

巨量資料探勘與統計應用 W13

# 行為順序檢定： 滯後序列分析

布丁布丁吃布丁

<http://blog.pulipuli.info/>



資料檢定級

類別⇨類別資料

序列分析





投籃

傳球

運球

# 籃球的三重威脅

## 防守者難以猜透下一步



## 奇跡世代籃球高手的下一步 如何預測？





唯有「天帝之眼」？



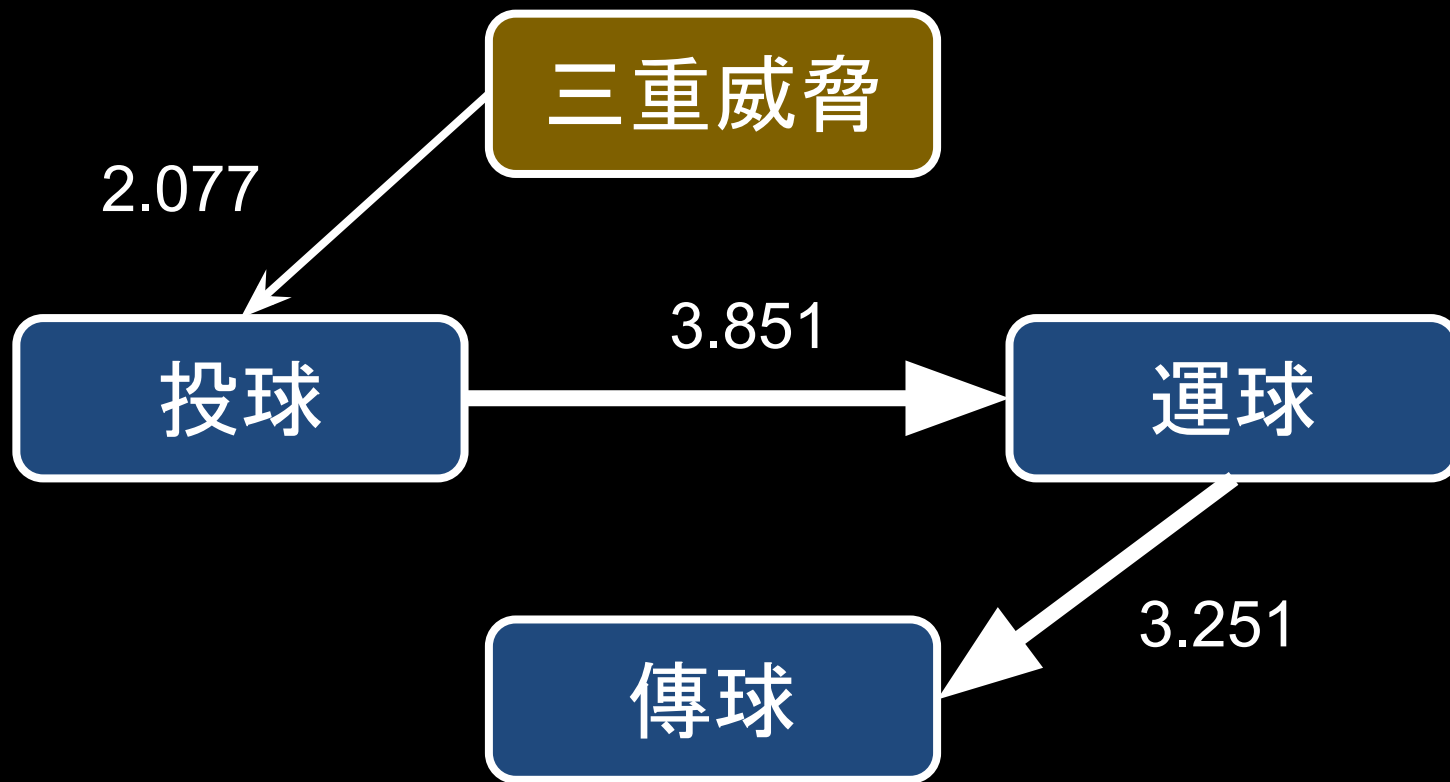
唯有「天帝之眼」？



## 資料分析的「天帝之眼」： 序列分析

		後續動作			
		三重	投球	傳球	運球
前一動作	三重威脅	0	14	3	0
	投球	0	0	1	3
	傳球	0	10	0	3
	運球	0	5	7	0

資料分析的「天帝之眼」：  
序列分析



資料分析的「天帝之眼」：  
序列分析



# 本週課程大綱

1. 行為序列資料表
2. 滯後序列分析
3. 滯後序列分析:樣本統計量  
事件轉移表
4. 滯後序列分析:檢定統計量  
調整後的殘差
5. 實作:序列分析
6. 深入探討序列分析
7. 課堂練習:  
幼兒平行遊戲行為研究
8. 序列分析的應用



我只是影子...  
未達顯著水準...

虛無方



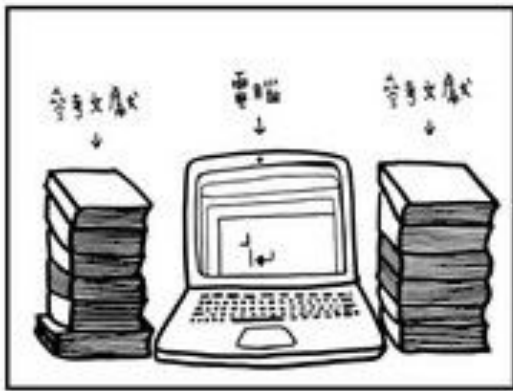
你抬頭的高度  
已達顯著水準!

對立方

Part 1.

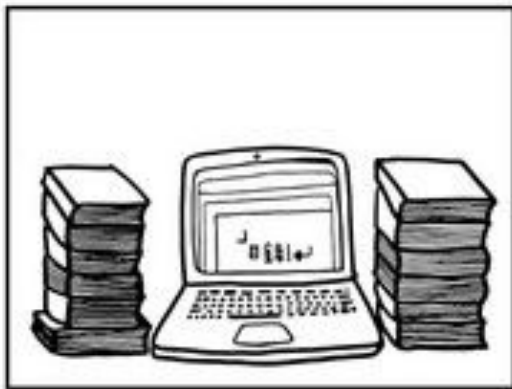
# 行為序列資料表

# 大學生寫報告的行為 (1/3)



**[20:00]**

首先把Word打開



**[23:00]**

三小時過後，  
只有「目錄」兩字

# 大學生寫報告的行為 (2/3)



**[23:05]**

你的腦袋一片空白，  
就像你做任何事情一樣。



**[23:10]**

這個報告檔案  
根本是你的人生寫照



# 大學生寫報告的行為 (3/3)



**[23:15]**

力不從心，一事無成。



**[23:20]**

而且沒有人愛你。

# 行為編碼：事件



事件編碼	說明
A	寫報告
B	停止寫報告
C	發愣
D	用手勢表現難過
E	用身體表現難過

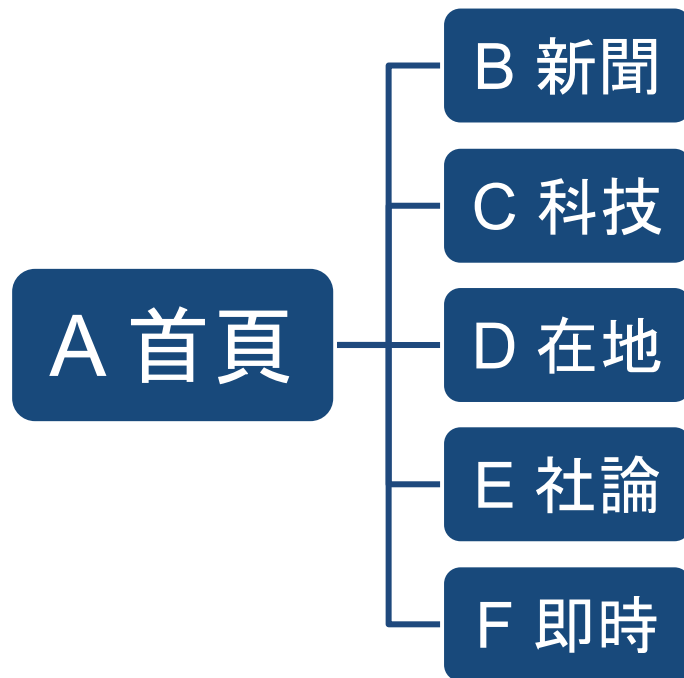
# 行為序列資料



研究對象 編號	序列 編號	事件
1	1	A
1	2	B
1	3	C
1	4	D
1	5	D
1	6	E

# 網站瀏覽行為：多研究對象分析

## MSNBC新聞網



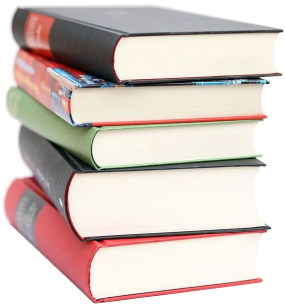
## 行為序列資料表

研究對象編號	序列編號	事件
1	1	A
1	2	A
2	1	B
3	1	C
3	2	B
3	3	B



# 圖書借閱行為：同時間多事件分析

## 圖書館藏



A 統計學的世界

B 量化研究法

C 資料探勘

D 壞壞總裁愛...

E 總裁一家逼...

F 報復總裁大...



## 行為序列資料表

研究對象編號	序列編號	事件
1	1	D
1	2	E
2	1	A;B
2	1	E
2	2	D
3	3	A;B;C

# 行為序列資料表如何而來？

## 觀察研究法



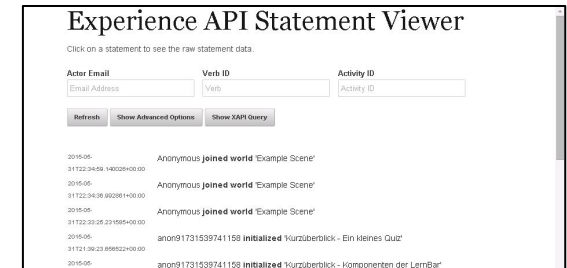
- 現場觀察與錄影記錄
- 研究者需即時與主動地蒐集資料

## 日誌研究法



- 受試者自行記錄行為資料

## 內容分析研究法



- 側錄或記錄受試者的行為歷程，事後分析
  - 交易資料
  - 借閱資料
  - 網站瀏覽記錄

# 網站內容分析

## Apache網站記錄

```
jkreps-mn:~ jkreps$ tail -f -n 20 /var/log/apache2/access_log
::1 - - [23/Mar/2014:15:07:00 -0700] "GET /images/apache_feather.gif HTTP/1.1" 200 4128
::1 - - [23/Mar/2014:15:07:04 -0700] "GET /images/producer_consumer.png HTTP/1.1" 200 86
::1 - - [23/Mar/2014:15:07:04 -0700] "GET /images/log_anatomy.png HTTP/1.1" 200 19579
::1 - - [23/Mar/2014:15:07:04 -0700] "GET /images/consumer-groups.png HTTP/1.1" 200 268
::1 - - [23/Mar/2014:15:07:04 -0700] "GET /images/log_compaction.png HTTP/1.1" 200 4141
::1 - - [23/Mar/2014:15:07:04 -0700] "GET /documentation.html HTTP/1.1" 200 189893
::1 - - [23/Mar/2014:15:07:04 -0700] "GET /images/log_cleaner_anatomy.png HTTP/1.1" 200
::1 - - [23/Mar/2014:15:07:04 -0700] "GET /images/kafka_log.png HTTP/1.1" 200 134321
::1 - - [23/Mar/2014:15:07:04 -0700] "GET /images/mirror-maker.png HTTP/1.1" 200 17054
::1 - - [23/Mar/2014:15:08:07 -0700] "GET /documentation.html HTTP/1.1" 200 189937
::1 - - [23/Mar/2014:15:08:07 -0700] "GET /styles.css HTTP/1.1" 304 -
::1 - - [23/Mar/2014:15:08:07 -0700] "GET /images/kafka_logo.png HTTP/1.1" 304 -
::1 - - [23/Mar/2014:15:08:07 -0700] "GET /images/producer_consumer.png HTTP/1.1" 304 -
::1 - - [23/Mar/2014:15:08:07 -0700] "GET /images/log_anatomy.png HTTP/1.1" 304 -
::1 - - [23/Mar/2014:15:08:07 -0700] "GET /images/consumer-groups.png HTTP/1.1" 304 -
::1 - - [23/Mar/2014:15:08:07 -0700] "GET /images/log_cleaner_anatomy.png HTTP/1.1" 304
::1 - - [23/Mar/2014:15:08:07 -0700] "GET /images/log_compaction.png HTTP/1.1" 304 -
::1 - - [23/Mar/2014:15:08:07 -0700] "GET /images/kafka_log.png HTTP/1.1" 304 -
::1 - - [23/Mar/2014:15:08:07 -0700] "GET /images/mirror-maker.png HTTP/1.1" 304 -
::1 - - [23/Mar/2014:15:09:55 -0700] "GET /documentation.html HTTP/1.1" 200 195264
```



# 網站內容分析

## Google分析





Part 2.

