Stockholms universitet
Statistiska institutionen
Raul Cano

SKRIVNINGSDATUM: 04-06-2019

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. måndagen den 17 juni. 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.

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

Obs! Som vi påpekade flera gånger på delkursen undersökningsmetodik: "av beräkningsmässiga skäl kommer vi att ge många exempel (inklusive tentamens uppgifter) där antal observationer n är lite, typ n = 5 eller n = 10 trots att vad som rekommenderas bör vara minst n = 20 eller n = 30 (alltså stor n)". Konkret: Vi använder normalfördelnings tabell trots att vi har små stickprov.

Obs! Använd minst 5 decimaltal i dina beräkningar. Förtydligande: Om du får t.ex. 0,000003645 då ska du använda 0,000003645 i dina beräkningar och inte avrunda till 0,00000!
Om du avrundar t.ex. 5,43689 till 5,4 i dina beräkningar då blir det poängavdrag!

Uppgift 1: (40 poäng)
Vår målpopulation består att alla 200 företag i en viss bransch och vi är ute efter den genomsnittliga omsättningen i samtliga företag under det senaste året. En stickprovsundersökning genomfördes för n=30 företag (OSU utan återläggning). Följande resultat i miljoner kronor erhölls:

3	6	5	4	1	5	3	5	8	4	7	2	6	4	7
9	7	2	10	5	6	8	4	5	6	9	1	5	8	3

a). Beräkna ett 95%-igt konfidensintervall för den genomsnittliga omsättningen i samtliga företag under det senaste året. (totalt 20 poäng: skattningen 5 poäng, skattade variansen 10 poäng, konfidensintervallet 5 poäng)

b). Beräkna ett 95%-igt konfidensintervall för den andel av företagen i den branschen vars omsättning var minst 6 miljoner kronor vid den aktuella tidpunkten. (totalt 20 poäng: skattningen 5 poäng, skattade variansen 10 poäng, konfidensintervallet 5 poäng)

Uppgift 2: (20 poäng)
De svenska hushållens dåliga sparande är ett ständigt bekymmer för landets ekonomer. För att undersöka sparbenägenheten hos en viss population bestående av 1000 familjer delade en bankorganisation in populationen i två strata beroende på familjens bruttoinkomst. Stratum 1 bestod av de familjer som hade en bruttoinkomst på högst 150 000 kr per år och Stratum 2 bestod av de familjer vars bruttoinkomst var högre. I stratum 1 fanns 600 familjer och i stratum 2 fanns 400 familjer. Man ansåg sig ha råd med 300 observationer. Dessa fördelades med hjälp av proportionellt stratifierat urval. Man fick sedan följande stickprovsmedelvärde och stickprovsstandardavvikelse för månadssparande hos de tillfrågade familjerna (enheten är kronor per månad):

Stratum	Stickprovsmedelvärde	Stickprovsstandardavvikelse
1	90	70
2	130	50

Beräkna ett 95%-igt konfidensintervall för det genomsnittliga månadssparandet i hela populationen.
(totalt 20 poäng: skattningen 5 poäng, skattade variansen 10 poäng, konfidensintervallet 5 poäng)

Uppgift 3: (10 poäng)
Vid Grönköpings universitet undersöker vi studentföreningars ekonomi under det senaste året. Det finns tjugo studentföreningar vid Grönköpings universitet. Då har vi en population som består av 20 grupper eller kluster. Det betyder att i den här uppgiften gäller gruppurval. Fem föreningar utvaldes slumpmässigt (OSU utan återläggning) och följande uppgifter om bl. a. föreningars tillgångar (i tusentals kronor) insamlades:

Tillgångar (i tkr)	Antal medlemmar	Antal medlemmar som kommer att rösta på Fridolin
32	95	8
86	192	22
52	126	11
75	155	18
28	83	6

a). Skatta de totala tillgångarna som studentföreningarna förfogar över. (5 poäng)

b). Skatta antalet medlemmar som kommer att rösta på Fridolin bland medlemmarna inom studentföreningarna. (5 poäng)

