《抽樣方法》

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（$\overline{Y}$）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）

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

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

試題評析	本題屬於簡單隨機抽樣法之樣本數計算題型，考古題中也有出現過，本題需要同時滿足兩種要求的計算方式，獲得滿分不難。
考點命中	《抽樣方法申論題完全制霸》，高點文化出版，趙治勳編著，第一章範題39，頁1-38。

答：

1.估計$\overline{Y}$

$$n_0=\frac{z_{\frac{\alpha}{2}}^2S^2}{B^2}=\frac{1.96^2 30^2}{3^2}=384.16$$

$$n=\frac{n_0}{1+\frac{n_0}{N}}=356.7502\approx 357$$

2.估計P

$$n_0=\frac{z_{\frac{\alpha}{2}}^2PQ}{B^2}=\frac{1.96^2(0.45)(0.55)}{0.05^2}=380.3184$$

$$n=\frac{n_0}{1+\frac{n_0}{N}}=353.4348\approx 354$$

故要同時滿足以上1. 2.，n取357個。

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）：

$$\sum_{1}^{30}x=22{,}500，\sum_{1}^{30}y=19{,}500，S_x^2=360{,}000，S_y^2=250{,}000，S_{xy}=240{,}000$$

已知120家醫院登記的總病床數為96,000。

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

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

試題評析	本題屬於簡單隨機抽樣法利用兩種不同的估計法進行參數估計，考題敘述跟98年地特與101地特雷同，考生有勤做考古題的話，獲得滿分並不難。
考點命中	《抽樣方法申論題完全制霸》，高點文化出版，趙治勳編著，第十三章範題8，頁13-13。

答：

(一) 1.比率

$$r=\frac{y}{x}=\frac{19500}{22500}$$

$$\bar{y}_R=r\bar{X}=\frac{19500}{22500}\times\frac{96000}{120}=693.33\quad(\text{萬元})$$

2.迴歸

$$\bar{y}_{lr}=\bar{y}+b_1(\bar{X}-\bar{x})=\frac{19500}{30}+\frac{240000}{360000}\left(\frac{96000}{120}-\frac{22500}{30}\right)=683.33\quad(\text{萬元})$$

其中 $b_1=\dfrac{s_{XY}}{s_X^2}=\dfrac{240000}{360000}$

(二) 單位均數估計量=簡單均數估計量

$$s_{\bar{y}}=\sqrt{(1-f)\frac{s_Y^2}{n}}=\sqrt{(1-\frac{30}{120})\frac{250000}{30}}=79.06\qquad(\text{萬元})$$

比率

$$s_r=\frac{1}{\bar{X}}\sqrt{(1-f)\frac{s_d^2}{n}}=\frac{1}{\frac{96000}{120}}\sqrt{(1-\frac{30}{120})\frac{104400}{30}}=0.06386$$

其中 $s_d^2=s_Y^2+r^2s_X^2-2rs_{XY}$

$$=250000+(\frac{19500}{22500})^2\times360000-2\times\frac{19500}{22500}\times240000=104400$$

$$s_{\bar{y}_R}=\bar{X}s_r=\frac{96000}{120}\times0.06386=51.09\quad(\text{萬元})$$

迴歸

$$s_{\bar{y}_{lr}}=\sqrt{(1-f)\frac{MSE}{n}}=\sqrt{(1-\frac{30}{120})\frac{93214.2857}{30}}=48.27$$

其中 $MSE=\dfrac{(n-1)s_Y^2-b_1^2(n-1)s_X^2}{n-2}=\dfrac{(30-1)250000-(\frac{240000}{360000})^2(30-1)360000}{30-2}$

$=93214.2857$

相對效率：$\dfrac{\text{比率}}{\text{簡單}}=\dfrac{s_{\bar{y}_R}^2}{s_{\bar{y}}^2}=\dfrac{51.09^2}{79.06^2}=0.42$ 與 $\dfrac{\text{迴歸}}{\text{簡單}}=\dfrac{s_{\bar{y}_{lr}}^2}{s_{\bar{y}}^2}=\dfrac{48.27^2}{79.06^2}=0.37$

由以上相對效率得知，比率估計量與迴歸估計量均比單位均數估計量次精確度為高。

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

(一)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（$\overline{Y}$）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。

(二)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）

(三)依(二)所建議之配置方式回答以下問題：（10分）

1.若針對估計均數$\overline{Y}$，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）

試題評析	本題屬於分層隨機抽樣法下參數估計與樣本配置之計算題型，考古題常命題，獲得滿分不難。
考點命中	《抽樣方法申論題完全制霸》，高點文化出版，趙治勳編著，第四章，頁34-37。

答：

h	N_h	n_h	a_h	$\overline{y}_h$	s_h	$p_h=\dfrac{a_h}{n_h}$
1	1000	10	5	15	40	0.5
2	3000	30	6	10	30	0.2
3	4000	40	8	8	50	0.2
4	2000	20	2	4	10	0.1

