



熱區分析 多重檢定校正

空間分析 2020.06.08
TA 杜承軒

多重檢定

多重檢定 (multiple testing) → 提高型一錯誤

--	--	--	--	--	--

兩兩獨立

判斷結果

		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>

α ：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

校正方法

* 做了 m 次實驗

校正後 p-value

概念

單一實驗： $\tilde{\alpha}_i$ vs. 整個實驗： α

Bonferroni

$$mp_i$$

透過直接對單一實驗校正
控制整個實驗錯誤率修正成 α
→ 太過嚴謹

$$p_i \leftrightarrow \tilde{\alpha}_i = \frac{\alpha}{m}$$
$$\tilde{p}_i = \min\{mp_i, 1\} \leftrightarrow \alpha$$

Sidak

$$1 - (1 - p_i)^m$$

透過各實驗獨立性計算校正
控制整個實驗錯誤率修正成 α
→ 仍太過嚴苛

$$p_i \leftrightarrow \tilde{\alpha}_i \text{ s.t. } \alpha = 1 - (1 - \tilde{\alpha}_i)^m$$
$$\Rightarrow \tilde{\alpha}_i = 1 - (1 - \alpha)^{\frac{1}{m}}$$
$$\tilde{p}_i = \min\{1 - (1 - p_i)^m, 1\} \leftrightarrow \alpha$$

FDR 錯誤發現率 False Discovery Rate

$$\frac{m}{i} p_i$$

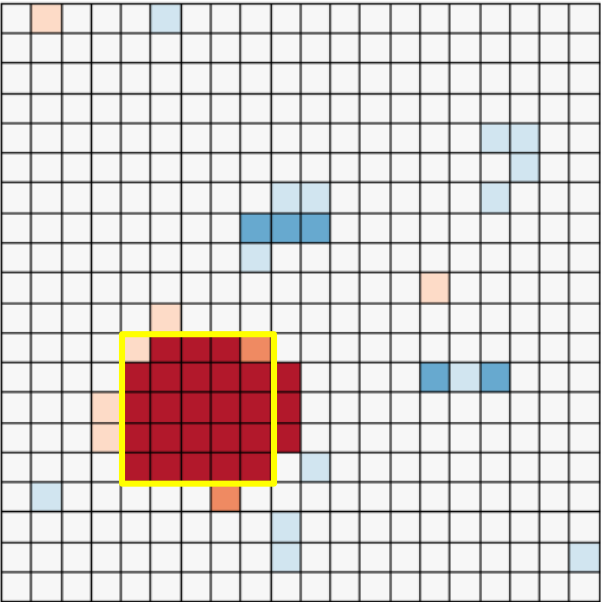
1. 由小到大排序 m 個 p-value
($p_1 \leq p_2 \leq \dots \leq p_m$)
2. 從 p_m 開始找到第一個 i
使得 $p_i \leq \frac{i}{m} \alpha$
3. 比 i 還小的實驗都視為顯著

