STK 4600

Solutions final exam Spring 2016

1a) $n = 200,\ N = 10000$

- Frauds are probably correlated to gross income

Propl allocation:

$$n_h = n \cdot \frac{N_h}{N}$$

$n_1 = 100$

$n_2 = 80$

$n_3 = 20$

b)

$$\hat{p} = \frac{1}{N} \sum_{h=1} N_h \cdot \hat{p}_h = \underline{.055}$$

$$V(\hat{p}) = \frac{1-f}{n} \sum_h \frac{N_h}{N} \sigma_h^2 \quad , \quad \sigma_h^2 = \frac{N_h p_h (1-p_h)}{N_h - 1}$$

$$\hat{\sigma}_h^2 = \frac{n_h}{n_h - 1} \hat{p}_h (1 - \hat{p}_h)$$

$$\Rightarrow \hat{V}(\hat{p}) = \frac{1-f}{n} \sum_h \frac{N_h}{N} \cdot \frac{n_h \hat{p}_h (1-\hat{p}_h)}{n_h - 1}$$

$$= \frac{.98}{200} (0.5 \times 0.0294 + 0.4 \cdot 0.0481 + 0.1 \cdot 0.1684)$$

$$= 2.4882 \times 10^{-4}$$

$$\Rightarrow SE = 1.577 \times 10^{-2} = \underline{0.0158}$$

and 95% $CI = 0.055 \pm 1.96 \times .0158 = \underline{.055 \pm 0.031}$

$$= \underline{(0.024,\ 0.086)}$$

c) We are 95% certain that true $p \in (.024, 0.086)$

Usual interpretation by repeated sampling

d) Optimal allocation: $n_h = n \cdot \dfrac{N_h \sigma_h^0}{\sum_{k=1}^{3} N_k \sigma_k^0}$

$$\sigma_h^\circ = \sqrt{\frac{N_h \cdot p_h^\circ (1-p_h^\circ)}{N_h - 1}}$$

$p_1^\circ = .01, \quad p_2^\circ = .05, \quad p_3^\circ = .15$

		$N_h \cdot \sigma_h^\circ$
$\Leftrightarrow$	$\sigma_1^\circ = .0995$	497.5
	$\sigma_2^\circ = .2180$	872.0
	$\sigma_3^\circ = .35725$	357.25
		sum: 1726.75

Optimal allocation

$n_1 = (497.5/1726.75) \times 200 = 57.6 \approx \underline{58}$

$n_2 = 200 \cdot 872.0/1726.75 = 101.0 = \underline{101}$

$n_3 = 200 \cdot 357.25/\text{"} \approx 41.4 = \underline{41}$

e) $\hat{p}_1 = .034, \quad \hat{p}_2 = .069, \quad \hat{p}_3 = .171$

$\hat{p} = \frac{1}{N} \sum N_h \cdot \hat{p}_h = \underline{.0617}$

$$\hat{V}(\hat{p}) = \frac{1}{N^2} \sum_h N_h^2 \cdot \frac{1-f_h}{\cancel{n_h}} \cdot \frac{\cancel{n_h}}{n_h - 1} \hat{p}_h (1-\hat{p}_h)$$

$$= \frac{1}{N^2} \Big[\underbrace{\cancel{N_1^2 \cdot 5 \times}}_{14238.16} + 10018.71 + 3398.67 \Big]$$

$= 27655.54 / N^2$

$\Rightarrow SE = 166.29955 / 10\,000 = .0166$

95% CI $= .0617 \pm .0325 = \underline{(.029, .094)}$

f) Propl allocation:

$$V(\hat{p}) = \frac{1-f}{n} \sum_h \frac{N_h}{N} \cdot \frac{N_h p_h^* (1-p_h)}{N_h - 1}$$

$= .049 \cdot (.00495 + .01900 + \overset{.01276}{\cancel{.2763}}) = \cancel{.06}$

$= 1.799 \times 10^{-3} = \underline{17.99 \times 10^{-4}}$

Optimal allocation.

$$V(\hat{p}_h) = \frac{p_h(1-p_h)}{n_h} \cdot \frac{N_h - n_h}{N_h - 1}$$

$$\Rightarrow V(\hat{p}) = \frac{1}{N^2} \sum N_h^2 \cdot V(\hat{p}_h)$$

$$= \frac{1}{N^2}(4048.137 + 7336.587 + 2985.24) = \frac{14369.96}{N^2}$$

$$= \underline{\underline{1.437 \times 10^{-4}}}$$

much smaller than for propl allocation

2a) Looks like variance is close to constant, so choose Model 2

Actually, may increase with x some so Model 1 is also ok

b) Model 1, ratio model

BLU est: $\hat{\bar{y}}_1 = \hat{\beta} \cdot \bar{x}$, $\hat{\beta} = \bar{y}_s / \bar{x}_s = 107.4/9.4$
$= 11.4255$