# 滯後序列分析

# Lag-Sequential Analysis

## 滯後序列分析介紹



Roger Bakeman

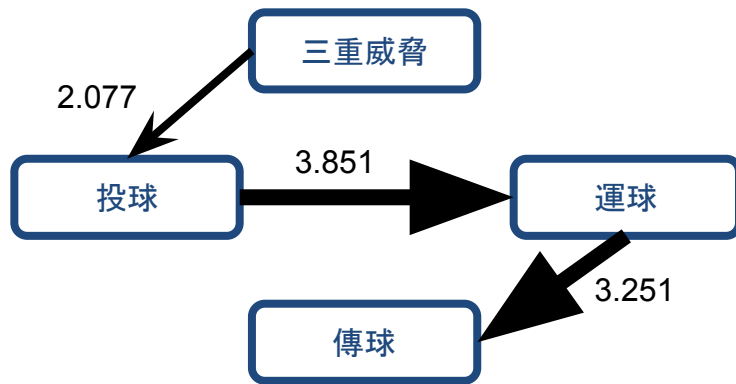
**貝克曼**

美國喬治亞州立大學教授

- 最早是由Sackett,G.P於1974年首次提出
- Bakeman等人出書詳細介紹滯後序列分析的用法
- 分析目的：
  - 分析行為轉移的序列出現次數是否多到呈現顯著
  - 繪製事件轉移圖呈現行為模式

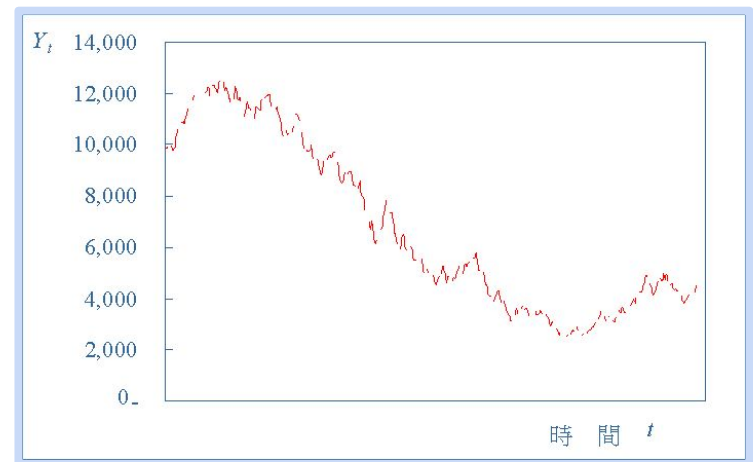
# 序列分析的差異

## 滯後序列分析



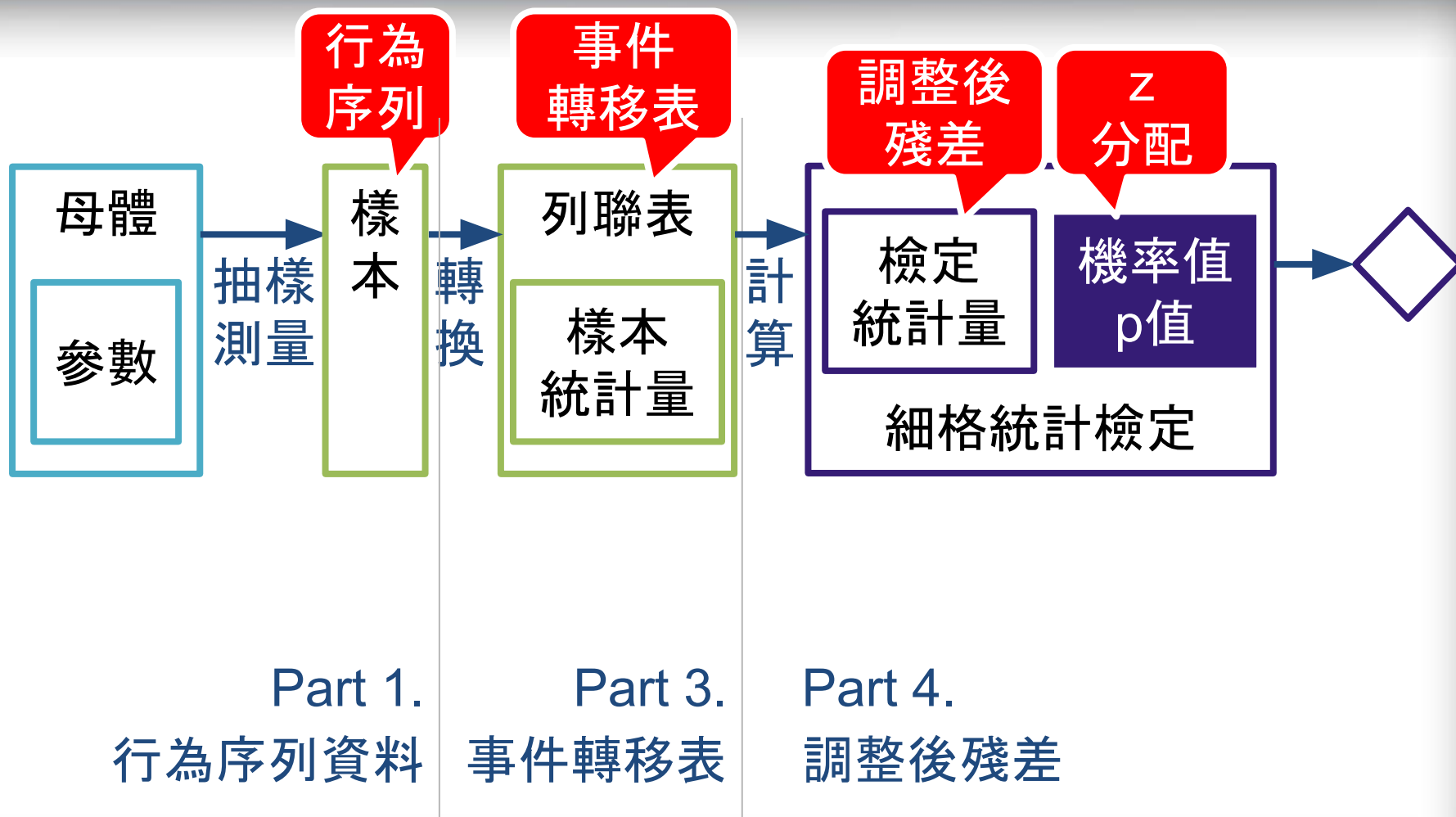
- 處理類別資料的「事件」
- 分析前一步與下一步的事件轉移關係

## 時間序列分析



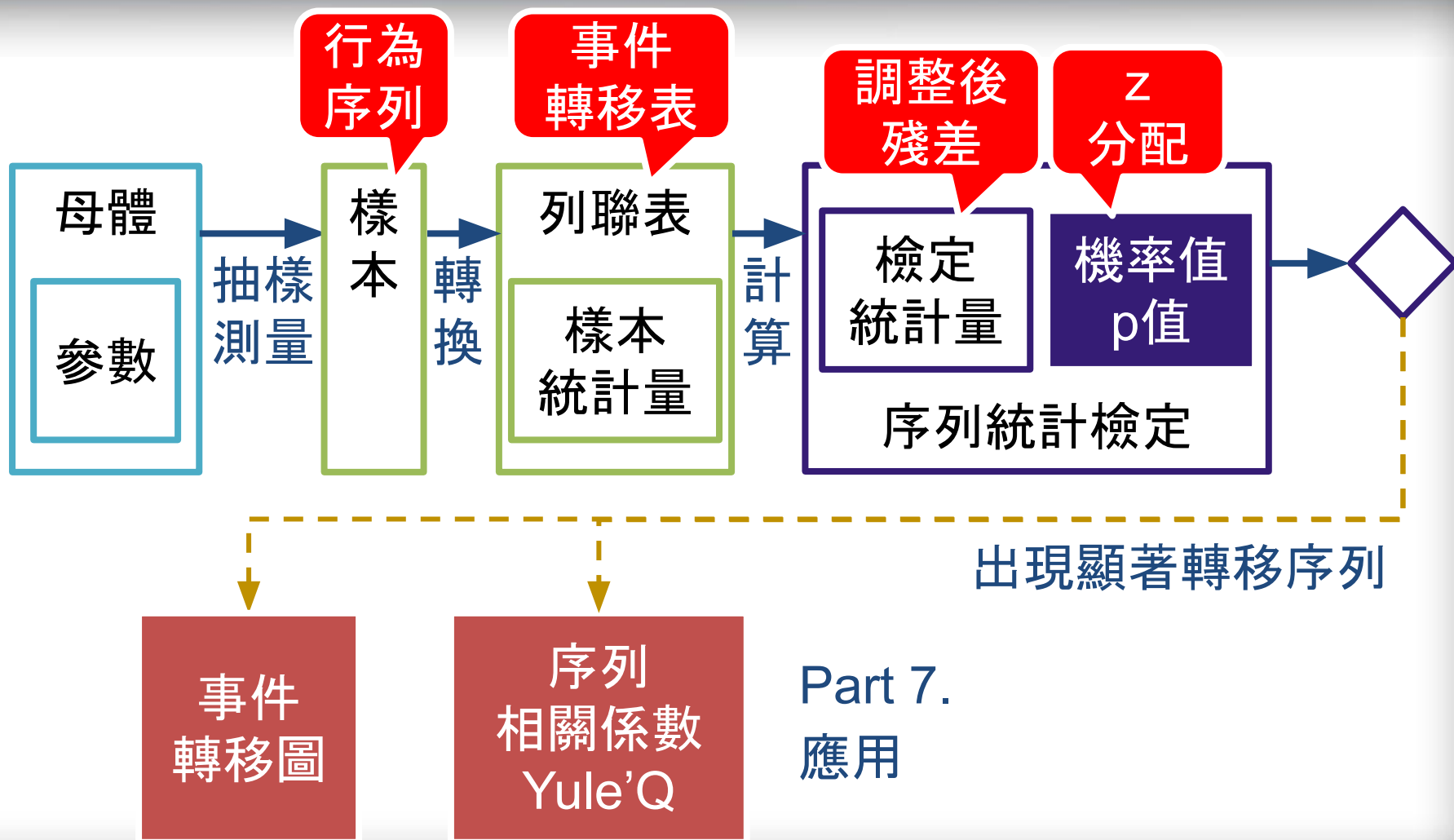
- 處理連續資料的「效果量」
- 分析趨勢、預測為主
- 考量季節與週期性變動

# 滯後序列分析的流程：統計檢定





# 滯後序列分析的流程：視覺化與應用



# 序列分析計算器

### Input

文字輸入 選擇檔案輸入 Google試算表發布連結

請上傳CSV檔案：(範例檔案1下載, 範例檔案2下載)

選擇檔案 未選擇任何檔案

### 動態分類資料下載

滯後序列長度:

動態分類器檔案 貝氏網路配置權(平均) 貝氏網路配置權(多層)

循序模式探勘: R腳本檔案: arulesSequences.r

Lag 1 (t)

A

C

D

Lag 0 (g)

A

B

C

D



## 序列分析 計算器

Part 3.