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

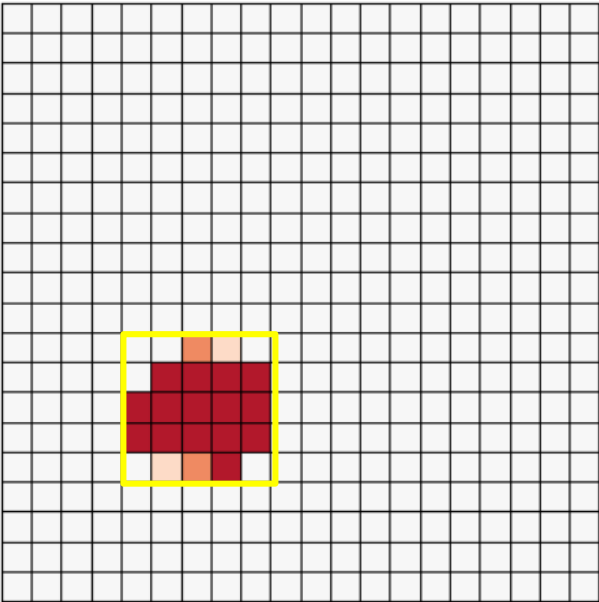
校正比較

模擬20×20的隨機網格，內含5×5的熱區

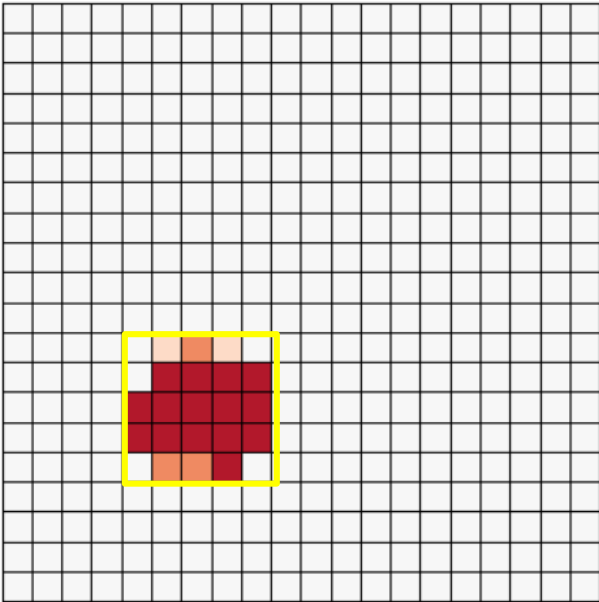
Gi* 原始



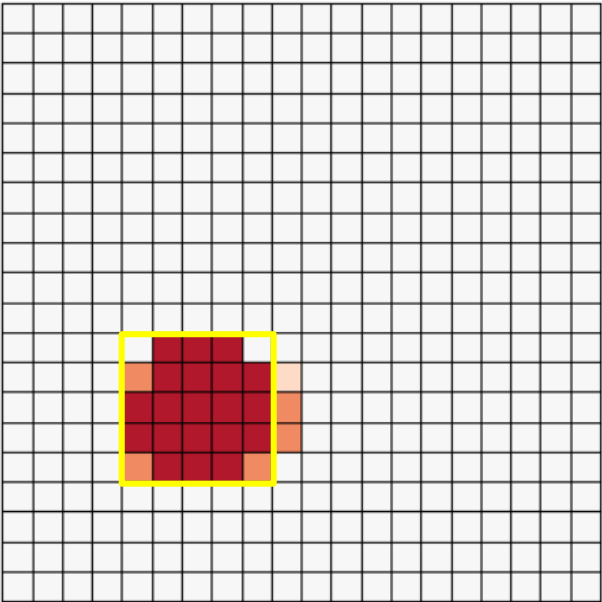
Bonferroni 校正



