104 年特種考試地方政府公務人員考試試題

等 別：三等考試
類 科：統計
科 目：抽樣方法

一、假設 $F=\{u_1,u_2,u_3,u_4\}$ 是一個緊緊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？

【擬答】：

⑴簡單隨機抽樣為抽後不放回，所以共可形成 $C_2^4=6$ 種樣本組合。

⑵機率均等隨機置回抽樣，可形成 $4^2=16$ 種樣本組合，但在此我們並不考慮抽出的順序，所以此方式共有 10 種樣本組合。

(3) $\mu=\dfrac{Y_1+Y_2+\ldots+Y_N}{N}=\dfrac{1+3+3+9}{4}=4$

(4) $\sigma^2=\dfrac{\sum_{i=1}^{N}(Y_i-\mu)^2}{N}=\dfrac{(1-4)^2+(3-4)^2+(3-4)^2+(9-4)^2}{4}=9$

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：

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

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

㈠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)。

㈡假設我們改用機率隨機置回抽樣，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。

【擬答】：

1. (1) $\hat{\sigma}^2_{(1)}=\dfrac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})^2$

單位編號	$\hat{\sigma}^2_{(1)}$	機率
{1, 2}	1	1/6
{1, 3}	1	1/6
{1, 4}	16	1/6
{2, 3}	0	1/6
{2, 4}	9	1/6
{3, 4}	9	1/6

所以抽樣分配如下

$\hat{\sigma}^2_{(1)}$	0	1	9	16

$f(\hat{\sigma}^2_{(1)})$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{1}{6}$

$E(\hat{\sigma}^2_{(1)})=0\times\frac{1}{6}+1\times\frac{2}{6}+9\times\frac{2}{6}+16\times\frac{1}{6}=6\neq\sigma^2$，為有偏估計量

$MSE_{(1)}=bias^2+Var\left(\hat{\sigma}^2_{(1)}\right)=(6-9)^2+34=43$

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

單位編號	$\hat{\sigma}^2_{(2)}$	機率
{1, 2}	2	1/6
{1, 3}	2	1/6
{1, 4}	32	1/6
{2, 3}	0	1/6
{2, 4}	18	1/6
{3, 4}	18	1/6

所以抽樣分配如下

$\hat{\sigma}^2_{(2)}$	0	2	18	32
$f(\hat{\sigma}^2_{(2)})$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{1}{6}$

$E(\hat{\sigma}^2_{(2)})=0\times\frac{1}{6}+2\times\frac{2}{6}+18\times\frac{2}{6}+32\times\frac{1}{6}=12\neq\sigma^2$，為有偏估計量

$MSE_{(2)}=bias^2+Var\left(\hat{\sigma}^2_{(2)}\right)=(12-9)^2+136=145$

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

單位編號	$\hat{\sigma}^2_{(3)}$	機率
{1, 2}	1.5	1/6
{1, 3}	1.5	1/6
{1, 4}	24	1/6
{2, 3}	0	1/6
{2, 4}	13.5	1/6
{3, 4}	13.5	1/6

所以抽樣分配如下

$\hat{\sigma}^2_{(3)}$	0	1.5	13.5	24
$f(\hat{\sigma}^2_{(3)})$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{1}{6}$

$E(\hat{\sigma}^2_{(3)})=0\times\frac{1}{6}+1.5\times\frac{2}{6}+13.5\times\frac{2}{6}+24\times\frac{1}{6}=9=\sigma^2$，為不偏估計量

$MSE_{(3)}=bias^2+Var\left(\hat{\sigma}^2_{(3)}\right)=(9-9)^2+76.5=76.5$

2.採用機率均等隨機置回抽樣時，$E(\hat{\sigma}^2_{(2)})=\sigma^2$ 為不偏估計量

而 $E(\hat{\sigma}^2_{(1)})=\frac{n-1}{n}\sigma^2$；$E(\hat{\sigma}^2_{(3)})=\frac{N-1}{N}\sigma^2$

所以 $\hat{\sigma}^2_{(1)}$ 與 $\hat{\sigma}^2_{(3)}$ 皆為有偏估計量

三、假設某一間大學共有 8000 位學生，其中 2000 人為男生，6000 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。

㈠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 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，如果我們想用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？

㈡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。

㈢先前估計 8000 位學生之平均身高問題，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 層中小樣本，萬一發生空事後層(empty post-stratum)之情形，應當如何處理或補救？

【擬答】：

*1.*⑴若欲估計 8000 名學生的平均身高，應以性別來作分層，因為分層隨機抽樣適用於層內同質性高，層外異質性高的母體型態，故性別之間的變異較高，故應採用男女性別作為分層。

⑵採用比例配置 $n_h = \dfrac{N_h}{\sum_{h=1}^{L} N_h} \times n$

故 $n_{男} = \dfrac{2000}{8000} \times 16 = 4$ ，男生抽取 4 人

$n_{女} = \dfrac{6000}{8000} \times 16 = 12$ ，女生抽取 12 人

⑶若欲估計 8000 名學生的平均收入，應採用日間、夜間部來作分層，因為日間部與夜間部的收入變異較大，即同樣日間部的學生收入或同樣夜間部的學生收入應同質性較高。故應採用日間、夜間部作為分層。

