

Meddelelse 18

Forelæsningerne i sidste uge

(Onsdag den 5. februar) Preben Blæsild gennemgik Afsnit 8.7 (Hotelling's T^2 -test for simpel middelværdihypotese) i alle detaljer.

(Fredag den 7. februar) Jeg afsluttede Kapitel 8 ved at omtale Afsnit 8.8 og Anneks til Kapitel 8 om beregninger i SAS.

Desuden begyndte jeg på Kapitel 9 om den flerdimensionale normalfordeling, hvor jeg nåede til midt på side 9.2.

Forelæsningerne i denne uge

(Onsdag den 12. februar): Jeg fortsatte med Kapitel 9 om den flerdimensionale normalfordeling, hvor jeg nåede til side 9.4 lige før Proposition 9.6.

(Fredag den 14. februar): Fortsætter gennemgangen af den flerdimensionale normalfordeling.

Øvelserne i de næste uger

(Uge 8: 17.-23. februar): 8.8 regnes færdig, 8.9, 8.10, 8.11, 8.12, 8.13, 8.14, 9.4 (gengivet på denne meddelelse)

Bemærkninger til opgaverne:

8.10: Spørgsmål 1) er identisk med opgave 4.9, som er regnet. Spørgsmål 2) handler om at indse at resultaterne i 1) er vigtige i lineære normale modeller). Det vigtigste her er at minde om resultaterne i 1), så 8.11 let kan regnes.

9.4: En liste af nyttige resultater, som vi ofte bruger. I når nok kun de første punkter ved øvelserne i uge 8.

(Uge 9: 24. februar - 2. marts): (foreløbigt program¹) 9.1, 9.2, 9.3 spørgsmål 4) og 5), 9.5, 9.17

Afleveringsopgaver

(Uge 9: 24. februar - 2. marts): De tre første spørgsmål af opgave 9.3 (gengivet på denne meddelelse)

¹Morten, Jan og jeg mødes onsdag og fastlægger den kommende uges program. Det vil være blandt disse opgaver at programmet vælges, undertiden suppleret med opgaver, som ikke er nået ved tidligere øvelser.

Opgave 9.1

Lad B være en $r \times p$ matriks med $p \geq r$. Hvis der eksisterer en matriks H , så

$$BH = I_r$$

hvor I_r er $r \times r$ identitetsmatricen, siges B at have en højre invers eller H at være en *højre invers* til B . På tilsvarende måde defineres en venstre invers til $p \times r$ ($p \geq r$) matricen A som en $r \times p$ matriks V så

$$VA = I_r.$$

- 1) Vis, at hvis B har en højre invers H så er $r(B) = r(H) = r$.
- 2) Vis, at hvis $r(B) = r$ så har B en højre invers H .

Opgave 9.2

Hvis alle de nødvendige inverse matricer eksisterer, så gælder, at hvis A er en $p \times p$ matriks, B en $p \times n$ matriks, C en $n \times n$ matriks og D en $n \times p$ matriks, så er

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}.$$

Opgave 9.3

Sporet (engelsk: trace) af en kvadratisk matriks A er defineret som summen af dens diagonalelementer og betegnes $\text{tr}A$.

Vis følgende:

- 1) $\text{tr}(A + B) = \text{tr}A + \text{tr}B$
- 2) $\text{tr}AB = \text{tr}BA$, når både AB og BA har mening
- 3) $\text{tr}B^{-1}AB = \text{tr}A$
- 4) $\text{tr}A = r(A)$, hvis $A^2 = A$ og $r(A)$ betegner rangen af A
- 5) $x^*Ax = \text{tr}xx^*A$, hvor x er en søjlevektor.

Opgave 9.4

Betragt en $p \times p$ matriks A , som er opdelt

$$A = \begin{Bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{Bmatrix}$$

så A_{11} er en $k \times k$ matriks og A_{22} er en $(p - k) \times (p - k)$ matriks. Lad $p \times p$ matricen

$$B = \begin{Bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{Bmatrix}$$

være opdelt på tilsvarende måde.

