

1**(a)**

i: för att få en totalskattning vid stratifierat urval får man summera över stratumens totalskattningar

$$\hat{\tau}_{ST} = \sum_{l=1}^L \hat{\tau}_l$$

ii: för att få en variansskattning för totalskattning vid stratifierat urval får man summera över stratumens variansskattningar

$$\hat{V}(\hat{\tau}_{ST}) = \sum_{l=1}^L \hat{V}(\hat{\tau}_l)$$

iii: dragningarna i strata är oberoende, varför det inte finns några kovarianstermer i uttrycket för variansskattning

$$Cov(\hat{\tau}_l, \hat{\tau}_k) = 0, \quad \forall k \neq l \text{ (där } k \text{ och } l \text{ betecknar olika strata)}$$

(b)

Se bladet utdelat vid F3

2

(a)

FR, s. 30

(b)

Ja, det borde hon: för att förhindra att primacy och recency effekter slår ut mer på vissa element i listan än på andra krävs randomisering (mer om primacy och recency effekter på t.ex. OH från F2).

3

Pop: $N = 15623$, $N_1 = 3593$, $N_2 = 5937$, $N_3 = 6093$.

Usv: y (en 0/1-variabel)

Param: total i undergrupp,

$$\tau_3 = \sum_{i=1}^{N_3} y_{3i}$$

Urv: OSU u.å., $n = 400$, $n_1 = 92$, $n_2 = 158$, $n_3 = 150$

(i) Fallet N_3 känt: Paramsk:

$$\hat{\tau}_{3i} = N_3 \frac{\sum_{i=1}^{n_3} y_{3i}}{n_3} = N_3 \bar{y}_3 = N_3 \hat{p}_3.$$

Varsk:

$$\hat{V}(\hat{\tau}_{3i}) \underset{\substack{= \\ \uparrow \\ \text{Från formelbladet}}}{=} N_3^2 \frac{N_3 - n_3}{N_3} \frac{s_{y3}^2}{n_3} \underset{\substack{= \\ \uparrow \\ y - 0/1\text{-variabel}}}{=} N_3^2 \frac{N_3 - n_3}{N_3} \frac{\hat{p}_3 \hat{q}_3}{n_3 - 1}$$

där $\hat{q}_3 = 1 - \hat{p}_3$

(ii) Fallet N_3 inte känt:

$$\hat{\tau}_{3ii} = N \frac{\sum_{i=1}^{n_3} y_{3i}}{n}.$$

(i)

(ii)

$$\hat{V}(\hat{\tau}_{3ii}) \stackrel{\substack{= \\ \uparrow \\ \text{Från formelbladet}}}{=} N^2 \frac{N-n}{N} \frac{\sum_{i=1}^{n_3} y_{3i}^2 - \frac{(\sum_{i=1}^{n_3} y_{3i})^2}{n}}{n(n-1)} \stackrel{\substack{= \\ \uparrow \\ y=0/1\text{-variabel}}}{=} \\ = N^2 \frac{N-n}{N} \frac{n_3 \hat{p}_3 - \frac{(n_3 \hat{p}_3)^2}{n}}{n(n-1)} = N^2 \frac{N-n}{N} \frac{n_3 \hat{p}_3 \left(1 - \frac{n_3 \hat{p}_3}{n}\right)}{n(n-1)}$$

Data

stratum	1	2	3	summa
Ni	3593	5937	6093	15623
"1"	64	94	67	
"0"	28	64	83	
ni	92	158	150	400
p="1"/ni	0.695652	0.594937	0.446667	

(i)

$$\hat{\tau}_{3i} = N_3 \hat{p}_3 = 6093 \times 0.4467 \approx 2721$$

$$\hat{V}(\hat{\tau}_{3i}) = N_3^2 \frac{N_3 - n_3}{N_3} \frac{\hat{p}_3 \hat{q}_3}{n_3 - 1} = 6093^2 \times \frac{6093 - 150}{6093} \times \frac{0.446666667 \times (1 - 0.446666667)}{150 - 1} = 60065$$

$$B_i = 2\sqrt{\hat{V}(\hat{\tau}_{3i})} = 2 \times \sqrt{60065} \approx 490$$

(ii)