(一) 1. $$\overline{y}_{st}=\sum_{h=1}^{L}W_h\overline{y}_h=\frac{1000}{10000}15+\frac{3000}{10000}10+\frac{4000}{10000}8+\frac{2000}{10000}4=8.5 \quad (萬元)$$

$$s_{\overline{y}_{st}}=\frac{1}{N}\sqrt{\sum_{h=1}^{L}N_h(N_h-n_h)\frac{s_h^2}{n_h}}$$

$$=\frac{1}{10000}\sqrt{\begin{array}{l}1000(1000-10)\dfrac{40^2}{10}+3000(3000-30)\dfrac{30^2}{30}\\ +4000(4000-40)\dfrac{50^2}{40}+2000(2000-20)\dfrac{10^2}{20}\end{array}}$$

$$=3.79 \quad (萬元)$$

2. $$p_{st}=\sum_{h=1}^{L}W_hp_h=\frac{1000}{10000}0.5+\frac{3000}{10000}0.2+\frac{4000}{10000}0.2+\frac{2000}{10000}0.1=0.21$$

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

$$=\frac{1}{10000}\sqrt{\begin{array}{l}1000(1000-10)\dfrac{\frac{10}{10-1}(0.5)(0.5)}{10}\\+3000(3000-30)\dfrac{\frac{30}{30-1}(0.2)(0.8)}{30}\\+4000(4000-40)\dfrac{\frac{40}{40-1}(0.2)(0.8)}{40}\\+2000(2000-20)\dfrac{\frac{20}{20-1}(0.1)(0.9)}{20}\end{array}}=0.04$$

(二) 建議採用紐門配置提高估計精確度

(三) 紐門配置：$n_h=\dfrac{N_hS_h}{\sum_{h=1}^{L}N_hS_h}\times n$

1. 估計$\overline{Y}$

S_h未知,以s_h代

$$\sum_{h=1}^{L}N_hS_h=1000\times40+3000\times30+4000\times50+2000\times10=350000$$

$$n_1=\frac{1000\times40}{350000}\times100=11.4286\approx11$$

$$n_2=\frac{3000\times30}{350000}\times100=25.7143\approx26$$

$$n_3=\frac{4000\times50}{350000}\times100=57.1429\approx57$$

$$n_4=n-n_1-n_2-n_3=6$$

2. 估計P

S_h未知,以$s_h=\sqrt{\frac{n_h}{n_h-1}p_hq_h}$代

$$\sum_{h=1}^{L}N_hS_h=1000\times0.527+3000\times0.4068+4000\times0.4051+2000\times0.3078=3983.4$$

$$n_1=\frac{1000\times0.527}{3983.4}\times100=13.2299\approx13$$

$$n_2=\frac{3000\times0.4068}{3983.4}\times100=30.6371\approx31$$

$$n_3=\frac{4000\times0.4051}{3983.4}\times100=40.6788\approx41$$

$$n_4=n-n_1-n_2-n_3=15$$

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 i	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

(一)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）

(二)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）

試題評析	本題屬於兩階段抽樣法之計算題型，考題跟101年地特類似，考生有勤練考古題的話，獲得滿分不難。
考點命中	《抽樣方法申論題完全制霸》，高點文化出版，趙治勳編著，第十六章範題8，頁16-10。

答：

(一) 此抽樣設計為兩階段群集抽樣法，第一階段之抽樣單位為區域，第二階段之抽樣單位為樹

(二)

	M_i	m_i	$p_i=\frac{a_i}{m_i}$	$s_{2i}^2=\frac{m_i}{m_i-1}p_iq_i$
1	200	10	$\frac{1}{10}$	$\frac{1}{10}$
2	300	15	$\frac{1}{3}$	$\frac{5}{21}$
3	300	15	$\frac{1}{3}$	$\frac{5}{21}$
4	400	20	$\frac{1}{5}$	$\frac{16}{95}$

$$\bar{\bar{p}}=\frac{1}{\hat{M}}\hat{A}=\frac{N}{\hat{M}n}\sum_{i=1}^{n}M_ip_i=\frac{10}{3000\times 4}[200\times\frac{1}{10}+300\times\frac{1}{3}+300\times\frac{1}{3}+400\times\frac{1}{5}]=0.25$$

其中 $\hat{M}=N\bar{M}=10\times 300=3000$

$$s_{\bar{\bar{p}}}=\frac{1}{\hat{M}}s_{\hat{A}}=\frac{1}{\hat{M}}\sqrt{N^2(1-f_1)\frac{s_{1b}^2}{n}+\frac{N}{n}\sum_{i=1}^{n}M_i(M_i-m_i)\frac{s_{2i}^2}{m_i}}$$

$$=\frac{1}{3000}\sqrt{10^2(1-\frac{4}{10})\frac{1433.3333}{4}+\frac{10}{4}\begin{bmatrix}200(200-10)\frac{\frac{1}{10}}{10}+300(300-15)\frac{\frac{5}{21}}{15}\\+300(300-15)\frac{\frac{5}{21}}{15}+400(400-20)\frac{\frac{16}{95}}{20}\end{bmatrix}}$$

$=0.06$

其中 $s_{1b}^2=\frac{\sum_{i=1}^{n}(M_ip_i-\hat{p}_t)^2}{n-1}=1433.3333$

$((\bar{\bar{p}}\pm 1.96s_{\bar{\bar{p}}})=(0.25\pm 1.96\times 0.06)=(0.1324,0.3676))$