1) Vis at,

$$A + B = \begin{Bmatrix} A_{11} + B_{11} & A_{12} + B_{12} \\ A_{21} + B_{21} & A_{22} + B_{22} \end{Bmatrix} \quad (1)$$

og

$$AB = \begin{Bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{Bmatrix}$$

Lad A være invertibel og lad A^{-1} være opdelt som A

$$A^{-1} = \begin{Bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{Bmatrix}.$$

2) Vis, ved hjælp af $AA^{-1} = I$, at hvis A_{22} og A^{11} er invertible, er

$$A^{21}(A^{11})^{-1} = -(A_{22})^{-1}A_{21} \quad (2)$$

og

$$A_{11 \cdot 2} = A_{11} - A_{12}A_{22}^{-1}A_{21} = (A^{11})^{-1}. \quad (3)$$

Tilsvarende, hvis A_{11} og A^{22} er invertible, er

$$A^{12}(A^{22})^{-1} = -(A_{11})^{-1}A_{12} \quad (4)$$

og

$$A_{11 \cdot 1} = A_{22} - A_{21}A_{11}^{-1}A_{12} = (A^{22})^{-1}. \quad (5)$$

3) Vis, at hvis A er invertibel og $A_{12} = 0$ og $A_{21} = 0$ er

$$A^{-1} = \begin{Bmatrix} A_{11}^{-1} & 0 \\ 0 & A_{22}^{-1} \end{Bmatrix}. \quad (6)$$

4) Vis derefter følgende resultater om determinant af en opdelt matriks.

$$\begin{vmatrix} A_{11} & 0 \\ 0 & A_{22} \end{vmatrix} = \begin{vmatrix} A_{11} & 0 \\ 0 & I \end{vmatrix} \begin{vmatrix} I & 0 \\ 0 & A_{22} \end{vmatrix} = |A_{11}| |A_{22}| \quad (7)$$

$$\begin{vmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{vmatrix} = \begin{vmatrix} I & 0 \\ 0 & A_{22} \end{vmatrix} \begin{vmatrix} A_{11} & A_{12} \\ 0 & I \end{vmatrix} = |A_{11}| |A_{22}| \quad (8)$$

Hvis A_{11} er invertibel

$$|A| = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} = |A_{11}| |A_{22} - A_{21}A_{11}^{-1}A_{12}| = |A_{11}| |A_{22 \cdot 1}|. \quad (9)$$

Hvis A_{22} invertibel, er tilsvarende

$$|A| = |A_{22}| |A_{11} - A_{12}A_{22}^{-1}A_{21}| = |A_{22}| |A_{11 \cdot 2}|. \quad (10)$$

Opgave 9.5

Lad A være en $n \times k$ matrix og B en $k \times n$ matrix. Vis, at

$$|I_n + AB| = |I_k + BA|.$$

Vink: Benyt (9) i Opgave 9.4 på matricen

$$\begin{vmatrix} I_n & -A \\ B & I_k \end{vmatrix}.$$

Opgave 9.6

Lad X være regulært normalfordelt $N_p(\mu, \Sigma)$, og lad $X^* = (X^{(1)*}, X^{(2)*})^*$ være en opdeling af X så $X^{(1)}$ er k -dimensional.

Lad

$$\Sigma = \begin{Bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{Bmatrix}$$

og

$$\Sigma^{-1} = \begin{Bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{Bmatrix}$$

være de tilsvarende opdelinger af henholdsvis kovariansmatricen og *præcisionen* for X . (Den inverse kovariansmatricen for en flerdimensional regulær normalfordeling kaldes præcisionen.)

Vis, at Σ^{11} er præcisionsmatricen for den betingede fordeling af $X^{(1)}$ givet $X^{(2)}$.

Opgave 9.7

(Projektioner.)

Lad $\mathbb{R}^p$ være direkte sum af underrummene L_1 og L_2 , det vil sige enhver vektor $x \in \mathbb{R}^p$ har en entydig fremstilling som en sum af vektorer fra L_1 og L_2 .

$$x = x_1 + x_2, \quad x_1 \in L_1, \quad x_2 \in L_2. \quad (11)$$

Definer afbildningen $P : \mathbb{R}^p \rightarrow \mathbb{R}^p$ ved

$$P(x) = P(x_1 + x_2) = x_1,$$

hvor $x_1 \in L_1$ er x 's komponent i L_1 fra fremstillingen (11).

