

STOCKHOLMS UNIVERSITET
Statistiska institutionen
Hans Nyquist

TENTAMEN I REGRESSIONSANALYS OCH UNDERSÖKNINGSMETODIK

DELKURS 2, UNDERSÖKNINGSMETODIK

2021-08-17

Skrivtid: 09.00-14.00, inlämningstid 14.00-15.00

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

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 31 augusti.

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)

Man vill uppskatta den totala intäkten hos en butikskedja med totalt 154 butiker. Man drar därför ett slumpmässigt urval om $n = 12$ butiker och noterade intäkterna under en vecka. De värden i tusentals kronor som observerades var

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
33,00	32,00	51,00	32,00	40,00	42,00	45,00	42,00	40,00	45,00	21,00	24,00

Uppskatta butikskedjans totala intäkter under observationsveckan och ange ett 99-procentigt konfidensintervall för uppskattningen.

b) Hur stort urval erfordras om ett precisionskrav på undersökningen är att den statistiska felmarginalen (halva bredden av ett 95 procentigt konfidensintervall) av butikskedjans totala intäkter är 500? Använd uppskattningar från uppgift a) för dina beräkningar.

Uppgift 2. (20 poäng)

Vid undersökningen i uppgift 1 noterades även antal kunder under observationsveckan. Observerat antal kunder i butikerna i urvalet redovisas i nedanstående tabell

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
120	90	160	120	130	140	140	140	130	150	90	110

Man vet att totalt hade butikerna i hela kedjan $\tau_x = 18248$ kunder.

a) Föreslå en metod för att uppskatta butikskedjans totala intäkter under observationsveckan där antal kunder används som hjälpinformation. Motivera ditt svar.

b) Använd den metod du föreslog i uppgift a) för att uppskatta den totala intäkten hos butikskedjan under observationsveckan.

c) Uppskatta variansen av uppskattningen i uppgift b).

d) Jämför precisionen i skattningen utan hjälpinformation (som i uppgift 1) och med hjälpinformation (som i denna uppgift).

Uppgift 3 (20 poäng)

I en kommun finns 300 elever i grundskolans årskurs 9. Av dessa är 146 flickor och 154 pojkar. Man vill nu uppskatta elevernas läsförståelse genom att låta ett urval av elever genomgå ett lästest. Ett stratifierat urval om 22 flickor och 22 pojkar drogs. De observerade medelpoängen var 78 bland flickorna och 75 bland pojkarna. Urvalsvarianserna var 42 respektive 110.

a) Uppskatta medelpoängen för läsförståelse bland eleverna i årskurs 9 i kommunen. Ändringskorrektur 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 4. (20 poäng)

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

- a) Olika urvalsdesign använder olika mycket information som finns tillgänglig vid urvalet. Ange för vart och ett av urvalsdesignen OSU, stratifierat urval och systematiskt urval, i vilken mån extra information används, hur den används och varför man använder den.
- b) Nämn två skattningstekniker som använder extra information vid uppskattning av ett populationsmedelvärde. Beskriv också hur den extra informationen används och varför man använder den.
- c) Förklara vad övertäckning och undertäckning är och vilka effekter det kan ha vid skattning av ett populationsmedelvärde.
- d) Förklara kortfattat begreppet snöbollsurval. Beskriv varför man ibland använder det samt vilka för och nackdelar det kan ha.

Uppgift 5. (20 poäng)

En population består av $N = 5$ element med följande värden på en undersökningsvariabel Y

$$U = \{1, 2, 5, 8, 9\}.$$

- a) Beräkna populationsmedelvärdet, μ_y .
- b) Lista alla möjliga urval av storleken $n = 3$ som kan dras från U med OSU utan återläggning. Beräkna sedan urvalsmedelvärdet, \bar{y} , för vart och ett av dessa urval. Observera att det finns $\binom{N}{n}$ olika urval, så du ska få så många urvalsmedelvärden, \bar{y} . Beräkna sedan medelvärdet av urvalsmedelvärdena $\bar{\bar{y}}$. Bekräftar det att \bar{y} är en väntevärdesriktig skattning av μ ?
- c) Beräkna variansen av de möjliga urvalsmedelvärdena \bar{y} .
- d) För vart och ett av urvalen i uppgift b, beräkna en uppskattning av variansen för urvalsmedelvärdet, $\widehat{V}(\bar{y})$, då ändlighetskorrektur tas med. Beräkna sedan medelvärdet av dessa uppskattningar. Bekräftar det att $\widehat{V}(\bar{y})$ är en väntevärdesriktig uppskattning av variansen för urvalsmedelvärdet, $V(\bar{y})$?