

Försättsblad till skriftlig tentamen vid Linköpings universitet



Datum för tentamen	2019-10-10
Sal (1)	<u>TER3(17)</u>
Tid	8-12
Utb. kod	732G38
Modul	TENA
Utb. kodnamn/benämning Modulnamn/benämning	Surveymetodik med teori Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	4
Jour/Kursansvarig Ange vem som besöker salen	Isak Hietala
Telefon under skrivtiden	013-281970
Besöker salen ca klockan	Endast telefonjour
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	Erika Larsson 013-28 18 68 erika.larsson@liu.se
Tillåtna hjälpmedel	- Miniräknare av valfri modell; - Kursbok Sharon L. Lohr SAMPLING: Design and Analysis (alla upplagor tillåtna - får innehålla markeringar, understrykningar och flärpar, men inte anteckningar) - A4-blad med vad som helst skrivet på fram- och baksida (för hand/skrivet på dator/kopierat etc. - inga begränsningar)
Övrigt	
Antal exemplar i påsen	

Tentamen 732G38

Linköpings universitet
Institutionen för datavetenskap, IDA
Avdelningen för Statistik och maskininläring, STIMA

Kurskod och namn:	732G38, Surveymetodik med teori
Datum och tid:	2019-10-10, 8-12
Jourhavande lärare:	Isak Hietala
Tillåtna hjälpmedel:	- Räknedosa av valfri modell; - kursboken <i>Sampling: Design and Analysis</i> av Lohr som får innehålla markeringar, överstrykningar samt små flärpar med anteckningar; - ett dubbelsidigt A4 med egna anteckningar
Betygsgränser:	Tentamen omfattar totalt 20 poäng; betyg G fås vid minst 12 poäng; betyg VG vid minst 16 poäng
Antalet uppgifter:	4 stycken
Annan information:	Siffrorna i uppgifterna är påhittade. Saknas någon siffra för att kunna lösa uppgiften? Skriv då tydligt ut att du saknar denna information, anta ett godtyckligt värde och lös uppgiften med detta antagande.

Redovisa, tolka och motivera tydligt alla dina lösningar!

Uppgift 1 (1.5p)

Motivera vilken av de tre bortfallsantaganden som gäller för följande tillfällen: (0.5p per deluppgift)

- En låda med postenkäter som ska skickas till norra Sverige försvinner vid transporten
- I en hälsoundersökning svarar en mindre andel män på frågan om deras längd
- I en partisympatiundersökning saknas svar från personer som vi *vet* stödjer Miljöpartiet

Uppgift 2 (3.5p)

En egenskap har uppmätts bland 2500 individer från två olika grupper och sammanställts i tabell 1.

Tabell 1: Beskrivande statistik över urvalet

	\bar{y}_h	s^2	n	N
Grupp 1	35.75	4.66	2500	14062
Grupp 2	24.62	4.72	2500	119021

- Beräkna ett 95-procentigt konfidensintervall för medelvärdet för båda grupperna. (2p)
- Vilken allokeringmetod användes i undersökningen? (0.5p)
- Beräkna nya urvalsstorlekar om proportionell allokering skulle ha använts. (1p)

Uppgift 3 (10p)

Under en vinter mäter en meteorolog slumpmässigt 8 dagars medeltemperatur och snödjupet. Den genomsnittliga temperaturen för hela vintern (bestående av 90 dagar) är -10.5 grader. En sammanställning av urvalet hittas i tabell 2. Alla konfidensintervall ska beräknas med 95 procent konfidens.

Tabell 2: Beskrivande statistik över urvalet

Temperatur ($^{\circ}\text{C}$)	Snödjup (cm)
-7.7	85
-12.5	94
-26.6	114
-19.5	114
-0.5	59
-12.6	99
-2.9	71
-7.2	86

- Skatta det genomsnittliga snödjupet med ett 95-procentigt konfidensintervall baserat på OSU. (2p)
- Skatta det genomsnittliga snödjupet med ett 95-procentigt konfidensintervall baserat på kvotskattning. (2p)
- Skatta det genomsnittliga snödjupet med ett 95-procentigt konfidensintervall baserat på regressionskattning givet modellen i tabell 3. (2p)

Tabell 3: Regressionsutskrift

	Estimate	Std. Error	t value	$\text{Pr}(> t)$
(Intercept)	66.4401	3.9013	17.03	0.0000
temp	-2.1283	0.2828	-7.53	0.0003

- Beräkna designeffekten för skattningarna från a)-c). Vilken metod är att föredra? Motivera! (4p)

Uppgift 4 (5p)

Ett slumpmässigt urval om 5 byggnaders tvättmaskiner i ett bostadsområde bestående av 20 byggnader gjordes. Tiden för en *normaltvätt* uppmättes bland alla maskiner i varje utvalda byggnad och sammanställs i tabell 4.

Tabell 4: Antalet timmar en tvätt tar

	M_i	m_i	\bar{y}_i	s_i^2
Byggnad A	2	2	2.17	0.83
Byggnad B	4	4	2.50	3.85
Byggnad C	7	7	1.11	4.04
Byggnad D	6	6	1.29	3.67
Byggnad E	7	7	2.85	1.80

- Skatta det totala antalet timmar en tvätt tar för varje maskin bland alla byggnader i området med ett 95-procentigt konfidensintervall. (2p)
- Beskriv den största skillnaden mellan enstegsklusterurval med lika och olika klusterstorlekar. (1p)
- Beskriv vad de olika varianskomponenterna i variansuträkningen för tvåstegsurval representerar, formel nedan, och beskriv vilken utav dem som är viktigast att hantera när antalet kluster är mycket stort. (2p)

$$N^2 \left(1 - \frac{n}{N}\right) \frac{S_t^2}{n} + \frac{N}{n} \sum_{i=1}^N \left(1 - \frac{m_i}{M_i}\right) M_i^2 \frac{S_i^2}{m_i}$$