《抽樣方法》

一、欲了解某一工業園區N家製造業者AI人才的需求狀況，下述三種抽樣設計可用以推估該工業園區有AI人才需求的業者家數比例及總需求人數：

(1) 如果園區各業者的營業規模已知，首先將業者依其營業規模分成L層（1,…,L），每層家數分別為$N_1,\cdots,N_L$，再由每層抽取一簡單隨機樣本，分別為$n_1,\cdots,n_L$，以調查業者的AI人才需求狀況。

(2) 如果園區各業者的營業規模未知，但已知園區業者營業規模的比例分別為$W_1,\cdots,W_L$，首先由N家業者抽取一簡單隨機樣本（n）調查業者規模及AI人才需求狀況後，再根據調查結果依其規模分成L層進行推估。

(3) 如果沒有園區業者營業規模的資訊，首先由N家業者抽取一簡單隨機樣本n'，取得營業規模資訊，而後根據營業規模資訊將n'家業者依其營業規模分成L層（$n'_1,\cdots,n'_L$），再由每層抽取一簡單隨機樣本，分別為$n_1,\cdots,n_L$，調查業者的AI人才需求狀況。

(一) 說明前述三種抽樣設計的抽樣方法為何？（10分）

(二) 若欲估計園區AI總需求人數（Y），分別列出對應前述三種抽樣方法的估計量（estimator）及該估計量之變異數的估計量。（15分）

試題評析	本題考到分層隨機、事後分層與雙重分層三種抽樣方法之點估計量與標準誤之公式，其中雙重分層抽樣法在歷屆考古題中只考過兩次，分別為104年與107年，考生容易忽略。
考點命中	《高點‧高上抽樣方法講義》第一回，趙治勳編撰，頁30～41，第4~5章

答：

(一) 抽樣設計(1)：分層隨機抽樣法
抽樣設計(2)：事後分層抽樣法
抽樣設計(3)：雙重分層抽樣法

(二) 抽樣設計(1)：分層隨機抽樣法

✓ 點估計 $\hat{Y}_{st} = N\bar{y}_{st} = N\sum_{h=1}^{L}\frac{N_h}{N}\bar{y}_h = \sum_{h=1}^{L}N_h\bar{y}_h$

✓ 點估計標準誤之估計值

$$v(\hat{Y}_{st}) = \sum_{h=1}^{L}N_h(N_h-n_h)\frac{s_h^2}{n_h} \quad \text{其中} \quad f_h = \frac{n_h}{N_h}$$

$$s_h^2 = \frac{\sum_{j=1}^{n_h}(y_{hj}-\bar{y}_h)^2}{n_h-1}$$

抽樣設計(2)：事後分層抽樣法

✓ 點估計 $\hat{Y}_{pst} = N\sum_{h=1}^{L}W_h\bar{y}_h$

✓ 點估計標準誤之估計值

$$v(\hat{Y}_{pst}) = N^2[\frac{1}{n}(1-f)\sum_{h=1}^{L}W_h s_h^2 + \frac{1}{n^2}\sum_{h=1}^{L}(1-W_h)s_h^2]$$

其中 $s_h^2 = \frac{\sum_{j=1}^{n_h}(y_{hj}-\bar{y}_h)^2}{n_h-1}$

抽樣設計(3)：雙重分層抽樣法

✓ 點估計 $\hat{Y}_{Dst}=N\bar{y}_{Dst}=N\sum_{h=1}^{L}w_h\bar{y}_h$ 其中 $w_h=\frac{n_h'}{n'}$

✓ 點估計標準誤之估計值

$$v(\hat{Y}_{Dst})=N^2\{\frac{g'}{n'-1}\sum_{h=1}^{L}w_h[(\frac{n'}{g'}w_h-1)\frac{s_h^2}{n_h}+(\bar{y}_h-\bar{y}_{Dst})^2]\}$$

其中 $g'=\frac{N-n'}{N-1}$ 及 $s_h^2=\frac{\sum_{j=1}^{n_h}(y_{hj}-\bar{y}_h)^2}{n_h-1}$

