MTH5119 : SOLUTIONS 3

Q1

h	N_h	S_h	c_h	W_h	$N_hS_h/\sqrt{c_h}$ (or $W_hS_h/\sqrt{c_h}$) $\times 105$	n_h
1	300	30	25	0.3	$9000/5 \to 9/5 \to 189$	$(189/559)n$
2	500	40	49	0.5	$20000/7 \to 20/7 \to 300$	$(300/559)n$
3	200	20	36	0.2	$4000/6 \to 2/3 \to 70$	$(70/559)n$
	$N = 1000$			1	559	n

(2)

(a) Substitute into cost equation

$$3000 = 500 + \frac{n}{559}(189\times25 + 300\times49 + 70\times36)$$

$$\longrightarrow n = (2500\times559)/21945 = 63.68$$

(2) so take $n = 63$ (round down) then $\underline{n_1 = 21, n_2 = 34, n_3 = 8}$ (round to nearest integer)
(otherwise budget exceeded)

(2) $\text{FPC} = \frac{1}{N}\sum_{h=1}^{H} W_h S_h^2 = 1.15$ so achieved variance $= \sum_{h=1}^{H}\frac{W_h^2S_h^2}{n_h} - \text{FPC}$

(2) $\longrightarrow \frac{.3^2\times900}{21} + \frac{.5^2\times1600}{34} + \frac{.2^2\times400}{8} - 1.15 = \underline{16.47}$ (4sf)

(b) proportional allocation is $n_h \propto N_h$ (or W_h) ie $n_1 = 0.3n$, $n_2 = 0.5n$, $n_3 = 0.2n$
With $n = 63$ this gives either $(19, 31, 13)$ or $(19, 32, 12)$
Put both into formula for achieved variance:

(1),(1) $\longrightarrow$ $\underline{17.25}$ (worse) and $\underline{16.95}$ (better) resp. to 4sf.

(2) (both are of course larger than the achieved variance for the opt. alloc^{n})

(c) require $\text{Var}[\bar{y}_{st}] = \sum \frac{W_h^2S_h^2}{n_h} - 1.15 = 15$ so that
(with optimal allocation)

$$n = \left(\frac{0.3^2\times900}{189} + \frac{0.5^2\times1600}{300} + \frac{0.2^2\times400}{70}\right)\times\frac{559}{16.15}$$

$= 68.89 \longrightarrow \underline{n = 69}$ round up (to satisfy variance constraint)

(1),(1),(1) $\longrightarrow n_1 = \frac{189}{559}\times69 = \underline{23}$ $\quad n_2 = \frac{300}{559}\times69 = \underline{37}$ $\quad n_3 = \frac{70}{559}\times69 = \underline{9}$

(rounding to nearest integer)

(1) now actual overall cost is $\sum c_h n_h$ + overheads

$= 23\times25 + 37\times49 + 36\times9 + 500 = \underline{\underline{£3212}}$

Q2 $\quad N = 64 \quad n = 4$ SRS
$y_1 = 50$ (men) $\quad y_2 = 20$ (men) $\quad y_3 = 15$ (women) $\quad y_4 = 10$ (women)

$\bar{y} = 95/4 = \underline{23.75}$ $\quad s^2 = \frac{1}{3}(50^2 + 20^2 + 15^2 + 10^2 - 95^2/4) = 322.91\dot{6}$ ($\to s = 17.97$ 4s.f.)

(4) $v[\bar{y}] = \frac{1-f}{n}s^2 = \underline{75.68}$ (4sf)

$\bar{y}_{men} = 35 \quad s^2_{men} = \frac{1}{2}(50-20)^2 = 450 \quad N_{men} = 30 \quad n_{men} = 2$

$\bar{y}_{women} = 12.5 \quad s^2_{women} = \frac{1}{2}(15-10)^2 = 12.5 \quad N_{women} = 34 \quad n_{women} = 2$ so

$\bar{y}_{st} = \frac{1}{64}(30\times35 + 34\times12.5) = \underline{23.046875}$ and $v[\bar{y}_{st}] =$

PTO.

