

Uppgift 1

(a)

Se FR kap. 4, KD avsn. 2.5.

(b)

Olika sätt finns och kan med fördel kombineras:

- ge respondenten tid att återkalla sina minnen,
- led respondenten till att minnas referensperiodens början; t ex nämna exakta datum/dag då perioden började, om möjligt få respondenten att minnas/prata om tiden kring periodens början (t ex förankra i personliga händelser vid tidspunkten)
- få respondenten att redogöra för samma typ av händelser som de undersökta men i perioden innan, d v s perioden som slutade vid referensperiodens början.

(c)

Ja, t ex om variabeln tenderar att underrapporteras (“känslig fråga”); se vidare material utdelat vid F3.

Uppgift 2

I

a)

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

b)

$$\binom{N-1}{n-1} = \frac{(N-1)!}{(n-1)!(N-n)!},$$

ty i 's plats är fix i både U och s .

c) se material utdelat vid F4.

d)

$$\pi_i = \frac{\# \text{ gynsamma fall}}{\# \text{ möjliga fall}} = \frac{\frac{(N-1)!}{(n-1)!(N-n)!}}{\frac{N!}{n!(N-n)!}} = \frac{n}{N}$$

II

a)

$$MSE(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right]$$

b)

$$Bias(\hat{\theta}) = E(\hat{\theta}) - \theta$$

c)

$$\begin{aligned} MSE(\hat{\theta}) &= E\left\{[\hat{\theta} - \theta]^2\right\} \\ &= E\left\{[\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta]^2\right\} \\ &= E\left\{[\hat{\theta} - E(\hat{\theta})]^2 + [E(\hat{\theta}) - \theta]^2 + 2[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta]\right\} \\ &= E\left\{(\hat{\theta} - E(\hat{\theta}))^2\right\} + E\left\{[E(\hat{\theta}) - \theta]^2\right\} + 2E\left\{[\hat{\theta}E(\hat{\theta}) - \hat{\theta}\theta - E(\hat{\theta})E(\hat{\theta}) + E(\hat{\theta})\theta]\right\} \\ &= Var(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2 + 2[E(\hat{\theta})E(\hat{\theta}) - E(\hat{\theta})\theta - E(\hat{\theta})E(\hat{\theta}) + E(\hat{\theta})\theta] \\ &= Var(\hat{\theta}) + Bias^2(\hat{\theta}) \end{aligned}$$

Uppgift 3

I

a) Vi utgår ifrån att $\hat{V}(\hat{p}) \approx \frac{\hat{p}(1-\hat{p})}{n}$. Samtidigt ställs kravet att $B = 2\sqrt{\hat{V}(\hat{p})}$ inte ska överstiga något visst värde, c : $2\sqrt{\hat{V}(\hat{p})} \leq c$, dvs $B \leq c$. Vi då betraktar det mest extrema acceptabla fallet, $B = c$.

$$\begin{aligned} 2\sqrt{\hat{V}(\hat{p})} &= c \\ \hat{V}(\hat{p}) &= \frac{c^2}{4} \\ \frac{\hat{p}(1-\hat{p})}{n} &= \frac{c^2}{4} \\ \frac{.5^2}{n} &= \frac{c^2}{4} \\ n &= \frac{1}{c^2} \end{aligned}$$

Alltså, n ska vara $1/c^2$ eller större.

b)

$$n = \frac{1}{c^2} = \frac{1}{.05^2} = \frac{1}{.0025} = 400$$

II

Enligt uppgiften, $\hat{V}(\hat{p}) \approx \frac{\hat{p}(1-\hat{p})}{n}$, $c = .05$, $\hat{p} = .7$. Vi har då

$$\begin{aligned} 2\sqrt{\hat{V}(\hat{p})} &= c \\ \hat{V}(\hat{p}) &= \frac{c^2}{4} \\ \frac{\hat{p}(1-\hat{p})}{n} &= \frac{c^2}{4} \\ \frac{.7 \times .3}{n} &= \frac{.05^2}{4} \\ n &= \frac{.7 \times .3 \times 4}{.05^2} \\ &= 336 \end{aligned}$$

III

Sannolikheten är 1, ty vi genomförde en totalundersökning och fann att $\mu_M > \mu_K$.

Uppgift 4

a

Pop.: $N = 531705$

Usv.: huruvida individ i målpopulationen åkte tunnelbana under tidsperioden, y - en 0/1-variabel

Param.: andel, $p = \frac{1}{N} \sum_{i=1}^N y_i$

Urv.: OSU, $n = 1100$

Punktsk.:

$$\hat{p} = \bar{y}$$

Varsk.:

$$\hat{V}(\hat{p}) = \frac{N-n}{N} \times \frac{\hat{p}(1-\hat{p})}{n-1}$$

Data:

(som i uppgiften)

