

第六單元 圖形的製作

梁文敏 副教授 葉懿諄 統計分析師

圖形製作的目的主要是利用視覺化將資料的特質更有效率地表現出來。如何選擇適當圖形，可依照資料類型為連續或類別來加以區分。針對單一變數的特質分布，當資料為連續變數，我們常利用盒形圖、直方圖來表現；當資料為類別變數，則常利用圓餅圖、條形圖(長條圖)來表現。此外，圖形還可表現多個變數的關係，例如：以條形圖來表現不同組別中，某一連續變數的平均值。

6.1 連續變數的分布

【範例 1】研究者欲瞭解受訓學員睡眠小時(sleepHR1)的分布。

◎ 範例資料：

id	sleepHR1
1	8.5
2	7.3
3	7
...	...
50	7.3

◎ 使用方法：

從問題中，我們要探討的變數為睡眠時數是一個連續(等距，間斷)變數。接下來，我們將介紹如何利用 SPSS 來繪製直方圖及盒形圖。

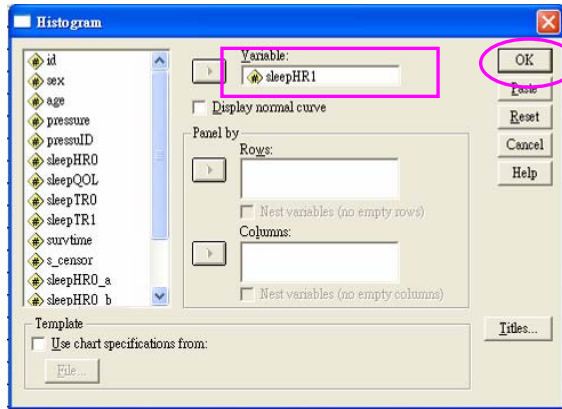
(1) 直方圖的繪製

◎ 軟體操作步驟：

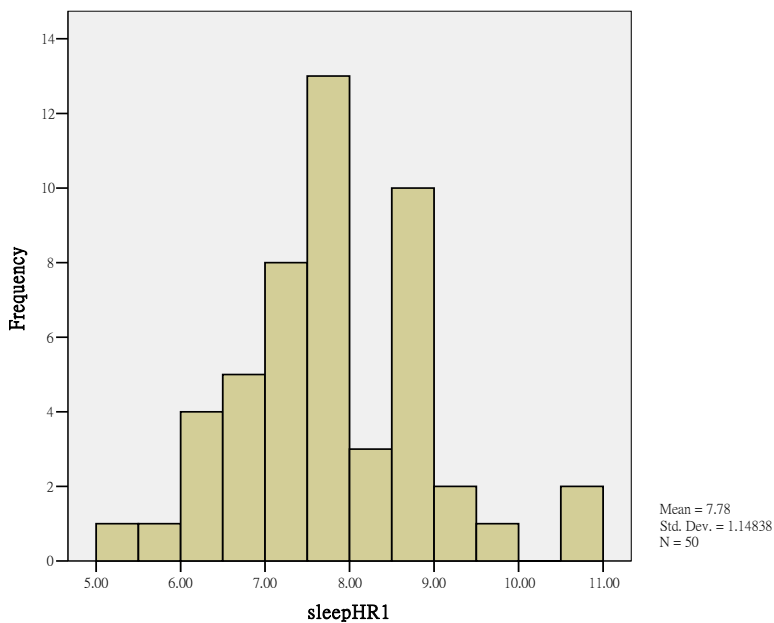
步驟一：選擇功能【Graphs 統計圖】→進入【Histogram 直方圖】。



步驟二：在 Variable 變數(選入變數：sleepHR1)→OK 確定。



輸出結果：



從上圖中我們可以發現每個長條是連接在一起的(此為直方圖的特質)，其中長條在 7.5 到 8 小時之間的長條最高，表示這區間的人最多，為 13 人；其次為 8.5 到 9 小時有 10 人，所有人的睡眠小時介在 5 至 11 小時，利用圖形可以快速地讓人瞭解資料的分布。

