

STOCKHOLMS UNIVERSITET

Statistiska institutionen

Regressionsanalys och undersökningsmetodik, vårterminen 2017

Jörgen Säve-Söderbergh

TENTAMEN I UNDERSÖKNINGSMETODIK

Datum	2017-08-14
Ansvarig Lärare:	Jörgen Säve-Söderbergh
Antal frågor:	5
Maxpoäng:	50
Hjälpmedel:	1) Språklexikon 2) Kalkylator utan lagrade formler eller lagrad text
Tentamensgenomgång	Fredag 26 augusti kl. 11.00 i sal B 705

Anvisningar

Redovisa dina lösningar i en form som gör det lätt att följa tankegången. Motivera alla väsentliga steg i lösningen. Ange alla antaganden och förutsättningar som du utnyttjar. Skriv endast på en sida av arket. Börja varje ny uppgift på nytt ark.

Lycka Till!

1. En forskare har föreslagit en metod för att göra urval. I följande enkla exempel då $N = 6$ och $n = 2$ går metoden ut på att vi kan dra följande stickprov. Antag att populationen är $\{1, 2, 3, 4, 5, 6\}$. Då är samtliga möjliga stickprov uppskrivna nedan med tillhörande sannolikheter:

Stickprov	Sannolikhet för att detta stickprov blir utvalt
$\{1, 4\}$	$1/3$
$\{2, 5\}$	$1/3$
$\{3, 6\}$	$1/3$

- a) Beräkna inklusionssannolikheter π_i och designvikterna d_i för samtliga individer i populationen. (5 p)
- b) Är det möjligt att tolka denna metod utifrån någon av de kända metoder som vi har gått igenom? (5 p)
2. En forskare önskade studera företags givmildhet och fragade ett urval om tio företag bland samtliga större företag i Stockholms län hur stora donationer de hade gjort under det senaste året för olika ändamål. Forskaren hade deltagit i en kurs i forskningsmetoder vid Stockholms universitet och hade där lärt sig att det är effektivt ur statistisk synvinkel att använda kvotskattningar. I den artikel som blev resultatet av forskningen användes företagets omsättning (i tusentals kronor) som hjälpvariabel. Mera exakt uttryckt var forskaren intresserad av den genomsnittliga donationsnivån bland Stockholms företag och urvalsprincipen som nyttjades var OSU. Antalet företag i populationen var $N = 504$. Från SCB kunde forskaren få uppgiften om den genomsnittliga omsättningen i population, vilken var $\mu_Z = 297$ (tusentals kronor).

Omsättning	Donerat belopp
440.4	4400
193.8	19840
296.3	130
316.3	3690
529.7	235
112.7	11650
107.2	660
363.9	6970
107.0	980
549.1	5420

Beräkna ett 95%-igt konfidensintervall för den genomsnittliga donationen bland alla större företag i Stockholms län med hjälp av en kvot-skattning genom att använda omsättningen som hjälpvariabel. (10 p)

3. Säkerhetsfrågor inom IT verkar bli viktigare för Sveriges framtid som företagarnation. En grupp tjänstemän vid rektorns kansli önskade undersöka hur medvetna studenterna vid Stockholms universitet var om dessa frågor. För att kunna redovisa eventuella skillnader mellan studenter på olika program stratifierades samtliga elever vid universitetet efter akademitillhörighet i tre stratum. En stickprovundersökning genomfördes för $n = 300$ individer. Frågan som ställdes var; har du vidtagit några åtgärder förutom att installera ett virusprogram för att höja säkerheten på din personliga dator? Följande resultat erhöles då:

Stratum	N_i	n_i	p_i
Ekonomi	1300	200	0.20
Lärarytildning	1080	50	0.09
Övriga institutioner	950	50	0.31

Beräkna ett 95%-igt konfidensintervall för andelen studenter vid Stockholms universitet med ett högt säkerhetstänkande. (10 p)

4. Tillgång till god hjälpinformation är många gånger väsentligt när man genomför en undersökning i praktiken.
- a) Ange två (olika) situationer när hjälpinformation kan användas i en undersökning. (4 p)
 - b) Ange för de två fall du angett ovan, om hjälpinformationen måste vara känd för alla element i populationen eller om det räcker för alla element i stickprovet. (4 p)
 - c) Var kan vi finna hjälpinformation? Ge minst ett tydligt exempel! (2 p)
5. Vad är *officiell statistik*? Ange exempel på officiell statistik, samt de krav som ställs på den. (10 p)

