貝氏迴歸模型偵測歧異點方法於氣溫校驗的應用

逢甲大學 統計學系 魏裕中

2018.02.05

頻率學派

Kendall Maurice
(1907-1983)
used the term 'frequentist'

貝氏學派

Thomas Bayes
(1702-1761)

1

頻率學派

貝氏學派

θ

未知參數

$x_1, x_2, \ldots$

樣本訊息

$x_1, x_2, \ldots$

樣本訊息

+

θ

事前資訊

1

頻率學派

貝氏學派

θ

未知參數

$postr(\theta|x_1,x_2,\ldots)$

$L(x_1,x_2,\ldots|\theta)$

$L(x_1,x_2,\ldots|\theta)$

+

$prior(\theta)$

1

頻率學派

貝氏學派

θ

未知參數

2013-2017
樣本訊息

2013-2017
樣本訊息

+

1998-2012
事前資訊

頻率學派

貝氏學派

θ

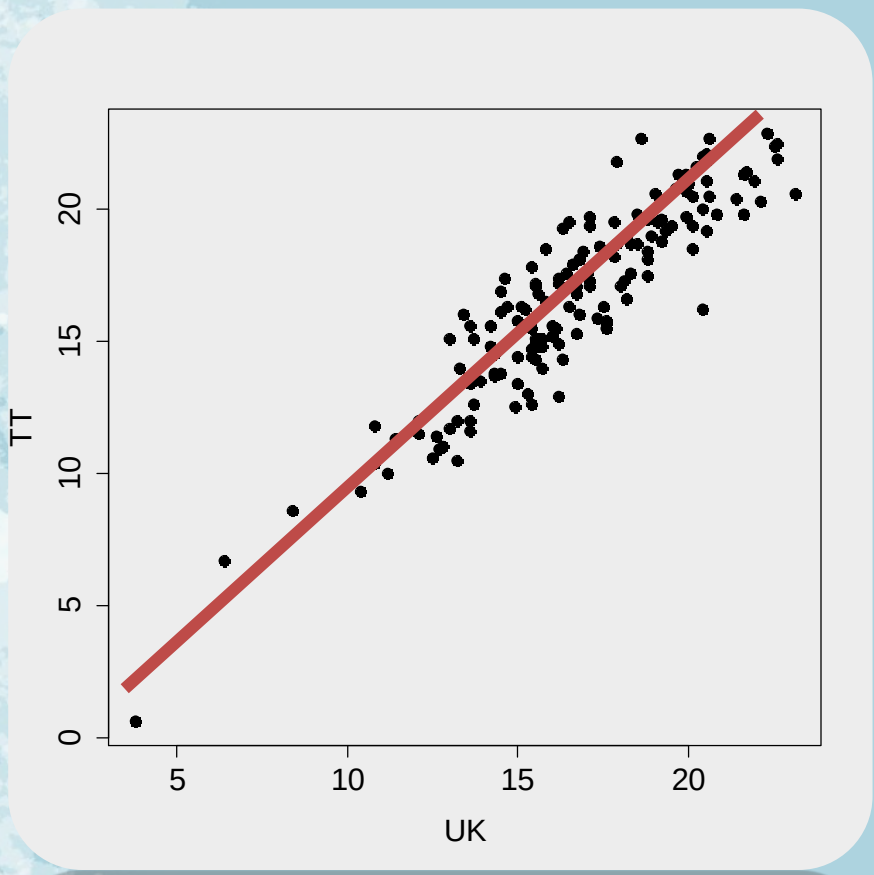
未知參數

Fixed Number

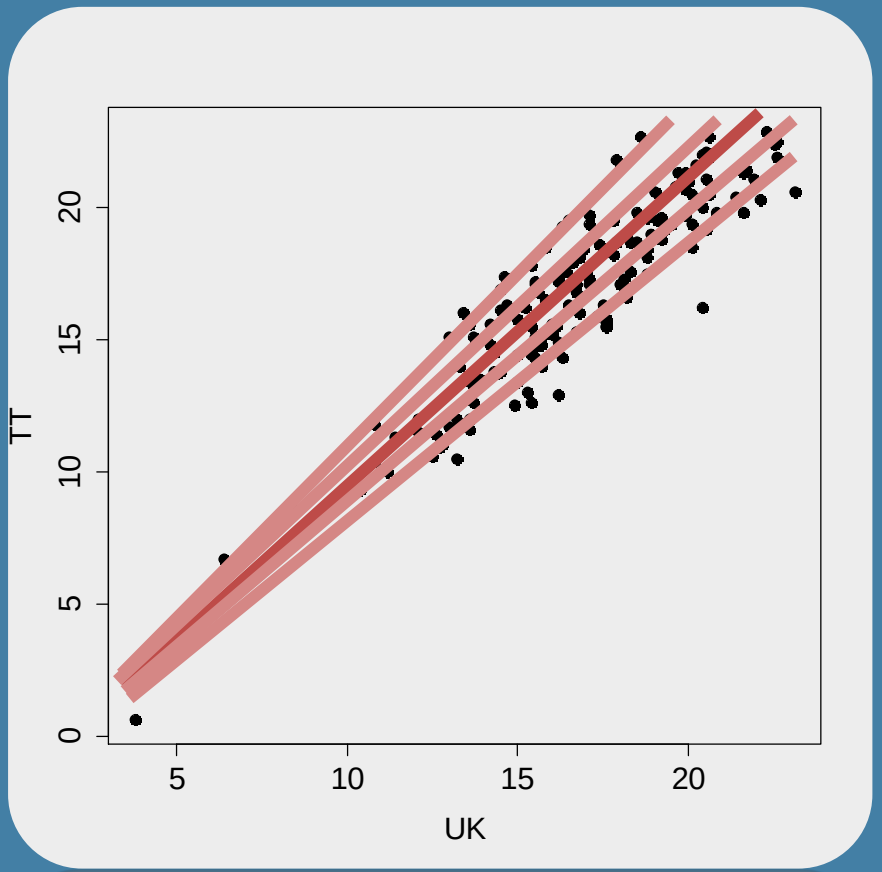
Random Variable

2

頻率學派



貝氏學派



貝氏迴歸模型偵測歧異點

選取欲探討的資料點

(x^*, y^*)

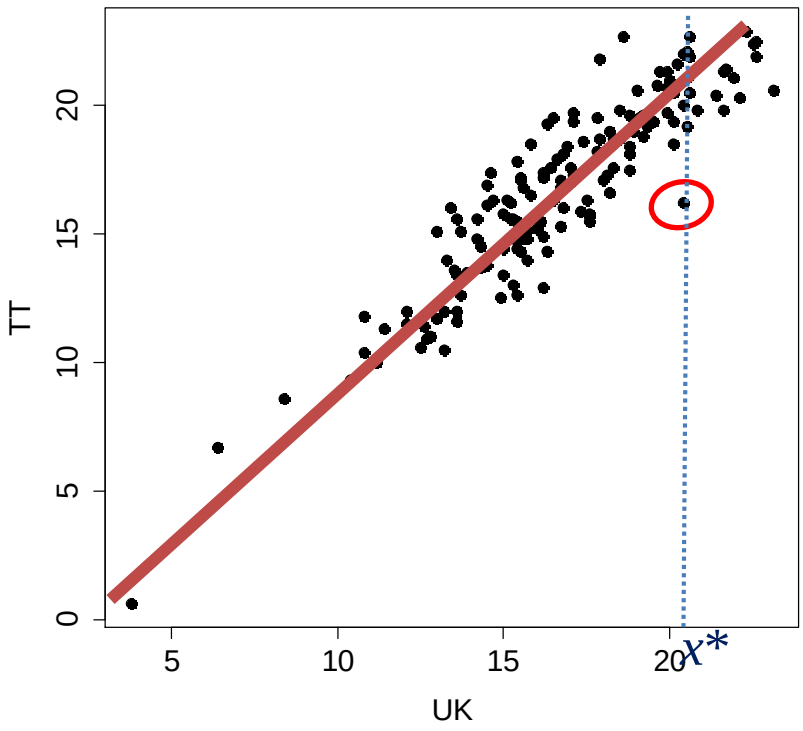
去除該點建構迴歸模型
得到迴歸參數事後機率

$postr(\boldsymbol{\beta}, \sigma^2 | \mathbf{x}, \mathbf{y})$

計算該點於迴歸模型中
的事後預測機率

$postr\ (\ y^* | \ x^*, \mathbf{x}, \mathbf{y})$

判斷是否為歧異點



貝氏迴歸模型偵測歧異點

選取欲探討的資料點

去除該點建構迴歸模型得到迴歸參數事後機率

計算該點於迴歸模型中的事後預測機率

判斷是否為歧異點

$$postr\,(y^*|\,x^*, \mathbf{x}, \mathbf{y})$$

$$=\iint p(y^*|\,x^*, \mathbf{x}, \mathbf{y}, \boldsymbol{\beta}, \sigma^2)\times p(\boldsymbol{\beta}, \sigma^2|\,x^*, \mathbf{x}, \mathbf{y})\,d\boldsymbol{\beta}\,d\sigma^2$$

