《抽樣方法》

一、假設 $F=\{u_1,u_2,u_3,u_4\}$ 是一個僅僅包含四個元素的小規模有限母體（finite population），而我們採用簡單隨機抽樣（simple random sampling）從母體F之中抽出樣本大小（sample size）為 $n=2$ 的樣本組合，那麼總共會有多少種不同的樣本組合？如果我們改用機率均等的隨機置回抽樣（sampling with replacement），那麼總共會有多少種不同的樣本組合？如果母體F之中四個元素的研究變數（study variable）值分別為 $y_1=1$、$y_2=3$、$y_3=3$、以及 $y_4=9$，那麼母體F的母體平均數（population mean）是多少？母體變異數（population variance）是多少？（5分）

試題評析	本題是簡單隨機抽樣法之可能樣本組合數，屬於基本計算考題，要拿到分數並不難。
考點命中	《高點・高上抽樣方法講義》，趙治勳編撰，頁10。

答：

採用簡單隨機抽樣法下抽取兩個樣本，共有 $C_n^N=C_2^4=6$ 種不同的樣本組合。而在機率均等的隨機置回抽樣下抽取兩個樣本，共有 $N^n=4^2=16$ 種不同的樣本組合。母體平均數：$\bar{Y}=\dfrac{Y}{N}=4$

母體變異數：$\sigma^2=\dfrac{\sum_{i=1}^{N}(Y_i-\bar{Y})^2}{N}=9$

(這裡使用無限母體的變異數公式，原因是對應到下面第二大題，若是有限母體的變異數即為

$S^2=\dfrac{\sum_{i=1}^{N}(Y_i-\bar{Y})^2}{N-1}=12$)

二、假設第一題之中從母體F抽出簡單隨機樣本的樣本數據為 Y_1 與 Y_2（註：採用大寫英文字母Y，表示樣本數據皆為隨機變數）。我們分別使用下列三種不同的點估計量（point estimator）來估計母體變異數 σ^2：

$$\hat{\sigma}^2_{(1)}=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})^2，\hat{\sigma}^2_{(2)}=\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2，\hat{\sigma}^2_{(3)}=\frac{N-1}{N(n-1)}\sum_{i=1}^{n}(Y_i-\bar{Y})^2，$$

其中 $\bar{Y}$ 為樣本平均數（sample mean）、N為母體大小（population size）。

(一)試計算以上三種估計量的抽樣分布（sampling distribution）與均方誤差（mean square error），並判別它們是否為不偏估計量（unbiased estimator）。（20分）

(二)假設我們改用機率均等的隨機置回抽樣，請重新判別三種估計量是否為不偏估計量。簡答即可，無需計算或證明。（3分）

試題評析	本題是考簡單隨機抽樣法中，抽後放回與抽後不放回兩種實驗規則下母體變異數之不偏估計量，屬於基本計算考題，考古題中也出現過很多相關題型，要拿到分數並不難。
考點命中	《迴歸分析熱門題庫2016版》，高點文化出版，第二篇第二章，趙治勳編著。

答：

若母體變異數為 $S^2=\dfrac{\sum_{i=1}^{N}(Y_i-\bar{Y})^2}{N-1}$ 的話，如此第二小題就沒有不偏估計量，故猜測出題老師之母體變異數為

$$\sigma^2=\frac{\sum_{i=1}^{N}(Y_i-\bar{Y})^2}{N}=9$$

(一)

樣本	機率	$\sum_{i=1}^{n}(y_i-\bar{y})^2$	$\hat{\sigma}^2_{(1)}$	$\hat{\sigma}^2_{(2)}$	$\hat{\sigma}^2_{(3)}$
1，3	4/12	2	4	2	1.5
1，9	2/12	32	16	32	24
3，3	2/12	0	0	0	0
3，9	4/12	18	9	18	13.5

<u>$\hat{\sigma}^2_{(1)}$</u>：

$\hat{\sigma}^2_{(1)}$	0	4	9	16
$P(\hat{\sigma}^2_{(1)})$	2/12	4/12	4/12	2/12

$E(\hat{\sigma}^2_{(1)})=7$　　$E([\hat{\sigma}^2_{(1)}]^2)=75$　　$V(\hat{\sigma}^2_{(1)})=26$

$MSN(\hat{\sigma}^2_{(1)})=26+[E(\hat{\sigma}^2_{(1)})-\sigma^2]^2=30$

<u>$\hat{\sigma}^2_{(2)}$</u>：

$\hat{\sigma}^2_{(2)}$	0	2	18	16
$P(\hat{\sigma}^2_{(2)})$	2/12	4/12	4/12	2/12

$E(\hat{\sigma}^2_{(2)})=12$　　$E([\hat{\sigma}^2_{(2)}]^2)=280$　　$V(\hat{\sigma}^2_{(2)})=136$

$MSN(\hat{\sigma}^2_{(2)})=V(\hat{\sigma}^2_{(2)})+[E(\hat{\sigma}^2_{(2)})-\sigma^2]^2=145$

