熱區分析
&
多重檢定校正

空間分析 2021.05.31

TA 杜承軒

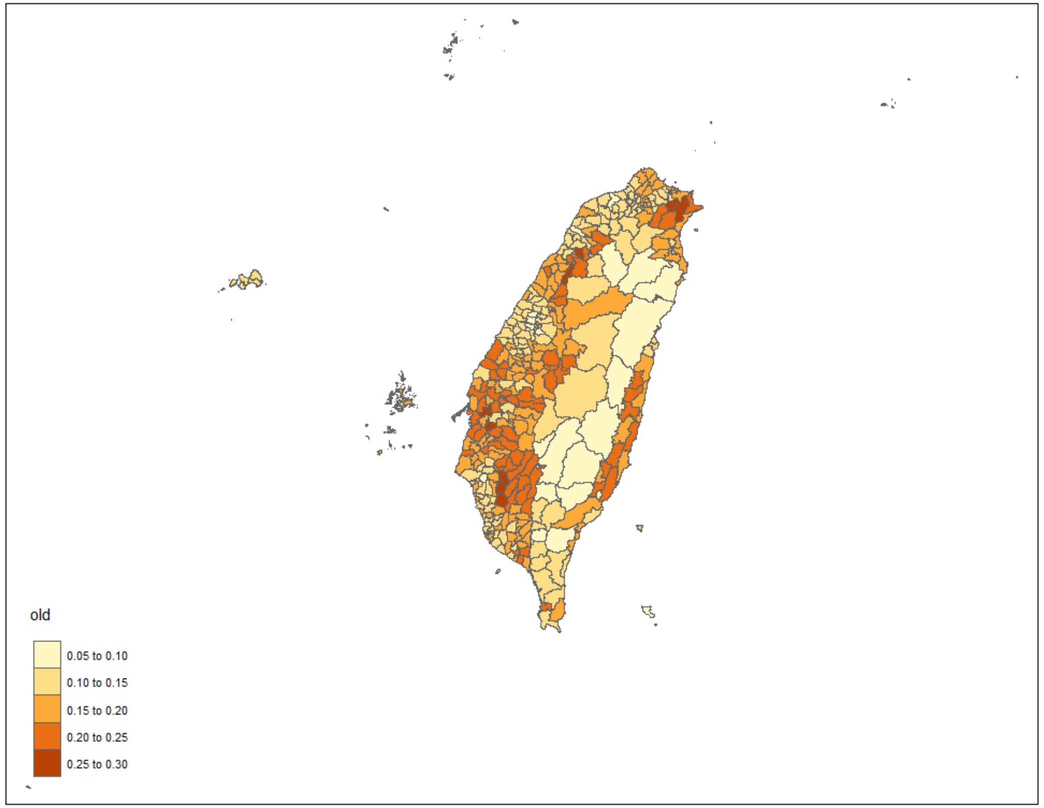
Lab 10 繪製台灣鄉鎮高齡人口比例的主題地圖 （定義：老年人口／全部人口）

【鄰近定義：Contiguity (Queen)】

1. 原始數值
2. LISA map
   (alpha=0.05, 區分 HH, LL, HL, LH)
3. Standardized Gi* values
   (alpha= 0.05, 區分 cluster, non-cluster)
4. 比較LISA進行FDR校正前後的HH熱區分布
   (alpha= 0.05, **校正前** HH)
   (alpha= 0.05, **校正後** HH)
5. 比較Gi*進行Bonferroni校正前後的熱區分布
   (alpha= 0.05, 校正後 cluster)

參考答案 → → →

- 資料：Popn_TWN2.shp



實作

LISA | Gi*

包含自己的鄰近定義

定義「鄰近」

```
TW.nb = poly2nb(TW)
```

```
TW.nb = poly2nb(TW)
TW.nb.in = include.self(TW.nb)
```

建立鄰近表

```
TW.nb.w = nb2listw(TW.nb,
                   zero.policy=T)
```

```
TW.nb.w.in = nb2listw(TW.nb.in)
```