$$=\iint p(y^*|\,x^*, \boldsymbol{\beta}, \sigma^2)\times p(\boldsymbol{\beta}, \sigma^2|\,\mathbf{x}, \mathbf{y})\,d\boldsymbol{\beta}d\sigma^2$$

$$prior(\boldsymbol{\beta}, \sigma^2)L(\mathbf{x}, \mathbf{y}|\boldsymbol{\beta}, \sigma^2)/c$$

$$prior(\boldsymbol{\beta}|\sigma^2)\;prior(\sigma^2)$$

Normal *Normal* *Inverse-Gamma* *Normal*

貝氏迴歸模型偵測歧異點

選取欲探討的資料點

去除該點建構迴歸模型得到迴歸參數事後機率

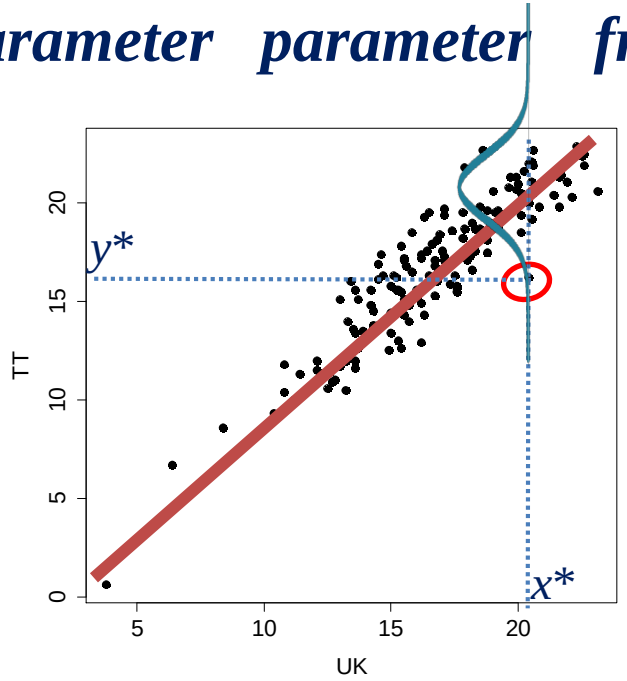
計算該點於迴歸模型中的事後預測機率

判斷是否為歧異點

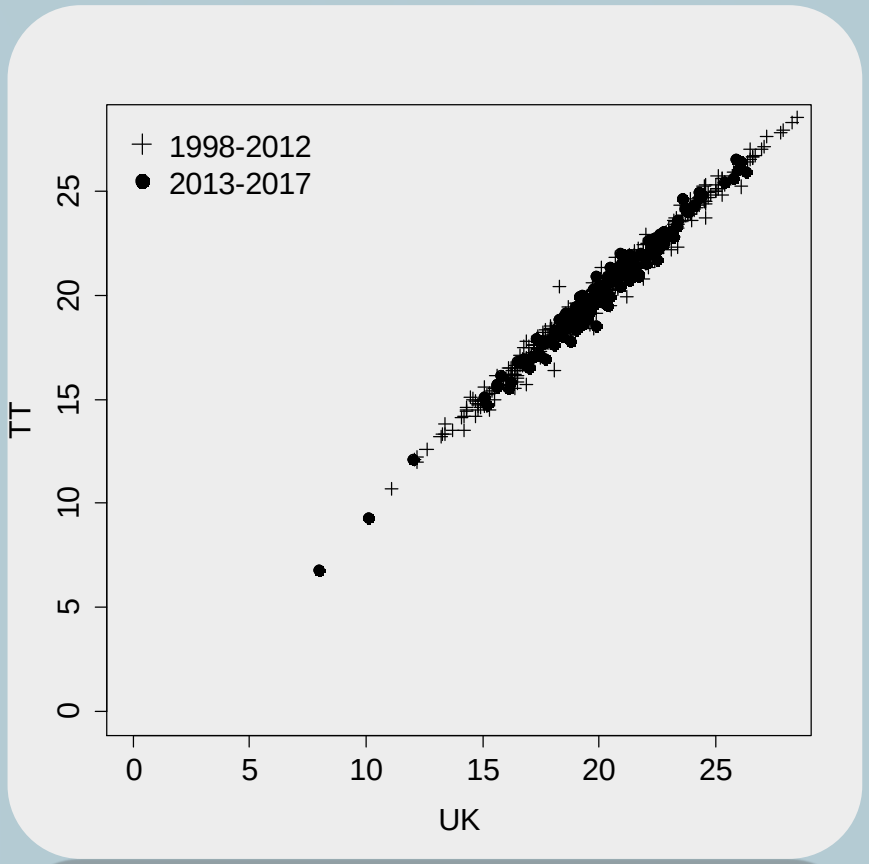
$$postr\,(y^*|\,x^*, \mathbf{x}, \mathbf{y})$$

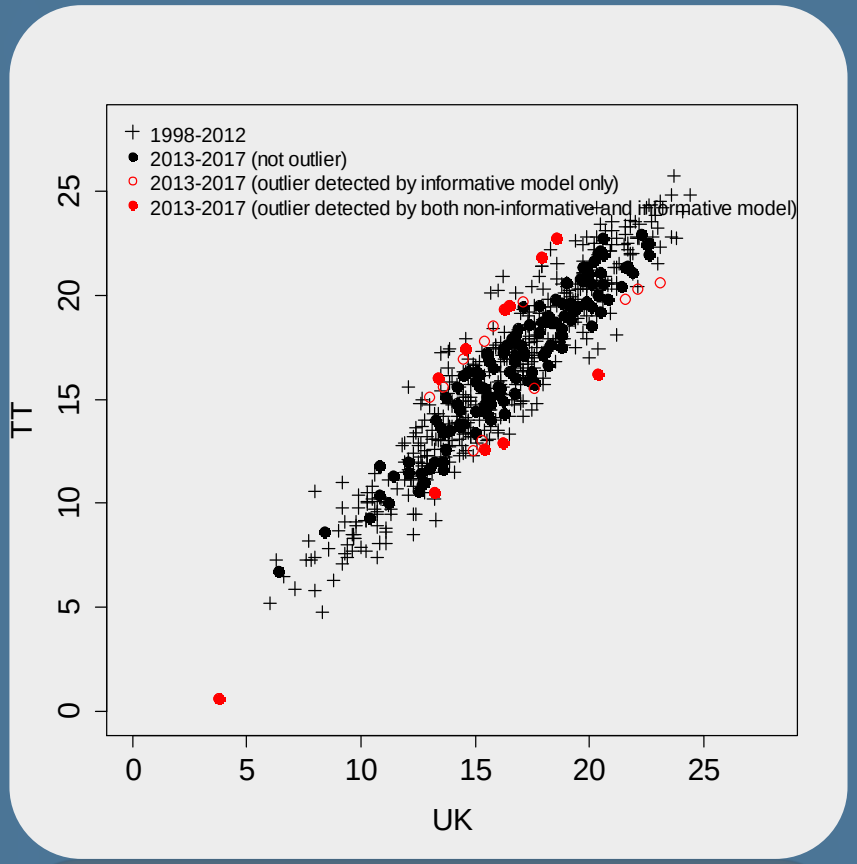
$$\sim t(x^*\,\bar{\beta},\ \overline{s^2}\,\{I + x^*\,\bar{V}\,x^*\},\ \bar{\nu})$$

Location parameter *Scale parameter* *Degree of freedom*



貝氏迴歸模型偵測歧異點





結論

資料

- 單一測站/ 單一月份單一時段
- 由氣溫參考值 找歧異的氣溫觀察值

資料處理與模型選擇

- 氣溫參考值 與 氣溫觀察值 差距> 單一變數
- 氣溫參考值 與 氣溫觀察值 關聯> 迴歸模型

結論

由迴歸模型偵測歧異點

- 頻率學派：
  - ⊙ Student化殘差&去點殘差
    - >Y軸離群
  - ⊙ Cook距離 & DFFITS
    - >X軸與Y軸離群
  - ⊙ 預測區間

