

## Tentamen i Statistisk analys

21 augusti 2020 kl. 9–15

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

*Tillåtna hjälpmedel:* Formel- och tabellsamling, kurslitteratur och dator. Även annan litteratur är tillåtet att använda, men försäkra dig om att den är tillförlitlig och ange referens.

*Återlämning:* Resultat och lösningar mailas till respektive skrivande senast fredag 28/8.

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.

---

Alla hjälpmedel i form av litteratur och beräkningsstöd är tillåtna. **Andra källor än kurslitteratur måste dock refereras till!!!** (för att jag ska kunna kolla att de stämmer).

Lösningar till uppgifterna måste göras självständigt, kommunikation med andra om uppgifterna är **ej** tillåtet. Tentamenslydelsen kommer innehålla en obligatorisk extrauppgift i form av en försäkran att du gjort lösningarna till uppgifterna självständigt.

Lösningar ska skickas in till tom.britton@math.su.se senast kl 15:00 samma dag. Lösningar inlämnade efter kl 15:00 kommer inte behandlas. För studenter med vidimerad förlängd skrivtid gäller 16.00 (bifoga intyg).

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 maila! 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 mail för eventuella frågor kl 10.00 resp 11.30. Om det skulle vara något jag behöver förtydliga kommer jag maila alla skrivande strax därefter, så kolla av din mail för detta ca 10.15 resp 11.45.

---

### 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.

## 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 totalsumman skulle bli negativ sätts poängen till 0).

- a) Föreställ dig en situation med hypotestest där nollhypotesen är sann (detta vet man förstås aldrig i realiteten, men låt oss likväl anta detta). I så fall är det troligt att  $p$ -värdet blir mindre ju större stickprov man har.
- b) ML-skattningen  $\hat{\theta}$  av en parameter  $\theta$  är det numeriska värde på  $\theta$  som gör att de observerade data har störst värde på sannolikhetsfunktionen (resp täthetsfunktionen för kontinuerliga variabler).
- c) Om en storhet av intresse uppmäts vid olika konsekutiva tidpunkter kan det vara olämpligt att använda regression eftersom två på varandra följande mätningar ofta är beroende.
- d) Ett mått att illustrera om lönerna i samhället går isär eller inte, är att studera utvecklingen av skillnaden mellan 90% lönepercentilen och 10% lönepercentilen. Att man inte studerar utvecklingen av lönernas standardavvikelse i stället, beror på att enstaka höga löner då skulle kunna ha för stor inverkan.
- e) Ett histogram baserat på ett stort stickprov kommer likna den fördelning (sannolikhetsfunktion eller täthetsfunktion) som datamaterialet kommer ifrån.

## Uppgift 2

Två Covid-19-drabbade kommuner hade 352 dödsfall bland sina 18352 invånare (kommun 1) respektive 421 dödsfall bland sina 33428 invånare (kommun 2).

- a) Går det att påstå att kommun 2 hade signifikant lägre dödlighet? (Motivera ditt svar med lämpligt test) (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 3

Betrakta följande parvisa sinsemellan oberoende observationer där andra komponenten är slumpmässig:

(56.7, 7.5), (66.6, -0.6), (55.7, 8.5), (63.5, 4.0), (63.1, 4.2), (57.6, 7.6), (57.8, 7.7), (60.4, 5.8), (57.4, 6.0), (59.2, 7.0), (59.3, 7.0).

- a) Ansätt lämplig modell, ange dess antaganden, och punktskatta **alla** parametrar. (4 p)
- b) Formulera statistiskt och testa hypotesen att andra komponenten i genomsnitt avtar med exakt 1 enhet om första komponenten ökar med en enhet. (3 p)
- c) En kommande observation kommer ha värdet 60.0 i första komponenten. Prediktera andra komponentens värde med ett 95% konfidensintervall. (3 p)

### Uppgift 4

Använd de 11 parvisa observationerna från föregående uppgift.

a) Beräkna Spearmans rangkorrelation.

(4 p)

b) Testa hypotesen att  $x$ - och  $y$ -variablerna (första och andra koefficienten) är oberoende mha icke-parametriska metoder på 99%-nivån.

(6 p)

### Uppgift 5

Arbetslösheten i 8 europeiska länder mättes i augusti 2019 och augusti 2020 (dvs före respektive efter Corona). Resultatet (procentsiffror) blev som följer

Land	Aug 2019	Aug 2020
A	10.6	12.5
B	4.9	7.1
C	8.6	9.2
D	6.5	8.1
E	7.1	11.5
F	4.3	5.9
G	7.5	9.1
H	9.1	10.9

a) Skatta effekten av Corona på arbetslösheten och ange vilka antaganden som görs. Bilda även ett 95% konfidensintervall för denna Coroneffekt.

(6 p)

b) Beräkna ungefärligt  $p$ -värde (eller inom vilka gränser  $p$ -värdet ligger) för hypotesen att Corona inte hade någon effekt alls.

(4 p)

### Uppgift 6

Gammafördelningen är en kontinuerlig fördelning på  $(0, \infty)$  som har två parameterar  $\alpha > 0$  och  $\beta > 0$ . Täthetsfunktionen ges av

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0,$$

där konstanten  $\Gamma(\alpha)$  är den s.k. gammafunktionen som för positiva heltal  $k$  uppfyller  $\Gamma(k) = (k-1)!$ .

Antag att ett stickprov  $x_1, \dots, x_n$  har samlats in från denna fördelning.

a) Antag att  $\alpha = 10$  är känt. Härled ML-skattningen för  $\beta$  under detta antagande.

(4 p)

b) Betrakta  $\alpha$  som en fri parameter igen. Man kan ganska lätt visa att  $E(X) = \alpha/\beta$  och att  $V(X) = \alpha/\beta^2$ . Använd detta för att konstruera momentskattningar för  $\alpha$  och  $\beta$  från stickprovet.

(6 p)

*Lycka till!*