

SPSS 市場調查與研究

大數據分析

- R/Python/Julia/SQL 程式設計與應用
(R/Python/Julia/SQL Programming and Application)
- 資料視覺化 (Data Visualization)
- 機器學習 (Machine Learning)
- 統計品管 (Statistical Quality Control)
- 最佳化 (Optimization)



李明昌博士

alan9956@gmail.com

<http://rwepa.blogspot.com/>

大綱

Day 1

1. SPSS軟體簡介
2. 敘述統計分析

補充篇：圖表建置器

Day 2

3. 卡方檢定應用
4. 平均數檢定與變異數分析

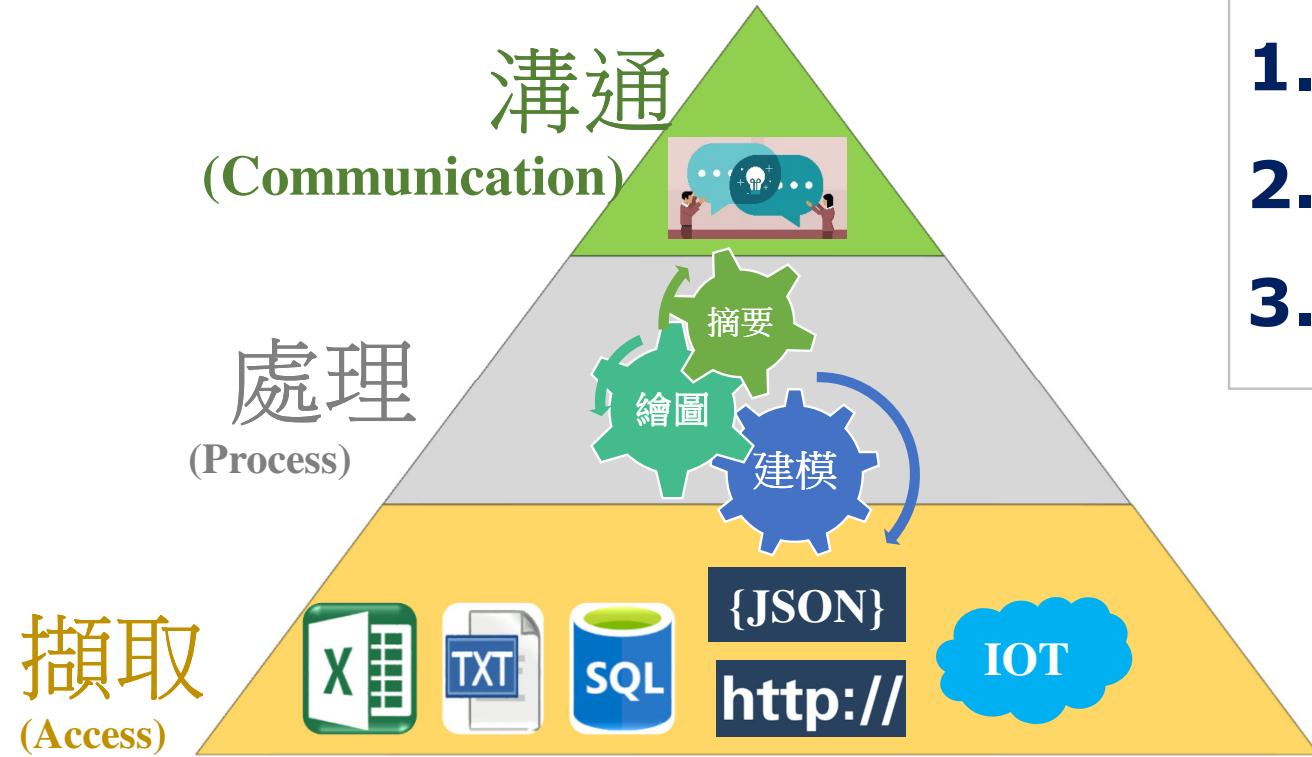
個人簡介 <http://rwepa.blogspot.com/>

- 姓名：李明昌 (ALAN LEE)
- 現職：中華R軟體學會 常務理事
臺灣資料科學與商業應用協會 常務理事
- 學歷：中原大學 工業與系統工程所 博士
- 經歷：
 - 淡江大學 兼任教師
 - 育達科技大學 兼任教師
 - 佛光大學 兼任教師
 - 國立台北商業大學 兼任教師
 - 東吳大學 兼任教師
 - 育達科技大學 資訊管理系(所) 專任助理教授
 - 崇友實業 行銷企劃專員
 - 國航船務代理股份有限公司 海運市場運籌管理員
- 大專院校、資策會、工業技術研究院、國家發展委員會、中央氣象局、公平交易委員會、各縣市政府與日本名古屋產業大學等公營單位演講達300餘場，2900小時以上。
- 連絡資訊：alan9956@gmail.com



- iPAS 巨量資料分析師 證照推廣
- iPAS 營運智慧分析師 證照推廣

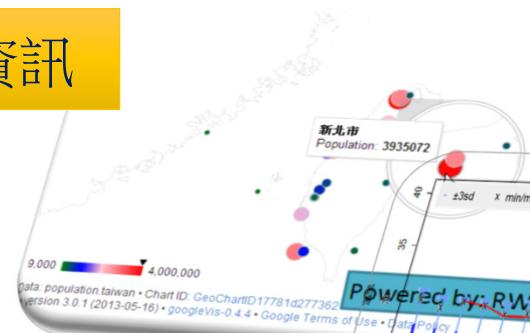
★★★資料分析架構→APC方法



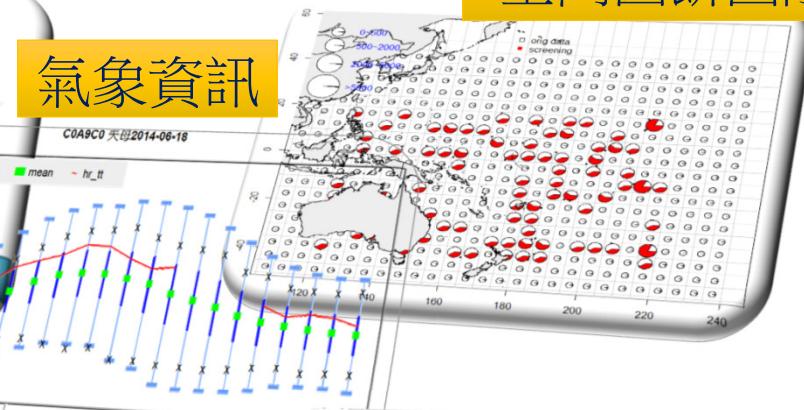
資料分析/視覺化應用

R + shiny → 互動式網頁

地理資訊



氣象資訊

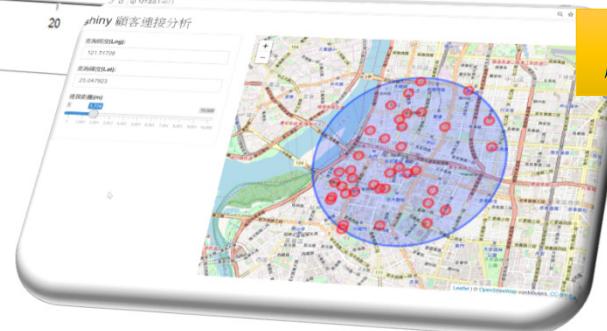


空間圓餅圖離群值分析

保險預測



顧客連結資訊



中央氣象局 1,600萬筆資料

網頁呈現



客製化選單

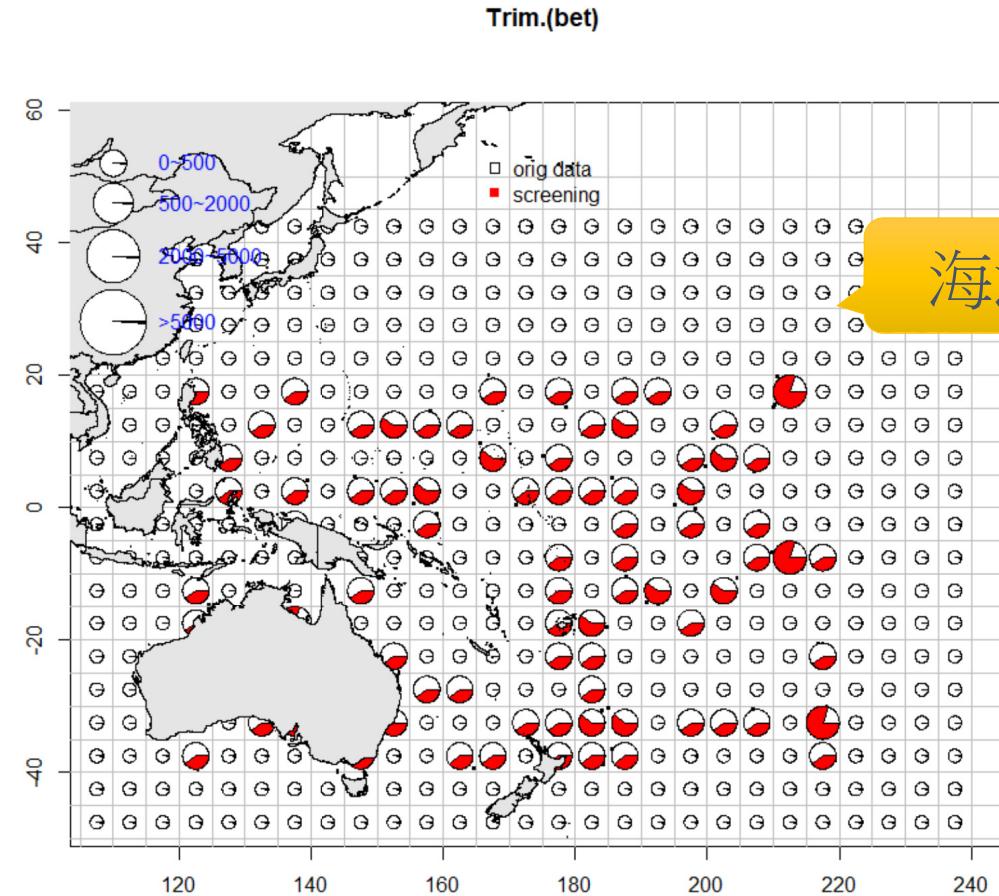
R統計運算

保險預測模型

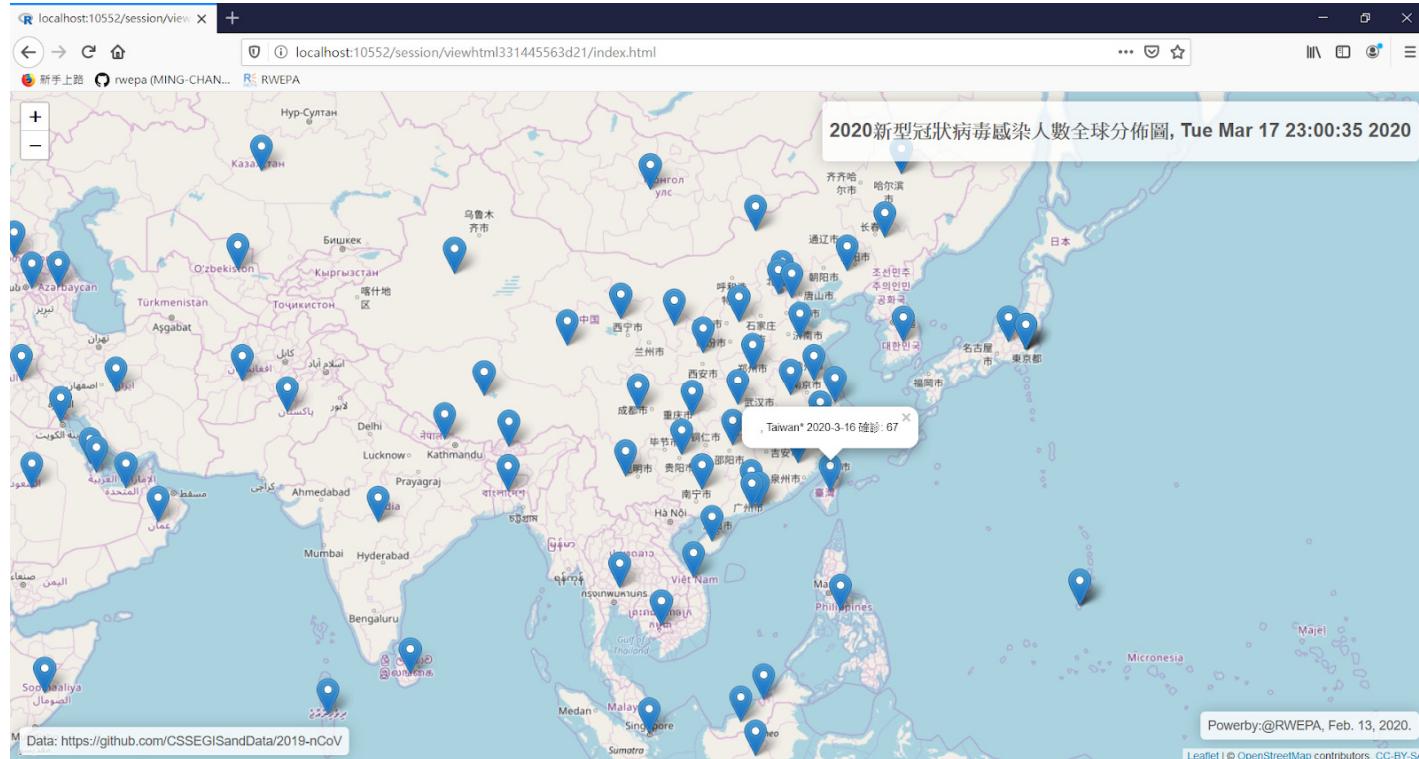
The screenshot shows a web-based application interface for insurance data analysis. At the top, there's a navigation bar with tabs like '檔案上傳', '資料處理', '統計圖表', '模型評估', and '預測模型'. The '預測模型' tab is highlighted with a red box. Below it, there's a section titled '機率模型閾值' with a slider set at 0.1. To the right, a yellow callout bubble says '機率模型閾值調整'. In the center, there's a table titled '檢視結果' (Review Results) with various columns including gender, vehicle type, exposure risk, age, and predicted probability. A yellow callout bubble on the right says '預測結果' (Prediction Results). The table has 12 rows of data. At the bottom, it says 'Showing 1 to 10 of 12 entries' and has navigation buttons for 'Previous', 'Next', and page numbers 1 and 2.