Tillägg till formelsamling undersökningsmetodik

Skattning av τ_X . Urval OSU

$$\hat{\tau}_{kvot} = \hat{R} \cdot \tau_Z = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n z_i} \cdot \tau_Z$$

$$\hat{V}(\hat{\tau}_{kvot}) = N^2 \left(\frac{N-n}{nN} \right) \frac{\sum_{i=1}^n (x_i - \hat{R}z_i)^2}{n-1}$$

$$\sum_{i=1}^n (x_i - \hat{R}z_i)^2 = \sum_{i=1}^n x_i^2 + \hat{R}^2 \sum_{i=1}^n z_i^2 - 2\hat{R} \sum_{i=1}^n x_i z_i$$

Skattning av μ_X . Urval OSU.

$$\hat{\mu}_{reg} = \bar{x} + b(\mu_Z - \bar{z})$$

$$b = \frac{\sum_{i=1}^n (z_i - \bar{z})(x_i - \bar{x})}{\sum_{i=1}^n (z_i - \bar{z})^2}$$

$$\hat{V}(\hat{\mu}_{reg}) = \left(\frac{N-n}{nN} \right) \left(\frac{1}{n-2} \right) \left[\sum_{i=1}^n (x_i - \bar{x})^2 - b^2 \sum_{i=1}^n (z_i - \bar{z})^2 \right]$$

Formelsamling undersökningsmetodik

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad \hat{t} = N\bar{X}$$

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n-1} = \frac{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n(n-1)}$$

Beräkning av stickprovsstorlek:

$$n \geq \frac{N\sigma^2}{D^2(N-1) + \sigma^2}$$

Stratifierat urval:

$$\bar{X}_{st} = \sum_{i=1}^L W_i \bar{X}_i \quad V(\bar{X}_{st}) = \sum_{i=1}^L W_i^2 V(\bar{X}_i) \quad \text{där } W_i = \frac{N_i}{N}$$

Optimal allokering:

$$n_i = n \frac{N_i \sigma_i}{\sum_{j=1}^L N_j \sigma_j}$$

Skattning av medelvärde samt proportion per element:

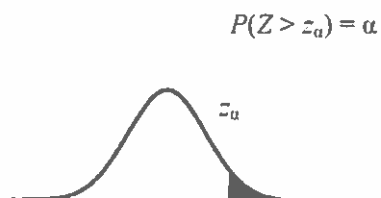
$$\bar{X}_{kvot} = \frac{\sum_{i=1}^n \tau_i}{\sum_{i=1}^n m_i} \quad \bar{X}_{VVR} = N \frac{\bar{\tau}}{M} \quad p_{kvot} = \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n m_i} \quad P_{VVR} = N \frac{\bar{a}}{M}$$

<i>Punktskattning</i>	<i>Varians</i>	<i>Variansskattning</i>	<i>Varians</i>	<i>Variansskattning</i>
<i>OSU</i>	<i>m. å.</i>	<i>m. å.</i>	<i>u. å.</i>	<i>u. å.</i>
\bar{X}	$\frac{\sigma^2}{n}$	$\frac{s^2}{n}$	$\frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right)$	$\frac{s^2}{n} \left(1 - \frac{n}{N} \right)$
\hat{t}	$N^2 \cdot \frac{\sigma^2}{n}$	$N^2 \cdot \frac{s^2}{n}$	$N^2 \cdot \frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right)$	$N^2 \cdot \frac{s^2}{n} \left(1 - \frac{n}{N} \right)$
P	$\frac{P(1-P)}{n}$	$\frac{p(1-p)}{n-1}$	$\frac{P(1-P)}{n} \left(\frac{N-n}{N-1} \right)$	$\frac{p(1-p)}{n-1} \left(1 - \frac{n}{N} \right)$
\hat{A}	$N^2 \cdot \frac{P(1-P)}{n}$	$N^2 \cdot \frac{p(1-p)}{n-1}$	$N^2 \cdot \frac{P(1-P)}{n} \left(\frac{N-n}{N-1} \right)$	$N^2 \cdot \frac{p(1-p)}{n-1} \left(1 - \frac{n}{N} \right)$

TABELL 1. Normalfördelningens kvantiler, standardiserad

$Z \in N(0, 1)$. Vilket värde har z_α om $P(Z > z_\alpha) = \alpha$ där α är en given sannolikhet.

