

STOCKHOLMS UNIVERSITET

Statistiska institutionen

Regressionsanalys och undersökningsmetodik, höstterminen 2016

Jörgen Säve-Söderbergh

Tentamen i Undersökningsmetodik

Datum	2017-02-08
Tid:	16.00-21.00
Ansvarig Lärare:	Jörgen Säve-Söderbergh
Antal frågor:	5
Maxpoäng:	50
Hjälpmedel:	1) Språklexikon 2) Kalkylator utan lagrade formler eller lagrad text
Tentamensgenomgång	Måndag 27 februari kl. 15.00 i sal B 419

Anvisningar

Redovisa dina lösningar i en form som gör det lätt att följa tankegången. Motivera alla väsentliga steg i lösningen. Ange alla antaganden och förutsättningar som du utnyttjar. Skriv endast på en sida av arket. Börja varje ny uppgift på nytt ark.

Lycka Till!

1. Priset på sockerbetor beror på sockerhalten. Att fastslå sockerhalten i ett stort parti sockerbetor är inte möjligt. I syfte att skatta denna parameter tar vi ett stickprov på tio sockerbetor. Med hjälp av en biokemist mäter vi vikten och sockerhalten bland de uttagna sockerbetorna. Därefter väger vi partiet sockerbetor och finner att det väger totalt 6000 kilo.

Sockerbeta	Sockerhalt (kg)	Vikt (kg)
1	0.2198	1.4
2	0.1716	1.2
3	0.1139	0.9
4	0.1452	1.1
5	0.1843	1.2
6	0.2281	1.6
7	0.1932	1.1
8	0.1382	0.9
9	0.1926	1.2
10	0.2311	1.4

- a) Skatta den totala sockerhalten i hela partiet av sockerbetor med hjälp av en kvotskattning. (7 p)
- b) Antag att vi får reda på att partiet består av 4800 sockerbetor, dvs $N = 4800$. Använd denna information för att skatta den totala sockerhalten i hela partiet sockerbetor. (3 p)
2. Det finns tjugo studentföreningar vid Örebro universitet. En tjänsteman på rektorns kansli läste en artikel om studentlivet i Uppsala och beslöt att studera hur livaktiga studentföreningarna i Örebro tedde sig. Tjänstemannen var gammal student från statistiska institutionen och undrade hur många statistikstudenter som var aktiva inom studentföreningarna. Fyra föreningar utvaldes slumpmässigt och följande uppgifter insamlades:

Tillgångar (kr)	Medlemmar	Medlemmar som studerar statistik
16000	34	4
96000	16	7
3600	172	24
795	119	2

- a) Skatta de totala tillgångarna som studentföreningarna förfogar över. (5 p)
- b) Skatta antalet statistiker bland medlemmarna inom studentföreningarna. (5 p)
3. Det finns olika möjligheter att se på tv med hjälp av en dator. En statlig myndighet önskade undersöka hur stor andel av invånarna i en stadsdel i Örebro som funderade på att skaffa den utrustning som krävs för att kunna se på tv via nätet och en dator. De beslutade om en stickprovsundersökning om $n = 700$ personer. I den aktuella stadsdelen bor $N = 5000$ personer. Därtill stratifierades populationen med avseende på ålder, eftersom tjänstemännen misstänkte att det fanns ett starkt samband mellan ålder och inställningen i denna fråga. OSU utan återläggning användes för att hämta stickprov ur respektive strata. Proportionell allokering nyttjades. Resultatet av undersökningen var som följer:

Stratum	N_i	n_i	p_i
15-29	900	126	0.88
30-49	1000	140	0.47
50 eller högre	3100	434	0.06

Beräkna ett 95%-igt konfidensintervall för andelen invånare som funderar på att se på tv via datorn. (10 p)

4. En population har följande värden:

561 401 931 120 801 739 661 118 942 984 710 430
 674 190 208 491 561 195 420 205 602 721 98 1054
 679 430 206 218 893 439 200 628 104 105 306 205
 1843 107 1154 721 502 200 499 1295 34 2054 183 228

Drag ett stickprov om $n = 4$ individer genom systematiskt urval. Använd följande utdrag ur en slumpantalstabell:

45740 41807 65651 33302 07051 93623

För full poäng måste du redogöra för hur du har använt slumpantalstabellen. (10 p)

5. Ett undersökningsföretag har fått ett uppdrag att uppskatta hur stor andel av svenska folket som spelar trumpet. Stickprovet ska göras med obundet slumpmässigt urval. Hur många individer måste minst intervjuas för att felmarginalen på ett 90%-igt konfidensintervall för den sökta andelen i populationen högst ska vara 10%? (10 p)

TABELL 1. 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 $P(Z \leq -z_\alpha)$.

$$P(Z > z_\alpha) = \alpha$$



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



Formelsamling undersökningsmetodik

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

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

Beräkning av stickprovsstorlek:

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

Stratifierat urval:

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

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:

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

Punktskattning	Varians	Variansskattning	Varians	Variansskattning
OSU	m. å.	m. å.	u. å.	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{t}	$N^2 \cdot \frac{\sigma^2}{n}$	$N^2 \cdot \frac{s^2}{n}$	$N^2 \cdot \frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right)$	$N^2 \cdot \frac{s^2}{n} \left(1 - \frac{n}{N} \right)$
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}	$N^2 \cdot \frac{P(1-P)}{n}$	$N^2 \cdot \frac{p(1-p)}{n-1}$	$N^2 \cdot \frac{P(1-P)}{n} \left(\frac{N-n}{N-1} \right)$	$N^2 \cdot \frac{p(1-p)}{n-1} \left(1 - \frac{n}{N} \right)$

