Stockholms universitet
Statistiska institutionen
Raul Cano

SKRIVNINGSDATUM: 17-08-2018

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: tisdagen den 28 augusti, 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: (20 poäng)

Antag att vi har en känd population bestående av $N = 5$ personer, A, B, C, D och E anställda i ett företag och med månadslönerna (i tkr)
$x_1 = 26,\ \ x_2 = 24,\ \ x_3 = 28,\ \ x_4 = 30,\ \ x_5 = 26$
Ur denna population skall man göra ett obundet slumpmässigt urval (OSU) utan återläggning om $n = 3$ element.

a). Beräkna samplingfördelningen för den stokastiska variabeln s^2. (10 poäng)
b). Beräkna $E(s^2)$. (5 poäng)
c). Beräkna $V(s^2)$. (5 poäng)

Uppgift 2: (20 poäng)

På Grönköpings universitet ville man undersöka hur stor andel av studenter som köper sina kursböcker på nätet. Man beslutade om en stickprovsundersökning om n=300 studenter. Det finns totalt 2000 studenter på Grönköpings universitet. 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
17-25 år	1000	150	0,90
26-30 år	800	100	0,80
31 år eller högre	200	50	0,60

Beräkna ett 95%-igt konfidensintervall för andelen studenter på Grönköpings universitet som köper sina kursböcker på nätet. (totalt 20 poäng: skattningen 5 poäng, skattade variansen 10 poäng, konfidensintervallet 5 poäng)

Uppgift 3: (20 poäng)
Globalisering har kommit till Grönköping, inklusive problemet med bluffakturor i olika företag. Population består av alla företag i Grönköping. Vår undersökningsvariabel är det fakturabelopp för den värsta bluffaktura som har kommit in till varje företag under det senaste året. Man vill skatta det genomsnittliga fakturabeloppet för samtliga företag under det senaste året. Populationen har stratifierats i tre stratum efter företagsstorlek. En stickprovsundersökning genomfördes för n=150 företag (OSU utan återläggning). Följande resultat erhölls:

Stratum	N_i	n_i	$\bar{x}_i$ (i tusentals kr.)	s_i
Stora	50	30	34	9
Mellanstora	100	60	23	4
Små	200	60	18	3

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

Uppgift 4: (20 poäng)
I Grönköping kommer ett företag i reklambranschen att publicera en lyxig broschyr för vegetarianer. Företaget vill ange en skattning på energiinnehåll i Chiquita-bananer (klass 1) och har anlitat Kalle Andersson för jobbet. Kalle har studerat 30 hp i statistik på Stockholms universitet och dessutom har några kompisar som är biokemister på Arrheniuslaboratoriet. Energiinnehållet ska ges i termer av kalorier. Kalle har redan skaffat ett stort parti på 1000 Chiquita-bananer (klass 1). Han vill ange en skattning på energiinnehåll i Chiquita-bananer (klass 1) genom att skatta det genomsnittliga energiinnehållet i hela partiet av bananer via en kvotskattning (med vikt som hjälpinformation). Första steget är redan gjort: Kalles sambo Kristina har vägt varenda banan i partiet och har fått 225000 gram som den totala vikten i hela partiet. Andra steget är också redan gjort: För att fastslå antal kalorier i fem slumpmässigt valda bananer har Kalle skickat de valda bananerna till sina kompisar på Arrheniuslaboratoriet.
Följande resultat erhölls (fem slumpmässigt valda bananer, OSU utan återläggning):

Antal kalorier	Vikt i gram
210	215
215	230
205	225
210	220
205	210

a). Beräkna ett 95%-igt konfidensintervall för det genomsnittliga energiinnehållet i hela partiet av bananer via en kvotskattning (med vikt som hjälpinformation).
(skattade variansen 10 poäng, konfidensintervallet 5 poäng) .

b). Ange den skattning på energiinnehåll som Kalle levererar till företaget. (skattningen 5 poäng)

Uppgift 5: (20 poäng)
Använd alla data som finns i **Uppgiften 4** för att beräkna ett 95%-igt konfidensintervall för det genomsnittliga energiinnehållet i hela partiet av bananer via en regressionsskattning (med vikt som hjälpinformation).
(totalt 20 poäng: skattningen 5 poäng, skattade variansen 10 poäng, konfidensintervallet 5 poäng)

Formelsamling undersökningsmetodik

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

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

Beräkning av stickprovsstorlek:

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

Stratifierat urval:

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

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:

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

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	$\frac{P(1-P)}{n}$	$\frac{p(1-p)}{n-1}$	$\frac{P(1-P)}{n}\left(\frac{N-n}{N-1}\right)$	$\frac{p(1-p)}{n-1}\left(1-\frac{n}{N}\right)$
$\hat{A}$	$N^2 \cdot \frac{P(1-P)}{n}$	$N^2 \cdot \frac{p(1-p)}{n-1}$	$N^2 \cdot \frac{P(1-P)}{n}\left(\frac{N-n}{N-1}\right)$	$N^2 \cdot \frac{p(1-p)}{n-1}\left(1-\frac{n}{N}\right)$

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.

