STOCKHOLMS UNIVERSITET
Statistiska institutionen
Regressionsanalys och undersökningsmetodik, höstterminen 2016
Jörgen Säve-Söderbergh

TENTAMEN I UNDERSÖKNINGSMETODIK

Datum	2017-01-09
Tid:	10.00-15.00
Ansvarig Lärare:	Jörgen Säve-Söderbergh
Antal frågor:	5
Maxpoäng:	50
Hjälpmedel:	1) Språklexikon
	2) Kalkylator utan lagrade formler eller lagrad text
Tentamensgenomgång	Onsdag 25 januari kl. 15.00 i sal B 413

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. För att bestämma en skattning av befolkningen 1 nov 2007 gjordes ett obundet slumpmässigt urval utan återläggning av 10 kommuner bland rikets 290 kommuner. Man tog reda på den definitiva befolkningen vid denna tidpunkt i dessa kommuner och även befolkningen ett år tidigare. Totalt i landet uppgick befolkningen 1 nov 2006 till 9 108 000 personer. Följande resultat erhölls, avrundade tal:

Kommun	1 nov 2006	1 nov 2007
Lidingö	42 300	42 700
Svedala	19 000	19 100
Helsingborg	123 200	124 800
Eslöv	30 400	30 700
Gullspång	5 500	5 400
Kil	11 800	11 700
Smedjebacken	10 700	10 700
Sandviken	36 700	36 800
Ånge	10 600	10 500
Vindeln	5 700	5 700

Källa: SCB:s befolkningsstatistik.

a) Skatta den totala befolkningen i riket 1 november 2007 med hjälp av en kvotskattning. (7 p)

b) Redogör för när en kvotskattning kan vara att föredra framför en vanlig medelvärdesskattning. (3 p)

2. För att undersöka har stor andel av svenska folket som har ett intresse för statistik önskar man skicka enkäter till ett antal personer från ett obundet slumpmässigt urval. Hur många personer måste man minst välja för att felmarginalen på ett 95%-igt konfidensintervall för den sökta andelen i populationen högst ska vara 5%? (10 p)

3. En riksdagsman hade hört talas om problemet med bluffakturor. Eftersom han hade erhållit anslag för att kunna söka kunskap på egen hand föreslog han för sin assistent att de skulle göra egen stickprovsundersökning. Var det som Svenskt Näringsliv påstår verkligen sant? Assistenten funderade en stund och sade att vi måste nog skilja på stora och små företag eftersom små företag får antagligen många fler

bluffakturor än stora. De stora har antagligen personal som kan åtgärda problemet. Populationen stratifierades i tre strata för att kunna redovisa dessa grupper separat. Ur företagsregistret togs OSU ur varje strata, totalt intervjuades $n = 300$ företag. Låt x_i beteckna fakturabeloppet för den värsta bluffaktura de erhållit. Följande resultat erhölls då:

Stratum	N_i	n_i	$\overline{x}_i$	s_i
Små företag	23000	200	18695	3029
Mellanstora företag	3000	50	23051	4782
Stora företag	1000	50	34592	9453

Beräkna ett 95%-igt konfidensintervall för det genomsnittliga fakturabeloppet. (10 p)

4. En population innehåller $N = 5$ element

$$2, 56, 34, 10, 27$$

Ur denna population skall man göra ett obundet slumpmässigt urval (OSU) om $n = 2$ element.

a) Hur många olika sådana urval kan man dra? (2 p)

b) Beräkna samplingfördelningen för den stokastiska variabeln estimatorn $\hat{\tau}$, d v s totalen. (8 p)

5. Din chef har givit dig i uppdrag att uppskatta kostnaderna för datakonsulter som din arbetsgivare måste inhandla. Ni har nämligen hittat ett gammalt magnetband som har använts som extern hårddisk för länge sedan. Chefen vill att du ska ta reda på hur många filer av de som magnetbandet innehåller som experter måste hjälpa er att lagra i ett nytt format. Magnetbandet innehåller 3000 filer och du gör ett urval om 400, bland dem visade sig 70 filer vara omöjliga att öppna. Beräkna ett 95%-igt konfidensintervall för det totala antalet filer som är omöjliga att öppna. (10 p)

