STOCKHOLMS UNIVERSITET
MATEMATISKA INSTITUTIONEN
Avd. Matematisk statistik, TB

MT4001
TENTAMEN
12 januari 2022

Tentamen i Statistisk analys

12 januari 2022 kl. 8–13

Examinator: Tom Britton, tel. 08-16 45 34, tom.britton@math.su.se
Tillåtna hjälpmedel: Formel- och tabellsamling och miniräknare.
Återlämning: Onsdag 19/1 kl 13.30 i sal 36 kommer alla få se sina tentor, följt av en kortare genomgång av vanliga fel. (Tentor kan tas med hem först senare om de hämtas ut på expeditionen - de måste då kopieras).

Varje korrekt löst uppgift ger 10 poäng. Eventuella bonuspoäng adderas till tentamensresultatet. 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. Texten ska vara väl läsbar och resonemang ska vara klara och tydliga.

Lösningar till uppgifterna måste göras självständigt och 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) I ickeparametrisk inferens är storleksrelationen mellan olika observationer viktigare än de exakta numeriska värdena.

b) I Bayesiansk inferens minskar valet av apriorifördelning i betydelse ju fler observationer som stickprovet baseras på.

c) Randomiserade studier är bättre än observationella studier eftersom risken för effekter från dolda variabler minimeras.

d) Beräkningsintensiva statistiska metoder används oftast i situationer där traditionella explicita lösningar inte kan erhållas.

e) Det mest lämpade diagrammet för data på nominalskala är histogram.

Uppgift 2

Ett stickprov av storlek åtta från en okänd kontinuerlig fördelning med täthet f samlades in. Följande värden erhölls: 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 medelvärden från stickprov. Till er hjälp har följande beräknats: $\sum_i x_i = 114.1$, samt $\sum_i x_i^2 = 1629.45$. Låt μ beteckna fördelningens väntevärde och σ dess standardavvikelse.

a) Konstruera ett 95% konfidensintervall för μ under förutsättning att $\sigma = 0.5$ är känt på förhand. (5 p)

b) Konstruera ett 95% konfidensintervall för μ under förutsättning att σ inte är känt på förhand. (5 p)

Uppgift 3

För att bilda sig en uppfattning om hur årsmedeltemperaturen sjunker norrut i Sverige används enkel linjär regression. Årsmedeltemperaturen har samlats in från $n = 11$ orter i Sverige med känd lattitud: Jokkmokk: latt: 66.6, medeltemp: -0.6, Umeå: 63.5, 4.0, Östersund: 63.1, 4.2, Gävle: 60.4, 5.8, Karlstad: 59.2, 7.0, Stockholm: 59.3, 7.6, Jönköping: 57.4, 6.0, Visby: 57.6, 7.6, Göteborg, 57.8, 7.7, Kalmar: 56.7, 7.5, Lund: 55.7, 8.5. Följande sammanfattande storheter beräknades från data (x är lattitud och y årsmedeltemperatur): $\sum_i x_i = 657.3$, $\sum_i x_i^2 = 39389.45$, $\sum_i y_i = 65.3$, $\sum_i y_i^2 = 455.95$ samt $\sum_i x_i y_i = 3820.38$.

a) Skatta och ange ett 99% konfidensintervall för hur mycket medeltemperaturen sjunker per lattitud (inom Sverige). Obs: endast konfidensintervall för efterfrågad storhet ska anges. (En korrekt lösning under antagandet att $\sigma = 1$ ger 3p.) (5 p)

b) Vilken förklaringsgrad R^2 har modellen? (2 p)

c) Skatta medeltemperaturen för Bollnäs med lattitud 61.34 (endast punktskattning). (3 p)

Uppgift 4

En förpackning tryffelsvamp anges väga 20 g och företaget planerar att aningen minska förpackningsstorleken aningen. Innan detta görs vill man försäkra sig om att nuvarande medelvikt verkligen överstiger 20 g. Vid en kontroll av 10 slumpvis valda paket uppmättes följande vikter: 19.7, 20.3, 20.1, 20.5, 19.9, 20.0, 20.3, 20.6, 20.1, 20.2. Summan av vikterna och deras kvadratsumma ges av: $\sum_{i=1}^{10} x_i = 201.7$ och $\sum_{i=1}^{10} x_i^2 = 4068.95$.

a) Testa på 95%-nivån om medelvikten överstiger 20 kg eller inte. Ensidigt test ska användas. (6 p)

b) Gör ett 90% 2-sidigt konfidensintervall för den genomsnittliga vikten på dessa tryffelförpackningar.

(4 p)

Uppgift 5

I DN/IPSOS opinionsmätning för december 2021 får socialdemokraterna 29.2%, baserat på 1500 (gissat värde som jag tror stämmer ungefär) tillfrågade och svarande personer. I föregående val hade socialdemokraterna 28,2%, och i föregående opinionsmätning (oktober 2021) hade socialdemokraterna 25,7%.

a) Testa nollhypotesen att socialdemokraterna i december 2021 ligger oförändrat från förra valet (du kan välja ensidig eller tvåsidig hypotes men precisera vilket du väljer, och valfri felrisk). Notera att valresultatet 28,2 % *inte* har någon slump i sig utan utgör nollhypotesens värde. (5 p)

b) Konstruera ett 95% tvåsidigt konfidensintervall för *skillnaden* mellan december och oktobersympatierna för socialdemokraterna. Verkar det vara en signifikant uppgång? Notera att oktoberundersökningen också är slumpmässig med samma urvalsstorlek, men oberoende eftersom olika individer tillfrågats. (5 p)

Uppgift 6

Gammastrålningen från 10 meteorer uppmättes till följande värden: 10,8, 3,9, 106,5, 12,9, 7,4, 23,8, 1,7, 7,1, 18,6, 9,0 (påhittade data). Antag att dessa data kan betraktas som ett oberoende stickprov från den bakomliggande kontinuerliga fördelningen av meteorers gamma-strålning. Man är intresserad av medianen $x_{0.5}$ i den bakomliggande fördelningen.

a) Skatta fördelningens median $x_{0.5}$. (2 p)

b) Skapa ett 95% konfidensintervall för medianen. *Ledning*: Använd lämpliga gränser från det ordnade stickprovet $x_{(1)}, \ldots, x_{(10)}$. (8 p)