統計行政

《抽樣方法》

試題評析
今年抽樣方法除了第四題比較難之外，其他三題尚屬基本題型。第一題屬於簡單隨機樣法的基本題型，第二題屬於比值估計法的重點題型，第三題則屬分層抽樣法的基本證明題。雖然有不少基本重點題型，平時若無熟練，臨場恐怕不易寫出完整的答案。因此預估一般考生約能拿40分左右；程度較高者，可拿60分以上。

一、假設母體總數$N=15000$，考慮以下放回簡單隨機抽樣方式（簡稱為SRS）自母體中抽出$n=15$個樣本。

(一)試問樣本空間中有多少種不同的結果。（五分）

(二)依據(一)說明滿足何種條件則稱之為SRS。（五分）

(三)依據(二)證明母體中任何一個體在樣本中的機率為$\frac{1}{1000}$。（十分）

答：

(一)樣本空間中共有$C_n^N=C_{15}^{15000}$種不同的結果。

(二)設母體為U，$S=\{s\mid\varnothing\neq s\subset U\}$，稱每一個S中的元素s為U的一組樣本。

令$Sn=\{s\in S\mid s$含有n個抽樣單位$\}$。定義一個機率函數$P: S\to[0,1]$如下

$$P(s)=\begin{cases}\frac{1}{C_n^N}, & 如果 s\in S_n\\ 0, & 如果 s\notin S_n\end{cases}$$

則P為一個抽樣設計，並稱之為SRS。

(三)P（母體中第i個個體在樣本中）$=\frac{C_{n-1}^{N-1}}{C_n^N}=\frac{n}{N}=\frac{15}{15000}=\frac{1}{1000}$

☆參考資料：高上《抽樣方法上課講義》第一回**P10~12**。

二、假設從全國250所大學中，以SRS抽樣抽出50所大學。學生人數（表示為Yi）和老師人數（表示為Xi）如下所示：

樣本數（n）	$\sum_{i=1}^{50}Y_i$	$\sum_{i=1}^{50}X_i$	$\sum_{i=1}^{50}Y_i^2$	$\sum_{i=1}^{50}X_iY_i$	$\sum_{i=1}^{50}X_i^2$
50	30.2×10^3	2.02×10^3	30.1×10^6	1.7×10^6	0.11×10^6

(一)利用上述資料說明如何估計$R=\frac{N_S}{N_T}$，其中N_S和N_T分別為250所大學學生和老師總人數。（十分）

(二)說明(一)所提出的估計值是否為R的不偏估計。（十分）

(三)利用上述資料建構R的95%區間估計。（十分）

答：

(一) $\hat{R}=\frac{\sum_{i=1}^{50}Y_i}{\sum_{i=1}^{50}X_i}=\frac{30.2\times10^3}{2.02\times10^3}=14.94$

(二) $\because E(\hat{R})=R-\frac{Cov(\hat{R},\bar{X})}{\bar{N}_T}$，其中 $\bar{X}=\frac{1}{50}\sum_{i=1}^{50}X_i$，$\bar{N}_T=\frac{N_T}{250}$。

$\therefore\hat{R}$不是R的不偏估計量。

(三) $S_Y^2=\frac{1}{n-1}\left[\sum_{i=1}^{n}Y_i^2-\frac{(\Sigma Y_i)^2}{n}\right]=\frac{1}{50-1}\left[30.1\times10^6-\frac{(30.2\times10^3)^2}{50}\right]=242024.49$

$S_X^2=\frac{1}{n-1}\left[\sum_{i=1}^{n}X_i^2-\frac{(\Sigma X_i)^2}{n}\right]=\frac{1}{50-1}\left[0.11\times10^6-\frac{(2.02\times10^3)^2}{50}\right]=579.43$

$$S_{YX}=\frac{1}{n-1}\left[\sum_{i=1}^{n}X_iY_i-\frac{(\sum X_i)(\sum Y_i)}{n}\right]$$

$$=\frac{1}{50-1}\left[1.7\times10^6-\frac{(2.02\times10^3)(30.2\times10^3)}{50}\right]=9794.29$$

$$v(R)=\frac{1}{(\bar{X}^2)}\cdot\frac{(1-f)}{n}\cdot[S_Y^2-2(R)S_{YX}+(\hat{R})^2\ S_X^2]$$

$$=\frac{1}{\left(\frac{2.02\times10^6}{50}\right)^2}\cdot\frac{\left(1-\frac{50}{250}\right)}{50}\cdot[242024.49-2(14.94)(9794.29)+(14.94)^2(579.43)]$$

$$=\frac{8}{(2.02\times10^6)^2}\cdot[78701.97]=0.1543\times10^{-6}$$

R的95%信賴區間為

$$\hat{R}\pm z_{0.025}\cdot\sqrt{v(R)}\Rightarrow14.94\pm1.96\times\sqrt{0.1543\times10^{-6}}\Rightarrow[14.9397,14.9403]$$

☆參考資料：高上《抽樣方法上課講義》第二回p.7-3、7-5。

三、假設母體分為兩層。令N_1和N_2分別表示第一層和第二層已知之母體總數，令σ_1^2和σ_2^2分別表示第一層和第二層之母體變方（σ_1^2和σ_2^2未知）。考慮以分層抽樣方式自母體兩層中分別以SRS抽出n1和n2個樣本並 $\bar{y}_{st}=\sum_{h=1}^{2}w_h\bar{y}_h$ 估計母體平均值，其中$\bar{y}_1$和$\bar{y}_2$分別表示第一層和第二層樣本平均值，$w_h=\frac{N_h}{N}(h=1,2)$，N ＝ N_1 ＋ N_2。我們希望在給定的預算c下選擇n_1和n_2使得$\bar{y}_{st}$的變方$\mathrm{Var}(\bar{y}_{st})$最小。

