

Tenta undersökningsmetodik 2020-08-17

Anonymkod: 0060-RGJ

Fråga 1:

Tenta Undersökningsmetodik 17/8
Anonymkod 0060-RGJ

① a) $\hat{v}(\hat{p}) = \hat{p}(1-\hat{p}/n-1) \left(1 - \frac{n}{N}\right)$

Konfidensintervall ges av $\hat{p} \pm 1,645 \sqrt{\hat{v}(\hat{p})}$

b) Där n ges av lösningen till ekvationen
 $2 \cdot 2,576 \cdot \sqrt{\hat{v}(\hat{p})} \leq 0,02$
 avrundat till närmaste heltal med $\hat{v}(\hat{p})$ definierad som i a

c)
$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n-1}$$

$$\hat{v}(\bar{x}) = \frac{\hat{\sigma}^2}{n} \left(1 - \frac{n}{N}\right)$$

Konfidensintervall $1,24 \pm 1,96 \sqrt{\hat{v}(\bar{x})}$

A)	Antal som inte åkt	14
	Antal som har åkt	36
	Andel som åkt	0,72
	N	200
	n	50
	\hat{p}	0,72
	$V^{\wedge}(\hat{p})$	0,003086
	Konfidensintervall 90% (+)	0,811378
	Konfidensintervall 90% (-)	0,628622

Svar **[0,628622, 0,811378]**

B)	N	200
	n	198
	p (hatt)	0,5
	$V^{\wedge}(p^{\wedge})$	1,26904E-05
	Konfidensintervall 99% (+)	0,50917662
	Konfidensintervall 99% (-)	0,49082338
	Längd	1,84%
	Svar:	198 personer

C)	n	50
	N	200
	Stickprovsmedelvärde	1,24 ($14*0 + 18*1 + 10*2 + 8*3 + 4*0$) / n
	Summan av Xi	62 ($14*0 + 18*1 + 10*2 + 8*3 + 4*0$)
	Kvadratsumman av Xi	130 ($14*0 + 18*1 + 10*4 + 8*9 + 16*0$)
	Summan av Xi i kvadrat	3844
	S2	1,08408163
	$V^{\wedge}(X^-)$	0,01626122
	Konfidensintervall 95% (+)	1,48993823
	Konfidensintervall 95% (-)	0,99006177
	Svar:	[0,99006177, 1,48993823]

Fråga 1D.

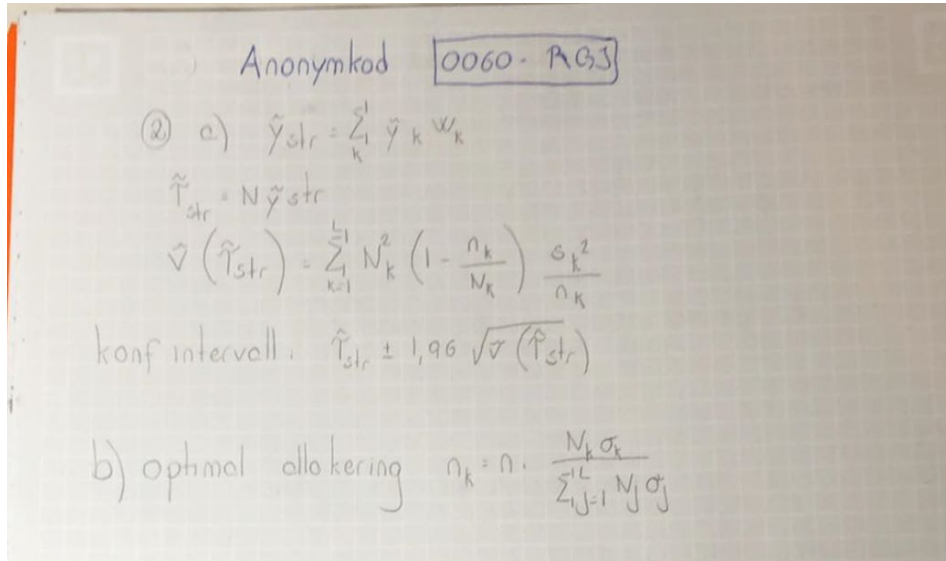
Det finns ett antal problem med att fråga folk på gatan.