Uppgift 4: (20 poäng)
Globaliseringen har kommit till Grönköping. Den största matkedjan i Grönköping har fått det första partiet av exotiska frukter från Majuro som är en atoll som fungerar som huvudstad för Marshallöarna i Stilla havet. Under åren 1946–58 sprängde USA flera vätebomber på atollerna Bikini och Enewetak i Marshallöarna. Majuro ligger nära atollen Bikini. Då Livsmedelsverket i Grönköping vill kontrollera om de frukterna som har kommit från Mjuro har en normal halt av radioaktivt cesium. Livsmedelsverket har fastställt ett gränsvärde om 300 Bq/kg (Becquerel/kilogram) för högsta tillåtna halt av radioaktivt cesium i all mat och frukter som säljs. Det finns 1000 exotiska frukter i det första partiet och Livsmedelsverket i Grönköping har redan genomfört en urvalsundersökning för att fastställa den genomsnittliga halten av radioaktivt cesium i hela partiet. Den totala vikten i hela partiet var 225000 gram.

Följande resultat erhölls (fem slumpmässigt valda exotiska frukter, OSU utan återläggning):

Halt av radioaktivt cesium i Bq/kg	Vikt i gram
260	210
265	220
275	225
280	230
255	215

Beräkna ett 95%-igt konfidensintervall för den genomsnittliga halten av radioaktivt cesium i hela partiet via en regressionsskattning (med vikt som hjälpinformation).
(totalt 20 poäng: skattningen 5 poäng, skattade variansen 10 poäng, konfidensintervallet 5 poäng)

Uppgift 5: (10 poäng)

Man har redan genomfört en urvalsundersökning om inkomst (månadsinkomst efter skatt) hos anställda i ett företag. Det finns 100 anställda i företaget. Eftersom man trodde att ålder var en av de faktorer som påverkade inkomsten, delade man in populationen i två åldersklasser samt drog man ett obundet slumpmässigt urval (utan återläggning) ur respektive åldersklass. Man hade råd och tid att dela ut en enkät till 20 personer och man fördelade stickprovet enligt optimal allokering (Neyman-allokering). Man antog i undersökningen att $\sigma_1 + \sigma_2 = 6{,}84$, där σ_i står för standardavvikelsen i stratum nummer i.

Följande data har erhållits:

Stratum Åldersklass	Population N_i	Urval n_i
18-30	60	15
31-64	40	5
Summa	100	20

Genom att använda all information som finns ovan, är din uppgift att beräkna de värdena på σ_1 och σ_2 som har använts i undersökningen. (10 poäng)

Formelsamling undersökningsmetodik

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

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

Beräkning av stickprovsstorlek:

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

Stratifierat urval:

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

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:

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

Punktskattning	*Varians*	*Variansskattning*	*Varians*	*Variansskattning*
OSU	*m.å.*	*m.å.*	*u.å.*	*u.å.*
$\overline{X}$	$\frac{\sigma^2}{n}$	$\frac{s^2}{n}$	$\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$	$\frac{s^2}{n}\left(1-\frac{n}{N}\right)$
$\hat{\tau}$	$N^2 \cdot \frac{\sigma^2}{n}$	$N^2 \cdot \frac{s^2}{n}$	$N^2 \cdot \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$	$N^2 \cdot \frac{s^2}{n}\left(1-\frac{n}{N}\right)$
p	$\frac{P(1-P)}{n}$	$\frac{p(1-p)}{n-1}$	$\frac{P(1-P)}{n}\left(\frac{N-n}{N-1}\right)$	$\frac{p(1-p)}{n-1}\left(1-\frac{n}{N}\right)$
$\hat{A}$	$N^2 \cdot \frac{P(1-P)}{n}$	$N^2 \cdot \frac{p(1-p)}{n-1}$	$N^2 \cdot \frac{P(1-P)}{n}\left(\frac{N-n}{N-1}\right)$	$N^2 \cdot \frac{p(1-p)}{n-1}\left(1-\frac{n}{N}\right)$

Tillägg till formelsamling undersökningsmetodik

Skattning av τ_X. Urval OSU

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

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

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

Skattning av μ_X. Urval OSU.

