

加權變項於統計軟體 SPSS 與 STATA 上之使用說明

蘇婉雯

加權是抽樣調查方法論中討論很久的一個議題，在母體人口基本結構可取得的前提下，當成功樣本的結構與母體有差異，而無法通過樣本代表性檢定時，加權是可以考慮的調整方法之一；近年來「學術調查研究資料庫」所蒐藏的一部份政府調查資料與大型面訪調查資料，例如：主計總處人力資源調查、台灣社會變遷基本調查、華人家庭動態調查與台灣教育長期追蹤資料庫等，都有針對釋出資料進行加權處理，以提供資料使用者進行分析時，能對母體有更好的推估。

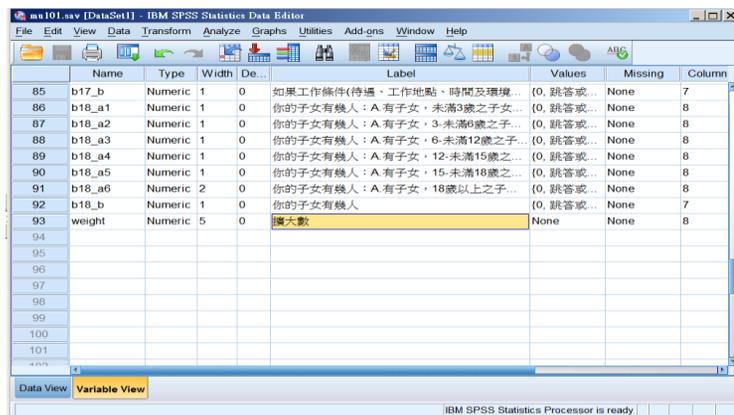
加權權值的計算有各種方法，如中選率的倒數、事後分層加權法(post-stratification)、多變項反覆加權法(raking)等等，這些通常是抽樣課程中的一部分，而本文將著重於利用資料庫現有的資料當範例，說明於統計軟體(SPSS、STATA)上如何使用權數變項，對加權權值的計算則不多著墨，請讀者多包涵。

調查資料檔中常見的加權權值大致包含兩種類型：

1. **擴大數**：意指每個個案所代表的母體人數，又稱為放大係數或膨脹係數，加權後樣本數相當於母體人數。如：主計總處人力資源調查。
2. **權值**：將調查資料的人口結構調整至與母體結構一致。常採用的結構有性別、年齡、地區、甚或教育程度等。加權後資料樣本數仍維持完訪案數。如：台灣社會變遷基本調查。

一、SPSS 介紹：

在 SPSS 中宣告擴大數或權值的方法一樣。以「行政院主計總處 101 年人力運用調查」的資料為例，以 SPSS 開啟檔案後資料畫面如圖 1。



	Name	Type	Width	Dec	Label	Values	Missing	Column
85	b17_b	Numeric	1	0	如果工作條件(待遇、工作地點、時間及環境...	{0, 跳答或...	None	7
86	b18_a1	Numeric	1	0	你的子女有幾人：A 有子女，未滿3歲之子女...	{0, 跳答或...	None	8
87	b18_a2	Numeric	1	0	你的子女有幾人：A 有子女，3-未滿6歲之子...	{0, 跳答或...	None	8
88	b18_a3	Numeric	1	0	你的子女有幾人：A 有子女，6-未滿12歲之子...	{0, 跳答或...	None	8
89	b18_a4	Numeric	1	0	你的子女有幾人：A 有子女，12-未滿15歲之...	{0, 跳答或...	None	8
90	b18_a5	Numeric	1	0	你的子女有幾人：A 有子女，15-未滿18歲之...	{0, 跳答或...	None	8
91	b18_a6	Numeric	2	0	你的子女有幾人：A 有子女，18歲以上之子...	{0, 跳答或...	None	8
92	b18_b	Numeric	1	0	你的子女有幾人	{0, 跳答或...	None	7
93	weight	Numeric	5	0	擴大數	None	None	8
94								
95								
96								
97								
98								
99								
100								
101								

