

STOCKHOLMS UNIVERSITET
Statistiska institutionen
Hans Nyquist

TENTAMEN I REGRESSIONSANALYS OCH UNDERSÖKNINGSMETODIK

DELKURS 2, UNDERSÖKNINGSMETODIK

2021-06-04

Skrivtid: 13.00-18.00, inlämningstid 18.00-19.00

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 21 juni.

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)

En förening som organiserar företag i hårvårdsbranschen har 288 medlemmar. Ett slumpmässigt urval om $n = 9$ företag drogs för att uppskatta den totala intäkten hos medlemsföretagen 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
16,50	16,00	26,00	21,00	20,00	20,50	22,50	21,00	20,00

Uppskatta totala intäkten hos medlemsföretagen under observationsveckan och ange ett 95-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 företagens totala intäkter är 400? Använd uppskattningar från uppgift a) i dina beräkningar.

Uppgift 2. (20 poäng)

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

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
11	10	16	13	13	14	14	14	12

Man vet att totalt hade företagen i föreningen $\tau_x = 3744$ kunder.

a) Föreslå en metod för att uppskatta den totala intäkten hos medlemsföretagen under observationsveckan där antal kunder används som hjälpinformation. Motivera ditt svar.

b) Använd den metod du föreslog i uppgift a) för att uppskatta medlemsföretagens totala intäkt under observationsveckan.

c) Uppskatta variansen av uppskattningen i uppgift b).

d) Jämför precisionen i skattningarna då hjälpinformation inte används (som i uppgift 1) och då hjälpinformation används (som i denna uppgift).

Uppgift 3. (20 poäng)

För att utvärdera effekten av reklam i tv undersöktes bl.a. hur lång tid invånarna i en kommun tittade på tv under en vecka. Kommunen består av tre byar, Norrbyn, Centrum och Sydbyn. I undersökningen använde man ett stratifierat urval med resultat som framgår av nedanstående tabell.

	Antal personer	Urvalsstorlek	Urvalsmedelvärde	Urvalsvarians
	N_k	n_k	\bar{y}_k	s_k^2
Norrbyn	155	20	16,2	34,2
Centrum	62	8	12,6	25,1
Sydbyn	93	12	9,1	19,0

- a) Uppskatta tiden i medeltal invånarna tittade på tv under observationsveckan 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 4. (20 poäng)

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

- a) I en undersökning om läsvanor kan en fråga vara "Vilka typer av genrer gillar du? Fler än ett alternativ kan väljas". Vad ska man tänka på när man definierar de olika svarsalternativen och hur kan svaren kodas?
- b) Vad står förkortningarna CAPI och CATI för? Ge exempel på någon fördel och någon nackdel för respektive metod.
- c) Beskriv vilka olika typer av bortfall som finns och vilka effekter det kan ha på en urvalsundersökning. Vad kan man göra för att motverka bortfall före och efter en undersökning?
- 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

$$U = \{1, 4, 5, 6, 9\}.$$

- a) Beräkna populationsmedelvärdet, μ_y .
- b) Lista alla möjliga urval av storleken $n = 2$ 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 ändlighetskorrektion tas med. Beräkna sedan medelvärdet av dessa uppskattningar. Bekräftar det att $\widehat{V}(\bar{y})$ är en väntevärdesriktig uppskattning av variansen för urvalsmedelvärdet, $V(\bar{y})$?