Svar:

$$\hat{p} = \bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{770}{1100} = .7$$

$$\begin{aligned} \hat{V}(\hat{p}) &= \frac{N-n}{N} \times \frac{\hat{p}(1-\hat{p})}{n-1} \\ &= \frac{531705-1100}{531705} \times \frac{.7(1-.7)}{1100-1} \\ &= 1.9069 \times 10^{-4} \end{aligned}$$

$$\begin{aligned} KI_{\hat{p}} &= \hat{p} \pm 2\sqrt{\hat{V}(\hat{p})} \\ &= .7 \pm 2 \times \sqrt{1.9069 \times 10^{-4}} \\ &= [.6724 ; .7276] \end{aligned}$$

b

Pop.: $N = 531705$

Usv.: huruvida en tunnelbaneåkande i målpopulationen erfarit trängsel under tidsperioden,
 y_1 - en 0/1-variabel

Param.: andel, $p_1 = \frac{1}{N_1} \sum_{i=1}^{N_1} y_{1i}$

Urv.: OSU, $n = 1100$, $n_1 = 770$.

Punktsk.:

$$\hat{p}_1 = \bar{y}_1 = \frac{\sum_{i=1}^{n_1} y_{1i}}{n_1}$$

Varsk.:

$$\hat{V}(\hat{p}_1) \underset{\substack{= \\ \uparrow \\ \text{Från formelbladet}}}{=} \frac{N_1 - n_1}{N_1} \frac{s_1^2}{n_1} \approx \frac{N - n}{N} \frac{s_1^2}{n_1} \underset{\substack{= \\ \uparrow \\ y - 0/1\text{-variabel}}}{=} \frac{N - n}{N} \frac{\hat{p}_1(1 - \hat{p}_1)}{n_1 - 1}$$

Data:

(som i uppgiften)

Svar:

$$\hat{p}_1 = \frac{\sum_{i=1}^{n_1} y_{1i}}{n_1} = \frac{616}{770} = .8$$

Varsk.:

$$\begin{aligned} \hat{V}(\hat{p}_1) &= \frac{N - n}{N} \frac{\hat{p}_1(1 - \hat{p}_1)}{n_1 - 1} = \frac{531705 - 1100}{531705} \times \frac{.8(1 - .8)}{770 - 1} \\ &= 2.0763 \times 10^{-4} \end{aligned}$$

$$\begin{aligned} KI_{\hat{p}_1} &= \hat{p}_1 \pm 2\sqrt{\hat{V}(\hat{p}_1)} \\ &= .8 \pm 2 \times \sqrt{2.0763 \times 10^{-4}} \\ &= [.7712 ; .8288] \end{aligned}$$

Uppgift 5

Denna uppgift kan lösas genom att urskilja konstanten c ($c = 800$ kr), vilket gjordes nedan, eller att bilda en ny datakolumn, summa belopp som fuskåkare är skyldiga på en buss, säg y , och räkna på dessa nya data; y -värden i datamaterialet vore då 4800, 1600, o s v. Båda dessa sätt ger samma resultat.

Pop.: bussar i distrikt, $N = 78$

Usv.: summa belopp som fuskåkare borde betala på en buss, y

Param.:

$$\tau_y = \sum_{i=1}^N y_i = \sum_{i=1}^N c \times z_i$$

där c är det belopp som SL förlorar per en passagerare ertappad utan biljett ($c = 800$ kr) och z blir antal passagerare ertappade utan biljett på bussen.

Urv.: OSU, $n = 8$

Punktsk.:

$$\hat{\tau}_\pi = N\bar{y} = N \frac{\sum_{i=1}^n y_i}{n} = N \frac{\sum_{i=1}^n c \times z_i}{n} = cN \frac{\sum_{i=1}^n z_i}{n} = cN\bar{z}$$

Varsk.:

$$\hat{V}(\hat{\tau}_\pi) = \hat{V}(cN\bar{z}) = c^2 \hat{V}(N\bar{z}) \underset{\substack{\uparrow \\ \text{från form-bl.}}}{=} c^2 N^2 \frac{N-n}{N} \frac{s^2}{n}, \text{ där } s^2 = \frac{\sum_{i=1}^n (z_i - \bar{z})^2}{n-1}$$

Data:

Buss	# fuskåkare	(zi-zbar)^2	
1	6	6.8906	
2	2	1.8906	
3	3	0.1406	
4	2	1.8906	
5	5	2.6406	
6	4	0.3906	
7	4	0.3906	
8	1	5.6406	
zbar=	3.375	19.8750	= s u m m a
		2.8393	= s 2

Svar:

$$\hat{\tau}_\pi = cN\bar{z} = 800 \times 78 \times 3.375 = 210600$$

Varsk.:

$$\hat{V}(\hat{\tau}_\pi) = c^2 N^2 \frac{N-n}{N} \frac{s^2}{n} = 800^2 \times 78^2 \times \frac{78-8}{78} \times \frac{2.8393}{8} = 1.2402 \times 10^9$$

$$\begin{aligned} KI_{\hat{\tau}_\pi} &= \hat{\tau}_\pi \pm 2\sqrt{\hat{V}(\hat{\tau}_\pi)} \\ &= 210600 \pm 2 \times \sqrt{1.2402 \times 10^9} \\ &= [140167 ; 281033] \end{aligned}$$