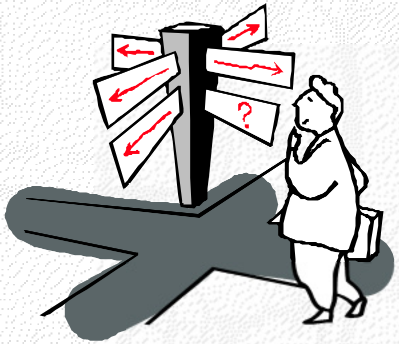


Testes Não Paramétricos

DAD (Duas Amostras Dependentes)

Prof. Lorí Viali, Dr.

http://www.mat.ufrgs.br/viali/

viali@mat.ufrgs.br

Testes para duas Amostras Dependentes

Os testes

- O teste de McNemar
- O teste de Wilcoxon
- O teste do sinais

O teste De McNemar

O teste de McNemar para a significância de mudanças é aplicável a experimentos do tipo "antes e depois" em que cada sujeito é utilizado como seu próprio controle e a medida é efetuada em escala nominal ou ordinal.

Para testar a significância de qualquer mudança observável, através deste método, é necessário construir uma tabela de frequências “2x2”. Veja exemplo a seguir:

A tabela 2x2

		Depois	
		-	+
Antes	+	A	B
	-	C	D

Note-se que aqueles casos que mostram mudanças entre a primeira e a segunda resposta aparecem nas células *A* e *D*. Um sujeito é contado na célula *A* se ele muda de + para - e é contado na *D* se ele muda de - para +.

Se nenhuma mudança é ocorre ele é contado nas células B (resposta + antes e depois) e C (resposta - antes e depois).

Hipóteses

Como A + D representa o número total de elementos que acusaram alguma modificação, a expectativa, sob a hipótese de nulidade, é de que 1/2 (A + D) acuse modificações em um sentido e 1/2 (A + D) no outro sentido.

Variável Teste

$$\chi^2_1 = \frac{\sum_{i=1}^{k} (O_i - E_i)^2}{E_i} = \frac{\left(A - \frac{A+D}{2}\right)^2}{\frac{A+D}{2}} = \frac{\left(D - \frac{A+D}{2}\right)^2}{\frac{A+D}{2}}$$

Simplificando vem:

$$\chi^2_1 = \frac{(A-D)^2}{A+D}$$

Correção de Continuidade

A correção torna-se necessária porque uma distribuição contínua, no caso, o qui-quadrado está sendo usada para aproximar uma distribuição discreta. Quando todas as frequências esperadas são pequenas, esta aproximação pode não ser boa.

A correção de continuidade (de Yates) é uma tentativa de remover esta fonte de erro. A expressão acima incluindo a correção de Yates fica:

$$\chi_1^2 = \frac{(|A - D| - 1)^2}{A + D}$$

E x e m p l o

Uma pesquisa realizada entre donos de automóveis sobre a necessidade do uso do cinto de segurança foi realizada antes e depois de um filme sobre acidentes, onde era enfocado os benefícios do uso do cinto.

Dos 70 motoristas entrevistados 20 eram a favor do uso do cinto antes e continuaram após, 30 eram contra antes e ficaram a favor após, 15 eram contra antes e continuaram contra após e 5 eram a favor e ficaram contra após. Teste, ao nível de 1%, a significância das mudanças.

