Solutions to Ex #5

1. Population: $N = 828$ clusters

SRS of clusters: $n = 85$

Size of clusters: $M_i = 215 = M_0$

t_i = no of errors in claim (cluster) i

Error rate in the population: $p = \dfrac{\sum_{i=1}^{N} t_i}{N \cdot M_0} = \dfrac{t}{N \cdot M_0}$

p = proportion of errors per field

a) $\hat{t}_R = M \cdot \dfrac{\sum_{s_I} t_i}{\sum_{s_I} M_i} = M \cdot \dfrac{\sum_{s_I} t_i}{n M_0} = \dfrac{N}{n} \sum_{s_I} t_i$

$\hat{p} = \dfrac{\hat{t}_R}{N \cdot M_0} = \dfrac{1}{n M_0} \sum_{i \in s_I} t_i = \dfrac{37}{85 \cdot 215} = \underline{.00202}$ $(= 0.2\%)$

$\hat{p} = \bar{y}_s$ since sample of fields is $n \cdot M_0 = 18275$.

$\hat{\mathrm{Var}}(\hat{p}) = \dfrac{1}{(N \cdot M_0)^2} \hat{\mathrm{Var}}(\hat{t}_R)$

From p.123: $\hat{V}(\hat{t}_R) = \dfrac{(M/N)^2}{(m/n)^2} \cdot N^2 \cdot \dfrac{1-f}{n} \cdot \dfrac{1}{n-1} \sum_{i \in s_I} (M_0 \bar{y}_i - M_0 \bar{y}_S)^2$

$= \underbrace{\left(\dfrac{M_0}{M_0}\right)^2}_{1} \cdot N^2 \cdot \dfrac{1-f}{n} \cdot \dfrac{1}{n-1} \sum_{i \in s_I} (t_i - \bar{t}_{s_I})^2$; $\bar{t}_{s_I} = \dfrac{1}{n} \sum_{s_I} t_i$

$\hat{V}(\hat{p}) = \dfrac{1}{M_0^2} \cdot \dfrac{1-f}{n} \cdot \dfrac{1}{n-1} \sum_{i \in s_I} (t_i - \bar{t}_{s_I})^2$, $\bar{t}_{s_I} = \dfrac{37}{85} = 0.435$

and

$SE(\hat{p}) = \sqrt{\hat{V}(\hat{p})} = \dfrac{1}{M_0} \sqrt{\dfrac{1-f}{n} \cdot \dfrac{1}{n-1} \sum_{i \in s_I} (t_i - \bar{t}_{s_I})^2}$

$= \dfrac{1}{215} \sqrt{\dfrac{1}{85}\left(1 - \dfrac{85}{828}\right) \cdot \dfrac{1}{84} \left\{ (4 - 0.435)^2 + (3 - 0.435)^2 + 4(2 - 0.435)^2 + 22(1 - 0.435)^2 + 57(0.435)^2 \right\}}$

$= \dfrac{1}{215} \sqrt{.005894} = \underline{.000357}$

b) $\hat{t}_R = \dfrac{N}{n} \cdot \sum_{s_I} t_i = N \cdot \bar{t}_{s_I} = \underline{360.4} \approx \underline{360}$

$SE(\hat{t}_R) = N \cdot M_0 \cdot SE(\hat{p}) = N \sqrt{.005894} = \underline{63.6}$

c) $\hat{p} = .00202$

$$\hat{V}_{SRS}(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n-1}\left(1-\frac{n}{N}\right) = \frac{.00202 \times .99798}{18274} \cdot \frac{159745}{178020}$$

$$= 9.9 \times 10^{-8} \qquad , SE = .000315$$

$$\text{deff}(SCS, \hat{p}) = \frac{.000357^2}{9.9\times 10^{-8}} = \frac{12.745\times 10^{-8}}{9.9 \times 10^{-8}} = \underline{1.29}$$

