《抽樣方法》

試題評析	1.此卷與104年地特應該是屬於同一位委員出題的，兩份考卷相似度很高。除了第一大題有些許的計算外，其他五大題都屬於觀念題，還需要寫下點估計量之計算公式。 2.此卷較難的是第五大題中兩階段抽樣法之比率估計，但其實已經在104年地特考過了，因此有熟讀考古題之考生應該可以獲得分數。
考點命中	第一題：《高點‧高上抽樣方法講義》，趙治勳編撰，第一章。 第二題：《高點‧高上抽樣方法講義》，趙治勳編撰，第四、五章。 第三題：《高點‧高上抽樣方法講義》，趙治勳編撰，第九章。 第四題：《高點‧高上抽樣方法講義》，趙治勳編撰，第十三章。 第五題：《高點‧高上抽樣方法講義》，趙治勳編撰，第十一、十六章。 第六題：《高點‧高上抽樣方法講義》，趙治勳編撰，第二章。

一、(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 $y_4=9$。我們採用不置回的簡單隨機抽樣（simple random sampling without replacement）從母體 F 之中抽出樣本大小（sample size）為 $n=2$ 的樣本組合，並以 Y_1，Y_2 來表示此樣本組合中的兩個樣本數據（註：採用大寫英文字母 Y，表示樣本數據皆為隨機變數）。試求 Y_1 與 Y_2 的聯合機率分佈（joint probability distribution）以及 Y_1 與 Y_2 的共變異數（covariance）。（15分）

(二)假設前一小題之中抽出樣本大小為 $n=3$ 的樣本組合，並以 Y_1,Y_2,Y_3 來表示此樣本組合中的三個樣本數據。試求 Y_1 與 Y_3 的共變異數。（5分）

答：

(一)

		Y_2		
	$f_{Y_1Y_2}(y_1,y_2)$	1	3	9
Y_1	1	0	$\frac{1}{4}\times\frac{2}{3}=\frac{1}{6}$	$\frac{1}{4}\times\frac{1}{3}=\frac{1}{12}$
	3	$\frac{2}{4}\times\frac{1}{3}=\frac{1}{6}$	$\frac{2}{4}\times\frac{1}{3}=\frac{1}{6}$	$\frac{2}{4}\times\frac{1}{3}=\frac{1}{6}$
	9	$\frac{1}{4}\times\frac{1}{3}=\frac{1}{12}$	$\frac{1}{4}\times\frac{2}{3}=\frac{1}{6}$	0

$$Cov(Y_1,Y_2)=E(Y_1Y_2)-E(Y_1)E(Y_2)=13-4\times4=-3$$

其中 $E(Y_1Y_2)=1\times3\times\frac{1}{6}+1\times9\times\frac{1}{12}+\cdots+9\times3\times\frac{1}{6}=13$

$$E(Y_1)=1\times\frac{1}{4}+3\times\frac{1}{2}+9\times\frac{1}{4}=4=E(Y_2)$$

(二)

$$P(Y_1=1,Y_3=3)=\frac{1}{4}\times\frac{1}{3}\times1+\frac{1}{4}\times\frac{2}{3}\times\frac{1}{2}=\frac{1}{6}$$

$$P(Y_1=3,Y_3=1)=\frac{2}{4}\times\frac{1}{3}\times\frac{1}{2}+\frac{2}{4}\times\frac{1}{3}\times\frac{1}{2}=\frac{1}{6}$$

$P(Y_1=9,Y_3=3)=\frac{1}{4}\times\frac{1}{3}\times 1+\frac{1}{4}\times\frac{2}{3}\times\frac{1}{2}=\frac{1}{6}$

$P(Y_1=3,Y_3=9)=\frac{2}{4}\times\frac{1}{3}\times\frac{1}{2}+\frac{2}{4}\times\frac{1}{3}\times\frac{1}{2}=\frac{1}{6}$