Hipóteses

H_0: A proporção de mudanças de A para B é igual a de B para A, isto é,

$$P_A = P_B = 1/2$$

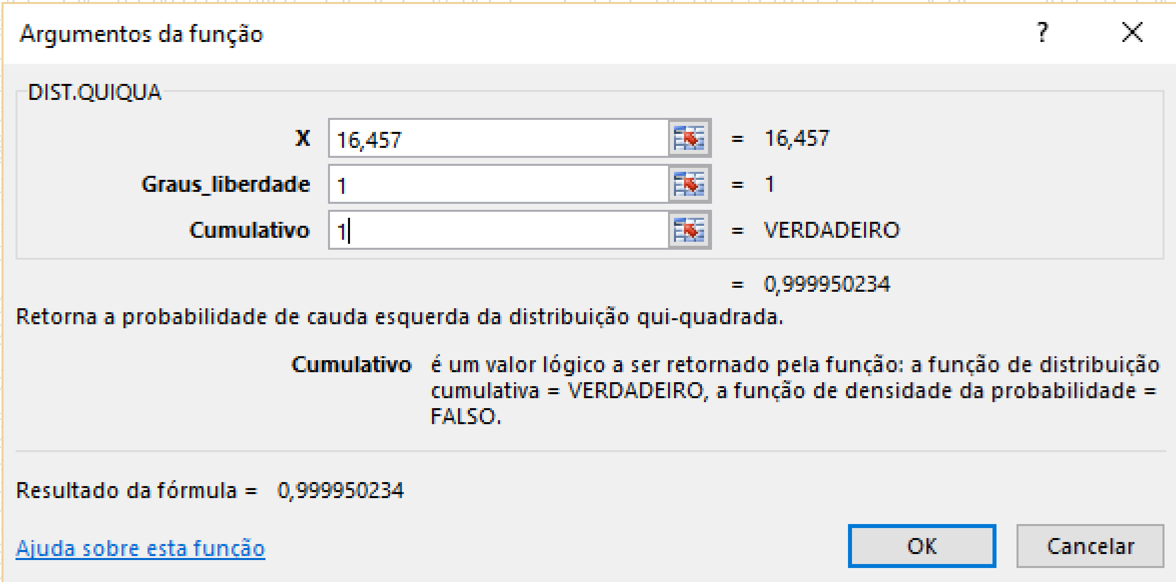
H_1: $P_A > P_B$

A evidência

		Depois	
		-	+
Antes	+	5	20
	-	15	30

A estatística teste

$$\chi_1^2 = \frac{(|5-30|-1)^2}{5+30} = 16{,}457$$



Significância do Resultado

O valor obtido pela função Qui-quadrado da planilha fornece $P(X \leq x) = 0{,}999950$. O que é necessário é $P(X > x)$ = valor-p = 1 - 0,999950 = 0,0050%, que é um resultado significativo a 1% ou até mesmo a 0,1%, portanto é possível concluir que as mudanças são significativas.

O teste de Wilcoxon

Objetivos

A prova de Wilcoxon de duas amostras emparelhadas é a equivalente não paramétrica ao teste t para duas amostras dependentes. As hipóteses são as mesmas, embora às vezes elas possam ser colocadas em termos da mediana e não da média.

Hipóteses

H_0: *A diferença entre as médias (ou medianas) populacionais é zero.*

H_1: *A diferença entre as médias (ou medianas) não é zero.*

Suposições

A suposição básica por trás deste teste é que as distribuições populacionais são simétricas (médias e medianas idênticas).

Metodologia

Inicialmente calcular d_i = diferença dentro do par "i". A seguir atribuir postos a cada d_i, independentemente de sinal. Ao menor d_i, atribuir o posto 1; ao próximo 2, etc. A cada posto atribuir o sinal da diferença, isto é, identificar quais postos decorrem de diferenças negativas e quais de diferenças positivas.

Se as duas classificações são equivalentes, isto é se Ho é verdadeira, é de se esperar que algumas das maiores diferenças sejam positivas e outras negativas. Desta forma, se forem somados os postos com sinal mais e os postos com sinal menos, deve-se esperar somas aproximadamente iguais.

Se houver diferença entre estas duas somas é sinal de que as duas classificações (ou tratamentos) não se equivalem e deve-se então rejeitar a hipótese nula.

Se as duas amostras foram extraídas da mesma população, então se espera que as distribuições acumuladas das amostras estejam próximas. Se as distribuições estão “distantes” isto sugere que as amostras provenham de populações distintas e um desvio grande pode levar a rejeição da hipótese de nulidade.

Empates

Eventualmente os escores de dois pares serão iguais. Neste caso eles **devem ser** excluídos da análise e o valor de n deve ser reduzido na mesma quantidade de valores em que a diferença for nula.

Pode ocorrer, ainda, um outro tipo de empate. Duas ou mais diferenças podem ter o mesmo valor absoluto. Neste caso, atribui-se o mesmo posto aos empates. Este posto é a média dos postos que teriam sido atribuídos se as diferenças fossem diferentes.

Por exemplo, se três pares acusam as diferenças: -1, -1 e +1, a cada par será atribuído o posto 2, que é a média entre 1, 2 e 3. O próximo valor, pela ordem, receberia o valor 4, porque já teriam sido utilizados os postos 1, 2 e 3.

Pequenas Amostras ($n < 25$)

Se T = a menor soma dos postos de mesmo sinal (negativos ou positivos) então T será significativo se não superar o valor dado na tabela, sob determinado nível de significância.

