統　計

《抽樣方法》

試題評析

今年抽樣方法出題是難得一見的平易近人，兼顧基本的計算題（50分）、基本觀念問答題（25分）和進階的觀念問答題（25分），而且四種基本的單段抽樣法均有出現。其中第一題屬於簡單隨機樣法的基本計算題。第二題屬於分層抽樣法的基本計算題。第三題是關於系統抽樣法和集體抽樣法的基本觀念問答題。第四題則屬進階的觀念問答題，平時若無深度思考，臨場恐不易寫出完整的答案。因此預估一考生約能拿45分左右；程度較高者，可拿70分以上。

一、設從某簽名請願的676張（即N=676）表中，用簡單隨機抽樣法從676張中抽出n=50張表為一組樣本。在母體（population）676張表中，每張表上設有42個空格可以簽名。在抽出的樣本50張表中，有的全部簽滿了，有的只簽了一部分，其分配情形如下表：

簽名人數	42	41	36	32	29	27	23	19	16	15	14	11	10	9	7	6	5	4	3
張　數	23	4	1	1	1	2	1	1	2	2	1	1	1	1	1	3	2	1	1

試計算：(一)母體平均數μ之估計值$\hat{\mu}$，及其變異數$\hat{V}ar(\hat{\mu})$。（10分）

(二)母體總和τ之估計值$\hat{\tau}$，及其變異數$\hat{V}ar(\hat{\tau})$。（10分）

(三)試分別求母體平均簽名人數μ及母體簽名總人數τ之95%的信賴區間。（10分）

答：

(一)令簽名人數為x_i，張數為f_i，則

1. $\hat{\mu}=\frac{1}{n}\sum_i x_i\cdot f_i=\frac{1}{50}(42\times23+41\times4+\cdots\cdots+3\times1)=29.42$（人／張）

2. $s^2=\frac{1}{n-1}\sum_i(x_i-\hat{\mu})^2\cdot f_i=\frac{1}{n-1}\left[\left(\sum x_i^2\cdot f_i\right)-\frac{\left(\sum x_i\cdot f_i\right)^2}{n}\right]=\frac{1}{49}\left[(42^2\times23+\cdots\cdots+3^2\times1)-\frac{(1471)^2}{50}\right]=228.98$

$\Rightarrow\ \hat{V}ar(\hat{\mu})\frac{s^2}{n}\cdot(1-f)=\frac{228.98}{50}\cdot\left(1-\frac{50}{676}\right)=4.241$

(二)1. $\hat{\tau}=N\cdot\hat{\mu}=676\times29.42=19888$（人）

2. $\hat{V}ar(\hat{\tau})=N^2\cdot\hat{V}ar(\hat{\mu})=(676)^2\times4.241=1938035.216$

(三)1. μ之95%信賴區間為$\hat{\mu}\pm z_{0.025}\cdot\sqrt{\hat{V}ar(\hat{\mu})}$

$\Rightarrow 29.42\pm1.96\cdot\sqrt{4.241}\Rightarrow[25.38,\ 33.46]$

2. τ之95%信賴區間為$N\cdot\hat{\mu}\pm N\cdot z_{0.025}\cdot\sqrt{\hat{V}ar(\hat{\mu})}$

$\Rightarrow 676\times(29.42\pm1.96\cdot\sqrt{4.241}\Rightarrow[17159,\ 22617]$

（講義命中事實：第一回p.15,16,17）

二、抽樣調查甲、乙、丙三村居民平均每週收看電視節目時數所得資料如下表：

甲村（第一層）	乙村（第二層）	丙村（第三層）
35 28 26 41 43 29	27 4 49	8 15 21 7 14
32 27 36 25 29 31	10 15 41	30 20 11 12 32
39 38 40 45 28 27	25 30	34 24
35 34		

甲、乙、丙三村居民人數分別為N_1=155人，N_2=62人，N_3=93人，總計母體有N=310人，預定查訪n=40人，現分別從三村中各抽一組簡單隨機樣本，樣本數分別為n_1=20人，n_2=8人，n_3=12人。

計算得下表：

甲村N_1=155	乙村N_2=62	丙村N_3=93
n_1=20 $\bar{x}_1$=33.9 s_1^2=35.358	n_2=8 $\bar{x}_2$=25.125 s_2^2=232.411	n_3=12 $\bar{x}_3$=19.0 s_3^2=87.636

