ERIC AGYEMANG

MAT 450 - EXAM3

① Given that we select 2 psus with probabilities of inclusion proportional to M_i, ~~the~~

ⓐ If the selected psus is $\{1,3\}$ by Lahiri's method, the estimate of $\bar{y}_U$ is given by $\hat{\bar{y}}_\psi$ where $\hat{\bar{y}}_\psi = \frac{\hat{t}_\psi}{\hat{M}_{0\psi}}$. where $\hat{t}_\psi = \frac{1}{n}\sum_{i\in R}\frac{t_i}{\psi_i}$, $\hat{M}_{0\psi} = \frac{1}{n}\sum_{i\in R}\frac{M_i}{\psi_i}$

But $n=2$; $M_0 = \sum_{i=1}^{N} M_i = 5+4+8+5 = 22$

$R = \{1,3\}$; $\psi_i = \frac{M_i}{M_0}$ and so $\psi_1 = \frac{5}{22}$, $\psi_3 = \frac{8}{22}$

Now $\hat{M}_{0\psi} = \frac{1}{2}\left[\frac{5}{(5/22)} + \frac{8}{(8/22)}\right] = 22$ $\therefore \hat{M}_{0\psi} = M_0 = 22$

Then $\hat{t}_\psi = \frac{1}{2}\left[\frac{20}{(\frac{5}{22})} + \frac{38}{(\frac{8}{22})}\right] = 96.250$

$\therefore \hat{\bar{y}}_\psi = \frac{\hat{t}_\psi}{\hat{M}_{0\psi}} = \frac{96.250}{22} = \boxed{4.375}$ as our estimate for the population mean

ⓑ For 95% CI for the population mean in a) we use the normal distribution $Z_{\frac{0.05}{2}} = 1.96$.

Then we got $\hat{\bar{y}}_\psi \pm Z_{\alpha/2}\, SE(\hat{\bar{y}}_\psi)$

But $SE(\hat{\bar{y}}_\psi) = \sqrt{\hat{V}(\hat{\bar{y}}_\psi)}$ where $\hat{V}(\hat{\bar{y}}_\psi) = \frac{1}{(\hat{M}_{0\psi})^2}\cdot\frac{1}{n}\cdot\frac{1}{n-1}\sum_{i\in R}\left(\frac{t_i}{\psi_i} - \frac{\hat{\bar{y}}_\psi \cdot M_i}{\psi_i}\right)^2$. Considering $\frac{1}{\hat{M}_{0\psi}^2}\cdot\frac{1}{n}\frac{1}{n-1}$ part we got

$\frac{1}{22^2}\cdot\frac{1}{2}\cdot\frac{1}{1} = \frac{1}{968}$

Now for the $\sum_{i\in R}\left(\frac{t_i}{\psi_i} - \frac{\hat{\bar{y}}_\psi\cdot M_i}{\psi_i}\right)^2$ part $= \left[\frac{20}{(5/22)} - \frac{4.375\times5}{(\frac{5}{22})}\right]^2 + \left[\frac{38}{(\frac{8}{22})} - \frac{4.375\times8}{(\frac{8}{22})}\right]^2$

$= 68.0625 + 68.0625 = 136.12$

$\therefore \hat{V}(\hat{\bar{y}}_\psi) = \frac{1}{968} \times 136.125 = 0.140625$ and $SE(\hat{\bar{y}}_\psi) = \sqrt{0.140625}$

$\Rightarrow \quad SE(\hat{\bar{y}}_\varphi) = 0.375$

$\therefore$ 95% CI of $\hat{\bar{y}}_\varphi$ is $4.375 \pm 1.96(0.375) = \underline{[3.64, 5.11]}$

If we want to use the t-distribution, we have ~~df~~ the df $= \Sigma(m_h - 1) = (5-1)+(8-1) = 11$ and so $t_{\alpha/2, n-1}$ gives $t_{0.025, 11} = 2.201$. Then the CI is given as

$4.375 \pm 2.201(0.375) = (3.55, 5.20)$ somehow close to that of the Z distribution.

c) Given that we selected psus $\{1, 3\}$ we estimate the $\bar{y}_U$ by $\hat{\bar{y}}_\varphi = \frac{1}{n}\sum_{i\in R} \bar{y}_i$ where $\bar{y}_i = \frac{\sum_{j\in S} y_{ij}}{m_i}$

Then $\bar{y}_1 = \frac{3+5}{2} = 4$, $\bar{y}_3 = \frac{7+2+9+4}{4} = 5.5$, $n = 2$

So $\hat{\bar{y}}_\varphi = \frac{1}{2}[4 + 5.5] = \boxed{4.75}$ as our estimate for $\bar{y}_U$.

For the 95% CI of $\hat{\bar{y}}_\varphi$ we get $\hat{\bar{y}}_\varphi \pm Z_{\alpha/2} SE(\bar{y}_\varphi)$ if we use ~~Z~~ Normal distribution.

