

民國 95 年國民生活狀況調查抽樣設計

壹、抽樣設計：

一、採分層隨機抽樣方法，將臺閩地區分為北部地區、中部地區、南部地區、東部地區和金馬地區。

(一) 北部地區：臺北市、基隆市、臺北縣、宜蘭縣、桃園縣、新竹縣、新竹市

(二) 中部地區：苗栗縣、臺中縣、臺中市、彰化縣、南投縣、雲林縣

(三) 南部地區：嘉義縣、嘉義市、臺南縣、臺南市、高雄縣、高雄市、屏東縣、澎湖縣

(四) 東部地區：臺東縣、花蓮縣

(五) 金馬地區：福建省金門縣、連江縣。

二、樣本配置：

預計有效樣本數 4,000 戶係依照各縣市 20 歲以上人口數占臺閩地區 20 歲以上總人口數的比例分配到各縣市。但為產生臺灣各縣市及金馬地區的獨立估計，比照往年，設各縣市最低樣本數為 138 人。金門和連江縣由於人口數很少，合併為金馬地區計。如此可保證各縣市估計百分比的抽樣誤差在 90%信心水準下，不超過 7%。而整體的估計百分比的最大抽樣誤差，在 95%信心水準下大約為 1.55%。

先設定較小縣市的樣本數為 138 人，再將剩餘的樣本依照各層(縣市)內 20 歲以上人口數占剩餘臺閩地區 20 歲以上人口數的比例分配樣本數。各縣市樣本分配約為如下：

各縣市樣本分配

縣 市	人口比例 (%)	依縣市人口比 例配置人數	預計有效 樣本人數
總 計	100.00	4,000	4,000
臺北縣	16.45	658	448
宜蘭縣	2.02	81	138
桃園縣	7.91	316	215
新竹縣	2.02	81	138
苗栗縣	2.44	98	138
臺中縣	6.51	261	178
彰化縣	5.70	228	156
南投縣	2.37	95	138
雲林縣	3.28	131	138
嘉義縣	2.52	101	138
臺南縣	4.97	199	138
高雄縣	5.55	222	151
屏東縣	4.00	160	138
臺東縣	1.06	43	138
花蓮縣	1.54	62	138
澎湖縣	0.42	17	138
基隆市	1.74	70	138
新竹市	1.66	66	138
臺中市	4.38	175	138
嘉義市	1.17	47	138
臺南市	3.33	133	138
臺北市	11.85	474	323
高雄市	6.73	269	183
金馬地區	0.38	15	138

貳、資料處理及統計分析方法

一、資料處理方式

以電腦處理為主，人工處理為輔。人工處理部分包括調查資料之審核、複查、檢誤、更正、結果表之核對、研判與分析等。電腦處理部分包括調查資料登錄、軟體程式設計、資料整理及結果表列印。

資料經審查和複查後，先以抽樣機率的倒數做基本權數進行加權調整。然後根據臺閩地區人口的性別、年齡與縣市分配檢定樣本的結構。若樣本的性別及年齡分配與母體分配有顯著差異，則以母體的性別與年齡分配調整樣本分配，以矯正偏差並減少抽樣誤差。

本調查採用 raking 方法，先以縣市分配，再以性別最後以年齡分配調整，如此反覆直到樣本分配與母體分配已無顯著差異，才停止 raking。

每一筆資料都乘以調整權數， $\frac{N_i}{N} / \frac{n'_i}{n'}$ ， N_i 和 n'_i 是第 i 交叉組的母體人數和樣本加權人數，而 N 和 n' 是母體總人數和樣本加權總人數，這樣使樣本與母體的分配在那一步驟調整後完全一致。最後的權數是各步驟調整權數累乘。

