111 年特種考試地方政府公務人員考試試題

等　別：三等考試
類　科：統計
科　目：抽樣方法

王瑋老師

甲、申論題部分：

一、欲針對園區內 5,000 個租屋的上班族進行抽樣調查，以了解其租屋支出及社會住宅需求。根據過去幾年的研究報告，月租屋支出之標準差為 S=30（千元），社會住宅需求比例（P）介於 35%到 45%之間。若採用簡單隨機抽樣方式，欲在 95%的信賴水準下估計月平均租屋支出（$\bar{Y}$）及社會住宅需求比例（P），同時達到下列估計精確度的要求，則需多少樣本數？（10 分）

$P(|p-P|\le 0.05)=0.95$，$P(|\bar{y}-\bar{Y}|\le 3)=0.95$

式中 p 及 $\bar{y}$ 分別為 P 及 $\bar{Y}$ 的估計量。

【解題關鍵】
1.《考題難易》★☆☆☆☆
2.《破題關鍵》同時滿足比例值與平均數估計的樣本數要求，在 111 年關務抽樣與 109 年高考皆有命題，需要留意題目所提供的抽樣誤差數值為何。
3.《命中特區》王瑋，抽樣方法，志光出版，頁 2-61 題 37 與頁 2-61 題 38。

【擬答】

㈠以社會住宅需求比例估計樣本數

$$n_0=\frac{Z^2\cdot PQ}{B^2}=\frac{1.96^2\times 0.45\times 0.55}{0.05^2}=380.3\fallingdotseq 381$$

$$n=\frac{n_0}{1+\frac{n_0}{N}}=\frac{381}{1+\frac{381}{5000}}=354.02$$，n 取 355

㈡以平均月租屋支出估計樣本數

$$n_0=\frac{Z^2\cdot S^2}{B^2}=\frac{Z^2\cdot S^2}{3^2}=\frac{1.96^2\times 30^2}{3^2}=384.16\fallingdotseq 385$$

$$n=\frac{n_0}{1+\frac{n_0}{N}}=\frac{385}{1+\frac{385}{5000}}=357.5$$，n 取 358

若要滿足上述兩者抽樣方式，需抽出樣本數 358

二、欲了解醫院的醫療採購情形，由 120 家醫院抽樣調查 30 家醫院，取得下列（x, y）的樣本統計資訊。x：登記的病床數為輔助變數（單位：床），y：醫療採購金額為興趣變數（單位：萬元）：

$\sum_{i=1}^{30}x_i=22,500$，$\sum_{i=1}^{30}y_i=19,500$，$s_x^2=360,000$，$s_y^2=250,000$，$s_{xy}=240,000$。

已知 120 家醫院登記的總病床數為 96,000。

㈠用比率估計量（ratio estimator, $\bar{y}_R$）及迴歸估計量（regression estimator, $\bar{y}_{lr}$）估計平均一家醫院之醫療採購金額（$\bar{Y}$）。（10 分）

㈡求算前述兩個估計量相對於單位均數估計量（mean per unit estimator, $\bar{y}$）的相對效率

（relative efficiency）並比較分析其精確度（precision）。（10 分）

（註：答案如有小數，請計算至第二位）

【解題關鍵】

1.《考題難易》★☆☆☆☆

2.《破題關鍵》迴歸估計、比率估計與簡單隨機抽樣比較是多年的考古題，本題很明顯是延續 111 年關務抽樣的命題而來，而 110 年關務亦有類似考題。

3.《命中特區》王瑋，抽樣方法，志光出版，頁 9-23 題 8~頁 9-28 題 10。

【擬答】

㈠比率估計量

$$\bar{y}_R = r\cdot\bar{X} = \frac{19500}{22500}\times\frac{96000}{120} = 693.33 \xrightarrow{\text{估計}} \bar{Y}$$

迴歸估計量

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{(n-1)s_{xy}}{(n-1)s_x^2} = \frac{240000}{360000} = \frac{2}{3}$$

$$\bar{y}_{lr} = \bar{y} + b_1(\bar{X}-\bar{x}) = \frac{19500}{30} + \frac{2}{3}\cdot(800-\frac{22500}{30}) = 683.33 \xrightarrow{\text{估計}} \bar{Y}$$

㈡簡單均數估計量變異數

$$\text{var}(\bar{y}) = (1-f)\cdot\frac{s_y^2}{n} = (1-\frac{30}{120})\cdot\frac{250000}{30} = 6250$$

