Journal of Propagations in Probability and Statistics
15(1), 55-62 June 2015

統 計 薪 傳
第十五卷 第一期
民國104 年6 月55-62 頁

Uncorrected Galley Proof (頁碼 55-62 為暫定!)

淺談調查研究的集群抽樣

陳宇宏
展欣科技企業公司

張光昭
輔仁大學

摘 要 在抽樣調查的理論中，集群抽樣是一種重要的基本方法，一般的抽樣調查教科書或專書也都會談到這種方法。不過，集群抽樣往往被一些調查工作者或是非統計專業的學者以及大專學生誤認爲是另一種較爲眾人所熟知的二階集群抽樣，以致於造成更進一步的誤解與誤用。有鑑於此，本文作者就集群抽樣做一觀念上的介紹與說明，以期對於廣大的調查工作者與莘莘學子有所助益。

關鍵字詞 集群抽樣、二階集群抽樣、比值估計、有限母體、研究變數、輔助變數、簡單隨機抽樣、抽樣單位、相關系數。

1. 前言

在抽樣調查的理論中，集群抽樣(cluster sampling)是一種重要的基本方法，從許多抽樣調查的教科書或專書裡都可以查閱得到這種方法。不過，雖說集群抽樣只是一種基本方法，卻往往被一些調查工作者或是非統計專業的學者以及大專學生誤認爲是另一種較爲眾人所熟知且經常被使用的二階集群抽樣(two-stage cluster sampling)，進而造成不知不覺中的誤解與誤用。有鑑於此，本文作者就集群抽樣的基本觀念，做一淺介與說明，以期對於廣大的調查工作者與莘莘學子有所助益。

集群抽樣之所以容易被許多人誤解，主要的原因之一，就是這個名詞本身從字面上來看，其含意十分模糊與籠統，讓人感覺好像任何一種與集群有關的抽樣方法都可以算是集群抽樣；而在所有與集群有關的抽樣方法之中，又以二階集群抽樣最爲一般人所熟知與使用，因此集群抽樣經常被誤認爲就是二階集群抽樣。這種誤解就一般非統計專業的調查工作者或大專學生而言，倒也是情有可原；但如果統計專業的學者專家也誤解集群抽樣，那可就是個大笑話啦! 那麼，集群抽樣到底是個什麼玩意兒呢? 在大多數的抽樣調查書籍裡，講到集群抽樣，是指一種只有單一階段的抽樣方法(請參閱 Scheaffer *et al.* 2012 一書第 8 章)，可直稱爲單階集群抽樣(single-stage cluster sampling)(請參閱 Cochran 1977 一書第 9 章與 9A 章)，或稱爲簡單一階集群抽樣(simple one-stage cluster sampling)(請參閱 Levy and Lemeshow 1999 一書第 9 章)。想要瞭解集群抽樣，必得先瞭解比值估計(ratio estimation)，如下一節之中的簡介。

□民國一百零三年八月收稿，一百零四年三月修訂、定稿。

□本文第一作者爲展欣科技企業有限公司負責人，電子郵址: techcom5054@hotmail.com；第二作者爲輔仁大學統計資訊學系專任教授；電子郵址: stat1016@mail.fju.edu.tw。

□本文附英文摘要於參考文獻之後。本文適合大專院校三年級以上(含)程度閱讀。

 ISSN 1607-7083

2. 比值估計

由於集群抽樣在抽樣之後會運用到比值估計的方法來推估母體的參數，因此比值估計可說是學習集群抽樣之前的預備知識。所謂比值估計，是適用於雙變數母體的一種估計方法，同時也是抽樣調查理論中的重要基本方法之一。假設某一個有限母體(finite population)具有雙變數:研究變數(study variable)以及輔助變數(auxiliary variable)，我們以 Y 來表示研究變數、X 表示輔助變數。假設這個母體的母體大小(population size)為 N，也就是說，母體之中總共含有 N 個元素(element)；那麼母體之中的第 i 個元素的兩個變數值就可以用數對 (y_i, x_i) 來表示，$i = 1, 2, \cdots, N$。所以，兩個變數之母體平均數(population mean)的數學式可分別表示如下:

$$\text{研究變數 } Y \text{ 的母體平均數} = \mu_Y = \frac{1}{N}\sum_{i=1}^{N} y_i \text{ ，}$$

$$\text{輔助變數 } X \text{ 的母體平均數} = \mu_X = \frac{1}{N}\sum_{i=1}^{N} x_i \text{ 。}$$