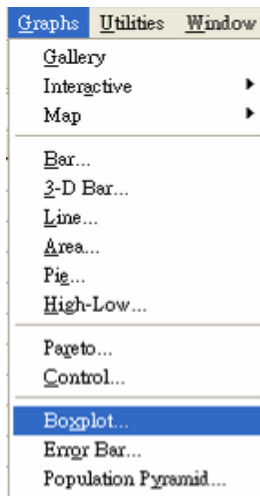
SAS 小教室

```
PROC UNIVARIATE PLOT;  
VAR sleepHR1;  
RUN;
```

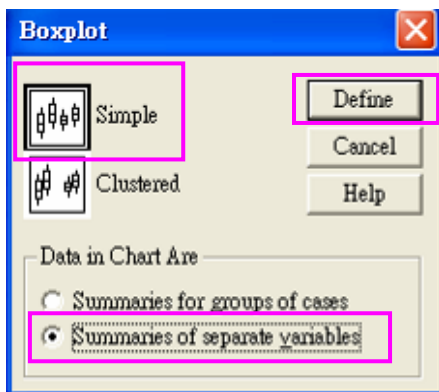
(2) 盒形圖的繪製

軟體操作步驟：

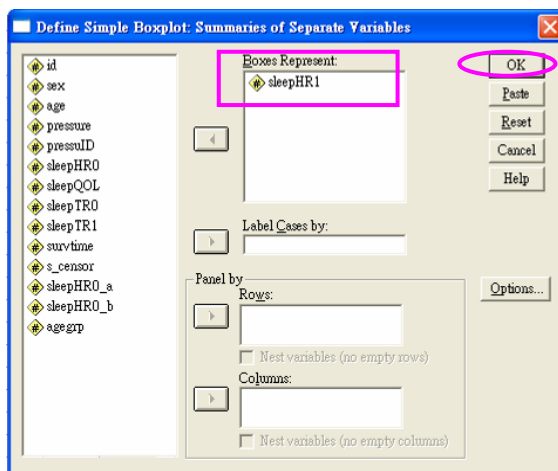
步驟一：選擇功能【Graphs 統計圖】→進入【Boxplot 盒形圖】。



步驟二：選擇Simple簡單圖形→在Data in Chart Are圖表中資料為方框中→點選：Summaries of separate variables各個變數之摘要→選擇進入：Define定義。



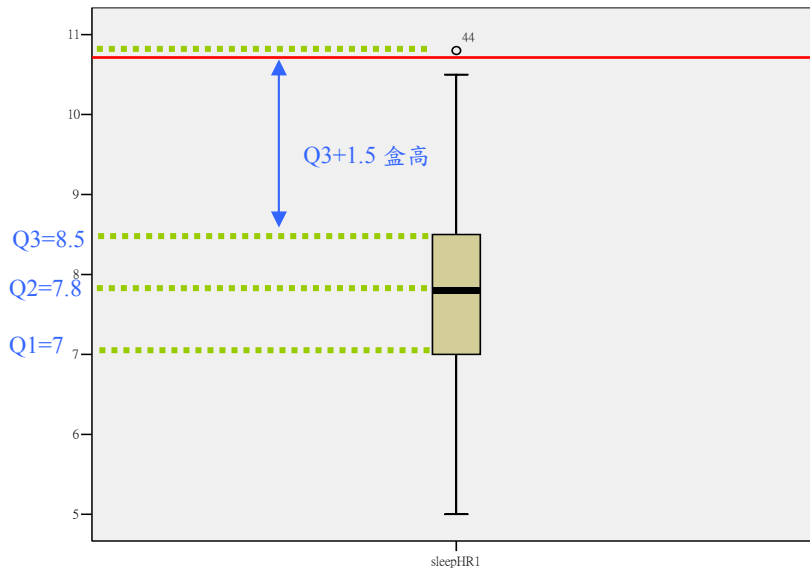
步驟三：在 Boxes Represent 盒形圖表示方框中（選入變數：sleepHR1）→OK 確定。



📌 輸出結果：

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
sleepHR1	50	47.6%	55	52.4%	105	100.0%



