Exemplo do teste estatístico não-paramétrico de soma das ordens de *Wilcoxon*

As empresas que fabricam e/ou comercializam produtos de grande consumo investem grandes quantidades de dinheiro em promoção e publicidade (P&P). Tem-se mesmo verificado que este tipo de investimento representa uma fatia crescente do orçamento daquelas empresas – com efeito, não basta ter um bom produto, é também necessário divulgar a sua existência junto dos potenciais clientes. Justifica-se pois que as empresas pretendam conhecer à *posteriori* a eficiência do seu esforço em P&P de forma a implementarem eventuais correcções futuras.

Em princípio, uma campanha de P&P para um determinado produto traduz-se, passado algum tempo, por um incremento das quantidades vendidas o qual se manterá por algumas semanas. O incremento cresce progressivamente, atinge um pico e decresce depois, devido ao "esquecimento" do mercado. Esta tendência pode ser contrariada ainda durante algum tempo concretizando alguns investimentos de manutenção.

Vejamos, através de um exemplo, como analisar a existência de uma relação causal entre um incremento de vendas e um esforço anterior em P&P.

Exemplo

Uma empresa lançou uma campanha publicitária há algum tempo e pretende avaliar se existe suficiente evidência estatística (para um nível de significância de 5%) de que as vendas de uma determinada linha de produtos aumentaram (ou diminuíram) em consequência.

Após uma pesquisa na base de dados, a empresa seleccionou 10 semanas de vendas antes e 10 semanas de vendas depois da campanha. As vendas verificadas durante estes períodos encontram-se representadas no próximo Quadro (ordenadas cronologicamente e ajustadas da sazonalidade e da tendência).

Vendas antes (unid./sem.)	**Vendas depois (unid./sem.)**
10	17
15	14
7	12
6	16
13	23
11	18
12	10
9	8
17	19
14	22
Média = 11,4	Média = 15,9

À primeira vista é-se tentado a concluir que as vendas aumentaram pois a média de vendas depois (15,9 unid./sem) é superior à média de vendas antes (11,4 unidades/semana). Contudo, pode não existir evidência estatística suficiente para tirar tal conclusão, devendo-se aqueles valores apenas à aleatoriedade dos dados. Neste caso, para obtermos uma resposta com um certo nível de confiança, realizamos um teste estatístico não-paramétrico – já que, à partida, desconhecemos os parâmetros das leis de distribuição daquelas duas populações – designado “método de *Wilcoxon*”.

Solução

1. Ordenam-se os valores de vendas, antes e depois do início da campanha, por ordem crescente e atribui-se-lhes o respectivo número de posição. No caso de existirem valores repetidos, a posição é calculada pela média aritmética das posições que receberiam se não fossem repetidos;

Vendas antes (unid./sem.)	**Posições**	**Vendas depois (unid./sem.)**	**Posições**
10	5,5	17	15,5
15	13	14	11,5
7	2	12	8,5
6	1	16	14
13	10	23	20
11	7	18	17
12	8,5	10	5,5
9	4	8	3
17	15,5	19	18
14	11,5	22	19
Média = 11,4	T_1 = 78	Média = 15,9	T_2 = 132

2. Se designarmos por D_1 e D_2 as distribuições de frequências relativas das vendas 1 e 2 (antes e depois da campanha, respectivamente) teremos como passos do método, os seguintes:

<u>Hipótese nula</u>: H_0: D_1 e D_2 são idênticos;

<u>Hipótese alternativa</u>: H_a: D_1 encontra-se desviada para a esquerda de D_2 ou D_1 encontra-se desviada para a direita de D_2.

<u>Estatística de teste</u>: T_1, se $n_1 < n_2$; T_2, se $n_2 < n_1$ (qualquer T pode ser usado se $n_1 = n_2$);

<u>Região de rejeição</u>: T_1: $T_1 \geq T_s$ (ou $T_1 \leq T_i$)
T_2: $T_2 \leq T_i$ (ou $T_2 \geq T_s$)

3. No caso do exemplo teremos:

   H_0: As duas distribuições de valores de venda são idênticas;
   H_a: A distribuição dos valores de venda depois encontra-se desviada para a esquerda ou para a direita da distribuição dos valores de venda antes.