Formelsamling undersökningsmetodik

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

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

Beräkning av stickprovsstorlek:

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

Stratifierat urval:

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

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:

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

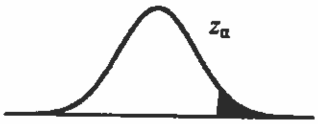
Punktskattning OSU	*Varians* m.å.	*Variansskattning* m.å.	*Varians* u.å.	*Variansskattning* u.å.
$\bar{X}$	$\frac{\sigma^2}{n}$	$\frac{s^2}{n}$	$\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$	$\frac{s^2}{n}\left(1-\frac{n}{N}\right)$
$\hat{\tau}$	$N^2 \cdot \frac{\sigma^2}{n}$	$N^2 \cdot \frac{s^2}{n}$	$N^2 \cdot \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$	$N^2 \cdot \frac{s^2}{n}\left(1-\frac{n}{N}\right)$
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 $P(Z \leq -z_\alpha)$.

$P(Z > z_\alpha) = \alpha$



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

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 9/1-17

Sal: Värtasalen

Tenta: Undersökningsmetodik

Kurs: Regressionsanalys och undersökningsmetodik

ANONYMKOD: UND-0018

☑ 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	8	10					

POÄNG	BETYG	Lärarens sign.
48	A	JSS

SU, STATISTIK

Skrivsal: Värtasalen **Anonymkod:** UND-0018 **Blad nr:** 1

1) $\boxed{n=10}$ $\boxed{N=290}$ $\boxed{\text{Total Befolkning } 2006 = 9108000.}$

	Kommun	2006	2007
1	Lidingö	42300	42700
2	Svedala	19000	19100
3	Hälsingborg	123200	124800
4	Eslöv	30400	30700
5	Gullspång	5500	5400
6	Kil	11800	11700
7	Smedjebacken	10700	10700
8	Sandviken	36700	36800
9	Ånge	10600	10500
10	Vindeln	5700	5700
	Totalt:	295900	298100
	Urvalsmedelvärde	29590	29810

a) Söker skatta totala befolkningen i riket 2007 med en kvotskattning.

- Totalbefolkning 2006 = $T_{2006} = 9108000$
- Skattad befolkning, $\hat{T}_x$, beräknas genom att multiplicera medelvärdet från urvalen med det totala antalet kommuner. Då fås en skattning av de totala befolkningsmängderna som baseras på de respektive urvalen.
- $\hat{T}_{2006} = 290 \cdot 29590 = 8581100$
- $\hat{T}_{2007} = 290 \cdot 29810 = 8644900$
- Kvotskattningen: $\boxed{T_{2007}} = \frac{T_{2006}}{\hat{T}_{2006}} \cdot \hat{T}_{2007} = \frac{9108000}{8581100} \cdot 8644900 = \boxed{9175717}$

R

b) En kvotskattning är att föredra framför en vanlig medelvärdesskattning om det finns data tillgänglig att jämföra med en skattning av samma data, och denna data korrelerar med den data som vi är intresserade av att skatta. Då kan man ta fram kvoten som visar ungefärligt hur felaktig skattningen är. Det är viktigt att inte de inbördes, relativa storlekarna bland kommunerna ändras nämnvärt under tiden mellan urvalen, i det här fallet. Ty detta skulle försämra precisionen för kvotskattningen.

Bra

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: UND-0018 Blad nr: 2

2) Obundet Slumpmässigt Urval.

Vi söker antalet personer, som andel av totalen, vi behöver skicka en enkät till för att felmarginalen skall vara högst 5% av andelen i populationen.

