



# 熱區分析 多重檢定校正

空間分析 2019.06.03  
TA 杜承軒

多重檢定 (multiple testing) → 提高型一錯誤

		判斷結果	
		拒絕 $H_0$	接受 $H_0$
真實狀況	$H_0$ 為真 (不應拒絕 $H_0$ )	<div>型一錯誤</div>	<div>正確</div>
	$H_0$ 為假 (應拒絕 $H_0$ )	<div>正確</div>	<div>型二錯誤</div>

$\alpha$ ：顯著水準，容許型一錯誤發生的機率上限

$p$ ：犯型一錯誤的機率

## 校正

### Bonferroni

把整個實驗的錯誤率修正回 $\alpha$   
→ 太過嚴謹

### Sidak

希望控制整個實驗的錯誤率(和Bonferroni一樣)  
→ 假設完全不能有型一錯誤, 仍太過嚴苛

### FDR

錯誤發現率(False Discovery Rate)

1. 由小到大排序 $m$ 個 $p$ -value ( $p_1 \leq p_2 \leq \dots \leq p_m$ )
2. 從 $p_m$ 開始找到第一個 $i$ , 使得 $p_i \leq p_{critical} = \frac{i}{m} \alpha = p_{FDR}$   
比 $i$ 還小的實驗都視為顯著

做了 $m$ 次實驗

$$\tilde{p}_i = \min\{p_i \times m, 1\}$$

單一實驗： $\alpha_c$       v.s.      整個實驗： $\alpha_e$

$$\alpha_e = 1 - (1 - \alpha_c)^m \xrightarrow{\text{移項}} \alpha_c = 1 - (1 - \alpha_e)^{\frac{1}{m}}$$

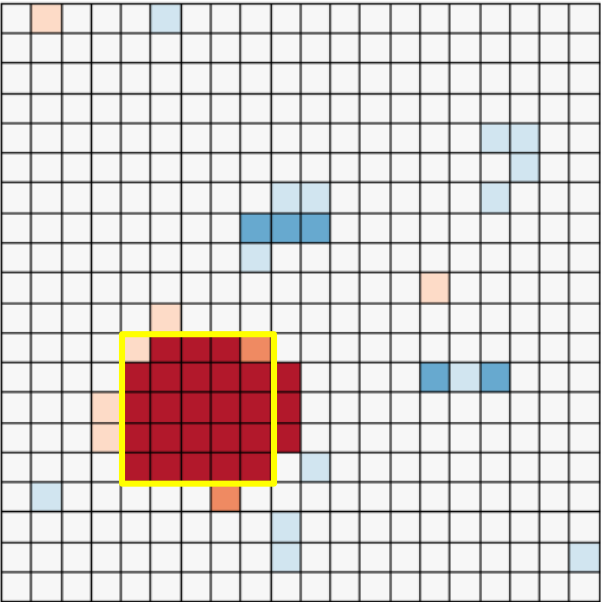
$$\tilde{p}_i = \min\{1 - (1 - p_i)^m, 1\}$$