$$\hat{\tau}_{3ii} = N \frac{\sum_{i=1}^{n_3} y_{3i}}{n} = 15623 \times \frac{67}{400} \approx 2617$$

$$\hat{V}(\hat{\tau}_{3ii}) = N^2 \frac{N-n}{N} \frac{n_3 \hat{p}_3 \left(1 - \frac{n_3 \hat{p}_3}{n}\right)}{n(n-1)} = 15623^2 \times \frac{15623 - 400}{15623} \times \\ \times \frac{150 \times 0.446666667 \times \left(1 - \frac{150 \times 0.446666667}{400}\right)}{400 \times (400 - 1)} = 83117$$

$$B_{ii} = 2\sqrt{\hat{V}(\hat{\tau}_{3ii})} = 2\sqrt{83117} \approx 577$$

4

Pop: $N = 15623$, $N_1 = 6093$, $N_2 = 5937$, $N_3 = 6093$.

Usv: y (en 0/1-variabel)

Param: medelvärde,

$$\mu = \sum_{i=1}^N y_i$$

Urv: OSU u.å., $n = 400$, $n_1 = 92$, $n_2 = 158$, $n_3 = 150$

Paramsk:

(a)

Utan kännedom om undergruppernas storlekar använder vi den enkla punktestimatoren

$$\hat{\mu} = \frac{\sum_{i=1}^n y_i}{n} = \hat{p}$$

till vilket hör variansestimatoren

$$\hat{V}(\hat{\mu}) \underset{\substack{= \\ \uparrow \\ \text{Från formelbladet, } y=0/1}}{\quad} \hat{V}(\hat{p}) = \frac{N-n}{N} \frac{\hat{p}\hat{q}}{n-1}$$

(b)

Den estimator som använder efterstratifiering uttrycks som

$$\hat{\mu}_{post} = \frac{1}{N} \hat{\tau}_{post} = \frac{1}{N} \sum_{g=1}^G N_g \hat{\mu}_g \underset{\substack{= \\ \uparrow \\ y=0/1\text{-variabel}}}{\quad} \frac{1}{N} \sum_{g=1}^G N_g \hat{p}_g$$

vars variansskattning är något komplex men ofta förenklas till

$$\hat{V}(\hat{\mu}_{post}) \approx \hat{V}(\hat{\mu}_{ST}) \stackrel{\substack{= \\ \uparrow \\ \text{Från formelbladet}}}{=} \frac{1}{N^2} \sum_{g=1}^G N_g^2 \frac{N_g - n_g}{N_g} \frac{s_g^2}{n_g} \stackrel{\substack{= \\ \uparrow \\ y = 0/1\text{-variabel}}}{=}$$

$$= \frac{1}{N^2} \sum_{g=1}^G N_g^2 \frac{N_g - n_g}{N_g} \frac{\hat{p}_g \hat{q}_g}{n_g - 1}$$

Data

stratum	1	2	3	summa
Ng	3593	5937	6093	15623
"1"	64	94	67	225
"0"	28	64	83	
ng	92	158	150	400
pg="1"/ng	0.695652	0.594937	0.446667	
Ng*pg	2499.478	3532.139	2721.54	8753.158
Ng^2	12909649	35247969	37124649	
(Ng-ng)/Ng	0.974395	0.973387	0.975382	
pg*ng/(ng-1)	0.002327	0.001535	0.001659	
Ng^2*((Ng-ng)/Ng)*pg*ng/(ng-1)	29266.47	52663.99	60064.94	141995.4

(a)

$$\hat{\mu} = \hat{p} = \frac{\sum_{i=1}^n y_i}{n} = \frac{225}{400} = 0.5625$$

$$\hat{V}(\hat{p}) = \frac{N-n}{N} \frac{\hat{p}\hat{q}}{n-1} = \frac{15623-400}{15623} \times \frac{0.5625 \times (1-0.5625)}{400-1} = 6.0098 \times 10^{-4}$$

$$B_a = 2\sqrt{\hat{V}(\hat{p})} = 2 \times \sqrt{6.0098 \times 10^{-4}} \approx 0.05$$