# 滯後序列分析：樣本統計量 事件轉移表

# 某動物園有 3 個地區

A 熱帶莽原區



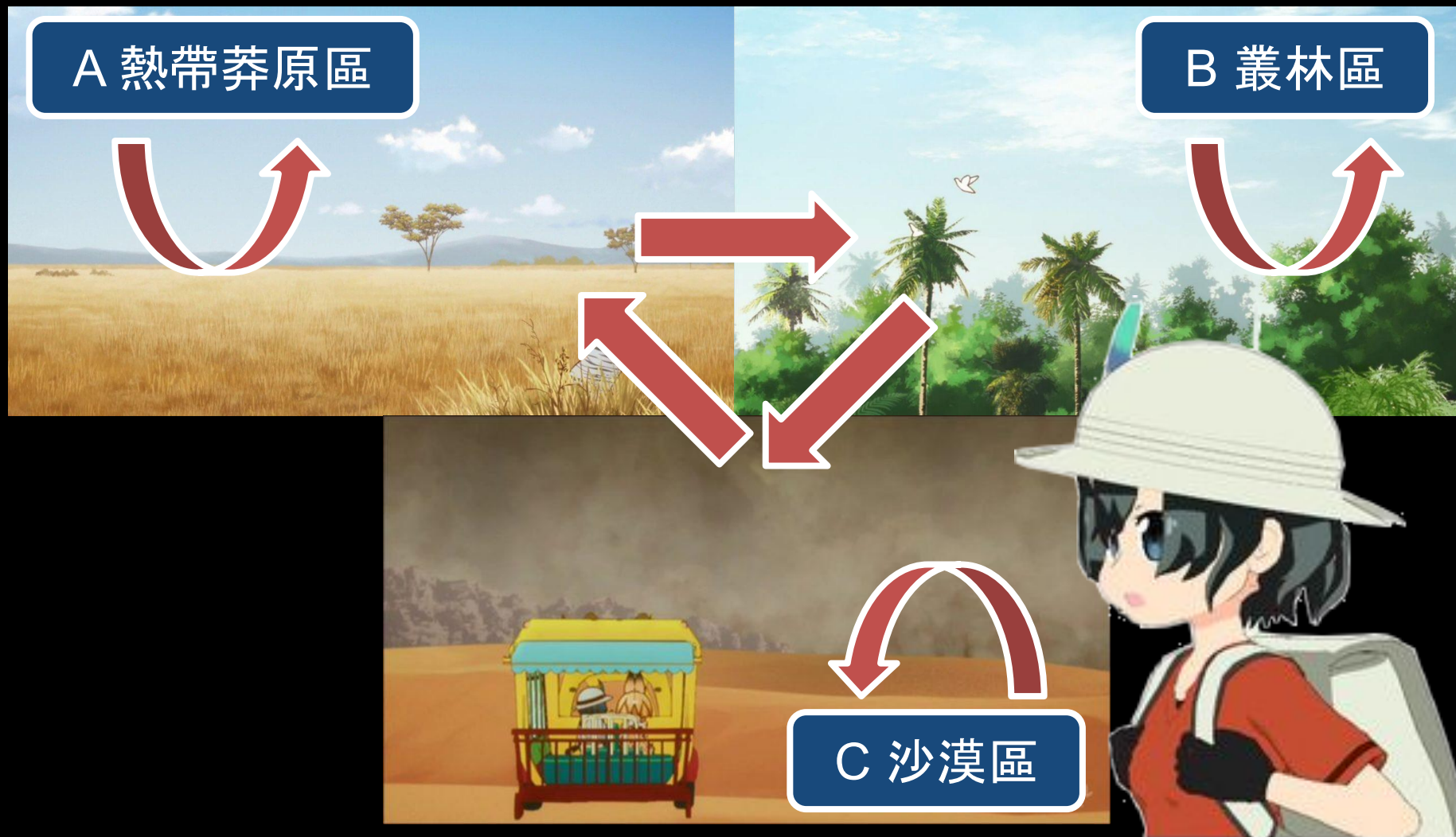
B 叢林區



C 沙漠區



# 遊客在動物園裡面走動





A 熱帶莽原區



B 叢林區



如果能夠分析遊客的路線的話  
就能規劃有效的導覽介紹了

研究對象編號	序列編號	事件
1	1	A
1	2	A
1	3	B
1	4	C
1	5	A
1	6	C
1	7	A
1	8	B
1	9	B
1	10	C

# 遊客所在區域 事件序列資料 表

案例數量:10筆



# 雙事件轉移序列次數統計

## 事件轉移表



- Lag 0 = g 給定事件  
= 目前事件
- Lag 1 = t 目的事件  
= 下一個事件

		Lag 1 (t)			Lag 0 總數
		A	B	C	
Lag 0 (g)	A	0	0	0	0
	B	0	0	0	0
	C	0	0	0	0
Lag 1 總數		0	0	0	0

# 事件

A

A

B

C

A

C

A

B

B

C

		Lag 1 (t)			Lag 0 總數
		A	B	C	
Lag 0 (g)	A	1	0	0	1
	B	0	0	0	0
	C	0	0	0	0
Lag 1 總數		1	0	0	1

A在  
Lag 0  
的總數

A在Lag 1的總數

序列總數



事件
A
A
B
C
A
C
A
B
B
C

		Lag 1 (t)			Lag 0 總數
		A	B	C	
Lag 0 (g)	A	1	1	0	2
	B	0	0	0	0
	C	0	0	0	0
Lag 1 總數		1	1	0	2

A在  
Lag 0  
的總數

B在Lag 1的總數

序列總數

事件
A
A
B
C
A
C
A
B
B
C

		Lag 1 (t)			Lag 0 總數
		A	B	C	
Lag 0 (g)	A	1	1	0	2
	B	0	0	1	1
	C	0	0	0	0
Lag 1 總數		1	1	1	3

B在  
Lag 0  
的總數

C在Lag 1的總數

序列總數

事件
A
A
B
C
A
C
A
B
B
C

		Lag 1 (t)			Lag 0 總數
		A	B	C	
Lag 0 (g)	A	1	2	1	4
	B	0	1	2	3
	C	2	0	0	2
Lag 1 總數		3	3	3	9

B在  
Lag 0  
的次數

C在Lag 1的次數

序列總數

# 事件

A

A

B

C

A

C

A

B

B

C

## 事件轉移表完成！



		Lag 1 (t)			Lag 0 總數
		A	B	C	
Lag 0 (g)	A	1	2	1	4
	B	0	1	2	3
	C	2	0	0	2
Lag 1 總數		3	3	3	9

# 計算序列出現機率

$$\text{序列機率 } p(g, t) = \frac{\text{序列次數 } f(g, t)}{\text{序列總數 } N_s}$$

$$= \frac{1}{9} = 0.111$$

		Lag 1 (t)			Lag 0 總數
		A	B	C	
Lag 0 (g)	A	1	2	1	4
	B	0	1	2	3
	C	2	0	0	2
Lag 1 總數		3	3	3	9

序列總數

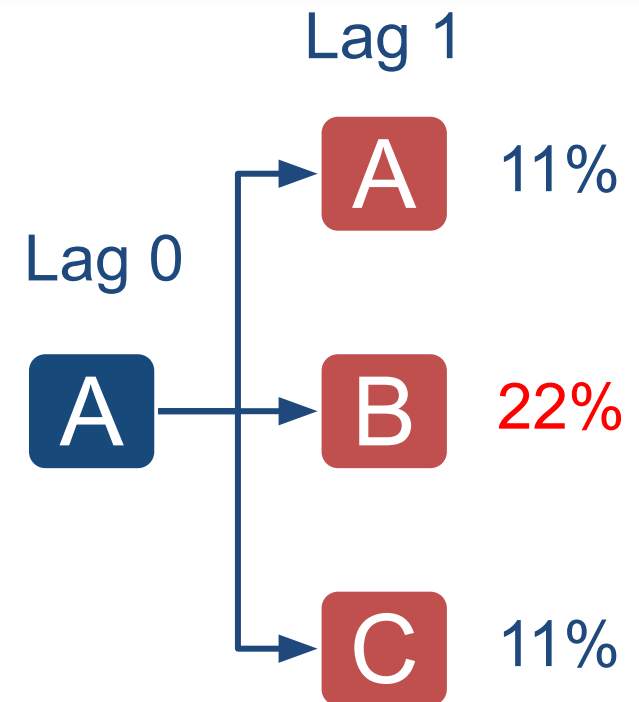


# 事件轉移機率表完成

			Lag 1 (t)			Lag 0 總數
			A	B	C	
Lag 0 (g)	A	頻率	1	2	1	4
		機率	0.11	0.22	0.11	
	B	頻率	0	1	2	3
		機率	0	0.11	0.22	
	C	頻率	2	0	0	2
		機率	0.22	0	0	
	Lag 1 總數		3	3	3	9

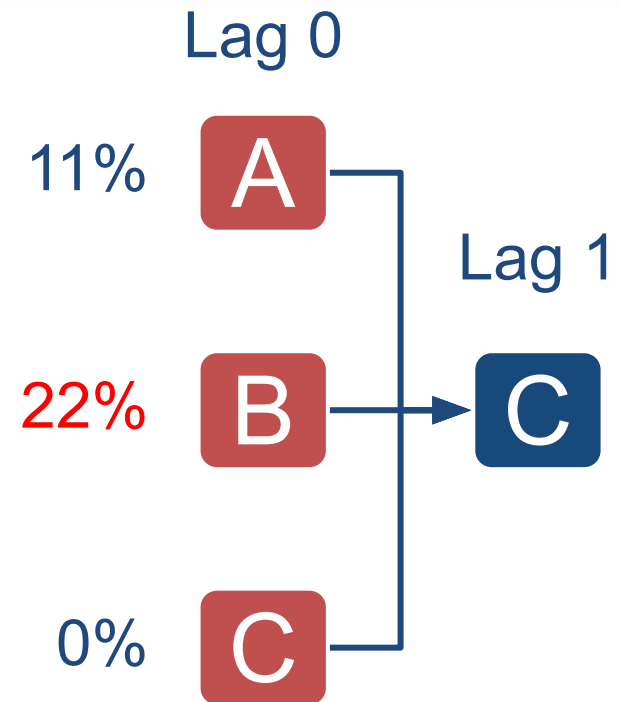
# 預測：已知 Lag 0, 猜測 Lag 1

			Lag 1 (t)		
			A	B	C
Lag 0 (g)	A	頻率	1	2	1
		機率	0.11	0.22	0.11
	B	頻率	0	1	2
		機率	0	0.11	0.22
	C	頻率	2	0	0
		機率	0.22	0	0



# 推理：已知 Lag 1, 猜測 Lag 0

			Lag 1 (t)		
			A	B	C
Lag 0 (g)	A	頻率	1	2	1
		機率	0.11	0.22	0.11
	B	頻率	0	1	2
		機率	0	0.11	0.22
	C	頻率	2	0	0
		機率	0.22	0	0



Part 4.

# 滯後序列分析：檢定統計量 調整後的殘差

BOSS有能!



