Stockholms universitet
Statistiska institutionen
Raul Cano

SKRIVNINGSDATUM: 31-05-2018

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: fredagen den 15 juni, kl. 16.00 i B705.
Därefter kan skrivningarna hämtas på studentexpeditionen, plan 7 i B-huset.

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: (30 poäng)
Antag att vi har en känd population bestående av $N = 6$ personer, A, B, C, D, E och F anställda i ett företag och med månadslönerna (i tkr)
$x_1 = 27,\ x_2 = 29,\ x_3 = 31,\ x_4 = 31,\ x_5 = 33,\ x_6 = 29$
Ur denna population skall man göra ett obundet slumpmässigt urval (OSU) utan återläggning om $n = 2$ element.

a). Beräkna samplingfördelningen för den stokastiska variabeln $\overline{X}$. (5 poäng)
b). Beräkna $E(\overline{X})$. (5 poäng)
c). Beräkna $V(\overline{X})$. (5 poäng)
d). Beräkna samplingfördelningen för den stokastiska variabeln s^2. (5 poäng)
e). Beräkna $E(s^2)$. (5 poäng)
f). Beräkna $V(s^2)$. (5 poäng)

Uppgift 2: (10 poäng)
I Grönköping och ur en skolkatalog på lågstadiet gör vi ett obundet slumpmässigt urval (OSU) utan återläggning bestående av tre klasser. Genom en enkät som samtliga elever besvarar i de tre klasserna samlar vi in uppgifter om mobiltelefonanvändning. Följande resultat erhölls:

Klass	Antal elever	Antal ringda samtal per dag	Antal mobiltelefonanvändare
1	25	10	4
2	28	12	6
3	30	18	8

a). Skatta antalet ringda samtal per elev och dag. (5 poäng)
b). Skatta andelen mobiltelefonanvändare. (5 poäng)

Uppgift 3: (20 poäng)
I Grönköping önskade man undersöka hur stor andel av invånarna i stadsdelen Vasa som funderade på att se på tv den kommande partiledardebatten. Man beslutade om en stickprovsundersökning om n=600 personer . I den aktuella stadsdelen bor N=4000 personer. Därtill stratifierades populationen med avseende på ålder. OSU utan återläggning användes för att hämta stickprov ur respektive strata. Följande resultat erhölls:

Stratum	N_i	n_i	p_i
15-29 år	500	100	0,30
30-49 år	1500	100	0,40
50 år eller högre	2000	400	0,60

Beräkna ett 95%-igt konfidensintervall för andelen invånare i Vasa som funderar på att se på tv den kommande partiledardebatten. (totalt 20 poäng: skattningen 5 poäng, skattade variansen 10 poäng, konfidensintervallet 5 poäng)

Uppgift 4: (20 poäng)
Antag att vår population består att alla företag i en viss bransch och att vi vill undersöka den totala omsättningen i samtliga företag under det senaste året. Populationen har stratifierats i tre stratum efter företagsstorlek. En stickprovsundersökning genomfördes för n=300 företag (OSU utan återläggning). Följande resultat erhölls:

Stratum	N_i	n_i	$\bar{x}_i$ (i milj. kr.)	s_i
Stora	20	10	500	30
Mellanstora	180	50	200	20
Små	1000	240	100	10

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

Uppgift 5: (20 poäng)
Skattemyndigheten i Grönköping är ute efter Don Corleone som äger den berömda och framgångsrika pizzerian "Mamma Mia". För inkomståret 2017 har Don Corleone deklarerat en pizzainkomst som motsvarar totalt 5 000 pizzor. Skattemyndigheten har redan genomfört en kontroll på antal sålda pizzor på fem slumpmässigt valda veckor under år 2017 (OSU utan återläggning, år 2017 har 52 veckor). Dessutom har Skattemyndigheten uppgifter på alla vetemjölleveranser till pizzerian "Mamma Mia" under det aktuella året (t.ex. det levererades totalt 600 kg. vetemjöl till pizzerian under 2017). Följande resultat erhölls:

Vecka	**Antal sålda pizzor**	**Vetemjöl som levererades (i kg.)**
35	80	10
12	90	11
46	120	15
8	70	8
29	75	8

a). Beräkna ett 95%-igt konfidensintervall (med hjälp av regressionsskattningen) för det totala antalet sålda pizzor på pizzerian "Mamma Mia" under 2017. (totalt 15 poäng: skattningen 3 poäng, skattade variansen 10 poäng, konfidensintervallet 2 poäng)

b). Enligt konfidensintervallet, har Don Corleone deklarerat rätt eller fel (ange varför) för inkomståret 2017? (5 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 OSU	*Varians* m.å.	*Variansskattning* m.å.	*Varians* u.å.	*Variansskattning* 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.