$$\tilde{p}_i = \min\{\frac{m}{i} p_i, 1\}$$

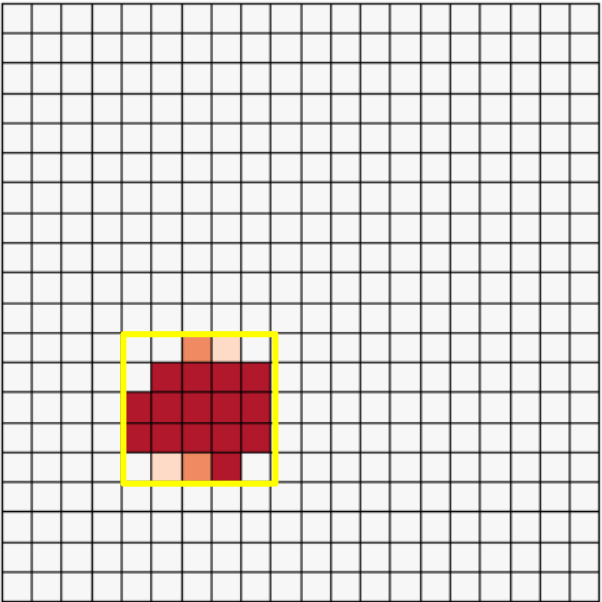
校正比較

模擬20×20的隨機網格， 內含5×5的熱區

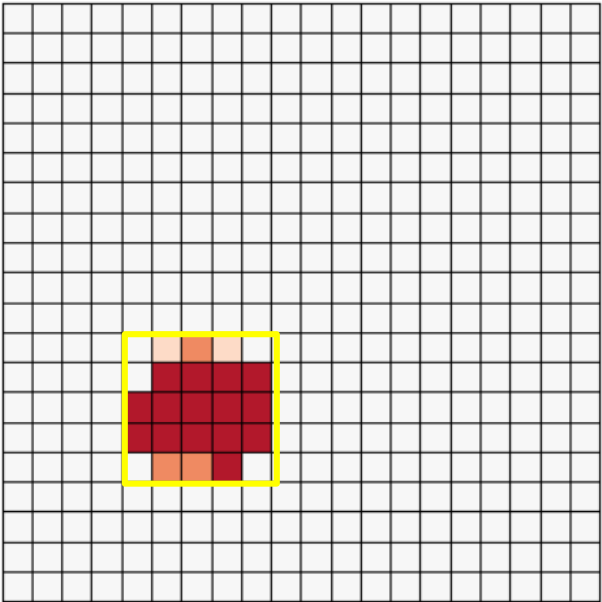
Gi\* 原始



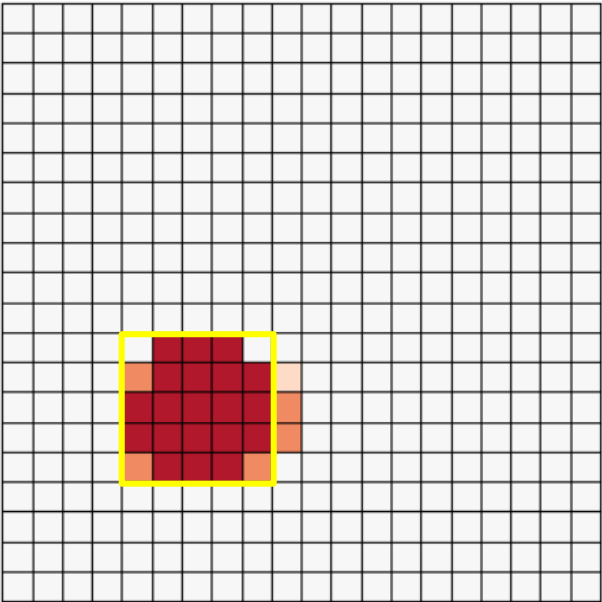
Bonferroni 校正



Sidak 校正



FDR 校正



Hotspots

90%	95%	99%
5	2	26

Hotspots

90%	95%	99%
1	2	15

Hotspots

90%	95%	99%
2	3	15

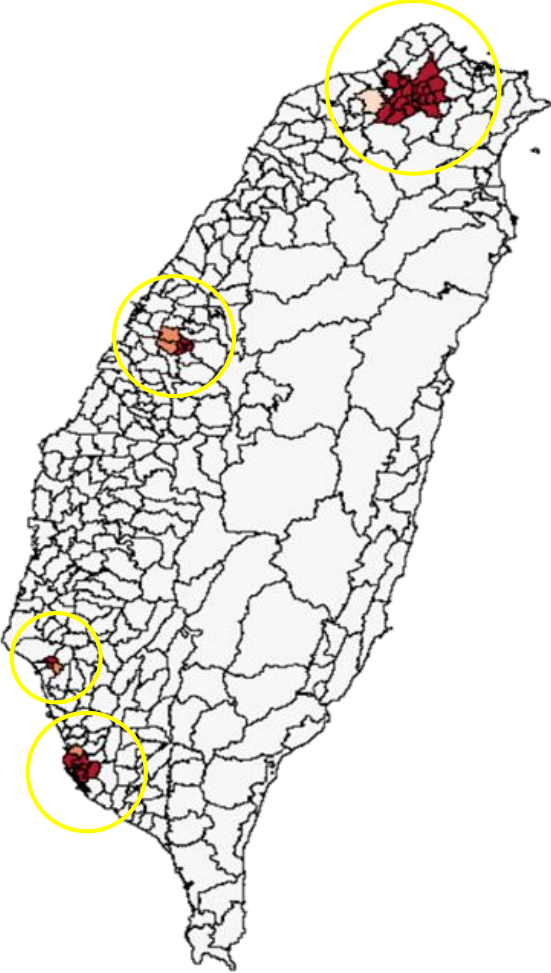
Hotspots

90%	95%	99%
1	5	20

校正比較

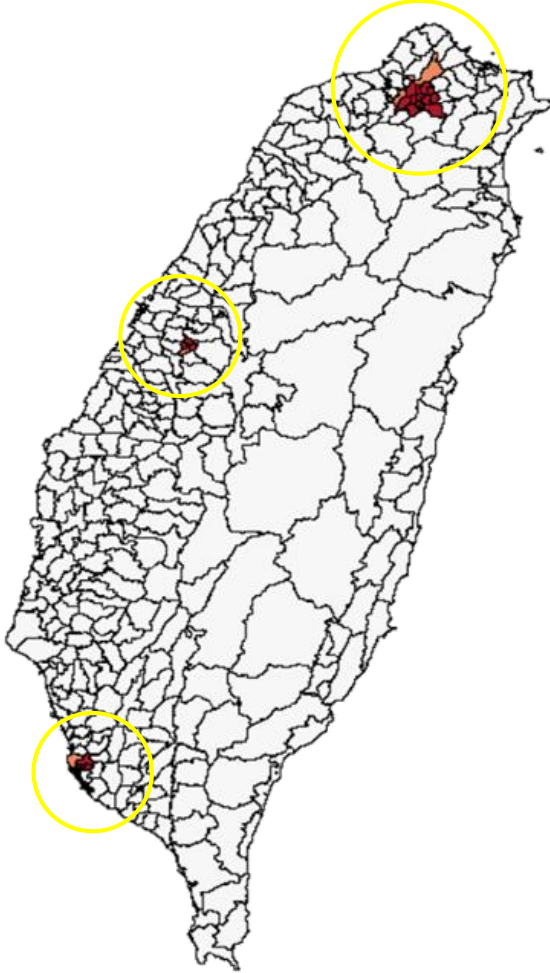
鄉鎮人口密度的空間自相關

Gi\* 原始



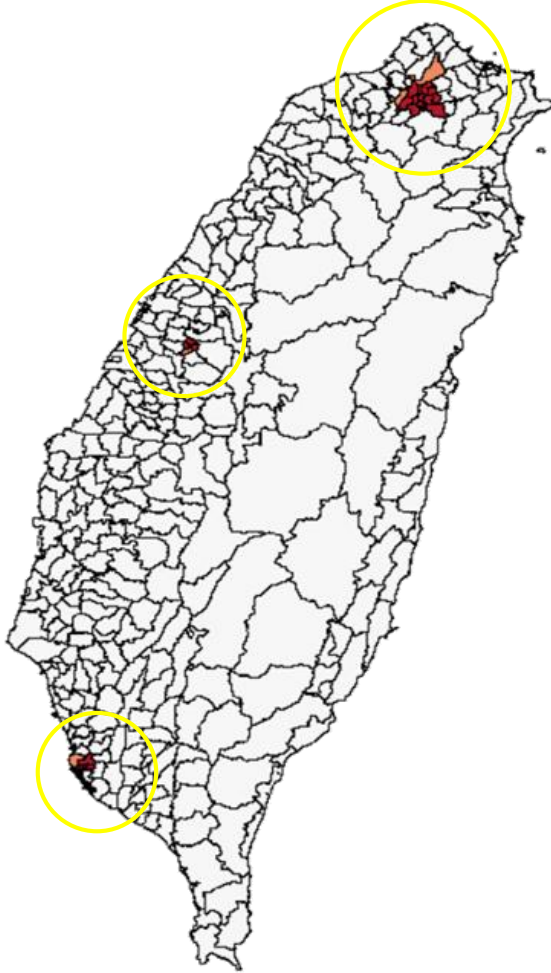
90%	95%	99%
1	4	35

Bonferroni 校正



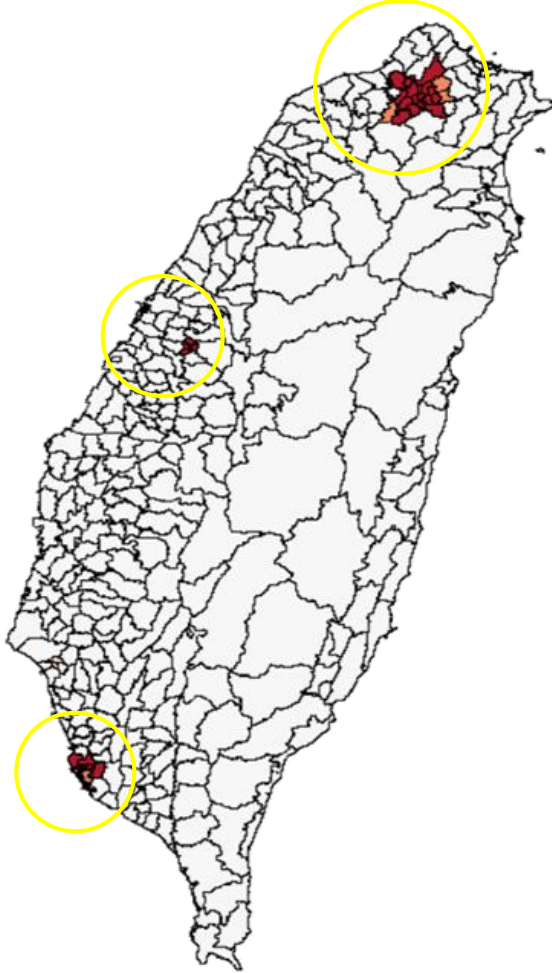
90%	95%	99%
0	4	21

Sidak 校正



90%	95%	99%
0	4	21

FDR 校正



90%	95%	99%
2	4	29



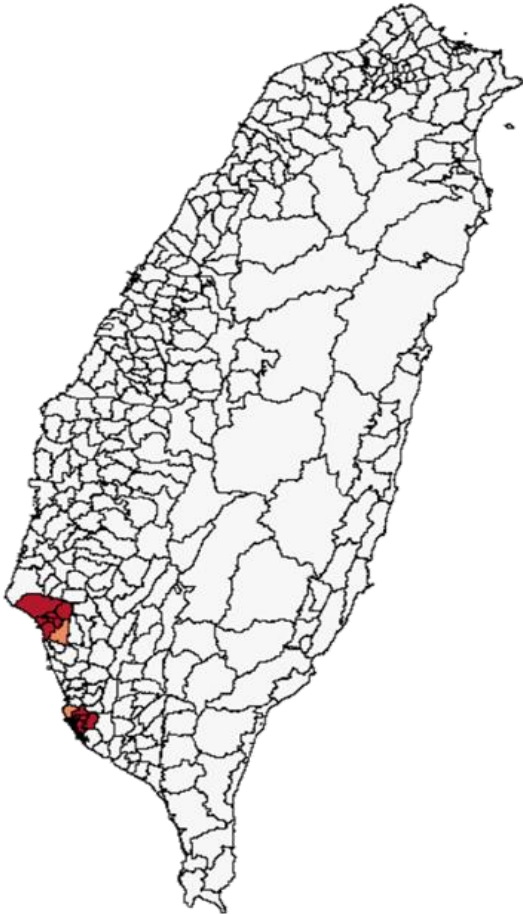
校正比較

2015年登革熱疫情：南高屏鄉鎮登革熱發生密度的空間自相關

Gi\* 原始



Bonferroni 校正



Sidak 校正



