Niels Richard Hansen
Steffen Lauritzen

Praktiske oplysninger

Opgaven består af to dele med i alt 13 delspørgsmål som alle skal besvares. For at få lov til at deltage i den skriftlige eksamen i Matematisk Statistik i juni 2021 er det et krav, at opgavebesvarelsen afleveres til tiden og godkendes.

Del 1: Lineære normale modeller

For at undersøge riboflavins indflydelse på vækst udførte man et rotteforsøg, hvor riboflavin blev givet i fire forskellige doser (2.5 μg , 5.0 μg , 10.0 μg hhv. 20.0 μg per dag). 16 hanrotter og 16 hunrotter blev tilfældigt inddelt i de fire dosisgrupper med fire rotter af hvert køn i hver gruppe. Rotterne fik den angivne dosis dagligt i fire uger, og efter de fire uger blev deres ugentlige vækstrate (vægtforøgelse i g per uge) målt. Vækstraterne for de 32 rotter findes i datasættet `growthdata.txt` som ligger på Absalon.

I delspørgsmål 1. og 2. bedes du inddrage faktorerne `koen` (med niveauerne `Male, Female`) og dosisgruppe (med niveauerne `I, II, III, IV`) i analysen.

1. Argumenter for at en tosidet variansanalysemodel med vekselvirkning giver en god beskrivelse af variationen i data. Du bedes blandt andet diskutere, om responsvariablen `growth_rate` bør transformeres med logaritmen.

2. Argumenter for at designet

$$\mathbb{G} = \{\mathtt{1}, \mathtt{koen}, \mathtt{dosisgruppe}, \mathtt{koen} \times \mathtt{dosisgruppe}\}$$

   opfylder betingelserne fra EH Sætning 14.21 og tegn et faktorstrukturdiagram. Udfør et test for, om en additiv model giver en rimelig beskrivelse af data.

3. Angiv et estimat og et 95%-konfidensinterval for den forventede vækstrate for en `Female` rotte som gives en dosis på 10 μg.

I forsøg af denne type ser man ofte at vækstraten er en lineær (retlinet) funktion af logaritmen til dosis.

4. Vælg en (udgangs-)model på baggrund af konklusionen fra testet i delspørgsmål 2. og test hypotesen om, at vækstraten er en lineær funktion af logaritmen til dosis. Angiv dimensionen af de to modeller du tester imod hinanden samt fordelingen af F-teststørrelsen under hypotesen. Angiv ligeledes F-teststørrelsen, p-værdien og skriv en konklusion på testet.

5. Angiv estimater for MLE under hypotesen fra delspørgsmål 4. og forklar, hvordan estimaterne skal fortolkes. Angiv desuden estimatorens fordeling.

6. Angiv et 95%-prædiktionsinterval for vækstraten for en `Female` rotte, som får en dosis på 10 μg riboflavin per dag.

Del 2: Eksponentielle familier

Lad $(X_1, Y_1, \ldots, Y_n, X_n)$ være indbyrdes uafhængige og eksponentialfordelte stokastiske variable hvor X_j og Y_j har tætheder

$$f_j(x;\theta) = j\theta e^{-j\theta x},\ x > 0, \quad g_j(y;\lambda) = \lambda e^{-\lambda y},\ y > 0, \quad j = 1, \ldots, n,$$

og $(\theta, \lambda) \in \mathbb{R}_+^2$ begge er ukendte. Bemærk, at $Y_1, \ldots, Y_n$ er identisk fordelte, mens dette *ikke* er tilfældet for $X_1, \ldots, X_n$.

1. Gør rede for, at ovennævnte specificerer en minimal og regulær eksponentiel familie af dimension 2, angiv grundmålet, kumulantfunktionen, og den kanoniske stikprøvefunktion.

2. Bestem maksimaliseringsestimatoren for (θ, λ) baseret på $(X_1, Y_1, \ldots, Y_n, X_n)$.

3. Betragt nu delfamilien givet ved restriktionen $(\theta, \lambda) = (\beta, 1/\beta)$ og gør rede for, at denne udgør en krum eksponentiel familie af dimension 1 og orden 2.

4. Bestem maksimaliseringsestimatoren $\hat{\beta}_n$ for β baseret på en stikprøve $(X_1, Y_1, \ldots, Y_n, X_n)$ i delmodellen.

5. Bestem Fisher informationen $i_n(\beta)$ for β baseret på disse observationer.

6. Undersøg ved simulation om det kan antages approximativt at $\hat{\beta}_n \overset{\text{as}}{\sim} \mathcal{N}(\beta, i_n(\beta)^{-1}$ på trods af, at observationerne (X_j, Y_j) ikke er identisk fordelte.

7. Bestem likelihood ratio teststørrelsen Λ_n for den sammensatte hypotese $H_0 : \theta\lambda = 1$ og undersøg ved simulation om den kan antages approximativt χ^2-fordelt og i så fald med hvor mange frihedsgrader.