		Lag 1 (t)			Lag 0 總數
		A	B	C	
Lag 0 (g)	A	1	2	1	4
	B	0	1	2	3
	C	2	0	0	2
Lag 1 總數		3	3	3	9

$[A \rightarrow B] [B \rightarrow C] [C \rightarrow A]$   
序列轉移次數真的比較多嗎？





# 卡方檢定!

- ☒ 在符合以下的情況時，使用費雪爾正確概率檢定 (Fisher's Exact Probability Test)：1. 2x2列聯表；2. 有細格期望值小於5個；3. 各類別行列總數皆小於20個。
- ☒ 在符合以下的情況時，使用葉氏連續性校正 (Yate's continuity correction)：1. 2x2列聯表；2. 有細格期望值小於5個；3. 樣本總數超過20個。
- ☐ 使用無樣式表格(容易複製到其他文件)

## Result

			Lag 1			列總合
			A	B	C	
Lag 0	A	個數	1	2	1	4
		期望個數	1.1	1.8	1.1	4
		殘差	-0.1	0.2	-0.1	
		標準化殘差	-0.1	0.1	-0.1	
		調整後殘差	-0.1	0.2	-0.1	
	B	個數	0	1	2	3
		期望個數	0.8	1.4	0.8	3
		殘差	-0.8	-0.4	1.2	
		標準化殘差	-0.9	-0.3	1.3	
		調整後殘差	-1.2	-0.5	1.8	
	C	個數	2	2	0	4
		期望個數	1.1	1.8	1.1	4
		殘差	0.9	0.2	-1.1	
		標準化殘差	0.9	0.1	-1.0	
		調整後殘差	1.3	0.2	-1.5	
行總合	個數	3	5	3	11	
	期望個數	3				



卡方檢定結果：

- 卡方檢定統計量 $\chi^2 = 4$   
此無法拒絕虛無假設，  
響。

## 卡方檢定 計算器



## 卡方檢定

		Lag 1 (t)			Lag 0 總數
		A	B	C	
Lag 0 (g)	A	1	2	1	4
	B	0	1	2	3
	C	2	0	0	2
Lag 1 總數		3	3	3	9

卡方檢定統計量  $\chi^2 = 4.522$  ,  $p$ 值 = 0.340  
 未達  $\alpha = 0.05$  的顯著水準

# 卡方檢定的限制

		Lag 1 (t)			Lag 0 總數
		A	B	C	
Lag 0 (g)	A	1	2	1	4
	B	0	1	2	3
	C	2	0	0	2
Lag 1 總數		3	3	3	9

卡方檢定適用場合

- 樣本數量最好在30以上
- 0細格數量不可超過細格總數的1/4

然而，

事件轉移表很多細格會是0

# 期望次數與殘差的計算



期望次數

$$f(g)p(t)$$

殘差

期望次數

$$f(g, t) - f(g)p(t)$$

期望次數：理論上這個序列  
應該要出現的次數

- **f(g)**: 事件g 的總數
- **p(t)**: 發生事件t的機率

殘差：這個序列實際次數跟  
理論次數之間的差距

- **f(g,t)**: 事件g 到事件t的次數

# 檢定統計量：調整後殘差

調整後殘差  $z = \frac{f(g, t) - f(g)p(t)}{\sqrt{f(g)p(t)[1 - p(t)][1 - p(g)]}}$

Lag 0 (g)

序列：



Lag 1 (t)

1. **f(g,t)**: 事件g 到事件t的次數
2. **f(g)**: 事件g 的總數
3. **p(g)**: 發生事件g的機率
4. **p(t)**: 發生事件t的機率

# 計算調整後殘差 (1/4) $f(g,t)$ 跟 $f(g)$

Lag 0 (g)



Lag 1 (t)

		Lag 1 (t)			Lag 0 總數
		A	B	C	
Lag 0 (g)	A	1	2	1	4
	B	0	1	2	3
	C	2	0	0	2
Lag 1 總數		3			9

$f(g,t)$  points to the cell (C, 2) in the table.

$f(g)$  points to the cell (C, 2) in the table.



# 計算調整後殘差 (2/4)

## $p(g)$

		Lag 1 (t)			Lag 0 總數
		A	B	C	
Lag 0 (g)	A	1	2	1	4
	B	0	1	2	3
	C	2	0	0	2
Lag 1 總數		3	3	3	9

$$p(g) = \frac{f(g)}{N_s}$$

$$= 2/9$$

$$= 0.222$$

$f(g)$

$N_s$  序列總數

# 計算調整後殘差 (3/4)

## $p(t)$

		Lag 1 (t)			Lag 0 總數
		A	B	C	
Lag 0 (g)	A	1	2	1	4
	B	2	1	2	3
	C	0	2	0	2
Lag 1 總數		3	3	3	9

$$p(t) = \frac{f(t)}{N_s}$$

$$= 3/9$$

$$= 0.333$$

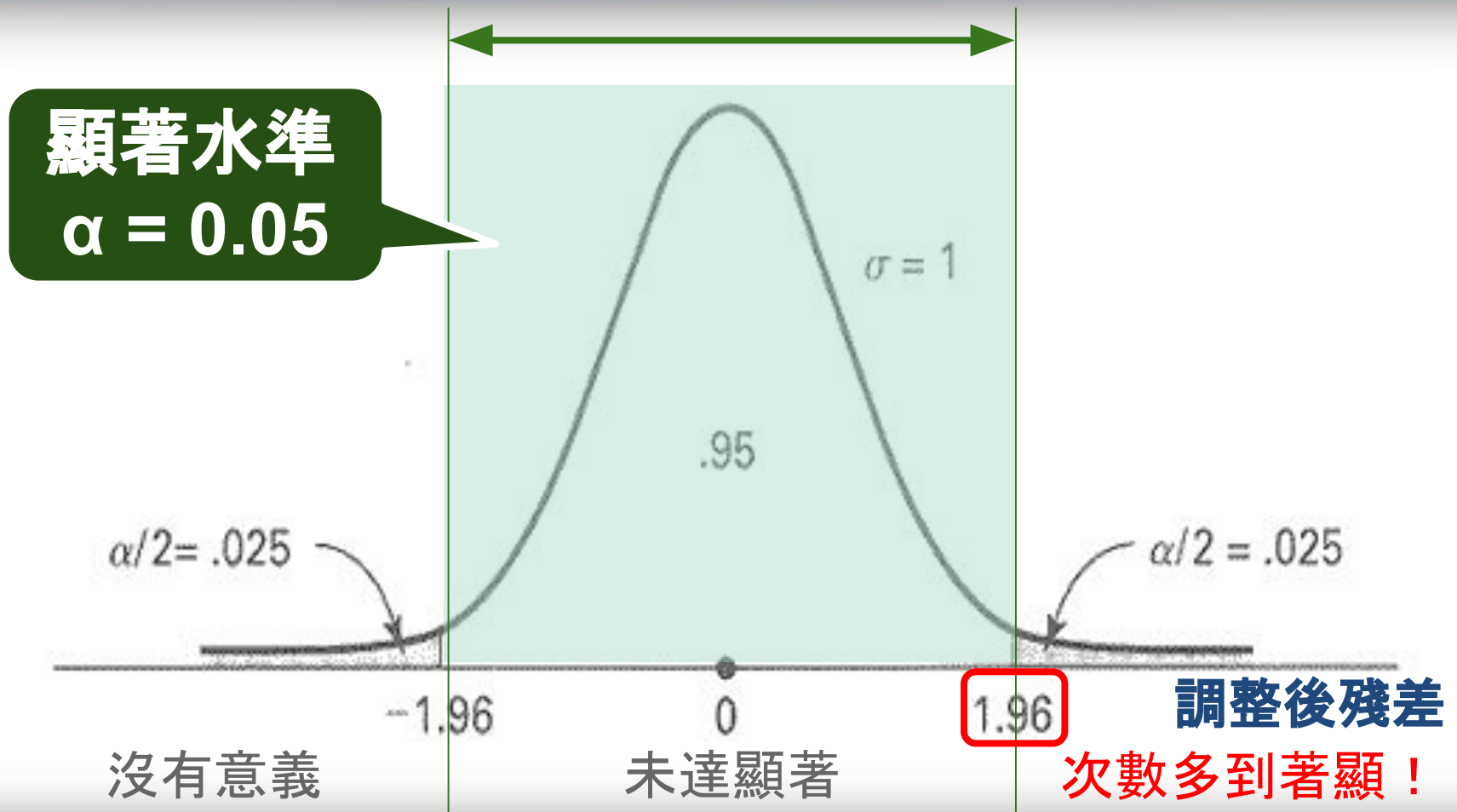
$N_s$

## 計算調整後殘差 (4/4)

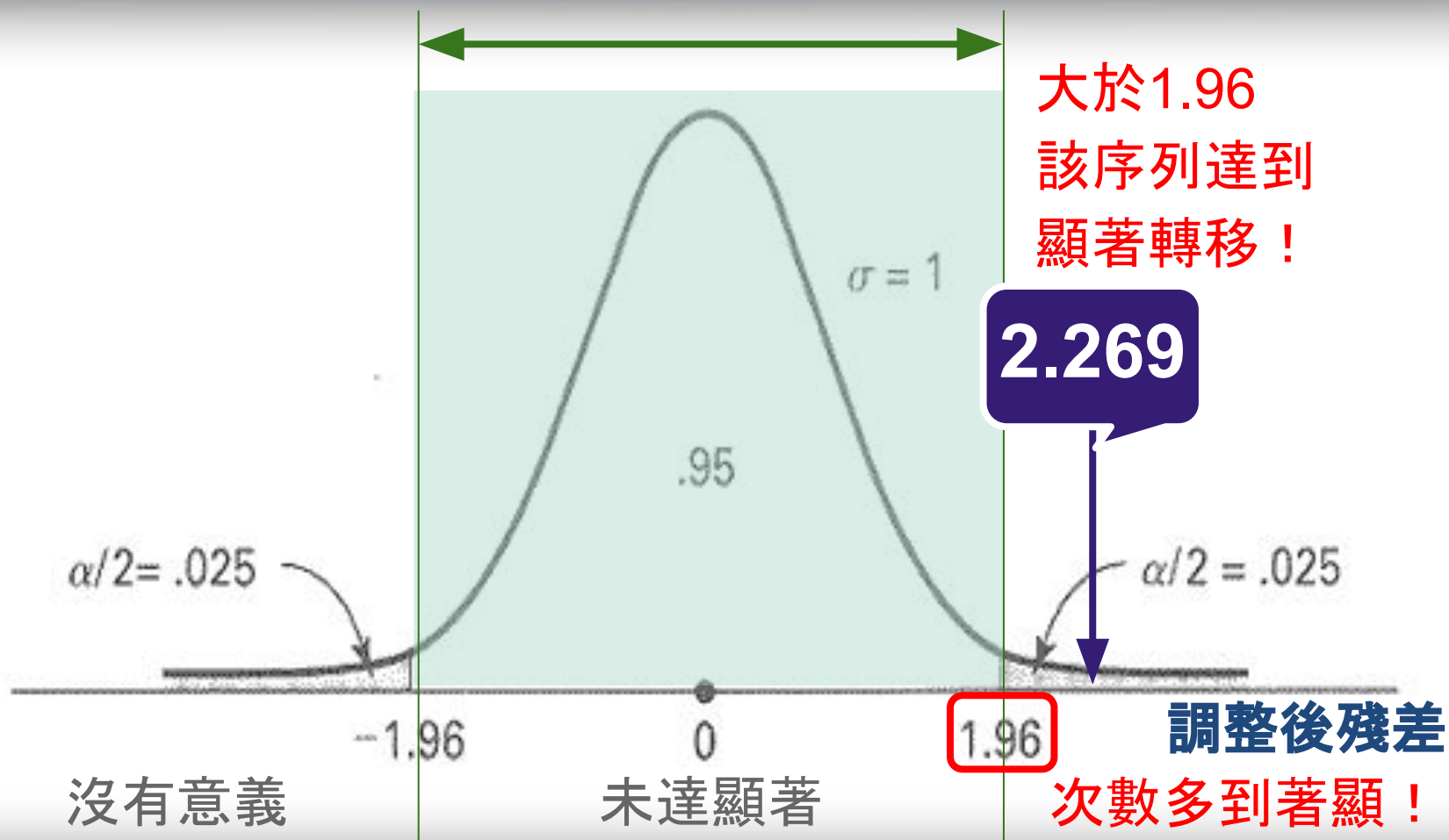
$$f(g,t)=2 \quad f(g)=2 \quad p(g)=0.222 \quad p(t)=0.333$$

$$\begin{aligned} \text{調整後殘差 } z &= \frac{f(g,t) - f(g)p(t)}{\sqrt{f(g)p(t)[1 - p(t)][1 - p(g)]}} \\ &= \frac{2 - 2 \times 0.333}{\sqrt{2 \times 0.333 \times [1 - 0.333] \times [1 - 0.222]}} \\ &= 2.269 \end{aligned}$$

