Ch 19 實習(1)

Agenda

- Nonparametric statistic 使用時機
- Wilcoxon Rank Sum Test
- Sign Test
- Wilcoxon Signed Rank Sum Test
- Kruskal-Wallis Test
- Friedman Test
- Spearman Rank Correlation Coefficient

1. Nonparametric statistic 使用時機

資料型態	Interval	Nominal	Ordinary
檢定方式	T Z ANOVA	Chi square goodness of fit 列連表	無母數
常態分配	Yes	Check	No

1. Nonparametric statistic 使用時機

有母數	無母數
Independent T	• Wilcoxon Rank Sum Test • Ordinal or Interval (非常態) • N>20 才會趨近常態 (或個別樣本都要>10)
Match Paired T	• Sign Test • Ordinal or Interval (非常態) • N>20 才會趨近常態 (或 Nd>10對）
Match Paired T	• Wilcoxon Signed Rank Sum Test • Interval (非常態) • N>30才會趨近常態 (或 Nd>15對）
One way ANOVA	• Kruskal-Wallis Test • Ordinal or Interval (非常態)
Random Block design	• Friedman Test • Ordinal or Interval (非常態)
Correlation	• Spearman Rank Correlation Coefficient • Ordinal or Interval (非常態)

1. 檢驗步驟

- S1: 先檢驗是否為常態分配（chi-square goodness of fit)
- S2: 非無母數，用無母數檢定（或 ordinal直接用無母數）
- S3: 設定假設
- S4: 找critical point
- S5: 求檢定量 T （or Z）
- S 6:結論

H_0: The population locations are the same （Ho: u1=u2)

H_1: (i) The locations differ, （Ho: u1=u2)

(ii) Population 1 is located to the right of population 2（Ho: u1>u2)

(ii) Population 1 is located to the left of population 2 （Ho: u1<u2)

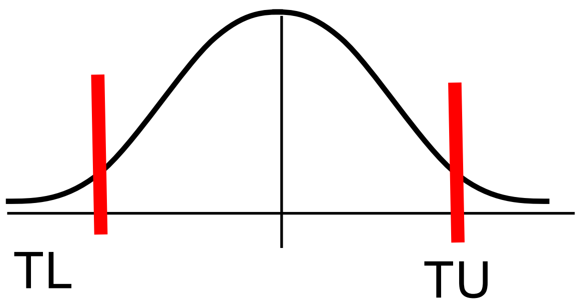
2. Wilcoxon Rank Sum Test (Mann- Whitney Test)

- The problem characteristics of this test are:
  - The problem objective is to compare two populations.
  - The data are either ordinal or interval (but not normal).
  - The samples are independent.
  - 類似獨立樣本T檢定

1. Wilcoxon Rank Sum Test

- (N各別<10)
- S1:設定假設
- H_0: The two population locations are the same

  H_1: The location of population 1 is different from the location of population 2
- S2: 求critical point
  - 查表$P(T \leqq T_L) = P(T \geqq T_U)$
- S3:算統計量 T
- S4:結論



- (N各別>10)
- S1:設定假設
- H_0: The two population locations are the same

  H_1: The location of population 1 is different from the location of population 2
- S2: 求critical point
  - 查表 Za/2, Za
- S3:算統計量 Z

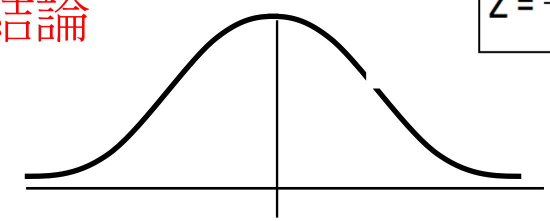
$$E(T) = \frac{n_1(n_1 + n_2 + 1)}{2}$$

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

Therefore,

$$Z = \frac{T - E(T)}{\sigma_T}$$

- S4:結論



Example 1

- Based on the two samples shown below, can we infer at 5% significance level that the location of population 1 is to the left of the location of population 2? (類似H0: u1<u2, 左尾）

  - **Sample 1:** 22, 23, 20
  - **Sample 2:** 18, 27, 26

Solution 1

- S1:
  - H_0: The two population locations are the same.