其次，兩個變數之母體總和數(population total)的數學式可分別表示如下:

$$\text{研究變數 } Y \text{ 的母體總和數} = \tau_Y = N\mu_Y = \sum_{i=1}^{N} y_i \text{ ，}$$

$$\text{輔助變數 } X \text{ 的母體總和數} = \tau_X = N\mu_X = \sum_{i=1}^{N} x_i \text{ 。}$$

在比值估計的方法之中，有一個重要的母體參數，可稱為比值參數(以 R 表示此參數)，它是兩個變數之母體平均數(或是母體總和數)的比值，如下所示:

$$R = \frac{\mu_Y}{\mu_X} = \frac{\tau_Y}{\tau_X} \text{ 。} \tag{1}$$

比值參數 R 通常是一個未知其值的母體參數，也是抽樣調查研究者想要估計的參數。假設我們使用簡單隨機抽樣(simple random sampling)從以上的雙變數母體之中抽出 n 個樣本，其樣本觀測值為 (Y_i, X_i) ，$i = 1, 2, \cdots, n$，那麼想要估計比值參數 R，就十分容易，只要將(1)式之中的兩個變數之母體平均數換成樣本平均數即可，如下所示:

$$\text{比值參數 } R \text{ 的估計量} = \hat{R} = \frac{\bar{Y}}{\bar{X}} = \frac{(1/n)\sum_{i=1}^{n} Y_i}{(1/n)\sum_{i=1}^{n} X_i} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i} \text{ 。} \tag{2}$$

以上比值參數 R 的估計量 $\hat{R}$ 是否能夠準確地估計 R，要取決於 Y 與 X 這兩個變數之間是否具有高度的正相關性，若正相關性很高，譬如說相關係數大於 0.9，那麼 $\hat{R}$ 就會是一個相當準確的估計量；但若相關性不夠高，譬如說相關係數小於 0.5，那麼就還是不要使用 $\hat{R}$ 為妙；如果 Y 與 X 這兩個變數之間不具有正相關性，卻具有負相關性，那可就千千萬萬不能使用 $\hat{R}$ 這個估計量啦！ 此外，估計量 $\hat{R}$ 還有一個理論上的小缺點: 它是一個具有偏差的估計量(biased estimator)；不過，只要 Y 與 X 這兩個變數之間具有高度的正相關性，那麼這個小缺點的負面影響也就很小。

(2)式之中估計量 $\hat{R}$ 的一個主要用途，就是可以被間接地用來估計研究變數 Y 的母體平均

數 μ_Y 或是 Y 的母體總和數 τ_Y 。我們首先將(1)式改寫成爲以下二式:

$$\mu_Y = R\mu_X \quad 與 \quad \tau_Y = R\tau_X 。 \tag{(3) 與 (4)}$$

接著，如果我們假設 μ_X 是一個已知其值的母體參數，那麼只要將(3)式之中等號右邊的比值參數 R 換成它的估計量 $\hat{R}$ ，即可得到等號左邊之參數 μ_Y 的估計量，如下所示:

$$\hat{\mu}_Y = \hat{R}\mu_X 。 \tag{5}$$

以上(5)式之中的估計量 $\hat{\mu}_Y$ 就是赫赫有名的比值估計量(ratio estimator)。同理，如果我們假設 τ_X 是一個已知其值的母體參數，那麼只要將(4)式之中等號右邊的比值參數 R 換成它的估計量 $\hat{R}$ ，即可得到等號左邊之參數 τ_Y 的比值估計量如下:

$$\hat{\tau}_Y = \hat{R}\tau_X 。 \tag{6}$$

在下一節所討論的集群抽樣，其估計方法與本節所討論的比值估計有密切的關聯性。

3. 集群抽樣

假設某一個有限母體由 N 個集群(cluster)所構成，其中第一個集群包含 M_1 個元素，第二個集群包含 M_2 個元素，餘類推；也就是說，母體之中的第 i 個集群之集群大小(cluster size)爲 M_i ，$i = 1, 2, \cdots, N$。那麼，整個母體之中的元素總個數(也就是母體大小)就可以被表示爲