- Felmarginalen för ett 95%-igt konfidensintervall är: $1{,}96^{*} \cdot \sigma$, där σ är $\sqrt{\text{Variansen}}$ *från Tabell 1

- Vi söker alltså felmarginalen D:

$$\boxed{D \leq 1{,}96 \cdot \sqrt{V}}$$

värden

Antar att alla respondenter i urvalet svarar, då inget annat angetts. Urvalet blir då lika stort som antalet utskickade enkäter.

Antar att ändlighetskorrektionen är onödig givet den stora målpop.

- Variansen:

$$\boxed{\hat{V}(p)} = \frac{p \cdot (1-p)}{n-1} \Rightarrow \sigma = \sqrt{p \cdot (1-p) \cdot \frac{1}{n-1}}$$

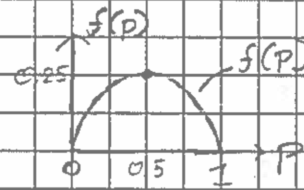
- Eftersom felmarginalen utgörs av

$$\boxed{D} = 1{,}96 \cdot \sqrt{\frac{p(1-p)}{n-1}}$$ och den får vara max 5%

så kan vi skriva ut detta kriterie som:

$$0{,}05 = 1{,}96 \cdot \sqrt{\frac{p(1-p)}{n-1}} \Rightarrow 0{,}0255^2 = \frac{p \cdot (1-p)}{n-1} \quad (\mathrm{I})$$

- Utan att gå in på detaljer så kan täljaren i ekvation (I) ej överstiga 0,25 då den når sitt maximala värde för $p = 0{,}5$.



$f(p) = p \cdot (1-p)$
$= p - p^2$

- Således uppnås det maximala värdet då $p = 0{,}5$, vilket ger följande ekvation.

$$0{,}0255^2 \cdot n - 1 = 0{,}25$$

R

$$n = 385{,}16 \rightsquigarrow \boxed{n = 386}$$ för att understiga 5% felmarginal.

- Notera avslutningsvis att 386 personer är en mycket liten andel av den totala befolkningen, och att vi därmed gjorde rätt i att skippa ändlighetskorrektionen! Ja, det är rimligt.

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: UND-0018 Blad nr: 3

3) $\boxed{n = 300}$, $\boxed{x_i = \text{Fakturabelopp för den värsta bluffakturan de erhållit.}}$

	Stratum	N_i	n_i	$\bar{X}_i$	S_i	$w_i = \frac{N_i}{N}$
1	Små	23000	200	18695	3029	0,8519
2	Mellanstora	3000	50	23051	4782	0,1111
3	Stora	1000	50	34592	9453	0,0370
	Total	27000	300			1

• Söker ett 95%-igt konfidensintervall för det genomsnittliga fakturabeloppet.

$$\boxed{\bar{X}_{st}} = \sum w_i \cdot \bar{x}_i = \frac{23000}{27000} \cdot 18695 + \frac{3000}{27000} \cdot 23051$$

$$+ \frac{1000}{27000} \cdot 34592 = 15925{,}37 + 2561{,}22 + 1281{,}19 =$$

$$= \boxed{19768} \text{ kr.}$$

R

$$\circ \boxed{V(\bar{X}_{st})} = \sum w_i^2 \cdot V(\bar{x}_i) = \sum w_i^2 \cdot \frac{S_i^2}{n} \cdot \left(1 - \frac{n}{N}\right) =$$

$$= 0{,}8519^2 \cdot 3029^2 \cdot \frac{1}{200} \cdot \left(1 - \frac{200}{23000}\right) +$$

$$+ 0{,}1111^2 \cdot 4782^2 \cdot \frac{1}{50} \cdot \left(1 - \frac{50}{3000}\right) +$$

$$+ 0{,}0370^2 \cdot 9453^2 \cdot \frac{1}{50} \cdot \left(1 - \frac{50}{1000}\right) = \boxed{40880}$$

R

• Vilket ger standardavvikelsen:

$$\sqrt{40880} = 202{,}19.$$

• Det 95%-iga konfidensintervallet ges av:

$$19768 \pm 1{,}96^{*} \cdot 202{,}19 \Rightarrow (19372 ; 20164)$$

R

* Från Tabell 1.

