《抽樣方法》

一、請敘述問卷設計之原則。（15分）

試題評析	本題只要把課本內容寫到考卷上就可以獲得高分。
考點命中	《高點．高上抽樣方法講義》第一回第一章，趙治勳編撰。

答：

(一)問卷設計的品質著重於研究者之經驗與技術，以下為問卷設計之步驟：

(1)決定所要蒐集之資訊　(2)決定問卷之類型　(3)決定問項的內容

(4)決定問項的型式　(5)決定問項的用語　(6)決定問項的先後順序

(7)決定檢驗可靠度的問項　(8)決定問卷版面之布局　(9)試查

(10)修訂及定稿　(11)正式調查

(二)注意事項：

(1)在語意上會引導受訪者傾向於假設某種結論的問題

(2)易引起誤解之問題

(3)混淆不清

(4)重複問題

(5)無關旨意之問題

(6)問題不要太長

(7)避免使用複句及形容詞，盡量使用肯定句

(8)避免下列問題：

(a)記憶性太多

(b)太普遍化問題

(c)需要太多自我分析的問題

(d)涉及社會地位與名譽

二、(一)何謂簡單隨機抽樣法？（5分）

(二)設一有限母體中有N個元素，若自母體中以抽出不放回的抽樣方式，抽出n個簡單隨機樣本，$y_1, y_2, ..., y_n$，請證明樣本平均數 $\bar{y}=\frac{\sum_{i=1}^{n} y_i}{n}$ 的變異數 $Var(\bar{y})=\frac{N-n}{N-1}\cdot\frac{\sigma^2}{n}$ （σ^2 為母體變異數）。（10分）

試題評析	本題只要把課本內容寫到考卷上就可以獲得高分。
考點命中	《高點．高上抽樣方法講義》第一回第二章，趙治勳編撰。

答：

(一)簡單隨機抽樣法定義：每一組可能的樣本均有相等的被抽出機率。

(二)

令 $u_i = \begin{cases} 1, \text{母體中第i單位出現在樣本中} \\ 0, \text{o.w.} \end{cases}$

$$E(u_i)=1\times\frac{n}{N}=\frac{n}{N} \qquad E(u_i^2)=1^2\times\frac{n}{N}=\frac{n}{N} \qquad E(u_iu_j)=1\times1\times\frac{n}{N}\frac{n-1}{N-1}=\frac{n(n-1)}{N(N-1)}$$

$$V(u_i)=\frac{n(N-n)}{N^2} \qquad Cov(u_i,u_j)=-\frac{n(N-n)}{N^2(N-1)}$$

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i=\frac{1}{n}\sum_{i=1}^{N}u_iY_i$$

$$V(\bar{y})=\frac{1}{n^2}[\sum_{i=1}^{N}V(u_i)Y_i^2+\sum_{i\neq j=1}^{N}\sum Cov(u_i,u_j)Y_iY_j]$$

$$=\frac{1}{n^2}[\sum_{i=1}^{N}\frac{n(N-n)}{N^2}Y_i^2+\sum_{i\neq j=1}^{N}\sum(-\frac{n(N-n)}{N^2(N-1)})Y_iY_j]$$

$$=\frac{1}{n^2}\frac{n(N-n)}{N}[\frac{1}{N}\sum_{i=1}^{N}Y_i^2-\frac{1}{N(N-1)}\sum_{i\neq j=1}^{N}\sum Y_iY_j]$$

$$=\frac{1}{n}\frac{N-n}{N}\frac{(N-1)\sum_{i=1}^{N}Y_i^2-\sum_{i\neq j=1}^{N}\sum Y_iY_j}{N(N-1)}=\frac{N-n}{N-1}\frac{\sigma^2}{n}$$