Utnyttja även $\Phi(-z) = 1 - \Phi(z)$ för $P(Z \leq -z_\alpha)$.



α	z_α
0,25	0,6745
0,10	1,2816
0,05	1,6449
0,025	1,9600
0,010	2,3263
0,005	2,5758
0,0025	2,8070
0,0010	3,0902
0,0005	3,2905
0,00025	3,4808
0,00010	3,7190
0,00005	3,8906
0,000025	4,0556
0,000010	4,2649
0,000005	4,4172



Stockholms
universitet

Statistiska institutionen

Rättningsblad

Datum: 14/8-2017

Sal: Brunnsvikssalen

Tenta: Undersökningsmetodik

Kurs: Regressionsanalys och undersökningsmetodik

ANONYMKOD:

RUO-0008

Jag godkänner att min tenta får läggas ut anonymt på hemsidan som studentsvar.

OBS! SKRIV ÄVEN PÅ BAKSIDAN AV SKRIVBLADEN

Markera besvarade uppgifter med kryss

1	2	3	4	5	6	7	8	9	Antal inl. blad
X	X	X	X	X					5
Lär.ant. 10	10	10	5	10					

POÄNG	BETYG	Lärarens sign.
45 A	A	

1.

$N = 6$

$n = 2$

$\{1, 2, 3, 4, 5, 6\}$

$\{1, 4\}$

$\frac{1}{3}$

$\{2, 5\}$

$\frac{1}{3}$

$\{3, 6\}$

$\frac{1}{3}$

a) $\pi_1 = \frac{\text{antal gynnsamma}}{\text{antal möjliga}} = \frac{1}{3}$

$\pi_2 = \pi_3 = \pi_4 = \pi_5 = \pi_6 = \frac{1}{3}$ OK

$d = N/n \Rightarrow d = 6/2 = 3$ varje element i urvalet ersätter 3 element i pop.

b) den är ingen o.s.u trots att alla element hör samma slk allt gå in i urvalet men enligt o.s.u $\binom{N}{n} \Rightarrow \binom{6}{2} = \frac{6 \times 5}{2} = 15$ vilket betyder att det finns 15 möjliga stickprov och inte bara 3. OK

den är ingen stratifierat urval heller, om vi säger ter att vi delar up populationen till två strata och ta $n = n \frac{N_i}{N} = 2 \frac{3}{6} = 1$ element ur stickprovet så har vi 9 möjliga stickprov d.v.s $\left. \begin{matrix} \{1, 2, 3\} \\ \{4, 5, 6\} \end{matrix} \right\} n=1 \Rightarrow \left. \begin{matrix} \{1, 4\} \\ \{1, 5\} \\ \{1, 6\} \\ \{2, 4\} \\ \{2, 5\} \\ \{2, 6\} \\ \{3, 4\} \\ \{3, 5\} \\ \{3, 6\} \end{matrix} \right\}$ stickprov

- Grupp urval nej ingen grupp urval heller. Ja, man kan se det som gruppurval, 3 grupper och 1 grupp väljs

denna metoden är en systematisk urval eftersom $r = N/n = 6/2 = 3$
vi drar varje tredje antal så om vi tar till exempel slumpat
1 så drar vi med 4; 2 drar vi med 5; 3 drar vi med 6. ok

10

SU, STATISTIK

Skrivsals: _____ Anonymkod: RVD-0008 Blad nr: 2

Uppgift 2:

X	Z	X_i^2	Z_i^2
4400	440400		
19840	193800		
130	296300		
3690	316300		
235	529700		
11650	112700		
660	107200		
6970	363900		
980	107000		
5420	549100		
53975	3016400	$\Sigma = 641749625$	$\Sigma = 1169508220000$

$\Sigma X_i Z_i = 14,113,000,000$

$\hat{\mu}_{kvot} = \hat{R} \frac{\Sigma X_i}{\Sigma Z_i} = \frac{\Sigma X_i}{\Sigma Z_i} \times \mu_Z \rightarrow 0,0179 \times 297000 = 5316,3$

OK, nu fattar jag hur den räknas

$\hat{V}(\hat{\mu}_{kvot}) = \frac{(N-n)}{nN} \frac{\Sigma (X_i - \hat{R}Z_i)^2}{n-1}$

Bra!

