

《抽樣方法》

與本試題有關的標準常態值可能有：

$$Z_{0.025} = 1.96, Z_{0.05} = 1.645, Z_{0.1} = 1.28,$$

$$t_{0.025}(2) = 4.303, t_{0.025}(3) = 3.182$$

計算題的答案要求：

1. 平均數取兩位小數（算到第三位小數然後四捨五入）。
2. 請以分數表示母體或是樣本比例的答案。
3. 至於次數或是個數的答案取整數（算到第一位小數然後四捨五入）。
4. 變異數取三位小數（算到第四位小數然後四捨五入）。
5. 信賴區間請代入最多小數的平均數與變異數，然後取兩位小數（算到第三位小數然後四捨五入）。
6. 變異數與變異數的比例取三位小數（算到第四位小數然後四捨五入）。
7. 計算過程保留越多小數越好，以確保滿足第8項要求。
8. 答案只要與標準答案差別在最後一位小數以內都算是正確答案。

一、某消防局為了瞭解平均每天多少人打電話進來報案，於是隨機抽樣30天的電話紀錄並且得到下表：

次數	0	1	2	3	4
天數	15	10	4	0	1

假設這是一個無限母體的問題，並且請根據上述表格回答以下的問題。

(一)如果該消防局想估計一年365天下來的總報案次數，請協助消防局提供這一項估計值。（5分）

(二)請提供上述總報案次數之變異數的不偏估計值。（10分）

(三)請提供總報案次數之近似的95%信賴區間。（10分）

試題評析	本題是無限母體之參數估計考題，這是一題統計學之考題，只要熟讀統計學，拿滿分很容易。
考點命中	《高點統計學講義》，趙治勳編撰。

答：

$$(一) \text{樣本平均數 } \bar{y} = \frac{0 \times 15 + 1 \times 10 + 2 \times 4 + 3 \times 0 + 4 \times 1}{30} = 0.73 \text{ 估計母體平均數 } \bar{Y}$$

$$\text{故以 } y = N\bar{y} = 365 \times 0.73 = 266.45 \approx 266 \text{ 次估計一年之總報案次數 } Y$$

$$(二) \text{樣本平均數之變異數爲 } s_y^2 = \frac{\sum_{i=1}^{30} (y_i - \bar{y})^2}{30 - 1} = 0.892 \text{ (無限母體下抽出率 } f \text{ 趨近 } 0)$$

$$\text{故 } s_y^2 = N^2 s_{\bar{y}}^2 = 365^2 \times 0.892 = 118836.7 \text{ 爲總報案次數 } Y \text{ 之變異數之不偏估計量}$$

(三)總報案次數 Y 之 95% 近似信賴區間爲

$$(y \pm 1.96 s_y) = (266 \pm 1.96 \sqrt{118836.7}) = (-409.66, 941.66)$$

【版權所有，重製必究！】

二、某校有45棵樹，為了方便管理，每一棵樹上都被貼上了一個編號，從「00」開始一直編到「44」。該校有一條榕樹大道，大道上的榕樹排列及它們的編號分別是04, 05, 07, 10, 15, 17, 19, 25, 27, 28, 29, 30, 31, 32, 37, 39。現在該校某一位求知若渴的學生希望引用「每5個取1個」的系統抽樣估計榕樹的比例。

(一)請提供所有「每5個取1個」之系統抽樣的樣本編號。(5分)

(二)請提供題(一)所抽中之樣本榕樹比例的抽樣分配。(10分)

(三)請證明樣本榕樹比例確實是真實榕樹比例的不偏估計量。(10分)

試題評析	本題是考系統抽樣法，這類給定母體資訊下，討論樣本估計量之不偏性，在考古題中也常常出現，只要有練習過考古題，拿到分數並不難。
考點命中	《高點抽樣方法講義》第一回，趙治勳編撰，第三章。

答：

(一)

開始編號	系統樣本編號	第i組系統樣本
0	00,05,10,15,20,25,30,35,40	1
1	01,06,11,16,21,26,31,36,41	2
2	02,07,12,17,22,27,32,37,42	3
3	03,08,13,18,23,28,33,38,43	4
4	04,09,14,19,24,29,34,39,44	5

(二)