    H_1: The location of population 1 is to the left of the location of population 2.
- S2:查表 T_L =6, T_U= 15
- S3計算T(由小到大排序，求T1)

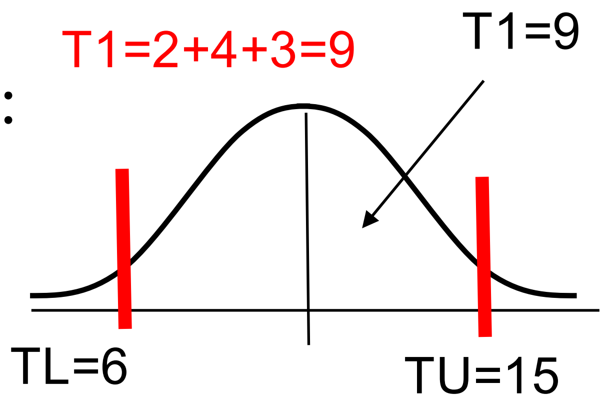
Sample 1	Rank
22	3
23	4
20	2

Sample 2	Rank
18	1
27	6
26	5

T1=2+4+3=9

T2=1+6+5=12

- 結論：



Don’t reject H0.
表示兩樣本的location沒有差異

1. Critical values of the Wilcoxon Rank Sum Test

α = .025 for one tail test, or α = .05 for two tail test

n2	n1						
	3	**4**	**5**	**.**	**.**	**.**	**10**
3	6 15						
4	6 18	11 25	17 33	.	.	.	61 89
5	6 21	12 28	18 37	.	.	.	64 96
.	T_L T_U	T_L T_U	T_L T_U				
.							
.							
10	9 33	16 44	24 56	.	.	.	79 131

T_L T_U

Using the table: For given two samples of sizes n_1 and n_2, $P(T \leqq T_L)=P(T \geqq T_U)=\alpha$.

A similar table exists for α = .05 (one tail test) and α = .10 (two tail test)

Example 2 (N<10)

- Use the Wilcoxon rank sum test on the following data to determine the two population locations differ. (Use a 10% significance level.)
- Sample 1: 15 7 22 20 32 18 26 17 23 30 (n=10)
- Sample 2: 8 27 17 25 20 16 21 17 10 18 (n=10)

Solution 2

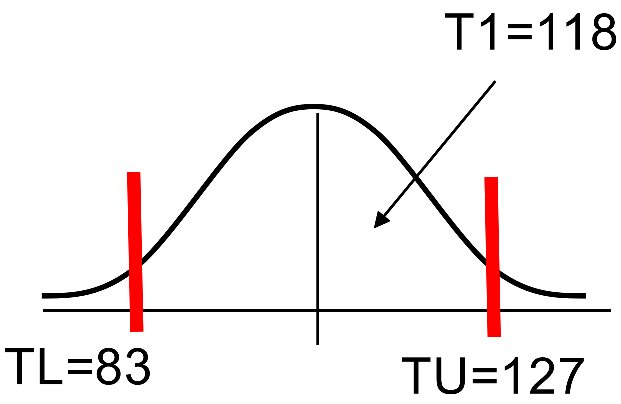
- H_0: The two population locations are the same

  H_1: The location of population 1 is different from the location of population 2

Rejection region: $T \geq T_U = 127$ or $T \leq T_L = 83$

Sample 1	Rank	Sample 2	Rank
15	4.0	8	2.0
7	1.0	27	18.0
22	14.0	17	7.0
20	.5	25	16.0
32	20.0	20	11.5
18	9.5	16	5.0
26	17.0	21	13.0
17	7.0	17	7.0
23	15.0	10	3.0
30	19.0	18	9.5
	$T_1 = 118$		$T_2 = 92$

Sample	Rank
7	1
8	2
10	3
15	4
16	5
17	(6+7+8)/3=7
18	(9+10)/2=9.5



There is not enough evidence to infer that the location of population 1 is different from the location of population 2

TABLE 9 Critical Values for the Wilcoxon Rank Sum Test

(a) $\alpha = .025$ one-tail; $\alpha = .05$ two-tail

n_2 \ n_1	3		4		5		6		7		8		9		10	
	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U
4	6	18	11	25	17	33	23	43	31	53	40	64	50	76	61	89
5	6	11	12	28	18	37	25	47	33	58	42	70	52	83	64	96
6	7	23	12	32	19	41	26	52	35	63	44	76	55	89	66	104
7	7	26	13	35	20	45	28	56	37	68	47	81	58	95	70	110
8	8	28	14	38	21	49	29	61	39	63	49	87	60	102	73	117
9	8	31	15	41	22	53	31	65	41	78	51	93	63	108	76	124
10	9	33	16	44	24	56	32	70	43	83	54	98	66	114	79	131

