

《抽樣方法》

試題評析	此卷共有四大題，第一大題是分層隨機抽樣法，屬於基礎之計算題型，考古題中也常出現這類計算題。第二大題是群集隨機抽樣法，考生可以用第七章簡單估計法或第十一章比率估計法回答，屬於基礎之計算題型。第三大題屬於 SRS 的不偏估計量，特別是(二)的不偏估計量，雖然考古題未曾出現過，但總複習課程老師已經精準預測到，考生對此估計量應該不會陌生的。第四大題混合了分層抽樣法與 SRS 兩種抽樣設計，只要考生小心題目之估計目標與抽樣方式，拿到高分應該不難。
考點命中	第一題：《高點·高上抽樣方法講義》，趙治勳編撰，第四章。 第二題：《高點·高上抽樣方法講義》，趙治勳編撰，第七章或第十一章。 第三題：《高點·高上抽樣方法講義》，趙治勳編撰，第一章與總複習。 第四題：《高點·高上抽樣方法講義》，趙治勳編撰，第一章與第四章。

95%信心水準請利用 $Z_{0.025}=1.96$ 計算

- 一、光明社區管理委員會主任委員希望能調查住戶對禁止住戶在社區中庭花園抽菸以免影響其他住戶的意見，因為該社區住戶高達 700 戶，分別分布在 A、B、C、D 四棟大樓，因此主任委員在各棟大樓各用簡單隨機抽樣取出不放回的方式選擇了 50 戶住戶加以調查其意向。各棟大樓戶數及贊成社區中庭花園禁菸之樣本比例如下：

大樓編號	戶數	贊成比例
A	100	0.8
B	300	0.4
C	200	0.5
D	100	0.7

請回答下列問題：

- (一) 請問光明社區住戶贊成禁止住戶在社區中庭花園抽菸比例之不偏估計推估值。(5 分)
- (二) 請問以上推估值在 95% 信心水準下的最大估計誤差。(10 分)
- (三) 在試行此一規定三個月後，主任委員想再做一次同樣的調查，請問在抽樣設計及總抽樣數均維持不變的情況下，可以如何來改進推估的精確程度？(10 分)

答：

h	N_h	n_h	p_h
1	100	50	0.8
2	300	50	0.4
3	200	50	0.5
4	100	50	0.7

$$(一) p_{st} = \sum_{h=1}^L W_h p_h = \frac{100}{700} \times 0.8 + \frac{300}{700} \times 0.4 + \frac{200}{700} \times 0.5 + \frac{100}{700} \times 0.7 = 0.5286$$

$$(二) s_{p_{st}} = \frac{1}{N} \sqrt{\sum_{h=1}^L N_h (N_h - n_h) \frac{n_h - 1}{n_h} p_h q_h}$$

$$= \frac{1}{700} \sqrt{100(100-50) \frac{50}{50-1} (0.8)(0.2) + 300(300-50) \frac{50}{50-1} (0.4)(0.6) + 200(200-50) \frac{50}{50-1} (0.5)(0.5) + 100(100-50) \frac{50}{50-1} (0.7)(0.3)}$$

$$= 0.03375$$

最大估計誤差為 $B = 1.96s_{p_s} = 0.06615$

(三)在各層樣本數上改採紐門配置之方式進行，如下：

[假設各層每單位之抽樣成本相同($c_1 = c_2 = c_3 = c_4$)]

$$n_h = \frac{N_h S_h}{\sum_{h=1}^L N_h S_h} \times n$$

$$\sum_{h=1}^L N_h S_h \stackrel{S_h \text{ 未知, 以 } s_h \text{ 代}}{=} 100 \times \sqrt{0.1633} + 300 \times \sqrt{0.2449} + 200 \times \sqrt{0.2551} + 100 \times \sqrt{0.2143}$$

$$= 336.1799$$

$$n_1 = \frac{100 \times \sqrt{0.1633}}{336.1799} \times 200 = 24.0409 \approx 24$$

$$n_2 = \frac{300 \times \sqrt{0.2449}}{336.1799} \times 200 = 88.323 \approx 88$$

$$n_3 = \frac{200 \times \sqrt{0.2551}}{336.1799} \times 200 = 60.0957 \approx 60$$

$$n_4 = n - n_1 - n_2 - n_3 = 28$$

二、某鎮農會欲估算該鎮豬農養豬飼料之花費，以簡單隨機抽樣取出不放回的方式於該鎮的 10 戶豬農選擇了 5 戶，詢問其過去一季平均養豬隻頭數及飼料花費。調查所得資料如下：