二、某一縣市共4000家養雞戶分散在20個村里（cluster），欲透過調查了解該縣市養雞戶的所得狀況，抽樣方法可採用一階段集體抽樣（single-stage cluster sampling）或二階段集體抽樣（two-stage cluster sampling）。

(一)若抽樣方法採一階段集體抽樣，首先由20個村里以簡單隨機抽樣（SRS）抽出3個村里，就抽得的3個村里之養雞戶全數調查，調查結果村里內養雞戶數及平均年所得列於下表：

村里（cluster）	養雞戶數（M_i）	村里內養雞戶平均每戶年所得（百萬元）（$\bar{y}_i$）
1	150	4
2	200	8
3	250	5.6

採用集體大小比率估計量（ratio-to-size estimator, $\bar{\bar{y}}_R$）估計該縣市養雞戶平均每戶年所得（以百萬元為單位）及該估計量之標準誤。（10分）

(二)若抽樣方法採二階段集體抽樣，首先由20個村里以簡單隨機抽樣抽出3個村里，再就抽得的3個村里之養雞戶以簡單隨機抽樣分別抽出1/10養雞戶進行調查，調查結果養雞戶之平均所得列於下表：

村里（cluster）	農戶數（M_i）	抽出養雞戶數（m_i）	村里內抽得之養雞戶平均每戶年所得及標準差（百萬元）（$\bar{y}_i$, s_i）	
			$\bar{y}_i$	s_i
1	150	15	4	1
2	200	20	7	3
3	250	25	6	2

試問：（15分）

(1)本抽樣設計第一階段的抽樣單位（primary sampling unit, PSU）及第二階段的抽樣單位（secondary sampling unit, SSU）分別為何？

(2)採用不偏估計量（unbiased estimator, $\bar{\bar{y}}$）估計該縣市養雞戶平均每戶年所得（以百萬元為單位）及該估計量之標準誤。

試題評析	本題考到比率群集與兩階段抽樣法，屬於常考範圍，不論是上課題庫或總複習課程都有幫考生複習過，考生應該容易獲得高分。
考點命中	《高點‧高上抽樣方法講義》第一回，趙治勳編撰，頁78～82，第11章與頁107～112第16章

答：

$\sum M_i = 600$，$\sum M_i^2 = 125000$，$\sum y_i = 3600$，$\sum y_i^2 = 4880000$，$\sum M_i y_i = 760000$

$ss_M = \sum M_i^2 - \dfrac{(\sum M_i)^2}{n} = 5000$，$ss_Y = \sum y_i^2 - \dfrac{(\sum y_i)^2}{n} = 560000$，

$ss_{MY} = \sum M_i y_i - \dfrac{(\sum M_i \sum y_i)}{n} = 40000$

(一) $\bar{\bar{y}}_R = \bar{y}_{rcl} = \dfrac{\bar{y}_t}{\bar{m}} = \dfrac{1200}{200} = 6$ (百萬元)

$$s_{\bar{\bar{y}}_R} = s_{\bar{y}_{rcl}} = \frac{N}{M}\sqrt{(1-f)\frac{s_c^2}{n}} = \frac{20}{4000}\sqrt{(1-\frac{3}{20})\frac{130000}{3}} = 0.9596 \quad \text{(百萬元)}$$

其中 $s_c^2 = \dfrac{\sum_{i=1}^{n}(y_i - \bar{y}_{rcl}M_i)^2}{n-1} = s_Y^2 + \bar{y}_{rcl}^2 s_M^2 - 2\bar{y}_{rcl} r' s_M s_Y$

$$= \frac{560000}{3-1} + 6^2 \times \frac{5000}{3-1} - 2\times 6\times\frac{40000}{3-1} = 130000$$

(二) (1) 第一階段抽樣單位為村里，第二階段抽樣單位為養雞戶

(2) $\bar{\bar{y}} = \hat{\bar{\bar{Y}}} = \dfrac{N}{Mn}\sum_{i=1}^{n} M_i \bar{y}_i = \dfrac{20}{4000\times 3}[150\times 4 + 200\times 7 + 250\times 6]$

