



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. onsdagen den 27 februari. 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 betygskriterier se kursbeskrivningen på kurshemsidan.

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

Uppgift 1: (40 poäng)

I Grönköping (som är en stor stad) ville man undersöka behovet av en ny simhall i ett bostadsområde med 500 personer. Därför tog man ett obundet slumpmässigt urval (utan återläggning) om 100 personer (i bostadsområdet) och frågade bland annat hur många besök i stadens simhall man gjort en viss månad. Resultatet blev följande:

Antal besök	0	1	2	3	4	5
Antal personer	35	20	25	12	5	3

a). Beräkna ett 95%-igt konfidensintervall för det genomsnittliga antalet simhallsbesök per person i området under den aktuella månaden. (20 poäng)

(totalt 20 poäng: skattningen 5 poäng, skattade variansen 10 poäng, konfidensintervallet 5 poäng)

b). Beräkna ett 95%-igt konfidensintervall för andelen personer i området som inte besökt simhallen under den aktuella månaden. (20 poäng)

(totalt 20 poäng: skattningen 5 poäng, skattade variansen 10 poäng, konfidensintervallet 5 poäng)

Uppgift 2: (10 poäng)

I ett stort bostadsområde i Dalaköping vill man uppskatta andelen (röstberättiga) personer i området som kommer att rösta på det politiska partiet XXX. Man vet att den andelen utgör högst 20% av alla (röstberättiga) personer som bor i detta område. Från denna population som består av 10 000 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 4 procentenheter. Hur stort urval måste man minst göra för att detta villkor skall vara uppfyllt? (10 poäng)

Uppgift 3: (10 poäng)

I Dala-Järna vill man uppskatta den genomsnittliga månadsinkomsten för alla inkomsttagare (månadsinkomsten i tusentals kronor). Från denna population som består av 1 000 individer tänker man göra ett obundet slumpmässigt urval utan återläggning. Urvalsstorleken bestäms med ledning av ett givet precisionsvillkor.

Man antar att populationsvariansen är lika med 25. Med hjälp av urvalet vill man uppskatta den sökta parametern med ett konfidensintervall med 95% konfidensgrad som har en felmarginal på högst 0,5 tusentals kronor. Hur stort urval måste man minst göra för att detta villkor skall vara uppfyllt? (10 poäng)

Uppgift 4: (20 poäng)

En population består av 1 000 företag i en viss bransch. Man vill undersöka den genomsnittliga omsättningen 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 omsättningen 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	Stickprovsmedel värde	Stickprovsstandardavvikelse
1	10	10	290	60
2	190	40	40	10
3	800	100	6	2

Beräkna ett 95%-igt konfidensintervall för den sökta parametern.

(totalt 20 poäng: skattningen 5 poäng, skattade variansen 10 poäng, konfidensintervallet 5 poäng)

Uppgift 5: (20 poäng)

I Grönköping (som är en stor stad) önskade man ett visst år studera personalkostnaderna för livsmedelsbutikerna i ett stort område. Bland dessa 100 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 14 10 8 2

Omsättning:

30 120 110 80 20

Totalt omsatte de 100 butikerna 5 000 miljoner kronor.

Beräkna ett 95%-igt konfidensintervall för den totala personalkostnaden för alla 100 livsmedelsbutiker i området via en regressionskattning (med omsättning som hjälpinformation).

(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:

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

Punktskattning	Varians	Variansskattning	Varians	Variansskattning
<i>OSU</i>	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.