區域空間自相關運算

```
LISA = localmoran(old, TW.nb.w,
       zero.policy = T,
       alternative = "two.sided")
```

```
Gi = localG(old,TW.nb.w.in)
```

```
> LISA
             Ii   E.Ii     Var.Ii          Z.Ii    Pr(z != 0)
220  0.8094220277 -0.025 0.17429168  1.998699187 4.564091e-02
221  0.6620073103 -0.025 0.22386090  1.452018784 1.464964e-01
222  1.3953564727 -0.025 0.17429168  3.402193655 6.684725e-04
223  0.5999538193 -0.025 0.14124553  1.662878712 9.633672e-02
224  1.5232521605 -0.025 0.14124553  4.119593286 3.795417e-05
225  1.3501517812 -0.025 0.17429168  3.293914418 9.880258e-04
226  2.3360250470 -0.025 0.14124553  6.282221450 3.337689e-10
227 -0.0299052525 -0.025 0.08616861 -0.016710399 9.866677e-01
228  0.0003684787 -0.025 0.11764114  0.073963051 9.410398e-01
229 -0.0043165576 -0.025 0.17429168  0.049543250 9.604864e-01
230 -0.0327045528 -0.025 0.06614064 -0.029958028 9.761005e-01
```

Local Moran's I LISA[,1] | Z score LISA[,4] | P value LISA[,5]

```
> Gi
 [1]  1.8911025  1.7181396  2.5357910  2.4823288
 [5]  3.7590712  2.4905072  4.3849408  1.7080833
 [9] -0.1426438  0.2470504  0.1209070 -1.7733190
[13]  2.4211648  2.8866465  2.4180649  2.9475747
[17]  0.9903472 -0.9465509  0.3367046 -0.9960144
[21] -1.4617826 -1.4423588 -1.6701713 -1.7999710
```

Z score of Gi*

LISA

```
LISA = localmoran(old, TW.nb.w, zero.policy = T, alternative = "two.sided" )
```

```
> LISA
              Ii    E.Ii     Var.Ii         Z.Ii     Pr(z > 0)
220 0.8094220277 -0.025 0.17429168 1.998699187 2.282046e-02
221 0.6620073103 -0.025 0.22386090 1.452018784 7.324819e-02
222 1.3953564727 -0.025 0.17429168 3.402193655 3.342363e-04
223 0.5999538193 -0.025 0.14124553 1.662878712 4.816836e-02
224 1.5232521605 -0.025 0.14124553 4.119593286 1.897709e-05
```

alternative = "greater"
預設：是否和鄰居相似(正相關)

HH	
LL	Not-Sig.

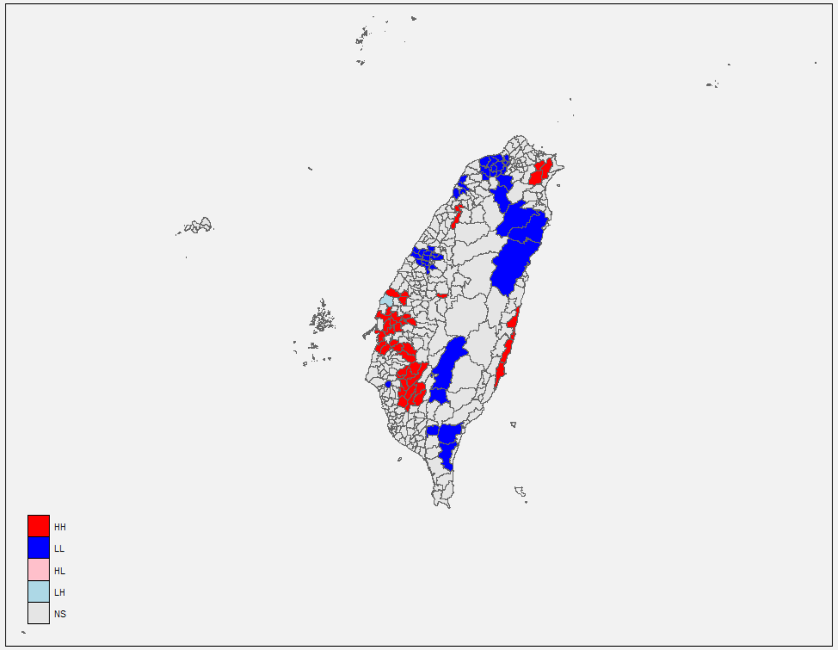
alternative = "two.sided"
我們要的：是否和鄰居有相關

```
Pr(z != 0)
4.564091e-02
1.464964e-01
6.684725e-04
9.633672e-02
3.795417e-05
```

HH	HL	
LL	LH	Not-Sig.

```
LISA = localmoran(old, TW.nb.w, zero.policy=T, alternative ="two.sided")
z = LISA[,4]
p = LISA[,5]
diff = old - mean(old) # 自己比平均是H/L
col = c()
col[diff>0 & z>0] = "red"       # H-H
col[diff<0 & z>0] = "blue"      # L-L
col[diff>0 & z<0] = "pink"      # H-L
col[diff<0 & z<0] = "lightblue" # L-H
col[p>0.05] = "grey90" # 不顯著
TW$colI=col

qtm(TW,'colI')
+tm_add_legend("fill",labels=c("HH","LL","HL","LH","NS"),
 col=c("red","blue","pink","lightblue","grey90"))
```