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

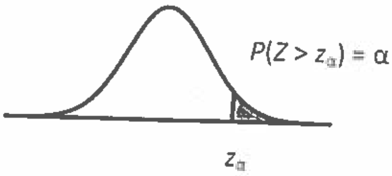
$$b = \frac{\sum_{i=1}^{n} \left(z_i - \bar{z}\right)\left(x_i - \bar{x}\right)}{\sum_{i=1}^{n} \left(z_i - \bar{z}\right)^2}$$

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

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

© 2013 Michael Carlson, Statistiska inst., Sthlm:s univ.

44

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 31/05/18

Sal: Ugglevikssalen

Tenta: Undersökningsmetodik

Kurs: Regressionsanalys och undersökningsmetodik

ANONYMKOD: 0047-CSS

☒ 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 iF
Lär.ant. 29p	10p	20p	15p	20p					

POÄNG	BETYG	Lärarens sign.
94p	A	RC

SU, STATISTIK

Skrivsal: UG **Anonymkod:** 0047-CSS **Blad nr:** 1

1) 29 p

7) $N=6 \quad n=2$

antal möjliga utfall $\binom{6}{2} = \frac{6!}{2!\,4!} = \frac{6 \cdot 5}{2 \cdot 1} = 15$

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

OSU	$\bar{x}$	s^2
27,29	28	$(27-28)^2+(29-28)^2/1=2$
27,31	29	$2^2+2^2=8$
27,31	29	$2^2+2^2=8$
27,33	30	$3^2+3^2=18$
27,29	28	$1^2+1^2=2$
29,31	30	$1^2+1^2=2$
29,31	30	$1^2+1^2=2$
29,33	31	$2^2+2^2=8$
29,29	29	$0^2+0^2=0$
31,31	31	$0^2+0^2=0$
31,33	32	$1^2+1^2=2$
31,29	30	$1^2+1^2=2$
31,33	32	$1^2+1^2=2$
31,29	30	$1^2+1^2=2$
33,29	31	$2^2+2^2=8$

R

a)

$\bar{x}$	28	29	30	31	32	
$P(\bar{x})$	2/15	3/15	5/15	3/15	2/15	$\sum P(\bar{x})$

R

b) $E(\bar{x}) = 28 \cdot \frac{2}{15} + 29 \cdot \frac{3}{15} + 30 \cdot \frac{5}{15} + 31 \cdot \frac{3}{15} + 32 \cdot \frac{2}{15} = 30$ Svar: 30

c) $V(\bar{x}) = (28-30)^2 \cdot \frac{2}{15} + (29-30)^2 \cdot \frac{3}{15} + (30-30)^2 \cdot \frac{5}{15} + (31-30)^2 \cdot \frac{3}{15} +$
$(32-30)^2 \cdot 2/15$

$= 8/15 + 3/15 + 0 + \frac{1}{5} + 8/15 \approx 1{,}86666$ → −1p Svar: 1,86666

R

d)

s^2	0	2	8	18
$P(s^2)$	2/15	8/15	4/15	1/15

R

R

e) $E(s^2) = 0 \cdot 2/15 + 2 \cdot 8/15 + 8 \cdot 4/15 + 18 \cdot 1/15 = 4{,}4$ Svar: 4,4

f) $V(s^2) = (0-4{,}4)^2 \cdot 2/15 + (2-4{,}4)^2 \cdot 8/15 + (18-4{,}4)^2 \cdot 1/15 +$
$(8-4{,}4)^2 \cdot 4/15$

$= \frac{[illegible]}{25} + \frac{384}{125} + \frac{[illegible]}{125} + \frac{1024}{375} + \frac{[illegible]}{25} = [illegible]$

Svar: [illegible]

R

② a) 10p

Klass	Antal elever (m_i)	ringde samtal (T_i)	mobil använt (a_i)
1	25	10	4
2	28	12	6
3	30	18	8
Σ	83	40	18

a) Andelet ringde samtal / elever / dag

$\frac{40}{83} \approx \underline{0,482}$ $\left(\frac{\Sigma T_i}{\Sigma m_i}\right)$ R

b) Andel mobil använt

$\frac{18}{83} \approx \underline{0,217}$ $\left(\frac{\Sigma a_i}{\Sigma m_i}\right)$ R

SU, STATISTIK

Skrivsal: UG   Anonymkod: 0047-CSS   Blad nr:

3) 20p

(3) $n=600$ $N=4000$

S	N_i	n_i	P_i
15-29	500	100	0,3
30-49	1500	100	0,4
50≤	2000	400	0,6
Σ	4000	600	

$\bar{x}_{st} = \sum W_i \bar{x}_i$ $W_i = \frac{N_i}{N}$

$\bar{x} = \frac{500}{4000} \cdot 0,3 + \frac{1500}{4000} \cdot 0,4 + \frac{2000}{4000} \cdot 0,6 = \frac{39}{80} = 0,4875$ R

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

$V(\bar{x}_1) = \frac{0,3 \cdot 0,7}{99}\left(1 - \frac{100}{500}\right) = 0,001696$

$V(\bar{x}_2) = \frac{0,4 \cdot 0,6}{99}\left(1 - \frac{100}{1500}\right) = 0,00226262$