假設$c=c_0+\sum_{h=1}^{2}c_hn_h$，其中$c_0$和ch（h ＝ 1，2）為已知之正數。

(一)當c1＝c2時，說明如何選擇n1和n2？（十分）

(二)由於σ_1^2和σ_2^2未知，因此(一)所選擇的n1和n2在實際應用上有何困難？（五分）

(三)另一簡單之樣本數配置方法稱之為比例配置。何謂比例配置並說明在何種情形下適用。（十分）

答：

(一)$\because \mathrm{Var}(\bar{y}_{st})=\sum_{h=1}^{2}w_h^2\cdot\frac{\sigma_h^2}{n_h}\cdot(1-f_h)=\sum_{h=1}^{2}w_h^2\cdot\frac{\sigma_h^2}{n_h}-\frac{\sum_{h=1}^{L}w_h\cdot\sigma_h^2}{N}$

$\therefore$欲求n_h使$\mathrm{Var}(\bar{y}_{st})$最小，等價於使$\sum_{h=1}^{2}w_h^2\cdot\frac{\sigma_h^2}{n_h}$最小。

由Cauchy-Schwarz不等式得知

$$\left(\sum_{h=1}^{2}c_hn_h\right)\cdot\left(\sum_{h=1}^{2}w_h^2\cdot\frac{\sigma_h^2}{n_h}\right)=\left[\sum_{h=1}^{2}(\sqrt{c_hn_h})^2\right]\cdot\left[\sum_{h=1}^{2}\left(\frac{w_h\sigma_h}{\sqrt{n_h}}\right)^2\right]$$

$$\geq\sum_{h=1}^{2}\left(\sqrt{c_hn_h}\cdot\frac{w_h\sigma_h}{\sqrt{n_h}}\right)^2=\sum_{h=1}^{2}(\sqrt{c_h}\cdot w_h\sigma_h)^2$$

而上述不等式之等號成立之充要條件為

$$\frac{\sqrt{c_1n_1}}{\frac{w_1\sigma_1}{\sqrt{n_1}}}=\frac{\sqrt{c_2n_2}}{\frac{w_2\sigma_2}{\sqrt{n_2}}}\Leftrightarrow\frac{n_1}{\frac{w_1\sigma_1}{\sqrt{c_1}}}=\frac{n_2}{\frac{w_2\sigma_2}{\sqrt{c_2}}}\Leftrightarrow\frac{n_1}{\frac{N_1\sigma_1}{\sqrt{c_1}}}=\frac{n_2}{\frac{N_2\sigma_2}{\sqrt{c_2}}}$$

當$c_1=c_2$時，可求得$\frac{n_1}{n_2}=\frac{N_1\sigma_1}{N_2\sigma_2}$。

(二)若σ_3^2和σ_2^2未知，則(一)所選擇的n_1和n_2無法求得。

(三)比例配置（proportional allocation）係採用各層的抽樣數與該層含有的個數成正比，其公式為$\frac{n_1}{n_2}=\frac{N_1}{N_2}$。此法適用於每層含有的個數相差大，且層間變異數相差不大的情況。

☆參考資料：高上《抽樣方法上課講義》第一回P10~12、p37、p.38、p.44。

四、今欲估計某一製鞋工廠在某一天所生產鞋子不良品所佔之比例（表示為P）。工廠共有24個生產線（$i = 1, \cdots; 24$）。第i個生產線每天生產M_i雙鞋子。考慮以比例集群抽樣方式（即第i個生產線被抽中之機率和M_i成正比）抽出一個樣本。

(一)說明如何利用均勻分配U(0, 1)亂數抽出一隨機樣本。（十分）

(二)假設抽中第3個生產線，$M_3 = 100$，不良品個數為2個。說明如何估計P。（十分）

(三)在實際應用上，當M_i太大時，無法每件檢查。有何處理方法？（五分）

答：

(一)假設母體之數值為$\{Y_1, \cdots; Y_N\}$，欲以簡單隨機抽樣法抽出樣本$\{y_1, \cdots; y_n\}$，則其電腦程式演算法則如下：

step 1：$N_1 = N$，$n_1 = n$

step 2：$i = INT(N_1 \times U) + 1$，U為由U(0, 1)分配抽出之一亂數

step 3：取Yi為樣本

step 4：若$n_i = 1$，則抽樣完成，否則執行step 5

step 5：若$i = N_i$，則執行step 7，否則執行step 6

step 6：將Y_{Ni}取代Yi

step 7：$N_1 = N_1 - 1$，$n_1 = n_1 - 1$，回到step 2

(二)令 $M = \sum_{i=1}^{n} M_i$，則

$$\hat{P}_{pps} = \frac{1}{M} \times \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{Z_i} = \frac{1}{M} \times \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{\frac{M_i}{M}} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{M_i} = \frac{1}{n}\sum_{i=1}^{n} p_i = \frac{1}{I} \times \frac{2}{100} = 2\%$$

(三)改從M_i個產品中抽出m_i個檢查，再用這m_i個產品的不良率代入上述公式中的p_i。

☆參考資料：高上《抽樣方法上課講義》第二回 p.5-12。