

## 民國九十一年國民健康促進知識態度與行為調查 縣市資料分析及加權補充說明

◎為能代表原抽樣母群體，有關平均值或比例數之計算，應加權處理說明：

本項調查以縣市為獨立選樣區，實地訪查過程不採用替代樣本，完訪率雖然高達八成以上，但完訪樣本資料經事後與母群體性別、年齡組成比較，發現不僅各縣市完訪率略有不同，甚至同一縣市之不同性別、年齡分組的完訪率亦不盡相同，因此不能假設未回應(non-response)的現象是完全隨機的(統計稱 missing completely at random, 簡稱 MCAR)，在此情況之下，若不按完訪率加權調整，則不能代表原抽樣群體。

基於未回應的現象具選擇性，為能代表原抽樣群體，權數不應依原始抽出率計算，而需以群體成員實際被抽出且完成訪問的機率來計算。換言之，令  $n$  代表抽出之樣本數， $m$  代表完訪樣本數， $N$  為母群體大小，則權數應依  $m/N$  而非依  $n/N$  計算。

本局人口與健康調查研究中心在製作權數檔時，就是依照統計專家上述建議，以完訪樣本數除以母群體總數為抽出機率。在計算抽出機率時，北、高兩市是以區為單位按性別、年齡分層，年齡分層方式為 15-24 歲、25-34 歲、35-44 歲、45-54 歲與 55 歲以上，其他縣市則是以縣市為單位按性別、年齡分層，年齡分層方式為 15-24 歲、25-34 歲、35-44 歲、45-54 歲、55-64 歲與 65 歲以上。採用組距較寬之 10 歲年齡組而未採用 5 歲年齡組，以及北、高兩市將 55-64 歲與 65 歲以上併組的理由，在於避免計算完訪率的個案數太少，因為若年齡分組過細，可能導致某些組的應訪個數太少，完訪率之計算就可能因多訪到一、兩案，或少訪到一、兩案而產生明顯變化，特別是北、高兩市各區應訪案數本來就不多，以應訪個案 20 人的細格為例，完訪 16 人的完訪率為 80%，若多訪 1 案，完訪率會增加為 85%，少訪 1 案就會減低為 75%，以加權值變化幅度來看，光多 1 案或少 1 案的變化幅度為 6.25%。

表一以台中市之完訪資料為例，說明權數設定過程，其餘各縣市或

北、高兩市的各區可依此類推。

表一 樣本權數設定過程－以台中市資料為例

分層		性別	母體						
			合計	15-24	25-34	35-44	45-54	55-64	65+
台灣省	台中市	男	365083	79438	78391	85654	60669	26678	34253
		女	387775	76707	90406	96103	63334	30511	30714
分層		性別	完訪案數						
			合計	15-24	25-34	35-44	45-54	55-64	65+
台灣省	台中市	男	557	116	102	131	91	49	68
		女	566	101	121	129	93	71	51
分層		性別	抽出機率：(完訪案數/母體數)						
			合計	15-24	25-34	35-44	45-54	55-64	65+
台灣省	台中市	男	0.00153	0.00146	0.00130	0.00153	0.00150	0.00184	0.00199
		女	0.00146	0.00132	0.00134	0.00134	0.00147	0.00233	0.00166
分層		性別	權數比值：(母體數/完訪案數)或(1/抽出機率)						
			合計	15-24	25-34	35-44	45-54	55-64	65+
台灣省	台中市	男	655.4452	684.8103	768.5392	653.8473	666.6923	544.4490	503.7206
		女	685.1148	759.4752	747.1570	744.9845	681.0108	429.7324	602.2353
分層		性別	台灣地區權數：(母體數/完訪案數) <sup>2</sup> (26755/17690404)(註)						
			合計	15-24	25-34	35-44	45-54	55-64	65+
台灣省	台中市	男	0.9913	1.0357	1.1623	0.9889	1.0083	0.8234	0.7618
		女	1.0362	1.1486	1.1300	1.1267	1.0300	0.6499	0.9108
分層		性別	按台灣地區權數加權後之加權樣本數						
			合計	15-24	25-34	35-44	45-54	55-64	65+
台灣省	台中市	男	552	120	119	130	92	40	52
		女	586	116	137	145	96	46	46

註：乘上此一比值係為使台灣地區加權樣本總數大致等於實際完訪樣本總數，

其中 17,690,404 為 15 歲以上母群體總數，26755 為完訪樣本數。

有關權數設定方法見人見智，本局係針對最基本且母群體比例容易取得的性別、年齡做考量，但目前設定權數的年齡分層方式不一定就是最理想的方式，使用者可由完訪樣本數自行依需要重設或再加上其他分層條件（如：城鄉別），只要是合理應該都可以接受，特別是資料使用者若需將本項分析結果與其他調查資料做比較時，則可能需考慮兩者加權方式異同是否影響資料間可比較性。

