

104年公務人員高考考試試題

類 科：統計

科 目：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. 數異數取三位小數（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。

(二)請提供上述總報案次數之變異數的不偏估計值。

(三)請提供總報案次數之近似的95%信賴區間。

【擬答】：

$$(一) \bar{y} = \frac{0 \times 15 + 1 \times 10 + \dots + 4 \times 1}{30} = \frac{11}{30} = 0.73$$

$$\hat{Y} = N\bar{y} = 365 \times 0.7333 = 267.667 \xrightarrow{\text{估計}} Y$$

$$(二) s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} = 0.892$$

$$\text{var}(\hat{Y}) = N^2(1-f) \frac{s^2}{n} \approx N^2 \frac{s^2}{n} = 365^2 \frac{0.892}{30} = 3961.019$$

(三)總報案次數之95%信賴區為

$$\hat{Y} \pm Z \cdot s_{\hat{Y}} \Rightarrow 267.667 \pm 1.96 \times \sqrt{3961.019}$$

$$\Rightarrow (144.31, 391.02)$$

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。

(二)請提供題(一)所抽中之樣本榕樹比例的抽樣分配。

(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。

【擬答】：

公職王歷屆試題 (104 高考)

(一)每5個取1個，可得5個系統，將原資料整理如下

	系統 1	系統 2	系統 3	系統 4	系統 5
	00	01	02	03	<u>04</u>
	<u>05</u>	06	<u>07</u>	08	09
	<u>10</u>	11	12	13	14
	<u>15</u>	16	<u>17</u>	18	<u>19</u>
編號	20	21	22	23	24
	<u>25</u>	26	<u>27</u>	<u>28</u>	<u>29</u>
	<u>30</u>	<u>31</u>	<u>32</u>	33	34
	35	36	<u>37</u>	38	<u>39</u>
	40	41	42	43	44

(二)系統抽樣 p_{sy} 之抽樣分配如下

p_{sy}	$\frac{1}{9}$	$\frac{4}{9}$	$\frac{5}{9}$
$f(p_{sy})$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$

$$(三) E(p_{sy}) = \sum_{i=1}^k p_{sy} \frac{1}{k} = \frac{1}{9} \times \frac{2}{5} + \frac{4}{9} \times \frac{1}{5} + \frac{5}{9} \times \frac{2}{5} = \frac{16}{45} = P = \frac{16}{45}$$

所以 P_{sy} 為母體比例值 P 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
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。

(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。

(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。

【擬答】：

$$(一) \bar{y} = \frac{48 + 47 + 57 + 52}{4} = 51 \xrightarrow{\text{估計}} \bar{Y}$$

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} = 20.6667$$

$$s_{\bar{y}} = \sqrt{(1-f) \frac{s^2}{n}} = \sqrt{(1-\frac{4}{21}) \frac{20.6667}{4}} = 2.0451$$

日均溫 μ 之95%信賴區間為

$$\bar{y} \pm Z \cdot s_{\bar{y}} \Rightarrow 51 \pm 1.96 \cdot 2.0451 \Rightarrow (46.99, 55.01)$$

$$(\Rightarrow) \bar{y}_{st} = \sum_{h=1}^2 W_h \bar{y}_h = \frac{14}{21} \times \frac{52+57}{2} + \frac{7}{21} \times \frac{57+55}{2} = 55 \xrightarrow{\text{估計}} \bar{Y}$$

$$\begin{aligned} s_{\bar{y}_{st}} &= \frac{1}{N} \sqrt{\sum N_h (N_h - n_h) \frac{s_h^2}{n_h}} \\ &= \frac{1}{21} \sqrt{14(14-2) \frac{12.5}{2} + 7(7-2) \frac{2}{2}} \\ &= \frac{1}{21} \sqrt{1085} = 1.5685 \end{aligned}$$

日均溫 μ 之 95% 信賴區間為

$$\bar{y}_{st} \pm Z \cdot s_{\bar{y}_{st}} \Rightarrow 55 \pm 1.96 \cdot 1.5685 \Rightarrow (51.93, 58.07)$$

(三) 簡單隨機抽樣與分層隨機抽樣皆為不偏估計量

所以相對有效性可直接比較變異數

$$RE = \frac{\text{var}_2(\bar{y}_{st})}{\text{var}_1(\bar{y})} = \frac{1.5685^2}{2.0451^2} = 0.588 < 1$$

相對效率小於 1，代表題(二)分層隨機抽樣優於題(一)簡單隨機抽樣

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：

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

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

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

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， $\mu_r = \mu_r / \mu_x$ 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确性並且下結論：

$$(\Rightarrow) (E1)/(E3) \geq 1。$$

$$(\Rightarrow) (E2)/(E3) \geq 1。$$

(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」。

【擬答】：

$$(\Rightarrow) (E1)/(E3)$$

$$\frac{\text{Var}(\bar{y})}{\text{Var}(\bar{y}_{reg})} = \frac{\frac{1}{n} \frac{N-n}{N-1} \sigma_y^2}{\frac{1}{n} \frac{N-n}{N-1} [1 - \rho_{xy}^2] \sigma_y^2} = \frac{1}{1 - \rho_{xy}^2} \geq 1$$

$$\text{因為 } \rho_{xy}^2 \leq 1 \Rightarrow 0 \leq 1 - \rho_{xy}^2 \leq 1$$

$$(\Rightarrow) (E2)/(E3)$$

$$\frac{\text{Var}(\bar{y}_r)}{\text{Var}(\bar{y}_{reg})} = \frac{\frac{1}{n} \frac{N-n}{N-1} [\sigma_y^2 + \mu_r^2 \sigma_x^2 - 2\mu_r \rho_{xy} \sigma_x \sigma_y]}{\frac{1}{n} \frac{N-n}{N-1} [1 - \rho_{xy}^2] \sigma_y^2}$$

$$\begin{aligned} & 1 + \frac{\bar{Y}^2}{\bar{X}^2} \frac{\sigma_x^2}{\sigma_y^2} - 2 \frac{\bar{Y}}{\bar{X}} \frac{\sigma_x}{\sigma_y} \rho_{xy} = 1 + \frac{CV_y^2}{CV_x^2} - 2 \frac{CV_y}{CV_x} \rho_{xy} \\ & = \frac{1 + \frac{\bar{Y}^2}{\bar{X}^2} \frac{\sigma_x^2}{\sigma_y^2} - 2 \frac{\bar{Y}}{\bar{X}} \frac{\sigma_x}{\sigma_y} \rho_{xy}}{1 - \rho_{xy}^2} = \frac{1 + \frac{CV_y^2}{CV_x^2} - 2 \frac{CV_y}{CV_x} \rho_{xy}}{1 - \rho_{xy}^2} \\ & = \frac{\left(\frac{CV_y}{CV_x} - \rho_{xy} \right)^2 + 1 - \rho_{xy}^2}{1 - \rho_{xy}^2} \\ & = 1 + \left(\frac{CV_y}{CV_x} - \rho_{xy} \right)^2 \geq 1 \end{aligned}$$

(三)由(一)簡單隨機抽樣之變異數大於迴歸估計式之變異數
故迴歸估計優於簡單隨機抽樣；

由(二)比率估計式之變異數大於迴歸估計式之變異數
故迴歸估計優於比率估計式。

所以綜合(一)(二)可知，迴歸估計式相對比較好