性別	女性	車輛種類	私家車	曝露風險		曝露風險對數		無索償折扣	被保險人年齡	私家車 —車齡 0	私家車 —車齡 1	私家車 —車齡 2	私家車 —車齡 0_1_2 組合	車齡 0_1_2 組合	預測 機率	理賠
				曝露風險	曝露風險對數	私家車 —車齡 0	私家車 —車齡 1									
M	0	A	1	0.9144422	-0.08944106	50	4	1	0	0	1	0	2	0.1069	有	
M	0	A	1	0.8158795	-0.20348856	20	4	0	0	1	1	2	2	0.1441	有	
3	M	0	A	1	0.8377823	-0.17699695	50	3	0	0	1	1	2	2	0.1866	有
4	M	0	A	1	0.4325804	-0.83798702	50	6	0	1	0	1	1	2	0.0944	無
5	M	0	A	1	0.7173169	-0.33223755	50	4	0	0	1	1	2	2	0.1218	有
6	M	0	A	1	0.8377823	-0.17699695	50	4	0	0	1	1	2	2	0.1495	有
7	M	0	A	1	0.8487337	-0.16400975	50	5	0	0	1	1	2	2	0.1422	有
8	F	1	A	1	0.8268309	-0.19015503	10	3	0	0	1	1	2	2	0.1733	有
9	M	0	A	1	0.7145791	-0.33606164	0	5	1	0	0	1	0	2	0.0694	無
10	M	0	A	1	0.3340178	-1.09656101	0	3	0	0	1	1	2	2	0.0783	無

空間圓餅圖離群值分析



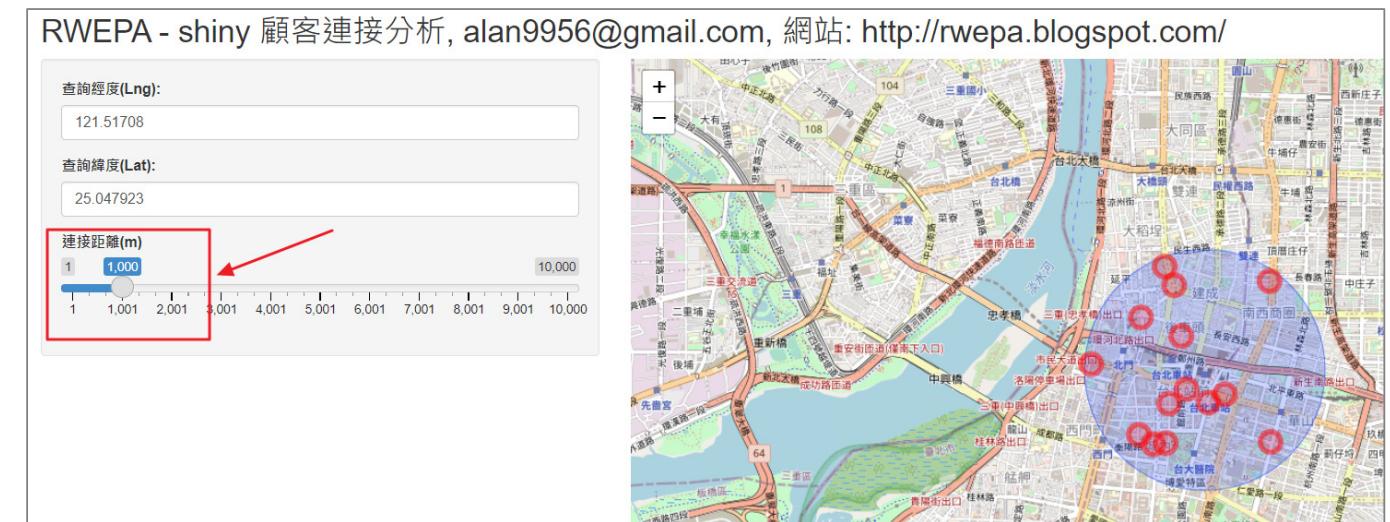
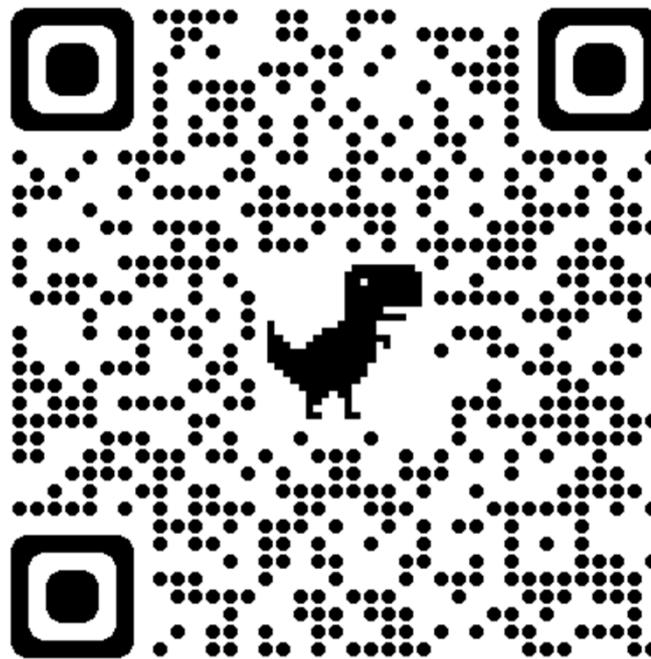
2020新型冠狀病毒視覺化



<http://rwepa.blogspot.com/2020/02/2019nCoV.html>

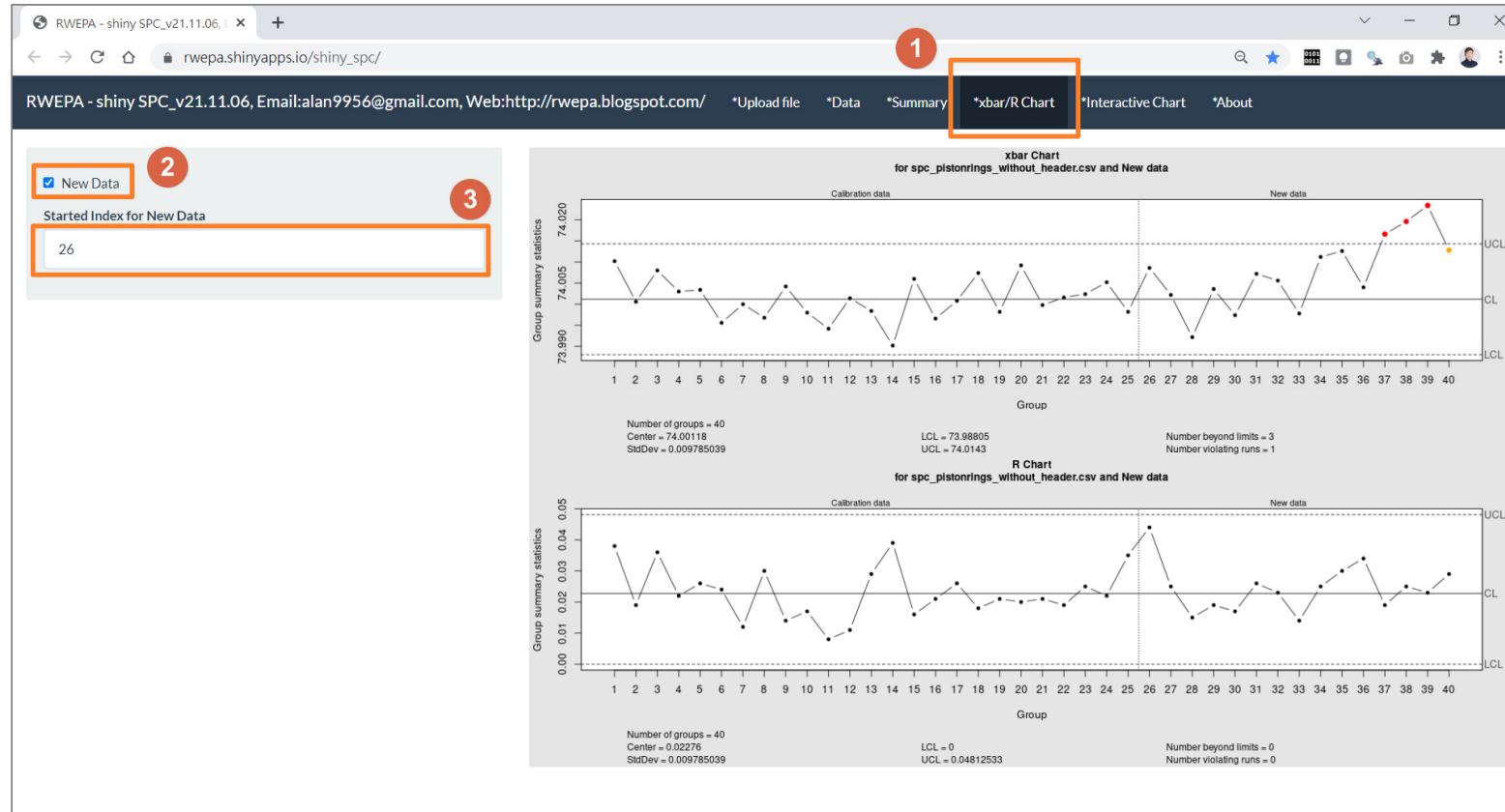
shiny 顧客連接分析

- <https://rwepa.shinyapps.io/shinyCustomerConnect/>



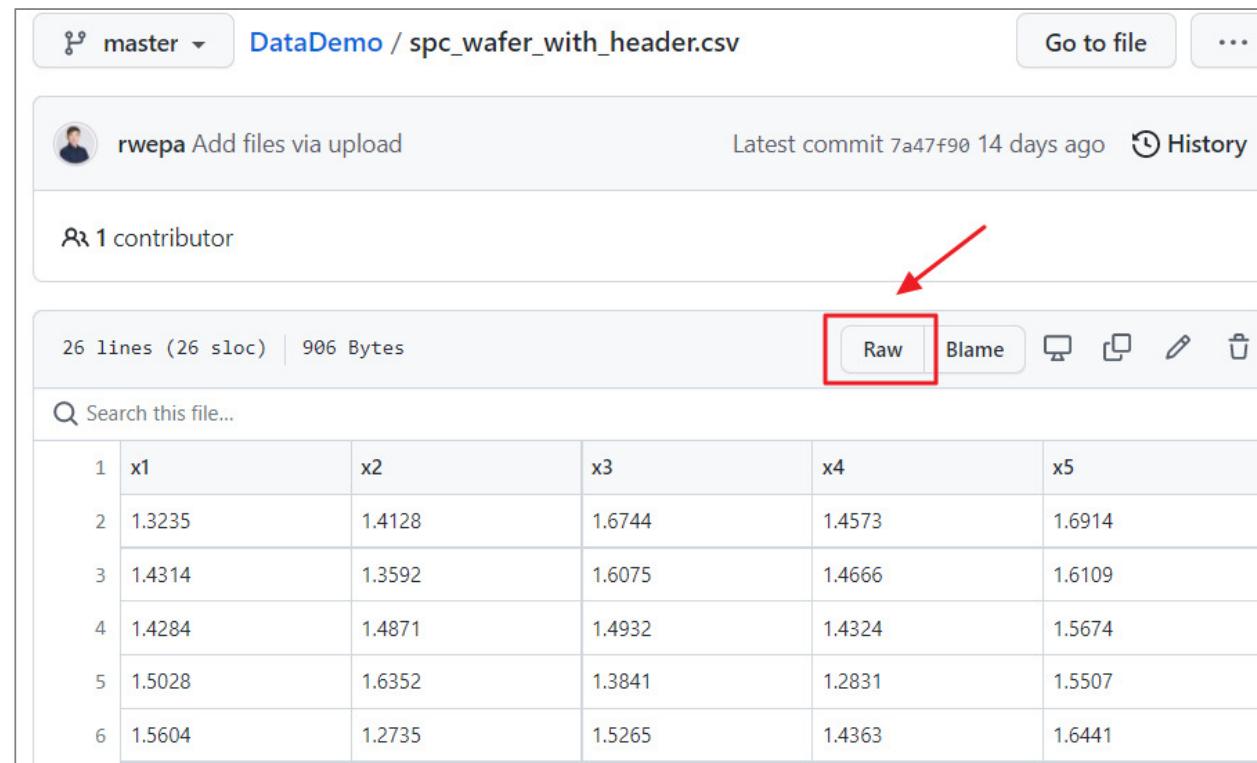
品質管制圖(quality control chart)應用

- <http://rwepa.blogspot.com/2021/10/r-shiny-quality-control-chart.html>



品質管制圖應用 - 資料集

- https://github.com/rwepa/DataDemo/blob/master/spc_wafer_with_header.csv
- https://github.com/rwepa/DataDemo/blob/master/spc_pistonrings_without_header.csv