盒形圖是一種用來表現單一連續變數分布的圖形，在盒子及上下鬚之間表示大部分資料的分布範圍，落在上下鬚之外的值則稱為界外值或極端值，會一一在圖形中標示出來。盒形圖中用到的指標為四分位數 Q1(即第 25 個百分位)、Q2(即第 50 個百分位)及 Q3(即第 75 個百分位)。從上圖中我們可以發現盒子的下界(Q1)為 7 小時，上界(Q3)為 8.5 小時，而中位數為 7.8 小時，從上圖中我們可以知道有 50%的人睡眠小時大約介在 7~8.5 小時之間。在圖的上方可以看到有一個圓圈標“44”的觀測值，這表示編號“44”是在上鬚(Q3+1.5 盒高=10.75)的界線之外，稱為界外值(outlier)，其數值為 10.8 小時。

SAS 小教室

```
PROC UNIVARIATE PLOT;  
VAR sleepHR1;  
RUN;
```

6.2 類別變數的分布

【範例 2】研究者欲瞭解壓力程度(pressure)的分布。

◎ 範例資料：

id	pressure
1	0
2	1
3	2
...	...
50	2

註：pressure (0:無壓力，1:壓力中等，2:壓力大)

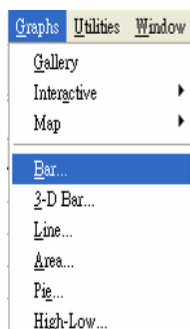
◎ 使用方法：

因壓力程度(pressure)為類別(序位)資料而非連續性資料。所以在範例中我們利用條形圖來表現壓力程度的分布。

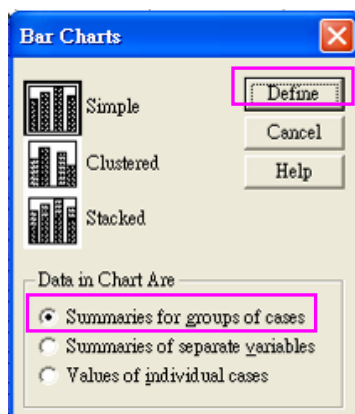
(1) 條形圖(長條圖)的繪製

◎ 軟體操作步驟：

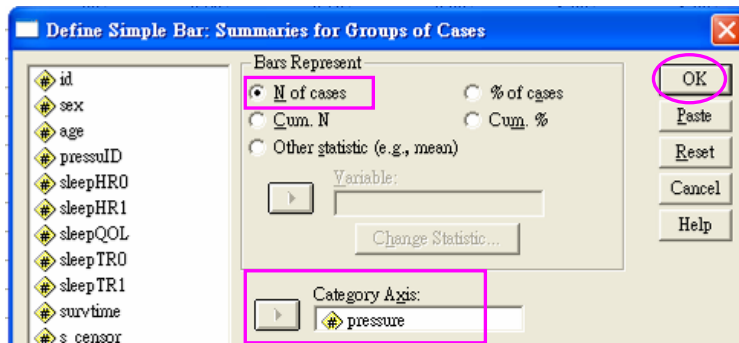
步驟一：選擇功能【**Graphs 統計圖**】→進入【**Bar 條形圖**】。



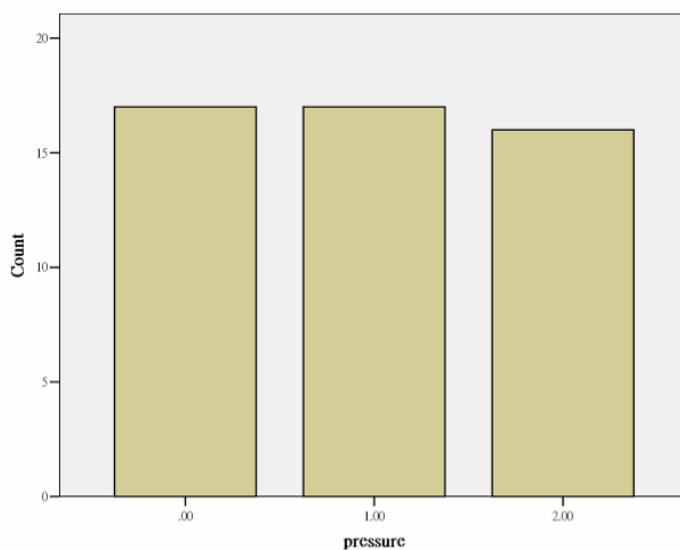
步驟二：選擇Simple簡單圖形→在下方Data in Chart Are圖表中資料為方框中→點選：Summaries for groups of cases 觀察值組別之摘要→選擇進入：Define 定義。