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

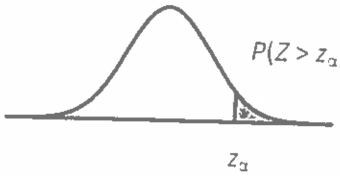
$$b = \frac{\sum_{i=1}^{n}\left(z_i - \bar{z}\right)\left(x_i - \bar{x}\right)}{\sum_{i=1}^{n}\left(z_i - \bar{z}\right)^2}$$

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

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

4

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 17/8-2018

Sal: Värtasalen

Tenta: Undersökningsmetodik

Kurs: Regressionsanalys och undersökningsmetodik

ANONYMKOD: 0031-DZT

☒ 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.	0p	20p	20p	20p	20p					

POÄNG	BETYG	Lärarens sign.
80p	B	RC

SU, STATISTIK

Skrivsal: VÅ Anonymkod: 0031-DZT Blad nr: 1

1) 0p

1

$N=5 \quad n=3 \quad \binom{5}{3} = \frac{5!}{3!(5!-3!)} = \frac{5\cdot4\cdot3\cdot2\cdot1}{3\cdot2\cdot1\cdot2\cdot1} = \frac{20}{2} = 10$

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

				$\bar{x}$	s^2
1	26	24	28	26	2
2	26	24	30	80/3	4,66666675
3	26	24	26	76/3	2,666667
4	26	28	30	28	2
5	26	28	26	80/3	2,666667
6	26	30	26	82/3	2,66666675
7	24	28	30	82/3	4,66666675
8	24	28	26	26	2
9	24	30	26	80/3	4,66666675
10	28	30	28	28	2

F (SE FACIT)

a) Givet att följande approximationer görs: $2{,}66666675 \approx 2{,}666667 \approx 2{,}6667$ samt $4{,}66666675 \approx 4{,}6667$ är samplingfördelningen för s^2:

2	4/10
2,6667	3/10
4,6667	3/10

F

b) $E(s^2) = 2\cdot(4/10) + 2{,}6667\cdot(3/10) + 4{,}6667\cdot(3/10) = \frac{30}{10} = \underline{3}$

F

c) $V(s^2) = (3-2)^2\cdot(4/10) + (3-2{,}6667)^2\cdot(3/10) + (3-4{,}6667)^2\cdot(3/10) =$

$= (4 + 0{,}3333 + 8{,}3332)/10 = \underline{1{,}2666}$

F

2

$n=300 \quad N=2000$ 2) 20p R

$\hat{P}_{st} = 0{,}90 \cdot \left(\frac{1000}{2000}\right) + 0{,}20 \cdot \left(\frac{200}{2000}\right) + 0{,}60 \cdot \left(\frac{200}{2000}\right) = 0{,}45 + 0{,}02 + 0{,}06 = 0{,}83$

$\hat{V}(P_1) = \frac{0{,}9 \cdot 0{,}1}{149}\left(1 - \frac{150}{1000}\right) = 0{,}0005134228188$

$\hat{V}(P_2) = \frac{0{,}8 \cdot 0{,}2}{99}\left(1 - \frac{100}{200}\right) = 0{,}00141414141414$

$\hat{V}(P_3) = \frac{0{,}6 \cdot 0{,}4}{49}\left(1 - \frac{50}{200}\right) = 0{,}0036734694$

$\hat{V}(\hat{P}_{st}) = \hat{V}(P_1) \cdot \left(\frac{1000}{2000}\right)^2 + \hat{V}(P_2) \cdot \left(\frac{200}{2000}\right)^2 + \hat{V}(P_3) \cdot \left(\frac{200}{2000}\right)^2 =$

$= 0{,}0001283557047 + 0{,}0000226262621 + 0{,}000036734694 =$

$= 0{,}0003915530227$ R

$\sqrt{\hat{V}(P_{st})} \approx 0{,}0198$

95% ki ges av

$P_{st} \pm 1{,}96 \cdot \sqrt{\hat{V}(P_{st})}$ R

$0{,}83 \pm 1{,}96 \cdot 0{,}0198$

$0{,}83 \pm 0{,}0388$

$[0{,}7912 ; 0{,}8688]$ Utgör ett 95%igt konfidensintervall för andelen studenter som köper kurslitteratur på internet.

