Sida 1 | 0076-JTS

Tenta undersökningsmetodik 210212

Uppgift 1 a) x = proportionen elever i kommunen som är positivt inställda till skolans salladsbord.

$$\hat{x}_{str} = \Sigma_k W_k \hat{x}_k$$

antal positivt inställda

$$W_k = N_k/N \,, \quad \hat{x}_k = a_k/n_k$$

$$a_1/n_1 = 48/60 = 0{,}8 \qquad W_1 = 350/900 \approx 0{,}389$$
$$a_2/n_2 = 25/50 = 0{,}5 \qquad W_2 = 300/900 \approx 0{,}333$$
$$a_3/n_3 = 8/40 = 0{,}2 \qquad W_3 = 250/900 \approx 0{,}278$$

$$\Sigma_k W_k \hat{x}_k = 0{,}389 \cdot 0{,}8 + 0{,}333 \cdot 0{,}5 + 0{,}278 \cdot 0{,}2 = 0{,}5333$$

$$\hat{V}(\hat{x}_{str}) = \Sigma_k W_k^2 \left(1 - \frac{n_k}{N_k}\right) \frac{\hat{x}_k(1-\hat{x}_k)}{n_k - 1}$$

$$W_1^2 = \left(\frac{350}{900}\right)^2 \approx 0{,}151$$

$$W_2^2 = \left(\frac{300}{900}\right)^2 = 1/9$$

$$W_3^2 = \left(\frac{250}{900}\right)^2 \approx 0{,}772$$

$$\Sigma_k W_k^2 \left(1 - \frac{n_k}{N_k}\right) \frac{\hat{x}_k(1-\hat{x}_k)}{n_k - 1} =$$

$$= 0{,}151\left(1 - \frac{60}{350}\right)\left(\frac{0{,}8 \cdot 0{,}2}{59}\right) + \frac{1}{9}\left(1 - \frac{50}{300}\right)\left(\frac{0{,}5 \cdot 0{,}5}{49}\right) +$$

$$+ 0{,}772\left(1 - \frac{40}{250}\right)\left(\frac{0{,}2 \cdot 0{,}8}{39}\right) = 0{,}151 \cdot 0{,}829 \cdot 0{,}00271 +$$

$$+ \frac{1}{9} \cdot 0{,}833 \cdot 0{,}00510 + 0{,}772 \cdot 0{,}84 \cdot 0{,}00410 =$$

$$= 0{,}000339 + 0{,}000472 + 0{,}00266 = 0{,}00347$$

$$\sqrt{\hat{V}(\hat{x}_{str}} = \sqrt{0{,}00347} \approx 0{,}0589$$

Tenta undersökningsmetodik 210212

fortsätter uppgift 1 a

$$95\%\ \text{K.I.} = \hat{x}_{str} \pm Z_{\frac{\alpha}{2}} \cdot \sqrt{\hat{V}(\hat{x}_{str})} =$$

$$= 0{,}533 \pm 1{,}96 \cdot 0{,}0589 \longrightarrow 0{,}42$$

$$\longrightarrow 0{,}65$$

Svar:
- Proportionen positivt inställda elever uppskattas till 0,53, dvs 53 %.
- Det 95-procentiga konfidensintervallet är mellan 0,42 och 0,65.

b) Vi kan göra ett proportionellt stratifierat urval, vilket innebär att urvalen i resp. stratum har samma proportioner som strata och population. I detta fall ska alltså lågstadieurvalet ha samma proportion som lågstadiet / hela grundskolan, osv.
Urvalsstorleken är fortfarande $60 + 50 + 40 = 150$
Urvalsstorlekarna blir alltså:

$$n_1 = W_1 \cdot 150 \approx 58$$

$$n_2 = W_2 \cdot 150 = 50$$

$$n_3 = W_3 \cdot 150 \approx 42$$

Svar: Den optimerade urvalsallokeringen är 58 lågstadieelever, 50 mellanstadieelever och 42 högstadieelever.