# 顯著轉移檢定：z分佈



# 顯著轉移檢定：z分佈



# 調整後殘差表完成

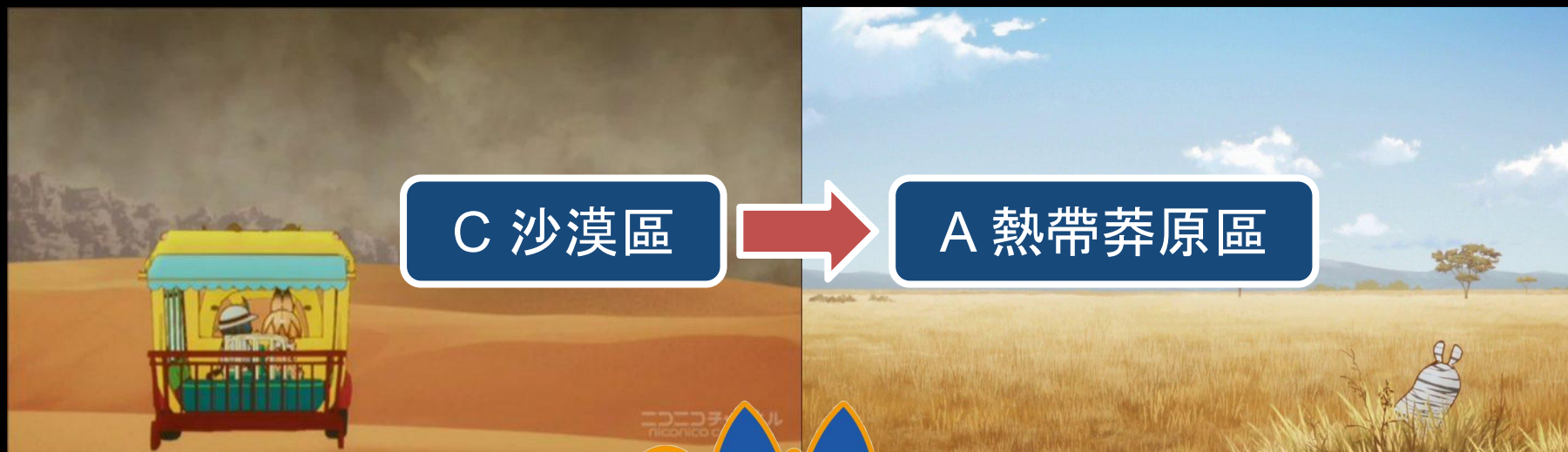
			Lag 1 (t)			Lag 0 總數
			A	B	C	
Lag 0 (g)	A	頻率	1	2	1	4
		調整後 殘差	-0.47	0.95	-0.47	
	B	頻率	0	1	2	3
		調整後 殘差	-1.50	0.00	1.50	
	C	頻率	2	0	0	2
		調整後 殘差	2.27	-1.13	-1.13	
Lag 1 總數			3	3	3	9



# 事件轉移圖

僅列出有顯著轉移的序列：





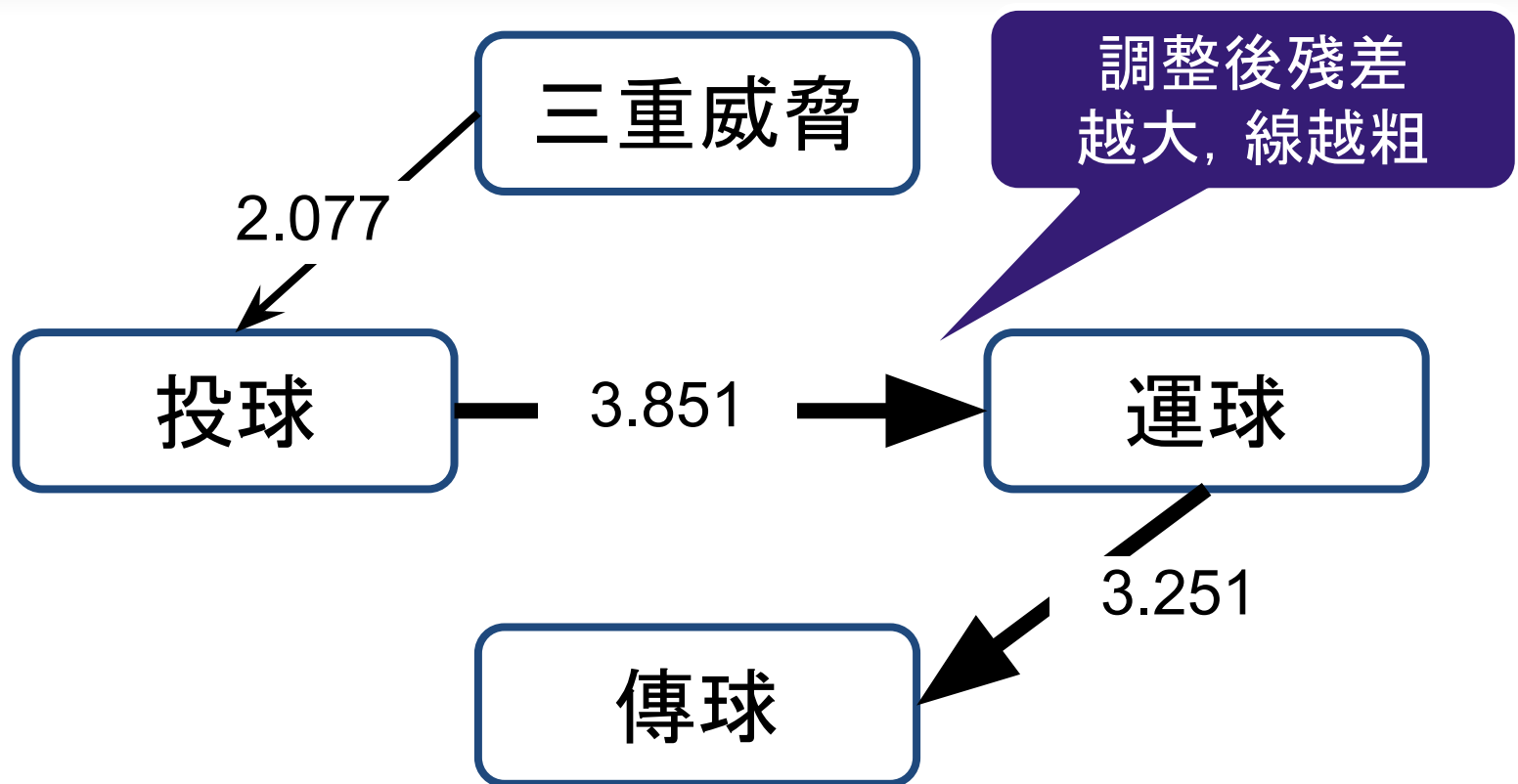
好~厲害！

這樣就可以規劃導覽了！

# 複雜的例子 分析三重威脅

			Lag 1 (t)			
			三重	投球	傳球	運球
Lag 0 (g)	三重	頻率	0	14	3	0
		調整後殘差	0.00	2.08	-0.76	-2.01
	投球	頻率	0	0	1	3
		調整後殘差	0.00	-2.73	0.05	3.85
	傳球	頻率	0	10	0	3
		調整後殘差	0.00	1.22	-2.39	1.27
	運球	頻率	0	5	7	0
		調整後殘差	0.00	-1.78	3.25	-1.56

# 複雜的例子 分析三重威脅



# 注意：只可解釋一階事件序列

## 正確的結論

- 事件「三重威脅」到事件「投球」
- 事件「投球」到事件「運球」
- 事件「運球」到事件「傳球」

以上序列出現顯著轉移

## 錯誤的結論

- 事件「三重威脅」到事件「投球」、再到事件「投球」、再到事件「運球」、最後是事件「傳球」

(未檢定序列  
是否出現顯著轉移)

Part 5.

# 實作:序列分析



# 序列分析計算器

### Input

文字輸入

選擇檔案輸入

Google試算表發佈連結

請上傳CSV檔案：(範例檔案1下載, 範例檔案2下載)

選擇檔案

未選擇任何檔案

### 動態分類資料下載

滯後序列長度: 3

動態分類器檔案

貝氏網路配置檔(平面)

貝氏網路配置檔(多層)

循序模式探勘: 

R腳本檔案: arulesSequences.r

Lag 1 (t)

A

C

D

Lag 0 (g)

A

B

C

D

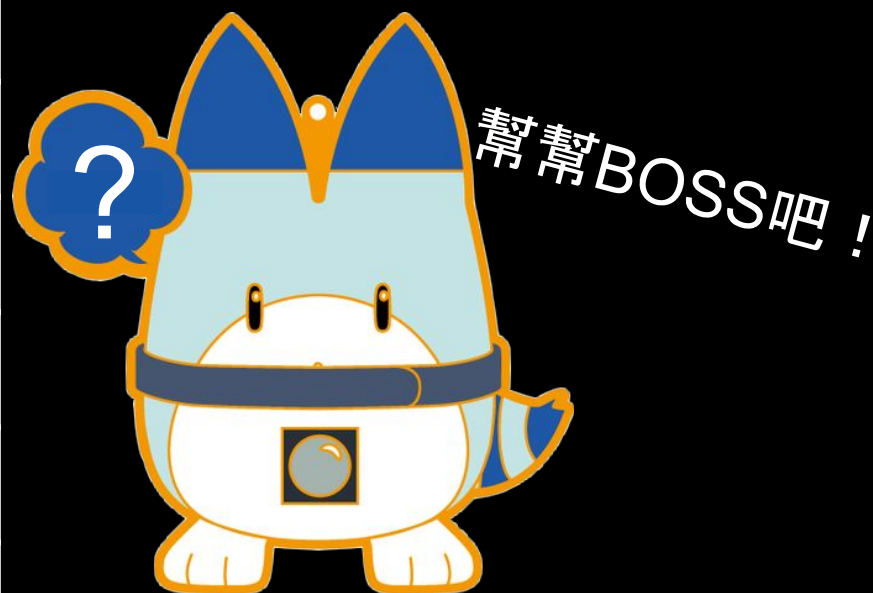


## 序列分析 計算器

研究對象 編號	序列 編號	事件
1	1	A 熱帶莽原區
1	2	A 熱帶莽原區
1	3	B 叢林區
1	4	C 沙漠區
1	5	A 熱帶莽原區
1	6	C 沙漠區
1	7	A 熱帶莽原區
1	8	B 叢林區
1	9	B 叢林區
1	10	C 沙漠區