But $\hat{V}(\hat{\bar{y}}_\varphi) = \frac{1}{n}\frac{1}{n-1}\sum_{i\in R}(\bar{y}_i - \bar{y}_\varphi)^2 = \frac{1}{2}\cdot\frac{1}{1}[(4-4.75)^2 + (5.5-4.75)^2]$
$= \frac{1}{2} \times 1.125 = 0.5625$

$SE(\hat{\bar{y}}_\varphi) = \sqrt{\hat{V}(\hat{\bar{y}}_\varphi)} = \sqrt{0.5625} = \underline{0.75}$, $z_{0.025} = 1.96$

$\therefore$ 95% CI $4.75 \pm 1.96(0.75) \Rightarrow \underline{[3.28, 6.22]}$

If a researcher uses t-distribution, $df = \Sigma(m_h - 1) = (2-1) + (4-1) = 4$ $\therefore t_{0.025, 4} = 2.776$ and so we got $4.75 \pm 2.776(0.75) = \underline{[2.668, 6.832]}$

(d) Assume that $\{3, 5, 7, 2, 9, 4\}$ are SRS we find the

$$\hat{V}(\bar{y}_{SRS}) = \left(1 - \frac{n}{N}\right)\frac{s^2}{n}$$ where $N = 22$, $n = 6$. and

$$s^2 = \frac{1}{n-1}\sum_{i \in S}(y_i - \bar{y})^2.$$ and $$\bar{y} = \frac{\sum y_i}{n} = \left\{\frac{3+5+7+2+9+4}{6}\right\} = 5$$

Then $$s^2 = \frac{1}{6-1}\left\{(3-5)^2 + (5-5)^2 + (7-5)^2 + (2-5)^2 + (9-5)^2 + (4-5)^2\right\}$$
$$= 6.8$$

$$\therefore \hat{V}(\bar{y}_{SRS}) = \left(1 - \frac{6}{22}\right)\frac{6.8}{6} = 0.8242$$

now comparing with c above, we realized that $\hat{V}(\hat{\bar{y}}_{\psi}) < V(\bar{y}_{SRS})$.

Findings shows that we are able to get a variability of 0.5625 when Lahiri's method is used to ~~sample~~ select psus in Cluster sampling while using the SRS we get a variability of 0.8242 and comparing these two methods, since $\hat{V}(\hat{\bar{y}}_{\psi}) < V(\bar{y}_{SRS})$, we say the method used in c) above is better with the smaller variance.

(e) $\hat{t}_{HT} = \sum_{i \in s} \frac{t_i}{\pi_i} = \sum z_i \frac{t_i}{\pi_i}$; $z_i = \begin{cases} 1, & \text{if psus } i \text{ in sample} \\ 0, & \text{if otherwise} \end{cases}$

But $\pi_i = P(\text{unit } i \text{ in sample}) = P(z_i = 1)$

$\pi_{ik} = P(\text{unit } i \text{ and } k \text{ are both in the sample})$

$\sum_{i=1}^{N} \pi_i = n$ and $\sum_{\substack{k=1 \\ k \neq i}}^{N} \pi_{ik} = (n-1)\pi_i$

But $\psi_1 = 5/22$, $\psi_2 = 4/22$, $\psi_3 = 8/22$, $\psi_4 = 5/22$

Completing the joint inclusion probabilities (π_{ik}) for samples of size 2. $\pi_{ik} = \left(\psi_i \frac{\psi_k}{1-\psi_i}\right) + \left(\psi_k \frac{\psi_i}{1-\psi_k}\right)$

	1	2	3	4	π_i
1	—	0.103981	0.236822	0.13369	0.474493
2	0.103981	—	0.184704	0.103981	0.392666
3	0.236822	0.184707	—	0.236822	0.658348
4	0.13369	0.103981	0.236822	—	0.474493
π_k	0.474493	0.392666	0.658348	0.474493	2.000

now for the estimate of the population total $\hat{t}_{HT}$

$\Rightarrow \hat{t}_{HT} = \frac{t_1}{0.474493} + \frac{t_3}{0.658348} = \frac{20}{0.474493} + \frac{38}{0.474493}$

$= \boxed{99.870486}$

$\hat{V}_{HT}(\hat{t}_{HT}) = \sum_{i \in s} (1-\pi_i)\frac{t_i^2}{\pi_i^2} + \sum_{i \in s}\sum_{\substack{k \in s \\ k \neq i}} \frac{\pi_{ik} - \pi_i \pi_k}{\pi_{ik}} \frac{t_i}{\pi_i}\frac{t_k}{\pi_k}$

$= \frac{(1-0.474493)(20^2)}{0.474493} + \frac{(1-0.658348)(38^2)}{0.658348} + \cancel{2}$

$+ 2 \cdot \left[\frac{0.236822 - 0.474493 \times 0.658348}{0.236822}\right] \times \frac{20}{0.474493} \times \frac{38}{0.658348}$

$$\Rightarrow \hat{V}_{HT}(\hat{t}_{HT}) = 519.4174503$$

$$\therefore SE(\hat{t}_{HT}) = \sqrt{\hat{V}_{HT}(\hat{t}_{HT})} = \sqrt{519.4174503} = 22.790732$$

$\therefore$ 95% CI using Z-distribution is given by

$$\hat{t}_{HT} \pm Z_{\alpha/2}\, SE(\hat{t}_{HT})$$

$$\Rightarrow 99.870486 \pm 1.96(22.790732) \cancel{= (55.22065, 144.540)}$$

$$= (55.17065192,\ 144.5103201)$$