FDR 校正



90%	95%	99%
0	2	18

90%	95%	99%
0	2	14

90%	95%	99%
0	2	14

90%	95%	99%
1	1	16

校正比較

2015年登革熱疫情：台灣鄉鎮登革熱發生密度的空間自相關

Gi\* 原始



90%	95%	99%
2	1	13

Bonferroni 校正



90%	95%	99%
0	3	6

Sidak 校正



90%	95%	99%
0	3	6

FDR 校正



90%	95%	99%
1	3	9

校正步驟

假設共有100個樣本， $\alpha = 0.05$

方法一

方法二

Bonferroni

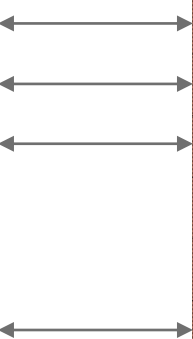
$p_i$ 和 $\tilde{\alpha}_i$ 比較

$\tilde{p}_i$ 和 $\alpha_i$ 比較

$$\tilde{\alpha} = \frac{\alpha}{100} = 0.0005$$

$$\tilde{p}_i = p \times 100$$

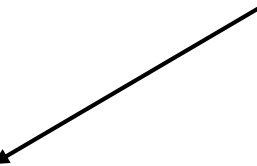
排序	原始p-val
1	0.00001
2	0.00002
3	0.00005
4	0.0001
5	0.0002
6	0.0005
7	0.001
8	0.002
9	0.005
10	0.01
11	0.02
12	0.05
13	0.1
14	0.2
.....	.....



修正後alpha
0.0005
0.0005
0.0005
0.0005
0.0005
0.0005
0.0005
0.0005
0.0005
0.0005
0.0005
0.0005
0.0005
0.0005
.....

校正後p-val