(1) Vis, at P er en lineær afbildning og at P er idempotent, det vil sige

$$P^2 = P.$$

P kaldes *projektionen på L_1 langs L_2* .

(2) Vis, at $(I - P)$ er projektionen på L_2 langs L_1 .

- (3) Vis, at hvis P er en lineær afbildning af $\mathbb{R}^p$ ind i $\mathbb{R}^p$, med $P^2 = P$, så findes lineære underrum L_1 og L_2 af $\mathbb{R}^p$, så P er projektionen på L_1 langs L_2 .

Lad $\mathbb{R}^p$ være forsynet med et indre produkt $(\cdot, \cdot)$, og lad P være projektionen på L_1 langs L_2 . P er en *ortogonal projektion*, hvis L_1 og L_2 er ortogonale, det vil sige

$$(x_1, x_2) = 0, \quad x_1 \in L_1, \quad x_2 \in L_2.$$

- (4) Vis, at P er ortogonal, hvis og kun hvis P er selvadjungeret, det vil sige

$$(Px, y) = (x, Py), \quad \forall x, y \in \mathbb{R}^p.$$

I praksis er P givet ved sin matriks $\underline{P}$ i den kanoniske basis $\{e_1, \dots, e_p\}$ i $\mathbb{R}^p$, og det indre produkt $(\cdot, \cdot)$ er givet ved en positiv definit matriks $\Sigma = \{(e_i, e_j)\}_{i,j=1}^p$, så

$$(x, y) = (x_1, \dots, x_p) \Sigma (y_1, \dots, y_p)^*,$$

hvor $x = \sum_{i=1}^p x_i e_i$ og $y = \sum_{i=1}^p y_i e_i$.

- (5) Vis, at projektionen P er en ortogonalprojektion med hensyn til det ved Σ givne indre produkt hvis og kun hvis matricen $\underline{P}$ for P opfylder

$$\underline{P}^* \Sigma = \Sigma \underline{P}.$$

- (6) Lad $L \subset \mathbb{R}^p$ være et q -dimensionalt underrum, som er frembragt af søjlerne i $p \times q$ matricen T . Vis, at den ortogonale projektion på L med hensyn til Σ har matriks

$$T(T^* \Sigma T)^{-1} T^* \Sigma$$

Bemærk specialtilfældet $\Sigma = I$.

- (7) Lad $L \subset \mathbb{R}^p$ være det q -dimensionale underrum

$$L = \{x \in \mathbb{R}^p \mid Rx = 0\},$$

hvor R er en $(p - q) \times p$ matriks af rang $p - q$. L er altså givet som nulrummet for den lineære afbildning bestemt ved R . Vis, at den ortogonale projektion på L med hensyn til Σ har matriks

$$I_p - \Sigma^{-1} R^* (R \Sigma^{-1} R^*)^{-1} R.$$

Opgave 9.8

(Multipel korrelationskoefficient.)

Lad $X \sim N_p(\mu, \Sigma)$ og lad $X^{(1)}, X^{(2)}$ være en k henholdsvis $p - k$ dimensional opdeling af X , med tilsvarende opdelinger af kovariansen

$$\Sigma = \begin{Bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{Bmatrix}.$$

Vi antager, at $\mu = 0$ da vi kun vil interessere os for spørgsmål om kovariansen. Yderligere antages, at Σ_{22} er invertibel.

Lad X_i være en komponent af $X^{(1)}$, og lad $\sigma_{(i)}$ være den i 'te søjle i Σ_{21} og lad β være den i 'te søjle i $\Sigma_{22}^{-1}\Sigma_{21}$ det vil sige

$$\beta = \Sigma_{22}^{-1}\sigma_{(i)}$$

β er regressionskoefficienterne for regressionen af X_i på $X^{(2)}$.

- (1) Vis, at $X_i - \beta \cdot X^{(2)}$ og $X^{(2)}$ er uafhængige.
- (2) Vis, at det lineære funktional $\alpha \cdot X^{(2)}$ af $X^{(2)}$ for hvilket $X_i - \alpha \cdot X^{(2)}$ har mindst varians er $\beta \cdot X^{(2)}$.
- (3) Vis, at den størst mulige korrelation mellem X_i og lineære funktioner $\alpha \cdot X^{(2)}$ af $X^{(2)}$ fås for $\alpha = \beta$.

