

中央研究院 人文社會科學研究中心 調查研究專題中心

2020 年電話調查抽樣分層報告

侯佩君

引用資料

侯佩君，2020。2020 年電話調查抽樣分層報告，中央研究院人文社會科學研究中心調查研究專題中心研究報告。

https://srod.survey.sinica.edu.tw/wp-content/uploads/2024/12/2020_phone_stratification.pdf

Hou, Pei-chun. 2020. 2020 Report on Sample Stratification in Phone Surveys. Center for Survey Research, Research Center for Humanities and Social Sciences, Academia Sinica.

https://srod.survey.sinica.edu.tw/wp-content/uploads/2024/12/2020_phone_stratification.pdf

一、前言

面對住宅電話用戶數的逐年下降及個人手機的高度普及率，採用電話調查來蒐集問卷資料勢必得因應時勢潮流而隨時調整研究方法與調查策略，而如何降低日益攀升的空號率即是首要面對的棘手問題。依據國家通信傳播委員會 (簡稱 NCC) 於 2020 年 5 月所公告的統計資料，¹ 市內電話與行動電話的空號率估計分別為 66.86% 及 56.69%。面對此等高比例的空號狀態，本專題中心採用等比例分層抽樣 (proportionate stratified sampling) 的設計來因應；亦即，除了在分類抽樣分層的集群時納入接觸高低的相關因子，將具有共通特質者劃入同一層級，以達「同層之內同質性高、異層之間差異性大」的特性之外，並運用等比例分配的機制來促成高接觸率者中選機率高的效應，裨益於有效提高接觸率、擷節調查成本。有關市話號碼與行動電話門號的抽樣分層作業說明，分述如下。

二、市話號碼局碼抽樣分層

(一)2019 年調研中心市話局碼抽樣分層

1. 分層變項：本研究於進行市話號碼集群分析時，嘗試放入多項可能影響接觸率高低的因子，包括市話千門用戶碼的使用率²、本專題中心於 2014 至 2018 年五年間各類型電話調查撥號結果³等。經多方測試後，最終篩選出相對影響力較高者，包括：①NCC 市話千門用戶碼的使用率；②過去五年間電話調查市話千門用戶碼空號和非住宅電話的比率；③前者於市話千門用戶碼的住戶接觸率。其中，將空號與非住宅電話合併為同一項是為了避免進行集群分析時彼此干擾，影響分層效果。

¹ 市話號碼用戶數約為 1,089 萬、行動電話用戶數約為 2,906 萬，而市話號碼數約為 3,286 萬、行動電話號碼數約為 6,710 萬。

² 市話號碼的組成區分為三個部分：區碼 + 局碼 + 用戶號碼。其中，區碼用以區別地區，由 2-3 碼組成，例如台北市為 02；局碼為局所交換機局碼，由 3-4 碼組成；用戶號碼則為後 4 碼。市話千門用戶碼的使用率係指用戶號碼以千門為計算單位的使用率，例如：02-2787-1000 ~ 02-2787-1999、02-2787-2000 ~ 02-2787-2999 等，以此類推。使用率的計算方式：使用數量/1000。

³ 千門用戶碼的空號率 (空號數/樣本數)、非住宅電話率 (非住宅電話數/樣本數)、住戶接觸率 (住戶接觸數/樣本數)。

2.分層方式

(1)首先，有鑑於非中華電信業者的市話號碼市佔率比較低，分層時先將非中華電信業者合併為單一分層。

市話電信業者共計四家，分別是中華電信、台固、亞太與新資通。依據各電信業者於 2018 年的統計資訊，住宅用戶佔各業者市話號碼的比率分別約為：73.69%、0.97%、0.04%、0.19%；有鑑於後三家業者市話住戶比率過低，故在分層時先將此三家的千門用戶碼獨立出來置於同一分層，不進行分層集群分析。

(2)中華電信市話千門用戶碼之分層集群分析：本研究進行集群分析時，所採用的測量方式為計算同一變項間相異性方式的華德最小變異數法 (Ward's Minimum Variance Method)，並以歐幾里德距離的平方 (Squared Euclidean distance) 作為測量距離的方法。為了避免不同計算單位的干擾，每一個因子都採用標準化後的數值 (Z scores) 來進行分析。經反覆多次分析，發現先排除低度使用量 (200 個以下) 的千門用戶碼後集群效果較好，共計可分成 3 群。分群後的相關統計資料與效果，請參考下表。

三因子於各集群之標準化平均數

集群	千門用戶碼 使用率	空號與非住宅 電話比率	住戶接觸率
1	0.8729	-1.0146	1.0998
2	0.0888	-0.1820	0.1163
3	-0.8408	1.0644	-1.0641