$P(Y_1=3,Y_3=3)=\frac{2}{4}\times\frac{1}{3}\times\frac{1}{2}+\frac{2}{4}\times\frac{1}{3}\times\frac{1}{2}=\frac{1}{6}$

		Y_3		
	$f_{Y_1Y_3}(y_1,y_3)$	1	3	9
Y_1	1	0	$\frac{1}{6}$	$\frac{1}{4}\times\frac{2}{3}\times\frac{1}{2}=\frac{1}{12}$
	3	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
	9	$\frac{1}{4}\times\frac{2}{3}\times\frac{1}{2}=\frac{1}{12}$	$\frac{1}{6}$	0

$Cov(Y_1,Y_3)=E(Y_1Y_3)-E(Y_1)E(Y_3)=13-4\times 4=-3$

其中 $E(Y_1Y_3)=1\times 3\times\frac{1}{6}+1\times 9\times\frac{1}{12}+\text{L}\ +9\times 3\times\frac{1}{6}=13$

$E(Y_1)=1\times\frac{1}{4}+3\times\frac{1}{2}+9\times\frac{1}{4}=4=E(Y_3)$

二、(一)假設某一個有限母體可被完整地切割為L個互不相交的層別（strata），其中第h層之層大小（stratum size）為 N_h，第 h 層之第 j 個元素的研究變數值為 y_{hj}，$h=1,\ldots,L$，$j=1,2,\ldots,N_h$。母體大小（population size）、母體平均數（population mean）以及母體變異數（population variance）的數學式分別為

母體大小：$N=N_1+N_2+\cdots+N_L$；

母體平均數：$\mu=\frac{1}{N}\sum_{h=1}^{L}\sum_{j=1}^{N_h}y_{hj}$；

母體變異數：$\sigma^2=\frac{1}{N}\sum_{h=1}^{L}\sum_{j=1}^{N_h}(y_{hj}-\mu)^2$。

其次，母體中第h層之層平均數（stratum mean）以及層變異數（stratum variance）的數學式分別為

第h層之層平均數：$\mu_h=\frac{1}{N_h}\sum_{j=1}^{N_h}y_{hj}$；

第h層之層變異數：$\sigma_h^2=\frac{1}{N_h}\sum_{j=1}^{N_h}(y_{hj}-\mu_h)^2$。

試證明下列等式：

$$\sigma^2=\sum_{h=1}^{L}W_h\sigma_h^2+\sum_{h=1}^{L}W_h(\mu_h-\mu)^2，$$

其中 $W_h=\frac{N_h}{N}$ 為第h層的層比重（stratum weight），$h=1,\ldots,L$。（10分）

(二)使用分層隨機抽樣（stratified random sampling）之後，抽樣者應如何推估（未知其值

的）母體平均數？需要什麼假設條件？請詳細說明。（4分）

（三）在使用分層隨機抽樣之前，抽樣者必須事先決定將要採用何種較為適當的配置，例如比例配置（proportional allocation）、尼門配置（Neyman allocation）。請詳細說明在使用事後分層（post-stratification）的方法之前，抽樣者是否也必須事先決定採用何種配置？（5分）

（四）請詳細說明在什麼情況下，抽樣者需要借助於雙重抽樣（double sampling）之方法，來進行分層隨機抽樣並推估母體的平均數（或是推估母體的其他參數）。（5分）

