2.3.16

STK4600

Solutions to Ex #4

1a) y_i = no of refereed publications for faculty member i, $i = 1, \dots, 807$.

$\sum_{i \in s} y_i = 88$; $\bar{y}_s = 88/50 = \underline{1.76}$

= estimate for μ, the mean per faculty member

$\hat{V}(\bar{y}_s) = \frac{s^2}{n}(1-f)$, $n = 50$, $f = \frac{n}{N} = \frac{50}{807} = .062$

$s^2 = \frac{1}{n-1}\sum_{i \in s}(y_i - \bar{y}_s)^2 = 337.12/49 = 6.88$

$\hat{V}(\bar{y}_s) = .12907$ & $\underline{SE(\bar{y}_s) = .36}$

$CV(\bar{y}_s) = \frac{SE(\bar{y}_s)}{\bar{y}_s} = \underline{.205 = 20.5\%}$

b) p = proportion of faculty members with no publications

$\hat{p} = 28/50 = \underline{0.56}$

95% CI : $\hat{p} \pm 1.96 \underbrace{\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}(1-f)}}_{.069}$ $= \underline{.56 \pm .135}$

$= \underline{(.425, .695)}$

2a) Prop'l allocation: $n_h = n \cdot \frac{N_h}{N}$; $h = 1, 2, 3, 4$.

N_h : 102 - 310 - 217 - 178 and $n = 50$

N_h/N : .1264 - .3841 - .2689 - .2205

$\Rightarrow$ $n_1 = 50 \frac{102}{807} = 6.3$ ≈ 7

$n_2 = 50 \times .3841 = 19.2 \Rightarrow 19$

$n_3 = 50 \times .2689 = 13.4 \Rightarrow 13$

$n_4 = 50 \times .2205 = 11 \Rightarrow 11$

2b) $\hat{t}_{st} = \sum_{h=1}^{4} N_h \cdot \bar{y}_h$ $\bar{y}_h = \frac{1}{n_h} \sum_{s_h} y_{hi}$

$\bar{y}_1 = 3.143$ $\bar{y}_2 = 2.105$ $\bar{y}_3 = 1.231$ $\bar{y}_4 = .4545$

$\underline{\hat{t}_{st} = 1321}$

(If you use the approx: $\hat{t}_{st} \approx N\bar{y}_s = 1340$)

Exact $\hat{V}(\hat{t}_{st}) = \sum_{h=1}^{4} N_h^2 \frac{1-f_h}{n_h} \cdot s_h^2$ $f_h = \frac{n_h}{N_h}$

$s_h^2 = \frac{1}{n_h-1} \sum_{s_h} (y_i - \bar{y}_h)^2$

$= \frac{1}{n_h-1} \left\{ \sum_{i \in s_h} y_i^2 - n_h \bar{y}_h^2 \right\}$

h:	1	2	3	4
s_h^2	6.808	8.212	4.358	.873

$\hat{V}(\hat{t}_{st}) = 65613$

$SE(\hat{t}_{st}) = \sqrt{65613} = \underline{256.15}$

Approximately: $\hat{V}(\hat{t}_{st}) = N^2 \cdot \frac{1-f}{n} \sum W_h s_h^2 = 65726$
(since appr. prop. all.) and $\underline{SE = 256.37}$

c) In the SRS: $\bar{y}_s = 1.76$ and $\hat{t} = 807 \times 1.76 = 1420$.

and $\hat{V}(\hat{t}) = 807^2 \cdot \hat{V}(\bar{y}_s)$

and $SE(\hat{t}) = 807 \times SE(\bar{y}_s) = 807 \times .36 = \underline{290.52}$

The precision is a little bit better with strat. sample : $256/290.5 = .88$: 12% reduction ~~in uncerta~~ in uncertainty

d) $\hat{p}_{st} = \sum W_h \hat{p}_h$, $W_h = N_h/N = .126 - .384 - .269 - .221$

$\hat{p}_h$: $.143 - .526 - .692 - .727$

$\Rightarrow \underline{\hat{p}_{st} = .567}$

Approx: $\hat{p}_{st} = \bar{y}_s = \frac{28}{50} = \underline{.560}$

$$[SE(\hat{p}_{st})]^2 = \sum_h \Big(\overset{W_n^2}{\frac{N_h}{N}}\Big)^2 \Big(1 - \frac{n_h}{N_h}\Big) \frac{\hat{p}_h(1-\hat{p}_h)}{n_h - 1}$$