# 遊客所在區域 事件序列資料 表

案例數量:10筆



# 實作學習單



實作：  
序列分析  
.docx

# W13-a. 序列分析



1. 下載CSV檔案

---



2. 上傳CSV檔案到  
「序列分析計算器」

3. 解讀報表

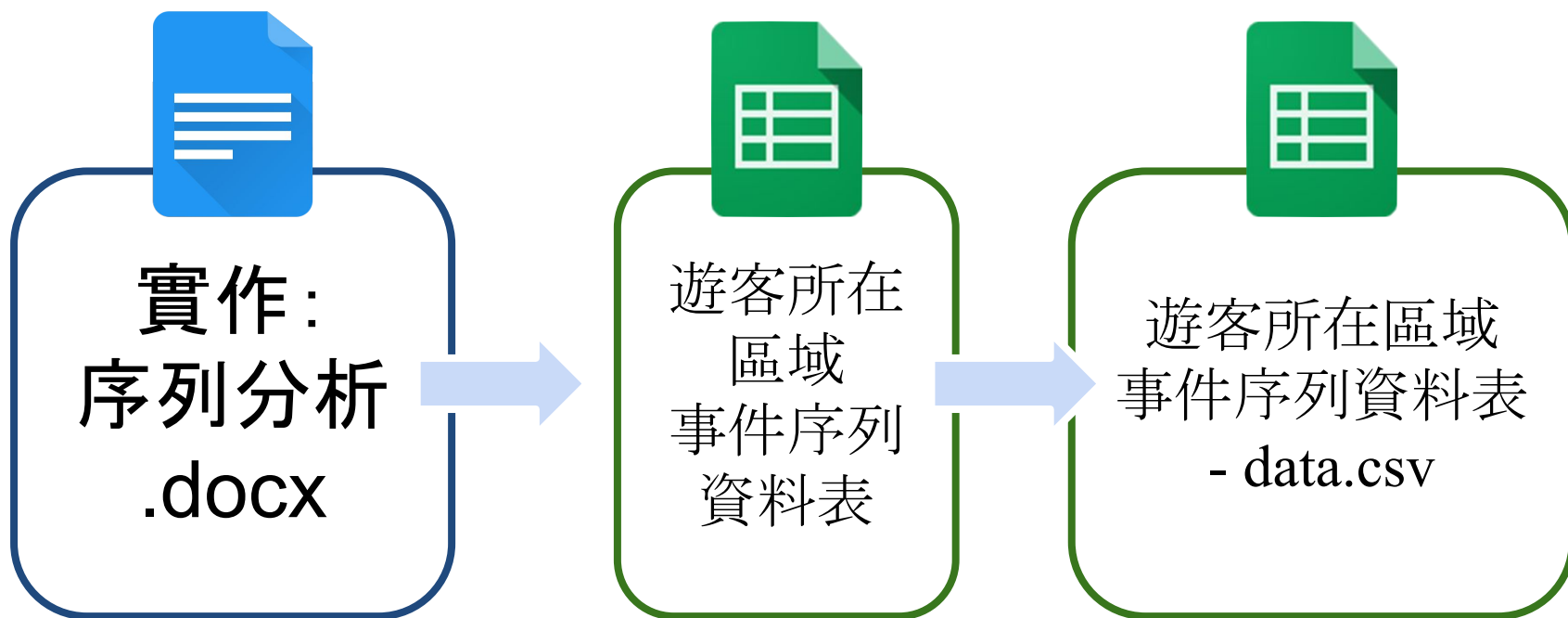
---



4. 撰寫結論



# 1. 下載CSV檔案





## 2. 上傳CSV檔案



遊客所在區域  
事件序列資料表  
- data.csv



**Input**

文字框輸入    選擇檔案輸入    Google試算表發佈連結

請上傳CSV檔案：(範例檔案1下載, 範例檔案2下載)

選擇檔案    未選擇任何檔案

**動態分類資料**

滯後序列長度: 3

動態分類器檔案

具氏網路配置檔(多層)





# 序列分析計算器功能介紹

- Input: 輸入事件序列資料表
- 分類與循序樣式探勘: 序列分析不需要使用此功能
- 事件轉移表編輯器: 可自訂事件轉移表的資料
- Setting: 設定細節
- Result: 顯示報表



### 3. 解讀報表

1. 觀察樣本統計摘要表
2. 事件統計表
3. 事件轉移表
4. 序列分析結果
5. 事件轉移圖

## Result

觀察樣本統計摘要表

研究對象數量	事件類別數量	事件總數	每位研究對象平均事件數量
1	3	10	10.000

事件統計表

事件編碼	出現頻率	出現百分比
A 熱帶莽原區	4	40.0%
B 叢林區	3	30.0%
C 沙漠區	3	30.0%
事件總數	10	100.0%

事件轉移表

			Lag 1 (t)			Lag 0 (g)
			A 熱帶莽原區	B 叢林區	C 沙漠區	總數
Lag 0 (g)	A 熱帶莽原區	出現頻率 f(g,t)	1	2	1	4
		調整後殘差 <sup>a</sup>	-0.474	0.949	-0.474	
	B 叢林區	出現頻率 f(g,t)	0	1	2	3
		調整後殘差 <sup>a</sup>	-1.500	0.000	1.500	
	C 沙漠區	出現頻率 f(g,t)	2	0	0	2
		調整後殘差 <sup>a</sup>	2.268	-1.134	-1.134	
Lag 1 (t)		出現頻率 f(t)	3	3	3	9
總數		出現機率 p(t)	0.333	0.333	0.333	1.000

a. 調整後殘差是用Allison & Liker (1982)的z分數計算公式，超過1.96即達0.05顯著水準，表示此序列次數顯著較多。



序列分析結果：

- 事件「C 沙漠區」到事件「A 熱帶莽原區」，調整後殘差為2.268。

以上序列出現顯著轉移。

事件轉移圖：(可拖曳事件調整位置)





# 觀察樣本統計摘要表 & 事件統計表

## Result

觀察樣本統計摘要表

研究對象數量	事件類別數量	事件總數	每位研究對象平均事件數量
1	3	10	10.000

事件統計表

事件編碼	出現頻率	出現百分比
A 熱帶莽原區	4	40.0%
B 叢林區	3	30.0%
C 沙漠區	3	30.0%
事件總數	10	100.0%



# 序列分析結果與事件轉移圖

回家再自己按按看



序列分析結果：

- 事件「C 沙漠區」到事件「A 熱帶莽原區」  
調整後殘差為2.268。

以上序列出現顯著轉移。





## 4. 撰寫結論：結論寫作框架

研究目的

XXXX(事件編碼)XXXXXX

樣本敘述統計量

序列分析結果

有顯著轉移序列

無顯著轉移序列

事件轉移圖



## 4-1. 研究目的 (框架)

### 研究目的

- 本研究使用序列分析來檢定(研究對象)的行為序列資料中(寫出每一個事件「編碼」)之間是否有顯著轉移。

事件編碼	出現頻率	出現百分比
A 熱帶莽原區	4	40.0%
B 叢林區	3	30.0%
C 沙漠區	3	30.0%
事件總數	10	100.0%



## 4-1. 研究目的 (填空)

### 研究目的

- 本研究使用序列分析來檢定遊客的行為序列資料中事件「A 熱帶莽原區」、事件「B 叢林區」與事件「C 沙漠區」之間是否有顯著轉移。

事件編碼	出現頻率	出現百分比
A 熱帶莽原區	4	40.0%
B 叢林區	3	30.0%
C 沙漠區	3	30.0%
事件總數	10	100.0%





## 4-2. 樣本敘述統計量 (框架)

### 樣本敘述統計量

- 研究對象共(研究對象數量)位，事件總數為(事件總數)次。
- (每個事件編碼都寫一次)
  - 事件「(事件編碼)」出現次數為(出現頻率)次，佔(出現百分比)。
- 雙事件轉移序列總數為(序列總數)次。

觀察樣本統計摘要表

研究對象數量	事件類別數量	事件總數	每位研
1	3	10	

事件統計表

事件編碼	出現頻率	出現百分比
A 熱帶莽原區	4	40.0%
B 叢林區	3	30.0%
C 沙漠區	3	30.0%
事件總數	10	100.0%

調整後殘差 <sup>a</sup>	-0.474	0.949	-0.474	
出現		1	2	3
調整		0.000	1.500	
出現頻率 f(g,t)	2	0	0	2
調整後殘差 <sup>a</sup>	2.268	-1.134	-1.134	
出現頻率 f(t)	3	3	3	9
出現機率 p(t)	0.333	0.333	0.333	1.000

事件轉移表



## 4-2. 樣本敘述統計量 (填空)

### 樣本敘述統計量

- 研究對象共1位，事件總數為10次。
- 事件「A 熱帶莽原區」出現次數為4次，佔40%；事件「B 叢林區」出現次數為3次，佔30%；事件「C 沙漠區」出現次數為3次，佔30%。
- 雙事件轉移序列總數為9次。

觀察樣本統計摘要表

研究對象數量	事件類別數量	事件總數	每位研
1	3	10	

事件統計表

事件編碼	出現頻率	出現百分比
A 熱帶莽原區	4	40.0%
B 叢林區	3	30.0%
C 沙漠區	3	30.0%
事件總數	10	100.0%

調整後殘差 <sup>a</sup>	-0.474	0.949	-0.474	
出現		1	2	3
調整		0.000	1.500	
出現頻率 f(g,t)	2	0	0	2
調整後殘差 <sup>a</sup>	2.268	-1.134	-1.134	
出現頻率 f(t)	3	3	3	9
出現機率 p(t)	0.333	0.333	0.333	1.000



## 4-3. 序列分析結果 (冗長)

### 序列分析結果

#### 有顯著轉移序列

(本實作的情況)

- 序列分析結果顯示，
  - 事件「C 沙漠區」到事件「A 熱帶莽原區」，調整後殘差為2.268。
- 以上序列出現顯著轉移。

有點冗長

#### 無顯著轉移序列

- 序列分析結果顯示，沒有序列達到顯著轉移。



## 4-3. 序列分析結果 (精簡摘要)

### 序列分析結果

#### 有顯著轉移序列

(本實作的情況)

- 序列分析結果顯示，事件「C 沙漠區」到事件「A 熱帶莽原區」調整後殘差為2.268，該序列呈現顯著轉移。

#### 無顯著轉移序列

- 序列分析結果顯示，沒有序列達到顯著轉移。



## 4-4. 事件轉移圖

(如果有顯著轉移序列才寫)

### 事件轉移圖

- 下圖為顯著轉移序列的事件轉移圖。



# W13-a. 序列分析

實作囉！

事件轉移表						
			Lag 1 (t)			Lag 0 (g)
			A 熱帶莽原區	B 叢林區	C 沙漠區	總數
Lag 0 (g)	A 熱帶莽原區	出現頻率 f(g,t)	1	2	1	4
		出現機率 p(g,t)	0.111	0.222	0.111	0.444
		期望個數 exp(g,t)	1.333	1.333	1.333	
		殘差	-0.333	0.667	-0.333	
		調整後殘差 <sup>a</sup>	-0.474	0.949	-0.474	
		相關係數 <sup>b</sup>	-0.333	0.600	-0.333	
	B 叢林區	出現頻率 f(g,t)	0	1	2	3
		出現機率 p(g,t)	0.000	0.111	0.222	0.333
		期望個數 exp(g,t)	1.000	1.000	1.000	
		殘差	-1.000	0.000	1.000	
		調整後殘差 <sup>a</sup>	-1.500	0.000	1.500	
		相關係數 <sup>b</sup>	-1.000	0.000	0.818	
	C 沙漠區	出現頻率 f(g,t)	2	0	0	2
		出現機率 p(g,t)	0.222	0.000	0.000	0.222
		期望個數 exp(g,t)	0.667	0.667	0.667	
		殘差	1.333	-0.667	-0.667	
		調整後殘差 <sup>a</sup>	2.268	-1.134	-1.134	
		相關係數 <sup>b</sup>	1.000	-1.000	-1.000	
Lag 1 (t)		出現頻率 f(t)	3	3	3	9
總數		出現機率 p(t)	0.333	0.333	0.333	1.000

Part 6.