應用此兩個表結果，試估計

(一)甲、乙、丙三村居民平均每週收看電視節目時數及其95%信賴區間。（10分）

(二)此三村居民每週收看電視總時數及其95%信賴區間。（10分）

答：

(一)1. $\bar{x}_{st}=\frac{1}{N}\sum_{h-1}^{3}N_h\bar{x}_h=\frac{1}{310}\left(155\times33.9+62\times25.125+93\times19.0\right)=27.675$

2. $\bar{X}$ 之95%信賴區間為 $\bar{x}_{st}\pm z_{0.025}\cdot\frac{1}{N}\sqrt{\sum_{h=1}^{3}N_h^2\cdot\frac{s_h^2}{n_h}\cdot(1-f_h)}$

$$\sum_{h=1}^{3}N_h^2\cdot\frac{s_h^2}{n_h}\cdot(1-f_h)=\sum_{h=1}^{3}N_h\cdot\frac{s_h^2}{n_h}\cdot(N_h-n_h)$$

$$=155\cdot\frac{35.358}{20}\cdot(155-20)+52\cdot\frac{232.411}{8}\cdot(52-8)+93\cdot\frac{87.636}{12}\cdot(93-12)$$

$$=164519.0385$$

$$\Rightarrow 27.675\pm1.96\cdot\frac{1}{310}\sqrt{164519.0385}\Rightarrow[25.11\ ,30.24]$$

(二)1. $\hat{X}_{st}=N\cdot\bar{x}_{st}=310\times27.675=8579.25$

2. X之95%信賴區間為 $N\cdot\bar{x}_{st}\pm z_{0.025}\cdot\sqrt{\sum_{h=1}^{3}N_h^2\cdot\frac{s_h^2}{n_h}\cdot(1-f_h)}$

$$\Rightarrow 310\times[25.11,30.24]\Rightarrow[7784,9374]$$

（講義命中事實：第一回p.35,38,39）

三、試分別定義：(一)系統抽樣法（systematic sampling）。又其應用時機為何？（10分）
(二)集體抽樣法（cluster sampling）。又其應用時機為何？（10分）

答：

(一)系統抽樣法：

將母體的每一個體從1至N加以編號，然後先隨機選出一個號碼作為第一個樣本，另外定一個自然數k（一個最方便的取法是取 $k=\left[\frac{N}{n}\right]$），再依此定距k循序抽出樣本，例如假設第一個樣本是第a號，則編號為a + k，a + 2k，a + 3k，…，a + (n − 1)k的個體就是剩下的(n − 1)個樣本，其中若 $a+ik(i=1,2,\cdots,n-1)$ 超過N，則將其減去N即可。所以只要決定出第一個樣本a以及定距k，則所有的樣本就可以決定，因此系統抽樣又稱為等間隔抽樣（equal interval sampling），其中定距k稱為抽樣間隔（sampling interval）。

應用時機：

當母體本身已經有按照某種次序排列的時候，則適合採用系統抽樣。例如戶政上的門牌號碼是不缺號地編排，此時若以戶為抽樣單位，則採用系統抽樣有其方便性。

(二)集體抽樣法：

先將母體分成相似的若干群，每一群稱為一個集體（cluster），分群的原則是，群內的變異大，而群間的變異小，使得每一群集宛如是母體的一個縮型（miniature），分群完後，再用簡單隨機抽樣法抽取少數的集體組成樣本，而對抽中的集體進行普查。

應用時機：

當完整的母體底冊有時不易獲得，或者母體底冊已過時尚未更新，此時可採用集體抽樣法。此外若抽樣考量目的在於降低抽樣成本；或者在實用上，需利用區域（例如台北市的行政區域）加以分群時，均宜採用集體抽樣法。

（講義命中事實：第二回p.4-1，p.4-12，p.5-1，p.5-8）

四、於簡單隨機抽樣法中，試分別說明(一)**比率估計與**(二)**迴歸估計有何異同之處。**（30**分**）

（**註：標準常態**$z_{.025}=1.96$）

答：

(一)**相同處：**

1.**兩種估計法均假設樣本資料為成對資料。**

2.**兩種估計法均可用來估計母體總數、母體平均數、母體單位總數與母體比例。**

3.**在估計上述四種參數時，就數學形式而言，比例估計量可以視為迴歸估計量的特例。**

(二)**相異處：**

1.**比例估計法可以用來估計母體比值**（population ratio），**迴歸估計法則無法估計母體比值。**

2.**迴歸估計法在理論上可以獲得性質較好的最小迴歸估計量，比例估計法則難以獲得較好的數理統計特性。**

（**講義命中事實：第二回**p.7-1，p.7-2，p.7-9，p.7-10）