master DataDemo / spc_wafer_with_header.csv Go to file ...

rwepa Add files via upload Latest commit 7a47f90 14 days ago History

1 contributor

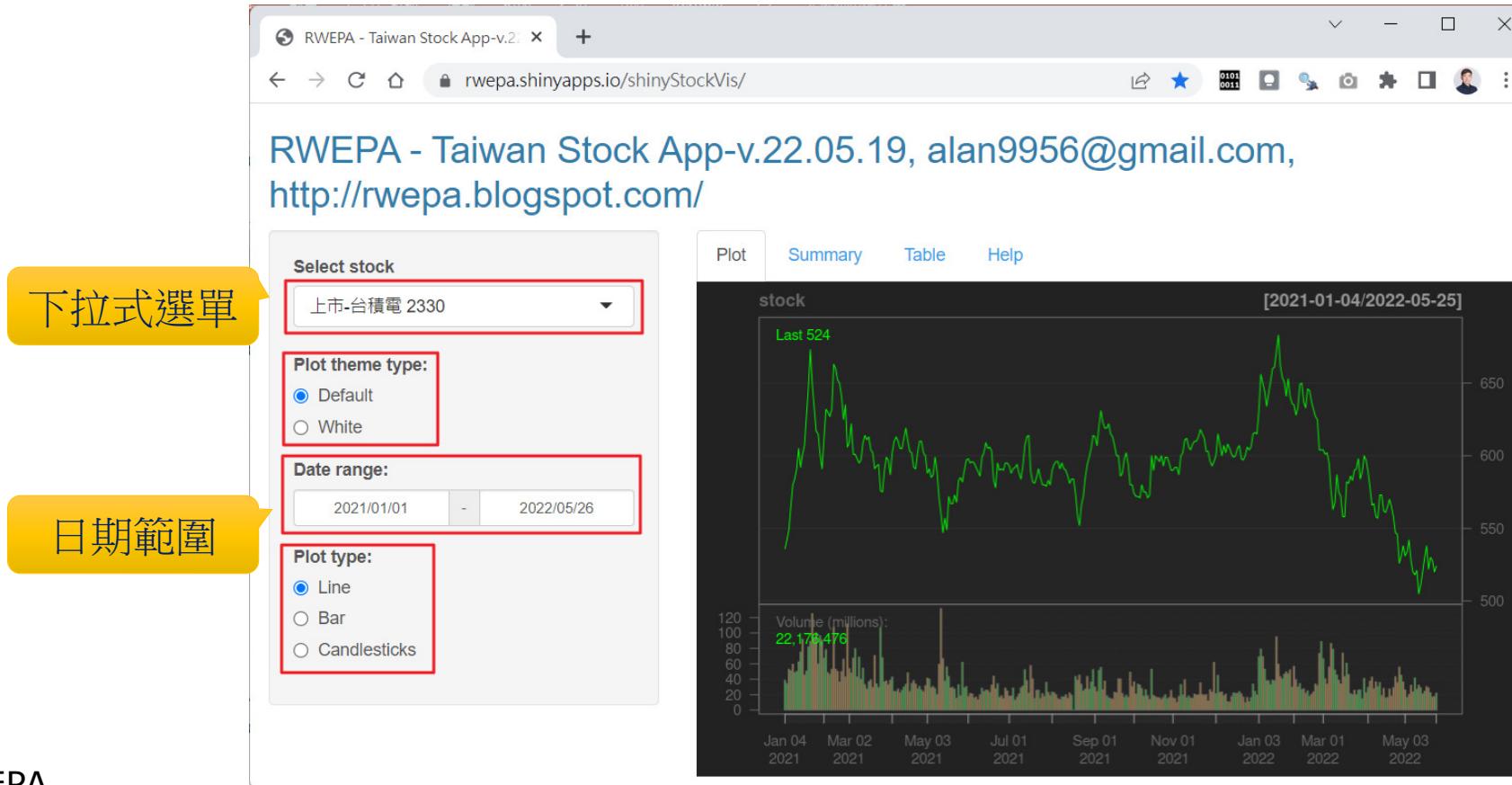
26 lines (26 sloc) | 906 Bytes Raw Blame

Search this file...

	x1	x2	x3	x4	x5
2	1.3235	1.4128	1.6744	1.4573	1.6914
3	1.4314	1.3592	1.6075	1.4666	1.6109
4	1.4284	1.4871	1.4932	1.4324	1.5674
5	1.5028	1.6352	1.3841	1.2831	1.5507
6	1.5604	1.2735	1.5265	1.4363	1.6441

Taiwan Stock App

- <https://rwepa.shinyapps.io/shinyStockVis/>

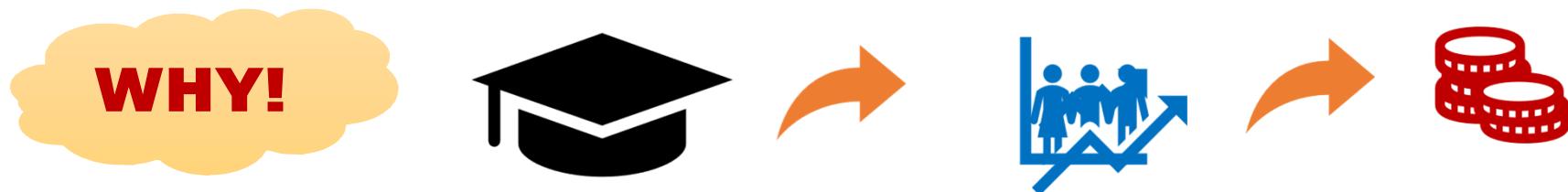


離子資料分析與視覺化應用

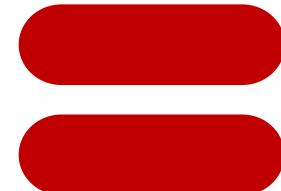
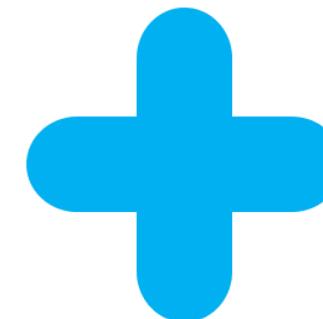


如何學習 SPSS

- 熟悉教材內容
- 將教材內容的資料集改為工作資料集
- 遇到問題時，想辦法尋找答案
- 掌握 APC方法
- 掌握 摘要, 繪圖, 建模
- 參考網路應用文章 (進階) & 學術論文



學習目標



Day 1

1. SPSS軟體簡介

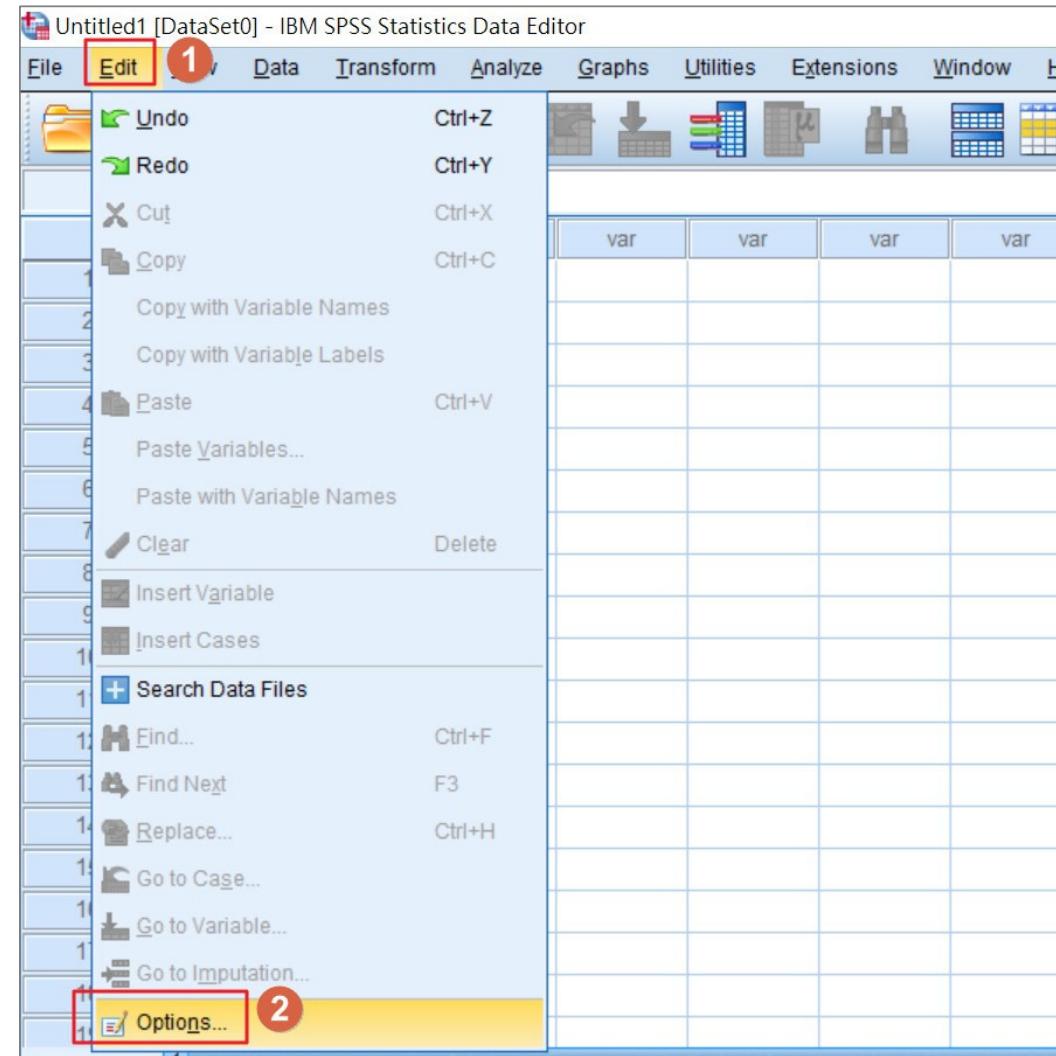
SPSS 簡介

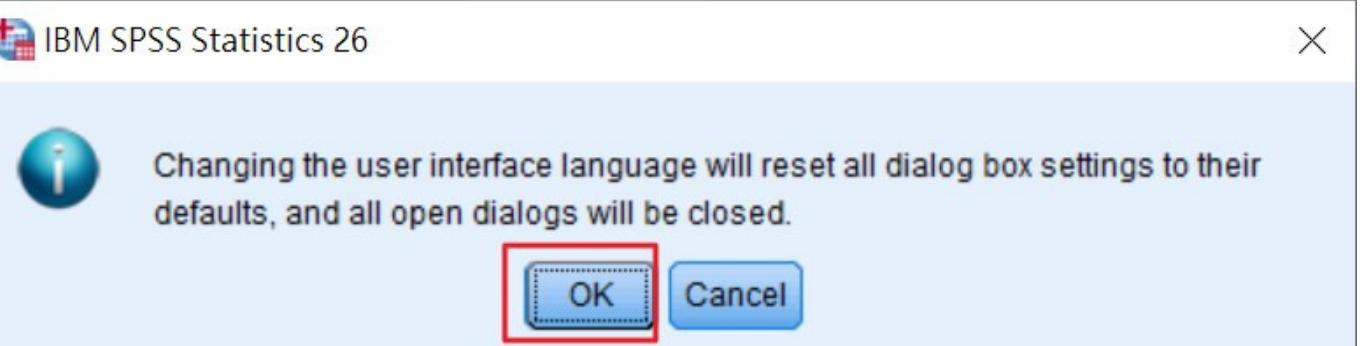
- SPSS是統計產品與服務解決方案（Statistical Product and Service Solutions）的簡稱。
- SPSS為IBM公司的一系列用於統計學分析運算、資料探勘、預測分析和決策支持任務的軟體產品及相關服務的總稱。
- SPSS二大軟體：
 - IBM SPSS Statistics → 課程主軸
 - IBM SPSS Modeler
- 參考: <https://www.ibm.com/spss>
- 參考: 楊世瑩, SPSS 26統計分析嚴選教材(適用SPSS 26~22), 茉峰, 2020.