# 深入探討序列分析





		Lag 1 (t)			Lag 0 總數
		A	B	C	
Lag 0 (g)	A	1	2	1	4
	B	0	1	2	3
	C	2	0	0	2
Lag 1 總數		3	3	3	9

都是2次  
為何只有 [C⇒A] 達到顯著？

# [C $\Rightarrow$ A] 達到顯著的原因

C $\Rightarrow$ A		Lag 1(t)			L0 總數
		A	B	C	
L0 (g)	A	1	2	1	4
	B	0	1	2	3
	C	2	0	0	2
L1 總數		3	3	3	9

C在Lag0次數  $f(g) = 2$

A在Lag1次數  $f(t) = 3$

稀有


A $\Rightarrow$ B		Lag 1(t)			L0 總數
		A	B	C	
L0 (g)	A	1	2	1	4
	B	0	1	2	3
	C	2	0	0	2
L1 總數		3	3	3	9

A在Lag0次數  $f(g) = 4$

B在Lag1次數  $f(t) = 3$

常見

# 序列出現顯著轉移的條件



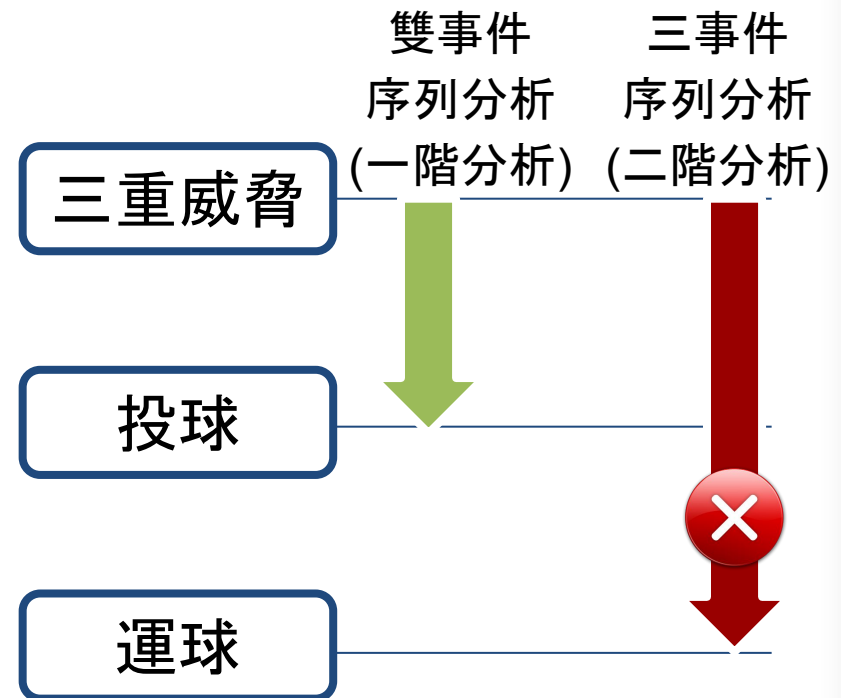
		Lag 1(t)			L0 總數
		A	B	C	
L0 (g)	A	1	2	1	4
	B	0	1	2	3
	C	2	0	0	2
L1		3	3	3	9

Diagram illustrating a sequence transition from C to A. The table shows the frequency of transitions between states A, B, and C at different lags. Red callouts highlight specific values: '1' for the L1 total and '2' for the transition counts from A to B and from B to C.

- 序列很常出現：序列C→A出現頻率很高
- 各別很少出現：C之後很少接其他事件，A之前很少來自其他事件

# 序列分析使用限制

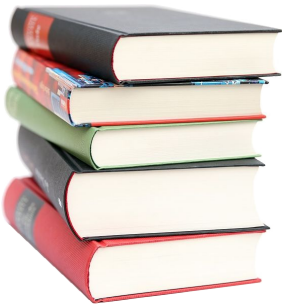
- 不容易達到顯著的情況：
  - 事件類別太多：  
例如A, B, C, D, E, .....
  - 事件總數太少
- 僅分析事件序列的轉移次數
- 限制：
  - 不考慮時間長度單位
  - 只能分析前一步與下一步 (一階分析)



# 事件太多的處理方法

## 合併事件編碼

### 圖書館藏



統計學的世界

量化研究法

資料探勘

壞壞總裁愛...

總裁一家逼...

報復總裁大...



### 事件編碼

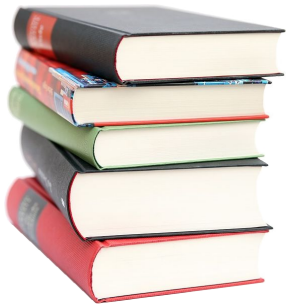
A 專業書籍

B 娛樂書籍

# 事件太多的處理方法

## ~~刪除~~非研究重點的事件編碼

### 圖書館藏



A 統計學的世界

B 量化研究法

~~C 資料探勘~~

D 壞壞總裁愛...

E 總裁一家逼...

~~F 報復總裁大...~~



### 行為序列資料表

研究對象編號	序列編號	事件
1	1	D
1	2	E
2	1	A;B
2	1	E
2	2	D
3	3	A;B; <del>E</del>

# 除了序列分析之外 ...

## W14 資料的分類與預測

### 動態貝氏網路

- 用前幾步動作預測下一步的機率
- 能夠建立預測下一步行為的規則機率模型

## W17 行為序列模式探勘

### 循序樣式探勘

- 找出多人共同的長序列行為模式
- 能考慮整體時間長度，不僅分析前後雙事件序列



Part 7.

課堂練習：

**幼兒平行遊戲行為研究**

# 幼兒平行遊戲行為研究

- Bakeman & Brownlee (1980) 對幼兒進行「平行遊戲」的記錄。
- 每15秒記錄一次嬰兒的主要狀態：
- 事件編碼
  - U 空閒：沒有玩玩具
  - S 一人：一個人玩玩具
  - T 兩人：兩個人一起玩玩具
  - P 各自：兩個人拿同樣的玩具，各玩各的
  - G 團體：三人以上一起玩玩具

幼兒平行遊戲事件序列資料表			
檔案 編輯 檢視 插入 格式 資料 工具			
NT\$ % .0 .00 123			
$f_x$	A	B	C
1	研究對象編號	序列編號	事件
2	1	1	U 空間
3	1	2	S 一人
4	1	3	P 各自
5	1	4	T 兩人
6	1	5	P 各自
7	1	6	T 兩人
8	1	7	P 各自
9	1	8	G 團體
10	1	9	T 兩人
11	1	10	P 各自
12	1	11	T 兩人
13	1	12	G 團體
14	1	13	T 兩人
15	1	14	P 各自
16	1	15	G 團體
17	1	16	T 兩人
18	1	17	G 團體

# 幼兒平行遊戲事件序列資料表

案例數量:128筆

※ 為教學需求, 本資料為虛構



# W13-b. 幼兒平行遊戲 行為研究



1. 下載CSV檔案

---



2. 上傳CSV檔案到  
「序列分析計算器」

3. 解讀報表

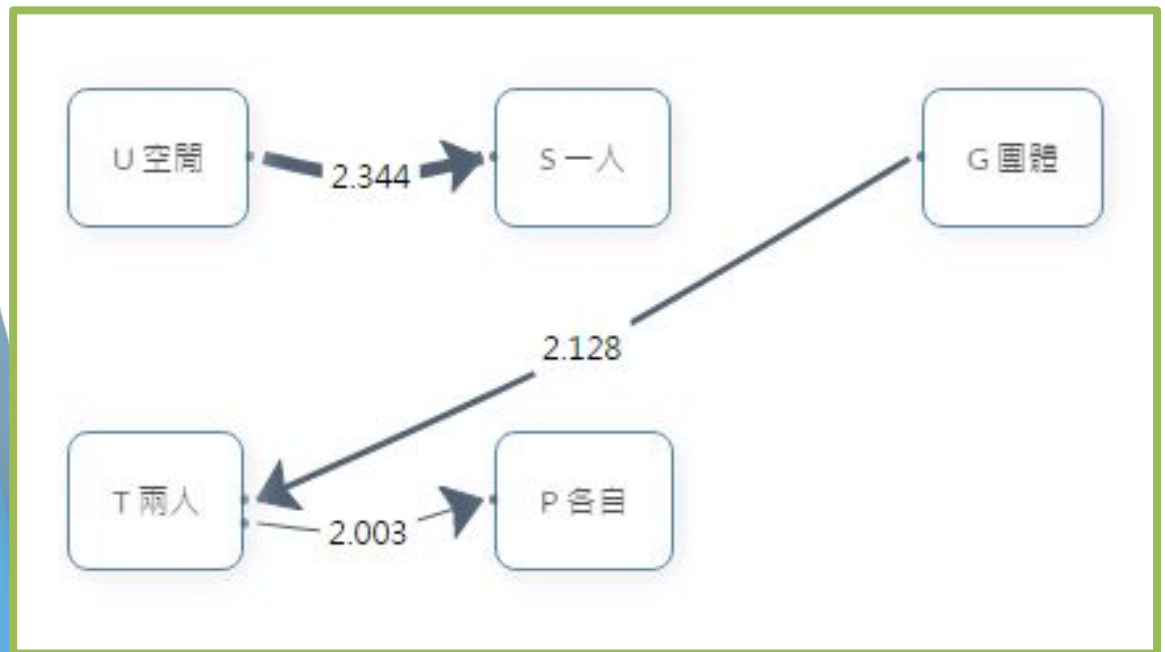
---



4. 撰寫結論

# W13-b. 幼兒平行遊戲 行為研究

練習啦！



Part 8.

# 序列分析的應用

# 序列分析的應用

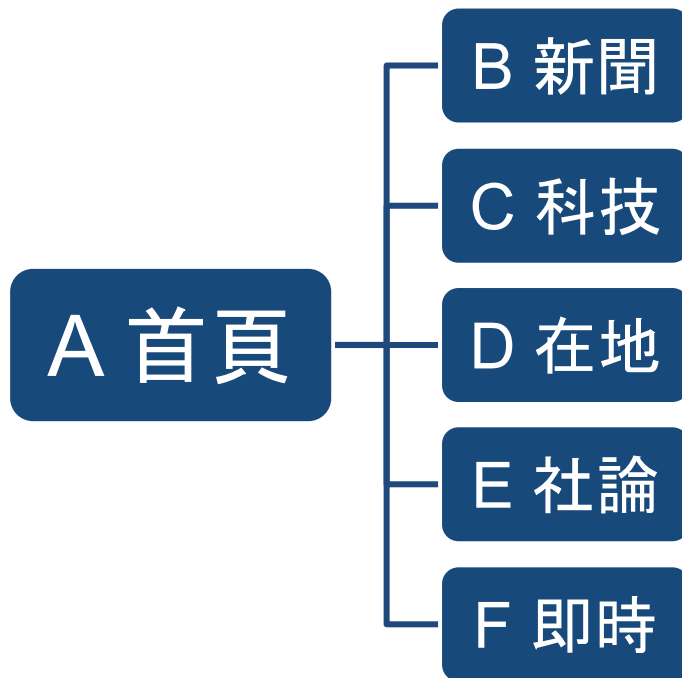
- 合併多研究對象
- 同時間多事件
- 僅分析有改變序列
- 不同研究對象之間的比較



# 增加事件總數的處理方法

## 合併多研究對象

MSNBC新聞網



行為序列資料表

研究對象編號	序列編號	事件
1	1	A
1	2	A
2	1	B
3	1	C
3	2	B
3	3	B

# 同一個研究對象的序列 可以計算



研究對象編號	序列編號	事件
1	1	A
1	2	A
2	1	B
3	1	C
3	2	B
3	3	B

		Lag 1 (t)			Lag 0 總數
		A	B	C	
Lag 0 (g)	A	1	0	0	1
	B	0	0	0	0
	C	0	0	0	0
Lag 1 總數		1	0	0	1

# 不同研究對象的序列 不可以計算



研究對象編號	序列編號	事件
1	1	A
1	2	A
2	1	B
3	1	C
3	2	B
3	3	B

		Lag 1 (t)			Lag 0 總數
		A	B	C	
Lag 0 (g)	A	1	0	0	1
	B	0	0	0	0
	C	0	0	0	0
Lag 1 總數		1	0	0	1

# 同時間多事件的序列分析

## 圖書館藏



A 統計學的世界

B 量化研究法



C 壞壞總裁愛...