(b)

$$\hat{p}_{post} = \frac{1}{N} \sum_{g=1}^G N_g \hat{p}_g = \frac{8753.157501}{15623} = 0.56$$

$$\hat{V}(\hat{p}_{post}) \approx \frac{1}{N^2} \sum_{g=1}^G N_g^2 \frac{N_g - n_g}{N_g} \frac{\hat{p}_g \hat{q}_g}{n_g - 1} = \frac{141995.3901}{15623^2} = 5.8176 \times 10^{-4}$$

$$B_b = 2\sqrt{\hat{V}(\hat{p}_{post})} = 2 \times \sqrt{5.8176 \times 10^{-4}} = 0.048$$

Inga speciella vinster med efterstratifieringen den här gången.

5

Pop: askar, $N = 10000$, där varje ask innehåller $k = 20$ artiklar

Usv: antal felaktiga artiklar i asken - y

Param: andel felaktiga artiklar i populationen

$$\theta = \frac{\mu}{k} = \frac{\sum_{i=1}^N y_i}{k \times N}$$

Urv: OSU u.å., $n = 30$

Paramsk:

$$\hat{\theta} = \frac{\hat{\mu}}{k} = \frac{\sum_{i=1}^n y_i}{k \times n} = \frac{\bar{y}}{k}$$

Varsk:

$$\hat{V}(\hat{\theta}) = \hat{V}\left(\frac{\hat{\mu}}{k}\right) = \frac{1}{k^2} \hat{V}(\hat{\mu}) \quad \underset{\substack{= \\ \uparrow \\ \text{Från formelbladet}}}{=} \frac{1}{k^2} \frac{N-n}{N} \frac{s_y^2}{n}$$

Data

Data kan lämpligen ställas upp på följande vis:

Ask nr:	Ant fel
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	1
15	1
16	1
17	1
18	1
19	1
20	1
21	1
22	1
23	1
24	1
25	1
26	1
27	1
28	1
29	2
30	3
ybar=	0.666667
s2=	0.505747

$$\hat{\theta} = \frac{\bar{y}}{k} = \frac{0.666666667}{20} = 0.033333$$

$$\hat{V}(\hat{\theta}) = \frac{1}{k^2} \frac{N-n}{N} \frac{s_y^2}{n} = \frac{1}{20^2} \times \frac{10000-30}{10000} \times \frac{0.505747126}{30} = 4.2019 \times 10^{-5}$$

Standardavvikelsen blir då $\sqrt{\hat{V}(\hat{\theta})} = \sqrt{4.2019 \times 10^{-5}} \approx 0.0065$

Ett alternativt sätt att ställa upp denna uppgift:

Beteskningsättet från föreläsningarna används.

Pop: askar, $M = 10000$, som innehåller $N = M \times k = 10000 \times 20 = 200000$ element.

Usv: huruvida elementet är felaktigt eller ej - y (en 0/1-variabel)

Param: andel felaktiga artiklar i populationen

$$\mu = \frac{\sum_{i=1}^N y_i}{N} = \frac{\sum_{i=1}^M \sum_{j=1}^{N_i} y_{ij}}{k \times M}$$

Urv: OSU u.å., $m = 30$

Paramsk:

$$\hat{\mu} = \frac{\sum_{i=1}^m \sum_{j=1}^{N_i} y_{ij}}{k \times m}$$

Varsk:

$$\hat{V}(\hat{\mu}) = \hat{V}\left(\frac{\hat{\tau}_{HT}}{N}\right) = \underset{\substack{\uparrow \\ \text{Från formelbladet}}}{=} \frac{1}{N^2} M^2 \frac{M-m}{M} \frac{s_{\hat{\tau}}^2}{m}$$

Data

(Som ovan)

$$\hat{\mu} = \frac{\sum_{i=1}^m \sum_{j=1}^{N_i} y_{ij}}{k \times m} = \frac{20}{20 \times 30} = 0.033333$$

$$\hat{V}(\hat{\mu}) = \frac{1}{N^2} M^2 \frac{M-m}{M} \frac{s_{\hat{\tau}}^2}{m} = \frac{1}{200000^2} \times 10000^2 \times \frac{10000-30}{10000} \times \frac{0.505747126}{30} = 4.2019 \times 10^{-5}$$