圖 1. SPSS 的資料編輯視窗畫面

此項調查資料的權數類型為擴大數，變項名稱為 weight。表一為加權前性別變項的次數分配。

表1. 未加權前人力運用調查性別變項(a2)次數分配

		a2 性別			
		次數	百分比	有效百分比	累積百分比
有效的	1 男	28933	49.9	49.9	49.9
	2 女	29100	50.1	50.1	100.0
	總和	58033	100.0	100.0	

【步驟 1】SPSS 工作列中點選 Data，開啟下拉選單後點選 Weight Cases，如圖 2。中文版介面則點選『資料』→『加權觀察值』，如圖 2-1。

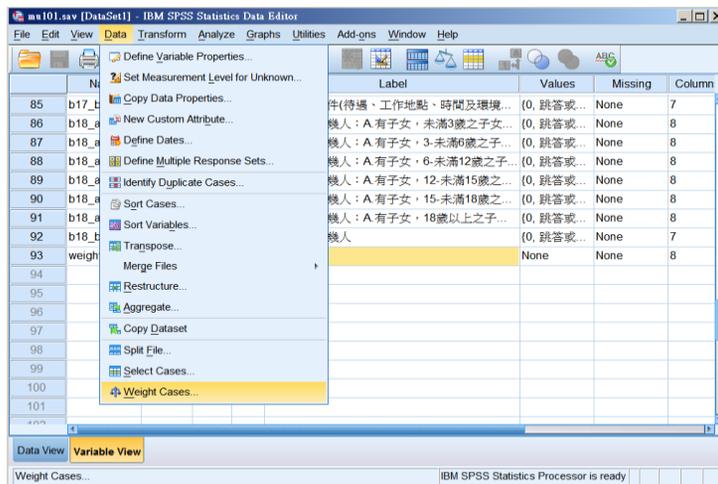


圖 2.SPSS 工作列中的 Data 選單

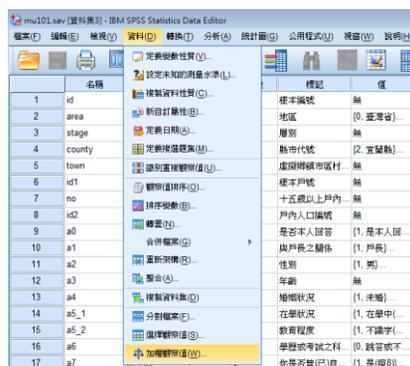


圖 2-1.SPSS 工作列中的 Data 選單（中文介面）

【步驟 2】在 Weight Cases 中點選 Weight cases by，將加權變項拉進 Frequency Variable 後按下 OK，如圖 3；中文介面則選擇「觀察值加權依據」，如圖 3-1。

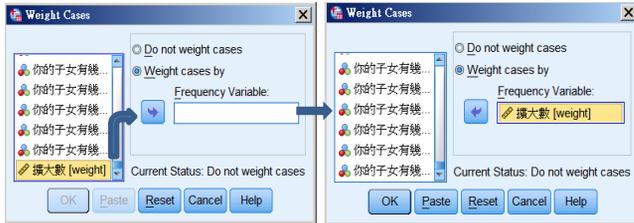


圖 3.將加權變項拉進 Frequency Variable 中

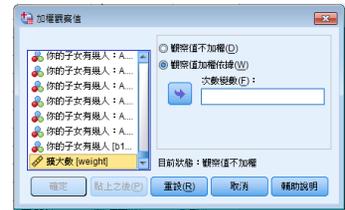


圖 3-1 加入加權變項

套用加權變項後，SPSS 右下角會出現 Weight On，代表此時進行任何分析，都是套用加權變項的狀況下。如圖 4。若不需加權，於圖 3 中點選 Do not weight cases 確認，或圖 3-1 中點選「觀察值不加權」即可。