步驟三：在 Bars Represent 條形圖表示方框中→點選：N of cases 觀察值個數→在 Category Axis 類別軸方框中（選入變數：pressure）→OK 確定。



🕒 輸出結果：



從上圖中的橫軸為壓力程度(pressure)，縱軸為個數。圖中可以看到 50 位參與者中無壓力組(pressure=0)有 17 人；壓力中等組(pressure=1)有 17 人；壓力大組 (pressure=2)有 16 人。

SAS 小教室

```
PROC GCHART DATA=new;
VBAR pressure / DISCRETE SNMVAR= pressure ;
RUN;
QUIT;
```

🕒 小小叮嚀

直方圖及條形圖常讓初學者覺得混淆不清，其分辨的秘訣在於其橫軸，直方圖的橫軸用以表現連續的尺度，所以每一個長條之間是沒有間隔的。而長條圖的橫軸則是用以表示類別或序位尺度，所以每一個長條之間有間隔。

6.3 兩個變數的關係

【範例 3】研究者欲瞭解在不同的壓力程度(pressure)下的平均睡眠小時(sleepHR1)。

◎ 範例資料：

id	pressure	sleepHR1
1	0	8.5
2	1	7.3
3	2	7
...
50	2	7.3

註：pressure (0:無壓力，1:壓力中等，2:壓力大)

◎ 使用方法：

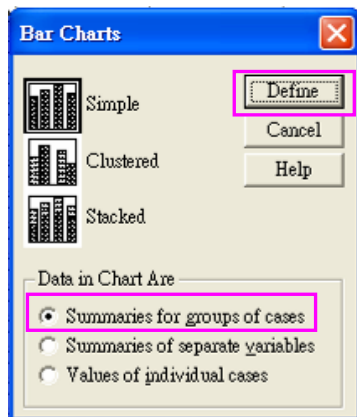
條形圖除了可以用來表現質性(類別)資料的分布外，還可用表現多個變數的關係，例如在本例中以條形圖(長條圖)來表現不同壓力程度下的平均睡眠小時。

◎ 軟體操作步驟：

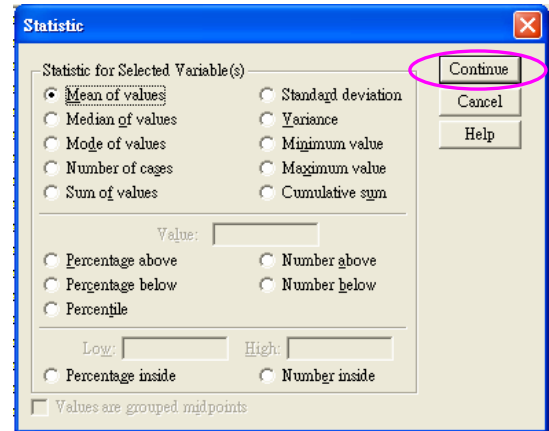
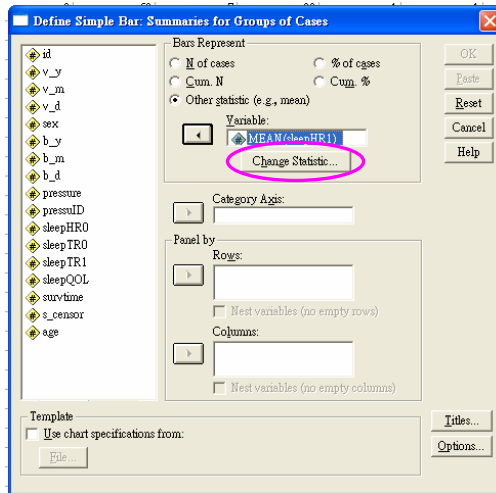
步驟一：選擇功能【Graphs 統計圖】→進入【Bar 條形圖】。



