Lösningsförslag

Uppgift 1

a) $\hat{t}_{unb} = \frac{15}{3} \sum M_i \bar{y}_i = \frac{15 \cdot 464}{3} = 2320$ böcker

$s_t^2 = \frac{1}{2} \left(\sum \hat{t}_i - \frac{2320}{15} \right)^2 = 709.33$

$\frac{s_i^2}{m_i} = \frac{\hat{p}_i(1-\hat{p}_i)}{m_i - 1}$, $\quad s_1^2 = \frac{\frac{37}{154}\left(1 - \frac{37}{154}\right)}{153} = 0.001193$

$s_2^2 = \frac{\frac{33}{146}\left(1 - \frac{33}{146}\right)}{145} = 0.001206$

$s_3^2 = \frac{\frac{46}{163}\left(1 - \frac{46}{163}\right)}{162} = 0.001250$

b) forts.

$$SE^2\left[\hat{t}_{unb}\right] = 15^2\left(1 - \frac{3}{15}\right) \cdot \frac{709.33}{3}$$

$$+ \frac{15}{3}\left(\sum \left(1 - \frac{m_i}{M_i}\right) \frac{\hat{p}_i(1-\hat{p}_i)}{m_i} \cdot M_i^2 \right.$$

$$= 218.61^2$$

95% KI $\quad 2320 \pm 428$

Uppgift 2

a) Beräkna först totalkostnaden i varje kluster: $t_i = m_i * \bar{y}_i$

	m_i	$\bar{y}_i$	t_i
1	10	2432.5	24325
2	23	2786.4	64087.2
3	45	2689.1	121009.5
4	15	2978.3	44674.5
5	79	3016.0	238264

Använd formler från sid. 179 i Lohr s.

$$\hat{t}_{unb} = \frac{N}{n}\sum_{i \in S} t_i = \frac{919}{5}(24325 + 64087.2 + 121009.5 + 44674.5 + 238264) = 90495804.76$$

Och

$$\hat{\bar{y}}_{unb} = \frac{\hat{t}_{unb}}{M_0} = \frac{90495804.76}{27392} = 3303.73$$

Variansen är

$$V(\hat{\bar{y}}_{unb}) = V\left(\frac{\hat{t}_{unb}}{M_0}\right) = \frac{1}{M_0^2}V(\hat{t}_{unb}) = N^2\left(1 - \frac{n}{N}\right)\left(\frac{s_t^2}{n}\right) = \frac{1}{27392^2}919^2\left(1 - \frac{5}{919}\right)\left(\frac{86058.13^2}{5}\right)$$
$$= 1658166.08$$

Där

$$s_t^2 = \frac{1}{n-1}\sum_{i \in S}\left(t_i - \frac{\hat{t}}{N}\right)^2 = 86058.13^2$$

Och standarfelet/medelfelet blir:

$$SE(\hat{\bar{y}}_{unb}) = \sqrt{V(\hat{\bar{y}}_{unb})} = \sqrt{1658166.08} = 1287.7$$

b)

Designeffekten beräknas enligt följande formel:

$$designeffekt = \frac{\hat{V}(\hat{\bar{y}}_{unb})}{\hat{V}_{OSU}(\hat{\bar{y}})}$$

V behöver alltså beräkna variansen för en situation med samma urvalsstorlek fast om OSU hade använts. I så fall är N=27392, n=172 och s=447.7. Variansen blir då:

$$\hat{V}_{OSU}(\hat{\bar{y}}) = \left(1 - \frac{n}{N}\right)\left(\frac{s^2}{n}\right) = \left(1 - \frac{172}{27392}\right)\left(\frac{447.7^2}{172}\right) = 1158.00$$

Och designeffekten blir då:

$$designeffekt = \frac{\hat{V}(\hat{\bar{y}}_{unb})}{\hat{V}_{OSU}(\hat{\bar{y}})} = \frac{1658166.08}{1158.00} = 1431.92$$

Vilket inte är någon bra designeffekt.

c)

Vi använder här resultaten på s. 180 i Lohr. Observera att nu när vi använder kvotestimatorn måste vi skatta både täljare och nämnare.

$$\hat{\bar{y}}_r = \frac{\hat{t}_{unb}}{\hat{M}_0} = \frac{\frac{N}{n}\sum \bar{y}_i M_i}{\frac{N}{n}\sum M_i} = \frac{\frac{919}{5}492360.2}{\frac{919}{5}172} = 2862.56$$

Variansen blir:

$$V(\hat{\bar{y}}_r) = \left(1-\frac{n}{N}\right)\left(\frac{1}{n\bar{M}^2}\right)\left(\frac{\sum M_i^2(\bar{y}_i-\hat{\bar{y}}_r)^2}{n-1}\right) = \left(1-\frac{5}{919}\right)\left(\frac{1}{5*34.4^2}\right)(58110901.095) = 9767.90$$

där

$$\frac{\sum M_i^2(\bar{y}_i-\hat{\bar{y}}_r)^2}{n-1} = \frac{1}{5-1}(10\verb|^|2*(2432.5-2862.56)\verb|^|2+23\verb|^|2*(2786.4-2862.56)\verb|^|2+45\verb|^|2*(2689.1-2862.56)\verb|^|2+15\verb|^|2*(2978.3-2862.56)\verb|^|2+79\verb|^|2*(3016.0-2862.56)\verb|^|2) = 58110901.095$$