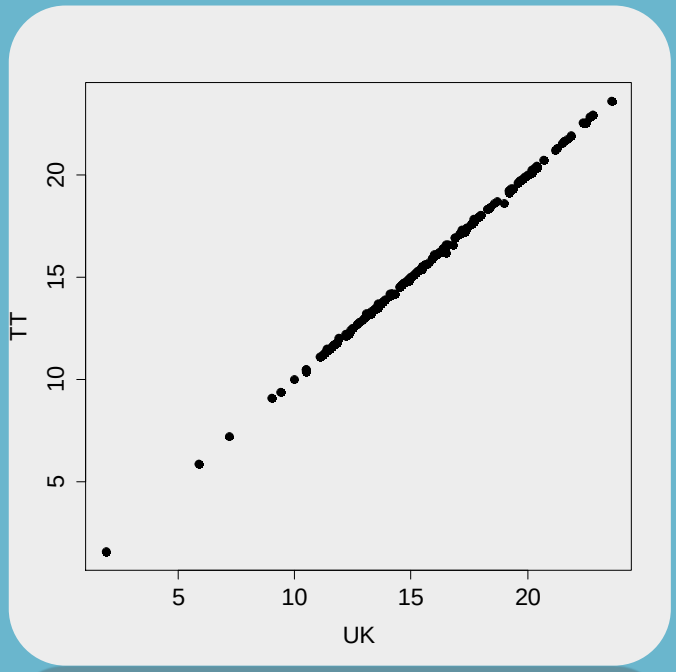
- 貝氏學派：
  - ⊙ 事後預測機率
    - >納入歷史資料訊息
    - >提供預測離群機率

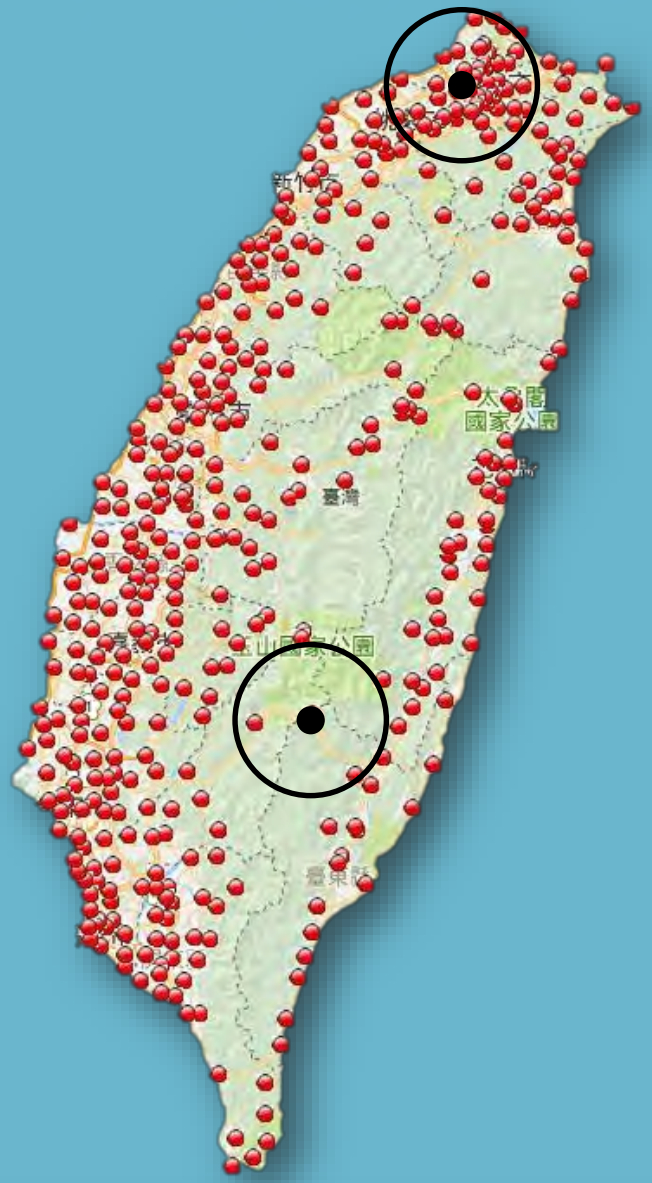
討論

1. 參考值選擇

找 氣溫觀察值異常點 (TT)

由 氣溫參考值當基準 (UK)
鄰近測站氣溫觀察值估計 (TT)





討論

1. 參考值選擇

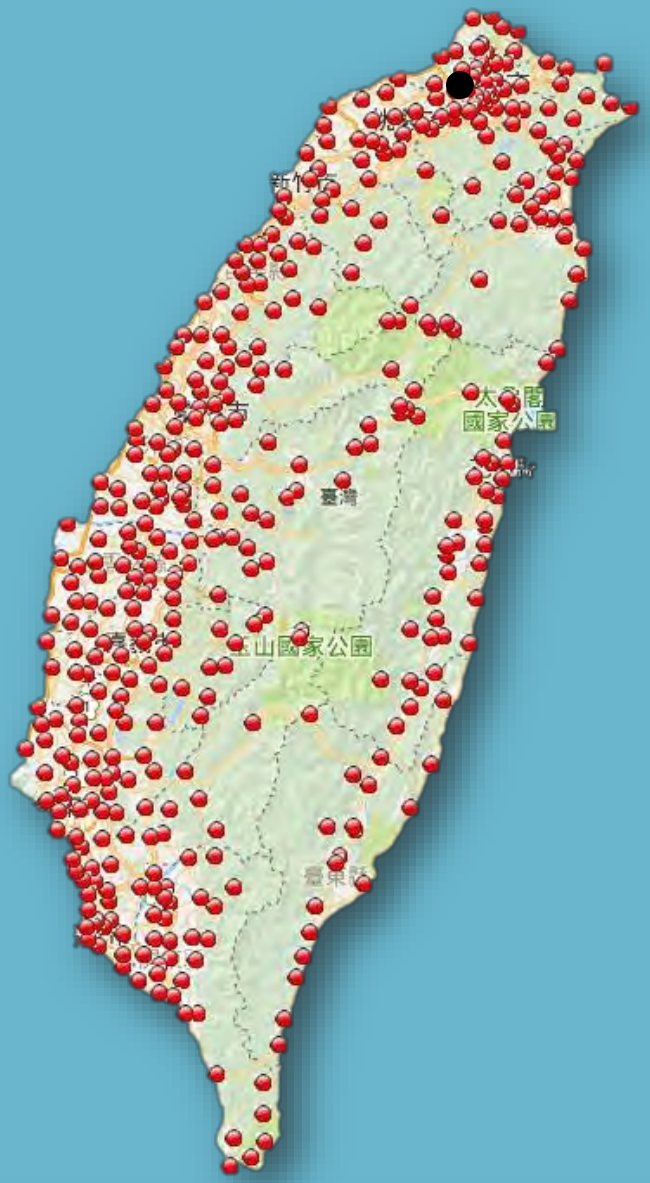
找 氣溫觀察值異常點 (TT)

由 氣溫參考值當基準 (UK)
鄰近測站氣溫觀察值估計 (TT)

(1) 由其他氣溫參考值當基準
(MOS, INI, DA, ...)