比率估計量變異數

$$s_d^2 = s_y^2 - 2r\cdot s_{xy} + r^2 s_x^2$$

$$= 250000 - 2\times\frac{195}{225}\times 240000 + (\frac{195}{225})^2\cdot 360000 = 104400$$

$$\text{var}(\bar{y}_R) = (1-f)\cdot\frac{s_d^2}{n} = (1-\frac{30}{120})\cdot\frac{104400}{30} = 2610$$

迴歸估計量變異數

$$s_{lr}^2 = \frac{SS_y - b_1^2\cdot SS_x}{n-1} = s_y^2 - b_1^2 s_x^2$$

$$= 250000 - (\frac{2}{3})^2\times 360000 = 90000$$

$$\text{var}(\bar{y}_{lr}) = (1-f)\cdot\frac{s_{lr}^2}{n} = (1-\frac{30}{120})\cdot\frac{90000}{30} = 2250$$

$$RE_1 = \frac{\text{var}(\bar{y}_R)}{\text{var}(\bar{y})} = \frac{2610}{6250} = 0.42 < 1$$

$$RE_2 = \frac{\text{var}(\bar{y}_{lr})}{\text{var}(\bar{y})} = \frac{2250}{6250} = 0.36 < 1$$

比率估計量與迴歸估計量的相對效率皆小於 1，代表兩者皆較簡單均數估計量精確，而迴歸估計量的相對效率又比比率估計量小，所以代表迴歸估計量最精確。

三、稽核單位欲了解申報資料的正確性，將申報資料共 10,000 件依其營業特性別分成四層，每層的案件數分別為 $N_1=1,000$；$N_2=3,000$；$N_3=4,000$；$N_4=2,000$，再由每層抽取一簡單隨機樣本 $n_1=10$；$n_2=30$；$n_3=40$；$n_4=20$ 進行查核。查核結果彙整如下表：

層別	1	2	3	4
短漏報件數	5	6	8	2
平均短漏報金額（萬元）	15	10	8	4
短漏報金額標準差（萬元）	40	30	50	10

㈠配合前述抽樣設計，回答以下問題：（20 分）

1. 估計平均短漏報金額（$\bar{Y}$）及該估計量之標準誤。
2. 估計短漏報比例（P）及該估計量之標準誤。

㈡假設前述資訊已知，原樣本配置方式改為何種配置方式將可提高估計精確度？（5 分）

㈢依㈡所建議之配置方式回答以下問題：（10 分）

1. 若針對估計均數 $\bar{Y}$，如何配置樣本 n=100 到各層？
2. 若針對估計比例 P，如何配置樣本 n=100 到各層？

（註：答案如有小數，請計算至第二位）

【解題關鍵】

1.《考題難易》★★☆☆☆

2.《破題關鍵》分層隨機抽樣與樣本最適配置是考試重要內容，本題是首次將分層平均值估計與比例值估計的問題結合在同一個題目中，過去僅有 108 年薦任抽樣有出過類似的概念。

3.《命中特區》王瑋，抽樣方法，志光出版，頁 3-12 題 5、頁 3-73 題 33 與頁 3-77 題 36。

【擬答】

㈠

1. $\bar{y}_{st}=\sum_{h=1}^{4}W_h\bar{y}_h=\frac{1000}{10000}\times15+\frac{3000}{10000}\times10+\frac{4000}{10000}\times8+\frac{2000}{10000}\times4=8.5$

$$s_{\bar{y}_{st}}=\frac{1}{N}\sqrt{\sum N_h(N_h-n_h)\frac{s_h^2}{n_h}}$$

$$=\frac{1}{10000}\sqrt{1000(1000-10)\frac{40^2}{10}+\ldots+2000(2000-20)\frac{10^2}{20}}$$

$$=3.79$$

2. $p_{st}=\sum_{h=1}^{3}W_hp_h=\frac{1000}{10000}\times\frac{5}{10}+\frac{3000}{10000}\times\frac{6}{30}+\ldots+\frac{2000}{10000}\times\frac{2}{20}=0.21$