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

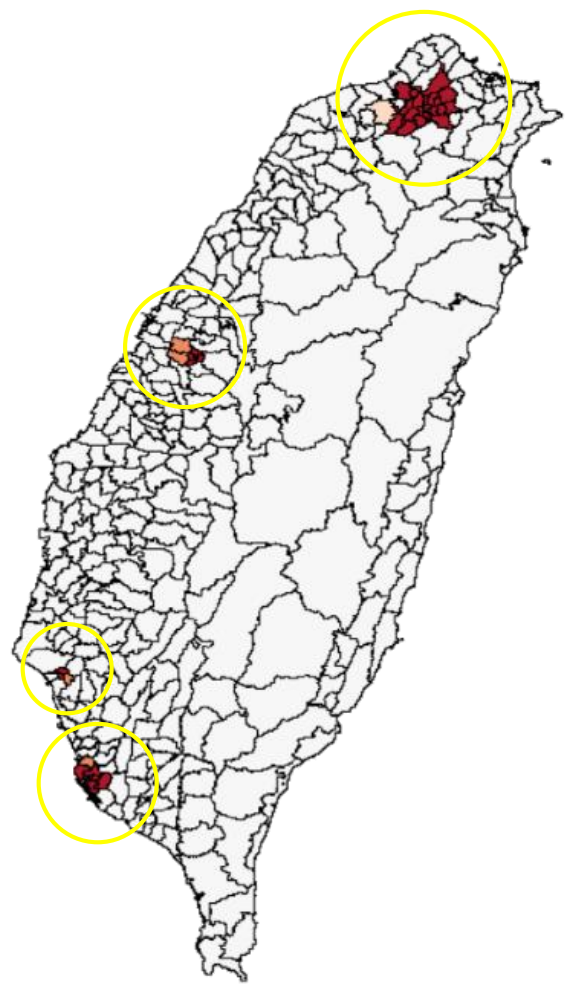
Hotspots

90%	95%	99%
1	5	20

校正比較

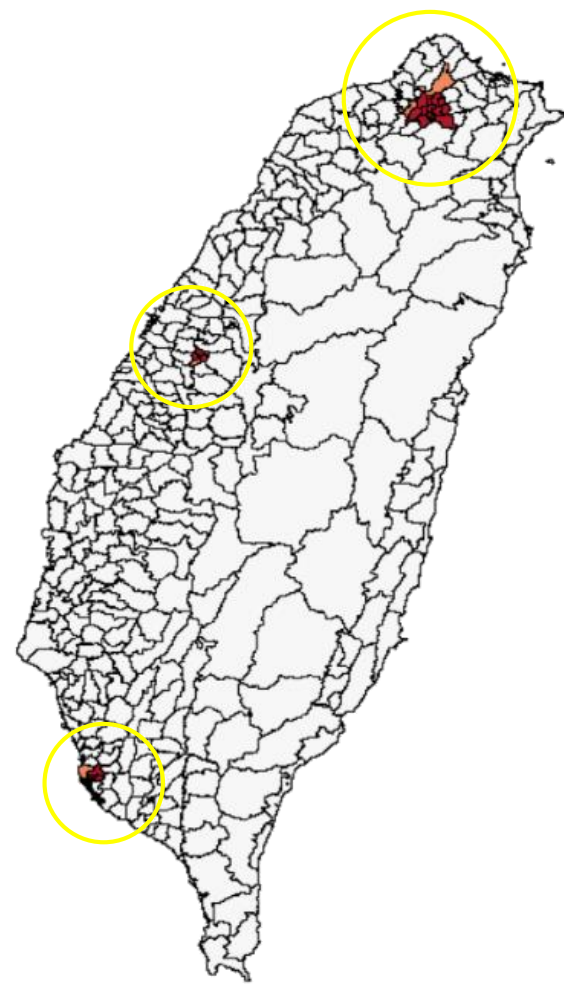
鄉鎮人口密度的空間自相關

Gi* 原始



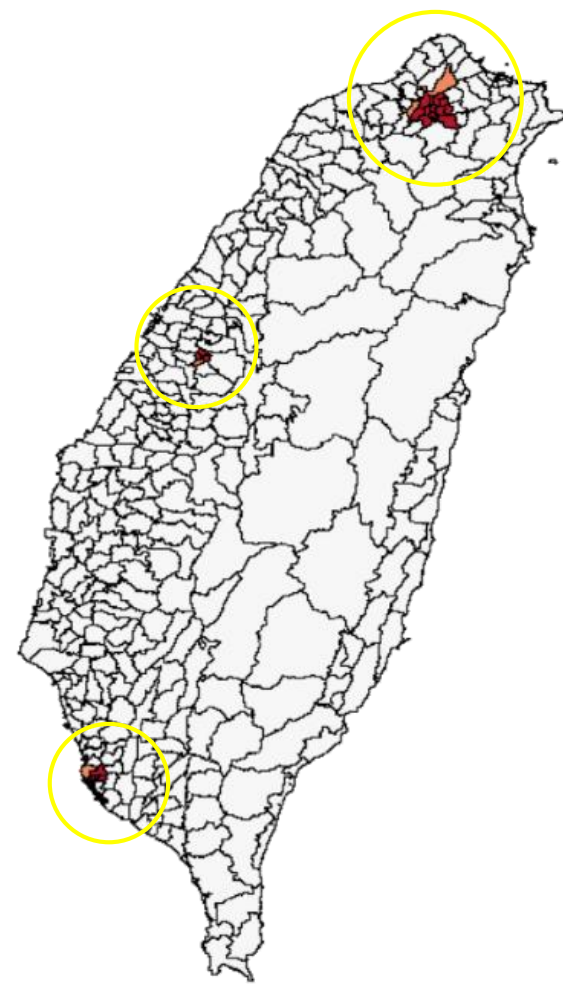
