Københavns Universitet
Det Natur- og Biovidenskabelige Fakultet

Reeksamen - Statistik 2
4 timer skriftlig eksamen
23. august 2018

4 timers skriftlig prøve. Alle hjælpemidler tilladt (inkl. computer uden netforbindelse). Besvarelsen må skrives med blyant. Opgavesættet består af fire opgaver med i alt 17 delopgaver. Ved bedømmelsen vægtes alle delopgaver ens. Data til Opgave 3 og 4 ligger på en USB-nøgle i filerne `stat2_aug_2018_opg3.txt` *og* `stat2_aug_2018_opg4.txt`. *USB-nøglen skal returneres efter eksamen, men udelukkende for at den kan genbruges. Filer på USB-nøglen vil således ikke kunne indgå som en del af besvarelsen.*

Opgave 1

Lad $X = (X_1,\, X_2,\, X_3)^T \in \mathbb{R}^3$ være normalfordelt $\mathcal{N}(\xi, \Sigma)$, hvor

$$\xi = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$
$$\Sigma = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}.$$

1. Afgør om X regulært normalfordelt. Begrund dit svar.
2. Argumenter for at $Y = (X_1, X_2)^T$ er regulært normalfordelt og angiv den tilhørende præcision $\langle \cdot, \cdot \rangle$. Beregn værdien af tætheden for Y i punktet $y = (1, 1)$.
3. Angiv fordelingen af $\begin{pmatrix} X_1 \\ X_3 \end{pmatrix}^T \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_3 \end{pmatrix}$.

Opgave 2

Lad $W \sim \mathcal{N}(\xi, \sigma^2 I_6)$ være normalfordelt og betragt en observation $w = (1,\ 4,\ 6,\ 7,\ 10,\ 12)^T$ af W.

1. Opskriv designmatricen A svarende til følgende parametrisering af middelværdivektoren

$$\xi = A\beta = \begin{pmatrix} \beta_1 + \beta_2 \\ \beta_1 + 2\beta_2 \\ \beta_1 + 3\beta_2 \\ \beta_1 + 4\beta_2 \\ \beta_1 + 5\beta_2 \\ \beta_1 + 6\beta_2 \end{pmatrix}$$

   og find maksimaliseringsestimatorerne for β og σ^2.

2. Angiv fordelingen af maksimaliseringsestimatorne for β og σ^2 og angiv et 95 %-konfidensinterval for β_2.

Opgave 3

Ved et klinisk forsøg indgår 24 personer, som hver har afprøvet det ene blandt to forskellige medikamenter. Medikamentet er blevet afprøvet i enten en høj eller en lav dosis. Forsøgsdesignet fremgår af følgende tabel

		Dosis	
		lav	høj
Medikament	A	6	2
	B	12	4

Der er således f.eks. 12 personer, som har afprøvet medikament B i lav dosis. Effekten af medikamenterne vurderes ved at foretage en måling på hver forsøgsperson 3 måneder efter behandlingens start. Lad X_i betegne målingen (=responsen) hørende til den i-te forsøgsperson.

Betragt faktorerne medikament M med niveauer $\{A, B\}$ samt dosis D med niveauer $\{lav, hoj\}$.

1. Argumenter for at faktorerne M og D er geometrisk ortogonale og angiv dimensionen af underrummet $L_M + L_D$.
2. Idet vi antager at de 24 målinger er givet ved en lineær normal model

$$X = (X_1, \ldots, X_{24})^T \sim \mathcal{N}(\xi, \sigma^2 I_{24}),$$

bedes du udføre F-testet der sammenligner modellerne $\xi \in L_{M\times D}$ og $\xi \in L_M + L_D$. Benyt nedenstående faktorstrukturdiagram (1) til beregning af teststørrelsen F.

$$\begin{tikzcd} & & M^{424690.054,\,2} \arrow[dr] & \\ I^{431533.199,\,24} \arrow[r] & M \times D^{429921.699,\,4} \arrow[ur] \arrow[dr] & & 1^{417985.972,\,1} \\ & & D^{423214.260,\,2} \arrow[ur] & \end{tikzcd} \tag{1}$$

Opgaven fortsætter på næste side.

