

Tentamen i Statistisk analys

17 augusti 2026 kl. 8-13

Examinator: Tom Britton, tel. 08-16 45 34, tom.britton@math.su.se. Besök på tentan: Ola Hössjer.

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

Återlämning: Tentan kommer vara rättad ca 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. I dessa poäng ingår eventuella 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) Parametern θ i en statistisk modell betraktas som en okänd konstant såväl i vanlig (frekventistisk) statistik som i Bayesiansk statistik.
- b) I regressionsanalys antas responsvariablerna $Y_1, \dots, Y_n$ vara likafördelade, men inte nödvändigtvis oberoende.
- c) Om nollhypotesen H_0 förkastas till förmån för alternativhypotesen H_1 , så är det statistiskt bevisat att nollhypotesen H_0 är falsk.
- d) Variansen för en skattning $\hat{\theta}_n$ baserad på n observationer går nästan alltid mot 0, tack vare stora talens lag.
- e) Stickprovsstandardavvikelse är ett bättre spridningsmått än vad stickprovsvarians är.

Uppgift 2

Vid en gäddfisketävling fick Moa 6 gäddor med vikter (x_i): 3.4, 3.7, 4.3, 4.7, 3.9, 4.0 kg.

- a) Beräkna medelvärde $\bar{x}$ och standardavvikelse s . (2 p)
- b) Antag i denna deluppgift att $\sigma = 0.5$ är känt sedan gammalt. Skapa ett 95% konfidensintervall för μ . (4 p)
- c) Gör om uppgift b), men nu under antagandet att σ inte är känt utan behöver skattas. Använd även detta konfidensintervall för att testa hypotesen $H_0 : \mu = 3.5\text{kg}$, på 5%-nivån. (4 p)

Uppgift 3

Tom upplever att hans löpstandard är på nedåtgående (trots träning). Han beslutar sig därför att springa samma 7km runda i mitten av varje månad under de 6 varmare månaderna. Vädret är lagom behagligt samtliga 6 löptillfällen så vädertrender kan bortses ifrån. Km-tiden de sex månaderna (april till september) blir i kalenderordning: 4.35, 4.41, 4.39, 4.49, 4.51 och i september 4.55.

- a) Ansätt lämplig modell för eventuell trend i Toms löpresultat och ange vilka antaganden som modellen gör, och skatta parametrarna (inklusive standardavvikelsen). (5 p)
- b) Toms peppande fru säger att den observerade trenden gott och väl kan förklaras av slumpmässig variation. Testa hennes hypotes på 5%-nivån (5 p)

Uppgift 4

50 slumpvis utvalda vuxna personer med heltidsanställning från 3 olika städer i Sverige tillfrågades om hur de tog sig till jobbet. Alternativen som gavs var: gå, cykel, bil eller kollektivtrafik (och ingen hade annat färdmedel). Utfallet blev som följer:

Stad	Gå	Cykel	Bil	Kollektivt	Summa
Xstad	6	12	13	19	50
Sveastad	10	8	20	12	50
Norrlandsborg	8	19	15	8	50
Antal	24	39	48	39	150

- a) Undersök om populationerna i de tre städerna kan antas välja färdmedel på liknande sätt, eller om färdmedel skiljer sig signifikant åt. Testa på 5%-nivån. (7 p)
- b) Antag i stället att du endast var intresserad av de två första städerna (Xstad och Sveastad) och endast huruvida bil används eller ej. Ange en skattning och 95% konfidensintervall för skillnaden i användandet av bil mellan Xstad och Sveastad. (3 p)

Uppgift 5

I Statistiska Centralbyråns partisynpatiuundersökning från maj 2026 hade socialdemokraterna 33.9%. Motsvarande siffra för undersökningen året innan (maj 2025) var 36.2%. Båda dessa undersökningar baseras på ca 4000 slumpvis valda svarande individer. I förra valet (år 2022) hade socialdemokraterna 30.3%. Sveriges röstberättigade befolkning är ca 8 miljoner vilket kan betraktas som oändligt i relation till slumpurvalet. Slumpurvalet kan således betraktas som ett oberoende urval där varje tillfrågad individ skulle rösta med socialdemokraterna med okänd sannolikhet p_S .

- a) Testa nollhypotesen att socialdemokraterna i maj 2026 ligger oförändrat från förra valet (du kan välja ensidig eller tvåsidig hypotes och valfri felrisk, men precisera vilket du väljer). Notera att valresultatet 30.3 % *inte* har någon slump i sig utan utgör nollhypotesens värde. (5 p)
- b) Konstruera ett 95% tvåsidigt konfidensintervall för *skillnaden* mellan maj 2026 och maj 2025 för socialdemokraterna. Verkar det vara en signifikant förändring? Notera att både dessa mätningar är slumpmässiga med samma urvalsstorlek, men oberoende eftersom olika individer tillfrågats. (5 p)

Uppgift 6

Låt Z vara för-första-gången-fördelad, dvs $p_Z(z) = p(1 - p)^{z-1}$, $k = 1, 2, \dots$. Antag att ett oberoende stickprov $z_1, \dots, z_n$ observeras.

- a) Härled ML-skattningen för p . (4 p)
- b) Vad blir ML-skattningen för $1/p$? (2 p)
- c) Denna deluppgift går ut på att konstruera ett 95% konfidensintervall för $1/p$. Centrala gränsvärdessatsen gör att $\bar{z}_n = (z_1 + \dots + z_n)/n$ kommer vara approximativt normalfördelad. Men med vilket väntevärde och vilken standardavvikelse? (Obs: väntevärde och varians för enskilda observationer Z står i formelsamlingen). Använd detta för att skapa ett approximativt 95% konfidensintervall för $1/p$. (4 p)

Lycka till!