(b) $\alpha = .05$ one-tail; $\alpha = .10$ two-tail

n_2 \ n_1	3		4		5		6		7		8		9		10	
	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U
3	6	15	11	21	16	29	23	37	31	46	39	57	49	68	60	80
4	7	17	12	24	18	32	25	41	33	51	42	62	52	74	63	87
5	7	20	13	27	19	37	26	46	35	56	45	67	55	80	66	94
6	8	22	14	30	20	40	28	50	37	61	47	73	57	87	69	101
7	9	24	15	33	22	43	30	54	39	66	49	79	60	93	73	107
8	9	27	16	36	24	46	32	58	41	71	52	84	63	99	76	114
9	10	29	17	39	25	50	33	63	43	76	54	90	66	105	79	121
10	11	31	18	42	26	54	35	67	46	80	57	95	69	111	83	127

...d R. A. Wilcox, "Some Rapid Approximate Statistical Procedures" (1964), p. 28. Reproduced with the permission of American Cyanamid Company.

Example 3 (N>10)

- Given the following statistics, calculate the value of the test statistic to determine whether the population locations differ. In addition, calculate the P-value. (a=0.05)
- $T_1=250 \ \ n_1=15$
- $T_2=215 \ \ n_2=15$

Solution 3

- S1:
  - H_0: The two population locations are the same
    H_1: The location of population 1 is different from the location of population 2
- S2:
  - $Z_{a/2}=1.96$
- S3:

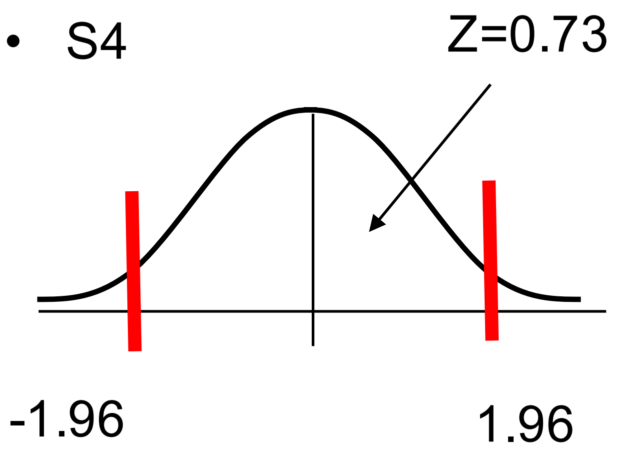
$$E(T) = \frac{n_1(n_1+n_2+1)}{2} = \frac{15(15+15+1)}{2} = 232.5$$

$$\sigma_T = \sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}} = \sqrt{\frac{(15)(15)(15+15+1)}{12}} = 24.11$$

$$z = \frac{T-E(T)}{\sigma_T} = \frac{250-232.5}{24.11} = 0.73$$

$$\text{p-value} = 2P(Z > .73) = 2(.5 - .2673) = .4654.$$

- S4



2. Sign Test

- The objective is to compare two populations.
- The data are either ordinal or interval (but not normal).
- The samples are matched by pairs（類似 paired T test).

2. The Sign Test

- S1: 假設檢定

H_0: The two population locations are the same

H_1: The two population locations are different

H_0: $p = .5$

H_1: $p \neq .5$

- S2:Critical point: Za or Za/2

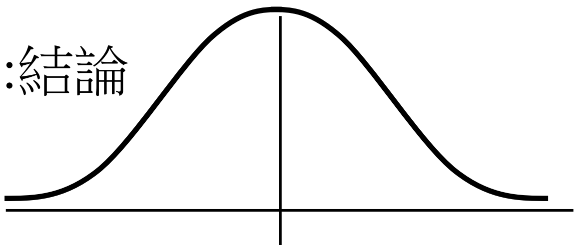
- S3:

$$z = \frac{x - 0.5n}{0.5\sqrt{n}}$$

$$where\ n \geq 10.$$

X: 有幾的正號
n:樣本數
（但不包含相減為0)

- S4:結論



The binomial variable can be approximated by a normal variable if np and n(1-p) >= 5.
The Z- statistic becomes

$$z = \frac{x - np}{\sqrt{np(1-p)}} = \frac{x - .5n}{\sqrt{n(.5)(.5)}} = \frac{x - .5n}{.5\sqrt{n}}$$

$$\text{where } n \geq 10.$$

Example 4

- 兩位評審員對12個參賽選美的候選人進行評分，其評分係依主觀偏好給予0-10分；而評分結果如下(α=5%)
- 評審員I 5 6 10 7 0 9 7 10 9 6 9 9
- 評審員II 4 1 7 5 8 5 5 6 8 10 5 4
- 試以sign test 來檢定兩位評審員，對12為參選者的評分是否有顯著差異

