

Tentamen

Linköpings Universitet, Institutionen för datavetenskap, STIMA

Kurskod och namn: 732G38 Surveymetodik med teori

Datum och tid: 2018-10-11, 8-12

Jourhavande lärare: Bertil Wegmann

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).

Betygsgränser: Tentamen omfattar totalt 20p. Godkänt från 12p, väl godkänt från 16p.

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! Lycka till!

Uppgift 1 (6p)

Motala kommun vill undersöka hur ofta invånarna i kommunen motionerar varje vecka. Man vet sedan tidigare att motionsvanor varierar mellan 3 områden som utgör hela kommunen. Därför tar man ett stratifierat urval baserat på område. Detta gav följande information:

Område	Stratumstorlek	Stickprovstorlek	Genomsnittligt antal motionsdagar per respondent	Kostnad per respondent	Standardavvikelse
1	12000	316	1.4	20	0.2
2	22000	1158	2.1	20	0.4
3	10000	526	2.5	20	0.4
Samtliga	44000	2000			0.5

- Visa att Neyman allokering har använts samt beräkna urvalsstorlekarna i strata om proportionell allokering hade använts. (2p)
- Skatta genomsnittligt antal motionsdagar per invånare i Motala kommun med 95 % konfidensintervall baserat på det stratifierade urvalet. (2p)
- Beräkna designeffekten för det stratifierade urvalet, tolka designeffekten i ord och avgör från den beräknade designeffekten om det stratifierade urvalet ger mera precisa skattningar än OSU. (2p)

Uppgift 2 (8p)

En lantbruksorganisation vill undersöka hur stor skörd en viss gröda ger i ett land som är uppdelat i 25 ungefär lika stora produktionsregioner. Man känner till sedan tidigare att högre halter av en luftförorening A minskar skördemängden. Den genomsnittliga halten av luftförorening A i landet är 5 enheter. Bland alla produktionsregioner tog man ett slumpmässigt urval om 5 stycken och noterade skördemängden och halten av luftförorening A. Detta gav följande tabell:

Luftförorening A per region i urvalet	Total skördemängd per region i urvalet (i kiloton)
4.2	9.2
5.1	8.5
5.8	8.0
3.5	9.9
5.2	8.3

- Skatta genomsnittlig skördemängd per region i landet med 95 % konfidensintervall baserat på OSU. (2p)
- Skatta genomsnittlig skördemängd per region i landet med 95 % konfidensintervall baserat på kvotskattning. (2p)
- Skatta genomsnittlig skördemängd per region i landet med 95 % konfidensintervall baserat på regressionskattning, givet följande skattade regressionsmodell: (2p)

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	12.78	0.254	50.31	0.000
Luftförorening A	-0.84	0.053	-15.97	0.001

- Beräkna designeffekten för regressionskattningen. Är den att föredra framför skattningen i a) som baseras på OSU? Motivera. (1p)
- Motivera varför regressionskattningen i c) är rimligare att använda sig av jämfört med kvotskattningen i b). (1p)

Uppgift 3 (4p)

En livsmedelskedja som är verksam i 30 stycken kommuner vill uppskatta det totala antalet påsar som sålts av en ny godissort. Därför tog man ett slumpmässigt urval av 5 kommuner och noterade för varje kommun antalet livsmedelsbutiker som sålt godissorten och totala antalet sålda påsar, vilket gav följande tabell:

Kommun	Antalet livsmedelsbutiker som sålt godissorten	Antalet sålda påsar
1	12	840
2	5	360
3	8	520
4	3	260
5	7	540

- a) Skatta det totala antalet sålda påsar av den nya godissorten för alla 30 kommuner med 95 % konfidensintervall. (2p)

I en mer omfattande undersökning vill Livsmedelsverket uppskatta totala antalet sålda chokladkakor från totalt 5850 butiker i alla Sveriges 290 kommuner under en viss period. Därför tog Livsmedelsverket ett slumpmässigt urval av 3 kommuner och sedan ytterligare ett slumpmässigt urval av butiker där chokladkakor säljs inom respektive utvald kommun. Detta gav följande tabell:

Kommun	Totala antalet butiker, M_i	Antal butiker i urvalet, m_i	Genomsnittligt antal sålda chokladkakor per butik i urvalet, \bar{y}_i	Urvalsvarians, s_i^2
1	32	8	20.25	16
2	12	3	18	10
3	16	4	24	20

- b) Skatta det totala antalet sålda chokladkakor i alla Sveriges kommuner under perioden med 95 % konfidensintervall. (2p)

Uppgift 4 (2p)

I en by vill man utföra en opinionsundersökning bland invånarna om hur många kronor dem spenderar på fritidsaktiviteter varje månad. Man tog därför ett slumpmässigt urval om 60 invånare av totalt 300. Resultatet av undersökningen redovisas nedan för byns nordliga och sydliga del:

Region	Antal invånare i urvalet	Totala antalet kronor spenderat av invånarna i urvalet	Urvalsvarians
Nordlig	42	44100	108900
Sydlig	18	21600	144400

Skatta genomsnittlig spending per invånare i byns sydliga del per månad med 95 % konfidensintervall.