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

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

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

Ur Formelsamling regressionsanalys

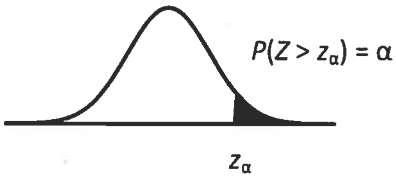
$$\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})=\sum_{i=1}^{n}x_i\,y_i-\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)\left(\sum_{i=1}^{n}y_i\right)$$

$$\sum_{i=1}^{n}(x_i-\overline{x})^2=\sum_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)^2$$

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

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 4/6-2019

Sal: Uggleviksalen

Tenta: Undersökningsmetodik

Kurs: Regressionsanalys och undersökningsmetodik

ANONYMKOD: 0020-HZK

[X] 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
Lär.ant. 40p	20p	10p	20p	10p					

IIC

POÄNG	BETYG	Lärarens sign.
100p	A	RC

SU, STATISTIK

Skrivsal: UG **Anonymkod:** 0020-HZK **Blad nr:** 1

1) 40p

Uppgift 1

$n=30 \quad N=200$

a) $\bar{x} = \frac{3+6+\ldots+3}{30} = \frac{158}{30} \approx 5{,}27$ R

$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{496 - \frac{158^2}{30}}{29} \approx 5{,}65057$ R

$\hat{V}(\bar{x}) = \frac{5{,}65057}{30}\left(1-\frac{30}{200}\right) = 0{,}16$

95% KI för $\bar{x}$

$\bar{x} \pm z_{0.025} \cdot \sqrt{\hat{V}(\bar{x})}$

$5{,}27 \pm 1{,}96\sqrt{0{,}16}$

$5{,}27 \pm 1{,}96 \cdot 0{,}4$

$\pm\ 0{,}784$

$[4{,}486 ; 6{,}054]$ R

b) $p = \frac{13}{30} \approx 0{,}433$ R

$\hat{V}(p) = \frac{0{,}433 \cdot 0{,}567}{29}\left(1-\frac{30}{200}\right) = 0{,}007196$ R

$p \pm z_{0.025} \cdot \sqrt{\hat{V}(p)}$

$0{,}433 \pm 1{,}96\sqrt{0{,}007196}$

$0{,}433 \pm 0{,}166$

$[0{,}267 ; 0{,}599]$ R

Uppgift 2

2) 20p

$N = 1000$ familjer

strata $L = 2$

$N_1 = 600 \quad N_2 = 400$

$n = 300$ PSU

$n_1 = 300 \cdot \frac{600}{1000} = 180$

R

$n_2 = 300 \cdot \frac{400}{1000} = 120$

$\bar{X}_{st} = \underbrace{\frac{600}{1000}}_{W_1 \; 0,6} \cdot 90 + \underbrace{\frac{400}{1000}}_{W_2 = 0,4} \cdot 130 = 106$

R

$V(\bar{X}_{st}) = 0,6^2 \cdot \frac{70^2}{180}\left(1 - \frac{180}{600}\right) + 0,4^2 \cdot \frac{50^2}{120}\left(1 - \frac{120}{400}\right) \approx 9,1933$

R

$\bar{X}_{st} \pm Z_{0.025} \sqrt{V(\bar{X}_{st})}$

$106 \pm 1,96 \cdot \sqrt{9,1933}$

$106 \pm 5,9428$

$[100,06 ; 111,94]$

R

SU, STATISTIK

Skrivsal: UG Anonymkod: 0020-H2K Blad nr: 2

3) 10p

Uppgift 3

N= 20 Grupper

n=5 OSU

x=tillgångar m=antal medlemmar a=rösta Fridolin

a) $\bar{x}_{kvot} = \frac{32+86+52+75+28}{95+192+126+155+83} = \frac{273}{651} = \frac{13}{31}$