$$母體大小 = M = M_1 + M_2 + \cdots + M_N = \sum_{i=1}^{N} M_i 。$$

令 y_{ij} 代表第 i 個集群之中第 j 個元素的研究變數值，那麼整個母體的平均數與總和數就如以下所示:

$$母體平均數 = \mu_Y = \frac{1}{M}\sum_{i=1}^{N} y_i = \frac{1}{M}\sum_{i=1}^{N}\sum_{j=1}^{M_i} y_{ij} ,$$

$$母體總和數 = \tau_Y = M\mu_Y = \sum_{i=1}^{N} y_i = \sum_{i=1}^{N}\sum_{j=1}^{M_i} y_{ij} ,$$

其中

$$y_i = \sum_{j=1}^{M_i} y_{ij} = 第\ i\ 個集群的集群總和數(cluster total)。$$

所謂集群抽樣，就是使用簡單隨機抽樣從母體的 N 個集群之中抽出 n 個集群，每一個被抽出的集群，就相當於一個樣本，而抽樣的過程到此就算是結束了，所以這是一種只有單一階段的抽樣方法，其過程十分簡易，只是將一般簡單隨機抽樣的抽樣單位(sampling unit)從元素換成集群而已。不過，抽樣之後接下來要估計母體參數的過程，就稍微有一點兒複雜啦! 首先，從每一個被抽出的 n 個樣本集群，我們可以取得一個數對型態的樣本觀測值，如後: $(Y_i, \tilde{M}_i)$ ，$i = 1, 2, \cdots, n$，其中 Y_i 代表被抽出 n 個樣本集群之中的第 i 個集群的集群總和數，$\tilde{M}_i$ 是這第 i 個集群的集群大小。有了這些樣本數據，接下來如果我們想要估計母體平均數 μ_Y ，

那麼可以使用的估計量可就不只一種啦！我們首先考慮一種或許是最常用也是最重要的估計量如下:

$$\hat{\mu}_Y^{(1)} = \frac{Y_1 + Y_2 + \cdots + Y_n}{\tilde{M}_1 + \tilde{M}_2 + \cdots + \tilde{M}_n} = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n \tilde{M}_i} \text{。} \tag{7}$$

以上(7)式之中的估計量 $\hat{\mu}_Y^{(1)}$，其數學式的型態與(2)式之中的估計量極爲相似，只不過是將(2)式之中分母連加符號之內的 X_i 換成 $\tilde{M}_i$ 而已！所以，$\hat{\mu}_Y^{(1)}$ 是一種屬於比值型態的估計量，其研究變數值是集群總和數 Y_i，輔助變數值是集群大小 $\tilde{M}_i$，而研究變數與輔助變數之間通常會具有高度的正相關性；舉例來說，如果究變數值 Y_i 是一個住戶之內的機車總數量，輔助變數值 $\tilde{M}_i$ 是該住戶之內成年人的總人數，那麼住戶之內的成年人愈多，其機車總數量通常也就愈多，所以一個住戶之內的機車總數量與成年人的總人數之間通常會具有相當成程度的正相關性；至於估計量 $\hat{\mu}_Y^{(1)}$ 所估計的母體參數，則是母體之中任何一個成年人所擁有機車數量的平均數(所以構成母體的元素是許許多多的成年人，而構成母體的集群則是許多住戶，元素隱藏於集群之內)。雖然 $\hat{\mu}_Y^{(1)}$ 是一種比值型態的估計量，我們也可以從“非比值”的觀點來解釋它的數學式，以機車的例子來說，分子的 Y_i 連加式相當於被抽出 n 個樣本住戶的機車總數量，分子的 $\tilde{M}_i$ 連加式則是被抽出 n 個住戶的成年人總人數，所以

$$\text{母體之中任一位成年人所擁有機車數量的平均數} \approx \frac{\text{樣本的機車總數量}}{\text{樣本的成年人總人數}} \text{，}$$

這是非常淺顯易懂的平均數概念。接著，如果母體大小 M 是一個已知其值的母體參數，我們就可以利用(7)式的估計量 $\hat{\mu}_Y^{(1)}$ 來估計母體總和數 τ_Y，如下:

$$\hat{\tau}_Y^{(1)} = M\hat{\mu}_Y^{(1)} = M\left(\frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n \tilde{M}_i}\right) \text{。} \tag{8}$$