Solution 4

$$z = \frac{x - 0.5n}{0.5\sqrt{n}}$$

$where\ n \geq 10.$

X: 有幾的正號
N:樣本數
（但不包含相減為0)

- S1:
  - H_0: 兩位評審員評分沒有差異(p=0.5)
    H_1: 兩位評審員評分有差異(p≠0.5)
- S2: 拒絕域:z>1.96 or z<-1.96
- S3:計算 Z
  - 評審員I 5 6 10 7 0 9 7 10 9 6 9 9
  - 評審員II 4 1 7 5 8 5 5 6 8 10 5 4

    ＋ ＋ ＋ ＋ - ＋ ＋＋ ＋ - ＋ ＋
  - 10個+,2個-,n=12(沒有為0)

$$z = \frac{10 - 12 \times 0.5}{\sqrt{12 \times 0.5 \times 0.5}} = 2.3094 > 1.96$$

- S4: 拒絕H_0,所以兩位評審員評分有差異

Example 5

- Suppose that in a matched pairs experiment, we find 28 positive differences, 7 zero differences, and 41 negative differences. Can we infer at the 10% significance level that the location of population 1 is to the left of the location 2?

Solution 5

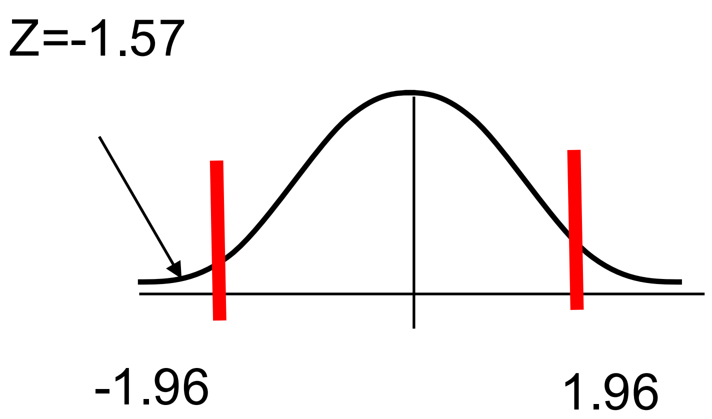
- H_0: The two population locations are the same

  H_1: The location of population 1 is to the left of the location of population 2 (左尾）

Rejection region: $z < -z_\alpha = -z_{.10} = -1.28$

$z = \frac{x - .5n}{.5\sqrt{n}} = \frac{28 - .5(69)}{.5\sqrt{69}} = -1.57$, p-value = $P(Z < -1.57) = .0582$. There is enough evidence to infer that the location of population 1 is to the left of the location of population 2.

+:28
-:41
0:7



3. Wilcoxon Signed Rank Sum Test

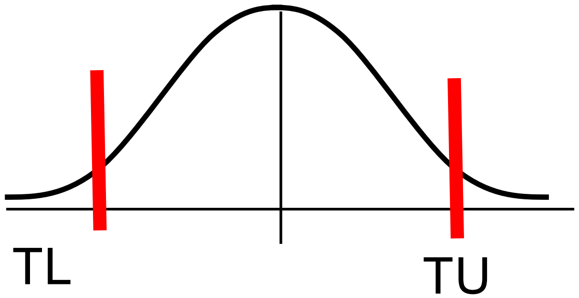
- This test is used when
  - the problem objective is to compare two populations,
  - **the data are interval but not normal**
  - the samples are matched pairs.

- The test statistic and sampling distribution
  - T is based on rank sum of the absolute values of the positive and negative differences
    - When n <=30, reject H_0 if $T>T_U$ or $T<T_L$(T_L and T_U tabulated values related to n, n is the number of nonzero).
    - When n > 30, T is approximately normally distributed. Use a Z-test.
    - E(T) = n(n+1)/4 standard deviation = [n(n+1)(2n+1)/24]^(1/2)

1. Wilcoxon Signed Rank Sum Test

- (N<30)
- S1:設定假設
- H_0: The two population locations are the same

  H_1: The location of population 1 is different from the location of population 2
- S2: 求critical point
  - 查表$P(T \leqq T_L) = P(T \geqq T_U)$
- S3:算統計量 T
- S4:結論