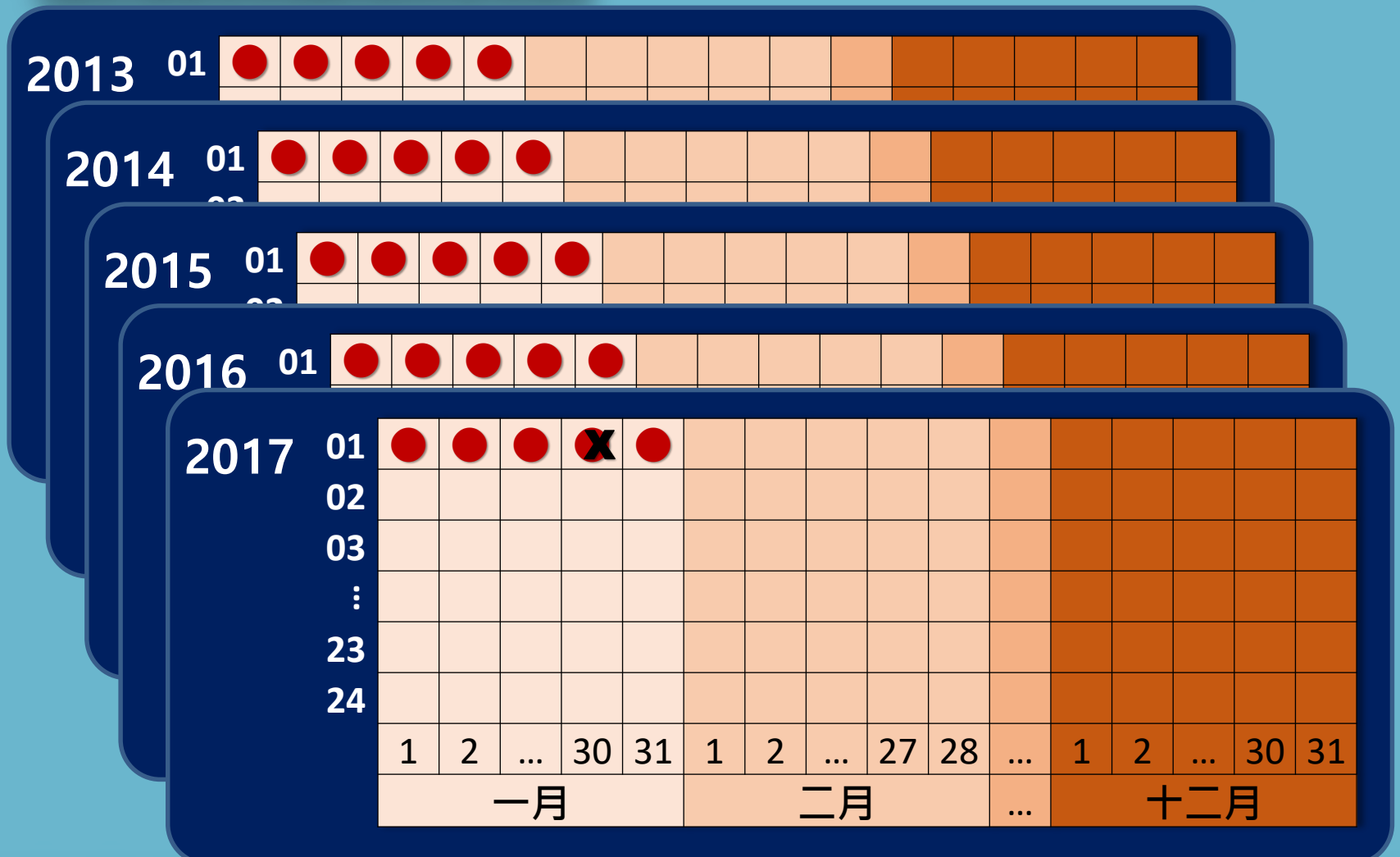
(2) 綜合各種參考值當基準
(UK+MOS+INI+DA+...)

- 各參考值的特性
- 綜合各種參考值的方法
(Bayesian Model Averaging)



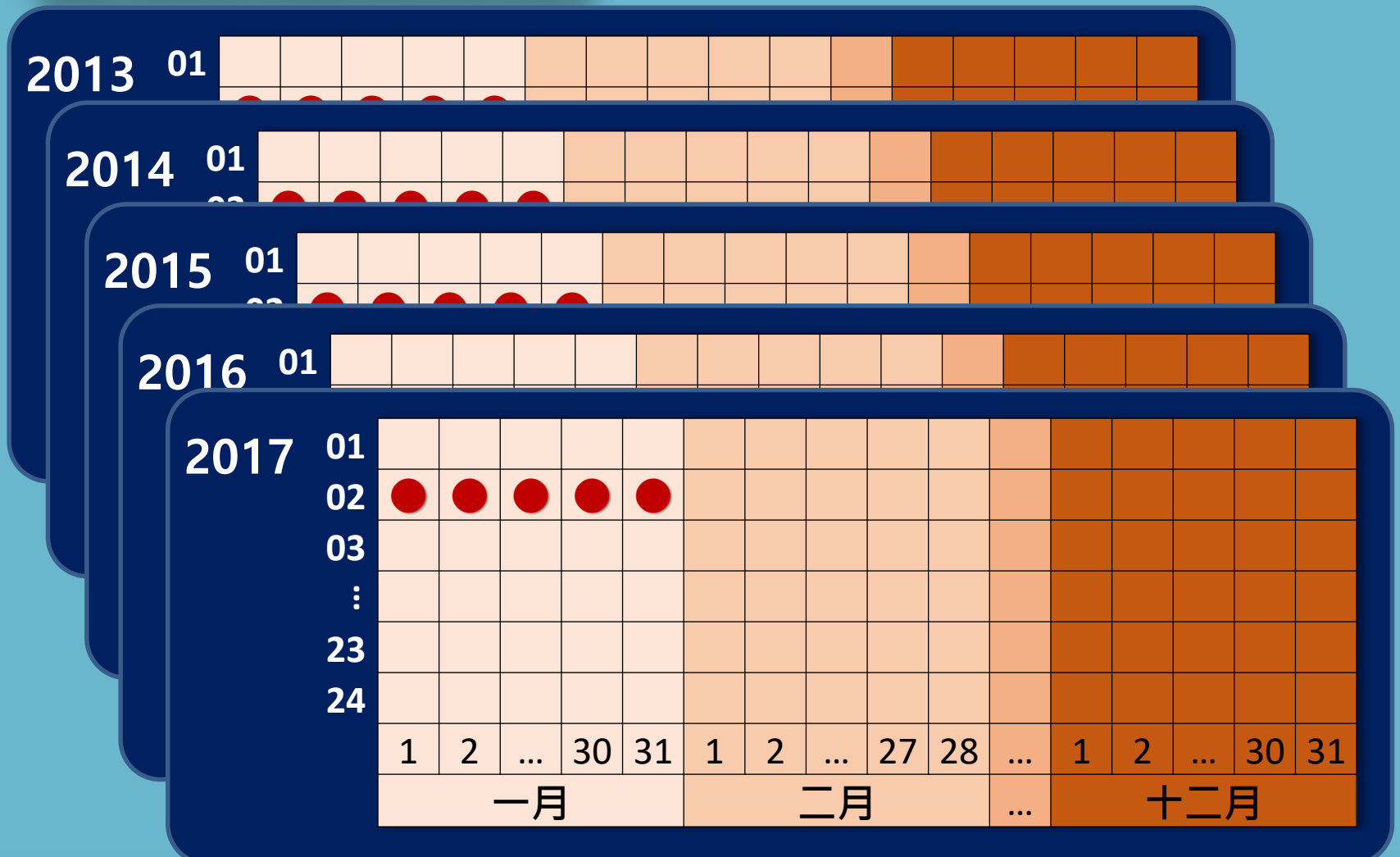
討論

2. 資料選取方式


2013 01
2014 01
2015 01
2016 01
2017 01
02
03
23
24
1 2 … 30 31 1 2 … 27 28 … 1 2 … 30 31
一月
二月
…
十二月