MTH5119 SOLUTIONS 3 continued

Q2. $v[\bar{y}_{st}] = \frac{1}{N^2}\left(N_{men}^2\left(\frac{1}{n_{men}} - \frac{1}{N_{men}}\right)s^2_{men} + N_{women}^2\left(\frac{1}{n_{women}} - \frac{1}{N_{women}}\right)s^2_{women}\right)$

(4) $= \frac{1}{64^2}\left(30^2(0.5-30^{-1})\,450 + 34^2(0.5-34^{-1})\,12.5\right)$

$= 195800/64^2 = \underline{47.80}$

(4) Although the alternative estimators ($\bar{y}$ and $\bar{y}_{st}$) are similar, the poststratified estimator seems to offer considerable improvement (about 40%) in terms of reduction in ~~estimated~~ the variance estimate. (This is a fairly strong indication that the stratification by sex is effective for estimating # texts)

(1) Q3 (a) (SRS) $Var[\bar{y}] = \frac{1-f}{n}S^2 = S^2/6 = 79.24\dot{6}2\dot{1}$ (Lab2, Task1) $= \underline{79.25}$ (4SF)

(1) Lab4 $Var[\hat{Y}_R] = 50.9597$ (6SF) $= \underline{50.96}$ (4SF)

(1) $Var[\bar{y}_{lr}] = 151.168$ (6SF) $= \underline{151.2}$ (4SF)

(2) The linear regression estimator is considerably worse than the other two in terms of precision [it is based on fitting a straight line with an intercept to the sample data; examination of some of the possible samples convinces me that the variability in these lines is very great] and the ratio estimator offers about a 37% improvement over the sample mean which ignores the covariate information.

(1) (b) coverage of $\bar{y} \pm 1.96\,v[\bar{y}]$: 84.65% (Lab1)

(1) Lab4 coverage of $\hat{Y}_R \pm 1.96\,v_0[\hat{Y}_R]$: 87.88%

(1) coverage of $\bar{y}_{lr} \pm 1.96\,v[\bar{y}_{lr}]$: 65.25%

(2) again the linear regression is much worse with unacceptably low coverage [~~perhaps~~ due to the huge negative bias in the variance estimator] but the other two are almost equal with the ratio estimator having the slight edge (nearer to 95%)

(1) (c) $n=4$ bias $= +0.1937$ (4SF) rel. bias $= \underline{0.02713}$ (2.7%)

(1) $n=3$ bias $= +0.1882$ rel. bias $= \frac{0.1882}{8.868} = 0.02122$ (2.12%)

(1) $n=6$ bias $= +0.0976$ rel. bias $= 0.0976/4.97238 = 0.01963$ (1.96%)

(1) $n=8$ bias $= +0.0490$ rel. bias $= 0.0490/3.48514 = 0.01406$ (1.41%)

(2) The relative bias increases slightly as n increases from 3 to 4, but thereafter seems to decrease steadily with sample size. All are comfortably less than 10%

(1) (d) $n=4$ bias in var. est. $v_0[\hat{Y}_R] = E[v_0] - Var[\hat{Y}_R] = 49.09 - 50.96 = \underline{-1.87}$

(1) $n=3$ bias in var est $v_0[\hat{Y}_R] = 73.8949 - (8.868)^2 = \underline{-4.74}$ (4SF)

(1) $n=6$ bias in var. est $v_0[\hat{Y}_R] = 24.3358 - (4.97238)^2 = \underline{-0.3888}$ (4SF)

(1) $n=8$ bias in var. est $v_0[\hat{Y}_R] = 12.0900 - (3.48514)^2 = -0.05620$ (4SF)

(2) (As expected) all biases are negative but fairly small except possibly for $n=3$, but it is still less than 10% of the true value, again the bias decreases with sample size. (Accuracy to 2 dp acceptable for (d))