



Skriftlig tentamen i **Undersökningsmetodik** (4,5 hp), ingående som moment 1 i kursen **Regressionsanalys och undersökningsmetodik, 15 hp.**

Skrivtid: 5 timmar

Hjälpmedel: Miniräknare utan lagrade formler eller lagrad text. Vidhäftade formel- och tabellblad (obs! vidhäftas endast de tabellsidor som behövs för den här tentamen).

Återlämning av tentamen: hämtas på studentexpeditionen, plan 7 i B-huset fr.o.m. tisdagen den 29 januari. Kolla på vår hemsida studentexpeditionens mottagningstider under terminstid.

Tentamen består av fem uppgifter som kan ge totalt 100 poäng. För betyget A gäller 90-100 p., för betyget B gäller 80-89 p., för betyget C gäller 70-79 p., för betyget D gäller 60-69 p., för betyget E gäller 50-59 p., för betyget Fx gäller 40-49 p. och för betyget F gäller 0-39 p. För detaljerade betygs-kriterier se kursbeskrivningen på kurshemsidan.

För full poäng på en uppgift krävs fullständiga och väl motiverade lösningar.

Uppgift 1: (20 poäng)

a). I Grönköping vill man uppskatta andelen personer (mellan 16 och 64 år) som är arbetslösa. Det finns 10 000 personer (mellan 16 och 64 år) i Grönköping. Man undersöker 2 000 individer (genom ett obundet slumpmässigt urval (OSU) utan återläggning) och undersökningen visar att 200 av dem är arbetslösa.

Ange ett 95%-igt konfidensintervall för den sökta andelen och formulera slutsatsen i ord. (10 poäng)

b). I Dalaköping och vid en lokal arbetskraftsundersökning vill man uppskatta andelen personer som är arbetslösa. Man vet att den utgör högst 30% av befolkningen mellan 16 och 64 år i detta område. Från denna population som består av 9 000 individer tänker man göra ett obundet slumpmässigt urval utan återläggning. Urvalsstorleken bestäms med ledning av ett givet precisionsvillkor.

Med hjälp av urvalet vill man uppskatta den sökta andelen med ett konfidensintervall med 95% konfidensgrad som har en felmarginal på högst 3 procentenheter. Hur stort urval måste man minst göra för att detta villkor skall vara uppfyllt? (10 poäng)

Uppgift 2: (20 poäng)

I Dalaköping och vid en urvalsundersökning om storleken av hushållens aktieinnehav i ett bostadsområde med 250 hushåll, uppgav de 32 tillfrågade hushållen (genom OSU, utan återläggning) värdet av sina innehav enligt följande (i enheten 1 000 kr):

3	17	37	8	0	19	0	108	0	42	24	9	64	0	13	22
19	0	6	18	28	59	81	41	0	11	0	0	29	34	0	0

Beräkna ett approx. 95%-igt konfidensintervall för hushållens genomsnittliga aktieinnehav i bostadsområdet vid den aktuella tidpunkten. (totalt 20 poäng: skattningen 5 poäng, skattade variansen 10 poäng, konfidensintervallet 5 poäng)

Beräkna ett 95%-igt konfidensintervall för den genomsnittliga personalkostnaden för alla 90 livsmedelsbutiker i området via en kvotskattning (med omsättning som hjälpinformation).
(totalt 20 poäng: skattningen 5 poäng, skattade variansen 10 poäng, konfidensintervallet 5 poäng)

Uppgift 3: (20 poäng)

I Grönköping finns 500 företag med högst 30 anställda. Man vill undersöka den genomsnittliga arbetsgivaravgiften i samtliga företag. Med ledning av register över antalet anställda indelas företagen i tre olika storleksklasser (strata). Ur varje stratum drar man ett obundet slumpmässigt urval (utan återläggning) och arbetsgivaravgiften för varje företag noteras. Resultatet av undersökningen sammanfattas i följande tabell (enheten är 100 000 kronor).