$\Sigma (X_i - \hat{R}Z_i)^2 = \Sigma X_i^2 + R^2 \Sigma Z_i^2 - 2\hat{R} \Sigma X_i Z_i = 641749625 + (0,0179)^2 (1169508220000) - 2(0,0179)(14113000000) = 511226353,8$

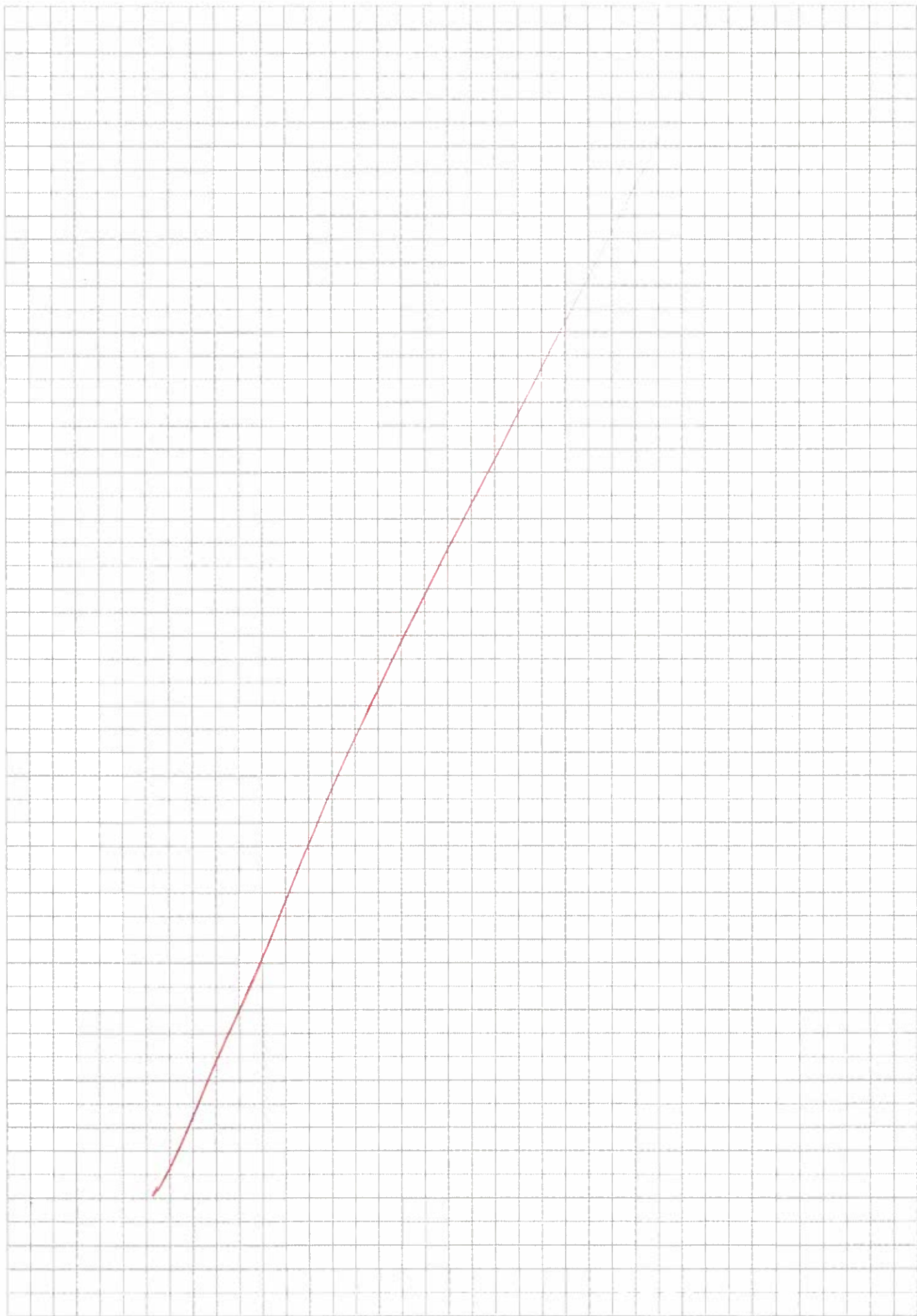
$\frac{\left(\frac{494}{5040}\right) 511226353,8}{9} = 5567588,597$

$\sqrt{V} = 2359,57 \Rightarrow \mu_{kvot} \pm 1,96(2359,57)$

$[5316,37 \quad 4624,76] \Rightarrow [691,54 ; 9941,06]$

OK

(10)



