáveis. Após a seleção aleatória das 12 pessoas, foi medido a temperatura de cada uma delas: 97,6; 97,5; 98,6; 98,2; 98,0; 99,0; 98,5; 98,1; 98,4; 97,9; 97,9; 97,7. Use α=0,05 para testar a afirmativa de que essas temperaturas do corpo provêm de uma população com mediana inferior ou igual a 98,6 ^{o}F.

Estatística de Teste e Valores Críticos Amostras Pequenas ($n<25$)

- Selecione um valor para α
- Mediana – valor amostral
  - Sinal - se mediana > valor
  - Sinal + se mediana < valor
- Estatística do teste:
  - A: Bilateral H_0: $M=M_0$ $H_{\acute{a}}$: $M \neq M_0$
    - x=O número de vezes que o sinal (de mais ou de menos) menos freqüente ocorreu.
  - B: Unilateral H_0: $M \leq M_0$ $H_{\acute{a}}$: $M>M_0$
    - x=O número de vezes que o sinal de menos ocorreu.
  - C: Unilateral H_0: $M \geq M_0$ $H_{\acute{a}}$: $M<M_0$
    - x=O número de vezes que o sinal de mais ocorreu
- O valor crítico pode ser obtido usando a tabela A-7 (Triola).

Hipóteses e Regiões Críticas

- A. Bilateral H_0: $M=M_0$ $H_á$: $M \neq M_0$
  - Rejeita-se a hipótese nula se é observado um número pequeno de sinais mais ou de sinais menos ou seja se a probabilidade de se observar um número pequeno de sinal mais ou de sinal menos for menor ou igual a $\alpha/2$. $x \leq C$ (valor crítico)
- B. Unilateral H_0: $M \leq M_0$ $H_á$: $M > M_0$
  - Rejeita-se a hipótese nula se é observado um número pequeno de sinal menos ou seja se a probabilidade de se observar um número pequeno de sinal menos for menor ou igual a α. $x \leq C$ (valor crítico)
- C. Unilateral H_0: $M \geq M_0$ $H_á$: $M < M_0$
  - Rejeita-se a hipótese nula se é observado um número pequeno de sinal mais ou seja se a probabilidade de se observar um número pequeno de sinal mais for menor ou igual a α. $x \leq C$ (valor crítico)

Estatística de Teste e Valores Críticos Amostras Grandes ($n \geq 25$)

- Selecione um valor para α
- Estatística do teste:

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

- O valor crítico pode ser obtido usando a tabela da normal padrão).

Solução do exemplo 1

- Unilateral H_0: M ≥98,6 $H_á$: M<98,6.

97,6	97,5	98,6	98,2	98,0	99,0	98,5	98,1	98,4	97,9	97,9	97,7
-	-		-	-	+	-	-	-	-	-	-

- Como n=11 pois houve um sinal de igual e n < 25, a estatística de teste é x=1 porque é número de vezes que o sinal de mais ocorreu.
- O valor crítico na tabela de teste de sinais é 2. Como x=1<2, então rejeita-se a hipótese nula.

Exemplo 2

- Uma cadeia de restaurantes foi processada por discriminação baseada em sexo, porque apenas 30 homens foram contratados juntamente com 70 mulheres. Use o teste de sinais com α=5% para testar a hipótese que homens e mulheres são contratados igualmente por essa companhia.

Teste dos Sinais para uma proporção

- Só existem dois resultados possíveis para o experimento.
- Determina-se o sucesso e registra-se como sinal de mais o sucesso e sinal de menos o fracasso.
- As hipóteses são construídas com respeito a proporção (sucesso).
- Hipóteses
  - Bilateral H_0: $p=p_0$ $H_á$: $p \neq p_0$
  - Unilateral H_0: $p \leq p_0$ $H_á$: $p > p_0$
  - Unilateral H_0: $p \geq p_0$ $H_á$: $M_d < p_0$
- Teste para duas proporções: teste quiquadrado de homogeneidade.

Solução do exemplo 2

- Considerando por + os homens e por – as mulheres. A estatística de teste x=30 (o menor entre 30 e 70).
- Como n > 25, calculamos a estatística z.

$$z = \frac{(x+0{,}5)-(n/2)}{\sqrt{n}/2} = \frac{30{,}5-50}{10/2} = -3{,}90$$

- Os valores críticos são z=-1,96 e z=1,96. Portanto, rejeitamos hipótese que a proporção de homens é igual a 0,5 ou seja existe diferença significativa entre as proporções de homens e mulheres.

Exemplo 3

- Considere uma amostra de 12 pessoas saudáveis. Após a seleção aleatória das 12 pessoas, foi medido a temperatura de cada uma delas: 97,6; 97,5; 98,6; 98,2; 98,0; 99,0; 98,5; 98,1; 98,4; 97,9; 97,9; 97,7. Use α=0,05 para testar a afirmativa de que essas temperaturas do corpo provêm de uma população com mediana inferior ou igual a 98,6 ^{o}F.

Teste de Postos com Sinais de Wilcoxon para uma mediana

- Um posto é número atribuído a um item amostral individual de acordo com a sua posição na lista ordenada.
- O teste tende a fornecer conclusões que refletem melhor a verdadeira natureza dos dados.
- Procedimentos:
  - 1.Calcule a diferença entre o valor da mediana da hipótese nula e os valores da amostra. Ignore os pares cuja diferença é zero.
  - 2. Ignore os sinais, ordene-os em ordem crescente e substitua os valores pelos postos.
  - 3. Atribua a cada posto o sinal da diferença que o originou.
  - 4. Ache a soma dos valores absolutos dos postos negativos e dos postos positivos.
  - 5. Seja n o número de pares de dados para os quais a diferença não é zero.
  - 6. Calcule a estatística de teste.

