

## Tentamen i Statistisk analys

5 januari 2021 kl. 13–18

*Examinator:* Tom Britton, tel. 08-16 45 34, tom.britton@math.su.se

*Tillåtna hjälpmedel:* Formel- och tabellsamling, kurslitteratur och miniräknare.

*Återlämning:* Resultat och lösningar meddelas via kursforumet senast 15/1.

Varje korrekt löst uppgift ger 10 poäng. Gränsen för godkänt är preliminärt 30 poäng. För D krävs 34 p, för C 40 p, för B 48 p och för A krävs 54 p. Eventuella bonuspoäng adderas till tentamensresultatet. Texten ska vara väl läsbar och resonemang ska vara klara och tydliga.

---

Lösningar till uppgifterna måste göras självständigt, kommunikation med andra personer är ej tillåtet och kommer anmälas vid uppdragande. Tentamen innehåller en obligatorisk extra-uppgift 0 i form av en försäkran att du gjort lösningarna till uppgifterna självständigt.

Lösningar laddas upp på kursforumet senast kl 18:00. För studenter med vidimerad förlängd skrivtid gäller 19.00 (bifoga intyg). För sent inlämnad tentamen kommer ej behandlas.

Lösningarna måste bestå av EN fil i filformatet PDF. Lösningarna kan tex skrivas i LaTeX eller Word eller dylikt, men även prydligt handskrivna lösningar som fotograferas eller scannas går bra. Men undvik stora filer som kan vara svåra att ladda upp! I PDF-filen måste ditt fullständiga namn tydligt framgå på varje sida! Försäkra dig i förväg om hur du skapar en pdf-fil.

Jag kommer checka av chatten i Zoom efter 1 timme resp 2.5 timmar. Eventuella förtydliganden/modifieringar meddelas strax därefter i samma forum.

---

### Uppgift 0

Dina lösningar måste inledas med följande formulering (om inte så kommer tentan inte att rättas): Jag, <ditt namn och personnummer>, intygar på heder och samvete att jag inte tagit hjälp eller kommunicerat med någon annan person för att lösa dessa uppgifter, samt att jag inte surfat eller använt andra otillåtna hjälpmedel.

### Uppgift 1

Nedan följer 5 påståenden att svara sant eller falskt på (eller ingenting om man inte vet). Korrekt svar på respektive påstående ger 2p, fel svar ger -2p och inget svar ger 0p (om total-

summan skulle bli negativ sätts poängen till 0).

- a) För samma data och hypoteser så har ett test med felrisk  $\alpha = 0.05$  lägre styrka än om felrisken  $\alpha = 0.01$  används.
- b) Huvudproblemet med bortfall vid statistiska undersökningar är att uteblivna observationer ofta avviker från erhållna observationer.
- c) För att skapa en Normalfördelningsplot transformeras data så att de (approximativt) kommer ligga längs en rät linje om den bakomliggande fördelningen är normalfördelningen.
- d) Två variabler som är starkt beroende kan likväl ha en korrelation nära 0.
- e) En skattningen som inte är väntevärdesriktig är alltid sämre än en skattning som är väntevärdesriktig

## Uppgift 2

Betrakta följande stickprov av 8 observationer: 3.13, 2.93, 2.81, 3.23, 2.99, 3.05, 3.10, 2.91. Det gäller att  $\sum_i x_i = 24.15$  och  $\sum_i x_i^2 = 73.0315$ . Antag att data är approximativt normalfördelade.

- a) Testa  $H_0 : \mu = 3.00$  mot  $H_1 \neq 3.00$  på 99%-nivån. (5 p)
- b) Gör ett 95% konfidensintervall för  $\mu$ . (5 p)

## Uppgift 3

I stad A testades 500 slumpvis utvalda människor för antikroppar mot covid-19, vilket resulterade i att 110 gav positivt resultat. I stad B drogs också ett slumpvis urval om 500 personer, och här hade 91 positivt resultat.

- a) Går det att påstå att det är statistiskt säkerställt att stad A har större populationsandel med antikroppar. (Motivera ditt svar med lämpligt test. Använd 5% signifikansnivå.) (7 p)
- b) Det test du genomförde i uppgift a är baserat på vissa antaganden. Är det något av antagandena som kan ifrågasättas för denna tillämpning? (3 p)

## Uppgift 4

Antalet nyinlagda på intensivvård (IVA) i Sverige var de 5 veckorna innan jul som följer (vecka, antal nyinläggningar): (47, 154), (48, 168), (49, 166), (50, 167), (51, 184). Antag att observationerna respektive vecka är oberoende av varandra (ett antagande som inte stämmer helt, men nästan). Till er hjälp kan ni använda följande:  $\sum_i x_i = 245$ ,  $\sum_i x_i^2 = 12015$ ,  $\sum_i y_i = 839$ ,  $\sum_i y_i^2 = 141241$ ,  $\sum_i x_i y_i = 41170$ .

- a) Ansätt lämplig modell och testa på 95%-nivån hypotesen att antalet nyinläggningar växer, mot nollhypotesen att så inte är fallet. Redovisa dina beräkningar och parameterskattningar. (6 p)
- b) Gör en prediktion för hur många som läggs in på IVA veckan därpå (vecka 52) inklusive

ett intervall som redovisar osäkerheten i skattningen (t ex 95%). (4 p)

### Uppgift 5

Fyra olika kundstockar i ett försäkringsbolag jämfördes vad gäller inträffade skador de senaste 10 åren. Bland 100 kunder i vardera kundstock var det 32, 47, 39 respektive 44 kunder som inte haft någon skada under 10-årsperioden.

a) Testa hypotesen att de fyra kundstockarna är likvärdiga (och därmed kan slås ihop) jämfört med alternativhypotesen att alla fyra kundstockar inte är likvärdiga. Använd felrisk 5%. (8 p)

b) Beräkna ungefärligt p-värde för testet. Motivera ditt svar. (2 p)

### Uppgift 6

Betrakta undersökningssituationen i uppgift 3, dvs att 500 individer från respektive population testas. I den här uppgiften kan ni betrakta skattningen av standaravvikelsen som icke-slumpmässig och lika med motsvarande korrekta teoretiska värde, så endast osäkerheten i själva skattningarna behöver beaktas.

a) Vilken styrka har testet om sanningen är att andelen med antikroppar i stad A är 25% och andelen i stad B är 20%? (6 p)

b) Hur många behöver undersökas i respektive population för att uppnå styrkan 80%? (Vi förutsätter att lika många undersöks i respektive population.) (4 p)

*Lycka till!*