Stratum	Antal företag i stratum	Antal företag i stickprovet	Stickprovsmedelvärde	Stickprovstandardavvikelse
1	300	50	30	4
2	150	30	20	3
3	50	20	10	2

Beräkna ett 95%-igt konfidensintervall för den genomsnittliga arbetsgivaravgiften i samtliga företag. (totalt 20 poäng: skattningen 5 poäng, skattade variansen 10 poäng, konfidensintervallet 5 poäng)

Uppgift 4: (20 poäng)

- a). I KD boken (kap. 12.3, sidan 353) tas upp de fel som inte specifikt orsakas av att man har gjort en urvalsundersökning. Namnge de felen som behandlas i kap. 12.3. (10 poäng)
- b). I KD boken (kap. 11.2, sidan 265) definieras vad vi menar med "ett slumpmässigt urval". Ange vad vi menar med "ett slumpmässigt urval". (10 poäng)

Uppgift 5: (20 poäng)

Man önskade ett visst år studera personalkostnaderna för livsmedelsbutikerna i ett område. Bland dessa 90 butiker utvaldes 5 st. med obundet slumpmässigt urval utan återläggning, och för dessa erhöles inte bara uppgift om personalkostnaden utan också om omsättningen, allt i miljoner kronor:

Personalkostnad:

3,6 11,4 10,0 8,6 3,2

Omsättning:

33,0 125,8 110,8 85,0 22,6

Totalt omsatte de 90 butikerna 4 500 miljoner kronor.

Beräkna ett 95%-igt konfidensintervall för den genomsnittliga personalkostnaden för alla 90 livsmedelsbutiker i området via en kvotskattning (med omsättning som hjälpinformation). (totalt 20 poäng: skattningen 5 poäng, skattade variansen 10 poäng, konfidensintervallet 5 poäng)

göra för att detta villkor skall vara uppfyllt? (10 poäng)

Uppgift 2: (20 poäng)

I Dalaköping och vid en urvalsundersökning om storleken av hushållens aktieinnehav i ett bostadsområde med 250 hushåll, uppgav de 32 tillfrågade hushållen (genom OSU, utan återläggning) värdet av sina innehav enligt följande (i enheten 1 000 kr):

3	17	37	8	0	19	0	108	0	42	24	9	64	0	13	22
19	0	6	18	28	59	81	41	0	11	0	0	29	34	0	0

Beräkna ett approx. 95%-igt konfidensintervall för hushållens genomsnittliga aktieinnehav i bostadsområdet vid den aktuella tidpunkten. (totalt 20 poäng: skattningen 5 poäng, skattade variansen 10 poäng, konfidensintervallet 5 poäng)

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:

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

Punktskattning	Varians	Variansskattning	Varians	Variansskattning
OSU	m. å.	m. å.	u. å.	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{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}	$N^2 \cdot \frac{P(1-P)}{n}$	$N^2 \cdot \frac{p(1-p)}{n-1}$	$N^2 \cdot \frac{P(1-P)}{n} \left(\frac{N-n}{N-1} \right)$	$N^2 \cdot \frac{p(1-p)}{n-1} \left(1 - \frac{n}{N} \right)$



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.

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

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

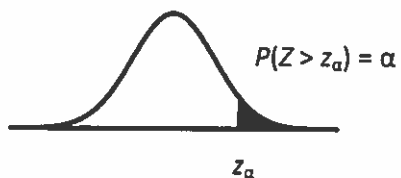
$$\hat{V}(\hat{\mu}_{reg}) = \left(\frac{N-n}{nN} \right) \left(\frac{1}{n-2} \right) \left[\sum_{i=1}^n (x_i - \bar{x})^2 - b^2 \sum_{i=1}^n (z_i - \bar{z})^2 \right]$$

\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 2. 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



1



Stockholms universitet

Statistiska institutionen

Rättningsblad

Datum: 16/1-2019

Sal: Ugglevikssalen

Tenta: Undersökningsmetodik

Kurs: Regressionsanalys och undersökningsmetodik

ANONYMKOD:

0004-SSC

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					3 2
Lärant. 10p	20p	20p	20p	20p					
10p									
POÄNG					BETYG			Lärarens sign.	
90p					A			RC	



Fråga 7.