SU, STATISTIK

Skrivsal: VÅ Anonymkod: 0031-DZT Blad nr: 2

3) 20p

3

$N = 350 \quad n = 150$

$\hat{\mu}_x = 34 \cdot \left(\frac{50}{350}\right) + 22 \cdot \left(\frac{100}{350}\right) + 18 \cdot \left(\frac{200}{350}\right) = 21{,}71428571$ R

$\hat{V}(\hat{\mu}_x) = \frac{9^2}{30}\left(1 - \frac{30}{50}\right) \cdot \left(\frac{50}{350}\right)^2 + \frac{4^2}{60}\left(1 - \frac{60}{100}\right) \cdot \left(\frac{100}{350}\right)^2 + \frac{3^2}{60}\left(1 - \frac{60}{200}\right) \cdot \left(\frac{200}{350}\right)^2 =$

$= 1{,}08 \cdot \left(\frac{50}{350}\right)^2 + 0{,}10666667 \cdot \left(\frac{100}{350}\right)^2 + 0{,}105 \cdot \left(\frac{200}{350}\right)^2 =$

$= 0{,}02204082 + 0{,}00870748 + 0{,}03428571 = 0{,}06503401$ R

$\sqrt{\hat{V}(\hat{\mu}_x)} = 0{,}2550$

95% ki ges av

$21{,}7143 \pm 1{,}96 \cdot 0{,}2550$

$21{,}7143 \pm 0{,}4998$

$[21{,}2145 ; 22{,}2141]$ R utgör ett 95%-igt konfidensintervall för det genomsnittliga fakturabeloppet för samtliga företag under det senaste året.

R

4) 2 p

$N = 1000 \quad n = 5 \quad \tau_z = 225\,000$

	x_i	z_i	$x_i z_i$	x_i^2	z_i^2
1	210	215	45150	44100	46225
2	215	230	49450	46225	52900
3	205	225	46125	42025	50625
4	210	220	46200	44100	48400
5	205	210	43050	42025	44100
Σ	1045	1100	229975	218475	242250

$\mu_z = \tau_z / N = 225$

$\hat{\mu}_{kvot} = \frac{1045}{1100} \cdot 225 = 213,75$ R

$\Sigma(x_i - R z_i)^2 = 218475 + 0,9025 \cdot 242250 - 2 \cdot 0,95 \cdot 229975 =$

$= 437105,625 - 436952,500 = 153,125$

$\hat{V}(\hat{\mu}_{kvot}) = \left(\frac{1000-5}{5 \cdot 1000}\right) \cdot \frac{153,125}{5-1} = 7,61796875$ R

a) 95%igt ki ges av

$\hat{\mu}_{kvot} \pm 1,96 \cdot \sqrt{\hat{V}(\hat{\mu}_{kvot})}$

$213,75 \pm 1,96 \cdot \sqrt{7,61796875}$

$213,75 \pm 5,4097$ R

$[208,3403 ; 219,1597]$ utgör ett 95%:igt konfidensintervall för det ~~genomsnittliga~~ kaloriinnehållet enligt kvotskattning.

b) Då energiinnehållet betecknas som antal kalorier i en banan kan energiinnehållet ~~skattas~~ som:

$\hat{\mu}_{kvot} = \underline{213,75}$ kalorier per banan enligt kvotskattningen. R

Detta motsvarar 95 kalorier per 100g. R

SU, STATISTIK

Skrivsal: VÅ Anonymkod: 0031-DZT Blad nr: 3

5) 20p

5 Data hämtas från uppgift 4 på föregående sida.

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

$$b = \frac{\sum x_i z_i - \frac{1}{n}(\sum x_i)(\sum z_i)}{\sum z_i^2 - \frac{1}{n}(\sum z_i)^2} = \frac{229975 - \frac{1}{5}(1045)(1100)}{242250 - \frac{1}{5}1100^2} = \frac{75}{250} = 0,3$$

$$\bar{x} = 1045/5 = 209$$

$$\mu_z = 225$$

$$\bar{z} = 1100/5 = 220$$

R

R

$$\hat{\mu}_{reg} = 209 + 0,3(225 - 220) = 210,5$$

$$\hat{V}(\hat{\mu}_{reg}) = \left(\frac{1000-5}{5 \cdot 1000}\right)\left(\frac{1}{5-2}\right)\left[\left(218475 - \frac{1}{5} \cdot 1045^2\right) - 0,3^2\left(242250 - \frac{1}{5} \cdot 1100^2\right)\right]$$

$$= \left(\frac{995}{5000}\right)\left(\frac{1}{3}\right)\left[70 - 22,5\right] = 3,15083333$$