$= 5.8333$ (百萬元)

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

$$= \frac{1}{4000}\sqrt{20^2(1-\frac{3}{20})\frac{243333.3333}{3} + \frac{20}{3}\left[\begin{array}{l} 150(150-15)\dfrac{1^2}{15} + 200(200-20)\dfrac{3^2}{20} \\ +250(250-25)\dfrac{2^2}{25} \end{array}\right]}$$

$= 1.3171$ (百萬元)

其中 $s_{1b}^2 = \dfrac{\sum_{i=1}^{n}(M_i\bar{y}_i - \bar{\hat{Y}}_t)^2}{n-1} = 243333.3333$

三、欲了解某區域養殖漁戶營運狀況，該區域共有800家養殖漁戶，首先由該區域抽得一個包含500戶（第一重樣本n' = 500）的簡單隨機樣本以取得養殖型態（是否為漁電共生的層別資訊），調查得知其中100戶為漁電共生戶，進而以簡單隨機抽樣由漁電共生及非漁電共生的養殖漁戶分別抽20%以調查其營運成本資訊，調查結果整理如下表：

養殖型態（層別）	第一重樣本（戶）n'_h	第二重樣本（戶）n_k	養殖漁戶年營運成本	
			平均年營運成本 $\bar{y}_i$（十萬元）	標準差 s_i（十萬元）
漁電共生（I）	100	20	120	100
非漁電共生（II）	400	80	50	60
合計	500（n'）	100（n）		75（s）

(一) 估計該區域養殖漁戶平均年營運成本（$\bar{y}$）（以十萬元為單位）及該估計量的標準誤。（10分）

(二) 如果此調查總預算為44,500元，取得養殖型態的單位成本為9元，調查營運成本的單位成本為400元（$c' = 9$, $c_k = 400$），有關標準差、各層權重、各層平均年營運成本及標準差之母體資

訊分別以前述樣本資料（s, $w'_i=\frac{n'_i}{n'}$, $\bar{y}_i$, s_i）取代。試求：（15分）

(1) 決定分層雙重抽樣的最佳抽樣設計（Optimum double sample plan），亦即求算n'、n_h。

(2) 就(1)的抽樣設計求算平均年營運成本估計量（以十萬元為單位）的變異數。

試題評析	本題考到雙重分層抽樣法的計算題型，考古題未曾出現過，雖然如此，老師在上課題庫裡面也有自編例題供考生練習，且計算樣本數的公式也有收錄在講義中，準備充分的考生可以輕鬆利用這一題跟其他人拉開距離。
考點命中	《高點・高上抽樣方法講義》第一回，趙治勳編撰，頁42～48，第6章 《抽樣方法申論題完全制霸》，高點文化出版，趙治勳編著，頁6-2~6-5，例1~3

答：

(一) $\hat{\bar{Y}}=\bar{y}_{Dst}=\sum_{h=1}^{L}w_h\bar{y}_h=\frac{100}{500}\times120+\frac{400}{500}\times50=64$ (十萬元)

$$s_{\bar{y}_{Dst}}=\sqrt{\frac{g'}{n'-1}\sum_{h=1}^{L}w_h[(\frac{n'}{g'}w_h-1)\frac{s_h^2}{n_h}+(\bar{y}_h-\bar{y}_{Dst})^2]}$$

$$=\sqrt{\frac{0.9387}{500-1}\left\{\begin{matrix}\frac{100}{500}[(\frac{500}{0.9387}\frac{100}{500}-1)\frac{100^2}{20}+(120-64)^2]\\+\frac{400}{500}[(\frac{500}{0.9387}\frac{400}{500}-1)\frac{60^2}{80}+(50-64)^2]\end{matrix}\right\}}$$

$=7.07932$ (十萬元)