*2.*假設男生身高變異數為 $s_1^2 = 1.96k$ ，男生身高變異數為 $s_2^2 = k$

採用紐門配置 $n_h = \dfrac{N_h S_h}{\sum_{h=1}^{L} N_h S_h} \times n$

$$\sum_{h=1}^{2} N_h S_h = 2000 \times \sqrt{1.96k} + 6000 \times \sqrt{k} = 8800\sqrt{k}$$

故 $n_{男} = \dfrac{2000 \times \sqrt{1.96k}}{8800\sqrt{k}} \times 16 = 5.09$ ，男生抽取 5 人

$n_{女} = \dfrac{6000 \times \sqrt{k}}{8800\sqrt{k}} \times 16 = 10.91$ ，女生抽取 11 人

*3.*⑴在實務上的抽樣問題，有時並無完整的分層母體資訊，我們只能在樣本被選取之後，才可能設定抽樣單位到正確的層，但此時我們若僅瞭解各分層的配置，在一組樣本選取後再進行分層，即事後分層。以本例來說，當我們不知道此大學男女實際人數時，僅可就過去的經驗大致推估母體的性別比例後，再作分層估計，即

$$\bar{y}_{pst} = \sum_{h=1}^{L} A_h \cdot \bar{y}_h = \frac{N_{男}}{N}\bar{y}_{男} + \frac{N_{女}}{N}\bar{y}_{女}$$

但當我們亦缺乏此資訊時，便僅能以合理的男性女性比例 1：1 來作估計。即

$$\bar{y}_{pst} = \sum_{h=1}^{L} A_h \cdot \bar{y}_h = 0.5 \times \bar{y}_{男} + 0.5 \times \bar{y}_{女}$$

⑵事後分層變異數為

$$\mathrm{var}(\bar{y}_{pst}) \approx \left(\frac{N-n}{Nn}\right)\sum_{i=1}^{L} A_h \cdot s_h^2 + \frac{1}{n^2}\sum_{i=1}^{L}(1-A_h)\cdot s_h^2$$

其中第一項是從一組比例配置的分層樣本的平均數中所獲得的變異數，所以事後分層可以視為比例配置。

⑶不論分層隨機抽樣或事後分層抽樣，皆不允許某個分層為空的分層。但事後分層的抽樣過程中，卻很容易因為層數過多而樣本數 卻不甚大，就很可能造成某些層為零。處理這個問題的方法，最常用的就是合併鄰層法，也就是將空事後層與鄰近層合併以減少總層數的方法。合併鄰層後，兩層的權重放大，並以鄰層的樣本平均數作為合併兩層的平均數來作後續的事後分層估計。

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

【擬答】：

雙重抽樣是指在抽樣時分兩次抽取樣本的一種抽樣方式。首先抽取一個初步樣本，並搜取一些簡單專案以獲得有關總體的資訊；然後，在此基礎上再進行深入抽樣。

1. 分層雙重抽樣是在分層隨機抽樣法中執行雙重抽樣的抽樣法。首先第一重是從抽樣母體中，採用簡單隨機抽樣抽出第一重抽樣單位 n'，其分布在不同的分層，各分層大小為 n_h'。然後再從第一重抽樣單位每一層 n_h' 中，以簡單隨機抽樣法分別抽出第二重抽樣單位 n_h，計算第二重各層樣本平均數 $\bar{y}_h$ 作分層估計。

$$\bar{y}_{st} = \sum g_h \bar{y}_h = \sum \frac{n_h'}{n'} \cdot \bar{y}_h \xrightarrow{\text{估計}} \bar{Y}$$

2. 比值估計雙重抽樣是在比值估計法中執行雙重抽樣的抽樣法。首先第一重是從抽樣母體中，採用簡單隨機抽樣抽出 n' 個配對資料作為第一重樣本，並以第一重樣本中的樣本平均數 $\bar{x}'$ 估計輔助變數的母體平均數 $\bar{X}$ 然後再從每一層第一重抽樣單位中以簡單隨機抽樣法分別抽出第二重抽樣單位做比率估計。

$$\bar{y}_r = r\bar{x}' = \frac{\bar{y}}{\bar{x}}\bar{x}' \xrightarrow{\text{估計}} \bar{Y}$$

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。

【擬答】：

比值集群抽樣法即是集群抽樣法中執行比值估計的方法，而二階段比值集群抽樣法即是二階段集群抽樣法中執行比值估計的方法。比值估計法最主要的想法是採用輔助變數來增加估計的準確度，所以在這兩種方法下，皆是將各個集群的母體大小 M_i 視為輔助變數，用以提升推估集群母體參數準確度的估計方法。

1. 比值集群抽樣法平均數估計

$$\bar{y}_{rc} = \frac{Y_1 + Y_2 + \cdots + Y_n / n}{M_1 + M_2 + \cdots + M_n / n} = \frac{\bar{y}_t}{\bar{M}} \xrightarrow{\text{估計}} \bar{\bar{Y}}$$

2. 二階段比值集群抽樣法平均數估計

$$\bar{y}_{rc} = \frac{N}{\hat{M}}\bar{y}_t = \frac{1}{\bar{M}}\frac{\sum M_i \bar{y}_i}{n} \xrightarrow{\text{估計}} \bar{\bar{Y}}$$