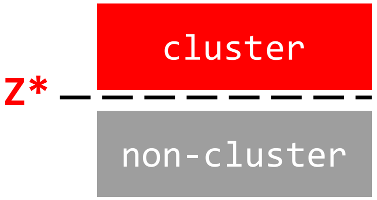


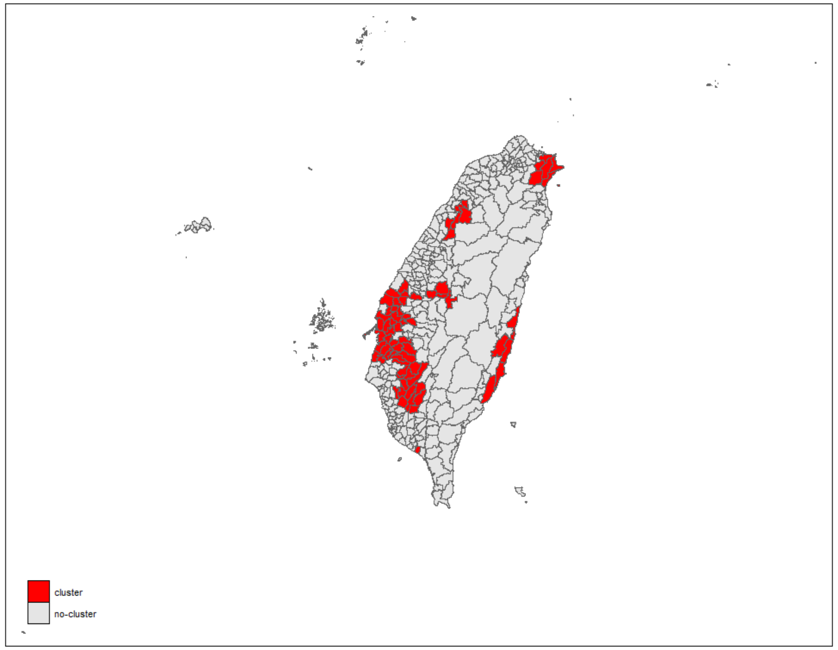
Gi*

```
Gi = localG(old, TW.nb.w.in)
```

※ 會列出Gi*的Z分數

```
> Gi
 [1]  1.8911025  1.7181396  2.5357910  2.4823288
 [5]  3.7590712  2.4905072  4.3849408  1.7080833
 [9] -0.1426438  0.2470504  0.1209070 -1.7733190
[13]  2.4211648  2.8866465  2.4180649  2.9475747
```

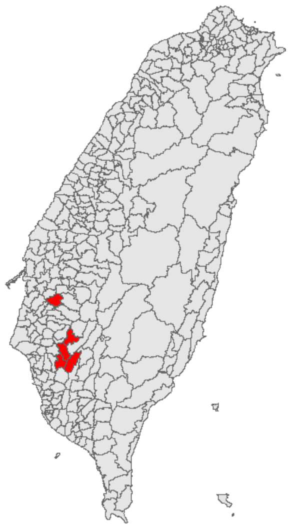





```
TW$Gi = localG(old,TW.nb.w.in)
TW$colG="grey90"
TW$colG[TW$Gi>=qnorm(.95)]="red"
qtm(TW,'colG')
```

Bonferroni校正: 1-0.05/n

```
qnorm(1-0.05)      → 1.64
qnorm(1-0.05/10)   → 3.09
qnorm(1-0.05/100)  → 3.29
qnorm(1-0.05/1000) → 3.89
```