## 行為序列資料表

研究對象編號	序列編號	事件
1	1	A;C
1	2	C
2	1	A;B
2	1	C
2	2	B
3	3	A;B;C

# 計算每次各種序列的組合

$A \Rightarrow C$

研究對象編號	序列編號	事件
1	1	A;C
1	2	C
2	1	A;B
2	1	C
2	2	B
3	3	A;B;C

		Lag 1 (t)			Lag 0 總數
		A	B	C	
Lag 0 (g)	A	0	0	1	1
	B	0	0	0	0
	C	0	0	0	0
Lag 1 總數		0	0	1	1

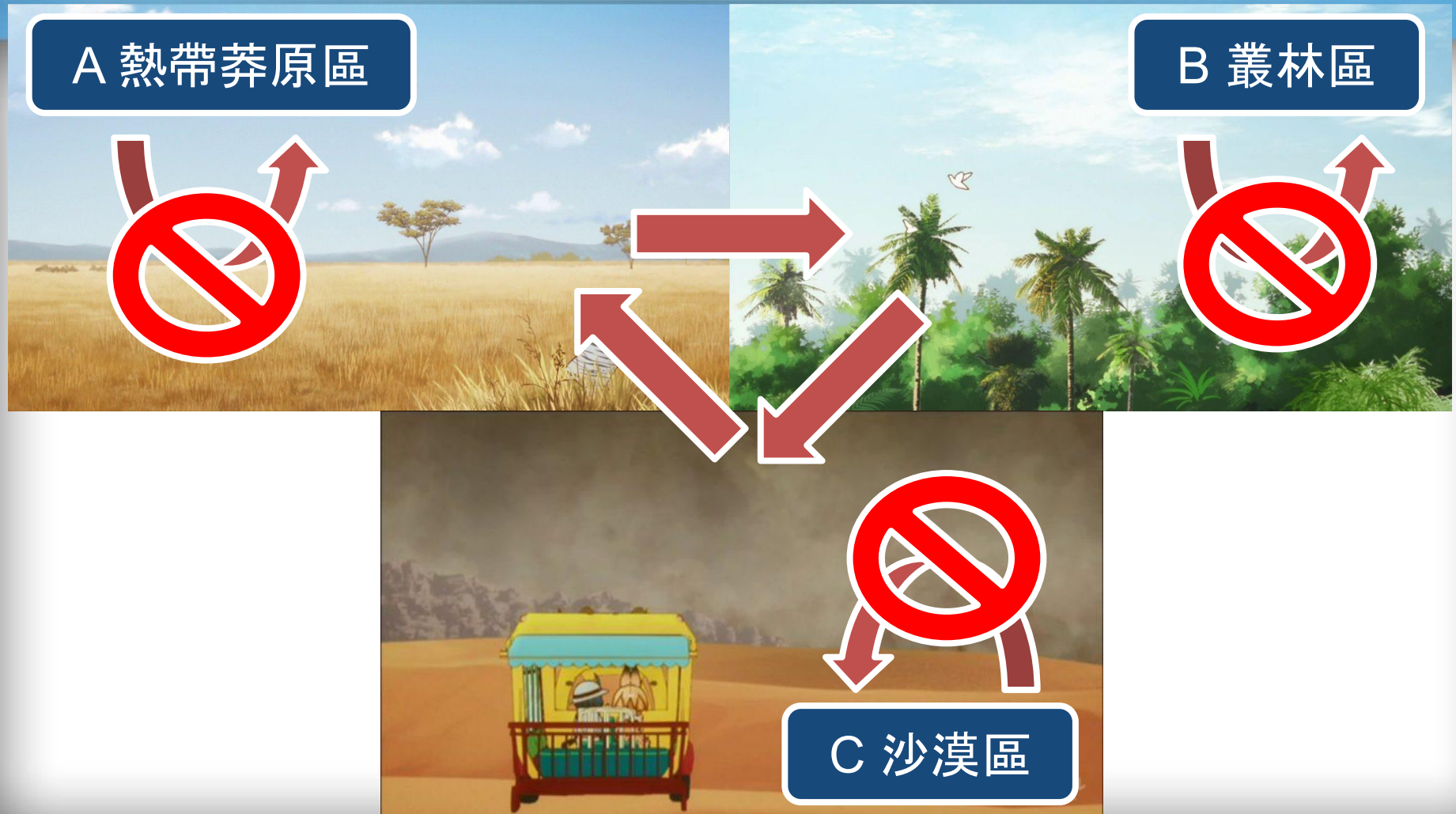
# 計算每次各種序列的組合

$C \Rightarrow C$

研究對象編號	序列編號	事件
1	1	A;C
1	2	C
2	1	A;B
2	1	C
2	2	B
3	3	A;B;C

		Lag 1 (t)			Lag 0 總數
		A	B	C	
Lag 0 (g)	A	0	0	1	1
	B	0	0	0	0
	C	0	0	1	1
Lag 1 總數		0	0	2	1

# 僅分析有改變序列





打勾啟用  
此分析

☒ 僅分析有改變的序列

☒ 顯示詳細資訊。

☐ 顯示完整的事件轉移圖。



☐ 使用無樣式表格(容易複製到其他文件)

## Setting

## Result

事件轉移表

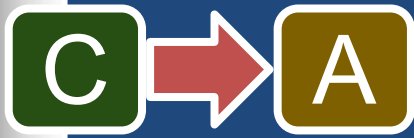
	Lag 1 (t)			Lag 0 (g)
	A	B	C	總數

研究對象編號	序列編號	事件
1	1	A
1	2	A 
1	3	B
1	4	C
1	5	A
1	6	C
1	7	A
1	8	B
1	9	B 
1	10	C

把前後重複的序列  
進行合併吧！

		Lag 1 (t)			Lag 0 總數
		A	B	C	
Lag 0 (g)	A	0	2	1	3
	B	0	0	2	2
	C	2	0	0	2
Lag 1 總數		2	2	3	7

# 僅分析有改變序列的調整後殘差計算 $p(t)$



		Lag 1 (t)			Lag 0 總數
		A	B	C	
Lag 0 (g)	A	0	2	1	7
	B	2	2	2	
	C	0	2	2	
Lag 1 總數		2	2	3	

$$p(t) = \frac{f(t)}{N_s - f(g)}$$

$$= 2/(7-2)$$

$$= 0.4$$

$N_s$

# 僅分析有改變序列的調整後殘差表

			Lag 1 (t)			Lag 0 總數
			A	B	C	
Lag 0 (g)	A	頻率	0	2	1	3
		調整後殘差	-2.29	0.76	-2.21	
	B	頻率	0	0	2	2
		調整後殘差	-1.37	-1.37	1.37	
	C	頻率	2	0	0	2
		調整後殘差	2.049	-1.37	-2.05	
	Lag 1 總數		2	2	3	7

用計算機  
比較差異  
看看吧

# 不同研究對象之間的比較

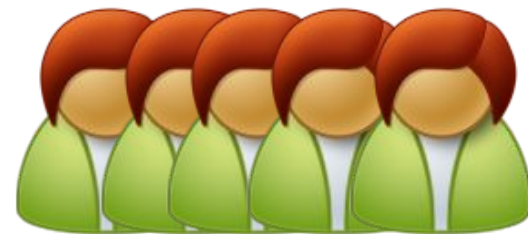
## 調整後殘差

### 實驗組的事件轉移表



(調整後 殘差)		Lag 1 (t)		
		A	B	C
Lag 0 (g)	A	0.57	1.51	-1.95
	B	-1.56	-1.66	2.78
	C	1.12	0.42	-1.20

### 控制組的事件轉移表



(調整後 殘差)		Lag 1 (t)		
		A	B	C
Lag 0 (g)	A	-2.00	0.31	1.46
	B	-0.78	-0.32	0.99
	C	2.68	0	-2.35

# 使用相關係數來比較



實驗組的  
事件轉移表

			Lag 1 (t)		
			A	B	C
Lag 0 (g)	A	調整後殘差	0.57	1.51	-1.95
		相關係數	0.36	0.69	-1.00
	B	調整後殘差	-1.56	-1.66	2.78
		相關係數	-1.00	-0.62	0.87
	C	調整後殘差	1.12	0.42	-1.20
		相關係數	0.60	0.18	-0.51

使用  
相關係數  
來比較！

# 相關係數 Yule'Q值

很像  
相關係數r

完全負相關  $-1.00 \leq$

相關係數

$\leq 1.00$  完全正相關

相關係數範圍(絕對值)	關聯程度	相關係數 > 0.7 容易顯著
1.00	完全(正/負)相關	
0.70 - 0.99	高度(正/負)相關	相關係數 > 0.7 容易顯著
0.40 - .069	中度(正/負)相關	
0.10 - 0.39	低度(正/負)相關	
0.10 以下	無相關	



# 相關係數Yule'Q值

## 1. 縮減列聯表

**B** → **C**

出現頻率		Lag 1 (t)		
		A	B	C
Lag 0 (g)	A	1	4	0
	B	0	3	7
	C	2	5	2

➔

**B** → **C**

出現頻率 二項實驗		Lag 1 (t)	
		是C	不是C
Lag 0 (g)	是B	7	3
	不是B	2	12

# 相關係數Yule'Q值

## 2. 計算相關係數



出現頻率 二項實驗		Lag 1 (t)	
		是C	不是C
Lag 0 (g)	是B	7 (a)	3 (b)
	不是B	2 (c)	12 (d)

$$\text{相關係數Yule's } Q \text{ 值} = \frac{ad - bc}{ad + bc}$$

$$= \frac{7 \times 12 - 3 \times 2}{7 \times 12 + 3 \times 2}$$

$$= 0.867$$

# 相關係數Yule'Q值 計算每一個細格的相關係數

實驗組的 事件轉移表			Lag 1 (t)		
			A	B	C
Lag 0 (g)	A	出現頻率	1	4	0
		相關係數	0.36	0.69	-1.00
	B	出現頻率	0	3	7
		相關係數	-1.00	-0.62	0.87
	C	出現頻率	2	5	2
		相關係數	0.60	0.18	-0.51

# 不同研究對象之間的 相關係數 比較

## 實驗組的事件轉移表

(相關係數)		Lag 1 (t)		
		A	B	C
Lag 0 (g)	A	0.36	0.69	-1.00
	B	-1.00	-0.62	0.87
	C	0.60	0.18	-0.51

## 控制組的事件轉移表

(相關係數)		Lag 1 (t)		
		A	B	C
Lag 0 (g)	A	-1.00	0.14	0.57
	B	-0.43	-0.15	0.42
	C	0.89	0	-0.85

## 組間比較結論

- 以事件「C」到事件「A」的序列來看，
- 控制組的調整後殘差為2.68，達到顯著轉移；實驗組的調整後殘差為1.12，未達顯著轉移。
- 控制組的相關係數為0.89，是高度相關，高於實驗組相關係數的0.6。

*Thank you for  
your attention*

