UPPSALA UNIVERSITET
Matematiska institutionen
Rolf Larsson

TENTAMEN I MATEMATISK STATISTIK
Inferens 1 1MS035
2017–04–18

Skrivtid: 14.00-19.00. För betygen 3, 4, resp. 5 krävs 18, 25 resp. 32 poäng, inklusive ev. bonuspoäng. Lösningarna ska vara väl motiverade. Skriv endast på ena sidan, börja ny uppgift på ny sida och använd ej rödpenna. Uppgifterna är inte ordnade i svårighetsgrad!

Tillåtna hjälpmedel: Räknedosa. Formelsamling för Inferensteori 1.

1. Vi har ett slumpmässigt stickprov $x_1 = 3.2, x_2 = 2.2, x_3 = 3.6$ från X, som är Rayleighfördelad med täthetsfunktion

$$f_X(x) = \frac{x}{a} e^{-x^2/(2a)}$$

för $x \geq 0$ (och 0 annars).
I deluppgifterna nedan får utan bevis användas att $E(X) = \sqrt{\pi a/2}$.

   (a) Beräkna minsta kvadratskattningen av a. (2p)

   (b) Beräkna maximum likelihoodskattningen av a. (3p)

2. Man har ett slumpmässigt stickprov x_1, x_2 från en slumpvariabel X med väntevärde μ och varians σ^2.

   Man vill skatta μ, och två skattningar är föreslagna:

$$\mu_1^* = \frac{2x_1 + x_2}{3}, \quad \mu_2^* = 2x_1 - x_2.$$

   (a) Visa att båda skattningarna är väntevärdesriktiga. (1p)

   (b) Vilken av dessa skattningar är effektivast? (2p)

   (c) Finns det någon effektivare skattning på formen $a_1x_1 + a_2x_2$ där a_1 och a_2 är konstanter? Motivera ditt svar! (2p)

3. Ett företag tillverkar elektroniska komponenter. Under en dag är 5 st av dessa felaktiga.

   Antag att antalet felaktiga komponenter under en dag är Poissonfördelat med parameter λ, där λ kan tolkas som det genomsnittliga (eller förväntade) antalet felaktiga komponenter under en dag.

   (a) Företaget hävdar i sin reklam att det i genomsnitt tillverkar högst 2 felaktiga komponenter under en dag. Undersök om detta kan vara sant genom att testa en lämplig hypotes. (2p)

   (b) Vilken styrka har testet i (a) om $\lambda = 7$? (3p)

v.g.v

4. En datainsamlare på Vägverket räknar antalet bilar som kör igenom det ensligt belägna samhället Dysterhogdal under slumpmässigt valda tidsintervall, vardera en timme långa och inte överlappande. Hen tror att antalet bilar som passerar under en timme är Poissonfördelat med parameter λ. Vid fem sådana tillfällen observeras antalen 10, 14, 6, 11 och 9.

   Beräkna ett 95%-igt konfidensintervall för λ. (5p)

5. En behandling för att sänka det långsiktiga blodsockervärdet, HBA1c, testades på åtta patienter med diabetes typ 2. Värdet före och efter behandling mättes upp för varje patient. Resultaten ges i nedanstående tabell.

   Avgör, genom att utföra ett lämpligt test, om behandlingen hade effekt. Var noga med att specificera vilka förutsättningar du gör. (5p)

Patient nr	1	2	3	4	5	6	7	8
Före behandling	66	72	43	52	71	47	63	59
Efter behandling	54	61	47	41	67	46	61	51

6. En biolog undersöker förekomster av miljögifter i sjön Storsjövattnet. Hon fångar 50 gäddor och 50 abborrar. Av dessa har 12 av gäddorna och 8 av abborrarna en skadligt hög halt av kvicksilver i sig.