- Gi* 原始數值

```
Gi = localG(old, TW.nb.w.in, return_internals = T)
```

※ 可以列出每個格子的Gi*, 以及期望值、變異數

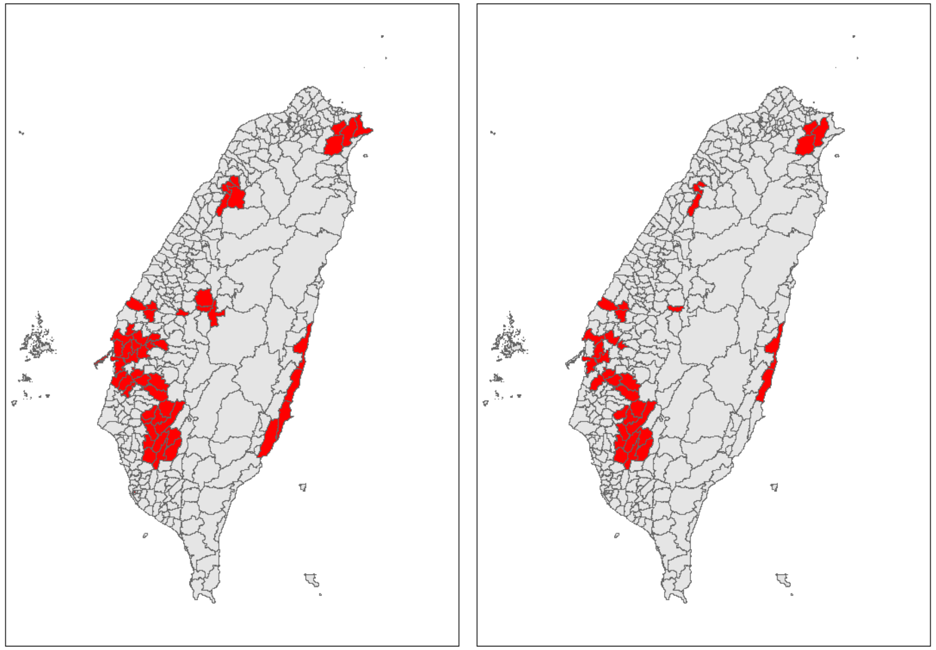
```
> attr(Gi,"internals")
```

	G	EG	VG
1	0.0443024793	0.02439024	1.108689e-04
2	0.0444890960	0.02439024	1.368440e-04
3	0.0510906836	0.02439024	1.108689e-04
4	0.0482406792	0.02439024	9.231537e-05

FDR校正

```r
LISA. = localmoran(old, TW.nb.w, zero.policy = T)
p = LISA.[,5]
p.adj = p.adjust(p,"fdr")

TW$colHHfdr="grey90"
TW$colHHfdr[p.adj<0.05 & diff>0]="red"
```



FDR概念

i	p_i	p_i^*
1	0.00001	0.0010
2	0.00002	0.0010
3	0.00005	0.0017
4	0.0001	0.0025
5	0.0002	0.0040
6	0.0005	0.0083
7	0.001	0.0143
8	0.002	0.0250
9	0.0050	0.0556
10	0.0051	0.0510
11	0.0052	0.0473
12	0.0062	0.0517
13	0.0123	0.0946
14	0.2	1.4 → 1
……	……	……

假設共有100個樣本，$\alpha = 0.05$

$$p_i^* = p_i \times \frac{100}{i}$$

1.從p-value數值大的開始搜尋

2.找到第一個熱區（顯著）

3.剩下的全部都是熱區

?

← 第一個熱區

Caldas de Castro, M., & Singer, B. H. (2006). Controlling the false discovery rate: a new application to account for multiple and dependent tests in local statistics of spatial association. *Geographical Analysis, 38*(2), 180-208.

多重檢定校正

https://youtu.be/5bqHT3Gp2W0

Moran's I

$$I=\frac{n}{W}\frac{\sum_i\sum_j w_{ij}(x_i-\bar{x})(x_j-\bar{x})}{\sum_i (x_i-\bar{x})^2}$$

$$\xrightarrow{\tilde{x}_i=x_i-\bar{x}} \frac{n}{W}\frac{\sum_i\sum_j w_{ij}\,\tilde{x}_i\,\tilde{x}_j}{\sum_i \tilde{x}_i^{\,2}}$$

- $W=\sum_i\sum_j w_{ij}$
- $\sum_i (x_i-\bar{x})^2 = n\,\sigma_x^2=(n-1)\,s_x^2=(n-1)\,s_{\tilde{x}}^2$

```
> TP.nb=poly2nb(TP)
> TP.nb.w=nb2listw(TP.nb)
> M=moran.test(x,TP.nb.w)
> M$estimate[1]
Moran I statistic
      -0.01261841
> TP.nb.M=nb2mat(TP.nb)
> xx=x-mean(x)
> sum(TP.nb.M*(xx%*%t(xx)))/sum(xx^2)
[1] -0.01261841
> sum(TP.nb.M*(xx%*%t(xx)))/(var(xx)*11)
[1] -0.01261841
```

$$I=\frac{n}{W}\frac{\sum_i\sum_j w_{ij}(x_i-\bar{x})(x_j-\bar{x})}{\sum_i (x_i-\bar{x})^2}$$

$$=\frac{n}{W}\frac{\sum_i\sum_j w_{ij}(x_i-\bar{x})(x_j-\bar{x})}{n\sigma^2}$$

$$=\frac{1}{W}\sum_i\sum_j w_{ij}\frac{(x_i-\bar{x})}{\sigma}\frac{(x_j-\bar{x})}{\sigma}$$

$$=\frac{1}{W}\sum_i\sum_j w_{ij}z_i z_j$$

$$=\frac{1}{W}\sum_i z_i\sum_j w_{ij}z_j=\frac{1}{W}\sum_i \boldsymbol{I_i}$$

Local Moran's I

$$I_i=z_i\sum_j w_{ij}z_j$$

$$I_i=\frac{x_i-\bar{x}}{s^2}\sum_{j\neq i} w_{ij}(x_j-\bar{x})=\mathbb{z}_i\sum_j w_{ij}\mathbb{z}_j$$

