

# 中央研究院調查研究工作室 臺灣地區性樣本初抽單位 (PSU)抽取原則 Q&A\*

張苙雲\*\* 許秀微\*\*\*

中央研究院內各社會科學相關研究所暨部分生命科學研究所辦理問卷調查研究已經有多年的歷史，近年間此類研究計畫更不斷增加。然而，每項調查研究均各自擬訂一套抽樣方案，並自行進行抽樣工作，造成研究員與研究助理人力的重疊與浪費。自本院成立調查研究工作室以後，多項問卷調查研究委由工作室統籌抽樣工作。工作室既負責統籌辦理抽樣工作，自需檢討如何整合、簡化各項調查研究工作，故而有抽樣顧問小組<sup>1</sup>的成立。此外，為與徵募訪員的工作相互配合、減

---

\* 本文是根據中央研究院調查研究工作室「抽樣顧問小組」會議之會議記錄改寫而成。

\*\* 作者現任中央研究院社會學研究所籌備處研究員。

\*\*\* 作者為中央研究院調查研究工作室助理。

1 抽樣顧問小組成立於民國八十二年八月四日，其成立之目的不僅協助調查研究工作室擬訂各個調查計畫之抽樣方案，更致力於抽樣方法的研究，以期降低抽樣誤差，提昇調查結果的可靠度。小組成員如下所列：

化抽樣工作，並使各項長期性研究有較高的可比較性，決定在一段時期內採用一組較為固定之樣本鄉鎮市區作為初抽單位 (Primary Sampling Unit, PSU)，每隔一段時期後再檢討這些 PSU 的妥適性，並重行抽取新一輪的 PSU。然，在抽樣顧問小組討論過程中，議及相當多的問題。茲於下節簡介抽取 PSU 的原則，第二部份則以答問方式將討論內容和決議事項陳明。

## 一、抽樣的程序

### 1. 各分層、各階段抽取數的決定：

抽樣方法採分層多段等機率抽樣原則，除台北市、高雄市為兩段式抽樣外，其餘各分層均為三段式抽樣，第一段抽樣單位為鄉鎮市區，第二段為村里，第三段抽人，台北、高雄則直接跳過區抽村里。各分層、各階段的抽取數決定如表一（樣本數以 1,600 案估計）。

### 2. 各層內第一階段鄉鎮市區的抽樣過程如下：（兩直轄市除外）

- A. 根據「79 年戶口普查」之資料計算各鄉鎮市區教育程度高中以上之百分比。
- B. 將同分層之鄉鎮市區按教育程度由小到大排列，並做人口數的累加。
- C. 以該層的總人口數  $T$  除以所需的鄉鎮市區數  $n$ ， $T/n = K$  ( $K$  取整數)。
- D. 在  $1 \sim K$  中取一個亂數  $R$ ，則  $R$  在累加人口數該欄的落點所在即是抽取的第一個鄉鎮市區。

---

召集人：張苙雲教授

委員：朱瑞玲教授  
張明正教授

李隆安教授  
楊文山教授

洪永泰教授  
齊力教授

黃景祥教授  
劉應興教授

表一

	層級	總人口數 T	比例 P	鄉鎮 區數	村里數	每里 案數	樣本數 $T \times P$
固定	一 新興鄉鎮	1,705,545	0.143	6	2	20	240
	三 工商市鎮	1,668,404	0.139	6	2	20	240
	四 綜合性市鎮	1,187,398	0.099	4	2	20	160
	七 服務性鄉鎮	1,266,735	0.106	4	2	20	160
	小計			20	40		800
非 固 定	二 山地鄉鎮	319,190	0.027				
	五 坡地鄉鎮	701,366	0.059				
	二、五合併	1,020,556	0.086	3	2	20	120
	六 偏遠鄉鎮	1,236,676	0.103	4	2	20	160
	八 省轄市	1,392,166	0.116	6	3	10	180
	九 台北市	1,659,531	0.139		25	9	225
	十 高雄市	827,074	0.069		22	5	110
	總計	11,977,945	1.000	33	119		1,595

