

STOCKHOLMS UNIVERSITET
Statistiska institutionen
Hans Nyquist

TENTAMEN I REGRESSIONSANALYS OCH UNDERSÖKNINGSMETODIK

DELKURS 2, UNDERSÖKNINGSMETODIK

2021-02-12

Skrivtid: 13.00-18.00, sista inlämningstid 19.00

Godkända hjälpmedel: Miniräknare, dator, kurslitteratur, föreläsninganteckningar och språklexikon, formelsamling (från Athena) och statistiska tabeller (från Athena)

Obs! Det är inte tillåtet att ta hjälp av andra personer under skrivningen

Tentamen består av fem uppgifter. För full poäng på en uppgift krävs tydliga, utförliga och väl motiverade lösningar.

Resultatet meddelas senast den 1 mars.

Kortfattade svar läggs ut strax efter tentamen på Athena

Kontakt med examinator under tentamen: För eventuella frågor om innehållet i tentan kan du kontakta examinator under pågående tentamen på mail: Hans.Nyquist@stat.su.se. Inkommande mailfrågor besvaras kontinuerligt under tentans gång. Om examinator behöver informera om någonting under tentan görs detta till din registrerade mailadress. Kontrollera därför din mail under tentans gång.

Observera att praktisk hjälp endast finns tillgänglig under tentans första timme på mailadressen expedition@stat.su.se. Läs noggrant bifogad instruktion för inlämning av tentan. Där finns all nödvändig information om inlämning, anonymkod etc. Om du trots instruktionerna skulle få problem att lämna in tentan, maila istället tentan till tenta@stat.su.se. Detta görs dock bara i undantagsfall.

Uppgift 1. (20 poäng)

I en kommun finns 900 grundskoleelever. För att undersöka om eleverna är positivt inställda till skolmatens salladsbord genomfördes ett stratifierat urval. I nedanstående tabell visas stratumstorlekar, urvalsstorlekar och observerat antal positivt inställda till skolans salladsbord.

Stratum	Antal elever	Urvalsstorlek	Observerat antal positivt inställda
Lågstadiet	350	60	48
Mellanstadiet	300	50	25
Högstadiet	250	40	8

- a) Uppskatta proportionen elever i kommunen som är positivt inställda till skolans salladsbord och ange ett 95 procentigt konfidensintervall. Ändlighetskorrektur ska användas.
- b) Man planerar att göra om undersökningen nästa år. Utgå från ovanstående data och föreslå en urvalsallokering som kan minimera den statistiska felmarginalen med samma totala urvalsstorlek.

Uppgift 2. (20 poäng)

Medelinkomsten per hushåll ska uppskattas i en kommun med 4260 hushåll. För detta har man gjort ett obundet slumpmässigt urval utan återläggning om 300 hushåll. Resultatet från undersökningen gav att $\sum y_i = 19575000$ och $\sum y_i^2 = 1,28 \cdot 10^{12}$, där y_i är observationer på hushållens inkomst.

- a) Uppskatta hushållens medelinkomst i kommunen och ange ett 95 procentigt konfidensintervall. Ändlighetskorrektur ska användas.
- b) Man planerar att göra om undersökningen nästa år. Utgå från ovanstående data för att beräkna erforderlig urvalsstorlek för att bredden av ett 90 procentigt konfidensintervall inte ska överstiga 10000 kr.

Uppgift 3. (20 poäng)

Svara kortfattat men uttömmande på var och en av följande deluppgifter. Komplettera gärna med bilder och skisser.

- a) Beskriv vilka åtgärder man kan vidta före, under och efter datainsamlingen för att minska effekterna av bortfall vid en urvalsundersökning.
- b) Vilka av följande urvalstekniker är slumpmässiga urval: (i) snöbollsurval, (ii) systematiskt urval, (iii) stratifierat urval, (iv) tvåstegsurval och (v) bekvämlighetsurval
- c) Förklara begreppet teleskopeffekt och vilka fördelar respektive nackdelar det för med sig vid en urvalsundersökning.
- d) Förklara begreppen över- och undertäckning och vilka fördelar respektive nackdelar de för med sig vid en urvalsundersökning.

Uppgift 4. (20 poäng)

En reklamfirma vill studera effekten av en reklamkampanj för en viss produkt. För att uppskatta produktens totala försäljningsvärde hos en butikskedja under det kvartal som reklamkampanjen genomfördes, τ_y , drog man ett OSU ur om $n = 20$ butiker bland butikskedjans $N = 452$ butiker. För varje butik observerades försäljningsvärdet i kr under kvartalet innan reklamkampanjen, x_i , och försäljningsvärdet i kr under kvartalet då reklamkampanjen genomförs, y_i . Man vet dessutom att totala försäljningsvärdet av produkten kvartalet före reklamkampanjen är $\tau_x = 2162560$ kr. De erhållna observationerna visas i tabellen nedan.

Butik i	Försäljningsvärde före kampanjen, x_i	Försäljningsvärde efter kampanjen y_i
1	2080	2390
2	4000	4280
3	4400	4720
4	2590	2760
5	3510	3630
6	8800	9420
7	2730	2940
8	4870	5140
9	1830	1950
10	8630	8970
11	5990	6260
12	5100	5380
13	8280	8880
14	4730	5100
15	9240	9980
16	1100	1710
17	8290	8890
18	2570	2650
19	3880	4190
20	2440	2570

- a) Vilken estimator är lämplig att använda för att uppskatta produktens totala försäljningsvärde under reklamkampanjen, τ_y ? Ge en välgrundad motivering.
- b) Uppsatta produktens totala försäljningsvärde under reklamkampanjen, τ_y . Använd estimatorn du valde i uppgift a.
- c) Uppskatta variansen av uppskattningen i uppgift b. Ändlighetskorrektur ska användas.

Uppgift 5. (20 poäng)

En population består av $N = 6$ element med följande värden på en undersökningsvariabel Y

$$U = \{1, 2, 3, 7, 8, 9\}.$$

- a) Beräkna populationsmedelvärdet, μ_y , och populationsvariansen σ_y^2 .
- b) Man föreslår en urvalsdesign där man endast tillåter nio möjliga urval

$$\begin{aligned} S_1 &= \{1, 7\}, S_2 = \{1, 8\}, S_3 = \{1, 9\}, \\ S_4 &= \{2, 7\}, S_5 = \{2, 8\}, S_6 = \{2, 9\}, \\ S_7 &= \{3, 7\}, S_8 = \{3, 8\}, S_9 = \{3, 9\}, \end{aligned}$$

Vilket urval som dras är slumpmässigt och sannolikheten för vart och ett är lika stor, dvs. $1/9$. Ange inklusionssannolikheterna för vart och ett av elementen i U . Är den föreslagna urvalsdesignen ett sannolikhetsurval? Motivera ditt svar.

c) Beräkna urvalsmedelvärdena för vart och ett av urvalen ovan. Beräkna sedan variansen för urvalsmedelvärdena.

d) Beräkna den teoretiska variansen för urvalsmedelvärdet om man i stället använder ett OSU utan återläggning med $n = 2$ som urvalsdesign.

e) Vilken av de två urvalsdesignerna ger lägst varians? Förklara varför varianserna skiljer sig åt.