答：

(一) $$\sigma^2=\frac{\sum_{h=1}^{L}\sum_{j=1}^{N_h}(y_{hj}-\mu)^2}{N}=\frac{\sum_{h=1}^{L}\sum_{j=1}^{N_h}[(y_{hj}-\mu_h)+(\mu_h-\mu)]^2}{N}$$

$$=\frac{\sum_{h=1}^{L}\sum_{j=1}^{N_h}[(y_{hj}-\mu_h)^2+(\mu_h-\mu)^2+2(y_{hj}-\mu_h)(\mu_h-\mu)]}{N}$$

$$=\frac{\sum_{h=1}^{L}\sum_{j=1}^{N_h}(y_{hj}-\mu_h)^2+\sum_{h=1}^{L}\sum_{j=1}^{N_h}(\mu_h-\mu)^2}{N}=\frac{\sum_{h=1}^{L}N_h\sigma_h^2+\sum_{h=1}^{L}N_h(\mu_h-\mu)^2}{N}$$

$$=\sum_{h=1}^{L}W_h\sigma_h^2+\sum_{h=1}^{L}W_h(\mu_h-\mu)^2 \quad \text{得證}$$

(二) $$\bar{y}_{st}=\sum_{h=1}^{L}W_h\bar{y}_h$$

假設各層之劃分彼此獨立，使得 $\bar{y}_i \perp \bar{y}_j, i\neq j$

(三)採用事後分層抽樣法的時候，研究者不需要事先決定配置方式，原因是事後分層抽樣法是先從抽樣母體中以簡單隨機抽樣法抽出n個，然後在這n個樣本中屬於每一層之樣本數就已經固定了，根本不需要研究者自行配置。

(四)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。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。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。

三、某水果商訂購了一卡車的椰子，並打算採用比值估計（ratio estimation）之方法來推估整輛卡車裝載所有的椰子如果全數被剖開之後的椰子汁總重量。該水果商採用不置回的簡單隨機抽樣之方法從整輛卡車的椰子中抽出了10顆椰子。試說明該水果商應該如何利用被抽出的10顆椰子來取得樣本數據以及如何進行比值估計，並寫出點估計量的數學式。（15分）

答：

假設椰子重量與椰子汁之重量具有高度相關，因此可以椰子之重量作為輔助變數提供估計椰子汁總重量（Y）之有用訊息。

該水果商應該先收集被抽出10顆椰子之椰子重量（x_i）與椰子汁重量（y_i），再利用地磅得到卡車重量，將之扣除卡車淨重後可得所有椰子之總重量（X）。

椰子汁總重量（Y）之點估計與標準誤估計量分別為

點估計$\hat{Y}_r = rX = N\bar{y}_r$，其中$\bar{y}_r = r\bar{X}$與$r = \frac{y}{x} = \frac{\bar{y}}{\bar{x}}$

點估計標準誤之估計值

$$s_{\hat{Y}r} = Xs_r = \frac{X}{\bar{X}}\sqrt{1-f}\,\frac{s_d}{\sqrt{n}} = N\sqrt{1-f}\,\frac{s_d}{\sqrt{n}} = Ns_{\bar{y}_r}$$

其中$s_d^2 = \frac{\sum_{i=1}^{n}(y_i - rx_i)^2}{n-1} = s_Y^2 + r^2 s_X^2 - 2rr' s_X s_Y$

s_X^2, s_Y^2分別為 x,y 之樣本變異數

r'為 x,y 之樣本相關係數

四、迴歸分析（regression analysis）課程中的最小平方估計量（least squares estimator），在抽樣方法的理論中有何可用之處？請詳細說明並寫出相關的估計量數學式。（10分）

答：

迴歸分析是以自變數（X）預測應變數（Y）之統計方法，這樣的觀念可以應用在抽樣方法中，也就是說利用一個與應變數相關的已知變數(輔助變數)製造出一些額外訊息以提升估計之精確度，此為抽樣理論中之迴歸估計法。我們以簡單隨機抽樣法為例說明如何進行迴歸估計法，如下：

欲估計母體平均數$\bar{Y}$

點估計$\bar{y}_{lr} = \bar{y} + b_1(\bar{X} - \bar{x})$

點估計標準誤之估計值

$$s_{\bar{y}_{lr}} = \sqrt{1-f}\,\frac{s_e}{\sqrt{n}} = \sqrt{1-f}\,\frac{\sqrt{MSE}}{\sqrt{n}}$$