<u>$\hat{\sigma}^2_{(3)}$</u>：

$\hat{\sigma}^2_{(3)}$	0	1.5	13.5	24
$P(\hat{\sigma}^2_{(3)})$	2/12	4/12	4/12	2/12

$E(\hat{\sigma}^2_{(3)})=9$　　$E([\hat{\sigma}^2_{(3)}]^2)=157.5$　　$V(\hat{\sigma}^2_{(3)})=76.5$

$MSN(\hat{\sigma}^2_{(3)})=V(\hat{\sigma}^2_{(3)})+[E(\hat{\sigma}^2_{(3)})-\sigma^2]^2=76.5$

由於 $E(\hat{\sigma}^2_{(3)})=9=\sigma^2$，故 $\hat{\sigma}^2_{(3)}$ 為 σ^2 之不偏估計量

(二)由於 $E(\hat{\sigma}^2_{(2)})=9=\sigma^2$，故 $\hat{\sigma}^2_{(2)}$ 為 σ^2 之不偏估計量

說明：(題目沒有要求呈現以下計算過程)

樣本	機率	$\sum_{i=1}^{n}(y_i-\bar{y})^2$	$\hat{\sigma}^2_{(1)}$	$\hat{\sigma}^2_{(2)}$	$\hat{\sigma}^2_{(3)}$
1，1	1/16	0	0	0	0
1，3	4/16	2	4	2	1.5
1，9	2/16	32	16	32	24
3，3	4/16	0	0	0	0

3，9	4/16	18	9	18	13.5
9，9	1/16	0	0	0	0

$\hat{\sigma}^2_{(1)}$：

$\hat{\sigma}^2_{(1)}$	0	4	9	16
$P(\hat{\sigma}^2_{(1)})$	6/16	4/16	4/16	2/16

$E(\hat{\sigma}^2_{(1)}) = 5.25$

$\hat{\sigma}^2_{(2)}$：

$\hat{\sigma}^2_{(2)}$	0	2	18	16
$P(\hat{\sigma}^2_{(2)})$	6/16	4/16	4/16	2/16

$E(\hat{\sigma}^2_{(2)}) = 9$

$\hat{\sigma}^2_{(3)}$：

$\hat{\sigma}^2_{(3)}$	0	1.5	13.5	24
$P(\hat{\sigma}^2_{(3)})$	6/16	4/16	4/16	2/16

$E(\hat{\sigma}^2_{(3)}) = 6.75$

三、假設某一間大學共有8000位學生，其中2000人為男生，6000人為女生。同時，8000位學生之中有7000人屬於日間部，1000人屬於夜間部。

(一)我們打算採用分層隨機抽樣（stratified random sampling）之方法從8000位學生之中抽出n＝16位學生，來估計8000位學生的平均身高。那麼應當依照男女性別來分層，抑或依照日間、夜間部別來分層較為恰當？如果依據比例配置（proportional allocation）來進行抽樣，須要分別從各層抽出多少人？抽出樣本並取得樣本觀測值之後，應如何估計8000位學生的平均身高？另外，如果我們想要採用分層隨機抽樣來調查8000位學生的月平均收入（例如校外打工或兼差之收入），那麼應當依照男女性別來分層，抑或依照日間、夜間部別來分層較為恰當？（8分）

(二)先前採用分層隨機抽樣來估計8000位學生之平均身高的問題，如果依據尼門配置（Neyman allocation）來進行抽樣，須要分別從各層抽出多少人？假設母體之中男生層之身高的變異數為女生層之身高變異數的1.96倍，日間部與夜間部學生身高的變異數比例數為1.1。（8分）

(三)先前估計8000位學生之平均身高的問題，如果我們改用事後分層（post-stratification）之方法來抽樣與推估，那麼整個調查過程應該如何進行？從樣本配置的觀點來看，事後分層可以被視為何種類型之配置？由於n＝16屬中小樣本，萬一發生空事後層（empty post-stratum）之情形，應當如何處理或補救？（12分）

試題評析	本題(一)、(二)是考分層隨機抽樣法下的樣本數配置，要拿到分數並不難，而(三)是考事後分層之抽樣法及觀念。
考點命中	《高點・高上抽樣方法講義》，趙治勳編撰，頁33，36。

答：

(一)由於分層原則應以「層與層間變異性大，而層內變異性小」，故在平均身高之估計上應該以男女性別去分層較為恰當。

比例配置：各層樣本數為 $n_h = \dfrac{N_h}{N} \times n$

男生$(h=1)$：$n_1 = \dfrac{N_1}{N} \times n = \dfrac{2000}{8000} \times 16 = 4$

女生$(h=2)$：$n_2 = n - n_1 = 16 - 4 = 12$

點估計量：$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h = \dfrac{2000}{8000} \times \bar{y}_1 + \dfrac{6000}{8000} \times \bar{y}_2$