Here, cluster sampling is almost as precise as RS

2. a) $\pi_{Ii} = 1/4$, for $i = 1,2,3,4$

$\pi_{j|i} = m_i/M_i$ where $m_i = 10$ and
M_i = size of PSU i = no of employees

$\Rightarrow$ Inclusion prob. for empl. j in establ. i:

$$\pi_{ij} = \frac{1}{4} \cdot \frac{m_i}{M_i} = \frac{10}{4M_i}$$

$$\Rightarrow \quad \pi_{ij} = \begin{cases} 1/4 & \text{for } i = 1 \\ 1/8 & \text{for } i = 2 \\ 1/40 & \text{for } i = 3 \\ 1/48 & \text{for } i = 4 \end{cases}$$

b) Self-weighting sample $\Leftrightarrow \pi_{ij} = 10/M = 10/250 = 1/25$
for all i, j

$$\Leftrightarrow \pi_{Ii} \cdot \pi_{j|i} = \frac{1}{25} \Leftrightarrow \pi_{Ii} \cdot \frac{10}{M_i} = \frac{1}{25}$$

$$\Rightarrow \underline{\pi_{Ii}} = \frac{M_i}{250} = \underline{M_i/M} \quad - \text{ PPS sampling on stage 1}$$

c) $\pi_{ij} = 25/250 = 1/10$.

Two strata: Stratum 1: PSU 1 & 2
Stratum 2: " 3 & 4

$$\pi_{Ii} = \frac{M_i}{M_1+M_2} \quad \text{for } i = 1,2$$

$$\pi_{Ii} = \frac{M_i}{M_3+M_4} \quad \text{for } i = 3,4$$

$\pi_{j|i} = m_i / M_i$; m_i = sample sizes to be determined

Hence for $i = 1, 2$.

$$\frac{M_i}{30} \cdot \frac{m_i}{M_i} = \frac{1}{10} \Leftrightarrow \underline{m_i = 3 \text{ for } i = 1,2}$$

For $i = 3,4$: $\frac{M_i}{220} \cdot \frac{m_i}{M_i} = \frac{1}{10} \Leftrightarrow \underline{m_i = 22 \text{ for } i = 3,4}$

d) $\underline{I}$: $\hat{t} = 4 \cdot \sum_{j \in s_i} \frac{M_i}{10} y_{ij} = 4 M_i \cdot \bar{y}_{si}$ – not a good estimator

$\underline{II}$: $\hat{t} = 25 \sum_{j \in s_i} y_{ij} = M \cdot \bar{y}_{si}$ – ok estimator

$\underline{III}$: Assume establ. i and k are selected at stage I.

$$\hat{t} = 10 \left(\sum_{j \in s_i} y_{ij} + \sum_{j \in s_k} y_{kj} \right) = 250 \cdot \bar{y}_s \text{ – ok!}$$