SPSS 發展歷史

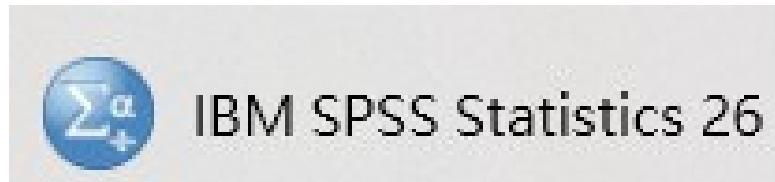
- SPSS原名為社會科學統計套件（Statistical Package for the Social Sciences）。
- 1968年，美國斯坦福大學三位研究生開發出最早的SPSS軟體，當時主要對象為中小型計算機和企業用戶，產品統稱SPSSx版。1975年，芝加哥成立SPSS公司。
- 1984年，SPSS公司首先推出世界上第一個可以在DOS上運行的統計分析軟體的PC版本，即SPSS/PC+版。後來又相繼推出Windows和Mac OS X等操作系統上的版本，並不斷擴展軟體的功能相關服務，形成了目前SPSS的基本面貌，同時該公司也被IBM收購。
- SPSS® 軟體平台提供進階統計分析、龐大的機器學習演算法、文字分析、開放程式碼延伸，並與大數據整合，還能無縫部署到應用程式中。

- 英文版 SPSS 更改為中文介面：Edit \ Options

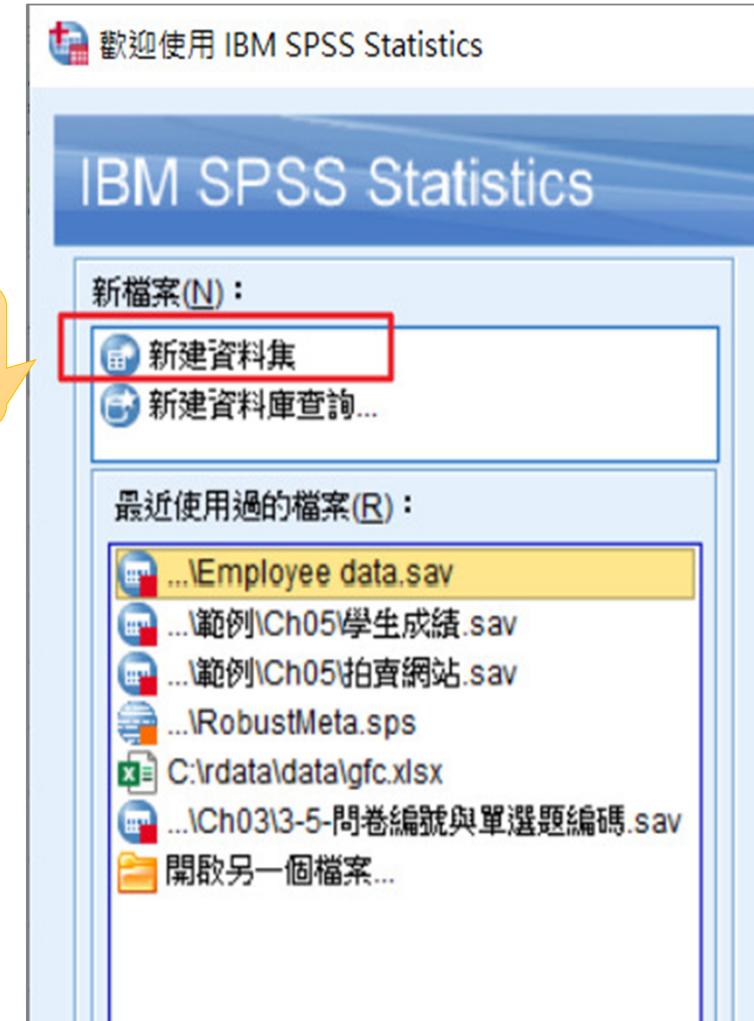




- 執行SPSS，可於Windows的「程式集」，點選「IBM SPSS Statistics 26」方形圖示：



新建資料集



- 選取[新建資料集]，按[確定]鈕後，可開啟『IBM SPSS Statistics 資料編輯器』畫面。此功能與「檔案(F)/新增(E)/資料(D)」相同。



1. 功能表列
2. 工具列
3. 資料區
4. 資料視窗／變數視窗

啤酒廠牌.sav

啤酒廠牌.sav [資料集1] - IBM SPSS Statistics 資料編輯器

檔案(E) 編輯(E) 檢視(V) 資料(D) 轉換(I) 分析(A) 圖形(G) 公用程式(U) 延伸(X) 視窗(W) 說明(H)

1:編號 1 顯示：4 個變數（共有 4 個）

	編號	是否飲用	啤酒廠牌	廠牌分組	變數	變數	變數	變數	
1	1	1	6	12					
2	2	1	1	1					
3	3	1	3	5					
4	4	1	1	1					
5	5	2	0	0					
6	6	1	5	5					
7	7	1	1	1					
8	8	1	1	1					
9	9	1	4	4					
10	10	1	5	5					
11	11	1	4	4					

資料視圖 變數視圖

IBM SPSS Statistics 處理器已備妥 Unicode:ON

啤酒廠牌.spv (SPSS Output Document, SPV)

The screenshot shows the IBM SPSS Statistics Viewer window for the file '啤酒廠牌.spv [文件2]'. The menu bar includes 檔案(E), 編輯(E), 檢視(V), 資料(D), 轉換(I), 插入(I), 格式(O), 分析(A), 圖形(G), 公用程式(U), 延伸(X), 視窗(W), and 說明(H). The toolbar contains various icons for file operations like Open, Save, Print, and Export.

The left pane displays a tree view of the output structure:

- 輸出 (Output)
 - 日誌 (Log)
 - 次數分配表 (Frequency Tables)
 - 標題 (Title)
 - 附註 (Notes)
 - 作用中資料集 (Active Dataset)
 - 統計量 (Statistics)
 - 次數表 (Frequency Tables)
 - 標題 (Title)
 - 是否飲用 (Drink Type)
 - 啤酒廠牌 (Beer Brand)
 - 廠牌分組 (Brand Group)

GET
FILE='C:\spss_datasets\啤酒廠牌.sav'.
DATASET NAME 資料集1 WINDOW=FRONT.
FREQUENCIES VARIABLES=是否飲用 啤酒廠牌 廠牌分組
/STATISTICS=STDEDEV MEAN MEDIAN
/ORDER=ANALYSIS.

→ 次數分配表

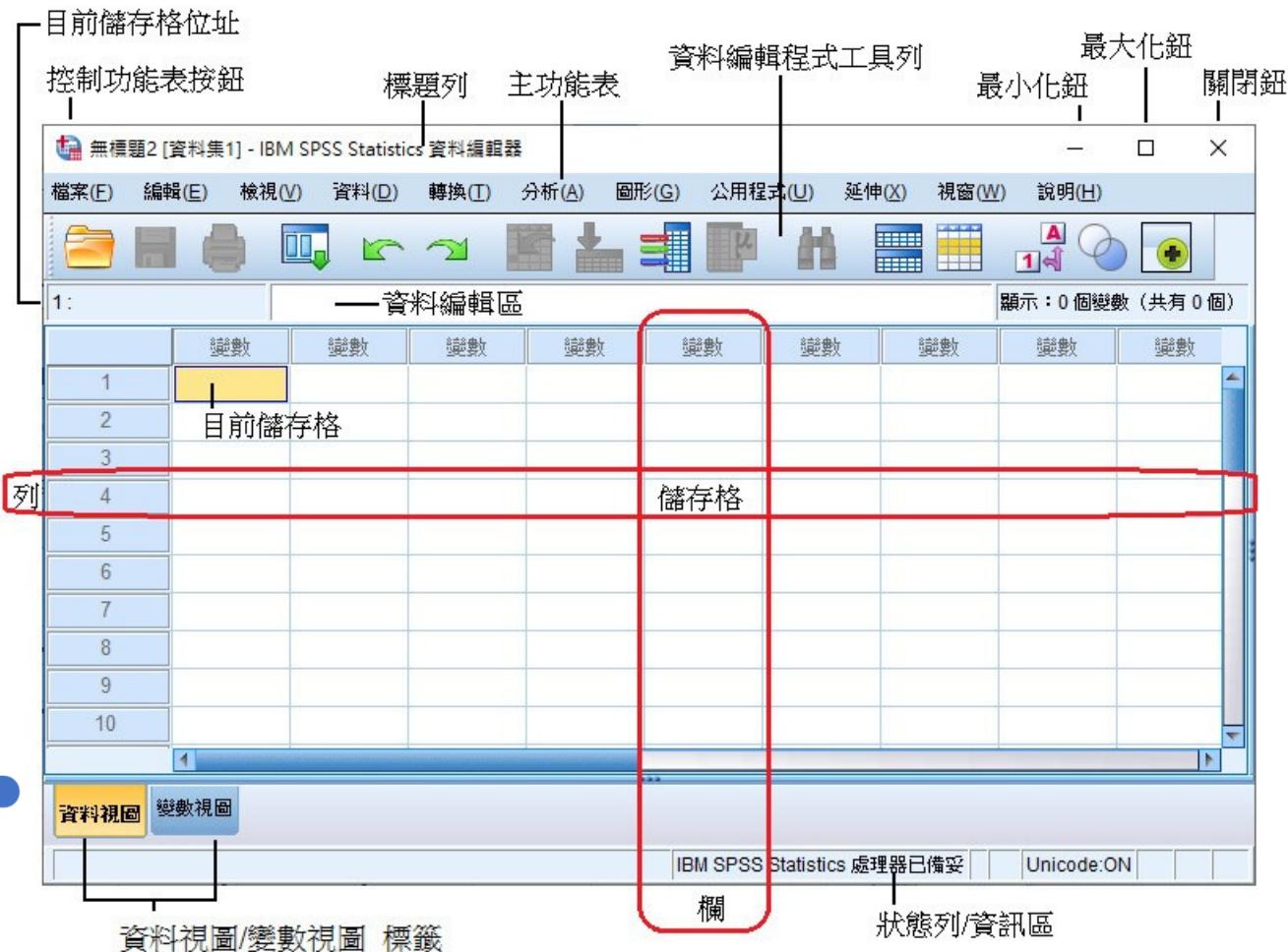
統計量

	是否飲用	啤酒廠牌	廠牌分組
N	有效	200	154
	遺漏	0	46
	平均數	1.23	2.36
	中位數	1.00	1.00
	標準差	4.22	2.536
			3.637