三因子的分群效果--變異數分析 (ANOVA)

變項名稱		平方和 (SS)	自由度 (df)	均方和 (MS)	F檢定	顯著水準
千門用戶碼 使用率 Eta ² =46.1%	組間 (組合)	12294.119	2	34.543	19935.749	<i>p</i> <0.001
	組內	5479.881	17772	0.427		
	總和	17774.000	17774			
空號與非住宅 電話比率 Eta ² =69.2%	組間 (組合)	8197.766	2	34.756	7606.889	<i>p</i> <0.001
	組內	9576.234	17772	0.423		
	總和	17774.000	17774			
住戶接觸率 Eta ² =73.6%	組間 (組合)	13080.448	2	44.099	24764.372	<i>p</i> <0.001
	組內	4693.552	17772	0.263		
	總和	17774.000	17774			

3.各分層樣本配置率：經集群分析後，抽樣分層共計分類為 4 層：前 3 層為中華電信市話千門用戶碼使用量在 200 個以上者，如上述所分類之 3 群；而第 4 層則是由非中華電信業者的市話局碼及中華電信市話千門用戶碼使用量低於 200 個以下者合併而成。在抽樣樣本配置時，採原中華電信 3 分層的前 5 年電話調查結果的平均住戶接觸率做為樣本配置率的估算標準，而第 4 分層因住戶比例不及 0.4%，故分派 0.5%。各分層樣本配置比率請參見下表。

層別	樣本配置率	住戶接觸率 平均數	千門用戶碼 總數	千門用戶碼 比率
1	0.500	0.51	5,113	0.231
2	0.350	0.35	6,651	0.301
3	0.145	0.15	6,011	0.272
4	0.005	NA*	4,346	0.196
總數	1.000		22,121	1.000

*：NA 表示無此項統計資訊。

(二)2020 年調研中心市話局碼抽樣分層

本專題中心於實際電話調查採行 2019 年市話調查抽樣分層後，發現第 1 層與第 2 層的住宅接觸率分別僅有 42.1%與 30.9% (如下表所示)，顯見因該抽樣分層在樣本配置上是採用過去 5 年電話調查的平均接觸結果，在市話住戶逐年下滑的態勢下，已經無法體現現況而勢必要加以調整。

層別	住宅接觸數 (列%)	空號 + 非住宅電話 (列%)	總撥號數
1	3032 (42.1)	2452 (34.1)	7200
2	1559 (30.9)	2302 (45.7)	5040
3	363 (17.4)	1307 (62.6)	2088
4	5 (6.9)	53 (73.6)	72
總數	4959	6114	14400

為有效反應抽樣分層的實質效用，本研究運用新電話調查的撥號結果重新調整原抽樣分層的配置，相關措施包括：(1)為提高第 1 層的住宅接觸率，新增篩選條件如下：千門用戶碼使用率 $\geq 70\%$ 、空號與非住宅電話比率 $\leq 30\%$ 、住宅接觸率 $\geq 50\%$ ，並將不符前述條件者移至第 2 層；(2)依前述條件而產生的新分層市話千門用戶碼的住宅接觸率來重新計算各分層的樣本分派配置比例，如下表所示。

層別	樣本配置比率	千門用戶碼總數	千門用戶碼比率
1	0.450	2,070	0.094
2	0.400	9,694	0.438
3	0.148	6,011	0.272
4	0.002	4,346	0.196
總數	1.000	22,121	1.000

二、行動電話門號抽樣分層

相較於市話千門用戶碼，行動電話號碼在執行抽樣分層分析時，僅具有千門用戶碼使用量的資訊，⁴而測試幾種切割使用量的分類方式後，以下表表現較優。後續擬於累積相當數量的撥號接觸結果時，再行檢測此抽樣分層的實際效益。

層別 (使用量)	千門用戶碼 數量	千門用戶碼 比率	總使用量	總使用量 比率	千門用戶碼 平均使用量
1 800-1000	6,544	0.104	5785,675	0.205	884.12
2 500-799	18,861	0.300	11553,360	0.408	612.55
3 300-499	19,424	0.309	7744,631	0.274	398.71
4 1-299	17,989	0.287	3200,987	0.113	177.94
	62,818	1.000	28284,653	1.000	

⁴ 行動電話門號的組成為 10 碼：OABC+DE+FGHI。其中，ABC 為服務類別接取碼，DE 為網路業者識別碼，FGHI 為用戶號碼。行動電話門號千門用戶碼 093387-1000~093387-1999、093387-2000~093387-2999 等，以此類推。使用率的計算方式：使用數量/1000。