$\Rightarrow \hat{\bar{y}}_1 = \underline{117.7}$

$$\hat{V}(\hat{\bar{y}}_1 - \bar{y}) = \frac{1-f}{n} \cdot \frac{\bar{x}_r \cdot \bar{x}}{\bar{x}_s} \hat{\sigma}_1^2 \qquad \hat{\sigma}_1^2 = 32.30$$

$$\bar{x}_r = \frac{N\bar{x} - n\bar{x}_s}{N-n} = 10.316$$

$\Rightarrow \hat{V} = 17.933$ and $\underline{SE = 4.235}$

95% CI $= 117.7 \pm 1.96 \times 4.235 = 117.7 \pm 8.3$
$= \underline{(109.4, 126.0)}$

c) Model 2

$$\hat{\beta} = \sum_s x_i y_i / v(x_i) \Big/ \sum_s x_i^2 / v(x_i) = \sum_s x_i y_i \Big/ \sum_s x_i^2$$

$$= 20980.4/1832.63 = 11.4484$$

BLU for T:

$$\hat{T}_2 = \sum_s y_i + \hat{\beta} \sum_{i \notin s} x_i \qquad \text{and } \hat{\bar{Y}}_2 = \hat{T}_2 / N$$

$$\hat{T}_2 = 20 \times 107.4 + 11.4484 \cdot (N-n)\bar{x}_r = 133477.08$$

$$\Rightarrow \hat{\bar{Y}}_2 = 117.9$$

$$V(\hat{T}_2 - T) = \sigma^2 \left(\frac{(\sum_{i \notin s} x_i)^2}{\sum_s x_i^2 / v(x_i)} + \sum_{i \notin s} v(x_i) \right)$$

$$= \sigma^2 \cdot \left(\frac{(N-n)^2 \bar{x}_r^2}{\sum_s x_i^2} + (N-n) \right)$$

$$\Rightarrow \hat{V}(\hat{T}_2 - T) = 321.85 \cdot (74411 + 1112) = 24307077.55$$

$$\Rightarrow \hat{V}(\hat{\bar{Y}}_2 - \bar{Y}) = \hat{V}(\hat{T}_2 - T)/N^2 = 18.9688$$

and $SE(\hat{\bar{Y}}_2 - \bar{Y}) = 4.355$

and 95% $CI = 117.9 \pm 8.5 = (109.4, 126.4)$

practically the same as for Model 1

d) $E(\bar{Y}_s - \bar{Y}) = \beta \cdot \bar{x}_s - \beta \bar{x} = \beta(\bar{x}_s - \bar{x})$

model-unbiased $\Leftrightarrow \bar{x}_s = \bar{x}$

Is the BLU estimator under Model 1:

$$\text{BLU}: \hat{\bar{Y}} = \hat{\beta} \cdot \bar{X} = \hat{\beta} \cdot \bar{x}_s = \bar{y}_s$$

e) Under model $E(Y_i) = \beta$, $V(Y_i) = \sigma^2$

$\bar{Y}_s$ is model-unbiased

and $V(\bar{Y}_s - \bar{Y}) = V\{(\frac{1}{n} - \frac{1}{N}) \sum y_i\} + \frac{1}{N^2}(N-n)\sigma^2$

$\Rightarrow$

$$V(\bar{Y}_s - \bar{Y}) = \left(\frac{N-n}{nN}\right)^2 n\sigma^2 + \frac{N-n}{N^2}\sigma^2$$

$$= \frac{N-n}{N} \cdot \frac{\sigma^2}{n}\left(\frac{N-n}{N} + \frac{n}{N}\right)$$

$$= (1-f) \cdot \frac{\sigma^2}{n}$$

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_s (y_i - \bar{y}_s)^2 = 821.52$$

$\Rightarrow \hat{V} = 40.35$ and $\underline{SE = 6.35}$ > for Model 1, 2

$\Rightarrow$ Including x is important, since clearly x and y are positively correlated. It corrects estimate (from 107.41) and reduces SE by at least 30%

f) $y_i = \beta x_i + \varepsilon_i, \quad V(\varepsilon_i) = \sigma^2 x_i^2$

$\Rightarrow Z_i = \frac{y_i}{x_i} = \beta + \varepsilon_i^* \quad \& \quad V(\varepsilon_i^*) = \sigma^2$

BLU est: $\hat{\beta} = \frac{\sum x_i Z_i / v(x_i)}{\sum x_i^2 / v(x_i)} = \bar{Z}_s$

and $\hat{T} = n\bar{Z}_s + (N-n)\bar{Z} = N \cdot \bar{Z}_s$

g) and BLU-est for $\bar{R} = \hat{\bar{R}} = \bar{Z}_s = \underline{11.388}$

$$V(\bar{Z}_s - \bar{R}) = (1-f)\frac{\sigma^2}{n}$$

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum (z_i - \bar{z}_s)^2 = \frac{1}{19}(2656.8 - 20 \times 11.388^2) = \underline{3.32}$$

$SE = \sqrt{.163} = \underline{.40}$

95% for $\bar{R}$: $11.39 \pm 1.96 \times .4 = 11.39 \pm 0.78 =$

$= \underline{(10.61, 12.17)}$

h) When $\underline{y_i = c x_i \ \forall i}$:

$$\bar{R} = \frac{1}{N} \sum_{1}^{N} \frac{y_i}{x_i} = c = \frac{\bar{y}}{\bar{x}}$$

Model 1, 95% for $\bar{y}/\bar{x}$

$$= \frac{1}{10.3} (109.4, 126.0) = (10.6, 12.2)$$

Model 2, 95% for $\bar{y}/\bar{x}$:

$$\frac{1}{10.3} (109.4, 126.4) = (10.6, 12.3)$$

Actually "identical" to 95% CI for $\bar{R}$

– – –