SU, STATISTIK

Skrivsal: Värtasalen **Anonymkod:** UND-0018 **Blad nr:** 4

4) $\boxed{\begin{matrix} N = 5 \\ n = 2 \end{matrix}}$, $\{2,\ 56,\ 34,\ 10,\ 27\}$

a) Ur denna population dras med OSU $n=2$ element.

- Detta kan göras i $\binom{5}{2}$ kombinationer.
- $\binom{5}{2} = \frac{5!}{2!\cdot(5-2)!} = \frac{5\cdot4\cdot3\cdot2\cdot1}{2\cdot1\cdot3\cdot2\cdot1} = \frac{5\cdot4}{2\cdot1} = \frac{20}{2} = \boxed{10}$ kombinationer R

b) Söker samplingfördelningen för estimatorn av totalen, $\hat{\tau}$. $\hat{\tau} = N^2\cdot\frac{s^2}{n}\cdot\left(1-\frac{n}{N}\right)$, $s^2 = \frac{\sum(x_i-\bar{x})^2}{n-1}$

- Utfallsrummet för urvalet av 2 element ur populationen.

#	u_1	u_2	$\bar{u}$	s^2	$\hat{\tau}$
1	2	56	29	1458	10935
2	2	34	18	512	3840
3	2	10	6	32	240*
4	2	27	14,5	312,5	2343,8
5	56	34	45	242	1815
6	56	10	33	1058	7935
7	56	27	41,5	420,5	3153,8
8	34	10	22	288	2160
9	34	27	30,5	24,5	183,8
10	10	27	18,5	144,5	1083,8
				✓	✓

- $\boxed{\bar{u}} = (u_1+u_2)/2$
- $\boxed{s^2} = \frac{\sum(u_i-\bar{u})^2}{2-1}$
- $\boxed{\hat{\tau}} = 5^2\cdot\frac{s^2}{2}\cdot\left(1-\frac{2}{5}\right)$

- *Exempel på uträkning i Tabell ovan:

#3: $\{u_1 = 2,\ u_2 = 10\} \Rightarrow \boxed{\bar{u}} = 6 \Rightarrow \boxed{s^2} = \frac{\sum(u_i-\bar{u})^2}{n-1} = \frac{(2-6)^2}{2-1} + \frac{(10-6)^2}{2-1} = \boxed{32} \Rightarrow$

$\Rightarrow \boxed{\hat{\tau}} = N^2\cdot\frac{s^2}{n}\cdot\left(1-\frac{n}{N}\right) = 5^2\cdot\frac{32}{2}\cdot\left(1-\frac{2}{5}\right) = \boxed{240}$

Svar: Samplingfördelningen för $\hat{\tau}$ ges i tabellform som så här:

$$\left\{\begin{matrix} 183{,}8 \\ 240{,}0 \\ 1083{,}8 \\ 1815{,}0 \\ 2160{,}0 \\ 2343{,}8 \\ 3153{,}8 \\ 3840{,}0 \\ 7935{,}0 \\ 10935{,}0 \end{matrix}\right\}$$

Tyvärr fel formel för $\hat{\tau}$.

6p

SU, STATISTIK

Skrivsal: Värtasalen **Anonymkod:** UND-0018 **Blad nr:** 5

5)

$N = 3000$, Antal Filer på bandet
$n = 400$, Urvalets storlek
$x = 70$, Antalet defekta filer i urvalet.
($P = 70/400 = 0{,}175$, Andelen defekta filer i urvalet)

- Vi söker ett 95%-igt konfidensintervall för antalet defekta filer.
- Skattar totalt antal defekta filer till:

$$\boxed{\hat{X}} = \frac{70}{400} \cdot 3000 = \boxed{525} \text{ st}$$

- Skattar variansen genom att skatta variansen för proportionen defekta filer.