$= .000302 + .001917 + .001208 + .000909$

$= .004336$ (with approx: $SE(\hat{p}_{st}) = \frac{1-f}{n} \sum W_h \overset{\frac{n_h \hat{p}_h(1-\hat{p}_h)}{n_h-1}}{s_h^2}$ ~~$= .0064$~~ $= .004335$)

$\underline{SE(\hat{p}_{st}) = .066}$

95% CI : $.567 \pm 1.96 \times .066 = .567 \pm .129$

$= \underline{(.438, .696)}$

e) Not much difference, only a slight improvement with stratification

f) Yes, because SE is smaller and CI's are shorter

g) We note that the ratios

NB! → s_h / s_4 are: 2.~~7~~9, 3.07, 2.23 with average 2.7

and assumption is equivalent with $\frac{\sigma_h}{\sigma_4} = 2.5$

So the assumption is a rather crude approximat

Optimal allocation:

$$n_h = 50 \cdot \frac{N_h \sigma_h}{\sum_{h=1}^{4} N_h \sigma_h} = 50 \cdot \frac{N_h \cdot 2.5 \sigma_4}{N_4 \sigma_4 + 2.5 \sigma_4 \sum_{h=1}^{3} N_h} \quad h = 1, 2, 3$$

$$n_h = 50 \cdot \frac{N_h \times 2.5}{N_4 + 2.5 \sum_{h=1}^{3} N_h} = 125 \frac{N_h}{1750.5}, \quad h = 1, 2, 3$$

$$\& \ n_4 = 50 \cdot \frac{N_4}{1750.5}$$

$n_1 = 7.3$ → $n_1 = 7$

$n_2 = 22.1$ → $n_2 = 22$

$n_3 = 15.5$ → $n_3 = 16$

$n_4 = 5.1$ → $n_4 = 5$

h)

$$\hat{p}_{st} = \sum W_h \hat{p}_h$$

$\hat{p}_h$: .143 – .545 – .688 – .800

$\hat{p}_{st} = .589$

$$\left[\hat{SE}(\hat{p}_{st})\right]^2 = \sum_h W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{\hat{p}_h(1-\hat{p}_h)}{n_h - 1}$$

$= .000302 + .001618 + .000959 + .001899$

$= .004778$

$SE(\hat{p}_{st}) = .069$

95% CI : $0.589 \pm 1.96 \times .069 = .589 \pm .135$

$= (.454, .724)$

i) Optimal allocation did not increase the precision

The reason: Optimal alloc. was with repect to: y_i = no of refereed publ., and not with respect to : $y_i = 1/0$ according to no refereed publ.

3. $C = 120000$, $c_0 = 30000$

2 strata : Stratum 1: have phone $W_1 = 0.9$

Stratum 2: does not have phone $W_2 = 0.1$

$\sigma_1^2 = \sigma_2^2$

a) $c_1 = c_2 = 180$

Choose n_1, n_2 to minimize variance of estimator given the cost.

$$n_h = \frac{C - c_0}{c_1} \cdot \frac{W_h}{W_1 + W_2} = W_h (C - c_0)/c_1$$

$\Rightarrow n_1 = 0.9 \times \frac{90000}{180} = 0.9 \times 500 = 450$

$n_2 = 0.1 \times \frac{90000}{180} = 50$

Proportional allocation $\sum n = 500$

b) Now $c_1 = 60$, $c_2 = 240$

& $n_h = (C - c_0) \dfrac{W_h}{\sqrt{c_h}} \cdot \dfrac{1}{(W_1\sqrt{c_1} + W_2\sqrt{c_2})}$

$= (C - c_0) \dfrac{W_h}{\sqrt{c_h}} \cdot \dfrac{1}{8.52056}$

$\Rightarrow n_1 = 1227.3 \Rightarrow \underline{n_1 = 1227}$

$n_2 = 68.2 \qquad \underline{n_2 = 68} \qquad \underline{n = 1295}$

c) Certainly the one in part b): Get to collect

795 more observations

d) Use sampl. method from part (b)