其中 $g'=\frac{N-n'}{N-1}=\frac{800-500}{800-1}=0.9387$

(二) (1)

$$n'=\frac{C\sqrt{c_nV_n'}}{\sqrt{c_nc_n'}(\sqrt{c_nV_n}+\sqrt{c_n'V_n'})}=\frac{44500\sqrt{400\times784}}{\sqrt{400\times9}(\sqrt{400\times4624}+\sqrt{9\times784})}=287.6270 \text{ 取 288 個樣本}$$

$$n=\frac{C\sqrt{c_n'V_n}}{\sqrt{c_nc_n'}(\sqrt{c_nV_n}+\sqrt{c_n'V_n'})}=\frac{44500\sqrt{9\times4624}}{\sqrt{400\times9}(\sqrt{400\times4624}+\sqrt{9\times784})}=104.7784$$

取105個樣本

其中 $V_n=(\sum_{h=1}^{L}W_hS_h)^2=(\frac{100}{500}\times100+\frac{400}{500}\times60)^2=4624$

$V_n'=\sum_{h=1}^{L}W_h(\bar{Y}_h-\bar{Y})^2=\frac{100}{500}(120-64)^2+\frac{400}{500}(50-64)^2=784$

各層樣本配置採紐門配置 $n_h=\frac{w_hS_h}{\sum_{h=1}^{L}w_hS_h}\times n$

$$n_1=\frac{\frac{100}{500}\times100}{\frac{100}{500}\times100+\frac{400}{500}\times60}\times105=30.8824 \text{ 取31個樣本}$$

$n_2=n-n_1=105-31=74$ 個樣本

(2) $V_{opt}=\frac{V_n}{n}+\frac{V_n'}{n'}=\frac{(\sqrt{c_nV_n}+\sqrt{c_n'V_n'})^2}{C}=\frac{(\sqrt{400\times4624}+\sqrt{9\times784})^2}{44500}$

$=46.8570$ (十萬元^2)

四、欲了解2022年國內汽車銷售概況，就2000家汽車銷售業者進行調查，汽車銷售業者分為兩大類：國產型（I）及進口型（II），業者家數分別為$N_1=1500$及$N_2=500$。抽樣方法採用分層隨機抽樣，依類別分層，從每一層分別隨機抽出10家業者進行調查。假設2021年（x）各類業者的年平均銷售量已知為：$\bar{X}_1=200$（輛）；$\bar{X}_2=140$（輛）。

調查結果20家業者在2021年（x）及2022年（y）的銷售量統計如下：

層別（h）	變數	樣本均數 $\bar{y}_i$，$\bar{x}_i$（輛）	比率（$\hat{R}_h$）	樣本共變異數 S_{xyh}	樣本標準差 S_h
I	y x	240 200	1.2	7200	100 80
II	y x	180 100	1.8	2200	60 40
合計	$S_y=110$，$S_x=90$，$S_{xy}=9000$				

(一) 利用下列估計量估計年平均銷售量（$\bar{Y}$）及該估計量的變異數：（15分）

(1) $\bar{y}_{st}$，分層隨機抽樣結合簡單均數估計量（mean per unit estimator）。

(2) $\bar{y}_{RS}$，分層隨機抽樣結合分開比率估計量（separate ratio estimator）。

(3) $\bar{y}_{RC}$，分層隨機抽樣結合混合比率估計量（combined ratio estimator）。

(二) 求算估計量$\bar{y}_{st}$、$\bar{y}_{RS}$、$\bar{y}_{RC}$對單位均數估計量（$\bar{y}$）之相對效率（relative efficiency），並說明那個估計量具有較佳精確度。（10分）

試題評析	本題抽樣法是分層隨機抽樣法，但是考到三種估計法，分別有簡單估計法、聯合比率估計法與個別比率估計法，其中比率分層只考過兩次（102年與108年），雖然跟考古題相似，但是考過的次數不多，容易被考生們忽略。
考點命中	《高點‧高上抽樣方法講義》第一回，趙治勳編撰，頁30～37，第4章與頁70～77，第10章