Denne størst mulige korrelation er *den multiple korrelationskoefficient* mellem X_i og $X^{(2)}$. Den betegnes med $\bar{R}_{i \cdot k+1, \dots, p}$.

- (4) Vis, at

$$\bar{R}_{i \cdot k+1, \dots, p} = \frac{\sqrt{\sigma_{(i)}^* \Sigma_{22}^{-1} \sigma_{(i)}}}{\sqrt{\sigma_{ii}}}.$$

- (5) Vis, at

$$1 - \bar{R}_{i \cdot k+1, \dots, p}^2 = \frac{\sigma_{ii \cdot k+1, \dots, p}}{\sigma_{ii}},$$

hvor $\sigma_{ii \cdot k+1, \dots, p}$ er den partielle kovarians for $X_{k+1}, \dots, X_p$ fast. $\bar{R}_{i \cdot k+1, \dots, p}^2$ har altså en fortolkning som den del af variansen på X_i som forklares af variationen af $X^{(2)}$.

Opgave 9.9

(Multipel korrelationskoefficient.)

Lad

$$P = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{pmatrix}$$

være en 3×3 korrelationsmatriks.

Vis, at

$$\bar{R}_{1 \cdot 23}^2 = \frac{\rho_{12}^2 + \rho_{13}^2 - 2\rho_{12}\rho_{13}\rho_{23}}{1 - \rho_{23}^2}.$$

Opgave 9.17

(Kronecker produkt af matricer.)

Lad $A = \{a_{ij}\}$ være en $m \times n$ matris og lad $B = \{b_{rs}\}$ være en $p \times q$ matris. Kronecker produktet $A \otimes B$ af A og B er en $mp \times nq$ matris hvis $((i, r), (j, s))$ 'te element er

$$(A \otimes B)_{(i,r)(j,s)} = a_{ij}b_{rs}.$$

Idet $\{(i, r) | i = 1, \dots, m, r = 1, \dots, p\}$ og $\{(j, s) | j = 1, \dots, n, s = 1, \dots, q\}$ ordnes leksikografisk kan $A \otimes B$ skrives som en opdelt matris

$$A \otimes B = \{a_{ij}B\}$$

hvor den (i, j) 'te delmatris er $a_{ij}B, i = 1, \dots, m, j = 1, \dots, n$.

Vis følgende egenskaber ved Kronecker produktet:

(1) $0 \otimes B = 0 \quad A \otimes 0 = 0$

(2) $(A_1 + A_2) \otimes B = A_1 \otimes B + A_2 \otimes B$

(3) $A \otimes (B_1 + B_2) = A \otimes B_1 + A \otimes B_2$

(4) $aA \otimes bB = abA \otimes B, \quad a, b \in \mathbb{R}$

(5) $A_1 A_2 \otimes B_1 B_2 = (A_1 \otimes B_1)(A_2 \otimes B_2)$

(6) Hvis A og B er invertible er $A \otimes B$ invertibel, med

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}.$$

(7) $(A \otimes B)^* = A^* \otimes B^*$

(8) Hvis A er 1×1 , det vil sige $A = a$, er

$$a \otimes B = aB \text{ og } B \otimes a = aB.$$

(9) Hvis A er en $n \times n$ matris og B er en $p \times p$ matris er

$$\text{tr}(A \otimes B) = \text{tr}A \text{tr}B.$$

(10) Hvis A er en $n \times n$ matris og B er en $p \times p$ matris er

$$|A \otimes B| = |A|^p |B|^n.$$

Opgave 9.10

(Partielle korrelationer.)

Lad

$$P = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{pmatrix}$$

være en 3×3 korrelationsmatris.

(1) Vis, at

$$\rho_{12 \cdot 3} = \frac{\rho_{12} - \rho_{13}\rho_{23}}{\sqrt{(1 - \rho_{13}^2)(1 - \rho_{23}^2)}} \quad (12)$$

(2) Vis, at $\rho_{12 \cdot 3}$ kan være forskellig fra 0 selv om $\rho_{12} = 0$ og omvendt.