90%	95%	99%
1	4	35

Bonferroni 校正



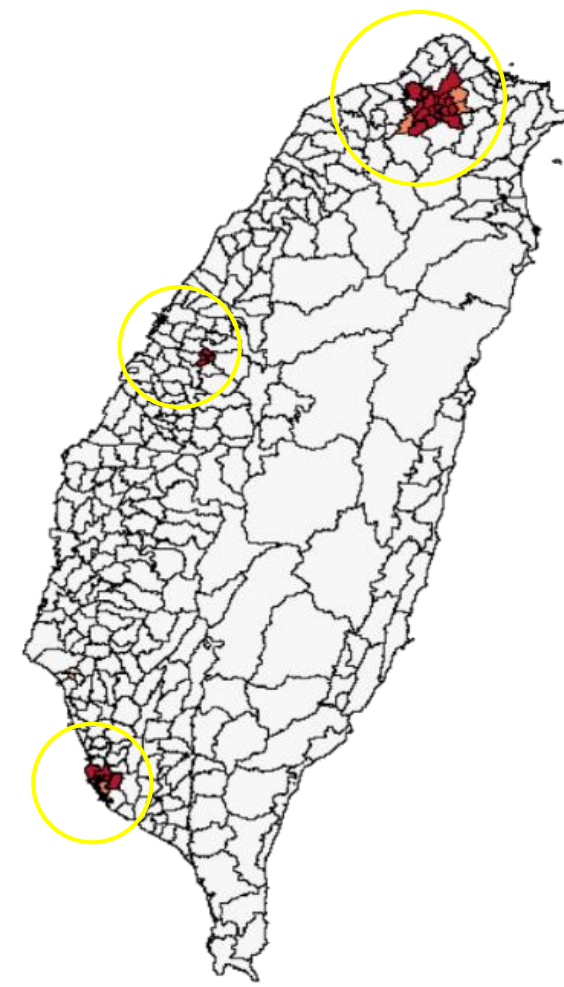
90%	95%	99%
0	4	21

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：台灣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