豬農編號	平均養豬隻頭數	飼料花費(仟元)
1	200	850
2	400	1600
5	100	450
8	50	300
10	250	800

根據此一調查結果，請問你會如何估算該鎮豬農平均每隻豬隻每季飼料花費？請解釋你的想法（可由估計及／或抽樣設計的觀點），並提出推估結果在 95%信心水準下的最大估計誤

差。(15分)

答：

$$\hat{M} = N\bar{m} = 10 \times \left(\frac{200 + 400 + 100 + 50 + 250}{5} \right) = 2000$$

$$\bar{y}_{cl} = \frac{N}{\hat{M}} \bar{y}_t = \frac{10}{2000} \times 800 = 4 \quad (\text{仟元})$$

本抽樣設計為群集隨機抽樣法，我認為本案例使用此抽樣方法是適當的，因為 10 戶豬農設置於同一個鎮內，因為一個鎮之涵蓋範圍通常不大，故養殖資源與成本應該相近(組間變異小)，但每一個豬農之養殖數量或豬隻品種都有所不同(組內變異大)，在這樣之情況下，使用群集隨機抽樣法會比 SRS 之精確度為高。

$$s_{\bar{y}_{cl}} = \frac{N}{\hat{M}} \sqrt{(1-f) \frac{s_t^2}{n}} = \frac{10}{2000} \sqrt{\left(1 - \frac{5}{10}\right) \frac{253750}{5}} = 0.7965 \quad (\text{仟元})$$

$$\text{其中 } s_t^2 = \frac{\sum_{i=1}^n (y_i - \bar{y}_t)^2}{n-1} = 253750$$

最大估計誤差為 $B = 1.96s_{\bar{y}_{cl}} = 1.5611$ (仟元)

三、簡單隨機抽樣可分為取出不放回及取出放回兩大類，樣本觀察值之平均在兩種設計中均為母體平均 μ 之不偏估計，若樣本數為 n ， S_0 為樣本集合，該不偏估計為

$$\bar{y} = \frac{1}{n} \sum_{i \in S_0} y_i$$

請回答下列問題：

- (一)如果均以 \bar{y} 作為 μ 之估計量，雖然均為樣本中觀察值之平均，但在相同的樣本數下，取出不放回的設計會比取出放回相對較有效 (more efficient)，請闡述其原因，並討論其相對有效性與樣本數及母體數之關係。(10分)
- (二)在取出放回的設計中有另外一個母體平均 μ 之不偏估計，該估計為樣本中不同單元觀察值之平均，亦即若有重複出現之單元，其觀察值只會被計入一次，該不偏估計量為

$$\bar{y}_v = \frac{1}{v} \sum_{i \in S} y_i$$

其中 S 為 S_0 中不同單元所構成之集合， v 為 S 中之單元數目，且 $v \leq n$ 。

抽樣理論中之一基本結果為在簡單隨機抽樣取出放回設計中， \bar{y}_v 亦較 \bar{y} 相對較有效，請說明其理由。(5分)

答：

令 \bar{y}_{SRSWR} 、 $\bar{y}_{SRSSWOR}$ 分別表使用取出放回與取出不放回下之樣本平均數

$$\bar{y}_{SRSWR} = \bar{y}_{SRSSWOR} = \frac{1}{n} \sum_{i \in S_0} y_i$$

(一)有限母體下變異數為 $S^2 = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N-1}$

$$\frac{V(\bar{y}_{SRSWOR})}{V(\bar{y}_{SRSWR})} = \frac{(1 - \frac{n}{N}) \frac{S^2}{n}}{\frac{N-1}{N} \frac{S^2}{n}} = \frac{N-n}{N-1} < 1, \text{ 因此 } \bar{y}_{SRSWOR} \text{ 相對有效}$$

由 $\frac{N-n}{N-1}$ 得知，當樣本數 n 越接近母體數 N 時， $N-n$ 會越大，此時取後不放回 (\bar{y}_{SRSWOR}) 之有效性越接近取出放回 (\bar{y}_{SRSWR})。

(二)估計之有效性會受到樣本數之影響，當樣本數越大時有效性會隨之提高。從 $\bar{y}_{SRSWR} = \frac{1}{n} \sum_{i \in s_0} y_i$ 可以知道

有 n 個觀察值，而 $\bar{y}_v = \frac{1}{v} \sum_{i \in s} y_i$ 有 v 個觀察值，由 $v < n$

可得 \bar{y}_v 也較 \bar{y}_{SRSWR} 更具相對有效性。

四、欣民里里長希望能將里內獲得整建該里休閒公園之經費用於該公園內規劃一狗公園，作為里民飼養家犬之活動空間，藉以減少里民於公園遛狗時可能造成之環境污染及人犬衝突，因此里長希望能先進行可行性評估，其中包括對里民的問卷抽樣調查，問卷內容包括：