※以上總人口數是來自「80 年台閩地區人口統計」，指的是介於 20～60 歲的人口數。

E.然後以 R 為基礎，再往下找第  $R+K$  的落點，以此類推，直到抽滿所需的鄉鎮市區數為止。

F.據上述所抽取之一組 PSU 附於後，其中並附有所抽取的鄉鎮市區之客家及山地的標示、高中以上百分比、總人口數及各年齡層人口百分比以供參考。

## 二、抽樣的確立

### 1. 抽樣範圍的討論：

Q1.是否將山地鄉及離島地區排除在抽樣範圍之外？

山地鄉及離島地區與一般城鄉性質差異甚大，訪問不易，且其數量不大，可以視調查需要定期做調查，故最後仍決定不將山地鄉及離島從抽樣地區中排除。

### 2. 分層的決定:

此分層抽樣方案前七層採羅啓宏先生所著「鄉鎮區發展類型之研究」的分層原則，加上省轄市、台北市及高雄市共分為以下十層：

- |          |          |
|----------|----------|
| (1)新興鄉鎮  | (6)偏遠鄉鎮  |
| (2)山地鄉鎮  | (7)服務性鄉鎮 |
| (3)工商市鎮  | (8)省轄市   |
| (4)綜合性市鎮 | (9)台北市   |
| (5)坡地鄉鎮  | (10)高雄市  |

Q1.為何決定使用羅啓宏先生所著「鄉鎮區發展類型之研究」中的分層為原則？

本小組曾考慮過整理主計處 79 年戶口普查資料以為分層之參考，但其資料並不齊全且變項不多，僅有教育程度、行業、職業勞動力與婚姻狀況等變項，實不足以架構一套分層，加之

又有時間的壓力，短時間內實不易收集足夠的資料，因此轉而考慮現有一些相關主題的研究報告。其中李朝賢所著「臺灣地方經濟發展差異之研究」、與羅啓宏所著「鄉鎮區發展類型之研究」中的分層原則，「社會變遷基本調查」與「社會意向調查」均用過，每次抽出之客家鄉鎮維持在 4~5 個之間，且由族群分佈的觀點看：閩南人約佔 75%±3%；外省人約佔 12%±2%；客家人約佔 11%±3%頗為穩定，然而由於李朝賢所著「臺灣地方經濟發展差異之研究」資料均來自民國 75 年或更早，基於時效的考量，故而決定採用羅啓宏（1992）所著「鄉鎮區發展類型之研究」中的分層為原則。

Q2.羅啓宏先生所著「鄉鎮區發展類型之研究」中所涵蓋的區域並未包含直轄市與省轄市，對於未涵蓋之部份如何處理？

台北市與高雄市雖均為直轄市，然其一南一北本身之地理環境已有差異，再加之北商南工，城市性質亦有所不同，故將之各置一層。至於省轄市則介於直轄市與其他鄉鎮之間，也獨立一層。

### 3. 抽樣單位的決定：

除台北市、高雄市外，其餘各層第一段抽樣單位均為鄉鎮市區，台北市、高雄市第一段抽樣單位則為里。

Q1.為何台北市與高雄市不以區而以里為第一抽樣單位？而省轄市為何不也以里為第一抽樣單位？

對於台北市與高雄市兩直轄市，考慮其里內人數可能都多過各鄉鎮市區總人數，且區數不多，沒有必要先抽區再抽里，故決定以里為第一抽樣單位。而省轄市雖較其他鄉鎮市區進步，但差異不致太大，故仍以區為第一抽單位。

## Q2. 為何以村里為第二段抽樣單位而不以鄰為第二段抽樣單位？

以往院內均以村里為第二段抽樣單位，而村里內案數則以一位訪員在調查期間內所能負荷的數量為準，目的是為了減少訪員負責的區域範圍，可是如此每個鄉鎮市區中就往往只能抽二個村里，而以兩個村里代表一個城鄉，其代表性似乎略嫌不足。若是以鄰為第二段抽樣單位，如此可在一個城鄉抽取較多的鄰，而每鄰的案數則以一個訪員一天所能負荷的案數為準，例如一個城鄉欲抽取 40 案，則可抽 6 個鄰每鄰 6~7 案，或抽 8 個鄰每鄰 5 案，視問卷長短而定。如此在一個城鄉抽取較多的鄰，其代表性將提高，且每個鄰的同質性較村里高許多，只要抽取較少的案數即足以代表該鄰。可是如此一來每位訪員所負責的區域將相對變大，而且若抽過多的鄰，每鄰案數相對減少，可能使第二段抽樣失去意義，而訪員也將因此疲於奔命，更重要的是，詢問戶政事務所後，發覺鄰的基本資料取得甚為困難，因此似乎抽村里較為可行。