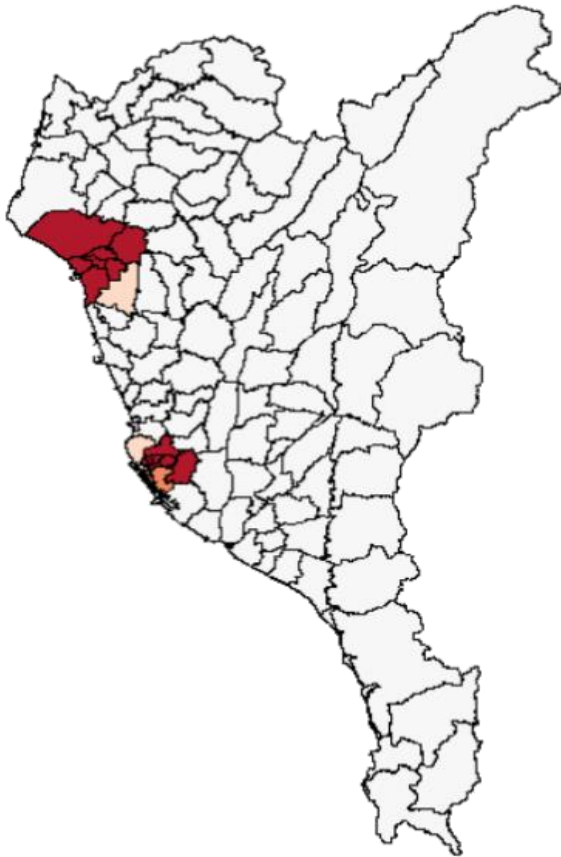
90%	95%	99%
0	2	14

90%	95%	99%
1	1	16

校正比較

2015年登革熱疫情：南高屏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

p_i 和 $\tilde{\alpha}_i$ 比較

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

排序 (i)	原始 p_i	修正後 $\tilde{\alpha}_i$
1	0.00001	0.0005
2	0.00002	0.0005
3	0.00005	0.0005
4	0.0001	0.0005
5	0.0002	0.0005
6	0.0005	0.0005
7	0.001	0.0005
8	0.002	0.0005
9	0.005	0.0005
10	0.01	0.0005
11	0.02	0.0005
12	0.05	0.0005
13	0.1	0.0005
14	0.2	0.0005
.....

方法二

\tilde{p}_i 和 α_i 比較

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

校正後 \tilde{p}_i
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， $\alpha = 0.05$

Sidak

方法一

p_i 和 $\tilde{\alpha}_i$ 比較

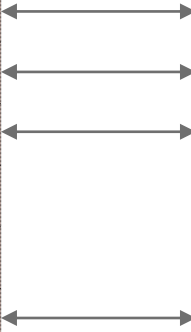
$$\tilde{\alpha}_i = 1 - (1 - \alpha)^{\frac{1}{100}}$$

方法二

\tilde{p}_i 和 α_i 比較

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

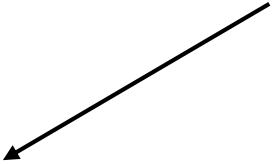
排序 (i)	原始 p_i
1	0.00001
2	0.00002
3	0.00005
4	0.0001
5	0.0002
6	0.0005
7	0.001
8	0.002
9	0.005
10	0.01
11	0.02
12	0.05
13	0.1
14	0.2
.....



修正後 $\tilde{\alpha}_i$
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 \tilde{p}_i
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

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

排序 (i)	原始 p_i	修正後 $\tilde{\alpha}_i$
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 α_i 比較

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

校正後 \tilde{p}_i
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

校正步驟

假設共有100個樣本， $\alpha = 0.05$

FDR

- 1. order the test statistics p-values (p_i) in ascending order ($p_1 \leq p_2 \leq \dots \leq p_m$)
- 2. starting from p_m find the first p_i for which $p_i \leq \left(\frac{i}{m}\right) \alpha$
- 3. regard all tests as significant for which $p_i \leq p_{critical} = \left(\frac{i}{m}\right) \alpha = p_{FDR}$

i	p_i	$\tilde{\alpha}_i$
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	0.0050	0.0045
10	0.0051	0.0050
11	0.0052	0.0055
12	0.0062	0.0060
13	0.0123	0.0065
14	0.2	0.0070
.....

3. 用 $\left(\frac{11}{100}\right) \times 0.05 = p_{FDR}$ 來當判斷標準
(前面的一定都會是顯著的)

?

2. 找到 $p_{11} \leq \left(\frac{11}{100}\right) \times 0.05$

1. 從最大開始找

整體p-value校正

```
p.adjust( )
```

based on the total number

- `p.adjust(p, method = "bonferroni")`
→ `p*length(p)`
- `p.adjust(p, method = "fdr")` *微調
→ `p*length(p)/rank(p)`

P.S. $\text{ifelse}(pi > 1, 1, pi)$

考慮鄰近—隨鄰居數調整

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

```
p.adjustSP(p, nb, method)
```

based on the number of neighbors (+1)

- `localmoran(x, TP.nb.w, p.adjust.method='fdr')`
- `p.adjustSP(p, TP.nb, "fdr")`