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

$$=\frac{1}{10000}\sqrt{1000(1000-10)\frac{0.5\times0.5}{10-1}+\ldots+2000(2000-20)\frac{0.1\times0.9}{20-1}}$$

$$=0.04$$

㈡原配置為比例配置，因為不論平均數或比例值的估計，其分層標準差皆已知，所以可採用紐門配置來提高估計精確度。

㈢1.採用紐門配置，各層樣本數為 $n_h=\frac{N_hs_h}{\sum_{h=1}^{L}N_hs_h}\times n$

$$\sum N_hS_h==1000\times40+3000\times30+4000\times50+2000\times10=350000$$

$n_1=100\times\frac{1000\times40}{350000}=11.43$，$n_1$ 取 11

$n_2=100\times\frac{3000\times30}{350000}=25.71$，$n_2$ 取 26

$n_3=100\times\frac{4000\times50}{350000}=57.14$，$n_3$ 取 57

$n_4=100\times\frac{2000\times10}{350000}=5.71$，$n_4$ 取 6

2.採用紐門配置，$n_h=\frac{N_hs_h}{\sum_{h=1}^{L}N_hs_h}\times n$，其中 $s_h=\sqrt{\frac{n_hp_hq_h}{n_h-1}}\approx\sqrt{p_hq_h}$

$$\sum_{h=1}^{L}N_hs_h\approx\sum_{h=1}^{L}N_h\sqrt{p_hq_h}$$

$$=1000\sqrt{0.5\times0.5}+3000\sqrt{0.2\times0.8}+\ldots+2000\sqrt{0.1\times0.9}=3900$$

$$=3900$$

$$n_1=\frac{1000\cdot\sqrt{0.5\times0.5}}{3900}\times100=12.82\text{，}n_1\text{取 }13$$

$$n_2=\frac{3000\cdot\sqrt{0.2\times0.8}}{3900}\times100=30.77\text{，}n_2\text{取 }31$$

$$n_3=\frac{4000\cdot\sqrt{0.2\times0.8}}{3900}\times100=41.03\text{，}n_3\text{取 }41$$

$$n_4=\frac{2000\cdot\sqrt{0.1\times0.9}}{3900}\times100=15.38\text{，}n_4\text{取 }15$$

公職

四、欲調查某國家公園內的樹感染病蟲害的情況，該國家公園內共分成 10 個區域，首先由 10 個區域隨機抽出 4 個區域，再就被抽出的區域隨機抽出 5%的樹進行調查，調查結果如下表：

區域	樹種植數量 M_i	樹抽出數量 m_i	樹感染病蟲害數量 a_i
1	200	10	1
2	300	15	5
3	300	15	5
4	400	20	4

㈠說明此抽樣設計的名稱及兩階段的抽樣單位（Primary Sampling Units 及 Secondary Sampling Units）。（15 分）

㈡試估計該國家公園內的樹感染病蟲害的比例（P），並提供 95%信賴區間。（20 分）（註：答案如有小數，請計算至第二位）

【解題關鍵】

1.《考題難易》★★☆☆☆

2.《破題關鍵》兩階段群集抽樣法雖屬課內基本內容，但過去較少詢問抽樣單位，是需要留意的地方。而本題要估計比例值，但總數未知，故應該採用二階段的比率群集估計，可參考 109 年關務抽樣與 108 年地特抽樣考題。

3.《命中特區》王瑋，抽樣方法，志光出版，頁 12-44 題 19。

【擬答】

㈠此為兩階段群集抽樣法，首先第一階段先從 10 個群集(區域)隨機抽出 4 個群集(區域)，第二階段再由被抽到的群集中，隨機抽出 5%的樹木成為最終的樣本。因此第一階段的抽樣單位是區域，而第二階段的抽樣單位則是樹。

㈡估計比例值，但 M 未知，採用二階段比率群集估計法

$$\hat{M} = \bar{M} \times N = \frac{200+300+300+400}{4} \times 10 = 300 \times 10 = 3000$$

$$p_{rc} = \frac{\bar{a}_t}{\bar{M}} = \frac{(200\times 0.1 + \ldots + 400 \times 0.2)/4}{300} = \frac{75}{300} = 0.25 \xrightarrow{\text{估計}} P$$

$$\text{var}(\bar{y}_{rc}) = \frac{1}{\hat{M}^2}\left(N^2(1-f_1)\frac{s_{1c}^2}{n} + \frac{N}{n}\sum_{i=1}^{n} M_i^2 (1-f_{2i})\frac{s_{2i}^2}{m_i} \right)$$

$$s_{1c}^2 = \frac{\sum_{i=1}^{n}(M_i p_i - M_i p_{rc})^2}{n-1}$$

$$= \frac{\sum_{i=1}^{n} M_i^2 p_i^2 - 2p_{rc}\sum_{i=1}^{n} M_i^2 p_i + p_{rc}^2 \sum_{i=1}^{n} M_i^2}{n-1}$$