$\uparrow$ final sample

3. Solutions for R-exercise 4

```
>y=UKDriverDeaths

#In order to get the units (months) numbered from 1 to 192:
>y=as.vector(y)
>y
  [1] 1687 1508 1507 1385 1632 1511 1559 1630 1579 1653 2152 2148 1752 1765 1717
 [16] 1558 1575 1520 1805 1800 1719 2008 2242 2478 2030 1655 1693 1623 1805 1746
 [31] 1795 1926 1619 1992 2233 2192 2080 1768 1835 1569 1976 1853 1965 1689 1778
 [46] 1976 2397 2654 2097 1963 1677 1941 2003 1813 2012 1912 2084 2080 2118 2150
 [61] 1608 1503 1548 1382 1731 1798 1779 1887 2004 2077 2092 2051 1577 1356 1652
 [76] 1382 1519 1421 1442 1543 1656 1561 1905 2199 1473 1655 1407 1395 1530 1309
 [91] 1526 1327 1627 1748 1958 2274 1648 1401 1411 1403 1394 1520 1528 1643 1515
[106] 1685 2000 2215 1956 1462 1563 1459 1446 1622 1657 1638 1643 1683 2050 2262
[121] 1813 1445 1762 1461 1556 1431 1427 1554 1645 1653 2016 2207 1665 1361 1506
[136] 1360 1453 1522 1460 1552 1548 1827 1737 1941 1474 1458 1542 1404 1522 1385
[151] 1641 1510 1681 1938 1868 1726 1456 1445 1456 1365 1487 1558 1488 1684 1594
[166] 1850 1998 2079 1494 1057 1218 1168 1236 1076 1174 1139 1427 1487 1483 1513
[181] 1357 1165 1282 1110 1297 1185 1222 1284 1444 1575 1737 1763
> N=192

# since n=24, k=8
> k=8
> start=sample(k,1)
> start
[1] 4
> s=seq(start,N,k)
> s
 [1]   4  12  20  28  36  44  52  60  68  76  84  92 100 108 116 124 132 140 148
[20] 156 164 172 180 188
> y[s]
 [1] 1385 2148 1800 1623 2192 1689 1941 2150 1887 1382 2199 1327 1403 2215 1638
[16] 1461 2207 1552 1404 1726 1684 1168 1513 1284

#Both estimates are the same:
> mean(y[s])
[1] 1707.417
# true value:
> mean(y)
[1] 1670.307
```

4

i) LP ⟹ SP

Assume $T(x_1) = T(x_2) = t$. We must show that
$I(E, x_1) = I(E, x_2)$ using LP.

We have $l_{x_1}(\theta) = P_\theta(X = x_1) = \underbrace{P_\theta(X = x_1 \mid T = t)}_{h(x_1)} \cdot P_\theta(T = t)$

$l_{x_2}) = P_\theta(X = x_2) = \underbrace{P(X = x_2 \mid T = t)}_{h(x_2)} P_\theta(T = t)$

Hence
$l_{x_1}(\theta) / l_{x_2}(\theta) = h(x_1)/h(x_2)$: proportional likelihood

and from LP it follows that $I(E, x_1) = I(E, x_2)$

ii) LP ⟹ CP

We must show that $I(E^*, (j, x_j)) = I(E_j, x_j)$

$$l^*_{(j,x_j)}(\theta) = \frac{1}{2} \cdot f^j_\theta(x_j) = \frac{1}{2} l^j_{x_j}(\theta)$$

$$\overset{LP}{\Rightarrow} I(E^*, (j, x_j)) = I(E_j, x_j)$$

5. $H_0: \theta = \frac{1}{2}$ against $H_1: \theta > \frac{1}{2}$

E_1: Reject H_0 if $Y_1 \geq c_1$.

$$\text{P-value} = P_{\frac{1}{2}}(Y_1 \geq 9) = \sum_{x=9}^{12} \binom{12}{x} \left(\frac{1}{2}\right)^{12} = \underline{.0730}$$

E_2: Reject H_0 if $Y_2 \geq c_2$

$$\text{P-value} = P_{\frac{1}{2}}(Y_2 \geq 9)$$

Let Y be the number of successes in the first 11 trials. Y is binomial $(11, \theta)$. Now,

$$Y_2 \leq 8 \iff Y \leq 8$$

Hence: $P_\theta(Y_2 \geq 9) = P_\theta(Y \geq 9)$ and P-value is:

$$P_{1/2}(Y \geq 9) = \sum_{x=9}^{11} \binom{11}{x} \left(\frac{1}{2}\right)^{11} = \underline{.0327}$$

 6

$$l_x(\underline{y}) = \begin{cases} p(s) & \text{if } \underline{y} \in \Omega_x \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} 1/252 & \text{if } \underline{y} \in \Omega_x \\ 0 & \text{otherwise} \end{cases} \qquad \binom{10}{5} = 252$$

$$\Omega_x = \{ \underline{y} : y_2 = 0, y_3 = 1, y_7 = y_8 = 0, y_{10} = 1 \}$$

$$\underline{y} = (y_1, \ldots y_{10})$$