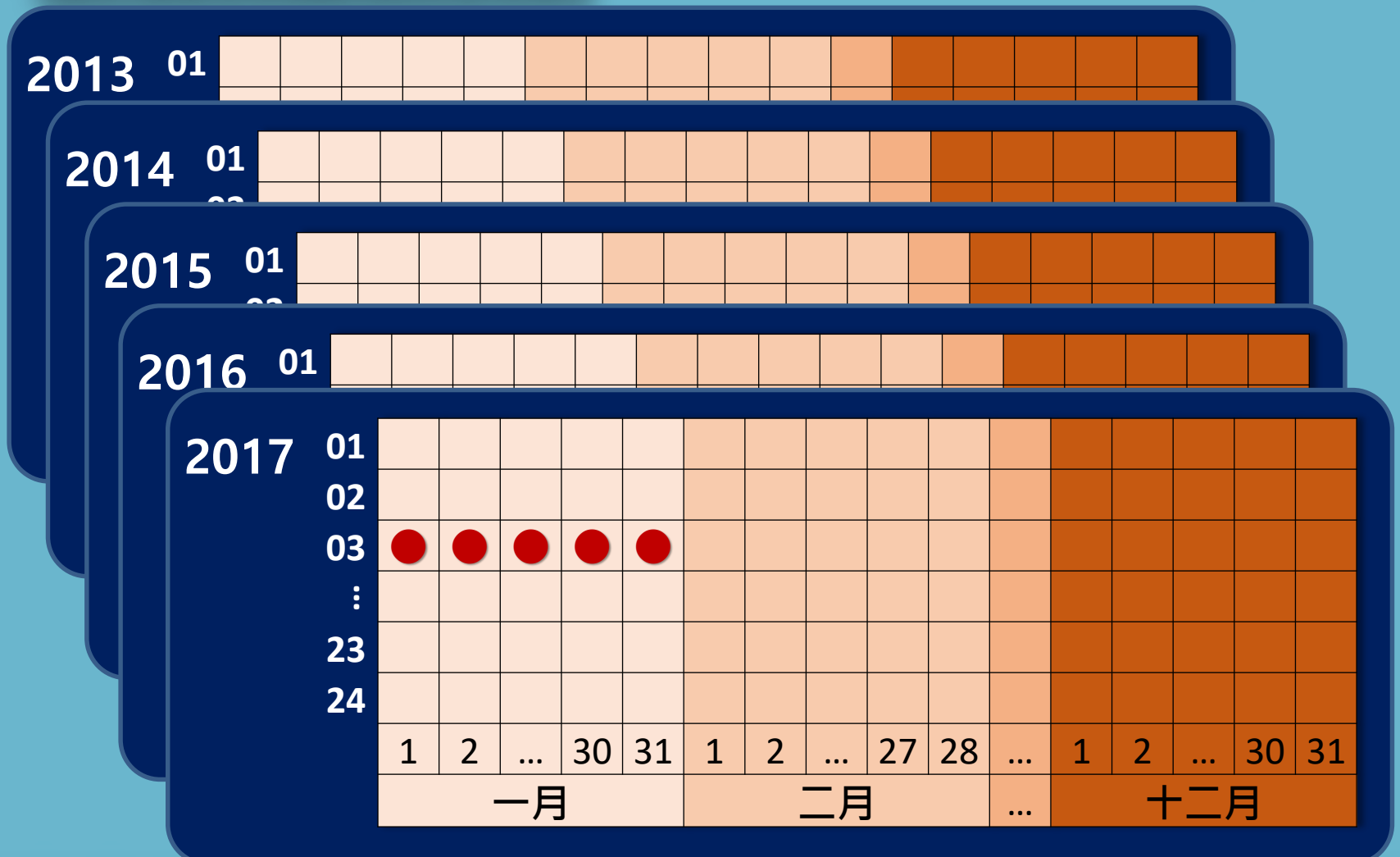
討論

2. 資料選取方式


2013
2014
2015
2016
2017
01
02
03
23
24
1 2 … 30 31 1 2 … 27 28 … 1 2 … 30 31
一月
二月
…
十二月

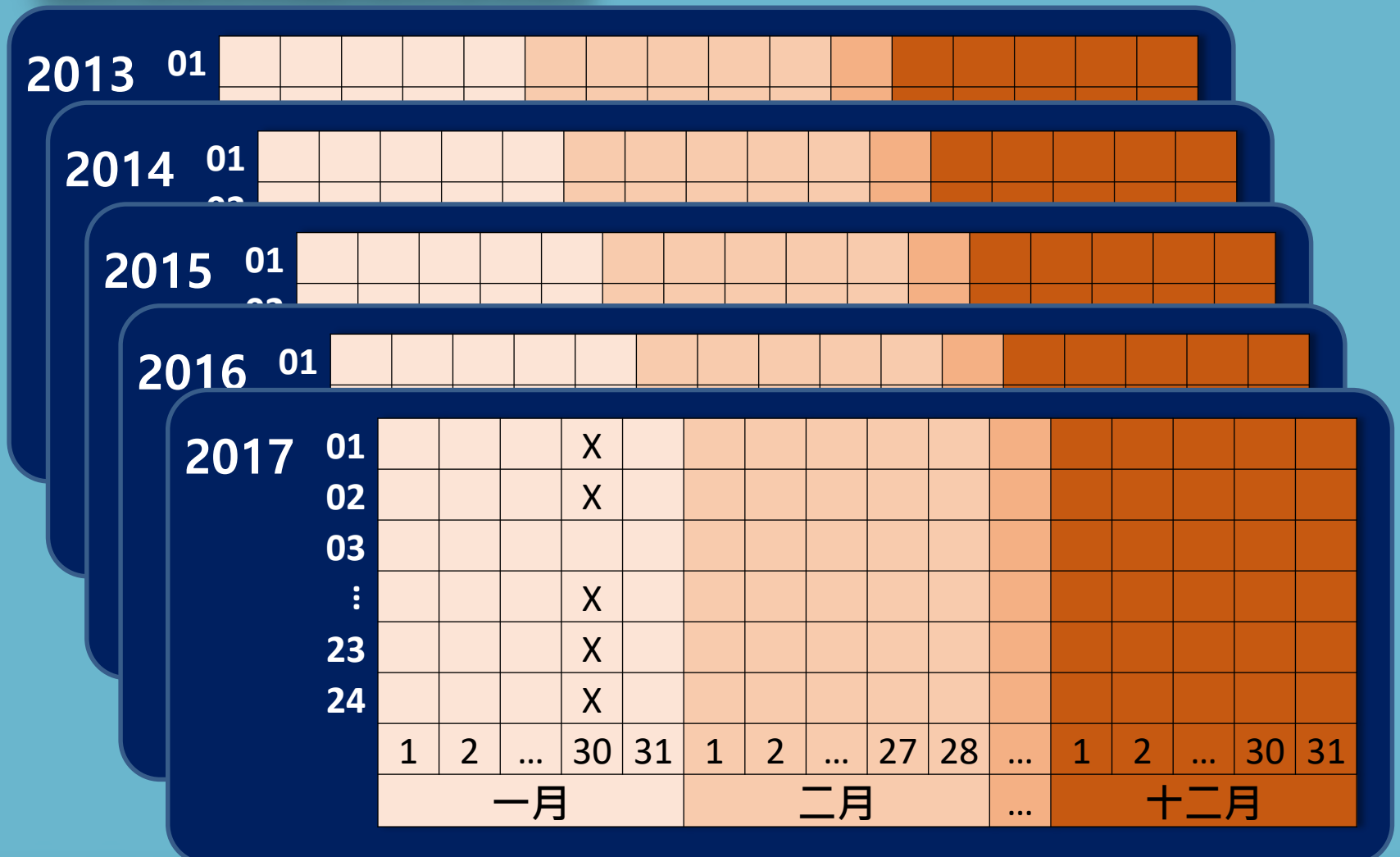
討論

2. 資料選取方式


2013
01
2014
01
2015
01
2016
01
2017
01
02
03
23
24
1 2 … 30 31 1 2 … 27 28 … 1 2 … 30 31
一月
二月
…
十二月