(3) Vis, at $\rho_{12 \cdot 3}$ og ρ_{12} har forskelligt fortegn når $\rho_{13}\rho_{23} > \rho_{12} > 0$ eller $\rho_{13}\rho_{23} < \rho_{12} < 0$.

Opgave 9.11

(Partielle korrelationer.)

Lad P være en $p \times p$ korrelationsmatriks. Vis, at

$$\rho_{ij \cdot k, \dots, p} = \frac{\rho_{ij \cdot k+1, \dots, p} - \rho_{ik \cdot k+1, \dots, p}\rho_{kj \cdot k+1, \dots, p}}{\sqrt{(1 - \rho_{ik \cdot k+1, \dots, p}^2)(1 - \rho_{kj \cdot k+1, \dots, p}^2)}} \quad (13)$$

Partielle korrelationer kan altså beregnes ud fra P ved først at benytte (12) i Opgave 9.10 og successivt ved hjælp af (13) føje variable til sættet af variable, man betinger med.

STAT1, 10. nr. 17

opgave 8.1

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{Bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{Bmatrix} \right)$$

$$\underline{z} = (x, y)^T, \quad \underline{\mu} = (\mu_1, \mu_2)^T, \quad \underline{\Sigma} = \begin{Bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{Bmatrix}$$

$$\text{Påstand: } f_{(X,Y)}(x,y) = \frac{1}{2\pi|\underline{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\underline{z}-\underline{\mu})^T \underline{\Sigma}^{-1}(\underline{z}-\underline{\mu})}$$

Beweis:

$$|\underline{\Sigma}| = \sigma_1^2 \sigma_2^2 - \rho^2 \sigma_1^2 \sigma_2^2 = \sigma_1^2 \sigma_2^2 (1 - \rho^2)$$

$$\text{dvs. } \sqrt{|\underline{\Sigma}|} = \sigma_1 \sigma_2 \sqrt{1 - \rho^2}$$

$$\underline{\Sigma}^{-1} = |\underline{\Sigma}|^{-1} \begin{Bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{Bmatrix}$$

$$\underline{\Sigma}^{-1}(\underline{z}-\underline{\mu}) = |\underline{\Sigma}|^{-1} \begin{Bmatrix} \sigma_2^2(x-\mu_1) - \rho\sigma_1\sigma_2(y-\mu_2) \\ -\rho\sigma_1\sigma_2(x-\mu_1) + \sigma_1^2(y-\mu_2) \end{Bmatrix}$$

$$(\underline{z}-\underline{\mu})^T \underline{\Sigma}^{-1}(\underline{z}-\underline{\mu})$$

$$= |\underline{\Sigma}|^{-1} \left((\sigma_2^2(x-\mu_1) - \rho\sigma_1\sigma_2(y-\mu_2))(x-\mu_1) + (-\rho\sigma_1\sigma_2(x-\mu_1) + \sigma_1^2(y-\mu_2))(y-\mu_2) \right)$$

$$= |\underline{\Sigma}|^{-1} (\sigma_2^2(x-\mu_1)^2 - 2\rho\sigma_1\sigma_2(x-\mu_1)(y-\mu_2) + \sigma_1^2(y-\mu_2)^2)$$

$$= \frac{1}{(1-\rho^2)} \left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x-\mu_1)}{\sigma_1} \frac{(y-\mu_2)}{\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right)$$

Det ønskede følger af def. (8.2)