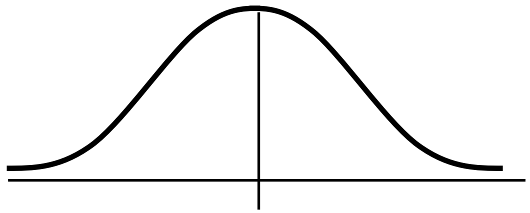


- (N>30)
- S1:設定假設
- H_0: The two population locations are the same

  H_1: The location of population 1 is different from the location of population 2
- S2: 求critical point
  - 查表 Za/2, Za
- S3:算統計量 Z
  - $E(T) = n(n+1)/4$
  - $\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$

$$Z = \frac{T - E(T)}{\sigma_T}$$

- S4:結論



Example 5 (N<30)

- Perform the Wilcoxon signed rank sum test to determine whether the location of population 1 differs from the location of population 2 given the data shown here. (Use α=.05)

Pair	1	2	3	4	5	6
Sample1	18.2	14.1	24.5	11.9	9.5	12.1
Sample2	18.2	14.1	23.6	12.1	9.5	11.3
Pair	7	8	9	10	11	12
Sample1	10.9	16.7	19.6	8.4	21.7	23.4
Sample2	9.7	17.6	19.4	8.1	21.9	21.6

Solution 5

- H_0: The two population locations are the same

  H_1: The location of population 1 is different from the location of population 2
- Reject region: TL<14; TU>64

Pair	Sample 1	Sample 2	Difference	|Difference|	Ranks	
1	18.2	18.2	0	0		
2	14.1	14.1	0	0		
3	24.5	23.6	.9	.9	6.5	
4	11.9	12.1	–.2	.2		2
5	9.5	9.5	0	0		
6	12.1	11.3	.8	.8	5	
7	10.9	9.7	1.2	1.2	8	
8	16.7	17.6	–.9	.9		6.5
9	19.6	19.4	.2	.2	2	
10	8.4	8.1	.3	.3	4	
11	21.7	21.9	–.2	.2		2
12	23.4	21.6	1.8	1.8	9	

(1+2+3)/3=2

(6+7)/2=6.5

$T^+ = 34.5$ $T^- = 10.5$

T = 34.5. There is not enough evidence to conclude that the population locations differ.

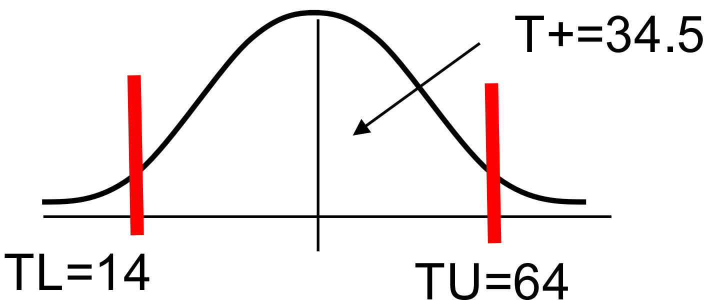


TABLE **10**
Critical Values for Wilcoxon Signed Rank Sum Test

	(a) $\alpha = .025$ one-tail; $\alpha = .05$ two-tail		(b) $\alpha = .05$ one-tail; $\alpha = .10$ two-tail	
n	T_L	T_U	T_L	T_U
6	1	20	2	19
7	2	26	4	24
8	4	32	6	30
9	6	39	8	37
10	8	47	11	44
11	11	55	14	52
12	14	64	17	61
13	17	74	21	70
14	21	84	26	79
15	25	95	30	90
16	30	106	36	100
17	35	118	41	112
18	40	131	47	124
19	46	144	54	136
20	52	158	60	150
21	59	172	68	163
22	66	187	75	178
23	73	203	83	193
24	81	219	92	208
25	90	235	101	224
26	98	253	110	241
27	107	271	120	258
28	117	289	130	276
29	127	308	141	294
30	137	328	152	313

...Statistical Procedures" (1964), p. 28. Reproduced with the permission of

Example 6 (N>30)

- A matched pairs experiment produced the following statistics. Conduct a Wilcoxon signed rank sum test to determine whether the location of population 1 is to the right of the location of population 2.(Use α=.01)

$T^{+}=3457 \qquad T^{-}=2429 \qquad n=108$

Solution 6

- H_0: The two population locations are the same

  H_1: The location of population 1 is to the right of the location of population 2 （右尾）

$T^+ = 3457 \quad T^- = 2429 \quad n=108$

Rejection region: $z > z_\alpha = z_{.01} = 2.33$

$$E(T) = \frac{n(n+1)}{4} = \frac{108(109)}{4} = 2943\,;\ \sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{108(109)(217)}{24}} = 326.25$$

$z = \frac{T - E(T)}{\sigma_T} = \frac{3457 - 2943}{326.25} = 1.58$, p-value $= P(Z > 1.58) = 1 - .9429 = .0571$. There is not enough evidence to conclude that population 1 is located to the right of the location of population 2.