- $z_i=\frac{x_i-\bar{x}}{\sigma}$
- $\mathbb{z}_i=\frac{x_i-\bar{x}}{s}$

```
> LISA=localmoran(x,TP.nb.w)
> LISA[1];          sum(LISA[,1])/12
[1] 0.005094452     [1] -0.01261841
> z=(x-mean(x))/(sd(x)*sqrt(11/12))
> z[1]*sum(TP.nb.M[1,]*z)
[1] 0.005094452
> LISA=localmoran(x,TP.nb.w,mlvar=F)
> LISA[1]
[1] 0.004669914
> z=(x-mean(x))/sd(x)
> z[1]*sum(TP.nb.M[1,]*z)
[1] 0.004669914
```

補充：用矩陣方法一次求得所有I_i

```
> z*(TP.nb.M%*%z)
```

P.S.

$$I_i=\frac{x_i-\bar{x}}{s_i^2}\sum_{j\neq i} w_{ij}(x_j-\bar{x})\,;\; s_i^2=\frac{\sum_{j\neq i} w_{ij}(x_j-\bar{x})^2}{n-1}$$

```
> lx=xx[1]*sum(TP.nb.M[1,]*xx)
> si2=var(x[-1])*10/11
> lx/si2
[1] 0.004670523
```

矩陣方法：

```
> xx*(TP.nb.M%*%xx)/ sapply(1:12,
     function(i) var(x[-i])*10/11)
```

Getis-Ord General G

$$G=\frac{\sum_i\sum_j w_{ij}x_i x_j}{\sum_i\sum_j x_i x_j}, j\neq i$$

$$\left(當w_{ii}=0 \xrightarrow{\text{ignore } j\neq i} G=\frac{\sum_i\sum_j w_{ij}x_i x_j}{\sum_i\sum_j x_i x_j-\sum_i x_i^2}\right)$$

```
> G=globalG.test(x,TP.nb.w)
> G$estimate[1]
Global G statistic
         0.09243927
> G.num=sum(TP.nb.M*(x%*%t(x)))
> G.den=sum(x%*%t(x))-sum(x^2)
> G.num/G.den
[1] 0.09243927
```

R package – spdep

$$I=\frac{n}{\sum_{i=1}^n\sum_{j=1}^n w_{ij}}\frac{\sum_{i=1}^n\sum_{j=1}^n w_{ij}(x_i-\bar{x})(x_j-\bar{x})}{\sum_{i=1}^n (x_i-\bar{x})^2}$$

$$I_i=\frac{(x_i-\bar{x})}{\sum_{k=1}^n (x_k-\bar{x})^2/(n-1)}\sum_{j=1}^n w_{ij}(x_j-\bar{x})$$

localmoran(mlvar=TRUE)

mlvar: values of local Moran's I are reported using the variance of the variable of interest (sum of squared deviances over n), but can be reported as the sample variance, dividing by (n-1) instead

Getis-Ord Gi*

$$G_i^*=\frac{\sum_j w_{ij}x_j}{\sum_j x_j}$$

>> 套件函數

>> 手動計算

Getis-Ord Gi

$$G_i=\frac{\sum_j w_{ij}x_j}{\sum_j x_j}, j\neq i$$

```
> Gi.=localG(x,TP.nb.w.in,return_internals=T)
> attr(Gi.,"internals")[,1]
0.0862 0.0885 0.0923 0.0868 0.0845 ……
> TP.nb.M.in%*%x/sum(x)
0.0862 0.0885 0.0923 0.0868 0.0845 ……
```

```
> Gi=localG(x,TP.nb.w,return_internals=T)
> attr(Gi,"internals")[,1]
0.0946 0.0966 0.0969 0.0948 0.0948 ……
> TP.nb.M%*%x/(sum(x)-x)
0.0946 0.0966 0.0969 0.0948 0.0948 ……
```

空間自相關計算

https://youtu.be/gOuFIxk8oFI

14:38 ~ 46:20