由於(7)式與(8)式之中的 $\hat{\mu}_Y^{(1)}$ 與 $\hat{\tau}_Y^{(1)}$ 皆爲比值型態的估計量，因此它們皆屬於具有偏差的估計量；不過如同前一節所述，只要 y_i 與 M_i 這兩個變數值之間具有高度的正相關性，那麼這個具有偏差的小缺點也就沒有太大的負面影響。

除了以上(7)式與(8)式的比值型態估計量，另有一種屬於“非比值”型態的估計量也很重要，其數學式如下:

$$\hat{\tau}_Y^{(2)} = N\left(\frac{1}{n}\sum_{i=1}^n Y_i\right) = \frac{N}{n}\sum_{i=1}^n Y_i \text{，} \tag{9}$$

$$\hat{\mu}_Y^{(2)} = \frac{1}{M}\hat{\tau}_Y^{(2)} = \frac{N}{Mn}\sum_{i=1}^n Y_i \text{。} \tag{10}$$

我們先來解釋(9)式之中的估計量 $\hat{\tau}_Y^{(2)}$: 首先，這個估計量完全沒有用到樣本觀測值 $(Y_i, \tilde{M}_i)$ 之中的 $\tilde{M}_i$，所以當然就屬於“非比值”型態的估計量。接著，再以機車的例子來說明，式中的 Y_i 連加式相當於被抽出 n 個樣本住戶的機車總數量，前面再以 n 除之，就成爲 n 個樣本住戶的單戶機車平均量；再以母體的住戶總間數，N，乘以此一單戶機車平均量，就得到母體之中所

有機車總數量的估計量。

在下一節，我們以機車總數量的例子，模擬製造出一個想像中的有限母體，來說明如何計算(8)與(9)式之中的估計量。

4. 計算範例

假設某社區大樓共有 100 間住戶，每一間住戶的機車總數量 y_i 以及戶內的成年人總人數 M_i，以數對 (y_i, M_i) 之型態列表如下:

表 1　某社區大樓 100 間住戶的數對 (y_i, M_i) 資料
(y_i: 第 i 間住戶的機車總數量；M_i: 第 i 間住戶的成年人總人數)

i	(y_i, M_i)	i	(y_i, M_i)	i	(y_i, M_i)	i	(y_i, M_i)	i	(y_i, M_i)
1	(2, 3)	21	(2, 4)	41	(3, 4)	61	(3, 5)	81	(3, 4)
2	(1, 2)	22	(3, 4)	42	(3, 5)	62	(2, 4)	82	(3, 5)
3	(2, 4)	23	(4, 5)	43	(4, 6)	63	(3, 4)	83	(4, 6)
4	(3, 4)	24	(3, 5)	44	(5, 6)	64	(3, 5)	84	(5, 6)
5	(1, 2)	25	(2, 3)	45	(1, 2)	65	(4, 6)	85	(1, 2)
6	(0, 1)	26	(1, 2)	46	(2, 3)	66	(4, 7)	86	(2, 3)
7	(2, 4)	27	(2, 3)	47	(2, 4)	67	(3, 5)	87	(2, 4)
8	(3, 4)	28	(3, 4)	48	(3, 5)	68	(2, 3)	88	(3, 5)
9	(3, 5)	29	(2, 4)	49	(3, 4)	69	(1, 2)	89	(3, 4)
10	(2, 4)	30	(4, 6)	50	(4, 5)	70	(2, 3)	90	(4, 5)
11	(3, 4)	31	(5, 6)	51	(3, 4)	71	(2, 3)	91	(0, 1)
12	(3, 5)	32	(1, 2)	52	(3, 5)	72	(3, 4)	92	(2, 3)
13	(4, 6)	33	(2, 3)	53	(2, 4)	73	(2, 4)	93	(3, 4)
14	(5, 6)	34	(2, 4)	54	(3, 4)	74	(4, 6)	94	(2, 4)
15	(1, 2)	35	(3, 5)	55	(3, 5)	75	(5, 6)	95	(4, 6)
16	(2, 3)	36	(0, 1)	56	(6, 8)	76	(1, 2)	96	(3, 5)
17	(2, 4)	37	(2, 3)	57	(1, 2)	77	(2, 3)	97	(3, 6)
18	(3, 5)	38	(3, 4)	58	(2, 3)	78	(3, 5)	98	(2, 4)
19	(3, 4)	39	(5, 7)	59	(2, 4)	79	(2, 4)	99	(1, 2)
20	(4, 5)	40	(4, 6)	60	(3, 5)	80	(3, 6)	100	(2, 3)