$V(\bar{x}_3) = \frac{0,6 \cdot 0,4}{399}\left(1 - \frac{400}{2000}\right) = 0,000481203$

$V(\bar{x}_{st}) = \left(\frac{500}{4000}\right)^2 \cdot 0,001696 + \left(\frac{1500}{4000}\right)^2 \cdot 0,00226262 +$
$\left(\frac{2000}{4000}\right)^2 \cdot 0,000481203 = 0,0000265 +$
$+ 0,000318180 + 0,000120 = 0,000464995$ R

95% KI ges av $\bar{x} \pm 1,96\sqrt{V(\bar{x}_{st})}$

$0,4875 \pm 1,96\sqrt{0,000464995} = 0,4875 \pm 0,0422$

$\boxed{0,445 ,\ 0,529}$ med 95% säkerhet ligger andelen som fundera[illegible] [illegible] [illegible] R

4) 15p

(4) $n = 300$

Stratum	N_i	n_i	$\bar{x}$	s_i
Stora	20	10	500	30
Med	180	50	200	20
Små	1000	240	100	10
Σ	1200	300		

a) totala omsättningen: $\sum (N_i \cdot \bar{x}_i)$

$= 20 \cdot 500 + 180 \cdot 200 + 1000 \cdot 100 = 146\,000$ (i milj kr)

$\hat{T}_{st} = 146\,000$ R

$\hat{V}(\hat{T}) = N^2 \cdot \frac{s^2}{n}\left(1 - \frac{n}{N}\right)$

$\hat{V}(\hat{T}_1) = 20^2 \cdot \frac{30^2}{10}\left(1 - \frac{10}{20}\right) = 18000$

$\hat{V}(\hat{T}_2) = 180^2 \cdot \frac{20^2}{50}\left(1 - \frac{50}{180}\right) = 187200$

$\hat{V}(\hat{T}_3) = 1000^2 \cdot \frac{10^2}{240}\left(1 - \frac{240}{1000}\right) \approx 316666$

$\hat{V}(\hat{T}_{st}) = \left(\frac{20}{1200}\right)^2 \cdot 18000 + \left(\frac{180}{1200}\right)^2 \cdot 187200 + \left(\frac{1000}{1200}\right)^2 \cdot 316666$

≈ 224124 −3p

95% k.i. för totala omsättningen i [illegible]

$146000 \pm 1.96\sqrt{224124} \approx 146\,000 \pm 928$

[145 072, 146 928] Vi kan med 95% säkerhet säga att totala omsättning ligger inom intervallet −2p

SU, STATISTIK

Skrivsal: UG **Anonymkod:** 0047-CSS **Blad nr:** 3

(5) $N = 52$ 5) 2 0p

Vedra	Sålt X	mil Z
35	80	10
12	90	11
46	120	15
8	70	8
24	75	8
Σ 5	435	52

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

$\bar{x} = 435/5 = 87$ $\bar{z} = 52/5 = 10{,}4$ kg

$T_z = 600$ kg $\mu_z = 600/52 \approx 11{,}54$

$b = \frac{\Sigma(z_i - \bar{z})(x_i - \bar{x})}{\Sigma(z_i - \bar{z})^2}$ $b = \frac{226}{33{,}2} \approx 6{,}807$

	$(z_i - \bar{z})(x_i - \bar{x})$	$(z_i - \bar{z})^2$	$(x_i - \bar{x})^2$
	$(10-10{,}4)(80-87)$ 2,8	$-0{,}4^2$	-7^2
	$(11-10{,}4)(90-87)$ 1,8	$0{,}6^2$	3^2
	$(15-10{,}4)(120-87)$ 151,8	$4{,}6^2$	33^2
	$(8-10{,}4)(70-87)$ 40,8	$-2{,}4^2$	-17^2
	$(8-10{,}4)(75-87)$ 28,8	$-2{,}4^2$	-12^2
Σ	226	33,2	1580

$\hat{\mu}_{reg} = 87 + 6{,}807(11{,}54 - 10{,}4) = 94{,}75998$ R

$\hat{\tau}_y = 94{,}75998 \cdot 52 \approx 4927{,}51896$

$\hat{V}(\hat{\mu}_{reg}) = \left(\frac{52-5}{52 \cdot 5}\right)\left(\frac{1}{3}\right)(1580 - (6{,}807^2 \cdot 33{,}2)) \approx 2{,}5108$ R

$\hat{V}(\hat{\tau}_y) = \hat{V}(N \cdot \hat{\mu}_{reg}) = 52^2 \cdot 2{,}5108 \approx 6789{,}3885$

95% konfidensintervall för totala skattningen

$4927 \pm 3{,}182 \cdot \sqrt{6789{,}3885} \approx 4927 \pm 262$

[4665, 5189] R

b) ~~Da~~ konfidensintervallet täcker 5000, Don Corleones angivna siffra, kan vi inte säga att han har deklarerat fel. Däremot kan vi inte heller säga att han har deklarerat rätt med ett så pass brett intervall.