- De kan ha glömt bort hur många gånger de åkt
- De kan i värsta fall ge oärliga svar
- Man riskerar alltid att få ett skevt urval när man väljer ut sina intervjuobjekt
- Definitionen av kvällstid kan vara otydlig i detta fallet
- Hur många gånger man åkt de busslinjer som finns besvarar inte direkt frågan om det finns behov för fler linjer utan endast hur utnyttjad den befintliga linjen är

Alternativa metoder för datainsamling skulle kunna vara att:

- Istället fråga hur många som åker med alternativa färdmedel (ex bil eller cykel) till den större byn på kvällstid och göra motsvarande övning
- Alternativt eliminera den mänskliga faktorn och mäta trafikbelastningen med kamera på vägen mellan byarna under kvällarna på en vecka

Fråga 2:



A)

Index	Land	nk	Nk	yk (medel)	Standardavvikelse	sk ²	Wk	S2 (mellan)
1	Danmark	15	75	21,2	12,8	163,84	0,56390977	49152
2	Norge	8	30	14,5	11,4	129,96	0,22556391	10721,7
3	Sverige	5	28	25,8	9,2	84,64	0,21052632	10901,632

N 133
 ystr 20,6571429
 Tao (skattning) 2747,4
 S2 (Tao) 70775,332

Konfidensintervall 95% (+) 3268,83122
 Konfidensintervall 95% (-) 2225,96878

Svar: [2225,96878, 3268,83122]