第i組系統樣本	系統樣本編號 * ☐ 表該樹是榕樹	樣本榕樹比例
1	00, ☐ 05, ☐ 10, ☐ 15, 20, ☐ 25, ☐ 30, 35, 40	5/9
2	01, 06, 11, 16, 21, 26, ☐ 31, 36, 41	1/9
3	02, ☐ 07, 12, ☐ 17, 22, ☐ 27, ☐ 32, ☐ 37, 42	5/9
4	03, 08, 13, 18, 23, ☐ 28, 33, 38, 43	1/9
5	☐ 04, 09, 14, ☐ 19, 24, ☐ 29, 34, ☐ 39, 44	4/9

故樣本榕樹比例 p 之抽樣分配如下：

p	1/9	4/9	5/9
機率	2/5	1/5	2/5

(三)母體榕樹比例 $P = \frac{16}{45}$

$$E(p) = \frac{1}{9} \times \frac{2}{5} + \frac{4}{9} \times \frac{1}{5} + \frac{5}{9} \times \frac{2}{5} = \frac{16}{45} = P$$

因為 $E(p) = P$ ，故樣本榕樹比例 p 確實是母體榕樹比例之不偏估計量

三、以下資料為紐約與倫敦兩城市的日均溫紀錄（以華氏計）：

都市	紐約	倫敦	紐約	紐約	倫敦	紐約	紐約
氣溫	48	54	52	47	57	54	49
都市	倫敦	紐約	紐約	紐約	紐約	倫敦	紐約
氣溫	59	53	50	52	57	55	54

【版權所有，重製必究！】

都市	倫敦	紐約	紐約	倫敦	紐約	紐約	倫敦
氣溫	68	49	51	61	55	53	50

請根據上述表格回答以下的問題。

- (一)如果挑選一組「4個單元」的「取後不放回」隨機樣本「48, 47, 57, 52」請提供這一項抽樣設計之母體平均數的95%信賴區間。(10分)
- (二)如果用地理位置把題意的母體切成「兩層」，然後每一層挑選一組「2個單元」的「取後不放回」隨機樣本，「52, 57」來自紐約，以及「57, 55」來自倫敦，請提供這一項抽樣設計之母體平均數的95%信賴區間。(10分)
- (三)計算「題(二)的抽樣設計」比上「題(一)的抽樣設計」的相對有效性。(5分)

試題評析	本題是SRS與分層隨機抽樣法之比較題，屬於常考之題型，要拿到分數並不難
考點命中	《高點抽樣方法講義》第一回，趙治勳編撰，第五章。

答：

$$(一) \bar{y} = \frac{y}{n} = \frac{48 + 47 + 57 + 52}{4} = 51$$

$$s_{\bar{y}} = \sqrt{(1-f) \frac{s^2}{n}} = \sqrt{(1-\frac{4}{21}) \frac{20.667}{4}} = 2.045 \quad (\text{假設母體資料未知})$$

母體平均數 $\bar{Y}$ 之 95% 近似信賴區間為

$$(\bar{y} \pm 1.96 s_{\bar{y}}) = (51 \pm 1.96 \times 2.045) = (46.992, 55.008)$$

$$(二) \bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h = \frac{14}{21} \times 54.5 + \frac{7}{21} \times 56 = 55$$

$$s_{\bar{y}_{st}} = \frac{1}{N} \sqrt{\sum_{h=1}^L N_h (N_h - n_h) \frac{s_h^2}{n_h}} = \frac{1}{21} \sqrt{14(14-2) \frac{12.5}{2} + 7(7-2) \frac{2}{2}}$$

$$= 1.569 \quad (\text{假設母體資料未知})$$

母體平均數 $\bar{Y}$ 之 95% 近似信賴區間為

$$(\bar{y}_{st} \pm 1.96 s_{\bar{y}_{st}}) = (55 \pm 1.96 \times 1.569) = (51.925, 58.075)$$

(三)因為母體資料已給，在樣本平均數之變異數中以母體變異數代入計算

$$V(\bar{y}) = (1-f) \frac{S^2}{n} = (1-\frac{4}{21}) \frac{23.714}{4} = 4.799$$

$$V(\bar{y}_{st}) = \frac{1}{N^2} \sum_{h=1}^L N_h (N_h - n_h) \frac{S_h^2}{n_h} = \frac{1}{21^2} [14(14-2) \frac{8.22}{2} + 7(7-2) \frac{33.238}{2}] = 2.885$$

$$\text{故相對有效性} \frac{V(\bar{y})}{V(\bar{y}_{st})} = \frac{4.799}{2.885} = 1.663 > 1$$