R

$$\sqrt{\hat{V}(\hat{\mu}_{reg})} \approx 1,7751$$

95%:igt ki ges av

$$\hat{\mu}_{reg} \pm 1,96 \cdot \sqrt{\hat{V}(\hat{\mu}_{reg})}$$

R

$$210,5 \pm 3,4792$$

$$[207,0208 ; 213,9792]$$

R

Utgör ett 95%:igt konfidensintervall för det genomsnittliga energiinnehållet för hela partiet bananer, enligt regressionsskattning.

R

1

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 17/8-2018

Sal: Värtasalen

Tenta: Undersökningsmetodik

Kurs: Regressionsanalys och undersökningsmetodik

ANONYMKOD: 0027-HHP

☒ 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. 0p	20p	20p	20p	20p					

SC

POÄNG	BETYG	Lärarens sign.
80p	B	RC

SU, STATISTIK

Skrivsal: Vårdasalen **Anonymkod:** 0027-HHP **Blad nr:** 1

1) Op

1. $x_1 = 26$, $x_2 = 24$, $x_3 = 28$, $x_4 = 30$, $x_5 = 26$

$N = 5$ varav $n = 3$ element ska väljas ut.

a) 3 bland 5 element kan väljas ut på: $\binom{N}{n} = \binom{5}{3} = \frac{5!}{3!(5-3)!} =$

$\frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1 \cdot 2 \cdot 1} = \frac{20}{2} = 10$ olika sätt

STORT FEL (SE FACIT)

$n = 3$

$n \neq 2$

Urval	$\bar{x}$	s^2
26, 24	25	1+1=2
26, 28	27	1+1=2
26, 30	28	4+4=8
26, 26	26	0+0=0
24, 28	26	4+4=8
24, 30	27	9+9=18
24, 26	25	1+1=2
28, 30	29	1+1=2
28, 26	27	1+1=2
30, 26	28	4+4=8
Summa	268	52
Medelvärde	26,8	5,2

$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

$P(s^2) = \frac{\text{antal gynnsamma}}{\text{antal möjliga}}$

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

$\bar{s}^2 = \frac{\sum s_i^2}{n}$

Samplingsfördelning för s^2:

s^2	0	2	8	18	
$P(s^2)$	0,1	0,5	0,3	0,1	1

F

b) $E(s^2) = \sum P(s^2) \cdot s^2 = 0{,}1 \cdot 0 + 0{,}5 \cdot 2 + 0{,}3 \cdot 8$
$+ 0{,}1 \cdot 18 = \boxed{5{,}2}$

F

c) $V(s^2) = \sum(s^2 - \bar{s}^2)^2 \cdot P(s^2) =$

$(0-5{,}2)^2 \cdot 0{,}1 + (2-5{,}2)^2 \cdot 0{,}5 + (8-5{,}2)^2 \cdot 0{,}3 +$

$(18-5{,}2)^2 \cdot 0{,}1 = 2{,}704 + 5{,}12 + 2{,}352 +$
$16{,}384 = \boxed{26{,}56}$

F

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: 0027-HHP Blad nr: 2

2) 20p

2. $N=2000$, $n=300$

Nr	Stratum	N_i	n_i	p_i	$W_i = \frac{N_i}{N}$
1	17-25 år	1000	150	0,90	0,50
2	26-30 år	800	100	0,80	0,40
3	31 år eller högre	200	50	0,60	0,10
	Summa	2000	300		1

Vi söker 95% KI kring p_{st} som fås genom:

$$p_{st} \pm z_{\alpha/2}\sqrt{\hat{V}(p_{st})}$$

R

(1) Vi börjar med att skatta p_{st}:

$$p_{st} = \sum_{i=1}^{L} W_i p_i = 0,5 \cdot 0,9 + 0,4 \cdot 0,8 + 0,1 \cdot 0,6 = 0,83$$