Uppgift 3

	N_i	n_i	P_i	w_i
Ekonomi	1300	200	0,2	$1300/3330$
Lärarutb.	1080	50	0,09	$1080/3330$
övrig.	950	50	0,31	$950/3330$
	$\Sigma 3330$			

$$P_{st} = \sum w_i P_i \Rightarrow \left(0,2 \frac{1300}{3330}\right) + \left(0,09 \frac{1080}{3330}\right) + \left(0,31 \frac{950}{3330}\right)$$

$$\Rightarrow P_{st} = 0,1957 \quad \text{OK}$$

$$V_{st} = \sum w_i^2 V(P_i) \Rightarrow \sum w_i^2 \frac{P(1-P)}{n-1} \left(1 - \frac{n}{N}\right) \Rightarrow$$

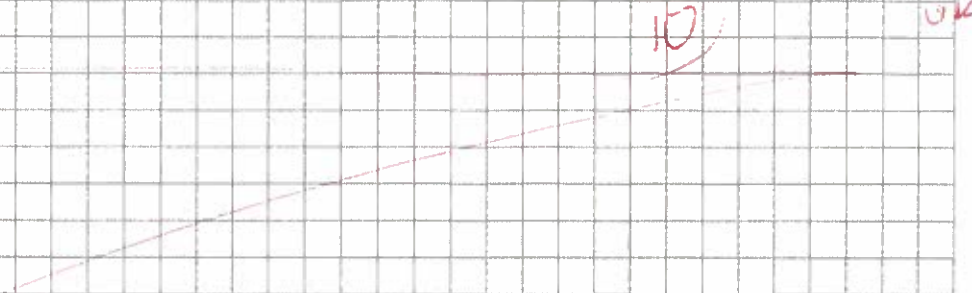
$$\left(\frac{1300}{3330}\right)^2 \frac{(0,2)(0,8)}{199} \left(1 - \frac{200}{1300}\right) + \left(\frac{1080}{3330}\right)^2 \frac{(0,09)(0,91)}{49} \left(1 - \frac{50}{1080}\right)$$

$$+ \left(\frac{950}{3330}\right)^2 \frac{(0,31)(0,69)}{49} \left(1 - \frac{50}{950}\right) = 0,0026 \quad \text{OK}$$

"mit fel!"

$$\sqrt{V} = 0,0247$$

$$P_{st} \pm 1,96 \sqrt{V} \Rightarrow [0,1957 \pm 0,0483] \Rightarrow [0,1474 ; 0,244] \quad \text{OK}$$



4) a) - hjälp variabel: vi använder hjälp variabel när vi tänker att den har påverkan på x variabel dvs att det finns ett samband mellan x variabel och hjälp variabel inklusion s.k. i x variabel är inte konstanta, och kan ändras för att vi får en bra skattning behöver det vara en positiv korrelation mellan de två variabler.
negativ går lika bra

16) Nej det räcker att den är känd för alla element i stickprovet. *Man måste ha något som gör att stickprovet korrelerar till populationen. Tex. $n_z = 297$ i uppgift 2.*
 om vi säger still exempel att vi vill undersöka lönen och vi ser att vilken utbildning man har kan påverka lönen. Så tar vi z utbildnings som hjälp variabel. *Bra ex, men du har inte skrivit om två olika situationer (se texten i 4a)*
etc saknas

uppgift 5.

officiell statistik är inte en beskrivning av folket eller en läse bok, lika lite är det en samlig detaljerade regionala beskrivning. Däremot menar man med officiellt statistik kan belysa förändring

förhållanden och ge en fullständig och trogen bild av hela landets tillstånd vid en viss tidpunkt.

Så den är berättar om våra samhället och något av allmänintresse. den är högkvalitet så den ska inte

^{BR} vara dålig eller opålitlig. Framställs regelbundet med långsiktplan. objektiv och transparent

den används till ex. i myndigheter, Arbetsmarknad o.s.v. värd ↗ OK

Krav: myndighet som ansvarar för officiell statistik skall upprätta aktuella beskrivningar av statistiken

beskrivningar skall innehålla en kvalitetsdeklaration och allmänt information om statistiken och den ska

dekumenteras i sådana omfattning att framtida användning är bevarade data material underlättas

(forts)

- officiell statistik ska offentliggöras så snart
den framställd användning denna information
som ett planerings och beslutsunderlag för regering
riksdag och statliga myndigheter (10)