其中$s_e^2 = MSE = \frac{\sum_{i=1}^{n}[(y_i - \bar{y}) + b_1(x_i - \bar{x})]^2}{n-2} = \frac{ss_Y - b_1^2 ss_X}{n-2}$

五、(一)某一間大學的學術部門打算採用二階段集群抽樣（two-stage cluster sampling）之方法來推估校內全體專任教師在前一個學年度的學術論文平均產量（亦即，推估任何一位專任教師在前一個學年度的學術論文平均發表篇數），並且將校內的每一個科系視為一個集群。試就比值估計（ratio estimation）以及不偏估計（unbiased estimation）兩種情況分別說明如何進行二階段集群抽樣以及如何推估，並寫出點估計量的數學式。（13分）

(二)假設前一小題之中的學術部門打算改變為採用集群柚樣（cluster sampling）之方法來推估校內全體專任教師在前一個學年度的學術論文平均產量，並且依舊將校內的每一個科系視為一個集群。試就比值估計（ratio estimation）以及不偏估計（unbiased estimation）兩種情況分別說明如何進行集群抽樣以及如何推估，並寫出點估計量的數學式。（13分）

[註]集群抽樣又稱之為單階段集群抽樣（single-stage cluster sampling）

答：

令 $\overline{\overline{Y}}$ 表全體專任教師在前一個學年度學術論文之平均產量

(一)二階段群集抽樣法

1.抽樣程序：

假設校內有N個科系，每一個科系有 M_i 位專任教師，第一階段抽樣是由N個科系中隨機抽出n個，第二階段是從被抽出之n個科系中的每一個科系隨機抽出若干位專任教師，並記錄其前一個學年度學術論文之篇數。

2.比值估計法：

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)，母體平均數之點估計為 $\hat{\overline{\overline{Y}}} = \frac{\sum_{i=1}^{n} M_i \bar{y}_i}{n\frac{\sum_{i=1}^{N} M_i}{N}} = \frac{\sum_{i=1}^{n} M_i \bar{y}_i}{n\overline{M}}$

3.不偏估計法：

不偏估計法也稱為簡單估計法，是一種沒有考慮輔助變數之估計方法，二階段群集抽樣法中母體平均數之點估計為

$$\hat{\overline{\overline{Y}}} = \frac{1}{M}\hat{Y} = \frac{N}{Mn}\sum_{i=1}^{n} M_i \bar{y}_i$$

(二)單階段群集抽樣法中

1.抽樣程序：

假設校內有N個科系，每一個科系有 M_i 位專任教師，由N個科系中隨機抽出n個，被抽出之n個科系中所有專任教師均記錄其前一個學年度學術論文之篇數

2.比值估計法：

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)，母體平均數之點估計為

$$\bar{y}_{rcl} = \frac{\bar{y}_t}{\bar{m}} \qquad \text{其中 } \bar{y}_t = \frac{1}{n}\sum_{i=1}^{n} y_i, \ \bar{m} = \frac{1}{n}\sum_{i=1}^{n} M_i$$

3.不偏估計法：

不偏估計法也稱為簡單估計法，是一種沒有考慮輔助變數之估計方法，單階段群集抽樣法中母體平均數之點估計為

$$\bar{y}_{\mathrm{cl}} = \frac{N}{M}\bar{y}_{\mathrm{t}} \qquad \text{其中 } \bar{y}_t = \frac{y}{n}$$

六、在什麼情況之下，使用系統抽樣（systematic sampling）之方法來推估母體的平均數，其推估之效果會優於使用簡單隨機抽樣的推估效果？請詳細說明。（5分）

答：

若母體變數呈現「直線趨勢性(順序性)」，即可預期系統抽樣法比簡單隨機抽樣法之精確度佳，其原因是在直線趨勢性之下，系統樣本內之變異數會與母體變異數具有較大之差異性。