STOCKHOLMS UNIVERSITET
MATEMATISKA INSTITUTIONEN
Avd. Matematisk statistik, TB

MT4001
TENTAMEN
8 januari 2025

Tentamen i Statistisk analys

8 januari 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 matematikexpeditionen.

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) Aposteriorifördelningen väger ihop apriorofördelningen och likelihooden, och ju fler observationer som görs desto mindre inflytande får apriorifördelningen.

b) En vanlig företeelse inom icke-parametriska metoder är att studera rangerna för observationerna snarare än deras numeriska värden.

c) Om en normalfördelningsplot visar att punkterna ligger som en "S-kurva" snarare än en rät linje, så bör icke-parametriska metoder användas.

d) Ett histogram baserat på ett stort stickprov $x_1, \ldots, x_{1000}$ från en okänd kontinuerlig fördelning X kommer inte att likna dess täthetsfunktion $f_X(x)$, men däremot dess fördelningsfunktion $F_X(x)$.

e) Styrkan för ett test definieras som sannolikheten att förkasta H_0 när H_0 är sann.

Uppgift 2

Åtta fullvuxna abborrar fångades med kastspö i Tomtasjön i Roslagen. Antag att de 8 abborrarna kan ses som ett slumpmässigt stickprov från Tomtasjöns tusentals fullvuxna abborrar. De åtta abborrarna hade följande vikter (i hektogram): 3.8, 7.2, 6.5, 4.9, 4.1, 9.0, 8.2, 7.7. Följande summar kan du använda dig av: $\sum_i x_i = 51.4$ och $\sum_i x_i^2 = 356.88$.

a) Punktskatta den förväntade vikten på sjöns fullvuxna abborrar μ och konstruera ett 99% konfidensintervall för μ. Ange vilka antaganden som görs. (5 p)

b) Konstruera ett (tvåsidigt) test av hypotesen $H_0: \ \mu = 5.0$ hg, med felrisk 5%. (5 p)

Uppgift 3

Längd och vikten på 10 (slumpvis utvalda) svenska fullvuxna män var: (x, y): (182 cm, 77 kg), (188 cm, 90 kg), (177 cm, 76 kg), (180 cm, 79 kg), (186 cm, 77 kg), (192 cm, 87 kg), (185 cm, 76 kg), (181 cm, 80 kg), (174 cm, 70 kg), (180 cm, 75 kg).

Följande summor får användas: $\sum_i x_i = 1825$, $\sum_i y_i = 787$, $\sum_i x_i^2 = 333319$, $\sum_i y_i^2 = 62245$, $\sum_i x_i y_i = 143852$.

a) Formulera en lämplig modell för hur männens vikt Y beror av männens längd x, inklusive vilka antaganden som görs. Skatta modellens tre parametrar. (5 p)

b) Testa hypotesen att männens vikt är oberoende av deras längd. *Ledning:* T ex kan detta göras genom att skapa ett konfidensintervall för viktens kvantitative beroende av längden. (5 p)

Uppgift 4

Sju gymnasieelever skrev högskoleprovet på hösten i årskurs 3 *och* på våren i årskurs 3. Resultaten blev som följer: (0.85, 0.95), (1.10, 1.05), (1.10, 1.25), (1.30, 1.45), (1.50, 1.40), (0.80, 0.80), (1.20, 1.25).

a) Skatta hur mycket bättre elever i genomsnitt tycks bli mellan höst och vår i årskurs 3 (även inräknat effekten av att ha skrivit en gång). (3 p)

b) Testa hypotesen H_0: eleverna är lika bra på våren som på hösten, mot H_1: eleverna har blivit bättre till våren. (7 p)

Uppgift 5

En slumpvariabel Y sägs vara gammafördelad med parametrar k och β, skrivs $Y \sim \Gamma(k, \beta)$, om dess täthetsfördelning ges av

$$f(y) = \frac{\beta^k}{k!} y^{k-1} e^{-\beta y}, \ y > 0.$$

Gammafördelningen är identisk med exponentialfördelningen om $k = 1$, och för större heltal k så kommer summan av k st oberoende $Exp(\beta)$ variabler ha denna gammafördelning. Dvs om $X_1, \ldots, X_k$ är oberoende $Exp(\beta)$ så kommer $Y := X_1 + \cdots + X_k \sim \Gamma(k, \beta)$.

Antag att k är känt, så β är den enda obekanta parametern. Ett stickprov $y_1, \dots, y_n$ från denna gammafördelning erhålls.

a) Skatta β från stickprovet $y_1, \dots, y_n$ med momentmetoden . Du kan antingen beräkna $E(Y)$ från täthetsfunktionen direkt (tar dock lite tid), eller också utnyttja beskrivningen ovan att $Y = X_1 + \cdots + X_k$. (4 p)

b) Skatta β från stickprovet med ML-metoden (för detta behöver inte $E(Y)$ beräknas). (4 p)

c) Beräkna någon av skattningarnas numeriska värde om $k = 3$ och följande 4 y-observationer har gjorts: 4.7, 8.6, 7.2, 10.8. (1 p)

d) Välj någon av skattningarna ovan och tolka skattningens uttryck mha av beskrivningen att gamma-fördelningen kan ses som summan av k exponentialfördelningar. (1 p)

Uppgift 6

Följande stickprov av storlek 10 från en okänd fördelning X observerades: 7.9, 4.5, 12.1, 20.4, 9.1, 3.1, 42.9, 10.0, 14.2, 8.5.

a) Skatta den teoretiska medianen $x_{0.5}$ för den okända fördelningen. Ange både formeln och dess numeriska värde. (3 p)

b) Beräkna ett konfidensintervall med konfidensgrad så nära 95% som möjligt. Ange både formeln och dess numeriska värde. (7 p)

Lycka till!