0.001
0.002
0.005
0.01
0.02
0.05
0.1
0.2
0.5
1
1
1
1
1
.....

直接和原始的0.05比較



p-val不會超過1



校正步驟

假設共有100個樣本， $\alpha = 0.05$

方法一

$p_i$ 和 $\tilde{\alpha}_i$ 比較

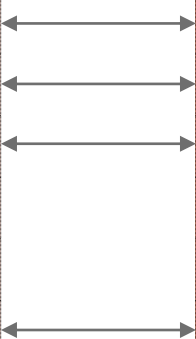
$$\tilde{\alpha} = 1 - (1 - \alpha)^{\frac{1}{100}}$$

方法二

$\tilde{p}_i$ 和 $\alpha_i$ 比較

$$\tilde{p}_i = 1 - (1 - p_i)^{100}$$

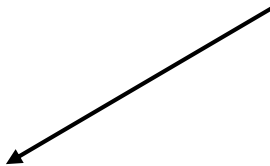
排序	原始p-val
1	0.00001
2	0.00002
3	0.00005
4	0.0001
5	0.0002
6	0.0005
7	0.001
8	0.002
9	0.005
10	0.01
11	0.02
12	0.05
13	0.1
14	0.2
.....	.....



修正後alpha
0.0005128
0.0005128
0.0005128
0.0005128
0.0005128
0.0005128
0.0005128
0.0005128
0.0005128
0.0005128
0.0005128
0.0005128
0.0005128
0.0005128
.....

校正後p-val
0.001
0.001998
0.004988
0.009951
0.019803
0.048782
0.095208
0.181433
0.39423
0.633968
0.86738
0.994079
0.999973
1