4. Kruskal-Wallis Test

- The problem characteristics for this test are:
  - The problem objective is to compare two or more populations. (比較兩個以上樣本）
  - The data are either ordinal or interval but not normal.
  - The samples are independent.
- The hypotheses are

H_0: The location of all the k populations are the same.

H_1: At least two population locations differ.

（一定是檢測K組相不相似，因為沒檢測誰比較大）

4. Kruskal-Wallis Test

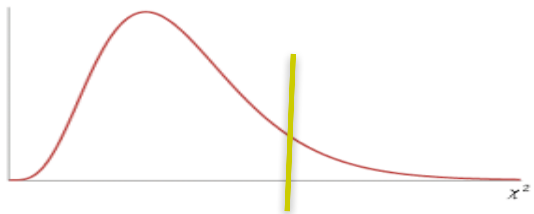
- S1: 假設檢定
  - H_0: The location of the 3 are the same
  - H_1: At least two locations are different
- S2:Critical point:

$$H > \chi^2_{\alpha,k-1}$$

- S3:計算H 值
  - Rank the data from 1 (smallest) to n (largest).
  - Calculate the rank sums $T_1, T_2, \ldots T_k$ for all the k samples.

$$H = \left[\frac{12}{n(n+1)}\sum_{j=1}^{k}\frac{T_j^2}{n_j}\right] - 3(n+1)$$

- S4:結論



Example 7

Use the Kruskal-Wallis Test on the following data to determine whether the population locations differ. (Use a=0.05)

Sample1	27	33	18	29	41	52	75
Sample2	37	12	17	22	30		
Sample3	19	12	33	41	28	18	

Solution 7

19.53 H_0 : The locations of all 3 populations are the same.

H_1 : At least two population locations differ.

Rejection region: $H > \chi^2_{\alpha,k-1} = \chi^2_{.05,2} = 5.99$

先找到最小值：12
"12": (1+2)/2=1.5
"17": 3
"18": (4+5)/2=4.5

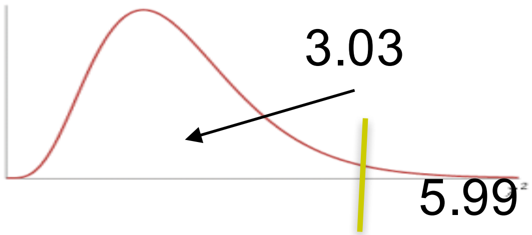
1	Rank	2	Rank	3	Rank
27	8	37	14	19	6
33	12.5	12	1.5	12	1.5
18	4.5	17	3	33	12.5
29	10	22	7	41	15.5
41	15.5	30	11	28	9
52	17			18	4.5
75	18				

$T_1 = 85.5$ $T_2 = 36.5$ $T_3 = 49$

$$H = \left[\frac{12}{n(n+1)}\sum_{j=1}^{k}\frac{T_j^2}{n_j}\right] - 3(n+1)$$

$$H = \left[\frac{12}{n(n+1)}\sum\frac{T_j^2}{n_j}\right] - 3(n+1) = \left[\frac{12}{18(18+1)}\left(\frac{85.5^2}{7}+\frac{36.5^2}{5}+\frac{49^2}{6}\right)\right] - 3(18+1) = 3.03, \text{ p-value} =$$

.2195. There is no evidence to conclude that at least two population locations differ.



無母數優缺點

- 無母數統計分析(Non-parametric Statistics)
  - 母體分配未知， 且樣本數不是很大
  - 優點：
    - 對母體的假設少， 不需要假設母體是什麼分配
    - 對小樣本的資料， 或是有偏斜分配的母體做推論比較合適
    - 可以分析順序資料
  - 缺點：
    - 檢定力(1-β)較弱
    - 對某些較複雜的模式如有交互作用的多因子設計無法做檢定
    - 處理方式不一(很難計算)