$\bar{m} = \frac{95+192+\ldots+83}{5} = 130{,}2$

$\hat{M} = \bar{m} \cdot N = 2604$

$\hat{T} = \hat{M} \cdot \bar{x}_{kvot} = 2604 \cdot \frac{13}{31} = 1092$

R

b) $p_{kvot} = \frac{8+22+11+18+16}{651} = \frac{65}{651} \approx 0{,}0998$

$\bar{m} = 130{,}2$

$\hat{M} = \bar{m} \cdot N = 130{,}2 \cdot 20 = 2604$

$\hat{A} = \hat{M} \cdot p_{kvot} = 2604 \cdot 0{,}0998 \approx 260$

R

Uppgift 4

4) 20p

μ_{reg} vikt = z x = halt av radioaktiv

$T_z = 225\,000$ N=1000 $\mu_z = \frac{225\,000}{1000} = 225$ n=5

$\hat{\mu}_{reg} = \bar{x} + b(\mu_z - \bar{z})$ $\bar{z} = \frac{1100}{5} = 220$ $\bar{x} = \frac{1335}{5} = 267$

	x	z	$x-\bar{x}$	$z-\bar{z}$	$(\bar{x}-x)(z-\bar{z})$	$(x-\bar{x})^2$	$(z-\bar{z})^2$
	260	210	-7	-10	70	49	100
	265	220	-2	0	0	4	0
	275	225	8	5	40	64	25
	280	230	13	10	130	169	100
	255	215	-12	-5	60	144	25
Σ					300	430	250

$b = \frac{300}{250} = 1{,}2$

$\hat{\mu}_{reg} = 267 + 1{,}2(225 - 220) = 273$ R

$\hat{V}(\hat{\mu}_{reg}) = \left(\frac{1000-5}{5 \cdot 1000}\right)\left(\frac{1}{5-2}\right) \cdot \left[430 - 1{,}2^2 \cdot 250\right] = 4{,}6433$ R

$\hat{\mu}_{reg} \pm z_{0.025}\sqrt{\hat{V}(\hat{\mu}_{reg})}$

$273 \pm 1{,}96 \cdot \sqrt{4{,}6433}$

$273 \pm 4{,}223$

$\boxed{[268{,}78 ; 277{,}22]}$ R

SU, STATISTIK

Skrivsal: OG Anonymkod: 0020-HZK Blad nr: 3

5) 10p

Uppgift 5

$N_1 = 60$ $N_2 = 40$ $N = 100$

$n_1 = 15$ $n_2 = 5$ $n = 20$

$$n_i = n \cdot \frac{N_i \sigma_i}{\sum N_j \sigma_j}$$

$\sigma_1 + \sigma_2 = 6{,}84$

$N_1 \sigma_1 = 60\sigma_1$ $N_2 \sigma_2 = 40\sigma_2$

$\sum N_j \sigma_j = 60\sigma_1 + 40\sigma_2$

$$\frac{20 \cdot \frac{60\sigma_1}{60\sigma_1 + 40\sigma_2}}{20} = 15 / 20$$

$$\frac{60\sigma_1}{60\sigma_1 + 40\sigma_2} = 0{,}75$$

$\sigma_2 = 6{,}84 - \sigma_1$

$60\sigma_1 = 0{,}75(60\sigma_1 + 40(6{,}84 - \sigma_1))$

$60\sigma_1 = 0{,}75(60\sigma_1 + 273{,}6 - 40\sigma_1)$

$60\sigma_1 = 15\sigma_1 + 205{,}2$

$$\frac{45\sigma_1}{45} = \frac{205{,}2}{45}$$

$\sigma_1 = 4{,}56$ R

$\sigma_2 = 6{,}84 - \sigma_1 = 6{,}84 - 4{,}56 = 2{,}28$ R

Svar: $\sigma_1 = 4{,}56$ $\sigma_2 = 2{,}28$ R

BCA