由於分層原則應以「層與層間變異性大，而層內變異性小」，故在平均收入之估計上應該以日間部與夜間部去分層較為恰當。

(二)由題意得知，$S_1^2 = 1.96 \times S_2^2 \Rightarrow S_1 = \sqrt{1.96} \times S_2$

尼(紐)門配置：各層樣本數為 $n_h = \dfrac{N_h S_h}{\sum_{h=1}^{L} N_h S_h} \times n$

男生$(h=1)$：$n_1 = \dfrac{N_1 S_1}{N_1 S_1 + N_2 S_2} \times n = \dfrac{N_1 \sqrt{1.96} \times S_2}{N_1 \sqrt{1.96} \times S_2 + N_2 S_2} \times n$

$= \dfrac{N_1 \sqrt{1.96}}{N_1 \sqrt{1.96} + N_2} \times n = \dfrac{2000 \times \sqrt{1.96}}{2000 \times \sqrt{1.96} + 6000} \times 16$

$= 5.0909 \approx 5$

女生$(h=2)$：$n_2 = n - n_1 = 16 - 5 = 11$

(三)先從抽樣母體中以簡單隨機抽樣法抽取 $n = 16$ 個抽樣單位，再觀察這些樣本中男生與女生之個數 (n_1, n_2)，再以事後分層下母體平均數 $\bar{Y}$ 之估計量 $\bar{y}_{pst} = \sum_{h=1}^{L} W_h \bar{y}_h$ 進行估計。

事後分層可以視為分層隨機抽樣法中之比例配置。

當出現「空事後層」時，應該合併層，又或者重新檢視分層標準，又或者是增加樣本數讓各層都有抽取到對應之觀察值。

四、何謂雙重抽樣（double sampling）？試就分層隨機抽樣以及比值估計兩種情況分別說明之，並寫出點估計量的數學式。（22分）

試題評析	本題是雙重抽樣法之觀念題，這種申論題沒有所謂之標準答案，只要考生把所學之所有觀念清楚地敘述，要拿到分數並不難。
考點命中	《高點・高上抽樣方法講義》第七章及第十三章，趙治勳編撰。

答：

(一)雙重抽樣法之定義：

先從抽樣母體中抽出一組大樣本，再由抽出的大樣本中抽出一組小樣本，此抽樣法稱為雙重抽樣法或雙面抽樣法。雙重抽樣法通常用在當一些重要變數未知就無法針對母體參數進行估計時，可以藉由第一次抽取之大樣本以瞭解該重要變數，再利用第二次抽取之小樣本估計母體參數。

(二)

1.雙重<u>分層隨機抽樣法</u>：

當分層隨機抽樣法中各層母體大小 N_h 未知時，就無法得到點估計值與標準誤之估計值。此時，可以採用雙重分層抽樣法。也就是說，藉由第一次抽取之大樣本瞭解 N_h，再利用第二次抽取之小樣本估計母體參數。

母體平均數 $\overline{Y}$ 之估計： 點估計 $\overline{y}_{Dst}=\sum_{h=1}^{L} w_h\overline{y}_h$ 其中 $w_h=\frac{n_h'}{n'}$

2.雙重比率簡單估計法：

若輔助變數 $\overline{X}$ 未知時，就得使用雙重抽樣法之技術，以第一重樣本平均數 $\overline{x}'$ 估計輔助變數 $\overline{X}$，再以第二重樣本估計母體參數。

母體平均數 $\overline{Y}$ 之估計：點估計 $\overline{y}_r=r\overline{x}'=\frac{\overline{y}}{\overline{x}}\overline{x}'$

五、集群抽樣（cluster sampling）以及二階段集群抽樣（two-stage cluster sampling）與比值估計有何關聯性？試就估計母體平均數之情況予以分別說明並寫出點估計量的數學式。（22分）

試題評析	本題是群集抽樣法與二階段群集抽樣法再使用比率估計法之相關考題， 比率群集抽樣法在課本中已經有介紹過，且考古題中也出現過不少相關題目，要拿到分數並不難，但是二階段群集抽樣法在使用比率估計實屬冷門考題，考生要在考試時自行推導出母體平均數之點估計量真的很不容易。
考點命中	《高點・高上抽樣方法講義》第八章、第十二章及第十七章，趙治勳編撰。

答：

(一)在群集隨機抽樣法中使用比率之觀念提升精確度，這時稱為比率群集估計法，比率群集估計法中之輔助變數即為各群集之母體大小數(M_i,M)

母體平均數 $\overline{\overline{Y}}$ 之估計：點估計 $\overline{y}_{rcl}=\frac{\overline{y}_t}{\overline{m}}$ 其中 $\overline{y}_t=\frac{1}{n}\sum_{i=1}^{n}y_i$ ，$\overline{m}=\frac{1}{n}\sum_{i=1}^{n}M_i$

(二)在二階段群集隨機抽樣法中使用比率之觀念提升精確度，這時稱為二階段比率群集估計法， 二階段比率群集估計法中之輔助變數即為各群集之母體大小數(M_i,M)

母體平均數 $\overline{\overline{Y}}$ 之估計：點估計 $\hat{\overline{\overline{Y}}}=\frac{\sum_{i=1}^{n}M_i\overline{y}_i}{\sum_{i=1}^{n}M_i}=\frac{\sum_{i=1}^{n}M_i\overline{y}_i}{n\overline{m}}$