R

(2) Vi beräknar $\hat{V}(p_{st})$:

$$\hat{V}(p_{st}) = \sum_{i=1}^{L} W_i^2 \cdot \hat{V}(p_i) \longrightarrow \hat{V}(p_i) = \frac{p_i(1-p_i)}{n_i - 1}\left(1 - \frac{n_i}{N_i}\right)$$

$$\hat{V}(p_1) = \frac{0,90(1-0,90)}{150-1}\left(1 - \frac{150}{1000}\right) = 0,0005134228188$$

$$\hat{V}(p_2) = \frac{0,80(1-0,80)}{100-1}\left(1 - \frac{100}{800}\right) = 0,0014141414$$

R

Vänd →

$$\hat{V}(p_3) = \frac{0{,}60(1-0{,}60)}{50-1}\left(1-\frac{50}{200}\right) = 0{,}0036734694$$

$$\hat{V}(p_{st}) = 0{,}50^2 \cdot 0{,}0005134228188 + 0{,}40^2 \cdot 0{,}0014141414 + 0{,}10^2 \cdot 0{,}0036734694$$

$$= 0{,}0001283557047 + 0{,}0002262626 24 + 0{,}000036734694 =$$

$$= 0{,}0003913530227$$

R

③ $z_{\alpha/2} = 1{,}96$ (från tabell)

95% igt KI fås såledles av:

$$p_{st} \pm z_{\alpha/2}\sqrt{\hat{V}(p_{st})} \longrightarrow 0{,}83 \pm 1{,}96\sqrt{0{,}0003913530227} \longrightarrow$$

$$0{,}83 \pm 0{,}0387739832 \longrightarrow \{0{,}7912260168 \leq p_{st} \leq 0{,}8687739832\}$$

$$\longrightarrow \{0{,}791;\ 0{,}869\}$$

R

R

R

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: 0027-HHP Blad nr: 3

3) 20p

3. $n=150$

Stratum	N_i	n_i	$\bar{x}_i$	s_i	$W_i = \frac{N_i}{N}$
Stora	50	30	34	9	50/350
Mellanstora	100	60	23	4	100/350
Små	200	60	18	3	200/350
Summa	350	150			1

Vi söker 95% kring $\bar{x}_{st}$ som ges av:

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

① Vi börjar med att skatta $\bar{x}_{st}$:

$$\bar{x}_{st} = \sum_{i=1}^{L} W_i \bar{x}_i = \left(\frac{50}{350}\right)\cdot 34 + \left(\frac{100}{350}\right)\cdot 23 + \left(\frac{200}{350}\right)\cdot 18 = 4,857142857 +$$

$$6,571428571 + 10,28571429 = 21,71428571$$

R

② Vi beräknar sedan $\hat{V}(\bar{x}_{st})$ som fås av:

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

$$\hat{V}(\bar{x}_1) = \frac{9^2}{30}\left(1 - \frac{30}{50}\right) = 1,08$$

$$\hat{V}(\bar{x}_2) = \frac{4^2}{60}\left(1 - \frac{60}{100}\right) = 0,1066666667$$

Vänd →

$$\hat{V}(\bar{x}_3) = \frac{3^2}{60}\left(1 - \frac{60}{200}\right) = 0{,}105$$

$$\hat{V}(\bar{x}_{st}) = \left(\frac{50}{350}\right)^2 \cdot 1{,}08 + \left(\frac{100}{350}\right)^2 \cdot 0{,}1066666667 + \left(\frac{200}{350}\right)^2 \cdot 0{,}105 =$$

R

$$0{,}0220408163 + 0{,}0087074830 + 0{,}0342857143 = 0{,}0650340136$$

(3) $z_{\alpha/2} = 1{,}96$ (från tabell)

95 % KI kring $\bar{x}_{st}$:

$$21{,}71428571 \pm 1{,}96\sqrt{0{,}0650340136} \longrightarrow 21{,}71428571 \pm 0{,}4998346393$$

$$\longrightarrow \left\{21{,}21445107 \leq \cancel{\bar{x}}^{\mu}_{st} \leq 22{,}21412035\right\} \longrightarrow \left\{21{,}214 \, ; \, 22{,}214\right\}$$

R

R

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: 0027-HHP Blad nr: 4

4) 20p

4.

x_i Antal kalorier	z_i Vikt i gram	x_i^2	z_i^2	$x_i z_i$
210	215	44100	46225	45150
215	230	46225	52900	49450
205	225	42025	50625	46125
210	220	44100	48400	46200
205	210	42025	44100	43050
1045	1100	218475	242250	229975