SPSS 支援檔案

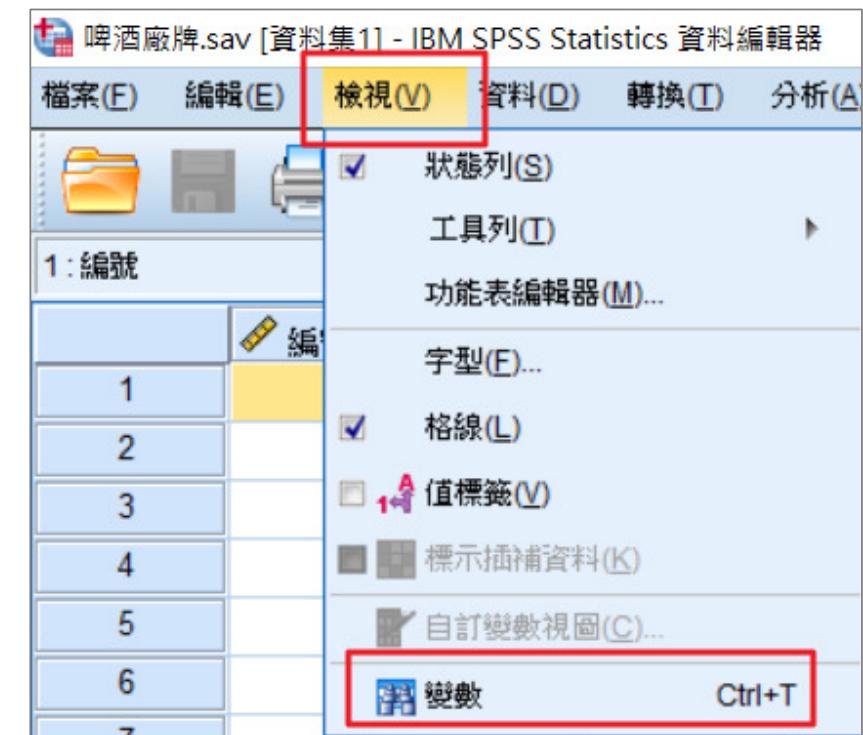
.CSPLAN	SPSS Sampling Plan File
.POR	SPSS Portable File
.SAS	SAS Program File
.SAV	SPSS Data File
.SBS	SPSS Script
.SGT	SPSS Chart Template
.SPD	SPSS Custom Dialog File
.SPE	SPSS Extension
.SPJ	SPSS Production Job File
.SPQ	SPSS Database Query File
.SPS	SPSS Program File
.SPV	SPSS Output Document
.STT	SPSS Table Template
.TLO	SPSS TableLooks File

資料視圖



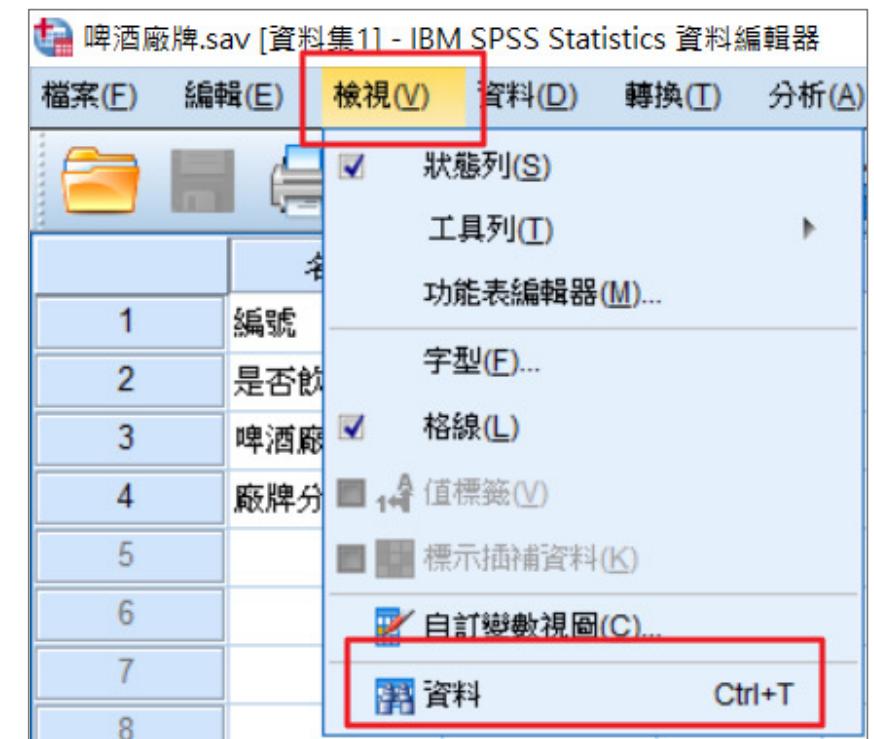
變數視圖

- 從「資料視圖」要切換到「變數視圖」畫面，使用方式為：
- 執行「檢視(V)/變數」
- 按左下角之 **變數視圖** 標籤來切換
- 按Ctrl +T 鍵
- 雙按任意之欄變數的標題按鈕



資料視圖

- 從「變數視圖」要切換到「資料視圖」畫面，使用方式為：
- 執行「檢視(V)/資料」
- 按左下角之 **資料視圖** 標籤來切換
- 按Ctrl+T鍵



定義變數

- 在進行輸入資料之前，得先定義各欄變數之名稱；
- SPSS 將依欄位順序給予 VAR00001 、 VAR00002 、 VAR00003, ... 。

	VAR00001	VAR00002
1	1.00	5.00
2	2.00	4.00
3	3.00	5.00

- 考慮輸入：編號、性別（1表男性；2表女性）與成績等三欄數值資料，可以使用以下步驟進行定義其變數名稱：

- 步驟1：執行「檢視(V)/變數」切換到『變數視圖』畫面
- 其內每一列，即用來定義一個欄變數之名稱及其資料類型。



- **步驟2：**於第1列之『名稱』處，輸入：編號，作為第1欄之變數名稱。按下Enter後，將自動補上其餘之相關定義的預設值。如：數字類型、寬度8、小數2。
- **步驟3：**編號並不須小數位，且寬度也不用那麼多。因此將其改為：寬度4、小數0。修改時，可直接鍵入新值或按其右側之上下箭頭進行調整。

	名稱	類型	寬度	小數
1	編號	數值	4	0

- 步驟4：依序於第2、3列之『名稱』處，輸入：性別與成績等作為各欄之變數名稱。且將其小數則均為0；寬度分別定為1與3。

	名稱	類型	寬度	小數
1	編號	數值	4	0
2	性別	數值	1	0
3	成績	數值	3	0
4				
	4			

資料視圖 變數視圖

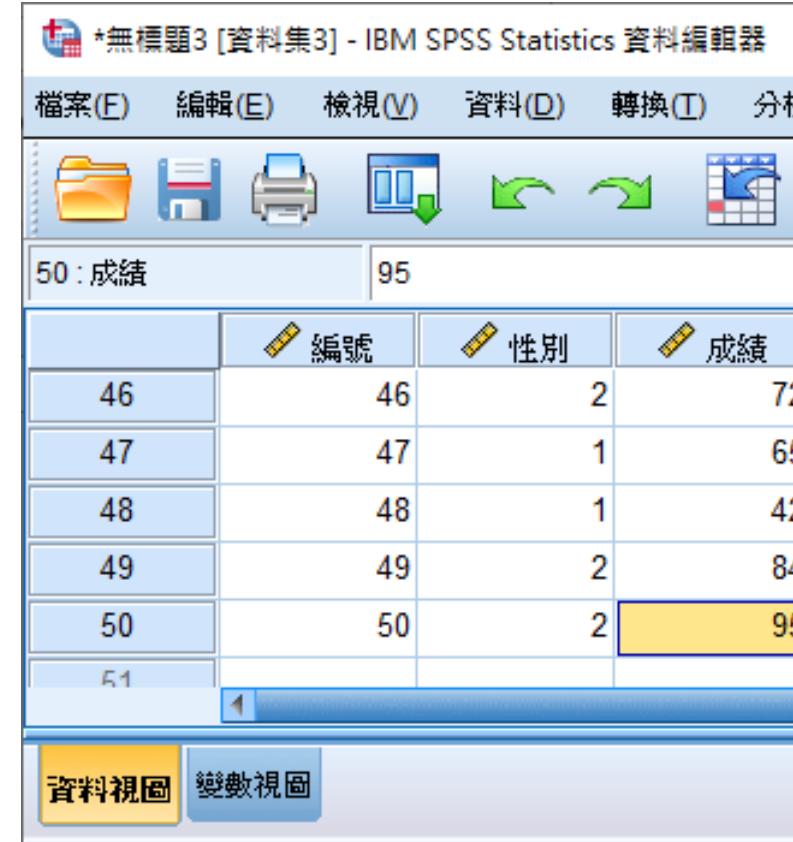
測量尺度(scale of measure)

- 名目尺度 (nominal scale) → SPSS 名義
 - 僅表示群或類別
 - 範例: 身分證號碼、性別、郵遞區號、縣市別
- 順序尺度 (ordinal scale) → SPSS 序數
 - 順序有別
 - 範例: 排名、年級、高度{高、中、短}
- 區間尺度;等距尺度 (interval scale)
 - 有絕對零點，差異或距離有意義
 - 範例: 日期、攝氏或華氏溫度
- 比例尺度;等比尺度(ratio scale)
 - 有自然零點，比率有意義
 - 範例: 克式(Kelvin)溫度、長度、時間、年齡

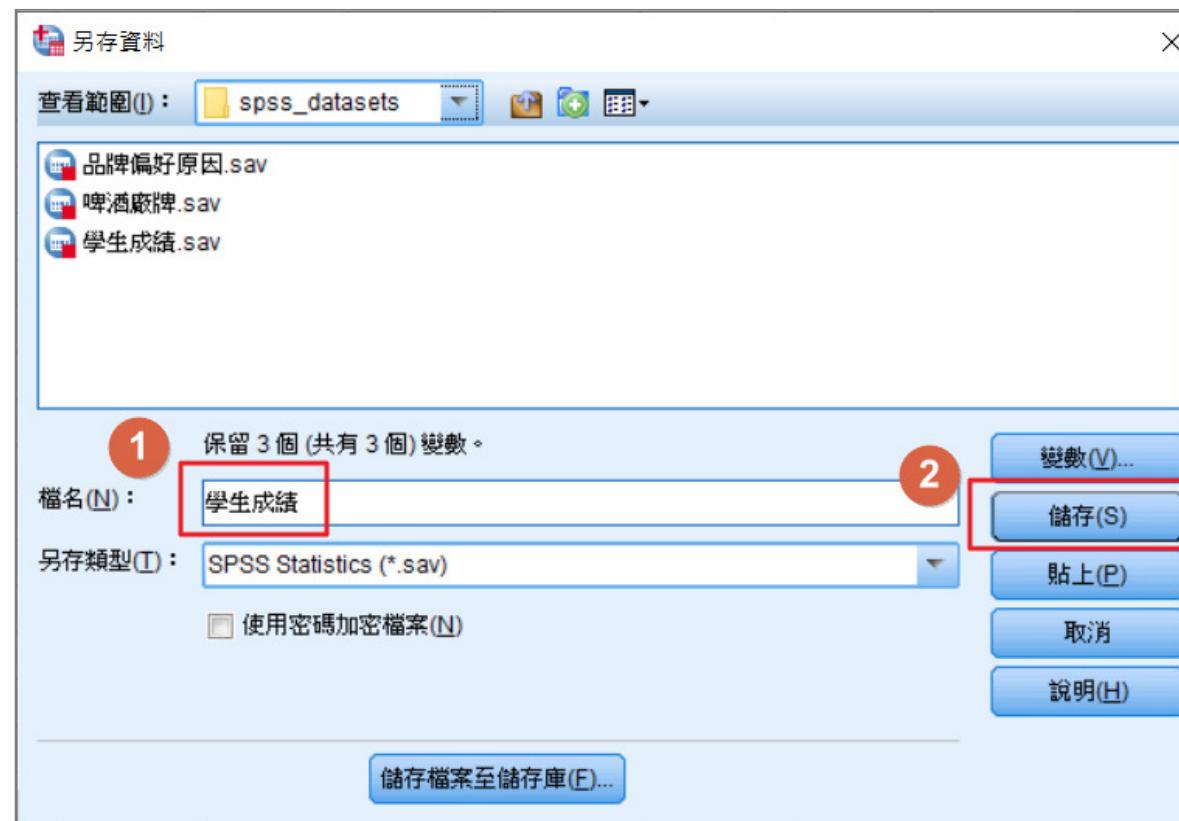


參考: https://en.wikipedia.org/wiki/Level_of_measurement

- **步驟5：**執行「檢視(V)/資料」切換回『資料視圖』畫面
 - 於各列輸入學生之編號、性別（1:男性；2:女性）與成績等三欄數值資料



- 步驟6：執行「檔案(F)/儲存」
- 或是按工具列上之「儲存此文件」
- 均將轉入『另存新檔』對話方塊，檔名：學生成績.sav



- 於SPSS，執行「檔案(F)/匯入資料(D)/Excel...」→未購原因.xlsx

實務應用可於 Excel
輸入完畢再匯入至 SPSS



2. 敘述統計分析

次數分配表

次數分配表 - 離散資料：性別

- 開啟「學生成績.sav」
- 分析 \ 敘述統計 \ 次數分配表



次數分配表



次數分配表結果

→ 次數分配表

[資料集1] C:\spss_datasets\學生成績.sav

統計量

性別

N	有效	50
	遺漏	0
平均數		1.54
中位數		2.00
標準差		.503
最小值		1
最大值		2
百分位數	25	1.00
	50	2.00
	75	2.00

右鍵

性別

	次數分配表	百分比	有效百分比	累積百分比
有效	1	23	46.0	46.0
	2	27	54.0	100.0
總計	50	100.0	100.0	

統計量		
性別		
N	有效	50
	遺漏	0
平均數		1.54
中位數		2.00
標準差		.503
最小值		1
最大值		2
百分位數	25	1.00
	50	2.00
	75	2.00

剪下

複製

複製為

貼上之後

建立/編輯自動 Script...