Forsøget udvides nu til også at omfatte 16 kontrolpersoner, der slet ikke modtager nogen behandling. For disse personer tillægges faktorerne M og D blot niveauet 0 således at det fulde forsøgsdesign bliver som anført i tabellen nedenfor

		Dosis		
		lav	høj	0
	A	6	2	0
Medikament	B	12	4	0
	0	0	0	16

Vi tænker i det følgende på M som en faktor med 3 niveauer $\{A, B, 0\}$. Tilsvarende er D en faktor med 3 niveauer $\{lav, hoj, 0\}$. Bemærk at visse kombinationer af $M \times D$ ikke giver mening og derfor ikke optræder i forsøgsdesignet.

Ved besvarelsen af de følgende delspørgsmål 3.-6. kan du benytte `R` til at regne på data fra filen `stat2_2018_aug_opg3.txt` på udleverede USB-nøgle.

3. Angiv de estimerede middelværdier for alle kombinationer af $M \times D$ som optræder i forsøget baseret på modellen $X \sim \mathcal{N}(\xi, \sigma^2 I_{40})$ hvor $\xi \in L_{M \times D}$.

4. Test om der er vekselvirkning mellem M og D, når der benyttes et signifikansniveau på 5 %.

5. Tag udgangspunkt i modellen $\xi \in L_M + L_D$ uanset hvilken konklusion du kom frem til i forrige delspørgsmål. Angiv dimensionen af underrummet $L_M + L_D$, og undersøg om modellen kan reduceres yderligere.

6. Angiv estimater og 95 % - konfidensintervaller for parametrene i modellen hvor det antages at $\xi \in L_M + L_D$. Benyt estimater og konfidensintervaller for parametrene til at drage konklusioner omkring effekten af medikament og dosis.

Opgave 4

I et markforsøg ønsker man at undersøge effekten af sprøjtning med bayleton på udbyttet af fire forskellige bygsorter (Lami, Lofa, Salka, Zita). I forsøget indgik 8 plots (=marker), hvoraf halvdelen blev sprøjtet med bayleton. Hvert plot blev inddelt i 4 mindre områder kaldet *parceller*, som blev dyrket med hver sin bygsort. I tabellen nedenfor er angivet udbyttet for hver af de 32 parceller i forsøget målt i enheden hekto (=100) kg per hektar.

		V			
P	B	Lami	Lofa	Salka	Zita
1	Ja	52.3836	49.6232	49.6232	49.7334
2	Ja	55.4953	52.7372	54.1466	52.8523
3	Ja	56.4953	48.8045	54.4419	53.1460
4	Ja	58.1991	48.3951	58.2607	55.5586
5	Nej	61.8205	52.1603	57.7914	59.1376
6	Nej	60.7161	53.1603	58.0875	57.3253
7	Nej	61.0146	58.3836	56.9758	57.6222
8	Nej	61.3131	55.8642	57.9758	60.0284

Data til opgaven findes på USB-nøglen i filen `stat2_2018_aug_opg4.txt`. Datasættet indeholder faktorerne P (=plot) med niveauer $\{1, \ldots, 8\}$, B (=bayleton) med niveauerne $\{Ja, Nej\}$ samt V (=sort) med niveauer $\{Lami, Lofa, Salka, Zita\}$.

1. Gør rede for at designet

$$\mathbb{G} = \{P, B \times V, B, V, 1\}$$

   er geometrisk ortogonalt og stabilt over for dannelse af minimum.

2. Med henvisning til notationen i lærebogen EH Sætning 14.21 om den ortogonale dekomposition bedes du organisere faktorerne fra $\mathbb{G}$ i et faktorstrukturdiagram og tilføje dimensionerne $dim(L_G)$ og $dim(V_G)$ for alle faktorer $G \in \mathbb{G}$. (Du skal ikke tilføje $||P_G X||^2$ og $||Q_G X||^2$).

3. Opstil en varianskomponentmodel for udbyttet på de enkelte parceller med $B \times V$ som fast effekt og P som tilfældig effekt. Angiv variansmatricen for samtlige målinger fra et vilkårligt plot udtrykt ved hjælp af parametrene i modellen.

4. Angiv estimaterne for variansparametrene i modellen fra spørgsmål 3.

Opgaven fortsætter på næste side.

5. Angiv på baggrund af modellen fra spørgsmål 3. estimatet for middelværdien af udbyttet på parceller svarende til følgende kombinationer af $B \times V$:

$$\{Nej, Lofa\}$$
$$\{Ja, Lami\}$$
$$\{Nej, Lami\}$$

6. Angiv 95 %-konfidensintervaller for middelværdierne hørende til grupperne anført under delopgave 5. Diskuter desuden om sprøjtning med bayleton lader til at påvirke udbyttet for sorten Lami.