.....

直接和原始的0.05比較



校正步驟

假設共有100個樣本， $\alpha = 0.05$

FDR

方法一

$p_i$ 和 $\tilde{\alpha}_i$ 比較

$$\tilde{\alpha} = \alpha \times \frac{i}{100}$$

排序(i)	原始p-val	修正後alpha
1	0.00001	0.0005
2	0.00002	0.001
3	0.00005	0.0015
4	0.0001	0.002
5	0.0002	0.0025
6	0.0005	0.003
7	0.001	0.0035
8	0.002	0.004
9	0.005	0.0045
10	0.01	0.005
11	0.02	0.0055
12	0.05	0.006
13	0.1	0.0065
14	0.2	0.007
.....	.....	.....

方法二

$\tilde{p}_i$ 和 $\alpha_i$ 比較

$$\tilde{p}_i = p \times \frac{100}{i}$$

校正後p-val
0.001
0.001
0.001667
0.0025
0.004
0.008333
0.014286
0.025
0.055556
0.1
0.181818
0.416667
0.769231
1
.....

直接和原始的0.05比較

p-val不會超過1

## 整體p值校正

`p.adjust( )`

*based on the total number*

(即前面校正步驟的方法二)

## 考慮鄰近 隨鄰居數調整

`localmoran(..., p.adjust.method='...')`

`p.adjustSP( )`

*based on the number of neighbors (+1)*