## 4. 鄉鎮區數、村里數及每村里案數的決定

原則上每村里案數是依一個訪員在調查期間內所能負荷的數量為準，約 20 案左右，而每個鄉鎮市區內至少須抽 2 個或 2 個以上的村里，據此原則在鄉鎮區數與村里數之間取得一平衡點，即得上表。至於台北市、高雄市因交通較為方便，在同區內，里與里間的距離相對較小，因此不在此限。

## 三、其他討論

### Q1. 此套 PSU 適用之調查性質為何？

此套 PSU 將適用於台灣地區的一般性調查。至於特殊主

題的調查研究則需視調查對象、目的等情況而定。

## Q2.目前各分層是否固定 PSU？

如下列所示：

(1)新興鄉鎮	固定	(6)偏遠鄉鎮	不固定
(2)山地鄉鎮	不固定	(7)服務性鄉鎮	固定
(3)工商市鎮	固定	(8)省轄市	不固定
(4)綜合性市鎮	固定	(9)台北市	不固定
(5)坡地鄉鎮	不固定	(10)高雄市	不固定

為何各分層有固定或不固定 PSU 的區分呢？基於實務的考量，若能抽取固定 PSU，我們便能在那些被抽出的固定鄉鎮區尋找當地的訪員，訪員對當地的地理環境較為熟悉，將有助於調查的進行。但是並非所有分層均適合抽取固定 PSU，而不將之固定的原因分述如下：

A.由於(2)山地鄉鎮、(5)坡地鄉鎮、及(6)偏遠鄉鎮實地調查進行困難，且其人口數佔總人口數的比例甚少，若不做適度的合併，所抽之樣本數往往不足以進行統計分析，因此不在該處設置固定的初抽單位，將由各計劃視其所需決定此部份之抽樣。

B.至於(8)省轄市、(9)台北市、(10)高雄市三層因都市化的結果以至受訪者防衛性較強，拒訪率相對較高，調查本就不易進行，若固定 PSU 一段時間，恐當地居民不勝其擾，使調查更難進行，因此亦不在此三層抽固定 PSU。

## Q3.為何目前抽樣名冊仍以戶籍資料為主？

本小組曾對戶籍資料、選舉名冊及電信局的地址資料各別進行探討，結果如下：

- A.戶籍資料：在中央研究院內較常被使用，但其資料的準確性及涵蓋性均有待質疑，且居住地與戶籍地常常有所出入，因此若使用戶籍資料而不進行追蹤，可能將會有相當比例的人根本沒機會被訪問到，而各計劃往往經費拮据更遑論追蹤了，退而以備取樣本替代應追蹤之正取樣本似乎也是不可避免的事。至於「戶政電腦化」也造成不少困擾，有些戶籍資料已電腦化的戶政事務所，能提供的只是久未更新的戶籍名冊，可更方便的電腦資料檔反而束諸高閣，以致所得之資料與實況出入更大，這是我們現在急於解決的大問題。然就整體而言，戶籍資料仍是目前較為可信的抽樣名冊。
- B.選舉名冊：由於此名冊抄錄自戶籍資料，因此戶籍資料的問題在此亦無法避免，且多了一個抄錄過程，發生錯誤的機會也較大。此外，選舉並非時常有，且不一定是全台灣地區的選舉，基於時效及涵蓋區域的考量，要長期使用選舉名冊似乎就不大恰當了。
- C.電信局的地址資料：依據以往經驗，約有 25%的人不是居住在戶籍所在地。因此，若以戶籍資料為抽樣依據，而對於不住在戶籍地者又不進行追蹤，則將有約 25%的樣本個案無法被訪問到。因此，李隆安教授提議試行以“地址”為準而不以戶籍為準的抽樣方式。如此不僅可克服戶籍資料之不完整，亦可藉機核對戶籍資料的準確性，累積一段時間後，尚可藉以推估由鄉村移至都市的流動戶口。由於現在電話的普及率已經高達 98%，若要以居住地為準取樣，可詢問電信局是否能提供現有電話及地址的檔案，用以協助抽樣的進行。唯目前尚未得到進一步消息。



Q4. PSU 固定了是否就一直沿用下去呢？

原則上此套固定 PSU 將沿用五年，可是爲使其適用於情勢多變的台灣地區，故本小組決定每年以 Kish 公式檢定一次，一但發現此套固定 PSU 已不再俱代表性則適時更換。