Grandes Amostras ($n \geq 25$)

Neste caso $\mathcal{T}$ (menor soma) é aproximadamente normal com os seguintes parâmetros:

$$\mu_{\mathcal{T}} = \frac{n(n+1)}{4}$$

$$\sigma_{\mathcal{T}} = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

Quando existem empates a variabilidade deve ser calculada corrigida por:

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum_{j=1}^{l}(t_j^3 - t_j)}{48}}.$$

Onde: n = *total de pares sem empates;*

l = *número de empates e*

t_j = *número de elementos no j-ésimo empate.*

Exemplo

Um grupo de 20 motoristas foi submetido a um teste para verificar o efeito do álcool na percepção de obstáculos. O número de cones derrubados antes e depois da ingestão de uma dose de destilado foi anotado.

M	A	D	M	A	D
1	0	2	11	0	2
2	1	1	12	1	3
3	2	2	13	2	2
4	1	3	14	4	3
5	2	3	15	3	2
6	3	4	16	5	4
7	4	5	17	1	3
8	3	2	18	3	6
9	4	4	19	1	5
10	1	2	20	3	4

Teste a hipótese de que o álcool não tem influência sobre a percepção dos motoristas.

S o l u ç ã o

Fazendo a diferença entre as duas variáveis, isto é, entre o número de cones derrubados antes e depois da ingestão de álcool, tem-se os valores da tabela.

M	D	M	D
1	-2	11	-2
2	0	12	-2
3	0	13	0
4	-2	14	1
5	-1	15	1
6	-1	16	1
7	-1	17	-2
8	1	18	-3
9	0	19	-4
10	-1	20	-1

Eliminando os pares com valor zero, tem-se uma nova tabela, agora com apenas 16 pares.

M	D	M	D
1	-2	11	1
2	-2	12	1
3	-1	13	-2
4	-1	14	-3
5	-1	15	-4
6	1	16	-1
7	-1		
8	-2		
9	-2		
10	1		

Ordenando as diferença em valores absolutos, mas mantendo a ordem em que os escores aparecem, isto é, se o positivo apareceu primeiro então ela será o primeiro na lista ordenada. Após, atribuindo escores a esses valores tem-se a seguinte tabela :

M	D	DO	E	M	D	DO	E
1	-2	-1	5	11	1	-2	12
2	-2	-1	5	12	1	-2	12
3	-1	-1	5	13	-2	-2	12
4	-1	1	5	14	-3	-2	12
5	-1	-1	5	15	-4	-3	15
6	1	1	5	16	-1	-4	16
7	-1	1	5				
8	-2	1	5				
9	-2	-1	5				
10	1	-2	12				

Assim a soma dos escores negativos será igual a $T_{-} = 116$ e a soma dos escores positivos será igual a $T_{+} = 20$. Logo o valor da estatística de Wilcoxon será o menor destes dois valores.

Aproximando pela Normal

Apesar de $n < 20$, vamos aproximar pela Normal apenas para comparar com o resultado do SPSS. A média e o desvio são dados por:

$$\mu_T = \frac{n(n+1)}{4} \qquad \sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

Como, neste exemplo, existem empates, é necessário corrigir a variabilidade.

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} =$$

$$= \sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum_{j=1}^{\ell}(t_j^3 - t_j)}{48}}.$$

A correção para empates, neste caso, será:

$$\frac{\sum_{j=1}^{\ell}(t_j^3 - t_j)}{48} = \frac{9^3 - 9 + 5^3 - 5}{48} = \frac{840}{48} = 17,50.$$

A variabilidade, para este exemplo, será então:

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum_{j=1}^{\ell}(t_j^3 - t_j)}{48}} =$$

$$= \sqrt{\frac{16(16+1)(2.16+1)}{24} - 80} =$$

$$= \sqrt{\frac{16.17.33}{24} - 80} = \sqrt{2.11.17} = 17,1464 .$$

Então, como n = 16, a média é:

$$\mu_T = \frac{n(n+1)}{4} = \frac{16(16+1)}{4} = $$
$$= 4.17 = 68.$$

Assim, para este caso, z será igual a:

$$z = \frac{20 - 68}{18,8812} = -2,542.$$

A significância unilateral deste resultado será:

$P(Z < -2{,}542) = 0{,}0055.$

E a bilateral será:

$P(Z < -2{,}542 \text{ ou } Z > 2{,}542) = 1{,}10\%.$

Resultados - SPSS

		N	Mean Rank	Sum of Ranks
Antes – Depois	Negative Ranks	4	5,00	20,00
	Positive Ranks	12	9,67	116,00
	Ties	4		
	Total	20		

	Antes - Depois
Z	-2,542 (a)
Asymp. Sign (2 tailed)	0,011
Exact Sig. (2-tailed)	0,010
Exact Sig. (1-tailed)	0,005
Point Probability	0,002

a Based on negative ranks.

b Wilcoxon Signed Ranks Test

O teste dos Sinais

O teste Binomial para duas amostras dependentes é essencialmente uma extensão do teste binomial (sinais) para uma amostra. Sempre que as suposições do teste t para duas amostras ou o de Wilcoxon forem violadas o teste dos sinais poderá ser utilizado.