其中 $\sigma^2=\frac{\sum_{i=1}^{N}(Y_i-\bar{Y})^2}{N}=\frac{\sum_{i=1}^{N}Y_i^2-\frac{(\sum_{i=1}^{N}Y_i)^2}{N}}{N}=\frac{\sum_{i=1}^{N}Y_i^2-\frac{\sum_{i=1}^{N}Y_i^2+\sum_{i\neq j=1}^{N}\sum Y_iY_j}{N}}{N}$

$$=\frac{(N-1)\sum_{i=1}^{N}Y_i^2-\sum_{i\neq j=1}^{N}\sum Y_iY_j}{N^2}$$

三、(一)臺灣自來水公司為了研究家戶用水行為，欲從家庭用水的水表資料檔中以抽樣方式抽出1,000戶來做家戶用水行為研究。在自來水公司的家庭水表資料檔中，各戶都有一個專門的水表號碼，也有每個月的用水度數。為了減低用水量多寡的用水行為差異，欲用用水度數分層，請敘述如何執行分層隨機抽樣。(5分)

(二)分層隨機抽樣之每一層樣本大小的配置(Allocation)，應注意那三項事項？(5分)

(三)什麼時候需要採用雙重抽樣方法(Double Sampling)？(5分)

試題評析	(二)(三)兩小題與考古題完全相同，第一小題於以往考古題中也常出現，只要有加強練習過考古題的考生應該能夠輕鬆獲得滿分。
考點命中	《高點・高上抽樣方法講義》第一回第五、七章，趙治勳編撰。

答：

(一)在以用水度數分層時，應該考量到分層標準是否滿足「層間變異大，層內變異小」，決定劃分各層用水度數後，再按照各層戶數比例計算出樣本中之樣本數，盡可能提高樣本與母體之相似度而降低變異。

(二)1.抽樣單位個數 N_h 越大的層應該抽出較多樣本

2.母體標準差 S_h 越大的層應該抽出較多樣本

3.每單位調整成本 c_h 越小的層應該抽出較多樣本

(三)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。

四、(一)若採用重複系統抽樣法（Repeated Systematic Sampling），在母體大小為N中抽出n_s個“k'取1”的系統樣本，可得n_s個樣本大小為n（$n=\frac{N}{k'}$）的系統樣本。若第i個系統樣本為y_i, $y_{i+k'}$, ..., $y_{i+(n-1)k'}$，令$\overline{y_i}=\frac{\sum_{j=1}^{n} y_{ij}}{n}$，（i=1, 2, ..., n_s）為第i個系統樣本的平均數，請寫出母體平均數$\hat{\mu}$的估計公式？及母體平均數估計式的估計變異數$\widehat{Var}(\hat{\mu})$？（10分）

(二)若採用群集隨機抽樣法（Cluster Sampling），在母體大小為M中先分成N個群集（cluster）（每一個群集之個數為m_1, m_2, ..., m_N，$M=\sum_{i=1}^{N} m_i$），再以群集為抽樣單位，以簡單隨機抽樣法抽出n個群集為一群集樣本，稱為群集隨機樣本。若y_i表示第i個群集中變數值的總和，請寫出母體平均數$\hat{\mu}$的估計公式？及母體平均數估計式的估計變異數$\widehat{Var}(\hat{\mu})$？（10分）

(三)在何種條件下，題(一)及(二)所述兩種抽樣法之母體平均數的估計值公式會相同？（5分）

試題評析	(一)(二)小題都只要列出估計式而已，不需要計算，考生應該能夠輕鬆獲得滿分；僅(三)小題需要考生熟悉抽樣方法間之關聯性，然而講義中已經有敘述過系統抽樣法與群集抽樣法之關聯性，只要考生加以推廣，就可以正確地回答。
考點命中	《高點・高上抽樣方法上課講義》第一回第四、八章，趙治勳編撰。

答：

(一) $\hat{\mu}=\overline{y}_{rsy}=\frac{1}{n_s}\sum_{i=1}^{n_s}\overline{y}_i$

$$\hat{V}(\hat{\mu})=(1-f)\frac{s_{re}^2}{n_s} \quad 其中 s_{re}^2=\frac{\sum_{i=1}^{n_s}(\overline{y}_i-\overline{y}_{rsy})^2}{n_s-1} \quad ,f=\frac{n}{N}$$

(二) $\hat{\mu}=\overline{y}_{cl}=\frac{N}{M}\overline{y}_t$　其中 $\overline{y}_t=\frac{y}{n}=\frac{\sum_{i=1}^{n} y_i}{n}$

$$\hat{V}(\hat{\mu})=\frac{N^2}{M^2}(1-f)\frac{s_t^2}{n} \quad 其中 s_t^2=\frac{\sum_{i=1}^{n}(y_i-\overline{y}_t)^2}{n-1}$$

