

## Tentamen i Statistisk analys

7 februari 2025 kl. 8-13

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

*Tillåtna hjälpmedel:* Formel- och tabellsamling (delas ut) och miniräknare.

*Återlämning:* Tentan kommer vara rättad senast två veckor efter tentamensdagen. När tentan är rättad meddelas detta via kursforumet och därefter kan tentan hämtas ut vid matematik-expeditionen.

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 (alla gränser gäller inkl ev bonuspoäng). Texten ska vara väl läsbar och resonemang ska vara klara och tydliga.

---

Det skall tydligt framgå hur beräkningar gjorts. Kommunikation med andra personer är **ej** tillåtet och kommer anmälas vid uppdagande.

---

### 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) Ett stort endimensionellt datamaterial med många olika möjliga observationsvärden illustreras väl med ett histogram.
- b) Ett hypotestest som förkastar  $H_0$  betyder att mothypotesen  $H_1$  är en betydligt mer rimlig förklaringsmodell till data än vad  $H_0$  är.
- c) Kvartilavståndet för ett datamaterial kan vara bredare än datamaterialets variationsbredd.
- d) Antalet frihetsgrader vid statistiska test är ofta antalet observationer subtraherat med antal parametrar som skattas.
- e) Om styrkan i ett test för  $\mu = 11$  mot  $H_0 : \mu = 10$  är 80% så är styrkan för  $\mu = 12$  och samma  $H_0$  mindre än 80%.

## Uppgift 2

Bruttonationalprodukten (bnp) i Sverige var de 9 åren mellan åren 2015 och 2023: 2015: 4.23, 2016: 4.40, 2017: 4.58, 2018: 4.79, 2019: 5.03, 2020: 5.02, 2021: 5.46, 2022: 5.87, 2023: 6.21 (enhet tusen miljarder kronor). Vi ansätter enkel linjär regression med år som förklarande variabel, även om alla antaganden inte är helt uppfyllda. Om år betecknas med  $x$  och bnp med  $y$  så gäller följande (som du får använda):  $\sum_i x_i = 18171$ ,  $\sum_i x_i^2 = 36687309$ ,  $\sum_i y_i = 45.59$ ,  $\sum_i y_i^2 = 234.51$ ,  $\sum_i x_i y_i = 92060.53$ .

- Redogör för vilka antaganden som görs i linjär regression, samt vilken av dessa som framför allt kan ifrågasättas i denna tillämpning. (3 p)
- Skatta den förväntade årliga tillväxttakten. (3 p)
- Gör ett 95% prediktionsintervall för bnp för 2024. (4 p)

## Uppgift 3

Ett stickprov  $x_1, \dots, x_{15}$  av storlek 15 samlades in från en viss ”snäll” fördelning med väntevärde  $\mu$ . Utfallet blev sådant att  $\bar{x} = (1/15) \sum_{i=1}^{15} x_i = 6.71$  och  $s^2 = (1/14) \sum_{i=1}^{15} (x_i - \bar{x})^2 = 0.074$

- Skapa ett 2-sidigt 95%-konfidensintervall för  $\mu$ . (4 p)
- Testa den ensidiga hypotesen  $H_0 : \mu = 7$  mot alternativet  $H_1 : \mu < 7$  på 99%-nivån. (4 p)
- Ge en kort motivering till varför t-fördelning får användas även om fördelningen för  $X$  ej är normalfördelad. (2 p)

## Uppgift 4

På ett ungdomsläger med 50 st 18-åringar från 3 olika EU-länder tillfrågades vad som var viktigast för landets välbefinnande bland tre alternativ: en väl fungerande skola tillgänglig för alla, väl fungerande vård och omsorg tillgänglig för alla, samt ett starkt försvar så att landet inte attackeras av främmande makt. Resultatet blev enligt följande (påkittade data).

Land	Skola	Vård/omsorg	Försvar	Summa
Estland	12	13	25	50
Belgien	18	16	16	50
Portugal	14	19	17	50
Antal	44	48	58	150

- Ansätt nollhypotesen att det inte föreligger någon skillnad mellan ungdomars inställning i de tre länderna (och vi förutsätter att urvalen är representativa från respektive land). Vad blir de förväntade antal ungdomar i respektive kategori för respektive land (skapa en motsvarande tabell). (3 p)
- Undersök om populationerna i de tre länderna skiljer sig signifikant åt. Testa på 5%-nivån. (7 p)

### Uppgift 5

Vid en undersökning av sexualvanor vid en ungdomsklinik angav de svarande hur många partners de hade haft sex med det senaste året. Man erhöll svar från 8 män och 10 kvinnor. Antal partners som de 8 männen angav var: 2, 6, 9, 0, 2, 0, 13, 0. För de 10 kvinnorna blev svaren 1, 3, 1, 0, 4, 1, 3, 3, 1, 7. En av frågorna som man vill se besvarade är om män och kvinnor verkar ha samma fördelning för antal sexpartners.

- a) Vilka skäl finns att använda icke-parametriska metoder i detta fall? Motivera. (2 p)
- b) Hur ska du hantera att det finns observationer med samma numeriska värden? (2 p)
- c) Använd lämplig icke-parametrisk metod för att på 95%-nivån testa hypotesen att män och kvinnor (för populationsgruppen som besöker ungdomskliniker!) har samma fördelning av antal partners. (6 p)

### Uppgift 6

Följande 8 observationer kan antas vara ett stickprov från en okänd fördelning  $F$  med väntevärde  $\mu$  och standardavvikelse  $\sigma$ : 13.4, 14.1, 15.1, 14.3, 14.9, 13.9, 14.4, 14.0. Vi kan anta att fördelningen  $F$  är tillräckligt snäll så att vi kan normalapproximera. Till er hjälp har följande beräknats:  $\sum_i x_i = 114.1$ , samt  $\sum_i x_i^2 = 1629.45$ . Vi är primärt intresserade av spridningen i data och inte  $\mu$ .

- a) Ange en lämplig punktskattning av  $\sigma^2$  och av  $\sigma$ . (3 p)
- b) Ange ett 90% konfidensintervall för  $\sigma$  (Obs: ej för  $\sigma^2$ !). Kommentera hur  $s$  ligger i intervallet. (7 p)

*Lycka till!*