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

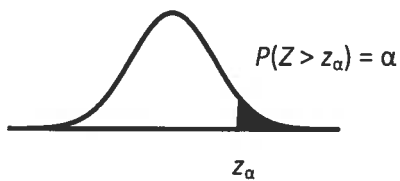
$$b = \frac{\sum_{i=1}^n (z_i - \bar{z})(x_i - \bar{x})}{\sum_{i=1}^n (z_i - \bar{z})^2}$$

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

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

2



Stockholms
universitet

Statistiska institutionen

Rättningsblad

Datum: 14/02/2019

Sal: Ugglevikssalen

Tenta: Undersökningsmetodik

Kurs: Regressionsanalys och undersökningsmetodik

ANONYMKOD:

0036-FDN

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					7
Lär.ant.	40p	10p	10p	18p	20p				

POÄNG	BETYG	Lärarens sign.
98p	A	RC

Uppgift 1

$N=500$

$n=100$

Utan återläggning

a)

besök	0	1	2	3	4	5
personer	35	20	25	12	5	3

$$0 \times 35 + 1 \times 20 + 2 \times 25 + 3 \times 12 + 4 \times 5 + 5 \times 3 = 141$$

$$\bar{x} = 141 \text{ besök}$$

$$\bar{X} = \frac{\bar{x}}{n} = \frac{141}{100} = 1,41$$

R

$$S^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2}{n-1} = \left(\frac{(0-1,41)^2}{100-1} \times 35 \right) + \left(\frac{(1-1,41)^2}{100-1} \times 20 \right) +$$

$$+ \left(\frac{(2-1,41)^2}{100-1} \times 25 \right) + \left(\frac{(3-1,41)^2}{100-1} \times 12 \right) + \left(\frac{(4-1,41)^2}{100-1} \times 5 \right) +$$

$$+ \left(\frac{(5-1,41)^2}{100-1} \times 3 \right) = 0,7028636364 + 0,033959596 +$$

$$+ 0,0879040404 + 0,3064363636 + 0,3387929293 +$$

$$+ 0,1087878788 = 1,578744444$$

$$S^2 = 1,578744444$$

$$\hat{V}(\bar{x}) = \frac{S^2}{n} \left(1 - \frac{n}{N} \right) = \frac{1,578744444}{100} \left(1 - \frac{100}{500} \right) = 0,0126299556$$

KI: $\bar{x} \pm z_{\alpha/2} \times \sqrt{\hat{V}(\bar{x})}$

$$1,41 \pm 1,96 \times \sqrt{0,0126299556}$$

$$1,41 \pm 0,22027$$

$$[1,18973; 1,63027]$$

R

Uppgift 7 b)

$$P_{kvot} = \frac{35}{100} = 0,35$$

35% av de
undersökta har
inte besökt simhallen

$$V(P_{kvot}) = \frac{P(1-P)}{n-1} \left(1 - \frac{n}{N}\right) = \frac{0,35(1-0,35)}{100-1} \left(1 - \frac{100}{500}\right) =$$

$$= 0,0022979798 \times 0,8 = 0,0018383838$$

$$P_{kvot} \pm Z_{\alpha/2} \times \sqrt{V(P_{kvot})}$$

$$0,35 \pm 1,96 \times \sqrt{0,0018383838}$$

$$0,35 \pm 0,0840377011$$

$$[0,26596; 0,43404]$$

Uppgift 2

2) 10p

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

$$\frac{10\,000 \times 0,16}{\left(\frac{0,04}{1,96}\right)^2 (10\,000-1) + 0,16} =$$

$$= \frac{1600}{4,324514786} = 369,98$$

$$N = 10\,000$$

$$P = 20\% = 0,2$$

$$|\text{felmargin}| = 0,04$$

$$z = 1,96$$

$$\sigma^2 = 0,2(1-0,2) = 0,16$$

$$D = \frac{\text{felmargin}}{z_{\alpha/2}} = \frac{0,04}{1,96} = 0,020408$$

(Minst)
 Minst måste urvalet bestå av 370
 individer för att positions vilkoret ska
 uppfyllas