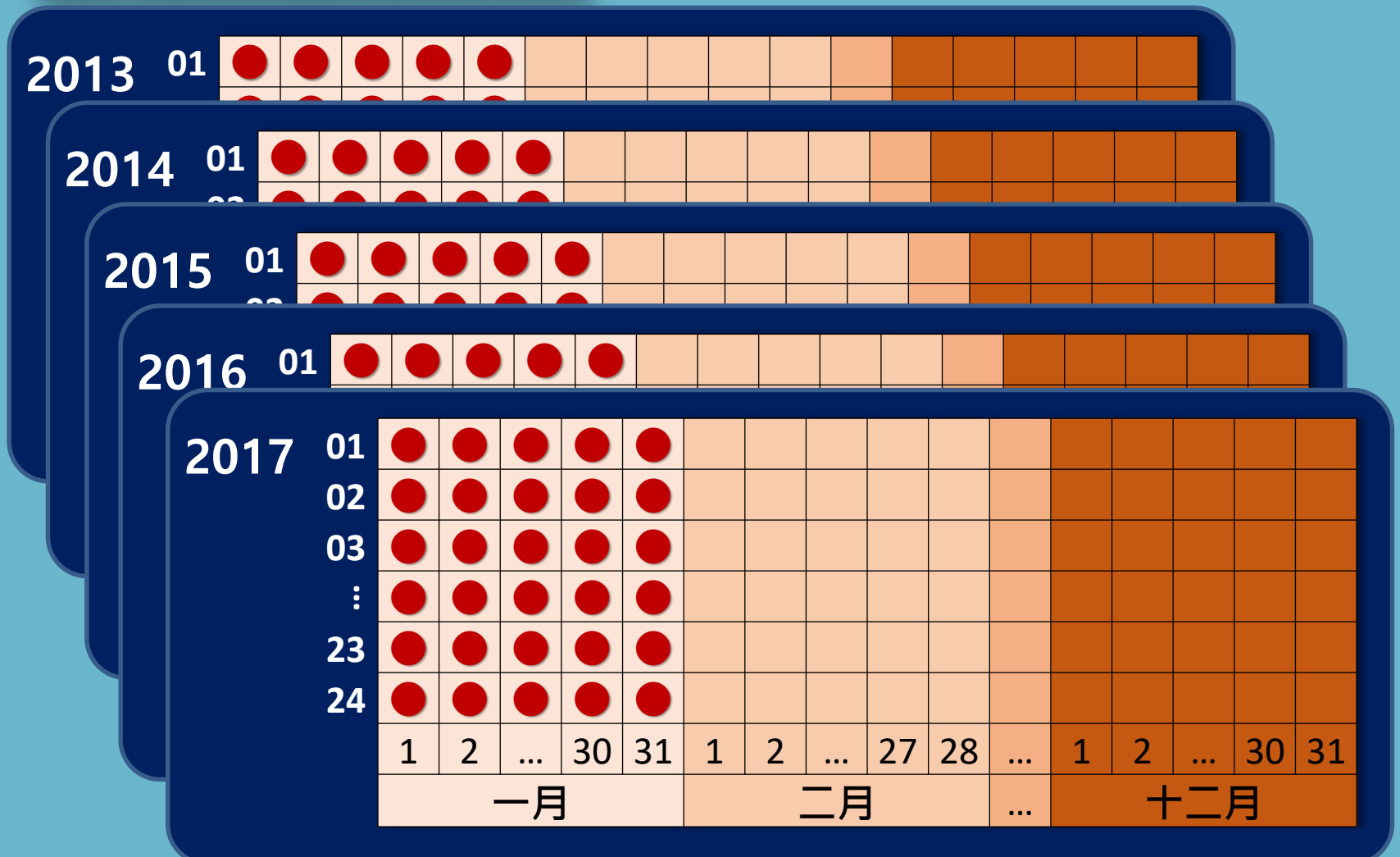
討論

2. 資料選取方式


2013 01
2014 01
2015 01
2016 01
2017
01
02
03
⋮
23
24
X
1 2 … 30 31 1 2 … 27 28 … 1 2 … 30 31
一月
二月
…
十二月

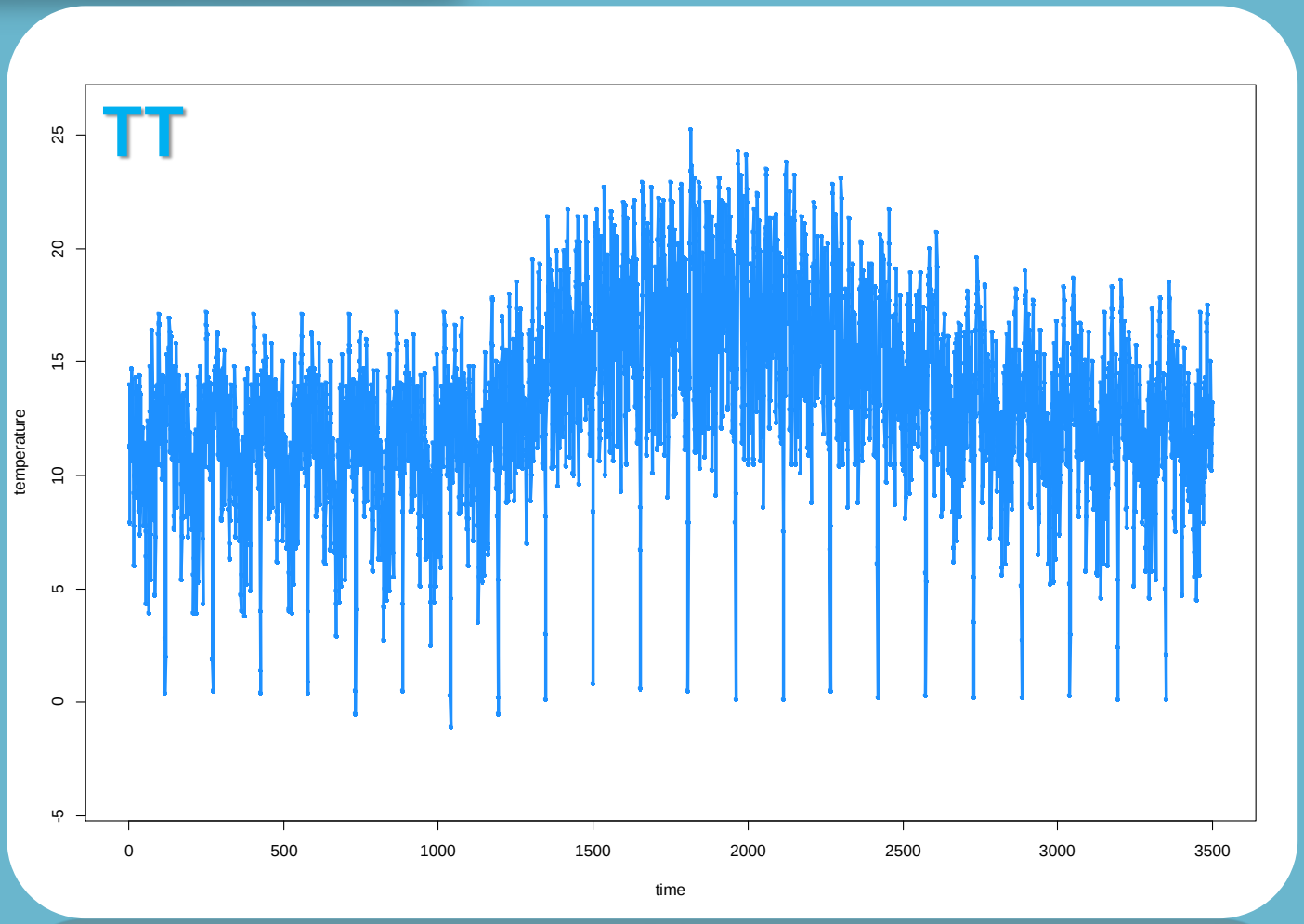
討論

2. 資料選取方式


2013
01
2014
01
2015
01
2016
01
2017
01
02
03
⋮
23
24
1 2 … 30 31 1 2 … 27 28 … 1 2 … 30 31
一月
二月
…
十二月

