

STOCKHOLMS UNIVERSITET
Statistiska institutionen
Hans Nyquist

TENTAMEN I REGRESSIONSANALYS OCH UNDERSÖKNINGSMETODIK

DELKURS 2, UNDERSÖKNINGSMETODIK

2021-01-13

Skrivtid: 09.00-15.00

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 27 januari.

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)

Du får i uppgift att planera och genomföra en urvalsundersökning bland eleverna i en skola. Syftet med undersökningen är att uppskatta andelen elever som är positivt inställda till skolmatsalens salladsbord.

- a) Hur stort urval erfordras om det finns $N = 300$ elever vid skolan och ett precisionskrav på undersökningen är att den statistiska felmarginalen (halva bredden av ett 95 procentigt konfidensintervall) av proportionen positivt inställda elever är 0,05.
- b) Budgetrestriktioner innebar att urvalsstorleken bestämdes till $n = 80$. Vid undersökningen svarade 64 elever att de var positivt inställda till skolmatsalens salladsbord. Uppskatta andelen och bestäm ett 90 procentigt konfidensintervall.
- c) Innebär undersökningens resultat att det är rimligt att tro att fler än 70 procent av alla elever vid skolan är positivt inställda till matsalens salladsbord? Motivera ditt svar!

Uppgift 2. (20 poäng)

Medelinkomsten per hushåll ska uppskattas i en kommun med tre byar, Norrbyn, Centrum och Sydbyn. För detta har man använt ett stratifierat urval. I nedanstående tabell visas antal hushåll, urvalsstorlekar och resultat från undersökningen.

| | Antal hushåll | Urvalsstorlek | Observerat medelvärde | Observerad standardavvikelse |
|---------|---------------|---------------|-----------------------|------------------------------|
| Norrbyn | 890 | 50 | 95000 | 40000 |
| Centrum | 2200 | 150 | 65000 | 30000 |
| Sydbyn | 1170 | 100 | 50000 | 20000 |

- 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 och föreslå en urvalsallokering som kan minimera den statistiska felmarginalen med samma totala urvalsstorlek.

Uppgift 3. (20 poäng)

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

- Det finns ett antal kvalitetskriterier som enligt lag ska beaktas och deklarerats i Sveriges officiella statistik. Välj ut två av dessa kriterier och beskriv dem kortfattat. Beskriv också vilka målkonflikter som kan uppstå mellan de två kriterier som du valt. Om du inte tror att det finns en målkonflikt så måste det motiveras.
- Förklara begreppet systematiskt urval. Beskriv varför man ibland använder systematiska urval samt vilka för och nackdelar det kan ha.
- Beskriv vilka effekter bortfall kan ha på en urvalsundersökning.
- Vad innebär ett sannolikhetsurval och vilka fördelar det har jämfört med icke-sannolikhetsurval.

Uppgift 4. (20 poäng)

En tillverkare av äppeljuice köper in äpplen till ett pris baserat på mängden juice som kan utvinnas från äpplena. Man behöver därför uppskatta mängden juice som kan utvinnas från ett parti äpplen innan man pressar fram juicen. För att uppskatta mängden juice från ett parti som väger 40 kg (40 000 g) drog man ett OSU utan återläggning om $n = 15$ äpplen och för varje äpple observerade man vikten, x_i och juicemängden, y_i , $i = 1, 2, \dots, 15$. De erhållna observationerna visas i tabellen nedan.

| Äpple, i | Vikt, x_i | Juicemängd, y_i |
|------------|-------------|-------------------|
| 1 | 10,0 | 7,3 |
| 2 | 11,8 | 6,8 |
| 3 | 14,1 | 9,1 |
| 4 | 16,1 | 11,3 |
| 5 | 12,7 | 7,3 |
| 6 | 17,2 | 12,2 |
| 7 | 18,1 | 12,7 |
| 8 | 9,5 | 7,3 |
| 9 | 8,2 | 5,0 |
| 10 | 13,2 | 7,3 |
| 11 | 11,8 | 7,7 |
| 12 | 14,5 | 10,9 |
| 13 | 15,0 | 9,5 |
| 14 | 7,3 | 5,0 |
| 15 | 15,9 | 10,0 |

- Vilken estimator är lämplig att använda för att uppskatta mängden juice i partiet? Ge en välgrundad motivering.

b) Uppsatta den totala juicemängden i partiet äpplen. Använd estimatorn du valde i uppgift a.

c) Uppskatta variansen av uppskattningen i uppgift b. Man kan anta att antalet äpplen i urvalet är litet i förhållande till totala antalet äpplen i partiet så att man kan bortse från ändlighetskorrektion.

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 , och populationsvariansen σ_y^2 .

b) Lista alla möjliga urval av storleken $n = 2$ som kan dras från U med OSU *med* återläggning. Beräkna sedan urvalsmedelvärdet, \bar{y} , och urvalsvariansen, s_y^2 för vart och ett av dessa urval. Beräkna sedan medelvärdet av urvalsmedelvärdena respektive urvalsvarianserna.

c) Bekräftar resultaten i uppgift b att urvalsmedelvärdet \bar{y} och urvalsvariansen s_y^2 är väntevärdesriktiga uppskattningar av populationsmedelvärdet μ_y respektive populationsvariansen σ_y^2 ?

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, $V(\bar{y})$?