又本項權數與性別、年齡標準化之意涵不同，加權後之資料如需作跨縣市或與全台灣地區作比較，仍應考慮性別、年齡標準化之問題。

### ◎描述性統計圖表的呈現，建議使用完訪樣本數

說明：

必須附帶提醒資料使用者的是，只有在做群體參數推論時需考慮加權問題，而為降低使用上之複雜度，本局只設定一套同時適用於跨縣市合併分析(含全台灣地區合併分析)以及各縣市單獨分析之權數，某些縣市因人口數較少，故抽出機率較其他縣市高，權數則相對較低，故其加權調整後之加權樣本數，會低於實際完訪樣數，例如澎湖縣實際完訪樣本數為 677 人，但加權完訪樣本數則為 115 人，為免讀者誤以為樣本數只有 115，在描述性統計圖或表的呈現上，建議使用完訪樣本數，若完訪樣本有某些未回應之問項(item non-response)，則應呈現該問項之回應人數（該問項之有效樣本數），而且必須在統計表及文字說明適當處加以註記，例如加註「百分比之計算經加權處理」，否則讀者將質疑百分比為何與各分項人數佔總完訪案數百分比不一致。

### ◎多變項模式之加權問題

說明：

多變項模式要否需加權以及如何加權，按本局請教統計專家的結果，若要對某個連續型反應做複迴歸分析、變異數分析或對某個二元(binary)反應做邏輯迴歸(logistic regression)分析，因為在做這類分析時，與描述性統計所強調的估計問題並不相同，研究者所關心的是解釋變數與依變數間的關係，而非反應變數的彙總結果是否能代表群體的特性。影響完訪率的因素(如：性別、年齡別)也可能影響所關心的反應，但是這些因素可以放到模型中，換言之，在普通複迴歸或邏輯迴歸中把它們當做解釋變數，在 ANOVA 中也把它們當做「處理或區集因子」。遇需要加權的情形，尚需看所用軟體是否有加權選項，而且在加權時需注意軟體是否能適當處理加權結果。一般說來，這並不是一個簡單的問題。例如：假設在 ANOVA 中不把(性別 × 年齡別)當做「處理或區

集因子」，則需加權，但標準理論並無加權的分析公式（事實上不是沒有，只是其加權的原因與這裡的問題不同，能否套用仍是個疑問）。

#### ◎除北、高兩市，其他縣市樣本不具鄉鎮市區代表性

##### 說明：

本項調查資料除北、高兩市以外，其他縣市並不具鄉鎮市區代表性，以台北縣為例，樣本區雖然有板橋市、中和市、三重市、新莊市、新店市、永和市、汐止市、蘆洲市、淡水鎮、五股鄉、貢寮鄉等十一個行政區，但只能將十一個樣本區的資料合併在一起代表台北縣的狀況，不能個別代表板橋市或中和市等，故資料之分析與呈現，不宜按鄉鎮市區別分層，其他與地理或行政區有關的分層方式，若分層結果會使某一分層只有單一個行政區，也會面臨到相同的問題，例如按醫療資源分層後，只有單一個鄉鎮市區落在其中某一層，在解釋或推論這一層的分析結果時，一樣會面臨該鄉鎮是否能反應該層狀況。

除北、高兩市以外之二十一縣市資料不具鄉鎮市區代表性的原因如下，本研究旨在收集建置全台灣地區及縣市代表性資料，故研究設計及抽樣設計均是以縣市為單位作考量，各鄉鎮市區之樣本數未經過估算，分析時極可能會有樣本數太少或檢力不足的問題，特別是在描述或探討健康相關議題時，通常會進一步再按樣本社會人口學背景特徵做分層比較，則樣本數又將更少，各項統計值的估算會相當不穩定，若由區間估計則會看到信賴區間很寬，比較不出不同社會人口學背景特徵之差異。另一方面則是權數的問題，本局所設定的權數除了北、高兩市是按行政區之外，其他縣市的權數都是以縣市為單位作設定，其目的在調整縣市完訪樣本性別、年齡與該縣市母群體之差異，而以縣市為單位所設定的權數，若應用在單一鄉鎮市區資料之加權上，並不能保證可將該鄉鎮市區的性別、年齡比例調整為與該鄉鎮市區母群體相同。