(三)1.重複系統抽樣法中k'等於群集抽樣法中N

2.重複系統抽樣法中n_s等於群集抽樣法中n

3.群集抽樣法中各群集之抽樣單位數均等$m_1=m_2=\cdots=m_N=\frac{N}{k'}$,且群集內之之觀察值為 $y_i, y_{i+k'}, \cdots, y_{i+(n-1)k'}$

五、在兩階段抽樣法中，第一階段先將母體分成N個抽樣單位，以簡單隨機抽樣法抽出n個抽樣單位，稱為第一抽樣單位。其次，將第一階段中的每一個抽樣單位分割成M_i(i=1, 2, ..., N)個抽樣單位，並從n個第一抽樣單位以簡單隨機抽樣法各抽出m_i(i=1, 2, ..., n)個抽樣單位（y_{i1}, y_{i2}, ..., y_{im_i}, i=1, 2, ..., n），稱為第二抽樣單位，$M=\sum_{i=1}^{N} M_i$，$\overline{M}=\frac{M}{N}$，$\overline{y_i}=\frac{1}{m_i}\sum_{j=1}^{m_i} y_{ij}$。

(一)在什麼情況下，兩階段抽樣法即為分層隨機抽樣法？（5分）
(二)在什麼情況下，兩階段抽樣法即為群集隨機抽樣法？（5分）
(三)若M未知，請寫出母體平均數$\hat{\mu}$的估計公式及$\widehat{Var}(\hat{\mu})$的估計公式。(10 分)
(四)若在第一階段中不以簡單隨機抽樣法抽出n個抽樣單位，改以第i（i=1, 2, ..., N）個被抽中的機率為$\frac{M_i}{M}$（Two-Stage Cluster Sampling with Probabilities Proportional to Size, pps），請寫出母體平均數$\hat{\mu}$的估計公式及$\widehat{Var}(\hat{\mu})$的估計公式。(10 分)

試題評析	(一)(二)小題需要考生熟悉抽樣方法間之關聯性，然課堂中已經有敘述過兩階段抽樣法與分層、群集間之關聯性。
考點命中	《高點・高上迴歸分析熱門題庫》第一回第十七、十八章，趙治勳編撰。

答：

(一)兩階段抽樣法中於第一階段時從 N 個群集中抽出 N 個出來($N=n$)，此時兩階段抽樣法即為分層隨機抽樣法。

(二)兩階段抽樣法中於第二階段時並沒有進行 SRS 抽樣程序，此時兩階段抽樣法即為群集隨機抽樣法。

(三)當 M 未知下，研究者可以$\hat{M}=N\bar{M}$估計，此時$\hat{\mu}=\hat{\bar{\bar{Y}}}=\frac{N}{N\bar{M}n}\sum_{i=1}^{n}M_i\bar{y}_i=\frac{1}{\bar{M}n}\sum_{i=1}^{n}M_i\bar{y}_i$

$$\hat{V}(\hat{\mu})=\frac{1}{N^2\bar{M}^2}[N^2(1-f_1)\frac{s_{1b}^2}{n}+\frac{N}{n}\sum_{i=1}^{n}M_i^2(1-f_{2i})\frac{s_{2i}^2}{m_i}]$$

其中 $f_1=\frac{n}{N}, f_{2i}=\frac{m_i}{M_i}$

$$s_{1b}^2=\frac{1}{n-1}\sum_{i=1}^{n}(\hat{Y}_i-\bar{\hat{Y}}_t)^2, s_{2i}^2=\frac{1}{m_i-1}\sum_{j=1}^{m_i}(y_{ij}-\bar{y}_i)^2$$

(四)

$$\hat{\mu}=\hat{\bar{\bar{Y}}}_{pps}=\frac{1}{n}\sum_{i=1}^{n}\bar{y}_i=\bar{\bar{y}}$$

$$\hat{V}(\hat{\mu})=\frac{1}{n(n-1)}\sum_{i=1}^{n}(\bar{y}_i-\bar{\bar{y}})^2$$