Uppskatta andelen arbetslösa

a) $N = 10\ 000$

$n = 2000$

OSU $\approx \hat{p}$

200 är arbetslösa

$P_{kvot} = \frac{\sum e_i}{\sum m_i} = \frac{200}{2000} = 0,1$ 10% arbetslösa

95% ki

N_i	n_i	P
10000	2000	0,1

SAKNAS $(1 - \frac{n}{N})$

95% ki =

$$0,1 \pm 1,96 \sqrt{\frac{p(1-p)}{n}} = 0,1 \pm 1,96 \sqrt{\frac{0,1 \times 0,9}{2000}}$$

$\underbrace{\hspace{10em}}_{0,01314}$

Svar: $0,1 \pm 0,013148$

Konfidensintervall = $[0,0869; 0,1131]$

med 95% konfidens är andelen personer arbetslösa mellan 8,69% 11,31% av de 10 000 personerna som är arbetslösa i Grönköping.

b)

$N = 9000$

95% ki

Felmarginal 3%

Beräkning stickprov: $n \geq \frac{N \sigma^2}{D^2}$

$\sigma^2 = 0,25$

$D^2(N+1) + \sigma^2$ (SE FACIT)

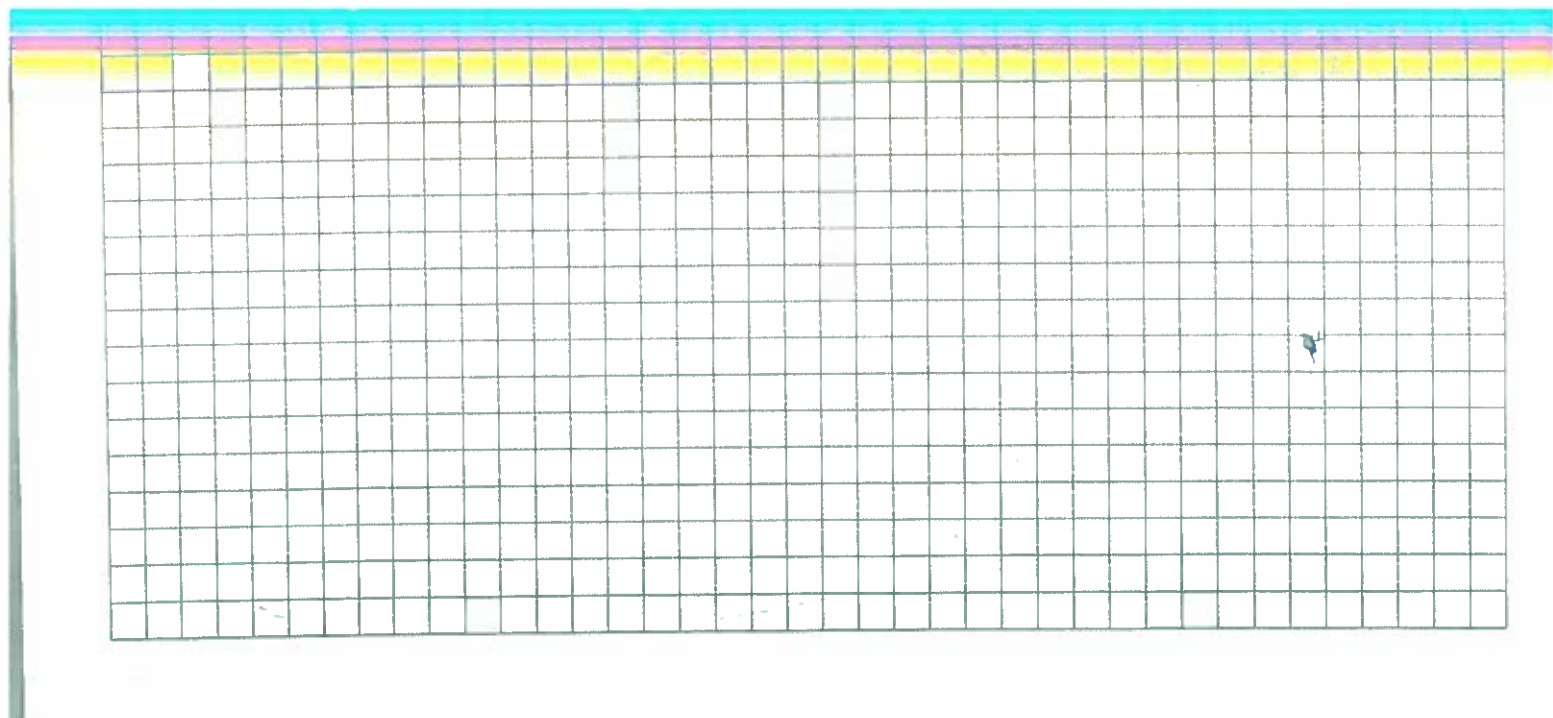
$\frac{9000 \cdot 0,25}{D^2(9000-1) + 0,25} = \frac{2250}{2,3582} = 954,175$