Tenta undersökningsmetodik 210212

Uppgift 2 $N = 4260 \quad n = 300 \quad Y = \text{inkomst}$

a) $$\hat{\mu}_Y = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{19575000}{300} = 65250$$

$$95\%\ \text{K.I.} = \mu \pm Z_{\frac{0{,}05}{2}} \cdot \sqrt{\left(1-\frac{n}{N}\right)\cdot\frac{s^2}{n}}$$

$$s^2 = \frac{\sum_{i=1}^{n} \cancel{x}^2 - n\bar{y}^2}{n-1} = \frac{1{,}28\cdot 10^{12} - 300\cdot 65250^2}{299} =$$

$$= \frac{2731250000}{299} = 9134615{,}385$$

$$\mu \pm Z_{\frac{0{,}05}{2}} \cdot \sqrt{\left(1-\frac{n}{N}\right)\cdot\frac{s^2}{n}} = 65250 \pm 1{,}96\cdot\sqrt{\left(1-\frac{300}{4260}\right)\cdot\frac{9134615}{300}} =$$

$$= 65250 \pm 1{,}96\cdot\sqrt{28304{,}44204} \approx$$

$$\approx 65250 \pm 1{,}96\cdot 168 \rightarrow 64920, \quad 65579$$

Svar: Medelinkomsten i kommunen uppskattas till 65250 och det 95-procentiga konfidensintervallet är mellan 64 920 och 65 579.

b)

Tenta undersökningsmetodik 210212

Uppgift 3

a) Nu tolkar jag alltså frågan som hur man minskar effekterna av bortfallet, inte omfattningen.

Om det har blivit bortfall trots försök att förebygga det kan man använda efterstratifiering för att kompensera för den snedvridning som uppstår om det finns systematiska fel (läs: mönster) i bortfallet, vilket det ofta gör.

b) Snöbollsurval är EJ slumpmässigt.
Systematiskt urval ÄR slumpmässigt.
Stratifierat urval ÄR slumpmässigt.
Tvåstegsurval ÄR slumpmässigt.
Bekvämlighetsurval är EJ slumpmässigt.

c) Teleskopeffekt är ett minnesfel som kännetecknas av att en respondent "flyttar en händelse i tiden", dvs antingen rapporterar att något hänt inom en specificerad tidsperiod som egentligen inträffade tidigare (teleskopeffekt framåt) eller att något inträffat under en period som egentligen inträffade senare (teleskopeffekt bakåt).

Tenta undersökningsmetodik 210212

uppgift 3 d)

Man kan kortfattat säga att övertäckning utgörs av element* som ingår i rampopulationen men inte i målpopulationen, och omvänt att undertäckning utgörs av element* som ingår i målpopulationen men inte i rampopulationen.

Undertäckning anses vara det allvarligare av de två täckningsfelen eftersom det ofta leder till systematiska fel då de avviker från målpopulationen på något särskilt sätt.

* Jag menar urvalsenheter

Man kan ofta avgränsa populationen vid övertäckning, vilket minimerar problematiken. Om det visar sig att övertäckningen är av betydande omfattning måste emellertid många urvalsenheter strykas, vilket kan leda till att urvalet som faktiskt blir kvar är mindre än planerat, och då förlorar man naturligtvis precision när det väl ska räknas på datan.

Tenta undersökningsmetodik 210212

uppgift 4

a) Jag väljer att använda kvotskattning

b) $N = 452 \quad n = 20 \quad \tau_x = 2162560$ kr

$$\hat{R} = \frac{\sum y_k}{\sum x_k} = \frac{101810}{95060} \approx 1{,}07101$$

$$\hat{\tau}_{krot} = \hat{R} \cdot \tau_x = 1{,}07101 \cdot 2162560 \approx 2316019$$