□

opgave 8.3

Lad $(X, Y) \sim N_2(\mu, \Sigma)$, hvor μ, Σ er som i opg. 8.2

$$\begin{aligned} f(x, y) &= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^2} \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_2^2(1-\rho^2)}(y-\mu_2-\frac{\rho\sigma_2}{\sigma_1}(x-\mu_1))^2} \\ &= \frac{1}{\sqrt{2\pi}\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^2 - \frac{1}{2\sigma_2^2(1-\rho^2)}(y-\mu_2)^2 + \frac{\rho^2\sigma_2^2}{\sigma_1^2}(x-\mu_1)^2 - 2\rho\frac{\sigma_2}{\sigma_1}(y-\mu_2)(x-\mu_1)} \\ &= \frac{1}{\sqrt{2\pi}\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2}\left(\frac{1}{\sigma_1^2} + \frac{\rho^2\sigma_2^2}{\sigma_1^2\sigma_2^2(1-\rho^2)}\right)(x-\mu_1)^2 - \frac{2\rho\sigma_2}{\sigma_1\sigma_2^2(1-\rho^2)}(x-\mu_1)(y-\mu_2) + \frac{1}{\sigma_2^2(1-\rho^2)}(y-\mu_2)^2} \\ &= \frac{1}{\sqrt{2\pi}\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)}{\sigma_1}\frac{(y-\mu_2)}{\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)} = f_{(X,Y)}(x, y) \end{aligned}$$

Juf. prop. 9.6 er $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$

Pr. def. af betingede tætheder gælder $f_{(X,Y)}(x, y) = f_X(x) f_{(Y|X)}(y)$, ud fra dette og

ovenstående opsplittning (8.4) kan vi dermed konkludere (idet $X \sim N(\mu_1, \sigma_1^2)$), at

$$(Y|X=x) \sim N\left(\mu_2 + \frac{\rho\sigma_2}{\sigma_1}(x-\mu_1); \sigma_2^2(1-\rho^2)\right)$$