A distribuição binomial com $p = 0,5$ fornece as probabilidades associadas à ocorrência, sob H_0, de valores tão pequenos quanto x para n dado. Nesse caso, faça x = número de sinais que ocorrem com maior frequência.

Objetivos

O teste dos sinais utiliza como dados os sinais “mais” e “menos”, ao invés de medidas quantitativas. É indicado em situações onde é inviável determinar medidas quantitativas, mas sendo possível a ordenação.

É aplicável ao caso de duas amostras relacionadas quando se deseja determinar se duas condições são diferentes. A única suposição que a prova exige é que a variável em estudo tenha distribuição contínua.

Hipóteses

$$\mathcal{H}_0\text{:}\;\; \mathcal{P}(X_{\mathcal{A}} > X_{\mathcal{B}}) = \mathcal{P}(X_{\mathcal{A}} < X_{\mathcal{B}}) = 1/2$$

$$\mathcal{H}_1\text{:}\;\; \mathcal{P}(X_{\mathcal{A}} > X_{\mathcal{B}}) = \mathcal{P}(X_{\mathcal{A}} < X_{\mathcal{B}}) \neq 1/2$$

Pequenas amostras

A probabilidade associada à ocorrência de um determinado número de sinais "mais" e "menos" pode ser obtida pela distribuição binomial com $p = q = 1/2$ onde n = número de pares.

Se um par não acusa diferença (i. é, se a diferença, sendo nula, não tem sinal), então ele será desprezado na análise, reduzindo-se, assim, o valor de n.

A prova de sinais pode ser unilateral ou bilateral. No segundo caso deve-se dobrar os valores de p obtidos na tabela D.

Empates

Para esse teste ocorre um "empate" quanto não é possível discriminar entre os dois valores do par ou quando os dois escores de cada par são iguais. Todos os casos empatados são eliminados da análise, reduzindo-se o valor de n.

Exemplo:

A tabela mostra as horas de alívio proporcionado por dois remédios contra a artrite. Existe evidência de que uma droga é mais eficaz do que a outra no alívio da dor?

Caso	R_A	R_B	Caso	R_A	R_B
1	2,0	3,5	7	14,9	16,7
2	3,6	5,7	8	6,6	6,0
3	2,6	2,9	9	2,3	3,8
4	2,6	2,4	10	2,0	4,0
5	7,3	9,9	11	6,8	9,1
6	3,4	3,3	12	8,5	20,9

Obtendo as diferenças, tem-se:

Caso	Dif.	Caso	Dif
1	-1,5	7	-1,8
2	-2,1	8	0,6
3	-0,3	9	-1,5
4	0,2	10	-2,0
5	-2,6	11	-2,3
6	0,1	12	-12,4

Nesse caso, tem-se:

$r^{+} = 9$

$r^{-} = 3$

$n = 12$

Empates = 0

Se $\mathcal{H}_1$ for unilateral então o valor-p será igual a $P(X \leq 3) = 7{,}30\%$.

Se $\mathcal{H}_1$ for bilateral então o valor-p será:

$$P(X \leq 3) + P(X \geq 9) = 7{,}30\% + 7{,}30\% =$$

$$= 14{,}60\%.$$

Grande amostras

Se $n > 25$ então a distribuição Binomial poderá ser aproximada pela normal de média:

$$\mu = E(X) = np$$

E desvio padrão igual a:

$$\sigma = \sqrt{npq} = \frac{\sqrt{n}}{2} = 0,5\sqrt{n}$$

Essa aproximação poderá ser melhorada com a introdução de uma correção de continuidade que consiste em somar ou subtrair o valor de 0,5 conforme a região de rejeição esteja à esquerda ou à direita da média.

MCNEMAR, Quinn. Note on the sampling error of the difference between correlated proportions or percentages. ***Psychometrika****, v. 12, n. 2, 1947.*

SHESKIN, David J. ***Handbook of Parametric and Nonparametric Statistical Procedures.*** *4th ed. Boca Raton (FL): Chapman & Hall/CRC, 2007.*

WILCOXON, F. Individual Comparisons by Ranking Methods. ***Biometrics Bulletin****, v. 1, p. 80-3, 1945.*