B) **Proportionell allokering**

Index	Land	nk	Nk
1	Danmark	16	75
2	Norge	6	30
3	Sverige	6	28
	Totalt	28	133

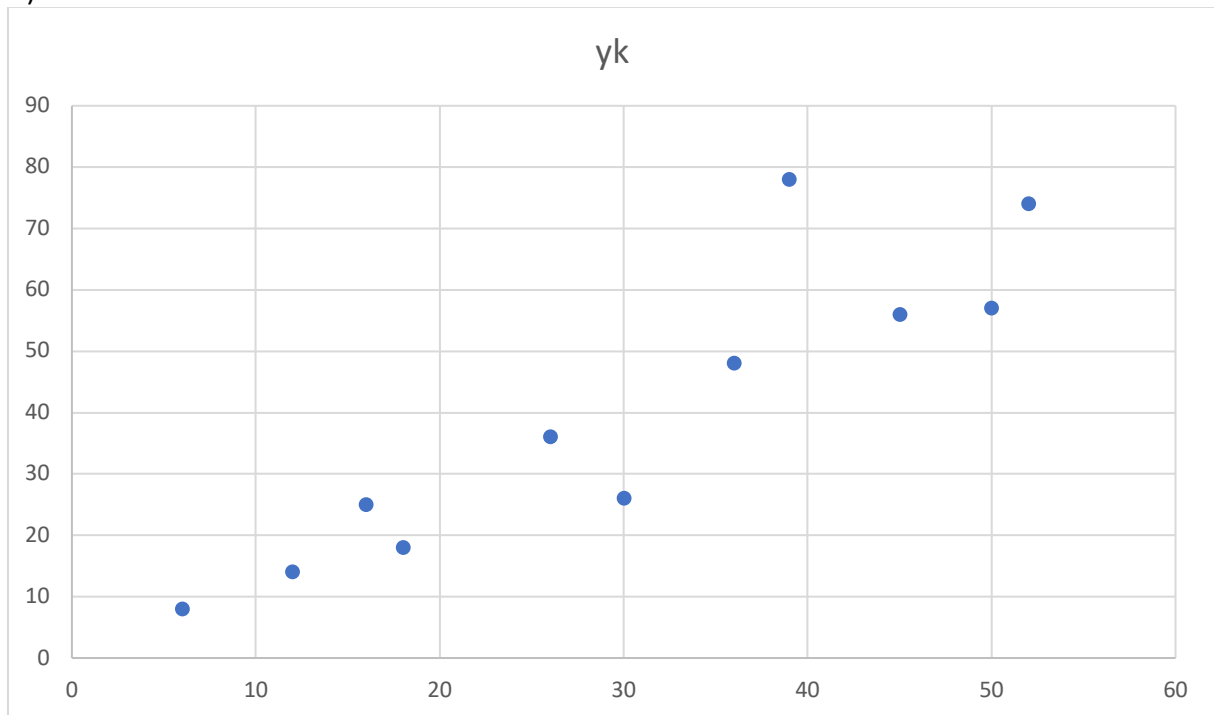
Optimal allokering

Index	Land	nk	Nk
1	Danmark	17	75
2	Norge	6	30
3	Sverige	5	28
	Totalt	28	133

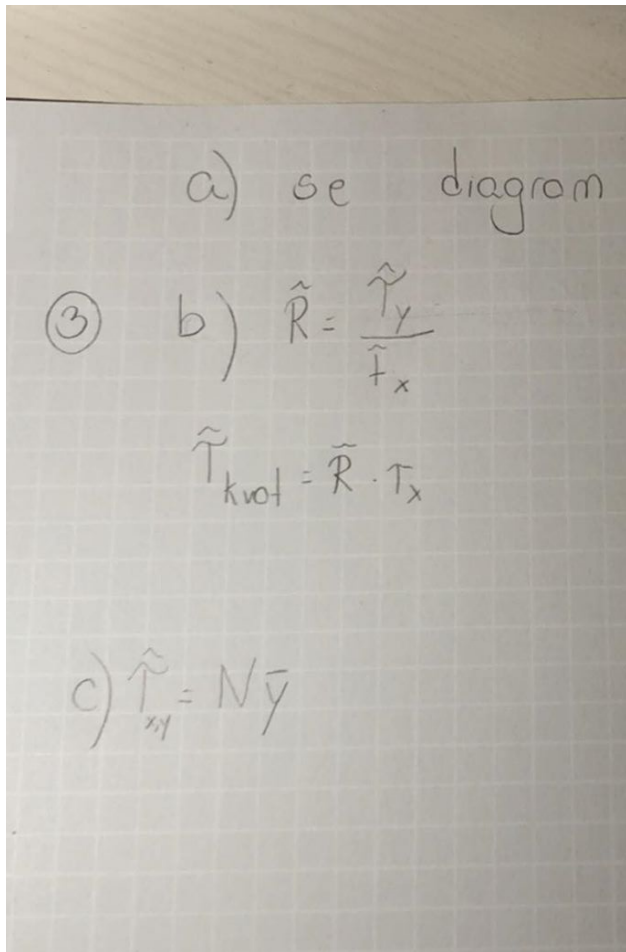
Vi ser här att det ursprungliga urvalet är överviktat i Norge men underviktat i Danmark.

Fråga 3:

A)



Jag skulle använda en kvotestimator eftersom diagrammet tydligt visar att det finns en stor och positiv korrelation mellan y och x samt att skärningspunkten för den linjära regressionslinjen skulle passera nära origo. Detta uppfyller alltså samtliga kriterier för att kunna använda en kvotestimator.



B)	Tao (x)	7680
	N	350
	n	11
	R [^]	1,33333333
	Tao (kvot)	10240
Svar:		Kvotestimatorn ger Tao (y) = 10240

C)	Tao (x)	10500
	Tao (y)	14000

Svar: Skillnaden ligger i att kvotskattningen tar hänsyn till den faktiska totalsumman för det avslutade året. Jämfört med den vanliga OSU skattningen av Tao (x) är denna betydligt lägre än vad urvalet ger sken av. Detta fångas av kvotskattningen vilket ger ett lägre värde på Tao (y) med kvotskattningen jämfört med den traditionella

Fråga 4:

A)

Inklusionssannolikhet är ett objekts sannolikhet att bli vald vid ett slumpmässigt urval. Denna sannolikhet är ofta n/N i tillämpningar och principen är att den bör vara lika stor för alla objekt vid ett obundet slumpmässigt urval samt att stickprovsstorleken n måste vara tillräckligt stor i förhållande till populationen. Om detta inte uppfylls riskerar estimat att icke vara representativa eller skeva om sannolikheten för vissa objekt att bli valda är större än för andra.

B)

Mixed mode innebär att man använder en blandning av flera olika insamlingsmetoder inom samma undersökning. Man anpassar undersökningen så att den fungerar i förhållande till respondenternas förutsättningar och beteenden.

