

抽樣設計

I、設計概要

- 1.抽樣方法：分層三段隨機抽樣法。
- 2.調查及抽樣群體：本調查以調查時臺灣地區年滿二十歲以上未滿六十歲之婦女(育齡婦女包括未婚已婚)為調查群體；以戶籍在臺灣地區，澎湖縣及山地鄉以外，年滿二十歲以上未滿六十歲之婦女為抽樣群體。
- 3.樣本大小：預計抽出 4500 個樣本個案。
- 4.分層標準：以 331 個行政單位(鄉鎮市區)為初抽單位(PSU)，並以之為分層的單位。分層標準有三，即都市化程度、整體教育水準、及生育水準。其中都市化程度為方便故，依行政區域劃分為準。
- 5.各段抽出單位：初抽單位為鄉鎮市區(PSU)，次抽單位(SSU)為鄰，終抽單位(TSU)為樣本個案。
- 6.樣本配置及各段抽出率：各層間之樣本依群體大小(終抽單位總數)比例配置，PSU 及 SSU 的抽樣均依「抽出機率和群體大小成比例(probability proportional to size, 簡稱 pps)」的原則抽出，而同一層內每一樣本 PSU(鄉鎮市區)及每一樣本 SSU(鄰)所含的樣本 TSU(樣本個案)數相同，其中每一樣本 SSU 所含的樣本個數為 4，以此換算各層及各樣本 PSU 中所需抽出的樣本鄰(SSU)數。
- 7.其他：本抽樣法所得每一調查對象最終被抽出機率相等。

II、抽出程序

- 1.分層：都市化程度依行政區域分為三類、教育水準及生育水準亦各分三類，三個分層標準交叉分類將 331 個鄉鎮市區分為 27 層。
- 2.第一段抽樣：由於此次調查與「民國七十五年臺灣地區家庭與生育率調查」之調查對象，其人數在各個鄉鎮市區的相對分布情形接近，而兩個調查均採用分層三段抽樣，且分層標準、初抽單位設計相似，乃採用該次調查的初抽單位為基礎，並依 Kish 與 Scott(1971)的方法調整以使各 PSU 被抽出的機率

與各該 PSU 所含的年滿二十歲以上未滿六十歲之婦女人數成比例。

其調整的基本方法是將層內各 PSU，依其抽出機率的變化，分為提高和降低兩類，設 $d=1, \dots, D$ 表示抽出機率降低的 PSU 代號， $i=D+1, \dots, D+I$ 表示抽出機率提高的 PSU 代號， p_i 及 p_d 表舊的抽出機率， P_i 及 P_d 表新的抽出機率。

(a) 若樣本 PSU 屬抽出機率提高者，則保留該樣本 PSU。

(b) 若樣本 PSU 屬抽出機率降低者，則以 P_d/p_d 的機率保留該樣本 PSU。

(c) 若在 (b) 中有某樣本 PSU 被剔除，則自抽出機率提高的 PSU 中，以和 P_i/p_i 成比例的機率抽出等數的 PSU 補充之。

不難證明上述程序可得到正確的抽出機率，即各 PSU 的抽出機率和其大小成比例(或等於預定的抽出機率，如本意不是要 pps)。實際執行時則採 Kish 與 Scott(1971)第三節的建議方法。

註: Leslie Kish and Alastair Scott (1971). "Retaining Units after Changing Strata and Probabilities.", Journal of the American Statistical Association 66, p461-470

註: 實務上因(a)依 Kish 與 Scott (1971)建議僅考慮被抽出機率變動大於一定數(如 10%)者，(b)並非各層獨立處理，(c)舊的樣本 PSU 採"隨機系統抽出法"抽出，因此新的抽出機率僅近似 pps，唯其差異不大。

3. 第二段抽樣：首先設計各樣本 PSU(鄉鎮市區)中各 SSU(鄰)所含的個案數，同時依各層大小比例配置樣本數，除以本層樣本 PSU 數，得各樣本 PSU 所需調查的樣本個案數。以每個樣本 SSU(鄰)調查 4 個個案為準，計算各樣本 PSU 所需調查的樣本鄰數，然後依 pps 抽樣的原則，隨機抽取所需數目的樣本鄰。

4. 第三段抽樣：自各樣本鄉鎮市區抄取各樣本鄰的調查對象(年滿二十歲但未滿六十歲的育齡女名冊)，然後獨立自各樣本鄰抽取 4 個個案組成調查樣本。若抽出之樣本個案已經不在原戶籍登記地，只要該個案仍居住在臺灣地區(澎湖縣和山地鄉除外)，則需至其現址調查，此類被調查者其現址一般可從其原戶籍地之戶政單位或其鄰居朋友親戚處查得。

III、群體參數估計及其誤差公式

1. 符號說明

$h =$ 層別 $= 1, \Lambda, L; L = 27$

$i =$ PSU(鄉鎮市區) $= 1, \Lambda, t_h(T_h)$

$t_h =$ 第 h 層樣本 PSU 數, $T_h =$ 第 h 層群體 PSU 數

$j = SSU(\text{鄰別}) = 1, \Lambda, m_{hi}(M_{hi})$

m_{hi} = 第 h 層第 i 個 PSU 的樣本 SSU 數, M_{hi} = 第 h 層第 i 個 PSU 的群體 SSU 數

$k = TSU(\text{個案別}) = 1, \Lambda, n_{hij}(N_{hij})$

n_{hij} = 第 h 層第 i 個 PSU 第 j 個 SSU 的樣本 TSU 數

N_{hij} = 第 h 層第 i 個 PSU 第 j 個 SSU 的群體 TSU 數

$$\hat{Y}_{hi} = \frac{1}{m_{hi}} \sum_{j=1}^{m_{hi}} \frac{1}{n_{hij}} \sum_{k=1}^{n_{hij}} y_{hijk}$$

$$\hat{Y}_h = \frac{1}{n_h} \sum_{i=1}^{t_h} \sum_{j=1}^{m_{hi}} \sum_{k=1}^{n_{hij}} y_{hijk}$$

2. 母群體平均數之估計和誤差公式

(A) 母群體平均數之估計:

$$\hat{Y} = \frac{1}{n} \sum_{h=1}^L \sum_{i=1}^{t_h} \sum_{j=1}^{m_{hi}} \sum_{k=1}^{n_{hij}} y_{hijk} = \bar{y},$$

$$\text{其中 } N = \sum_{h=1}^L N_h, n = \sum_{h=1}^L n_h \circ$$

*若要考慮各層完成率之差異,可改用下式估計:

$$\hat{Y} = \frac{1}{N} \sum_{h=1}^L N_h \hat{Y}_h = \frac{1}{N} \sum_{h=1}^L \frac{N_h}{n_h} \sum_{i=1}^{t_h} \sum_{j=1}^{m_{hi}} \sum_{k=1}^{n_{hij}} y_{hijk} \circ$$

(B) 母群體比率估計: 同母群體平均數之估計, 唯其觀測值 y_{hijk} 僅 0 和 1 兩種可能值。

(C) 平均數(及比率)變異數之估計:

$$\text{var} \left[\hat{Y} \right] = \sum_{h=1}^L \left(\frac{N_h}{N} \right)^2 \frac{1}{t_h(t_h-1)} \sum_{i=1}^{t_h} \left(\hat{Y}_{hi} - \hat{Y}_h \right)^2,$$

其中 \hat{Y}_{hi} 及 \hat{Y}_h 定義如前。