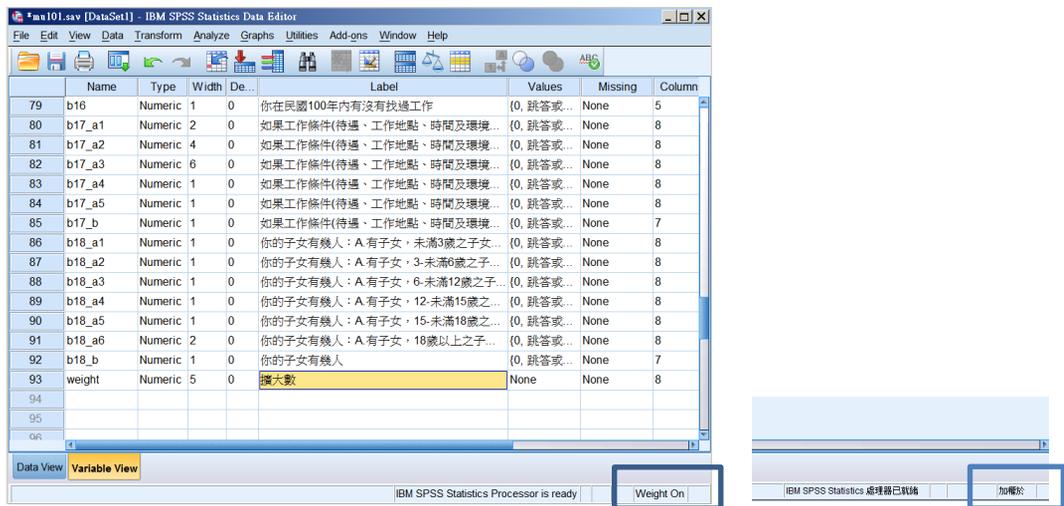


圖 4.SPSS 套用權值後的畫面（左：英文介面；右：中文介面）

進行與前述相同的分析報表（表 2），可以比較加權後灰格內的次數、總和、以及百分比與加權前的差異。加權後總和為母體總數，灰格百分比亦為母體的比例。

表2.加權後人力運用調查性別變項次數分配

		a2 性別			
		次數	百分比	有效百分比	累積百分比
有效的	1 男	9811031	49.7	49.7	49.7
	2 女	9910639	50.3	50.3	100.0
總和		19721670	100.0	100.0	

二、STATA 介紹：

STATA 與 SPSS 在使用權數時的兩個最大差異是：

1. 每次分析均需宣告，在分析語法後面以中括號[]來宣告權數類型與加權變項名稱。
2. 宣告擴大數與權數，指令不同。常用的指令如下。

擴大數：指令為 `fweight`。

權數：指令為 `aweight`。

(一) 以「行政院主計總處 101 年人力運用調查」做擴大數的語法說明範例。

- 次數分配語法（未加權），分析結果如表 3。

`tab a2`

表3. 未加權前人力運用調查性別變項次數分配

性別	Freq.	Percent	Cum.
男	28,933	49.86	49.86
女	29,100	50.14	100.00
Total	58,033	100.00	

- 次數分配語法（採用擴大數加權），分析結果如表 4。

`tab a2 [fweight= weight]`

表4. 加權後人力運用調查性別變項次數分配

性別	Freq.	Percent	Cum.
男	9,811,031	49.75	49.75
女	9,910,639	50.25	100.00
Total	19,721,670	100.00	

(二) 以「臺灣社會變遷基本調查計畫第六期第二次問卷一」作為**權值**的語法說明範例。

- 次數分配語法（未加權），分析結果如表 5。

`tab a1`

表5.未加權前臺灣社會變遷基本調查計畫第六期第二次性別變項次數分配

a1 性別	Freq.	Percent	Cum.
男	1,109	51.94	51.94
女	1,026	48.06	100.00
Total	2,135	100.00	

- 次數分配語法（採用權數加權）。

`tab a1 [aweight=wfinal]`

表6.加權後臺灣社會變遷基本調查計畫第六期第二次性別變項次數分配

a1 性別	Freq.	Percent	Cum.
男	1,063.2366	49.80	49.80
女	1,071.7634	50.20	100.00
Total	2,135	100.00	

三、結語

進行資料分析是否需使用加權，需視研究目的、調查主題以及分析所使用的變項而定，如果已知樣本與母體結構有明顯差異的狀況下，能使用加權進行分析，則其研究結果應該會更接近母體的表現；而在統計軟體進行加權分析作業，日益方便的狀況下，研究者可以利用此一便利性，做出更精緻的研究成果。