$$\boxed{\hat{V}(\hat{P})} = \frac{P \cdot (1-P)}{n-1} \cdot \left(1 - \frac{n}{N}\right) = \frac{0{,}175 \cdot (1-0{,}175)}{399} \cdot \left(1 - \frac{400}{3000}\right)$$

$$= \frac{0{,}175 \cdot 0{,}825 \cdot 0{,}867}{399} = \boxed{0{,}0003141} \Rightarrow \boxed{\sigma} = \boxed{0{,}01772}$$

- Med $\sigma = 0{,}01772$ skapas konfidensintervallet för proportionen P till:

* från Tabell 1

$$0{,}175 \pm 1{,}96^{*} \cdot 0{,}01772 \Rightarrow (0{,}1403 \, , \, 0{,}2097).$$

- Vilket i antal defekta filer blir:

$$(421, 629).$$

Svar: Det totala antalet filer som är defekta på bandet skattas ligga, med 95% sannolikhet, mellan 421 och 629.

R

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 9/1-17

Sal: Värtasalen

Tenta: Undersökningsmetodik

Kurs: Regressionsanalys och undersökningsmetodik

ANONYMKOD: UND-0010

☒ 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	7	10					

POÄNG	BETYG	Lärarens sign.
47	A	JSS

SU, STATISTIK

Skrivsal: Värtasalen **Anonymkod:** UND-0010 **Blad nr:** 1/5

Uppgift 1

Kvotskattnings formula:

$J_x = \frac{\bar{X}}{\bar{Z}} \cdot \tau_z$

[Skattningen]

$\hat{J}_x$ = totala befolkningen 1 Nov 2007

J_z = totala befolkningen 1 Nov 2006

* Obundet slumpmässigt urval utan återläggning
* $n = 10$ och $N = 290$
* $J_z = 9108\,000$
* Z_j = Befolkningen för varje kommun 2006 $j = 1, 2, \ldots, 10$
* X_i = Befolkningen för varje kommun 2007 $i = 1, 2, \ldots, 10$

kommun	1 Nov Z_j 2006	1 Nov X_i 2007
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
$\sum_{i=1}^{n=10}$	295900	298100

$Z = \sum_{i=1}^{10} Z_j = 42\,300 + 19\,000 + 123\,200 + 30\,400 + 5500$

$+ 11\,500 + 10\,700 + 36\,700 + 10\,600 + 5\,700$

$= 295900$

$X = \sum_{i=1}^{10} X_i = 42\,700 + 19\,100 + 124\,800 + 30\,700 + 5\,400$

$+ 11\,700 + 10\,700 + 36\,800 + 10\,500 + 5\,200$

$= 298100$

$\frac{Z}{10} = \bar{Z} = \frac{\sum_{j=1}^{n=10} Z_j}{n} = \frac{295900}{10} = 29590$

$\frac{X}{10} = \bar{X} = \frac{\sum_{i=1}^{n=10} X_i}{n} = \frac{298100}{10} = 29810$

A/ Antar att vi approximerar τ_x med $\hat{\tau}_x$

Alltså blir det

$\frac{29810}{29590} \cdot 9108\,000 = 9175717{,}472$ personer 2007 Nov 1

R B På baksidan av detta ARK

B) En kvotskattning är att föredra om man har tillgång till en korrelerande variabel som påverkar vår undersöknings variabel. Andra exempel är ålder och lön, föregående värde av en aktie till dagens värde och vårt nuvarande exempel när vi misstänker att 2006 population direkt påverkar 2002 population. ✓

SU, STATISTIK

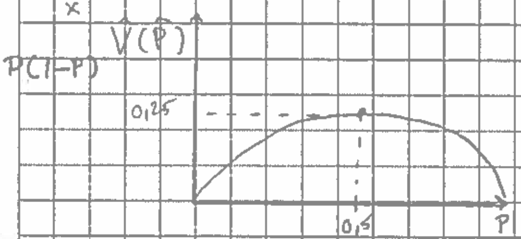
Skrivsal: Vårtasalen **Anonymkod:** UND-0010 **Blad nr:** 2/5

