

Wilcoxon Signed Rank Test

Group 8

數據所 黃亮臻 RE6124035 統計所 張伊萱 R26121023

統計所 施文千 R26124055 企管所 陳冠瑜 R46121213















Table of content

- 1. Introduction P.3
- 2. Theory P.6
- 3. Example P.14

1. Introduction















What is Wilcoxon Signed Rank Test

- A non-parametric test
- The type of variable could be continuous or ordinal (rank)
- Situations:
 - 1. Population medium (η) vs. Constant (η_0)
 - 2. Difference between Paired samples (η_A, η_B)
- Non-parametric alternative to One sample t-test and Paired t-test









Properties of Nonparametric Test

- Distribution-Free
- Allow smaller sample size
- Less sensitive to outliers

Why don't we always use non-parametric tests?

Low statistical power





2. Theory















Hypothesis Testing

One – sample test

The one-sample Wilcoxon signed-rank test can be used to test whether data comes from a population with a specified median (η_0) .

$$\begin{cases} H_0: \eta = \eta_0 \\ H_1: \eta \neq \eta_0 \end{cases}$$







Hypothesis Testing

Paired data test

The paired data Wilcoxon signed-rank test can be used to test that is difference between two sample median(η_A and η_B).

$$\begin{cases} H_0: \eta_A = \eta_B \\ H_1: \eta_A \neq \eta_B \end{cases}$$







Test Statistic

· Calculate the differences:

If we have one sample $x_1, x_2, ..., x_n$,

we define $d_i = x_i - \eta_0$, for i = 1,2,...,n

If we have paired data (x_i, y_i) , for i = 1, 2,...,n

we define $d_i = x_i - y_i$, for i = 1,2,...,n











Test Statistic

- Take the absolute value of each difference $(|d_i|)$ and rank them from smallest to largest, define it as R_i
- Let $W^+ = \sum_{i=1}^n R_i \times I_{\{d_i > 0\}}$, $W^- = \sum_{i=1}^n R_i \times I_{\{d_i < 0\}}$

Note: I is indicator function

$$\begin{cases} I_{\{d_i>0\}} = 0, & \text{if } d_i < 0 \\ I_{\{d_i>0\}} = 1, & \text{if } d_i > 0 \end{cases} \begin{cases} I_{\{d_i<0\}} = 1, & \text{if } d_i < 0 \\ I_{\{d_i<0\}} = 0, & \text{if } d_i > 0 \end{cases}$$











Test Statistic

The test statistic is

$$W = Min\{W^+, W^-\}$$

If the sample size is large enough, the test statistic approximates a normal distribution with

$$\mu = \frac{n(n+1)}{4} \text{ and } \sigma = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

$$Z = \frac{W - E(W)}{\sqrt{Var(w)}} = \frac{W - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \sim N(0,1)$$









Level Of Significance & Rejection Region

• Level of significance(α)

We usually use $\alpha = 0.05$

Rejection region(RR)

 $RR = \left\{W < W_{\frac{\alpha}{2}}(n)\right\}$, find Wilcoxon signed-rank table

• If the sample size is large, then RR= $\left\{Z<-Z_{\frac{\alpha}{2}}\right\}$, find Standard Normal Distribution Table













Wilcoxon Signed-Rank Table

alpha values								
n	0.001	0.005	0.01	0.025	0.05	0.10	0.20	
5						0	2	
6					0	2	3	
7				0	2	3	5	
8			0	2	3	5	8	
9		0	1	3	5	8	10	
10		1	3	5	8	10	14	
11	0	3	5	8	10	13	17	
12	1	5	7	10	13	17	21	
13	2	7	9	13	17	21	26	
14	4	9	12	17	21	25	31	
15	6	12	15	20	25	30	36	
16	8	15	19	25	29	35	42	
17	11	19	23	29	34	41	48	
18	14	23	27	34	40	47	55	
19	18	27	32	39	46	53	62	
20	21	32	37	45	52	60	69	
21	25	37	42	51	58	67	77	
22	30	42	48	57	65	75	86	
23	35	48	54	64	73	83	94	
24	40	54	61	72	81	91	104	
25	45	60	68	79	89	100	113	
26	51	67	75	87	98	110	124	
27	57	74	83	96	107	119	134	

alpha values								
n	0.001	0.005	0.01	0.025	0.05	0.10	0.20	
28	64	82	91	105	116	130	145	
29	71	90	100	114	126	140	157	
30	78	98	109	124	137	151	169	
31	86	107	118	134	147	163	181	
32	94	116	128	144	159	175	194	
33	102	126	138	155	170	187	207	
34	111	136	148	167	182	200	221	
35	120	146	159	178	195	213	235	
36	130	157	171	191	208	227	250	
37	140	168	182	203	221	241	265	
38	150	180	194	216	235	256	281	
39	161	192	207	230	249	271	297	
40	172	204	220	244	264	286	313	
41	183	217	233	258	279	302	330	
42	195	230	247	273	294	319	348	
43	207	244	261	288	310	336	365	
44	220	258	276	303	327	353	384	
45	233	272	291	319	343	371	402	
46	246	287	307	336	361	389	422	
47	260	302	322	353	378	407	441	
48	274	318	339	370	396	426	462	
49	289	334	355	388	415	446	482	
50	304	350	373	406	434	466	503	