$\sigma_2^2 = 2\sigma_1^2$

$\Rightarrow$ ~~$\sum_h$~~ $\sum W_h \sigma_h \sqrt{c_h} = W_1 \sigma_1 \sqrt{c_1} + W_2 \sigma_1 \sqrt{2 c_2}$

$= \sigma_1 (W_1 \sqrt{c_1} + W_2 \sqrt{2c_2})$

$\Rightarrow n_1 = \dfrac{W_1}{\sqrt{c_1}} \cdot \dfrac{C - c_0}{(W_1\sqrt{c_1} + W_2\sqrt{2c_2})}$

$n_2 = \dfrac{\sqrt{2}\, W_2}{\sqrt{c_2}} \cdot \dfrac{C - c_0}{(W_1\sqrt{c_1} + W_2\sqrt{2c_2})}$

$W_1\sqrt{c_1} + W_2\sqrt{2c_2} = \cancel{8.52056}\ 9.16226$

$\Rightarrow n_1 = 30000 \dfrac{W_1/\sqrt{c_1}}{\cancel{8.52056}\ 9.16226} = 1141.3 \approx \underline{1141}$

$\underline{n_2 = 89.7 \approx 90}$

4. For unit $i \in U_h$: $\pi_{hi} = n_h / N_h$

$\Rightarrow \hat{t}_{HT} = \sum_{h=1}^{H} \sum_{i \in s_h} \dfrac{y_{hi}}{\pi_{hi}} = \sum_{h=1}^{H} \sum_{i \in s_h} \dfrac{N_h}{n_h} y_{hi}$

$= \sum_{h=1}^{H} N_h \cdot \dfrac{1}{n_h} \sum_{i \in s_h} y_{hi} = \sum_{h=1}^{H} N_h \cdot \bar{y}_h = \hat{t}_{st}$

Solution for R-exercise 3

a.

```
>y=trees$Height
>x=trees$Girth
> x1=as.numeric(x<11.1)
> x2=as.numeric(x<14.6)-x1
> x3=as.numeric(x>14.5)
> stratum=x1+2*x2+3*x3
> stratum
 [1] 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3
> tapply(y,stratum,mean)
       1        2        3
71.87500 76.13333 79.87500
```

b.

```
>y1=y[stratum==1]
>y2=y[stratum==2]
>y3=y[stratum==3]

> N1=8
> N2=15
> N3=8
> n1=3
> n2=5
> n3=3
> s1=sample(N1,n1)
> s2=sample(N2,n2)
> s3=sample(N3,n3)
> y1s=y1[s1]
> y2s=y2[s2]
> y3s=y3[s3]
> t_hat1=N1*mean(y1[s1])
> t_hat2=N2*mean(y2[s2])
> t_hat3=N3*mean(y3[s3])
> t_hat=t_hat1+t_hat2+t_hat3
> muhat=t_hat/31
>muhat
[1] 73.66667

>varest1=N1^2*var(y1s)*(N1-n1)/(N1*n1)
>varest2=N2^2*var(y2s)*(N2-n2)/(N2*n2)
>varest3=N3^2*var(y3s)*(N3-n3)/(N3*n3)
>se=sqrt(varest1+varest2+varest3)
>semean=se/31
>semean
[1] 1.049641

>CI=muhat+qnorm(c(0.025,0.975))*semean
> CI
[1] 71.60941 75.72392
#sjekk:
> mean(y)
[1] 76  # CI inkluderte ikke sann verdi!
```

c.

```
> z=c(y1s,y2s,y3s)
> mean(z)
[1] 73.54545
> sesrs=sqrt(var(z)*(31-11)/(31*11))
> sesrs
[1] 1.384301
# much larger that the stratified SE

> CI=mean(z)+qnorm(c(0.025,0.975))*sesrs
> CI
[1] 70.83227 76.25863
# Since wider than stratified CI: does include the true mean 76.
```

d.

```
>s=sample(31,11)
> mu_srs=mean(y[s])
> mu_srs
[1] 77.45455
> varest=31^2*var(y[s])*(31-11)/(31*11)
> se=sqrt(varest)
> semeans=se/31
> semeans
[1] 1.227083
> CI_srs=mu_srs+qnorm(c(0.025,0.975))*semeans
> CI_srs
[1] 75.04951 79.85958

# CI_srs is wider but does include the true value
```