$N = 1000$ x_i = antal kalorier = undersökningsvariabel

$n = 5$ z_i = vikt i gram = hjälpvariabel

$T_z = 225\,000$

a) Kvotskattning från formelblad:

$$\hat{T}_{kvot} = \hat{R} \cdot T_z = \frac{\sum x_i}{\sum z_i} \cdot T_z \longrightarrow \hat{\mu}_{kvot} = \frac{\hat{T}_{kvot}}{N}$$

R

$$\hat{T}_{kvot} = \frac{1045}{1100} \cdot 225\,000 = 213\,750 \longrightarrow \hat{\mu}_{kvot} = \frac{213\,750}{1000} = 213,75$$

95% KI kring $\hat{\mu}_{kvot}$ fås av:

$$\hat{\mu}_{kvot} \pm z_{\alpha/2} \sqrt{\hat{V}(\hat{\mu}_{kvot})} \longrightarrow \hat{V}(\hat{\mu}_{kvot}) = \left(\frac{N-n}{nN}\right) \frac{\sum_{i=1}^{n} \left(x_i - \hat{R} z_i\right)^2}{n-1}$$

vänd →

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

$$\hat{R} = \frac{\sum x_i}{\sum z_i} = \frac{1045}{1100} = 0{,}95$$

$$\sum x_i^2 + \hat{R}^2 \sum z_i^2 - 2\hat{R}\sum x_i z_i = 218475 + \underbrace{0{,}95^2 \cdot 242250}_{218630{,}625} - \underbrace{2 \cdot 0{,}95 \cdot 229975}_{436952{,}5}$$

$$= 153{,}125$$

$$\hat{V}(\hat{\mu}_{kvot}) = \left(\frac{1000-5}{5 \cdot 1000}\right) \cdot \frac{153{,}125}{5-1} = 7{,}61796875$$

R

95% KI:

$$213{,}75 \pm 1{,}96\sqrt{7{,}61796875} \longrightarrow 213{,}75 \pm 5{,}409730937 \longrightarrow$$

R

$$\left\{208{,}3402691 \leq \mu_{kvot} \leq 219{,}1597309\right\} \longrightarrow \left\{208{,}34 \, ; \, 219{,}16\right\}$$

R

b) Kalle kommer leverera $\hat{\mu}_{kvot} = 213{,}75$ vilket betyder att en chiquita-banan i genomsnitt innehåller 213,75 kalorier.

R

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: 0027-HHP Blad nr: 5

5. Regressionsskattning från formelblad: 5) 20p

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

Vi använder oss av tabellen i föregående blad men lägger till uträkning av $\bar{x}$, $\bar{z}$, μ_z samt b:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1045}{5} = 209 \quad , \quad \bar{z} = \frac{\sum z_i}{n} = \frac{1100}{5} = 220$$

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

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

R

$$\frac{229975 - \frac{1100 \cdot 1045}{5}}{242250 - \frac{1100^2}{5}} = \frac{75}{250} = 0{,}3$$ R

Nu har vi allt vi behöver för att räkna ut $\hat{\mu}_{reg}$:

$$\hat{\mu}_{reg} = 209 + 0{,}3(225 - 220) = 210{,}5$$ R R

95% KI fås genom:

$$\hat{\mu}_{reg} \pm z_{\alpha/2}\sqrt{\hat{V}(\hat{\mu}_{reg})}$$ vänd →

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

$$\left(\frac{N-n}{nN}\right)\left(\frac{1}{n-2}\right)\left[\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right) - b^2\left(\sum z_i^2 - \frac{(\sum z_i)^2}{n}\right)\right] =$$

$$\left(\frac{1000-5}{5\cdot 1000}\right)\left(\frac{1}{5-2}\right)\left[\left(218475 - \frac{1045^2}{5}\right) - 0,3^2\left(242250 - \frac{1100^2}{5}\right)\right] =$$ R

$$0,0663333333\,(70 - 22,5) = 3,150833332\,,$$ R

95% KI kring $\hat{\mu}_{reg}$:

R

$$210,5 \pm 1,96\sqrt{3,150833332} \rightarrow 210,5 \pm 3,479115027 \rightarrow$$ R

$$\{207,020885 \leq \hat{\mu}_{reg} \leq 213,979115\} \rightarrow \{207,02\ ;\ 213,18\}$$