3. Example





















Example

Example 5.1.1. In computer memory manufacturing, chips that come off the assembly line can actually have widely varying properties, even from the same piece of silicon. To counter for this, manufacturer will measure the properties of the chips and bin them (as in product binning) to choose the product line best suits the measured value of some specification. In some application, for the example of running RAM chips in dual- or quad-channel architecture for better performance, it is important to have RAM's running at matching speeds. From one of the production line, consecutive two DDR4 RAM's are sampled to measure their memory clock speed (MHz). Tests made on 13 randomly selected pairs of RAM's are reported in the table below, and the production manager are wondering if the overall process can produce consecutive RAM sticks in pairs of two with identical memory clock speed before further analysis. Denote by X_i the clock speed of the first component, Y_i of the second of pair i, and D_i the absolute difference of $Y_i - X_i$. Is the production making pairs of RAM's with matching clock speeds?

Table 5.1: Memory clock speed measured in MHz

Pair i	1	2	3	4	5	6	7	8	9
X_i	300.02	299.99	300.00	300.00	300.01	300.01	299.99	299.98	299.99
Y_{i}	299.97	299.99	299.99	300.04	299.98	299.95	300.03	300.02	299.99
$ Y_i - X_i $	0.05	0.00	0.01	0.04	0.03	0.06	0.04	0.04	0.00
Pair i	10	11	12	13					
X_{i}	299.99	299.98	299.97	300.01					
Y_{i}	300.01	300.00	299.98	300.00					
$ Y_i - X_i $	0.02	0.02	0.01	0.01					

Example from applied nonparamatric statistics wayne W.Daniel















Pair i	Xi	Yi	di = Xi – Yi	Sign	Rank	
1	300.02	299.97	0.05	+	10	
2	299.99 299.99 0		0			
3	300.00	299.99	0.01	+	2	
4	300.00	300.04	-0.04	_		8
5	300.01	299.98	0.03	+	6	
6	300.01	299.95	0.06	+	11	
7	299.99	300.03	-0.04	_		8
8	299.98	300.02	-0.04	_		8
9	299.99 299.99 0					
10	299.99	300.01	-0.02	_		4.5
11	299.98	300.00	-0.02	_		4.5
12	299.97	299.98	-0.01	_		2
13	300.01	300.00	0.01	+	2	
	Sum					35

$$\begin{cases} H_0: \eta_X = \eta_Y \\ H_1: \eta_X \neq \eta_Y \end{cases}$$

By Wilcoxon table W = min{ |31|, |35| }

By Z table $E(W) = \frac{13*14}{4} = 45.5$

$$Var(W) = \frac{13*14*27}{24} = 204.75$$

$$Z = \frac{(31-14)}{\sqrt{204.75}} = 1.188$$

$$P$$
-value = 0.119















alpha values									
n	0.001	0.005	0.01	0.025	0.05	0.10	0.20		
5						0	2		
6					0	2	3		
7				0	2	3	5		
8			0	2	3	5	8		
9		0	1	3	5	8	10		
10		1	3	5	8	10	14		
11	0	3	5	8	10	13	17		
12	1	5	7	10	13	17	21		
13	2	7	9	13	17	21	26		
14	4	9	12	17	21	25	31		
15	6	12	15	20	25	30	36		
16	8	15	19	25	29	35	42		
17	11	19	23	29	34	41	48		

$$W = min\{ |31|, |-35| \} = 31$$

Use library MASS

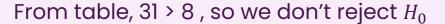
Wilcoxon signed rank test with continuity correction

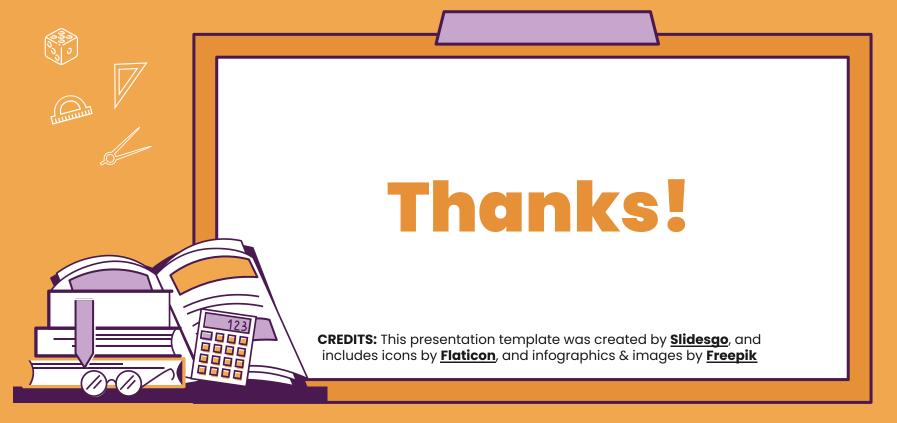
data: df\$X and df\$Y V = 31, p-value = 0.8936

alternative hypothesis: true location shift is not equal to \emptyset

















Introduction: 黃亮臻

Theory: 張伊萱

Example:施文千

PPT 彙整: 陳冠瑜、施文千