樣式輸出(E)...

匯出...

編輯內容(O)

可以複製到 Word

次數分配表 - 連續資料：成績

➔ 次數分配表

統計量

成績	
N	有效 50 遺漏 0
平均數	71.12
中位數	74.00
標準差	17.157
最小值	32
最大值	95
百分位數	25 59.50 50 74.00 75 85.25

成績

次數分配表		百分比	有效百分比	累積百分比
有效	32	2	4.0	4.0
	34	1	2.0	6.0
	39	1	2.0	8.0
	42	1	2.0	10.0
	47	2	4.0	14.0
	55	1	2.0	16.0



- 由於成績資料相同的並不多，因此成績分配較分散，故求其次數分配表已無多大意義！
- 改善方式：將資料分組為：『~59』、『60~79』與『80~』等三組，再進行求算分組後之次數分配表。

連續資料轉換為離散資料

- 執行「轉換(T)/重新編碼成不同變數(R)...」。



重新編碼成不同變數

- 左側選取「成績」，按 鈕，將其送到右側「數值變數 -> 輸出變數：」方塊，出現「成績 --> ?」，表示尚未進行轉換。



舊值與新值

- 按 **舊值與新值(O)...**，選「範圍，從最低到值(G)：」，於方塊輸入59。
- 於『新值』之「值(L)：」處輸入1，表示由最低分到59分將被歸到新值1。



舊值 → 新值(D)：
Lowest thru 59 → 1

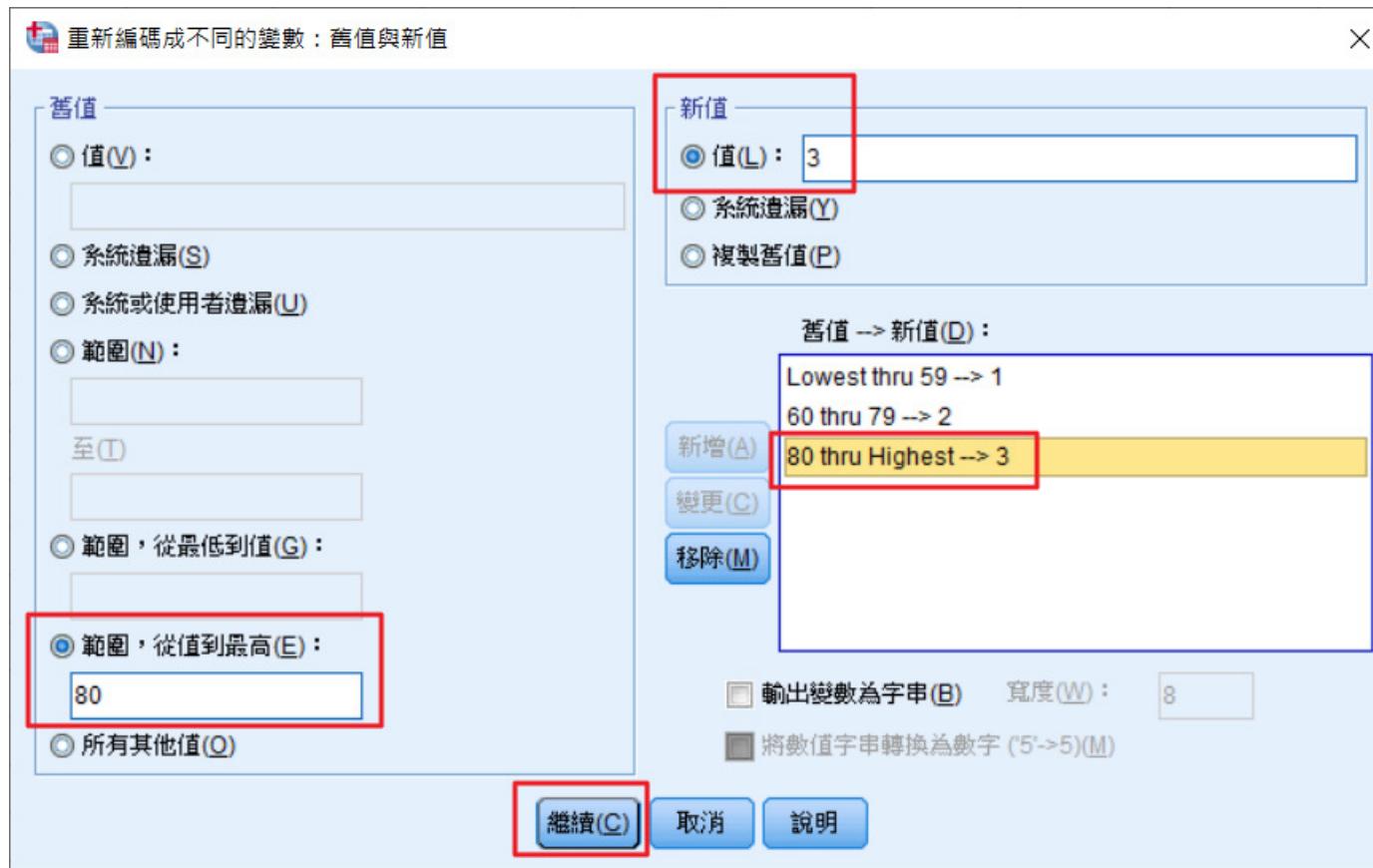
舊值與新值 (續)

- 選「範圍(N)」，定義60~79將被歸到新值2。



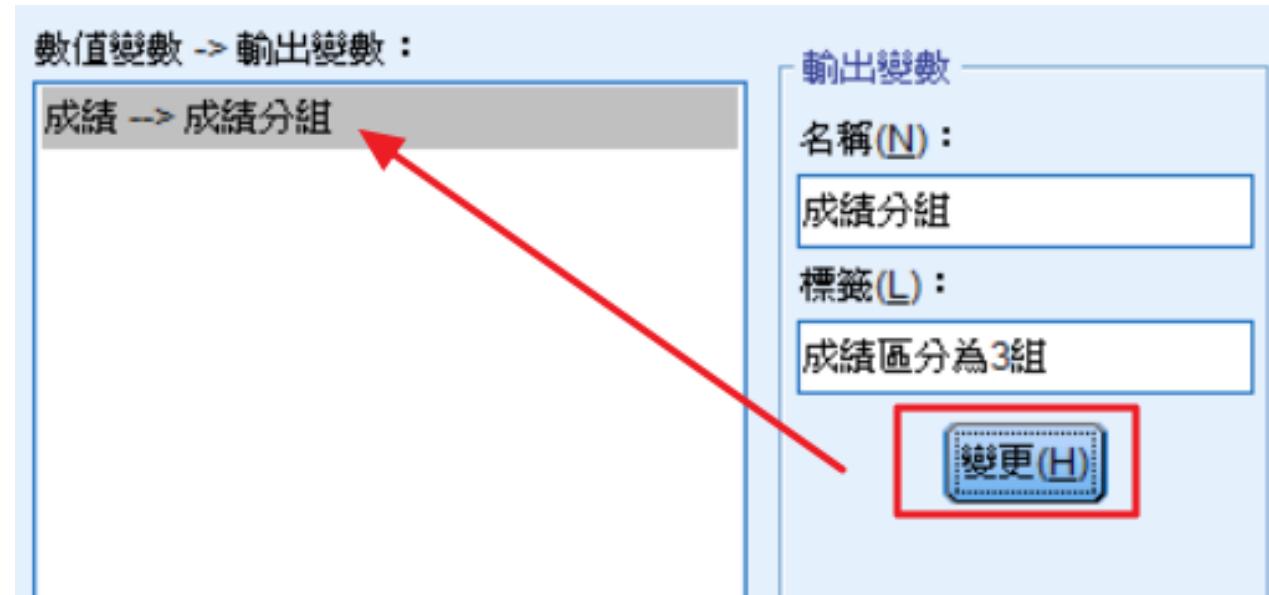
舊值與新值 (續)

- 選「範圍，從值到最高(E)」，定義80分以上，將被歸到新值3。



舊值與新值 (續)

- 按〔變更〕鈕，出現「成績 --> 成績分組」，表示已進行變更動作，將新值置入「成績分組」變數。

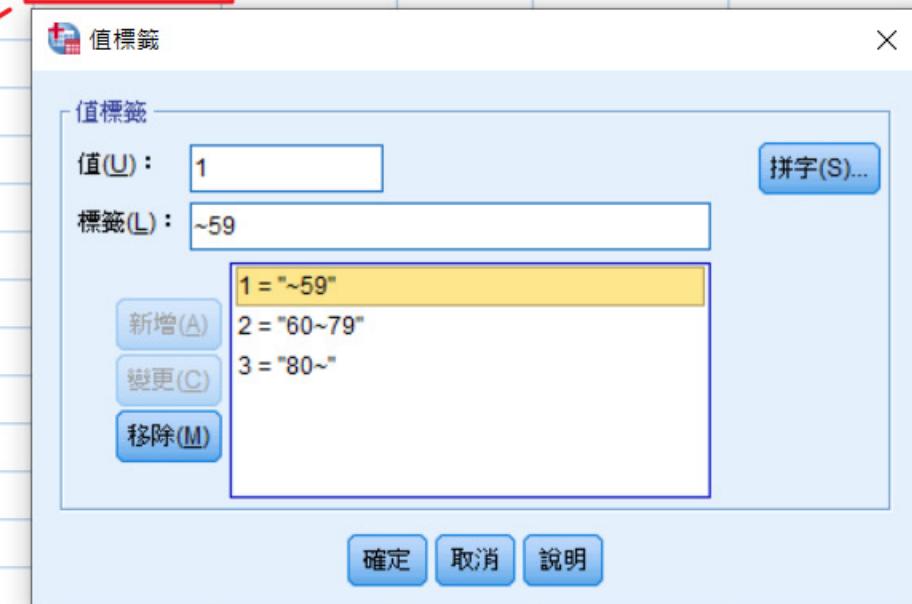


離散資料完成版

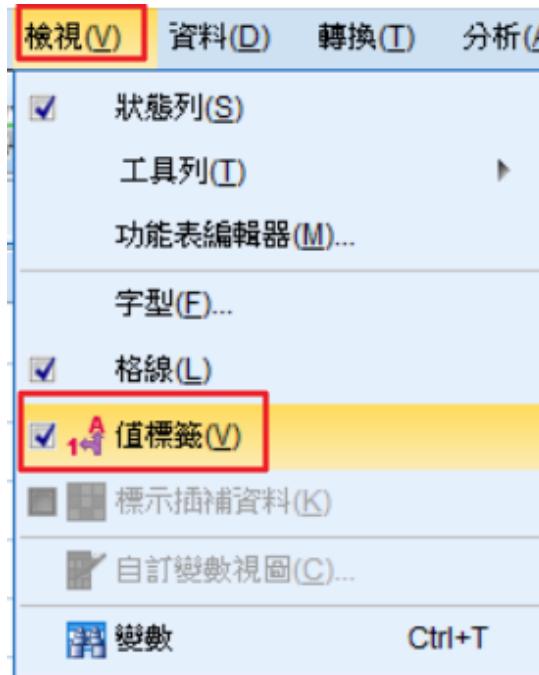
	編號	性別	成績	成績分組
1	1	1	34	1
2	2	2	71	2
3	3	1	83	3
4	4	2	65	2
5	5	2	58	1

設定值標籤

	名稱	類型	寬度	小數	標籤	值	遺漏	欄	對齊	測量	角色
1	編號	數值	4	0		無	無	8	靠右	尺度	輸入
2	性別	數值	1	0		無	無	8	靠右	名義	輸入
3	成績	數值	3	0		無	無	8	靠右	尺度	輸入
4	成績分組	數值	8	0	成績區分為3組	{1, ~59}...	無	14	靠右	名義	輸入
5											
6						{1, ~59}...	...				
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											



顯示值標籤



	編號	性別	成績	成績分組
1	1	1	34	~59
2	2	2	71	60~79
3	3	1	83	80~
4	4	2	65	60~79
5	5	2	58	~59

連續轉換為離散資料 - 次數分配表

→ 次數分配表

統計量

成績區分為3組

N	有效	50
	遗漏	0
平均數		2.18
中位數		2.00
標準差		.800
最小值		1
最大值		3
百分位數	25	1.75
	50	2.00
	75	3.00