Uppgift 3

$$N=1000$$

$$\sigma^2=25$$

$$\text{felmarginal}=0.5$$

$$z=1.96$$

$$\frac{1000 \times 25}{\left(\frac{0.5}{1.96}\right)^2 (1000-1) + 25} = \frac{25000}{90.0119748} = 277.741$$

för att precisionsvillkoret ska uppfyllas
måste minst 278 individer delta

Uppgift 4

4) 18p

Stratum	N	n	\bar{X}	G	w
1	10	10	290	60	0,01
2	190	40	40	10	0,19
3	800	100	6	2	0,8
Σ	1000	150			

$$\bar{X}_{st} = \sum_{i=1}^3 w_i \bar{X}_i = 0,01 \times 290 + 0,19 \times 40 + 0,8 \times 6 = 15,3$$

$$\hat{V}(\bar{X}_{st}) = \sum_{i=1}^3 w_i^2 \left(\frac{S_i^2}{n_i} \left(1 - \frac{n_i}{N_i} \right) \right) = 0,01^2 \left(\frac{60^2}{10} \left(1 - \frac{10}{10} \right) \right) +$$

$$+ 0,19^2 \left(\frac{10^2}{40} \left(1 - \frac{40}{190} \right) \right) + 0,8^2 \left(\frac{2^2}{100} \left(1 - \frac{100}{800} \right) \right) =$$

$$= 0 + 0,07125 + 0,0224 = 0,306022875$$

$$= 0,09365$$

$$\bar{X}_{st} \pm z_{\alpha/2} \sqrt{\hat{V}(\bar{X}_{st})}$$

$$15,3 \pm 1,96 \times \sqrt{0,306022875}$$

$$15,3 \pm 1,084258953$$

$$[14,215741105; 16,38425895]$$

(SE FACIT)

-2p

5) 20p

Uppgift 5

$N=100$
 $n=5$

Söker totala
regretionskattning

$P_K(x)$	$U_S(z)$
3	30
14	120
10	110
8	80
2	20

$\Sigma z = 5000$ miljoner

$$\bar{x} = \frac{\Sigma x_i}{n} = 7,4$$

$$\bar{z} = \frac{\Sigma z_i}{n} = 72$$

$$\mu_z = \frac{\Sigma z_i}{N} = \frac{5000}{100} = 50$$

$$b = \frac{\Sigma (z_i - \bar{z})(x_i - \bar{x})}{\Sigma (z_i - \bar{z})^2} = \frac{885,6}{8286} \approx 0,1069565217$$

$$\mu_{reg} = \bar{x} + b(\mu_z - \bar{z})$$

$x_i - \bar{x}$	$z_i - \bar{z}$	$(x_i - \bar{x})(z_i - \bar{z})$	$(z_i - \bar{z})^2$	$\frac{(x_i - \bar{x})(z_i - \bar{z})}{(z_i - \bar{z})^2}$
-4,4	-42	184,8	1764	19,36
6,6	48	316,8	2304	43,56
2,6	38	98,8	1444	6,76
0,6	8	4,8	64	0,36
-5,4	-52	280,8	2704	29,16
Σ		885,6	8286	99,2

$$\mu_{reg} = 7,4 + 0,1069565217(50 - 72) = 7,4 + (-2,353043478) = 5,046956522$$

$$\Sigma_{reg} = N \cdot \mu_{reg} = 100 \cdot 5,046956522 = 504,6956522$$

Nästa blad

Uppgift 5 fortsättning

$$\hat{V}(\mu_{reg}) = \left(\frac{N-n}{nN}\right) \left(\frac{1}{n-2}\right) \left[\sum (x_i - \bar{x})^2 - \frac{1}{n} \sum (z_i - \bar{z})^2 \right]$$

$$= \left(\frac{100-5}{5 \cdot 100}\right) \left(\frac{1}{5-2}\right) \left[99,2 - 0,1069565217^2 \cdot 8280 \right] =$$

$$= 0,19 \times \frac{1}{3} \times 4,479305657 = 0,2836893583$$

$$\hat{V}(\hat{\tau}_{reg}) = N^2 \times \hat{V}(\mu_{reg}) = 100^2 \times 0,2836893583 = 2836,893583$$

$$\hat{\tau}_{reg} \pm z_{\alpha/2} \times \sqrt{\hat{V}(\hat{\tau}_{reg})}$$

$$504,6956522 \pm 1,96 \times \sqrt{2836,893583}$$

$$504,6956522 \pm 104,3944941$$

$$[400,3011581; 609,0901463]$$