二、統計分析方法

本調查樣本採分層隨機抽樣，母體參數以如下估計：

$$\text{設 } Y_{hi} = \begin{cases} 1, & \text{第 } h \text{ 層(縣市)第 } i \text{ 樣本戶具有某特徵} \\ 0, & \text{第 } h \text{ 層(縣市)第 } i \text{ 樣本戶不具有某特徵} \end{cases}$$

$$I_{hi} = \begin{cases} 1, & \text{第 } h \text{ 層第 } i \text{ 樣本屬於某次群體} \\ 0, & \text{第 } h \text{ 層第 } i \text{ 樣本不屬於某次群體} \end{cases}$$

$$W_{hi} = \text{調整權數 (含基本權數)}$$

(一) 百分比之估計

某次群體具有某項特徵(如覺得非常滿意)的百分比可估計如下：

$$\hat{P}_1 = \frac{\sum_h \sum_i Y_{hi} W_{hi} I_{hi}}{\sum_h \sum_i W_{hi} I_{hi}}$$

(二) 百分比變異數之估計

本調查的樣本係採分層隨機抽樣，居住在不同縣市的民眾抽樣機率不一樣，不能以簡單隨機公式來約略估計變異數。應採用分層隨機抽樣之變異數公式：

$$\hat{\sigma}_{\hat{p}}^2 = \frac{1}{N^2} \sum_{h=1}^{23} N_h^2 \frac{\hat{P}_h (1 - \hat{P}_h)}{n_h}$$

N 是母體總人數， N_h 是第 h 縣市的母體人數， n_h 是第 h 縣市樣本人數， \hat{P}_h 是第 h 縣市具有某特徵的民眾的估計百分比。用此公式計算出臺閩地區以及各縣市的各項估計百分比的標準誤($\hat{\sigma}_{\hat{p}}$)，並與簡單隨機公式所算出之標準誤比較，即可算出標準誤是簡單隨機抽樣標準誤的 D 倍， D 是本次調查的抽樣設計效應係數。

(三) 百分比的分析

1. 次數分配

根據各題加權的樣本比例進行比較選項間的差異，用下列 Z -檢定，檢驗兩選項間百分比(P_1 和 P_2)的差異：

$$z_1 = \frac{\hat{P}_1 - \hat{P}_2}{D \sqrt{\frac{1}{n} [\hat{P}_1 + \hat{P}_2 - (\hat{P}_1 - \hat{P}_2)^2]}}$$

D : 本次抽樣設計效應係數

2. 交叉分析

以各題與基本資料的交叉表來分析民眾對各議題的看法與他們基本特徵間的相關。交叉表第一步採用卡方檢

定。交叉表的卡方顯著水準小於5%時才認定兩變數間相關。第二步是在有相關的交叉表內，以 z 檢定找出有顯著差異的地方。檢視兩個獨立的次群體(Subgroups)對同一議題看法的百分比間的差異，採用下列的 z 檢定。

$$z_2 = \frac{\hat{P}_1 - \hat{P}_2}{D \sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}}$$

此外，當需要比較數個次群體時，則採用 CHAID(Chi-squared Automatic Interaction Detector) 做分組比較分析。

(四) 重要度分析

當選項可以複選三次，但必須排序第一主要、第二主要和第三主要時，我們將分別以1，2/3，1/3的重要度分數來表示被選為第一主要，第二主要，第三主要的重要度。因此各選項的重要度亦可以下列推估：

重要度=被選為一主要的百分比×1+被選為第二主要的百分比×2/3+被選為第三主要的百分比×1/3

(五) 抽樣誤差

本調查主要在推估母體比例 P (民眾對某項生活滿意之比例)。故以樣本比例 \hat{P} 推估母體比例 P 在 $(1-\alpha)\%$ 信賴係數下之信賴區間(Confidence Interval)為：

$$(\hat{P}-d, \hat{P}+d)$$

其中， $d = Z_{(1-\alpha/2)} D \sqrt{\hat{P}(1-\hat{P})/n}$ 為可能之抽樣誤差。

有效樣本4,000份，於95%之信賴係數下，本調查之抽樣誤差應在1.55%以內。

每一個統計資料分析均經過正確的統計檢定。當我們必須並列無顯著差異的估計時，我們必定聲明兩者之間的差異，在統計上，並無顯著差異。