成績區分為3組

次數分配表		百分比	有效百分比	累積百分比
有效	~59	12	24.0	24.0
	60~79	17	34.0	58.0
	80~	21	42.0	100.0
	總計	50	100.0	100.0

次數分配表較合理 ^_^

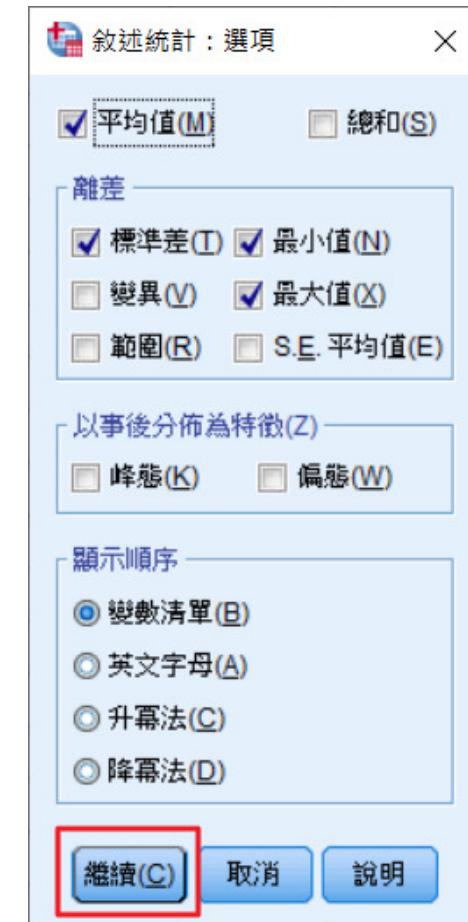
敘述統計

敘述統計

- 分析 \ 敘述統計(E) \ 敘述統計(D)



敘述統計選項



敘述統計完成圖

→ 敘述統計

敘述統計

	N	最小值	最大值	平均值	標準偏差
成績	50	32	95	71.12	17.157
有效的 N (listwise)	50				

補充篇：圖表建置器

圖表建置器簡介

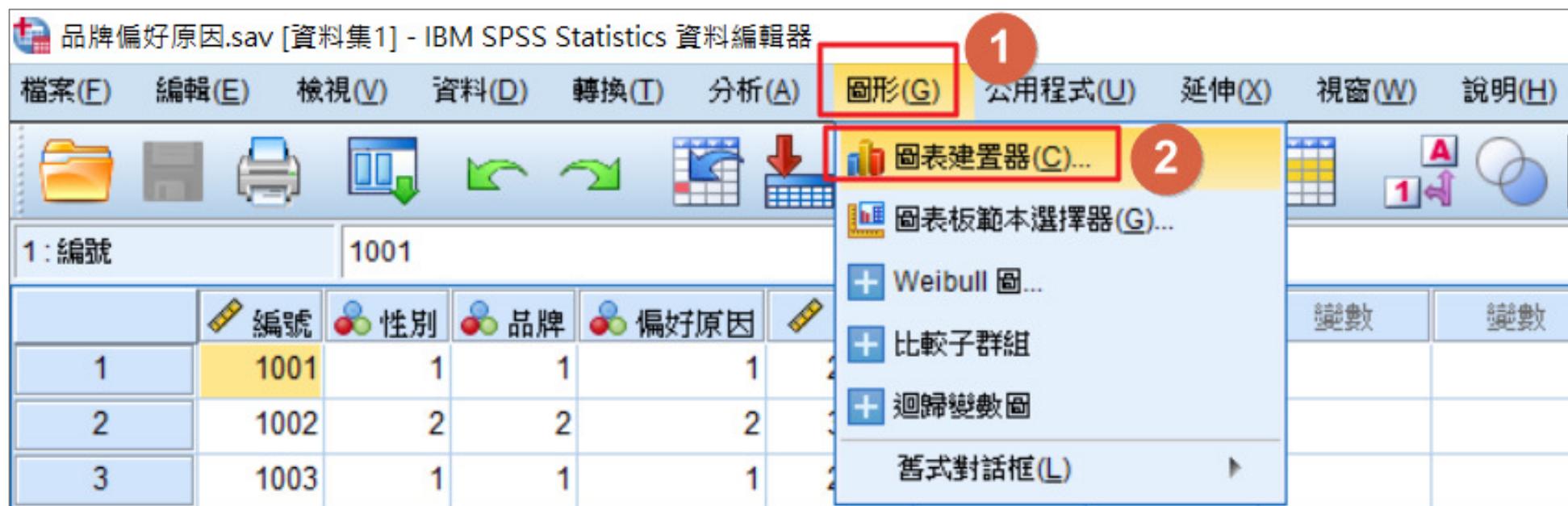
圖表建置器簡介

- 「圖表建置器」可讓您從預先定義的展示區圖表或個別元素（例如軸及長條）建立圖表。
- 可以將展示區圖表或基本元素拖放到畫布（為「圖表建置器」對話框中「變數」清單右邊的大塊區域），以建置圖表。
- 長條圖，線條圖，區域圖，圓餅圖，散佈圖，直方圖，高低圖，箱型圖（盒鬚圖），雙軸圖

圖表建置器操作

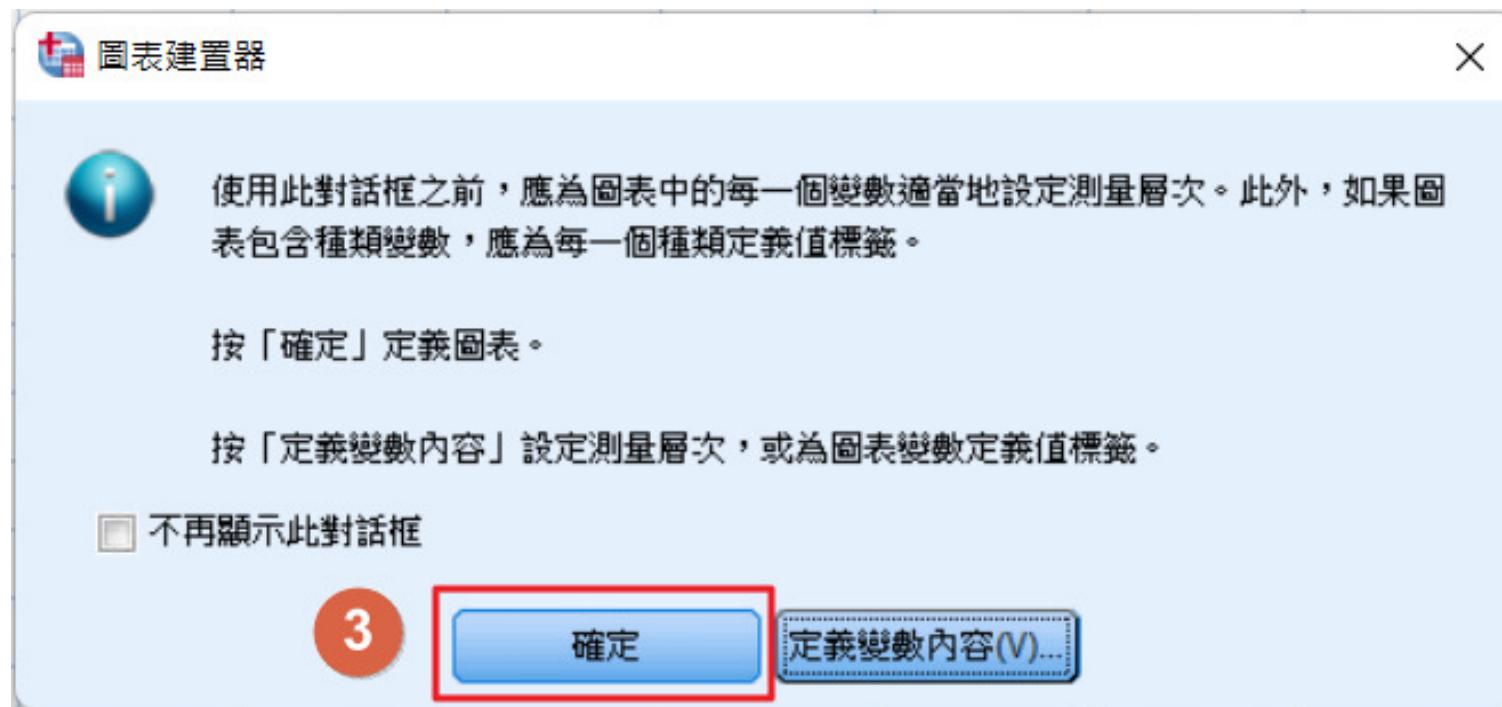
圖表建置器操作

- 檔案: 品牌偏好原因.sav
- 圖形 \ 圖表建置器



圖表建置器操作 (續)

- 按 確定



集群長條圖



圖表預覽使用範例資料

X上的集群: 設定顏色

性別

編號

性別

品牌

偏好原因

所得

A牌

B牌

C牌

展示區

基本元素

群組/點 ID

標題/註腳

選擇來源(C):

我的最愛

長條圖

線條圖

區域圖

圓餅圖/極座標圖

散佈圖/點狀圖

直方圖

「高-低」圖

箱型圖

雙重軸

確定

貼上(P)

重設(R)

取消

說明

元素內容

圖表外觀

選項

編輯此項目的內容(D):

長條圖1

X軸1 (長條圖1)

Y軸1 (長條圖1)

文字類型：標題 1

自動(U)

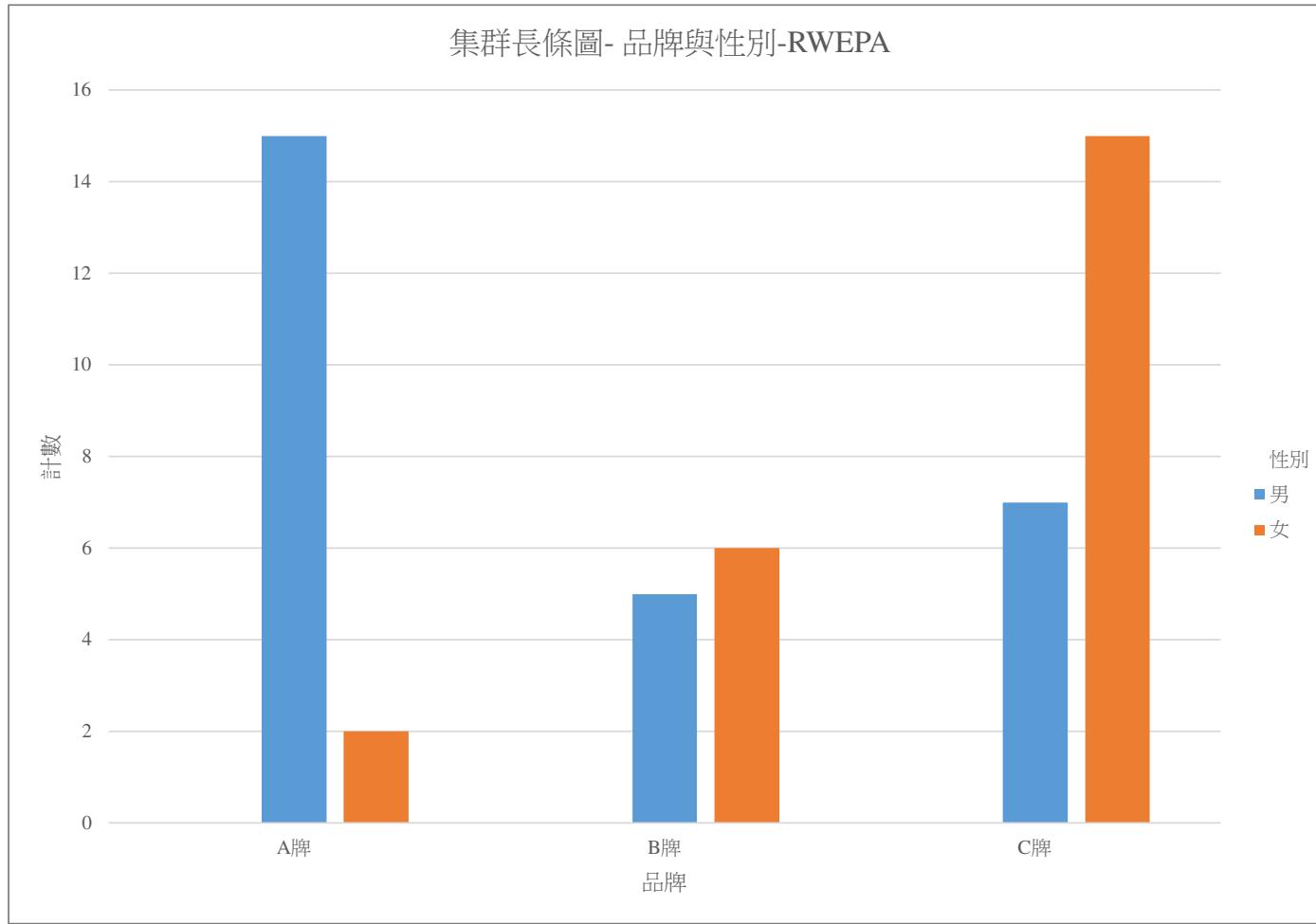
集群長條圖 計數 / 品牌 依據 性別

自訂(C)