Uppgift 2

* punktskattning ± Felmarginalen "Bredd"
* OBundet slumpmässigt urval (OSU), utan återläggning (UÅ)
* Vi antar att vårt stickprov $n \geq 30$, så stort att vi kan använda Centrala gränsvärdessatsen (CGS)

$\text{Bin} \sim (\hat{P}; \hat{V}(\hat{P})) \longrightarrow N \sim (\hat{P}; \hat{V}(\hat{P}))$

*



* Vid Binära fördelningar tar Variansen sitt maximala värde vid $\hat{P} = 0,5$ och $\hat{V}(\hat{p}) = 0,25$

* 1 har statistiskt intresse $P_1 = 0,5$ $\hat{V}(\hat{P}) = 0,25$
  0 har Ej statistiskt intresse $P_0 = 0,5$ $\hat{V}(\hat{P}) = 0,25$

* Vi skattar därför våra andelar och varianser med $P = \hat{P}$ och $V(P) = S^2$
* Vi skattar med maximal varians S^2
* 95%-igt konf intervall med bredden $0,05 \geq 1,96 \cdot \sqrt{\frac{P(1-P)}{n-1}} \cdot \left(1 - \frac{n}{N}\right)$
* Eftersom vi ej känner till N antar vi att ändlighetskorrektionen är approximativt lika med ≈ 1
* $0,05 = Z_{1-0,05/2} \cdot \sqrt{\frac{P(1-P)}{n-1}} \cdot 1$ $\quad 0,05 = 1,96 \cdot \sqrt{\frac{0,25}{n-1}}$

$$\left(\frac{0,05}{1,96}\right)^2 = \frac{0,25}{n-1} \rightarrow \quad n = \frac{0,25}{\left(\frac{0,05}{1,96}\right)^2} + 1 \quad n = 385,16$$

Svar: För att felmarginalen inte får överskrida 0,05 måste n minst "Avrundat uppåt" vara mer eller lika med 386 urvalsenheter.

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: UND-0010 Blad nr: 3/5

Uppgift 3

* Obundet slumpmässigt urval
* Utan återläggning (UÅ)
* $N = \sum_{i=1}^{L=3} N_i = 23\,000 + 3000 + 1000 = 27\,000$
* $n = \sum_{i=1}^{L=3} n_i = 200 + 50 + 50 = 300$

NR	Stratum	N_i	$W_i = N_i/N$	n_i	$\bar{x}_i$	s_i
1	små	23 000	0,851851851 9	200	18695	3029
2	mellan	3000	0,1111111111	50	23051	4782
3	stora	1000	0,037037037	50	34592	9453
$\sum$:		27 000	1	300		

* Vi skattar det stratifierade medelvärdet

$$\bar{x}_{st} = \sum_{i=1}^{L=3} W_i \cdot \bar{x}_i = \frac{N_i}{N} \cdot \bar{x}_i = \frac{23\,000}{27\,000} \cdot 18695 + \frac{3000}{27000} \cdot 23051 + \frac{1000}{27000} \cdot 34592$$

$$\bar{x}_{st} = 19767,77778 \text{ kr}$$

$$\hat{V}(\bar{x}_{st}) = \frac{s_i^2}{n_i} \cdot \left(1 - \frac{n_i}{N_i}\right) \cdot W_i^2$$

$$\hat{V}(\bar{x}_{st}) = \underbrace{\frac{3029^2}{200} \cdot \left(1 - \frac{200}{23000}\right) \cdot \left(\frac{23000}{27000}\right)^2}_{A} + \underbrace{\frac{4782^2}{50} \cdot \left(1 - \frac{50}{3000}\right) \cdot \left(\frac{3000}{27000}\right)^2}_{B}$$

$$+ \underbrace{\frac{9453^2}{50} \cdot \left(1 - \frac{50}{1000}\right) \cdot \left(\frac{1000}{27000}\right)^2}_{C} = A + B + C = 32999,22236 +$$