$\frac{2250}{2,3582} = 954,175$

$D = \left(\frac{\text{Felmarginal}}{Z_{\alpha/2}} \right)^2 = \left(\frac{0,03}{1,96} \right)^2$

Svar: 955 personer

måst minst göras för att villkor ska vara uppfyllt.



2) 20p

Fråga 2

$N = 250$

$n = 32$ OSU. vå

Värde i (1000kr)

$\sum x_i = 692$

aktie innehav = X

$\bar{x} = \frac{\sum x_i}{n} = \frac{692}{32} = 21,625$

$$V(x) = \frac{s^2}{n} \left(1 - \frac{n}{N}\right) \quad s^2 = \frac{\sum x^2 - \frac{(\sum x_i)^2}{n}}{n-1}$$

$\sum x^2 = 3^2 + 17^2 + 37^2 + \dots + 29^2 + 34^2 = 36272$

$$s^2 = \frac{36272 - \frac{(692)^2}{32}}{32-1} = 687,3387$$

$$V(x) = \frac{687,3387}{32} \times \left(1 - \frac{32}{250}\right) = 18,7297958$$

95% ki intervall =

$$21,625 \pm 1,96 \sqrt{18,7299}$$

$$\approx 8,4825$$

*Svar: $21,625 \pm 8,4825$ (tusentals kr)

Intervall $[3,1425; 30,1075]$

(Med 95% konfidens höjer det sanna μ inom detta intervall med (tusentals kronor) 1000 kr)

$$0.675560$$

Svar: 25 ± 0.675560 (100 000 kroner)

$$[24.3244; 25.67556]$$

Svar. med 95% konfidens ligger et sæt data
inom dette interval i (et set
100 000 kr)

Oppgitt 3

$N = 500$

0500 kr

200

(100 000 kr)

Stratum:

	N_i	n_i	\bar{x}_i	s_i^2
1	300	50	30	4
2	150	30	20	3
3	50	20	10	2

$$\sum N_i = 500 \quad \sum n_i = 100$$

$$\bar{X}_{st} = \sum_{i=1}^3 w_i \cdot \bar{X}_i = \frac{300}{500} \cdot 30 + \frac{150}{500} \cdot 20 + \frac{50}{500} \cdot 10 = 18 + 6 + 1 = 25$$

$$V(\bar{X}_{st}) = \sum_{i=1}^3 w_i^2 \cdot \frac{s_i^2}{n_i} \cdot \left(1 - \frac{n_i}{N_i}\right) = \left(\frac{300}{500}\right)^2 \cdot \frac{4}{50} \cdot \left(1 - \frac{50}{300}\right) + \left(\frac{150}{500}\right)^2 \cdot \frac{3}{30} \cdot \left(1 - \frac{30}{150}\right) + \left(\frac{50}{500}\right)^2 \cdot \frac{2}{20} \cdot \left(1 - \frac{20}{50}\right) = 0,1188$$

95% KI =

$$25 \pm 1,96 \sqrt{0,1188}$$

$$0,675560$$

Svar: $25 \pm 0,675560$ (100 000 kroner)

$$[24,3244; 25,67556]$$

Svar: med 95% konfidens ligger det sannsynlige
innom dette intervallet i (inntil
100 000 kr)

95% KI intervall =

$$21,625 \pm 1,96 \sqrt{18,7299}$$
$$\approx 8,4825$$

Svar: $21,625 \pm 8,4825$ (tusentals kr)