集群長條圖- 品牌與性別-RWEPA

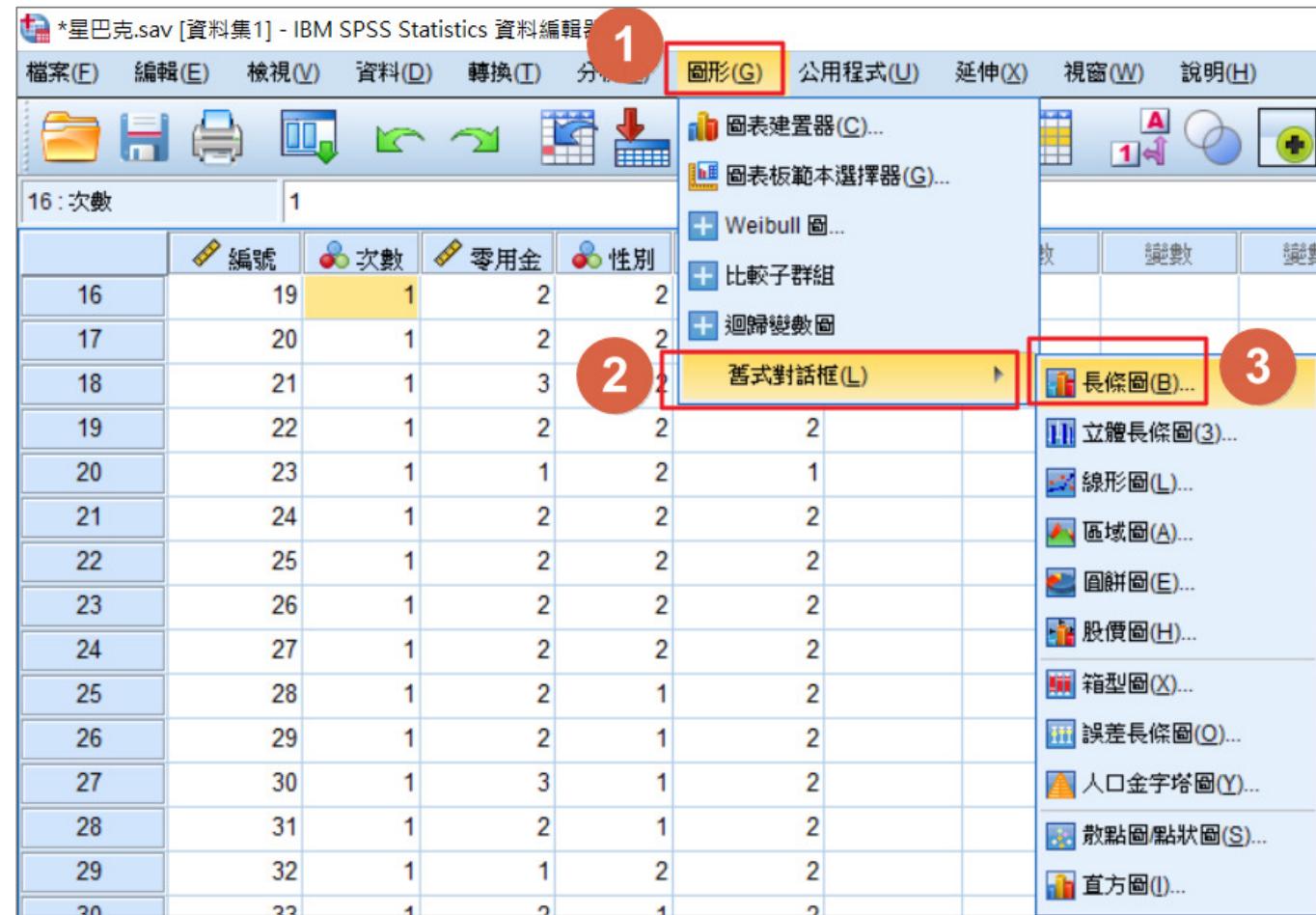
無(N)

集群長條圖 (完成版)

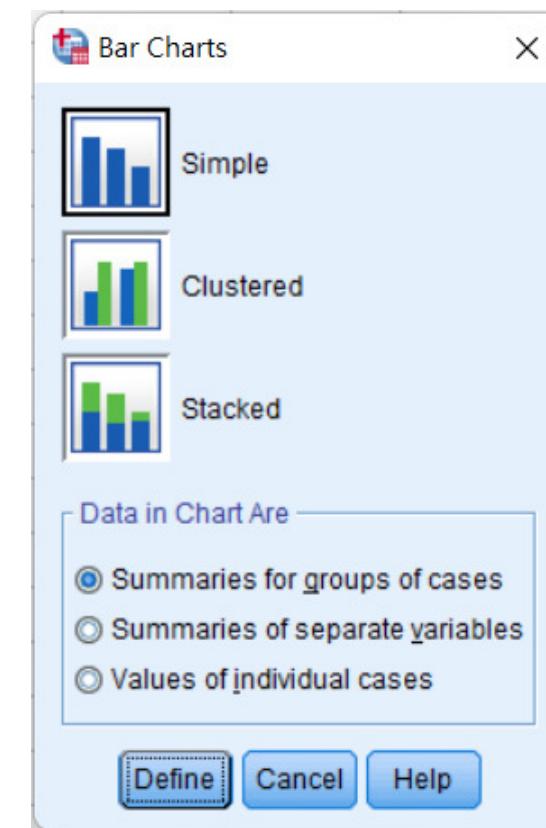


圖形 \ 舊式對話框 \ 長條圖

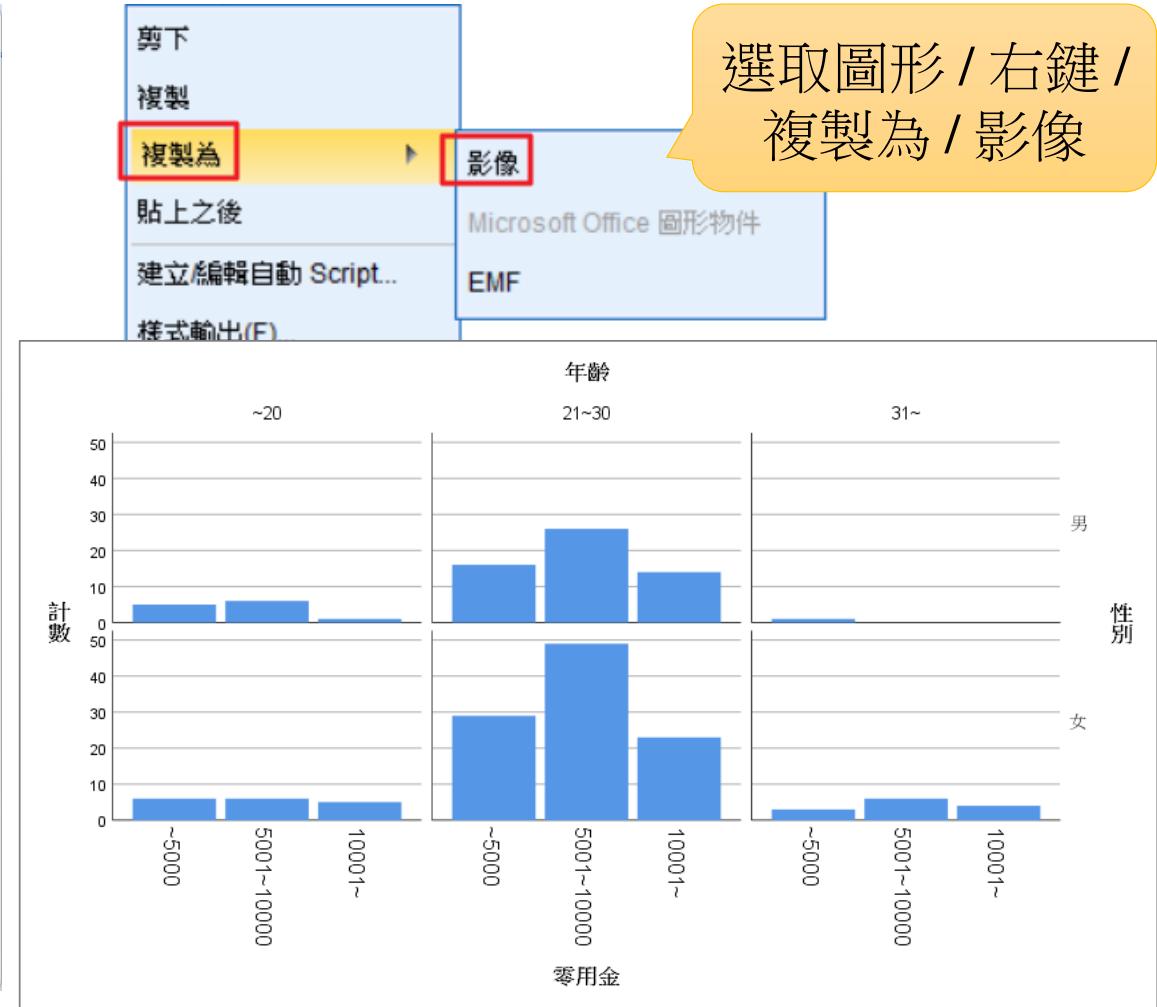
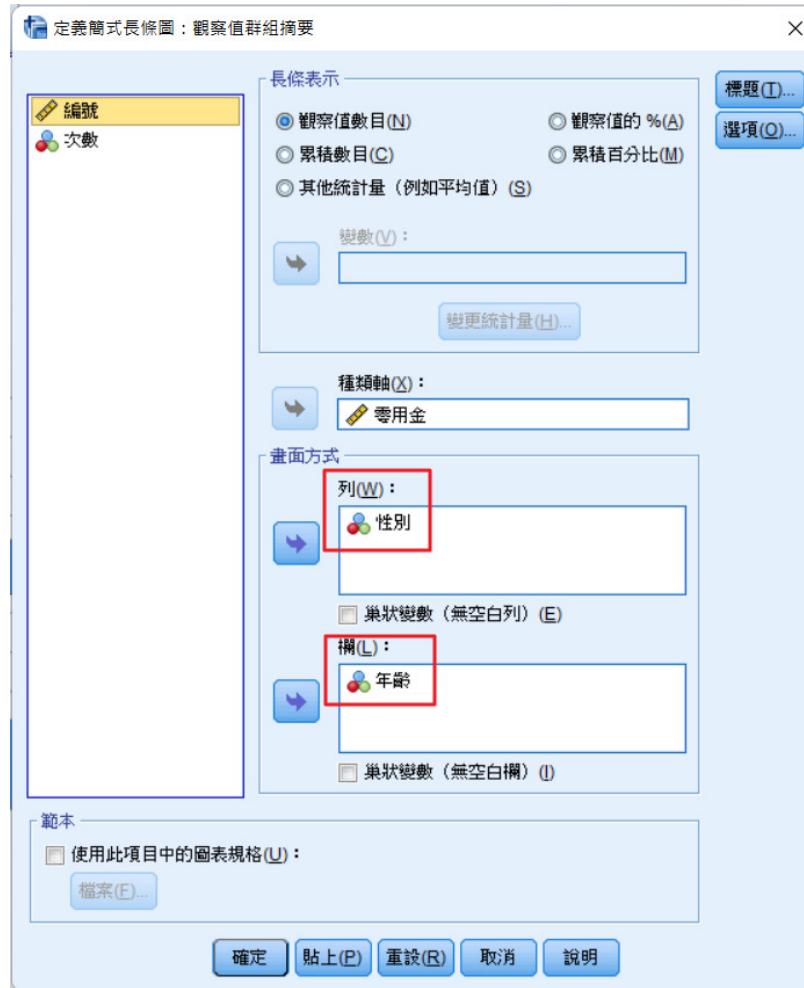
圖形 \ 舊式對話框 \ 長條圖



長條圖 \ 簡式 \ 觀察值群組摘要 \ 定義



觀察值群組摘要 (結果)



參考資料

- RWEPA: <http://rwepa.blogspot.com/>
- 楊世瑩, SPSS 26統計分析嚴選教材(適用SPSS 26~22), 暮峰, 2020.
- 線上說明:
<https://www.ibm.com/docs/zh-tw/db2woc?topic=spss-tutorial-analyzing-statistics>

謝謝您的聆聽

Q & A



李明昌

alan9956@gmail.com

<http://rwepa.blogspot.tw/>