$$+ 5552,197185 + 2328,978012$$

$$= 40880,39756 \text{ kr}$$

95 % -igt konf intervall

$$\bar{x}_{st} \pm z_{1-0,05/2} \cdot \sqrt{\hat{V}(\bar{x}_{st})}$$

$$19767,77778 \pm 1,96 \cdot \sqrt{40880,39756}$$

upper: 20164,06825

kr

lower: 19371,48731

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: UND-0010 Blad nr: 4/5

Uppgift 4

A/ (I) MED återläggning har vi $5^2 = 25$ urval

(II) UTAN återläggning har vi $\binom{5}{2} = \frac{5!}{2!(5-2)!} = \frac{5\cdot4\cdot3\cdot2\cdot1}{2\cdot1(3\cdot2\cdot1)} = \frac{5\cdot4}{2\cdot1} = \frac{20}{2} = 10$

10 urval

B/ Antar att vi har utan återläggning

A = 2, B = 56, C = 34, D = 10, E = 27

Vi sätter upp en korstabell eftersom vi bara har n = 2 dimensioner.

Vi vet även att utan återläggning finns det 10 st urval.

	A	B	C	D	E
A	~~A,A~~	A,B	A,C	A,D	A,E
B	B,A	~~B,B~~	B,C	B,D	B,E
C	C,A	C,B	~~C,C~~	C,D	C,E
D	D,A	D,B	D,C	~~D,D~~	D,E
E	E,A	E,B	E,C	E,D	~~E,E~~

X = kryss betyder omöjligt eftersom utan återläggning

O = möjliga urval "vi ser att vi får 10 möjliga urval

Samplingfördelningen

n	kombin	Värden	$\bar{x}_i$	$P(\bar{x}_i)$
1	B, A	56, 2	29 ,5	0,1
2	C, A	34, 2	18	0,1
3	C, B	34, 56	45	0,1
4	D, A	10, 2	6	0,1
5	D, B	10, 56	33	0,1
6	D, C	10, 34	22	0,1
7	E, A	27, 2	14,5	0,1
8	E, B	27, 56	41,5	0,1
9	E, C	27, 34	30,5	0,1
10	E, D	27, 10	18,5	0,1
				1

$\hat{\tau}$

-3p

$\hat{\tau} = N \cdot \bar{x}_{st} = N \cdot \sum_{i=1}^{i=10} P(\bar{x}_i) \cdot \bar{x}_i = 5 \cdot [0,1\cdot 29 + 0,1\cdot 18 + 0,1\cdot 45 + 0,1\cdot 6 + 0,1\cdot 33 + 0,1\cdot 22 + 0,1\cdot 14,5 + 41,5\cdot 0,1 + 0,1\cdot 30,5 + 18,5\cdot 0,1]$

$= 5\cdot 25,8 = 129$

$\hat{\tau} = 129$

SU, STATISTIK

Skrivsal: Värtasalen **Anonymkod:** VND-0010 **Blad nr:** 5/5

Uppgift 5

* Antar att vi gör ett Obundet slumpmässigt Urval, utan återläggning (u, ä)

* $n = 400$ $N = 3000$

* $\frac{70}{400}$ omöjliga att öppna $P(\text{omöj}) = 0{,}175$

* $\hat{P} = 0{,}175$

* Vi skattar en Varians $\hat{s}^2 = \frac{\hat{P}(1-\hat{P})}{n-1} \cdot \left(1 - \frac{n}{N}\right)$

$\hat{s}^2 = 3{,}1359649\,12 \cdot 10^{-4}$

95 % - Igt konfintervall av totala antalet filer omöjliga att öppna

$N \cdot \hat{P} \pm 1{,}96 \cdot \sqrt{N^2 \cdot s^2}$

$3000 \cdot 0{,}175 \pm 1{,}96 \cdot \sqrt{3000^2 \cdot s^2}$

$525 \pm 1{,}96 \cdot 53{,}12596748$

$525 \pm 104{,}1268963$

upper: $629{,}1268963$

lower: $420{,}8731037$