Teste de Postos com Sinais de Wilcoxon para uma mediana para $n \leq 30$

- Incorpora e usa mais informação que o teste de sinais. Usa postos de dados amostrais.
- Suposição: Amostra aleatória
- Selecione um valor para α
- Estatística do teste:
  - A. Bilateral H_0: $M=M_0$ $H_á$: $M \neq M_0$
    - T= à menor de duas somas seguintes: soma de de postos negativos e soma de postos positivos.
  - B. Unilateral H_0: $M \leq M_0$ $H_á$: $M>M_0$
    - T= à soma de de postos negativos.
  - C. Unilateral H_0: $M \geq M_0$ $H_á$: $M<M_0$
    - T= à soma de de postos positivos.
- Valores críticos da tabela

Tabela

n	Two-Tailed Test		One-Tailed Test	
	α = .05	α = .01	α = .05	α = .01
5	--	--	0	--
6	0	--	2	--
7	2	--	3	0
8	3	0	5	1
9	5	1	8	3
10	8	3	10	5
11	10	5	13	7
12	13	7	17	9
13	17	9	21	12
14	21	12	25	15
15	25	15	30	19
16	29	19	35	23
17	34	23	41	27
18	40	27	47	32
19	46	32	53	37
20	52	37	60	43
21	58	42	67	49
22	65	48	75	55
23	73	54	83	62
24	81	61	91	69
25	89	68	100	76
26	98	75	110	84
27	107	83	119	92
28	116	91	130	101
29	126	100	140	110
30	137	109	151	120

Teste de Postos com Sinais de Wilcoxon para uma mediana para n > 30

- Estatística do teste

$$z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

- Valores críticos da tabela z (normal padrão).
- n=número de pares de dados para os quais a diferença não é 0.

Solução do exemplo 3

97,6	97,5	98,6	98,2	98,0	99,0	98,5	98,1	98,4	97,9	97,9	97,7
1,0	0,11	0	0,04	0,06	0,04	0,01	0,05	0,02	0,07	0,07	0,09
11	10		3,5	6	3,5	1	5	2	7,5	7,5	9
-11	-10		-3,5	-6	3,5	-1	-5	-2	-7,5	-7,5	-9

- T=3,5 e n=11 , 25.
- O valor crítico da tabela é 13.
- Como T=3,5 < 13, rejeita-se a hipótese nula.

Exemplo 4

- Fazem medidas mentais de crianças dando-lhes blocos e pedindo-lhes que construam uma torre o mais alto possível. O experimento de blocos foi repetido um mês depois com os tempos listados em segundos. Use α=5% para testar a afirmativa de que não há diferença entre os tempos da primeira e da segunda tentativas.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1ª	30	19	19	23	29	178	42	20	12	39	14	81	17	31	52
2ª	30	6	14	8	14	52	14	22	17	8	11	30	14	17	15
d	0	13	5	15	15	126	28	-2	-5	31	3	51	3	14	37
Postos		6	4,5	8,5	8,5	14	10	1	4,5	11	2,5	13	2,5	7	12
Sinal	0	6	4,5	8,5	8,5	14	10	-1	-4,5	11	2,5	13	2,5	7	12

Teste de Postos com Sinais de Wilcoxon para amostras combinadas

- Procedimentos para encontrar a estatística de teste:
  - 1.Para cada par encontre a diferença , subtraindo o primeiro valor do segundo. Conserve os sinais mas ignore quaisquer pares para os quais d=0.
  - 2. Faça os procedimentos 2 a 6 do teste de postos com sinais para uma mediana.

Solução do exemplo 5

- Soma valores absolutos dos postos negativos:5,5.
- Soma dos valores absolutos dos postos positivos: 99,5
- Fazendo $T=5,5$ e $n=14 < 25$ temos, pela tabela que o valor crítico é 21. Portanto rejeita-se H_0

Teste de Postos com Sinais de Wilcoxon para amostras independentes

- Suposições
  - Amostras independentes e aleatórias.
  - Equivalente ao Teste U MannWhitney.
- A idéia é: se duas amostras são extraídas de populações idênticas e se associam postos a todos os valores individuais combinados em uma única coleção de valores, então os postos altos e baixos devem se distribuir igualmente entre as duas amostras.

Teste de Postos com Sinais de Wilcoxon para amostras independentes

- Procedimento para encontrar a estatística de teste.
  - 1. Combine temporariamente as duas amostras em uma única amostra. Substituta cada valor amostral pelo seu posto. Se houver valores empatados, associe a cada um deles a média dos postos envolvidos no empate.
  - 2. Ache a soma de cada uma das amostras.
  - 3. Calcule o valor da estatística z.

Teste de Postos com Sinais de Wilcoxon para amostras independentes

- Hipóteses:
  - H_0: as amostras têm valores de medianas iguais
  - $H_á$: as amostras têm valores de medianas diferentes
- Notação
  - n_1 = tamanho da amostra 1
  - n_2 = tamanho da amostra 2
  - R_1= soma dos postos da amostra 1
  - R_2 = soma dos postos da amostra 2
  - $R=R_1$ (menor soma)
  - μ_R = média dos valores de R
  - σ_R = desvio padrão dos valores de R

Teste de Postos com Sinais de Wilcoxon para amostras independentes

- Estatística de teste

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

$$\mu_R = \frac{n_1(n_1 + n_2) + 1}{2} \qquad \sigma_R = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

- Valores críticos da tabela da normal padrão