答：

(一) (1)分層隨機抽樣法-簡單估計法

$$\bar{y}_{st}=\sum_{h=1}^{L}W_h\bar{y}_h=\frac{1500}{2000}\times 240+\frac{500}{2000}\times 180=225\ (\text{輛})$$

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

$$=\frac{1}{2000^2}[1500(1500-10)\frac{100^2}{10}+500(500-10)\frac{60^2}{10}]=580.8\ (\text{輛}^2)$$

(2)分層隨機抽樣法-個別比率估計法

$$X=1500\times 220+500\times 40=400000$$

$$r_s=\sum_{h=1}^{L}r_h\frac{X_h}{X}=\frac{240}{200}\times\frac{1500\times 220}{400000}+\frac{180}{100}\times\frac{500\times 140}{400000}=1.305$$

$$v(r_s)=\frac{1}{X^2}\sum_{h=1}^{L}N_h(N_h-n_h)\frac{s_{sh}^2}{n_h}$$

$$=\frac{1}{400000^2}[1500(1500-10)\frac{1936}{10}+500(500-10)\frac{864}{10}]=0.002837$$

其中 $s_{s1}^2=s_{Y1}^2+r_1^2s_{X1}^2-2r_1s_{XY1}=100^2+1.2^2\times 80^2-2\times 1.2\times 7200$
$=1936$

$s_{s2}^2=s_{Y2}^2+r_2^2s_{X2}^2-2r_2s_{XY2}=60^2+1.8^2\times 40^2-2\times 1.8\times 2200$
$=864$

$$\bar{y}_{Rs}=r_s\bar{X}=1.305\times\frac{400000}{2000}=261\text{ (輛)}$$

$$v(\bar{y}_{Rs})=\bar{X}^2v(r_s)=(\frac{400000}{2000})^2\times0.002837=113.48\text{ (輛}^2\text{)}$$

(3)分層隨機抽樣法-聯合比率估計法

$$X=1500\times220+500\times140=400000$$

$$r_c=\frac{\sum_{h=1}^{L}W_h\bar{y}_h}{\sum_{h=1}^{L}W_h\bar{x}_h}=\frac{\frac{1500}{2000}\times240+\frac{500}{2000}\times180}{\frac{1500}{2000}\times200+\frac{500}{2000}\times100}=1.2857$$

$$v(r_s)=\frac{1}{X^2}\sum_{h=1}^{L}N_h(N_h-n_h)\frac{s_{sh}^2}{n_h}$$

$$=\frac{1}{400000^2}[1500(1500-10)\frac{2065.2767}{10}+500(500-10)\frac{587.7592}{10}]$$

$$=0.002975$$

其中 $s_{s1}^2=s_{Y1}^2+r_c^2s_{X1}^2-2r_cs_{XY1}=100^2+1.2857^2\times80^2-2\times1.2857\times7200$

$=2065.2767$

$s_{s2}^2=s_{Y2}^2+r_c^2s_{X2}^2-2r_cs_{XY2}=60^2+1.2857^2\times40^2-2\times1.2857\times2200$

$=587.7592$

$$\bar{y}_{Rc}=r_c\bar{X}=1.2857\times\frac{400000}{2000}=257.14\quad\text{(輛)}$$

$$v(\bar{y}_{Rc})=\bar{X}^2v(r_c)=(\frac{400000}{2000})^2\times0.002975=119\text{ (輛}^2\text{)}$$

(二) $v(\bar{y})=(1-f)\frac{s_Y^2}{n}=(1-\frac{20}{2000})\frac{110^2}{20}=598.95\text{ (輛}^2\text{)}$

相對效率分別為

$\frac{v(\bar{y}_{st})}{v(\bar{y})}=\frac{580.8}{598.95}=0.9697$ $\quad$ $\frac{v(\bar{y}_{Rs})}{v(\bar{y})}=\frac{113.48}{598.95}=0.1895$

$\frac{v(\bar{y}_{Rc})}{v(\bar{y})}=\frac{119}{598.95}=0.1987$

比較後可得個別比率(分開比率)估計法具有較佳之精確度