$$= \frac{26800 - 2\times 0.25 \times 96000 + 0.25^2 \times 380000}{4-1} = 850$$

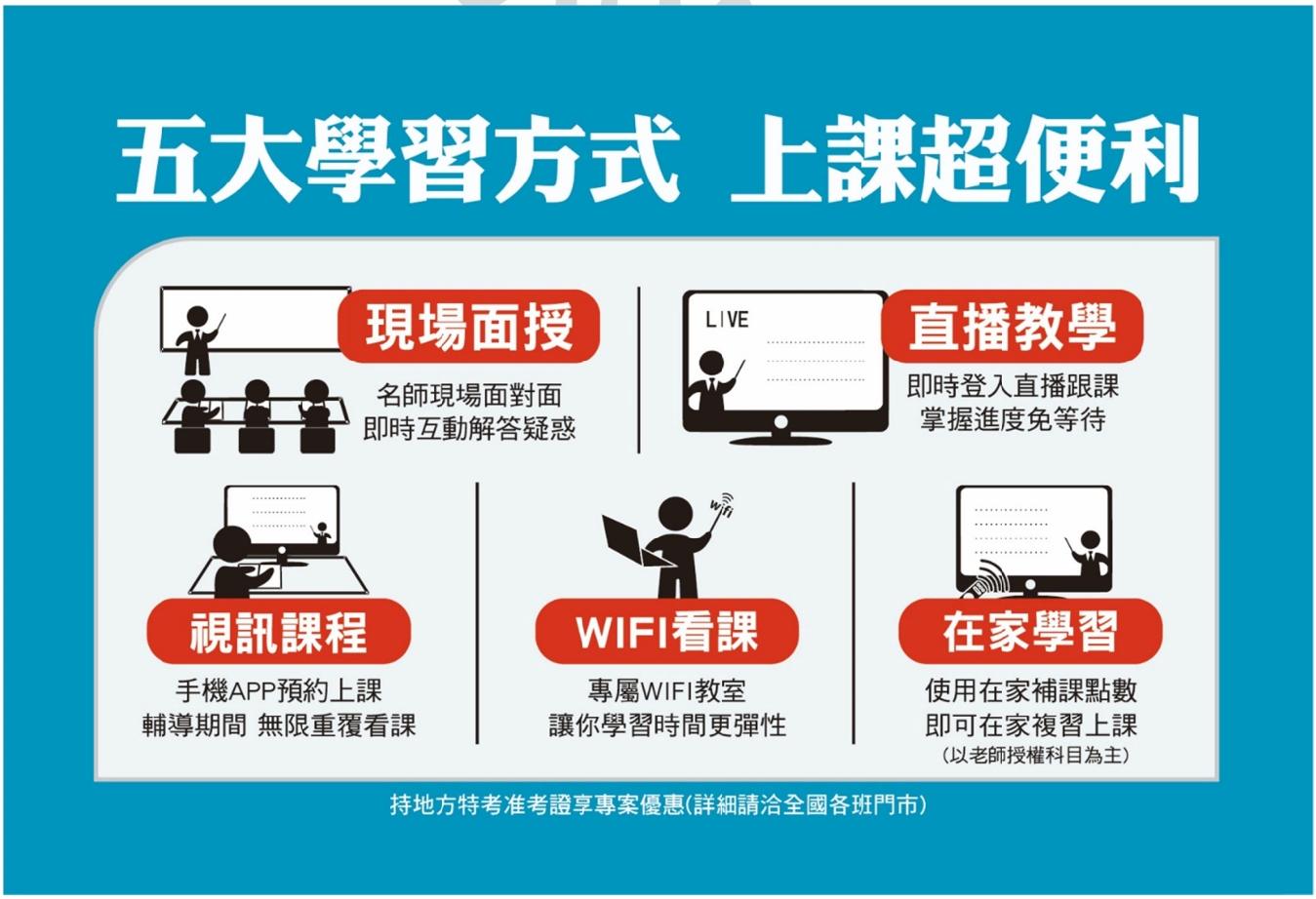
$$\sum_{i=1}^{n} M_i^2 (1-f_{2i})\frac{s_{2i}^2}{m_i}$$

$$= 200^2(1-\frac{10}{200})\frac{\frac{10}{10-1}\cdot 0.1\cdot 0.9}{10} + \cdots + 400^2(1-\frac{20}{400})\frac{\frac{20}{20-1}\cdot 0.2\cdot 0.8}{20}$$

$$= 4374.2857$$

故 $s_{p_{rc}} = \frac{1}{3000}\sqrt{10^2(1-\frac{4}{10})\frac{850}{4}+\frac{10}{4}\times 4374.2857} = 0.0513$

故該國家公園內的樹感染病蟲害的比例 95%信賴區間為

$0.4167 \pm 1.96 \times 0.2297 \Rightarrow (0.03, 0.87)$

公職