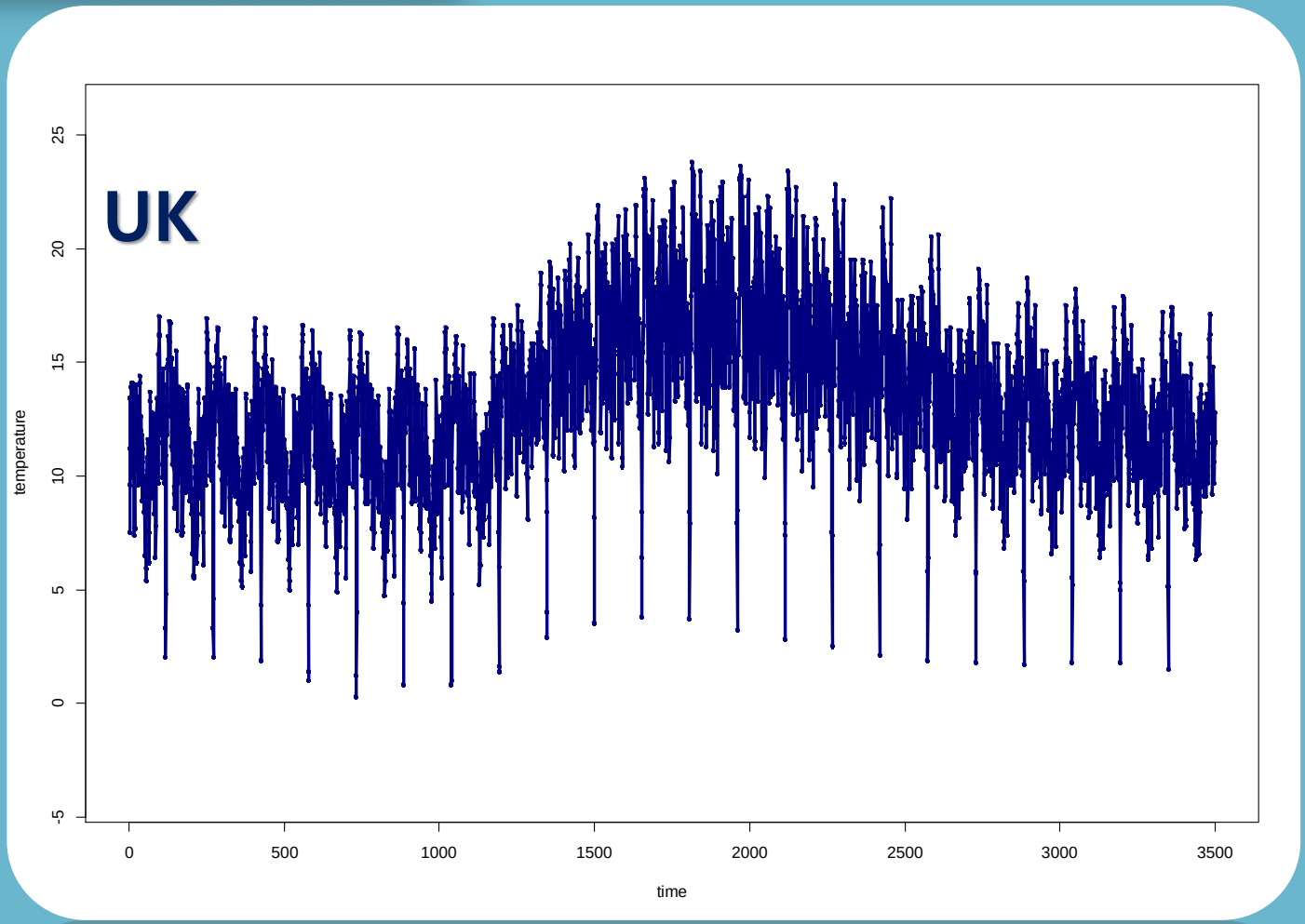
討論

2. 資料選取方式


TT
temperature
25
20
15
10
5
0
-5
0
500
1000
1500
2000
2500
3000
3500
time

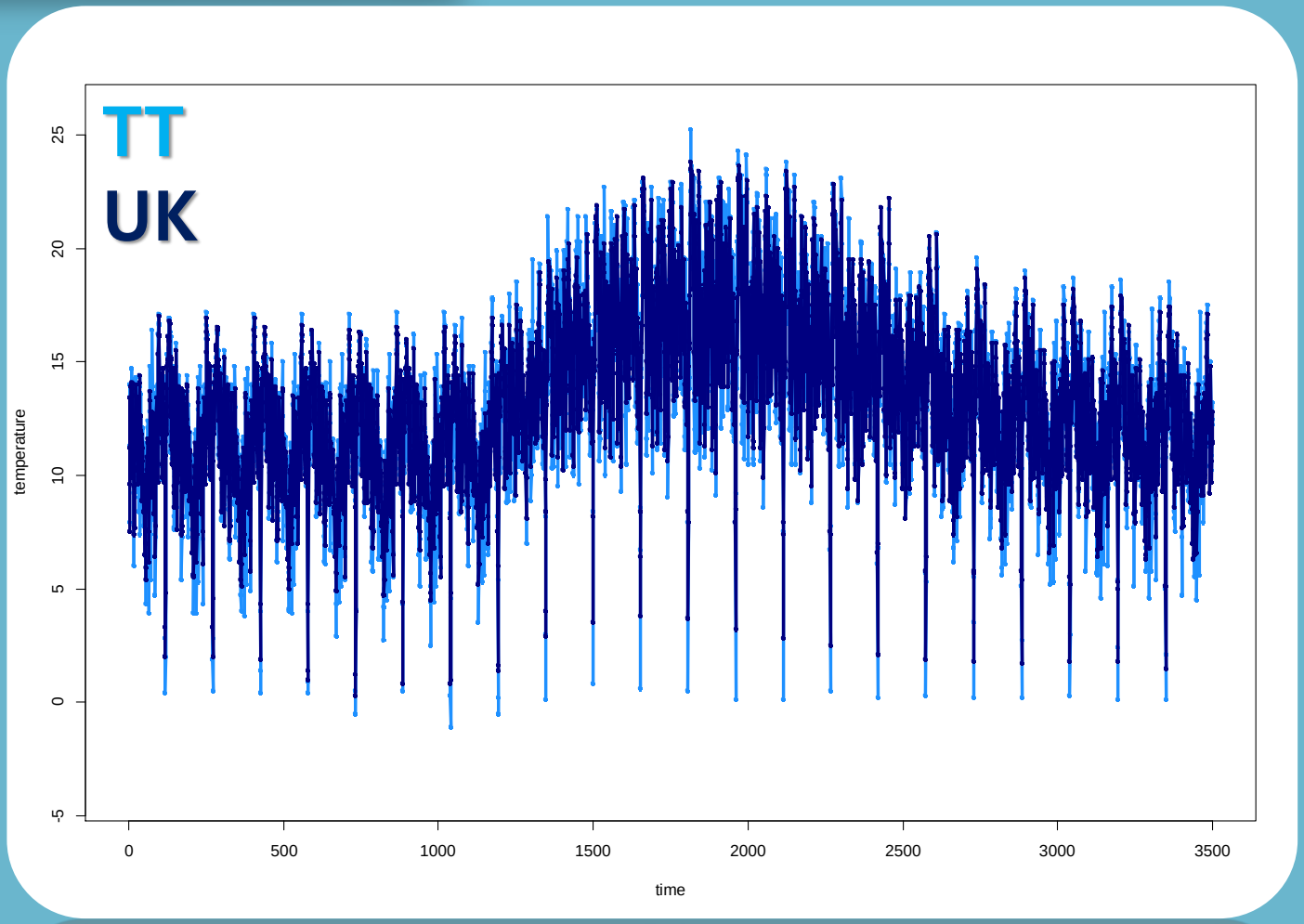
討論

2. 資料選取方式


UK
temperature
time
-5
0
5
10
15
20
25
0
500
1000
1500
2000
2500
3000
3500

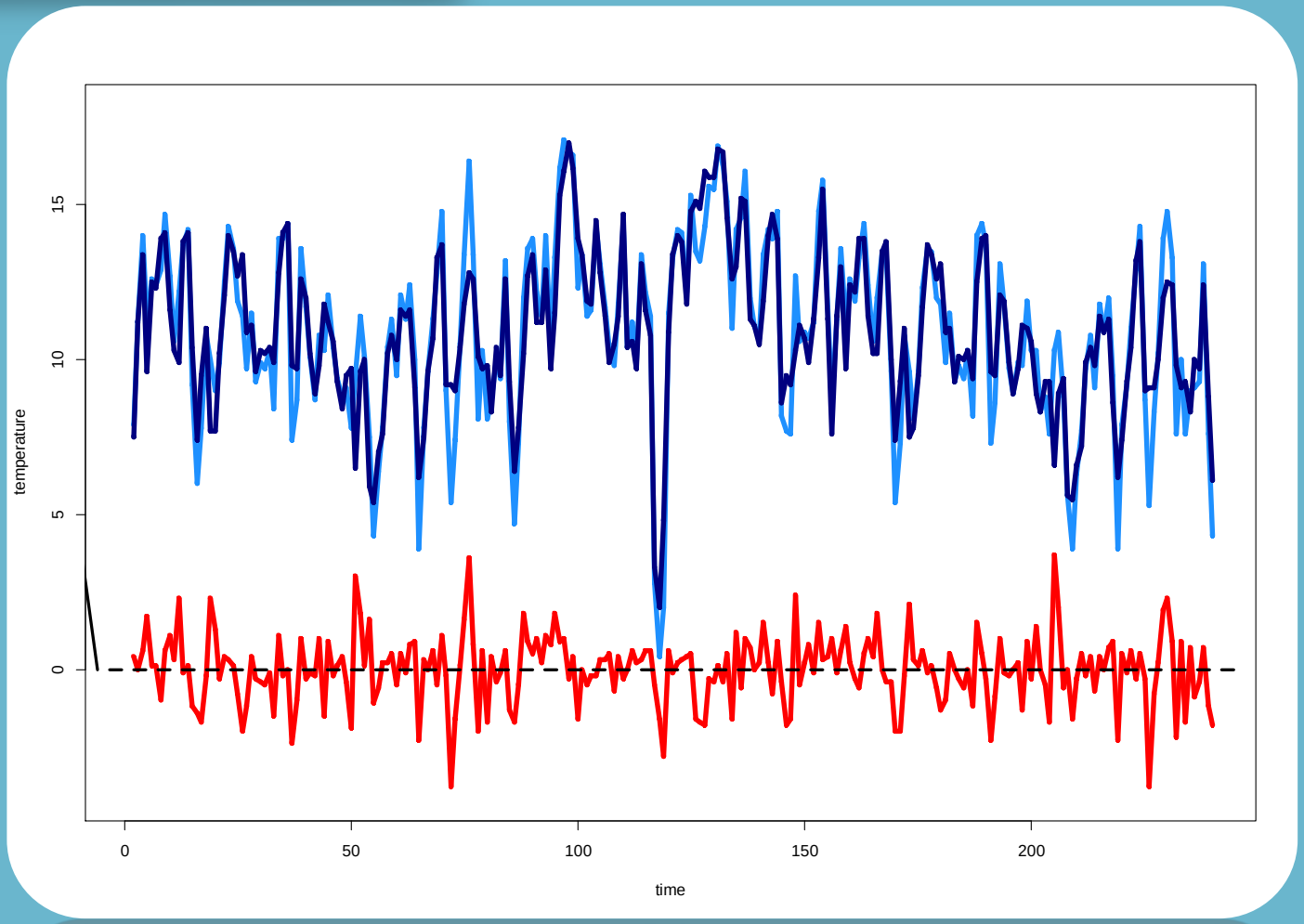
討論