Fördelar:

- Man kan få en lägre bortfallsfrekvens
- Ibland kan man även få högre kvalitet på svaren

Nackdelar:

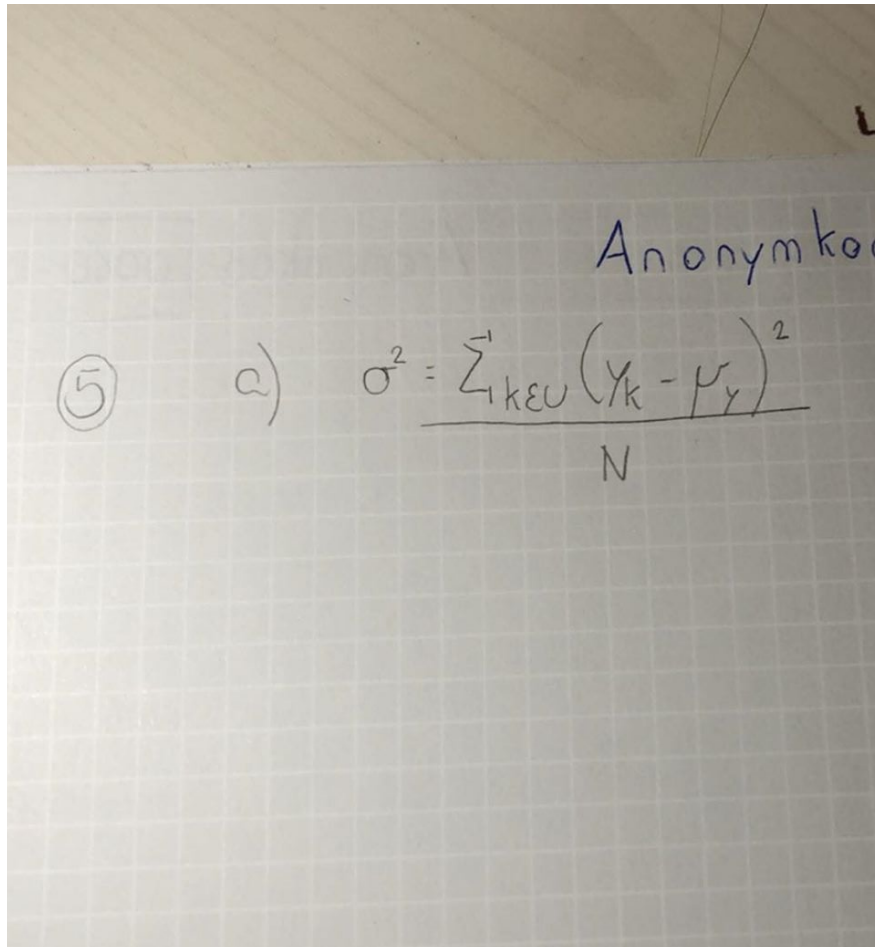
- Kan ofta vara dyrt då kompletterande datainsamling kan ta tid och pengar
- Mixed-mode undersökningar är rent praktiskt svårare att genomföra

C)

- n_s : Antalet i stickprovet som tillhör målpopulationen som lämnat svar
- n_b : Antalet i stickprovet som tillhör målpopulationen med inga eller ofullständiga svar, bortfallet
- n_0 : Antalet i stickprovet med oklar målpopulationsstatus, antingen tillhör de målpopulationen eller överteckningen

Måttet används för att belysa bortfallsandelen, och är i förlängningen ett kvalitetsmått på statistiska undersökningar.

Fråga 5:



A)	N	4
	Summa Y	18
	Medelvärde	4,5
	Varians	5,25

B)	Element 1	Element 2	Stickprovsmedelvärde		
Stickprovsutfall 1	2	3	2,5	Medelvärde	4,5
Stickprovsutfall 2	2	5	3,5	Varians	1,75
Stickprovsutfall 3	2	8	5		
Stickprovsutfall 4	3	5	4		
Stickprovsutfall 5	3	8	5,5		
Stickprovsutfall 6	5	8	6,5		

C)	Element 1	Element 2	Stickprovsmedelvärde		
Stickprovsutfall 1	2	2	2	Medelvärde	4,5
Stickprovsutfall 2	3	3	3	Varians	3,15
Stickprovsutfall 3	5	5	5		
Stickprovsutfall 4	8	8	8		
Stickprovsutfall 5	2	3	2,5		
Stickprovsutfall 6	2	5	3,5		
Stickprovsutfall 7	2	8	5		
Stickprovsutfall 8	3	5	4		
Stickprovsutfall 9	3	8	5,5		
Stickprovsutfall 10	5	8	6,5		

D) Utfallsrummet blir större med återläggning eftersom det möjliggör stickprov där samma element väljs två gånger. Ett större utfallsrum leder naturligt till större varians.

Där varianserna är beräknade på stickprovsmedelvärdena.

Utfallsrummen i D) anges som stickprovsutfall i bilderna.