四、假設樣本來自某一個簡單隨機抽樣設計，而且樣本數與母體都夠大。已知以下三種母體平均數之估計量的變異數：

【版權所有，重製必究！】

$$(E1) \quad \text{Var}(\bar{Y}) \frac{1}{n} \frac{N-n}{N-1} \sigma_y^2$$

$$(E2) \quad \text{Var}(\bar{Y}_{ratio}) \frac{1}{n} \frac{N-n}{N-1} [\sigma_y^2 + \mu_r^2 \sigma_x^2 - 2\mu_r \rho_{xy} \sigma_x \sigma_y]$$

$$(E3) \quad \text{Var}(\bar{Y}_{regression}) \frac{1}{n} \frac{N-n}{N-1} [1 - \rho_{xy}^2] \sigma_y^2$$

其中 N 是母體規模 (population size)， n 是樣本規模 (sample size)， σ_y^2 是母體 Y 的變異數， σ_x^2 是母體 X 的變異數， ρ_{xy} 是母體 X 跟母體 Y 之間的相關係數， $\mu_r = \mu_y / \mu_x$ 是母體 Y 的母體平均數除以母體 X 的母體平均數。

請證明以下比例的正確性並且下結論：

(一) $(E1)/(E3) \geq 1$ 。(10分)

(二) $(E2)/(E3) \geq 1$ 。(10分)

(三)根據上述兩項結果，請問 $(E1)$ ， $(E2)$ 及 $(E3)$ 那一種估計量可以被認為「比較好」。(5分)

試題評析	本題是SRS，比率簡單與迴歸簡單之精確度比較，上課講義中已經有完全相同之證明，要拿到滿分並不難。
考點命中	《高點抽樣方法講義》第一回，趙治勳編撰，頁91。

答：

$$(一) \frac{E1}{E3} = \frac{V(\bar{Y})}{V(\bar{Y}_{reg})} = \frac{\frac{1}{n} \frac{N-n}{N-1} \sigma_y^2}{\frac{1}{n} \frac{N-n}{N-1} (1 - \rho_{xy}^2) \sigma_y^2} = \frac{1}{1 - \rho_{xy}^2} \geq 1$$

($\because 0 \leq \rho_{xy}^2 < 1$) 得證

$$(二) \text{令 } \frac{E2}{E3} = \frac{V(\bar{Y}_r)}{V(\bar{Y}_{reg})} = \frac{\frac{1}{n} \frac{N-n}{N-1} (\sigma_y^2 + \mu_r^2 \sigma_x^2 - 2\mu_r \rho_{xy} \sigma_x \sigma_y)}{\frac{1}{n} \frac{N-n}{N-1} (1 - \rho_{xy}^2) \sigma_y^2}$$

$$= \frac{\sigma_y^2 + \mu_r^2 \sigma_x^2 - 2\mu_r \rho_{xy} \sigma_x \sigma_y}{(1 - \rho_{xy}^2) \sigma_y^2} \geq 1$$

$$\Rightarrow \sigma_y^2 + \mu_r^2 \sigma_x^2 - 2\mu_r \rho_{xy} \sigma_x \sigma_y \geq (1 - \rho_{xy}^2) \sigma_y^2$$

$$\Rightarrow \sigma_y^2 + \mu_r^2 \sigma_x^2 - 2\mu_r \rho_{xy} \sigma_x \sigma_y - \sigma_y^2 + \rho_{xy}^2 \sigma_y^2 \geq 0$$

$$\Rightarrow \mu_r^2 \sigma_x^2 - 2\mu_r \rho_{xy} \sigma_x \sigma_y + \rho_{xy}^2 \sigma_y^2 \geq 0$$

【版權所有，重製必究！】

$$\Rightarrow (\mu_r \sigma_x)^2 - 2(\mu_r \sigma_x)(\rho_{xy} \sigma_y) + (\rho_{xy} \sigma_y)^2 \geq 0$$

$$\Rightarrow (\mu_r \sigma_x - \rho_{xy} \sigma_y)^2 \geq 0 \text{ 成立，故得證}$$

(三)由(一)(二)得知，E3之精確度最高，故E3被認為「比較好」。

高
點
·
高
上

【版權所有，重製必究！】