1. 家中是否養狗。
2. 若有養狗，請問飼養隻數。
3. 是否贊成狗公園之規劃。

在該里的 500 戶中，里長以簡單隨機抽樣取出不放回的方式選擇了 50 戶作為樣本，調查資料顯示其中有 10 戶養狗。10 戶有養狗的樣本戶對另外兩題之答案彙整如下：

家戶編號	68	124	136	158	248	314	379	401	452	479
家犬隻數	1	1	1	3	1	1	4	1	1	1
是否贊成	是	是	是	否	是	否	是	否	是	否

(一)該里家戶養狗比例之不偏估計推估值為何？及此一比例之 95%信賴區間？(5 分)

(二)請問該里養狗之家戶平均養狗隻數之不偏估計，以及其 95%信賴區間？(10 分)

(三)請問該里養狗總隻數之不偏估計，以及其 95%信賴區間？(10 分)

(四)調查資料中沒有養狗的家戶贊成設置狗公園的戶數是 16 戶，但是有人質疑此一抽樣設計或許沒有考慮家戶年齡結構，因為有高齡長者的家戶可能希望規劃友善樂齡運動空間，經重新審視資料，樣本中有 20 戶為家中有 65 歲以上高齡里民，而其中有 2 戶贊成，而戶籍資料顯示欣民里中有 350 戶家中有 65 歲以上長者，請問根據此一資料，在考量家戶年齡結構之要求下，贊成設置狗公園之比例不偏估計為何？你是否認為此一估計量較為恰當？請說明理由。(10 分)

(五)若有里民質疑樣本數過少以至於養狗戶數之推估精確度不足，若欲重新以相同抽樣設計執行此一調查，但要將推估養狗戶數比例控制在 95%信心水準下不超過正負 5%，請問樣

本戶數至少需要多少？（10分）

答：

(一) 令 P 表該里家戶有養狗之比例

$$p = \frac{a}{n} = \frac{10}{50} = 0.2$$

$$s_p = \sqrt{(1-f) \frac{\frac{n}{n-1} pq}{n}} = \sqrt{\left(1 - \frac{50}{500}\right) \frac{\frac{50}{50-1} (0.2)(0.8)}{50}} = 0.05421$$

$$(p \pm 1.96s_p) = (0.09375, 0.30625)$$

(二) 令 \bar{Y} 表該里有養狗家戶中之平均養狗隻數

$$\bar{y} = \frac{y}{n} = \frac{15}{10} = 1.5 (\text{隻})$$

$$s_{\bar{y}} = \sqrt{(1-f) \frac{s^2}{n}} = \sqrt{\left(1 - \frac{10}{500}\right) \frac{1.1667}{10}} = 0.3381 (\text{隻})$$

$$(\bar{y} \pm 1.96s_{\bar{y}}) = (0.8373, 2.1627) (\text{隻})$$

(三) 令 τ 表該里養狗之總隻數

$$\hat{\tau} = (Np)\bar{y} = 500 \times 0.2 \times 1.5 = 150 (\text{隻})$$

$$s_{\hat{\tau}} = Ns_p s_{\bar{y}} = 500 \times 0.05421 \times 0.3381 = 9.1642 (\text{隻})$$

$$(\hat{\tau} \pm 1.96s_{\hat{\tau}}) = (132.0382, 167.9618) (\text{隻})$$

(四) 令 P 表該里贊成設置狗公園之比例

	h	N_h	n_h	p_h
65 歲以上	1	350	20	$\frac{20}{200} = 0.1$
65 歲以下	2	150	30	$\frac{20}{30} = 0.6667$

$$P_{st} = \sum_{h=1}^L W_h p_h = \frac{350}{500} \times 0.1 + \frac{150}{500} \times 0.6667 = 0.27$$

$$s_{P_{st}} = \frac{1}{N} \sqrt{\sum_{h=1}^L N_h (N_h - n_h) \frac{\frac{n_h}{n_h-1} p_h q_h}{n_h}}$$

$$\begin{aligned}
&= \frac{1}{500} \sqrt{350(350-20) \frac{\frac{20}{20-1}(0.1)(0.9)}{20} + 150(150-30) \frac{\frac{30}{30-1}(0.6667)(0.3333)}{30}} \\
&= 0.05235
\end{aligned}$$

$$\therefore v(p) = 0.05421^2 > v(p_{st}) = 0.05235^2$$

∴ 考慮家戶年齡結構之估計量較適合本案

(五) 令 P 表該里家戶有養狗之比例

$$n_0 = \frac{z_{\alpha}^2 pq}{B^2} = \frac{1.96^2 \times 0.2 \times 0.8}{0.05^2} = 245.824$$

$$n = \frac{n_0}{1 + \frac{n_0}{N}} = 164.8175 \approx 165$$

【版權所有，重製必究！】