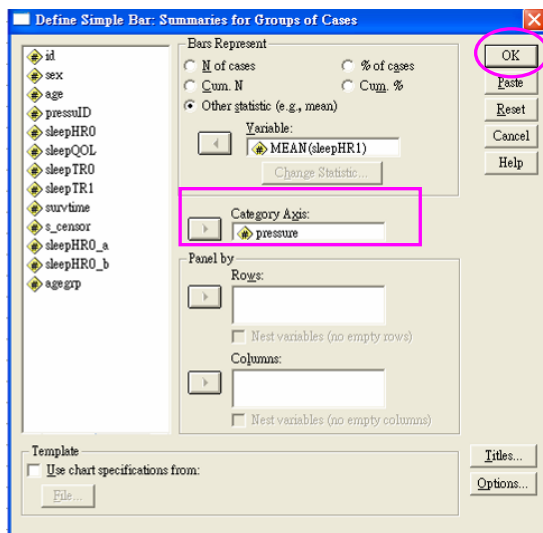
步驟二：選擇 Simple 簡單圖形→在下方 Data in Chart Are 圖表中資料為方框中→點選：Summaries for groups of cases 觀察值組別之摘要→選擇進入：Define 定義。



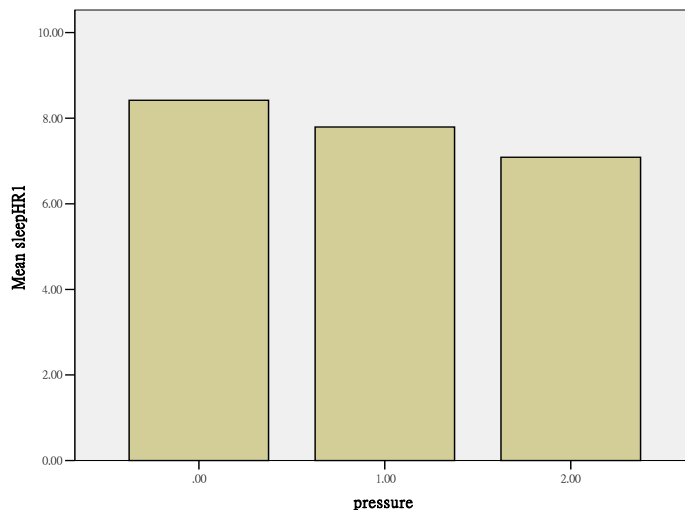
步驟三：在 Bars Represent 條形圖表示方框中 → 點選：Other statistics 其它摘要函數（選入變數：sleepHR1）→ 點選：Change Statistic... 變更摘要 → 點選 Mean of values 平均值 → Continue 繼續。



步驟三：在 Category Axis 類別軸方框中（選入變數：pressure）→ OK 確定。



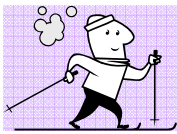
📍 **輸出結果：**



上圖中橫軸為壓力程度(pressure)，縱軸為平均睡眠時數(Mean sleepHR1)。圖中可以看到一個很有趣的現象，平均睡眠時數會隨著壓力程度的增加而減少。無壓力組(pressure=0)的平均睡眠時數大約為 8.2 小時；壓力中等組(pressure=1)的平均睡眠時數大約為 7.9 小時；壓力大組(pressure=2)的平均睡眠時數大約為 7 小時。上圖中我們可以發現隨著壓力的增加，平均睡眠時數減少。

SAS 小教室

```
PROC GCHART DATA=new;  
VBAR pressure / DISCRETE MEANVAR= sleepHR1 ;  
RUN; QUIT;
```



練習

Q1: 請繪製年齡(age)的分布圖。

- 1) 你會選擇哪一種圖形來表示? _____
- 2) 中間 50%的人年齡所在範圍? _____

Q2: 請繪製睡眠品質(sleepQOL)的分布圖。

- 1) 你會選擇那一種圖形來表示? _____
- 2) 哪一組人數最少? _____

Q3: 繪製不同壓力程度(pressure)的平均年齡(age)。

- 1) 你會選擇那一種圖形來表示? _____
- 2) 哪一組最年輕? _____
- 3) 哪一組最年長? _____