Intervall $[13,1425; 30,1075]$

(Med 95% konfidens ligger det sanna μ inom detta

Intervall med (tusentals kronor)
1000 kr

4) 20 p

Fråga 4

- a)
1. Täckningsfel - Detta kan leda till under eller övertäckning.
 2. Bortfallsfel - Med detta kan du ha individuellt bortfall då du räknat med att ha fler personer som genomför undersökningen än vad som det blev.
Partiellt bortfall - Exempel om svarsalternativet "Vet ej finns"
 3. Mätfel - Här ingår om man har frågat det som undersökningen undersöker och kan deprimera detta i om undersökningen har hög eller låg validitet och reliabilitet.
 4. Bearbetningsfel - Fel när man tex får över informationen på papper till datorn.

- b)
- Slumpmässigt urval - När man inte har någon påverkan på urvalet. Alla urval i rampopulationen måste ha en sannolikhet att bli valt som är större än 0.
- Slumpmässiga urval kan delas in i olika grupper
1. OSU -
 2. Stratifierat urval - Delas in i stratum (so homogena som möjligt)
 3. Systematiskt urval - Genom att välja var r:e i ett register $r = \frac{N}{k}$
 4. Klustururval - Delas in i grupper (heterogena). motsats är icke slumpmässiga urval

$$(E x_i - R z_i)^2 = 327,12 + (0,09756)^2 \cdot 36927,09 - 2 \cdot 0,09756 \cdot$$

$$R = \frac{T_{\text{kvot}}}{T_z} = 0,09756$$

$$3464,25 = 2,64540$$

$$K_{0,95\%} = 4,8780 \pm 1,96 \sqrt{0,1249}$$

$$0,692687$$

Svar: $4,8780 \pm 0,692687$ (millontals kr)

intervall = $[4,1853, 5,570687]$

Med 95% konfidens ligger den genomsnittliga personalkost i detta intervall (millontals kr)

Fråga 5

N=90

n=5

personalkostnad = X
försäkring = Z
OSU u&

x_i	z_i	$x_i \cdot z_i$	x_i^2	z_i^2
36	33	1188	1296	1089
11,4	125,8	1434,12	129,96	15825,64
10	110,8	1108	100	12276,64
8,6	85	731	73,96	7225
3,2	22,6	72,32	10,24	510,76

$\sum x_i = 36,8$ $\sum z_i = 377,0$ $\sum x_i \cdot z_i = 3464,25$

$\bar{x} = \frac{36,8}{5} = 7,36$

$\sum x_i^2 = 327,12$

$\sum z_i^2 = 36927,04$

$\bar{z} = 75,44$

$T_z = 4500$

$N_z = 50$

$N_{kvot} = \frac{\sum x_i}{\sum z_i} \cdot T_z = \frac{\bar{x}}{\bar{z}} \cdot N_z = \frac{7,36}{75,44} \cdot 50 = 4,8780$

$N(N_{kvot}) = \frac{(N-n)}{nN} \sum (x_i - R z_i)^2$

$\frac{(90-5)}{90 \cdot 5} \cdot \frac{2,64540}{5-1} = 0,1249$

$\sum (x_i - R z_i)^2 = 327,12 + (0,09756)^2 \cdot 36927,04 - 2 \cdot 0,09756 \cdot 3464,25 = 2,64540$

$R = \frac{T_{kvot}}{T_z} = 0,09756$

$K_i 95\% = 4,8780 \pm 1,96 \sqrt{0,1249}$

$0,692687$

Svar: $4,8780 \pm 0,692687$ (millontals kr)

intervall = $[4,1853; 5,570687]$

Med 95% konfident nivåer
dengensnittliga
personalkost i detta intervall

populationen måste ha en sannolikhet att bli valt som är större än 0.

Slumpmässiga urval kan delas in i olika grupper

1. OSU-

2. stratifierat urval = Delas in i stratum (ca homogena som möjligt)

3. Systematiskt urval = Genom att välja var r:e i ett ~~reguljärt~~ $r = \frac{N}{n}$

4. klustrurval = Delas in i grupper (heterogena).

Motsats är icke slumpmässiga urval