

Tentamen

Institutionen för datavetenskap, STIMA



Kurskod och namn:	732G38 , Surveymetodik med teori
Datum och tid:	2018-03-26, 8-12
Jourhavande lärare:	Annika Tillander
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 20 poäng, G från 12p, VG från 16p
Övrigt:	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. Lösningförslag till denna tentamen läggs upp på kursens LISAM-sida den 27/3.

Redovisa, tolka och motivera tydligt alla dina lösningar!

Lycka till!

Uppgift 1 (6p)

Ett företag med total 1475 anställda önskar skatta totalt antal sjukdagar. Eftersom antal sjukdagar (y) varierar med inkomst görs ett stratifierat urval baserat på inkomstgrupper.

Strata	N_h	n_h	\bar{y}_h	s_h
< 25000 kr	600	60	3.38	3.5
25000 – 35000 kr	456	46	0.48	0.64
> 35000 – 45000 kr	270	27	2.01	1.91
> 45000 kr	149	15	0.37	0.61
Samtliga	1475	148	1.37	2.22

Tabell 1: Genomsnittligt antal sjukdagar per stratum.

- Skatta totalt antal sjukdagar med 95% konfidensintervall baserat på det stratifierade urvalet. 2p
- Nu användes proportionell allokering, hur skulle urvalet på 148 fördelats om istället Neyman allokering används (utgå från s_h i tabellen)? 2p
- Om skattningen istället skulle baseras på OSU, hur stort urval skulle krävas för att uppnå samma precision (e) som i a)? 2p

Uppgift 2 (8p)

I ett villaområde med 25 villor värderades samtliga objekt 2012 och det genomsnittliga värdet var 2.1 miljoner kronor. Sedan 2017 gjordes ett slumpmässigt urval på 4 villor för värdering, se tabell 2.

	pris2012	pris2017
1	1.7	4.0
2	1.8	4.2
3	2.5	4.6
4	2.8	3.8

Tabell 2: Värdering 2012 och 2017 i miljoner kronor för de 4 villorna i urvalet

- Skatta det genomsnittliga värdet 2017 med 95% konfidensintervall baserat på OSU. 2p
- Skatta det genomsnittliga värdet 2017 med 95% konfidensintervall med kvotskattning. 2p
- Skatta det genomsnittliga värdet 2017 med 95% konfidensintervall med regressionskattning givet modellen: 2p

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.10	0.53	2.1	0.05
pris2012	1.25	0.24	5.2	0.00

- Beräkna designeffekten för kvotskattning och regressionskattning. Det sanna genomsnittliga värdet 2017 är 3.7 miljoner kronor, utifrån detta och designeffekterna vilken skattning är att föredra? 2p

Uppgift 3 (2p)

Ett företag med 20 industrier är intresserade av totala tiden driftstopp (y). Först görs ett slumpmässigt urval på 3 industrier och sedan ytterligare ett slumpmässigt urval av maskiner. Skatta totala tiden driftstopp med 95% konfidensintervall 2p

Industri	M_i	m_i	Driftstopp i timmar	\bar{y}_i	s_i^2
A	45	9	5,6,4,11,12,0,1,8,4	5.67	16.75
B	42	8	3,7,6,7,4,3,2,8	5.00	5.14
C	40	8	6,4,7,3,9,1,4,5	4.88	6.13

Tabell 3: Driftstopp per industri i urvalet

Uppgift 4 (4p)

En förening som finns på tre orter ska undersöka inställning till höjning av medlemsavgiften. För att effektivisera beslutas att göra ett slumpmässigt urval av två orter.

	Ort	M_i
1	Linköping	139
2	Motala	136
3	Norrköping	144
Totalt		419

Tabell 4: Antal medlemmar per ort

- a) I det slumpmässiga urvalet fås Linköping där 136 är positiva till en höjning och Motala där 88 är positiva. Skatta hur många totalt i föreningen är positiva höjning om urvalet gjorts med återläggning (*pps* Hansen-Hurwitz) respektive utan återläggning (*π ps* Horwitz-Thompson). 3p
- b) Visa att Hansen-Hurwitz variansskattningen för totalen, $\hat{V}(\hat{t}_\psi) = \frac{1}{n} \frac{1}{n-1} \sum_{i \in \mathcal{R}} \left(\frac{t_i}{\psi_i} - \hat{t}_\psi \right)^2$, då $\psi = \frac{1}{N}$ förenklas till: 1p

$$\hat{V}(\hat{t}_\psi) = \frac{N^2}{n} \frac{1}{n-1} \sum_{i \in \mathcal{R}} (t_i - \bar{t})^2$$