4. Os valores críticos da distribuição das somas das classificações, designados por T_i (inferior) e T_s (superior), respectivamente, encontram-se na tabela seguinte (até n = 10):

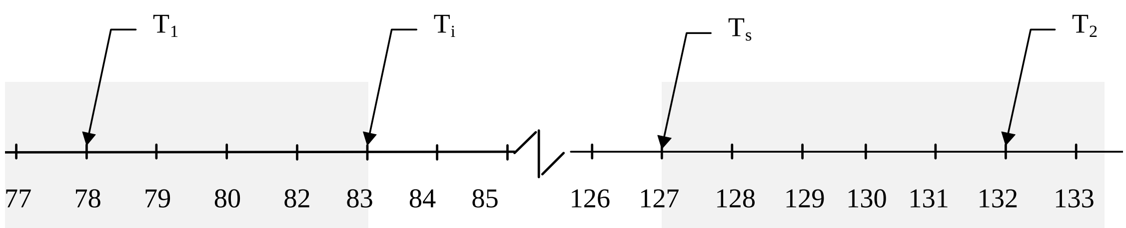
Nível de significância = 0,025

n_1 / n_2	**3**	**3**	**4**	**4**	**5**	**5**	**6**	**6**	**7**	**7**	**8**	**8**	**9**	**9**	**10**	**10**
	Ti	**Ts**	**Ti**	**Ts**	**Ti**	**Ts**	**Ti**	**Ts**	**Ti**	**Ts**	**Ti**	**Ts**	**Ti**	**Ts**	**Ti**	**Ts**
3	5	16	6	18	6	21	7	23	7	26	8	28	8	31	9	33
4	6	18	11	25	12	28	12	32	13	35	14	38	15	41	16	44
5	6	21	12	28	18	37	19	41	20	45	21	49	22	53	24	56
6	7	23	12	32	19	41	26	52	28	56	29	61	31	65	32	70
7	7	26	13	35	20	45	28	56	37	68	39	73	41	78	43	83
8	8	28	14	38	21	49	29	61	39	73	49	87	51	93	54	98
9	8	31	15	41	22	53	31	65	41	78	51	93	63	108	66	114
10	9	33	16	44	24	56	32	70	43	83	54	98	66	114	79	131

Nível de significância = 0,050

n_1 / n_2	**3**	**3**	**4**	**4**	**5**	**5**	**6**	**6**	**7**	**7**	**8**	**8**	**9**	**9**	**10**	**10**
	Ti	**Ts**	**Ti**	**Ts**	**Ti**	**Ts**	**Ti**	**Ts**	**Ti**	**Ts**	**Ti**	**Ts**	**Ti**	**Ts**	**Ti**	**Ts**
3	6	15	7	17	7	20	8	22	9	24	9	27	10	29	11	31
4	7	17	12	24	13	27	14	30	15	33	16	36	17	39	18	42
5	7	20	13	27	19	36	20	40	22	43	24	46	25	50	26	54
6	8	22	14	30	20	40	28	50	30	54	32	58	33	63	35	67
7	9	24	15	33	22	43	30	54	39	66	41	71	43	76	46	80
8	9	27	16	36	24	46	32	58	41	71	52	84	54	90	57	95
9	10	29	17	39	25	50	33	63	43	76	54	90	66	105	69	111
10	11	31	18	42	26	54	35	67	46	80	57	95	69	111	83	127

5. No caso do exemplo, temos que, para n1 = n2 = 10 e para α =0,05, T_i = 83 e T_s = 127. Podemos representar graficamente esta região de rejeições:



6. Logo, rejeitaremos H_0 se $T_2 \geq T_s$ ($T_2 \geq 127$). Como $T_2 = 132 > 127$, rejeitamos H_0 e concluímos que (para um nível de significância de 0,05) a distribuição das vendas depois da campanha se encontra localizada à direita da distribuição das vendas antes da campanha (pois a média de vendas antes 11,4 unidades/semana é inferior à média de vendas depois 15,9 unidades/semana), ou seja, **as vendas aumentaram**. Poder-se-ia chegar à mesma conclusão se tivéssemos seleccionado T_1 como estatística de teste (visto que $n_1 = n_2 = 10$). Com efeito, como $T_1 = 78 < 83$, rejeitávamos H_0.

Rui Assis
rassis@rassis.com
http://www.rassis.com