Vilket ger medelfelet:

$$SE(\hat{\bar{y}}_r) = \sqrt{9767.9} = 98.8$$

Uppgift 3

Detta är ett enstegs klusterurval av passagerare, där bussturerna utgör kluster. Vi vet inte totala antalet passagerare på hela linjen, varför andelen bältade p måste skattas med kvotskattningen $\hat{p}_r = \frac{\sum t_i}{\sum M_i} = \frac{4+8+..}{120+80+..} = \frac{112}{560} = 0.2$

Den skattade variansen är $(1-\frac{n}{N})\frac{1}{n\bar{M}^2}\cdot\frac{1}{n-1}\sum m_i^2(p_i-\hat{p}_r)^2 =$

$$= (1-\tfrac{6}{60})\frac{1}{6\cdot(\frac{560}{6})^2}\frac{1}{5}(120^2(\tfrac{4}{120}-0.2)^2+80^2(\tfrac{8}{80}-0.2)^2+\ldots) = 0.9\frac{1}{52266.67}\frac{1}{5}\cdot 2760 = 0.0095$$

Ett 95 % konfidensintervall blir då $p = 0.2 \pm 1.96\sqrt{0.0095} = 0.2 \pm 0.19$.

Uppgift 4

Detta är ett 2-stegs klusterurval. Först beräknas punktskattningen enligt formeln nedan.

$$\hat{p} = \hat{\bar{y}}_{unb} = \frac{\hat{t}_{unb}}{M_0}$$

Avrunda stickprovsstorlek till närmaste heltal.

$$\hat{t}_{unb} = \frac{N}{n}\sum M_i\bar{y}_i = \frac{44}{3}\left(11251 * \frac{380}{1125} + 4782 * \frac{148}{478} + 6127 * \frac{201}{613}\right) = 106919.61$$

vilket ger:

$$\hat{p} = \hat{\bar{y}}_{unb} = \frac{\hat{t}_{unb}}{M_0} = \frac{106919.61}{315211} = 0.3392$$

Variansen för $\hat{p}$ blir:

$$V(\hat{p}) = \frac{V(\hat{t}_{unb})}{M_o^2} = \frac{89203919.07}{315211^2} = 0.0008978$$

där

$$\begin{aligned}
V(\hat{t}_{unb}) &= N^2\left(1-\frac{n}{N}\right)\left(\frac{s_t^2}{n}\right) + \frac{N}{n}\sum\left(1-\frac{m_i}{M_i}\right)M_i^2\left(\frac{s_i^2}{m_i}\right) \\
&= 44^2 * \left(1-\frac{3}{44}\right)\left(\frac{147818.84}{3}\right) \\
&+ \frac{44}{3}\Bigg((1-0.1) * 11251^2 * \frac{\left(0.3378(1-0.3378)\right)}{1125-1} + (1-0.1) * 4782^2 \\
&* \frac{\left(0.3096(1-0.3096)\right)}{478-1} + (1-0.1) * 6127^2 * \frac{\left(0.3279(1-0.3279)\right)}{613-1}\Bigg) \\
&= 88888395.79 + 315523.28 = 89203919.07
\end{aligned}$$

där

$$\begin{aligned}
s_t^2 &= \frac{1}{n-1}\sum\left(\hat{t}_i - \frac{\hat{t}_{unb}}{N}\right)^2 \\
&= \frac{1}{3-1}\Bigg(\left(11251 * \frac{380}{1125} - \frac{106919.61}{44}\right)^2 + \left(4782 * \frac{148}{478} - \frac{106919.61}{44}\right)^2 \\
&+ \left(6127 * \frac{201}{613} - \frac{106919.61}{44}\right)^2\Bigg) = 147818.84
\end{aligned}$$

och där:

$$\frac{s_i^2}{m_i} = \frac{\hat{p}_i(1-\hat{p}_i)}{m_i - 1}$$

och:

$$\hat{p}_1 = \frac{380}{1125} = 0.3378$$

$$\hat{p}_2 = \frac{148}{478} = 0.3096$$

$$\hat{p}_3 = \frac{201}{613} = 0.3279$$

Alltså blir medelfelet:

$$SE(\hat{p}) = \sqrt{V(\hat{p})} = \sqrt{0.0008978} = 0.02997$$

Vilket ger konfidensintervallet:

$$0.3392 \pm 0.0587 = 1.96 * 0.02997$$

$$(\mathbf{0.280}\ ,\mathbf{0.398})$$