若將該社區大樓視爲一個有限母體，每一間住戶視爲一個集群，大樓每一間住戶的每一位成年人視爲一個元素，機車數量視爲研究變數，則

大樓住戶的總間數 = 母體之中的集群總個數 = N = 100，

社區大樓所有住戶的成年人總人數 = 母體大小 $= M = \sum_{i=1}^{N} M_i = 3 + 2 + \cdots + 3 = 415$，

社區大樓所有住戶的機車總數量 = 母體總和數 $= \tau_Y = \sum_{i=1}^{N} y_i = 2 + 1 + \cdots + 2 = 266$，

如果我們使用集群抽樣之方法從母體抽出 n = 8 個集群，來估計母體總和數 τ_Y，那麼我們首先要從 100 間住戶之中隨機地抽出 8 間住戶，這就需要介於 1 與 100 之間的 8 個亂數，來選出這 8 間住戶。我們不妨利用許多統計學教科書附錄之中的亂數表，來挑選 8 個亂數。從某一本統計學教科書(Walpole 1982)的附錄之中，我們節錄了亂數表的一小部份，如以下表 2 所示:

表 2 從統計學書籍中節錄的一小部份亂數表

	column							
line	1-5	6-10	11-15	16-20	21-25	26-30	31-35	35-40
1	62956	95735	70988	86027	27648	65155	46301	27217
2	17143	50118	41681	87224	75674	43371	09846	83403
3	99285	01369	94610	71099	69207	01999	23931	34711
4	12940	81308	40436	82916	74245	70324	88555	82182
5	28089	80216	08681	83524	00583	55179	31911	68484

爲了能夠簡便且迅速地抽出介於 1 與 100 之間的 8 個亂數，我們就從表 2 的第一個橫列由左至右依序選取 8 個二位數字如後: 62、95、69、57、35、70、98、88，所以依據集群抽樣之方法而抽得 8 間住戶的數對型態樣本數據爲

(2, 4)、(4, 6)、(1, 2)、(1, 2)、(3, 5)、(2, 3)、(2, 4)、(3, 5)。

如果母體大小 M 是一個已知其值的母體參數，我們就可以利用(8)式的估計量 $\hat{\tau}_Y^{(1)}$ 來估計母體總和數 τ_Y ，如下:

$$\hat{\tau}_Y^{(1)} = M\left(\frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} \tilde{M}_i}\right) = 415 \cdot \frac{2+4+\cdots+3}{4+6+\cdots+5} = 415 \cdot \frac{18}{31} \approx 240.97 \text{ 。} \tag{11}$$

如果母體之中的集群總個數，N，是一個已知其值的母體參數，我們就可以利用(9)式的估計量 $\hat{\tau}_Y^{(2)}$ 來估計母體總和數 τ_Y ，如下:

$$\hat{\tau}_Y^{(2)} = \frac{N}{n}\sum_{i=1}^{n} Y_i = \frac{100}{8}(2+4+\cdots+3) = \frac{100}{8} \cdot 18 = 225 \text{ 。} \tag{12}$$

就(11)與(12)二式的估計值來比較，很顯然(11)式的估計值其誤差較小(因爲母體總和數的標準答案爲 $\tau_Y = 266$)。一般而言，如果 y_i 與 M_i 這兩個變數之間具有相當高度的正相關性，那麼(11)式的估計方法會優於(12)式的估計方法。我們不妨利用抽得 8 間住戶的數對型態樣本數據來推估母體的相關系數(correlation coefficient)，看看 y_i 與 M_i 這兩個變數之間是否具有相當高度的正相關性。就前一節所討論的比值估計而言，母體相關系數的定義如下:

$$\rho_{XY} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_X)(y_i-\mu_Y)}{\sqrt{\sigma_X^2\sigma_Y^2}} = \frac{\sum_{i=1}^{N}(x_i-\mu_X)(y_i-\mu_Y)}{\sqrt{\left(\sum_{i=1}^{N}(x_i-\mu_X)^2\right)\left(\sum_{i=1}^{N}(y_i-\mu_Y)^2\right)}}$$