Svar: ca 2 316 000 kronor

c) $$\hat{V}(\hat{\tau}_{krot}) = N^2\left(1 - \frac{n}{N}\right) \cdot \frac{1}{n} \cdot \frac{\sum (y_i - \hat{R}x_i)^2}{n-1} =$$

$$= 452^2\left(1 - \frac{20}{452}\right) \cdot \frac{1}{20} \cdot \frac{\sum (y_i - \hat{R}x_i)^2}{19}$$

$$\sum (y_i - \hat{R}x_i)^2 = \sum y_i^2 - 2\hat{R}\sum x_i y_i + \hat{R}^2 \sum x_i^2$$

$$\sum y_i^2 = 660902900$$

$$\sum x_i^2 = 580896200$$

$$\sum x_i y_i = 619400100$$

$$\sum y_i^2 - 2\hat{R}\sum x_i y_i + \hat{R}^2\sum x_i^2 = 660902900 - 2 \cdot 1{,}07101 \cdot$$

$$\cdot 619400100 + 1{,}07101^2 \cdot 660902900 =$$

$$= 660902900 - 1326764658 + 758093743{,}6 =$$

$$= 92231986$$

$$452^2\left(1 - \frac{20}{452}\right) \cdot \frac{1}{20} \cdot \frac{92231986}{19} = 9763{,}2 \cdot 4854315 \approx \boxed{4{,}74 \cdot 10^{10}}$$

0076-JTS

Tenta undersökningsmetodik 210212

Uppgift 5 a) $\mu_Y = \frac{\sum_{i=1}^{N} y_i}{N} = (1+2+3+7+8+9)/6 = 5$

$$\sigma_Y^2 = \frac{\sum_{i=1}^{N} y_i^2 - n\bar{y}^2}{N} = \frac{(1^2+2^2+3^2+7^2+8^2+9^2) - 6\cdot 5^2}{6} = \frac{58}{6} \approx 9{,}67$$

$\boxed{\text{Svar: } \mu_Y = 5,\ \sigma_Y^2 = 9{,}67}$

b) inklusionssannolikheten är lika med antal urval där elementet är med delat på det totala antalet urval.
Samtliga element förekommer i tre av urvalen, så inklusionssannolikheten blir 3/9 = 1/3 för alla element.
Ja, det är ett sannolikhetsurval eftersom varje element har en känd inklusionssannolikhet som är större än 0.

c) $\bar{y}_1 = (1+7)/2 = 4$, $\bar{y}_2 = (1+8)/2 = 4{,}5$, $\bar{y}_3 = (1+9)/2 = 5$,
$\bar{y}_4 = (2+7)/2 = 4{,}5$, $\bar{y}_5 = (2+8)/2 = 5$, $\bar{y}_6 = (2+9)/2 = 5{,}5$,
$\bar{y}_7 = (3+7)/2 = 5$, $\bar{y}_8 = (3+8)/2 = 5{,}5$, $\bar{y}_9 = (3+9)/2 = 6$

$$\sigma_{\bar{Y}}^2 = \frac{\sum_i (\bar{y}_i - \bar{\bar{y}})^2}{N}$$

$$\bar{\bar{y}} = (4+4{,}5+5+4{,}5+5+5{,}5+5+5{,}5+6)/9 = 5$$

$$\frac{\sum_i (\bar{y}_i - \bar{\bar{y}})^2}{N} = \frac{1^2+0{,}5^2+0+0{,}5^2+0+0{,}5^2+0+0{,}5^2+1^2}{9} = \frac{1}{3}$$

d) Antal möjliga stickprov med OSU uå är $\binom{6}{2} = 15$

Forts på nästa sida

Tenta undersökningsmetodik 210212

Forts. 5d

stickprov	y_i
1,2	1,5
1,3	2
1,7	4
1,8	4,5
1,9	5
2,3	2,5
2,7	4,5
2,8	5
2,9	5,5
3,7	5
3,8	5,5
3,9	6
7,8	7,5
7,9	8
8,9	8,5
	75

$$\sigma^2 = \frac{\sum_{k \in U} Y_k^2 - N\mu_Y^2}{N} = \frac{433 - 15 \cdot \left(\frac{75}{15}\right)^2}{15} = \frac{433 - 375}{15} \approx 3,867$$

Svar: 3,867

e) OSU uå ger lägre varians.