2. 資料選取方式


TT
UK
temperature
time
-5
0
5
10
15
20
25
0
500
1000
1500
2000
2500
3000
3500

討論

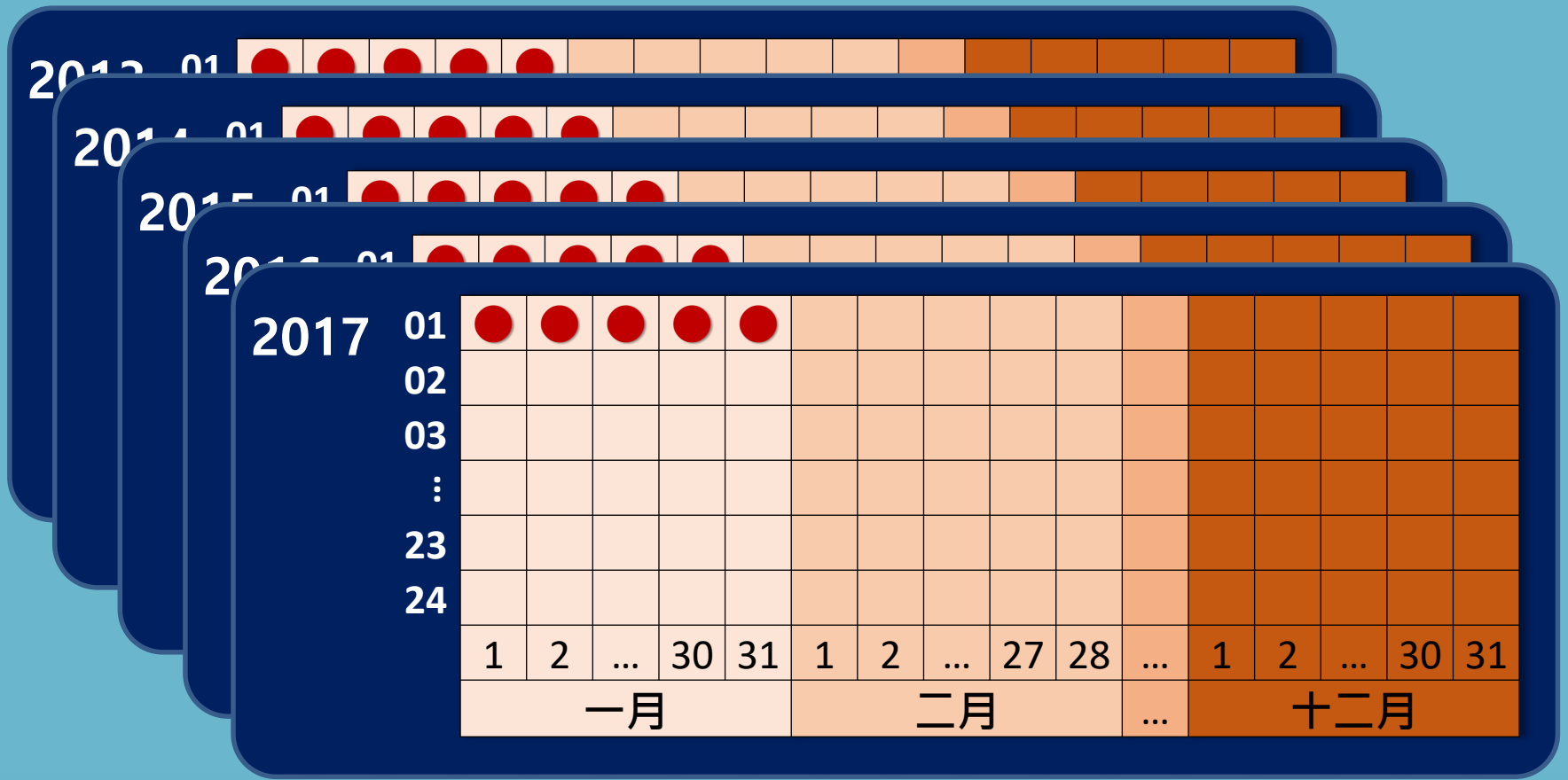
2. 資料選取方式



討論

3. 統計模型的假設

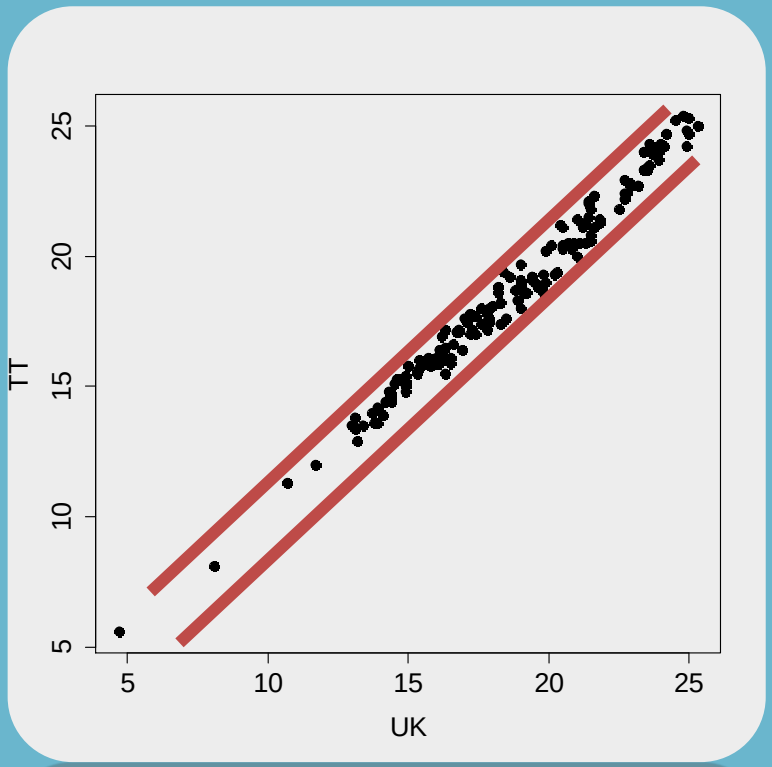
- 迴歸模型的假設: 資料點獨立性

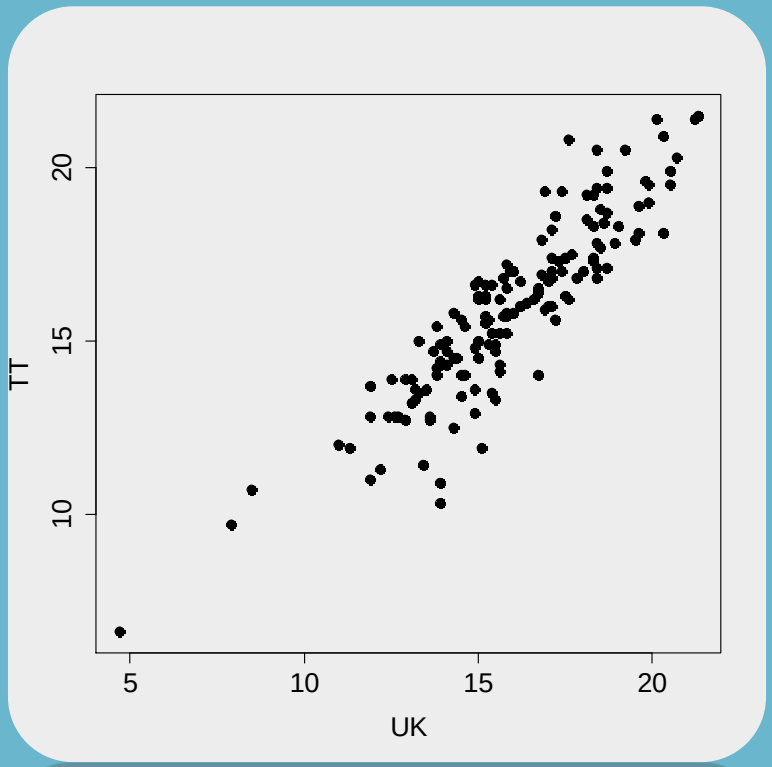


討論

3. 統計模型的假設

- 迴歸模型的假設: 變異同質性





Thank you