其中 σ_Y^2 與 σ_X^2 分別爲研究變數 Y 與輔助變數 X 的母體變異數。至於母體相關系數 ρ_{XY} 的估計量，則爲

$$\hat{\rho}_{XY} = r_{XY} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right)\left(\sum_{i=1}^{n}(Y_i-\bar{Y})^2\right)}} = \frac{\sum_{i=1}^{n}(X_iY_i)-n\bar{X}\bar{Y}}{\sqrt{\left(\sum_{i=1}^{n}X_i^2-n(\bar{X})^2\right)\left(\sum_{i=1}^{n}Y_i^2-n(\bar{Y})^2\right)}}$$

(請參閱 Levy and Lemeshow 1991 一書的 pp.193-197)。因此，y_i 與 M_i 這兩個變數之間的母體相關系數估計值爲

$$\hat{\rho}_{MY}=\frac{\sum_{i=1}^{n}(\tilde{M}_i Y_i)-n\bar{\tilde{M}}\bar{Y}}{\sqrt{\left(\sum_{i=1}^{n}\tilde{M}_i^2-n(\bar{\tilde{M}})^2\right)\left(\sum_{i=1}^{n}Y_i^2-n(\bar{Y})^2\right)}}$$

$$=\frac{(4\cdot 2+\cdots+5\cdot 3)-(1/8)(4+\cdots+5)(2+\cdots+3)}{\sqrt{(4^2+\cdots+5^2)-(1/8)(4+\cdots+5)^2}\sqrt{(2^2+\cdots+3^2)-(1/8)(2+\cdots+3)^2}}$$

$$=\frac{80-(1/8)(31)(18)}{\sqrt{135-(1/8)(31)^2}\sqrt{48-(1/8)(18)^2}}=\frac{10.25}{\sqrt{14.875}\sqrt{7.5}}\approx 0.97\text{ 。}$$

由於 $\hat{\rho}_{MY}\approx 0.97$ 已十分接近相關系數的極大值 1， y_i 與 M_i 這兩個變數之間可說是極高度的正相關，無怪乎(11)式的估計誤差遠小於(12)式的估計誤差。其實，只要母體相關系數之值大於 0.8，(11)式的估計方法通常都會優於(12)式的估計方法。

5. 結語

本文作者藉由一個模擬想像製造出的有限母體，來說明抽樣調查理論中較不爲人所熟知的集群抽樣，希望對於廣大的調查工作者與莘莘學子有所助益。

參考文獻

Cochran, W. G. (1977). Sampling Techniques, 3rd ed., John Wiley & Sons, INC.

Levy, P. S. and Lemeshow, S. (1999). Sampling of Populations, 3rd ed., John Wiley & Sons, INC.

Scheaffer, R. L., Mendenhall, W., Ott, R.L., and Gerow, K. G. (2012). Survey Sampling, 7th ed., Brooks/Cole, Cengage Learning.

Walpole, R. E. (1982). Introduction to Statistics, 3rd ed., Macmillan Publishing Co., Inc., New York.

Journal of Propagations in Probability and Statistics
15(1), 55-62 June 2015

Teaching "Cluster Sampling" in Survey Research

Ardor Chen
Techcom Information Corp.

Kuang-Chao Chang
Fu Jen Catholic University

ABSTRACT Cluster sampling is one of the basic and important methods in the theory of survey sampling. Somehow cluster sampling has been misused and confused with two- stage sampling by many practitioners of survey sampling and college students. In this article, we introduce the basic concept of cluster sampling with a simulated numerical example, and we hope the contents of this article can be useful for practitioners of survey sampling and college/university students who take statistics courses.

Keywords Cluster sampling; Two-stage cluster sampling; Ratio estimation; Finite population; Study variable; Auxiliary variable; Simple random sampling; Sampling unit; Correlation coefficient.

□ Received August 2014, revised March 2015, in final form March 2015.

□ Ardor Chen is the founder and CEO of Techcom Information Corp., Taipei, Taiwan, ROC; email: techcom5054@hotmail.com. Kuang-Chao Chang is a Professor in the Department of Statistics and Information Science at Fu Jen Catholic University, New Taipei City, Taiwan, ROC; email: stat1016@mail.fju.edu.tw.

ISSN 1607-7083