   (a) Kan man av detta dra slutsatsen att gäddor och abborrar är förgiftade (innehåller skadligt höga halter av kvicksilver) i olika utsträckning? Undersök detta genom att utföra ett lämpligt hypotestest på nivån 5%. (2p)

   (b) Antag att de sanna proportionerna förgiftade gäddor och abborrar är som i a)-uppgiften. Om man fångar lika många gäddor och abborrar, hur många måste man då minst fånga av varje sort för att kunna påvisa att de är förgiftade i olika utsträckning? (3p)

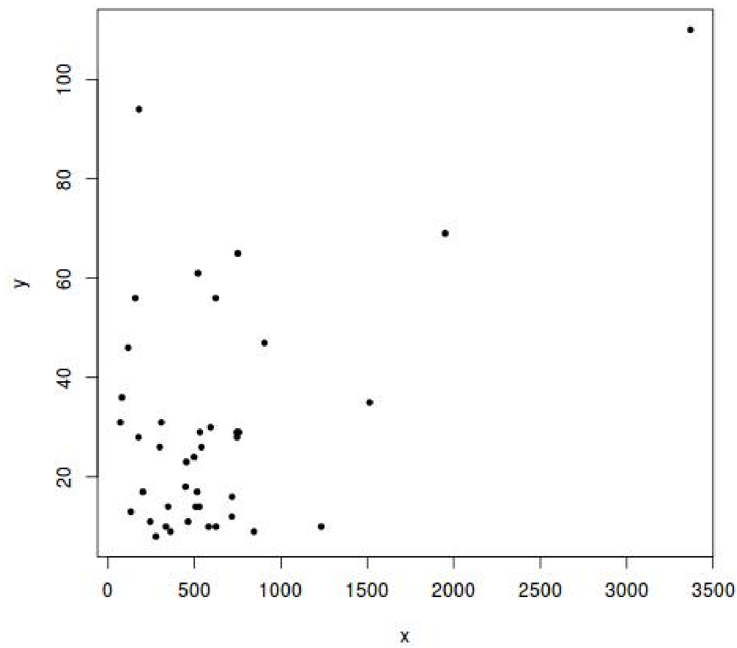
7. Zerblatt gör 100 oberoende mätningar av tyngdkraftsaccelerationen, g i m$/s^2$, på den för oss på Jorden avlägsna planeten QZ34. Mätningarna har medelvärde 15.20 och standardavvikelse 0.8.

   (a) Beräkna ett konfidensintervall för g med konfidensgrad 99%. (3p)

   (b) Det gäller att $G = m * g$ där G är ett föremåls tyngd i enheten Newton (N) och m är dess massa i kg. På planeten QZ34 hittar Zerblatt en sten med tyngd 8.0 N. Beräkna ett 99%-igt konfidensintervall för stenens massa. (2p)

V.g.v.

8. Under åren 1969-71 mättes svaveldioxidhalten (y) i luften upp i 41 städer i USA. I figur 1 är y plottad mot städernas folkmängd i tusental, x.



Figur 1: Plot av SO2-koncentration mot folkmängd, uppgift 8.

Räknehjälp: $\bar{x} = 608.6$, $\bar{y} = 30.05$, $S_{xx} = 13420000$, $S_{yy} = 22040$, $S_{xy} = 268500$.

Ansätt regressionsmodellen

$$Y_i = \alpha + \beta x_i + \varepsilon_i,$$

där $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, 2, ..., 41$, alla ε_i oberoende.

(a) Skatta parametrarna α och β. (1p)

(b) Beräkna förklaringsgraden. (1p)

(c) Två av observationsparen är $(x, y) = (3369,\ 110)$ och $(x, y) = (1950,\ 69)$. Om man uteslöt dessa, vilken effekt skulle det få på skattningen av β samt på förklaringsgraden? (Blir de större eller mindre än i (a) och (b)?) Det räcker med ett resonemang som svar, du behöver inte räkna! Tänk inte på kvadratsummorna utan geometriskt! (3p)

LYCKA TILL!