Opgave 8.5

Ud fra figuren tyder $(\bar{X}_{\text{son}}, \bar{X}_{\text{far}})$ på at være 2-dim normalfordelt, med middelværdier $\sim (67, 67)$ og jvf. bn. s. 8.3 at korrelationskoeff. er negativ og spredningen er noget større for sønne end for fædre.


```

OPTIONS LS=100 PS=70;
DATA opg_05;
INPUT soen far antal @@;
DO i=1 TO antal;
    filius=soen;
    pater=far;
    OUTPUT;
END;
KEEP filius pater;
LABEL filius='sons hoejde i tommer'
      pater='fars hoejde i tommer';
DATALINES;
           60 63 1 60 65 1
        61 63 1        61 67 1
62 60 1           62 64 1           62 66 1 62 67 1
           63 62 3 63 63 2 63 64 2 63 65 4 63 66 4 63 67 2 63 68 1 63 70 1
64 59 1 64 61 2 64 62 4 64 63 3 64 64 4 64 65 8 64 66 10 64 67 3 64 68 1 64 69 2 64 70
65 59 2 65 60 2 65 62 2 65 63 3 65 64 9 65 65 13 65 66 11 65 67 8 65 68 6 65 69 3 65 70
66 60 1 66 61 1 66 62 3 66 63 5 66 64 9 66 65 10 66 66 16 66 67 17 66 68 16 66 69 5 66 70
           67 61 2 67 62 5 67 63 4 67 64 14 67 65 20 67 66 26 67 67 26 67 68 19 67 69 13 67 70
           68 61 1 68 62 2 68 63 8 68 64 10 68 65 10 68 66 25 68 67 31 68 68 24 68 69 29 68 70
           69 61 1 69 63 5 69 64 5 69 65 13 69 66 19 69 67 17 69 68 24 69 69 29 69 70 21 69 71
70 63 1 70 64 2 70 65 6 70 66 18 70 67 12 70 68 20 70 69 22 70 70 20 70 71 14 70 72 6 70 73
71 64 4 71 65 5 71 66 9 71 67 10 71 68 19 71 69 15 71 70 21 71 71 11 71 72 8 71 73 5 71 74
72 65 3 72 66 2 72 67 7 72 68 7 72 69 11 72 70 11 72 71 10 72 72 9 72 73 3
73 65 1 73 66 1 73 67 2 73 68 8 73 69 7 73 70 6 73 71 7 73 72 6 73 73 3 73 74 1
74 63 1 74 65 2 74 66 1 74 68 5 74 69 2 74 70 2 74 71 7 74 72 4 74 73 3 74 75 2
75 69 1 75 70 2 75 72 3 75 74 2
76 68 2 76 73 1 76 74 1
77 68 1 77 70 1 77 73 2
78 69 1 78 70 1 78 72 1

;

PROC MEANS DATA=opg_05 N SUM USS;
RUN;

/*

PROC FREQ DATA=opg_05 COMPRESS;
TABLES pater*filius /NOPERCENT
           NOROW
           NOCOL
           NOCUM;

RUN;
*/

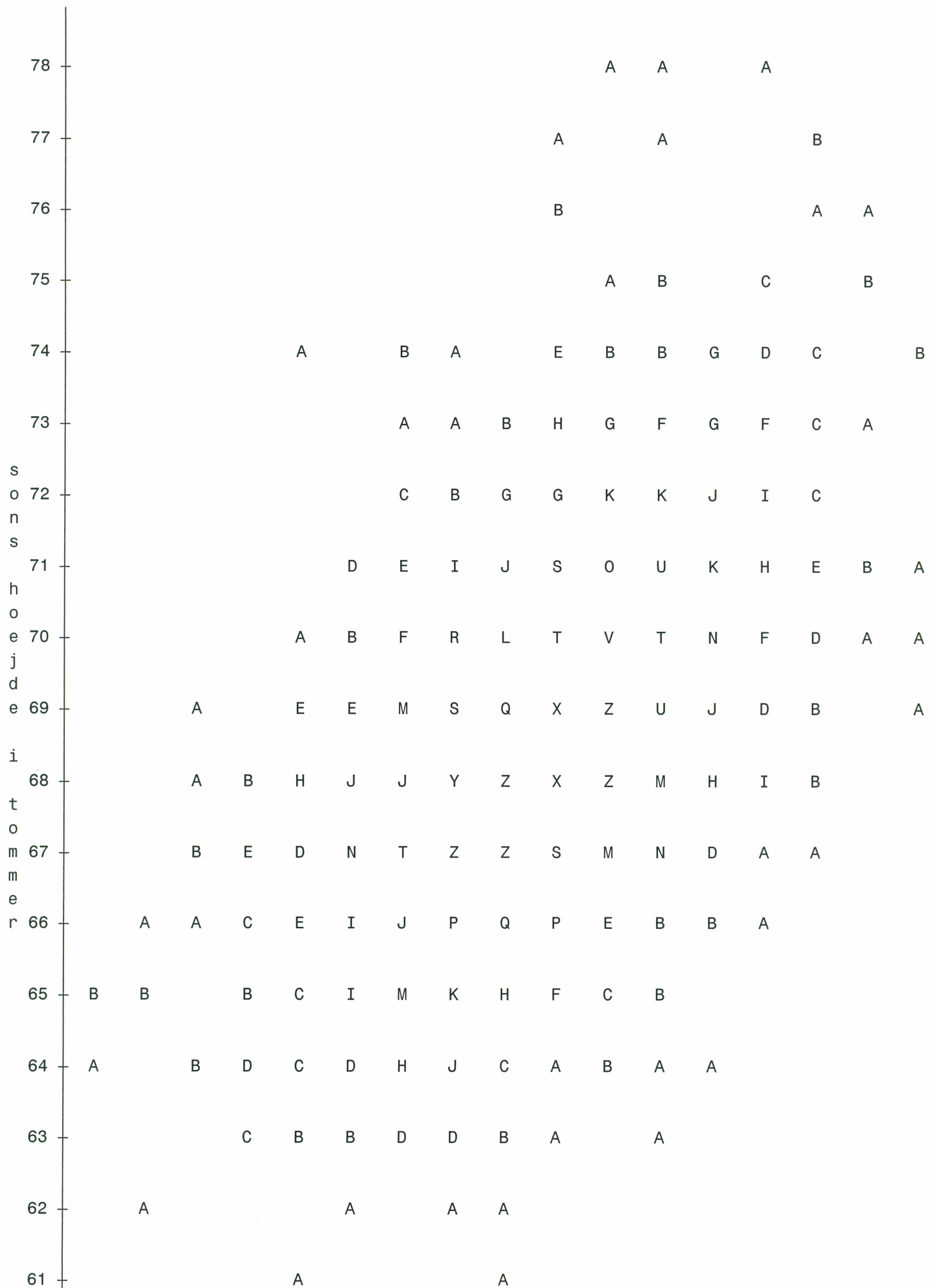
PROC PLOT DATA=opg_05;
PLOT filius*pater;
RUN;

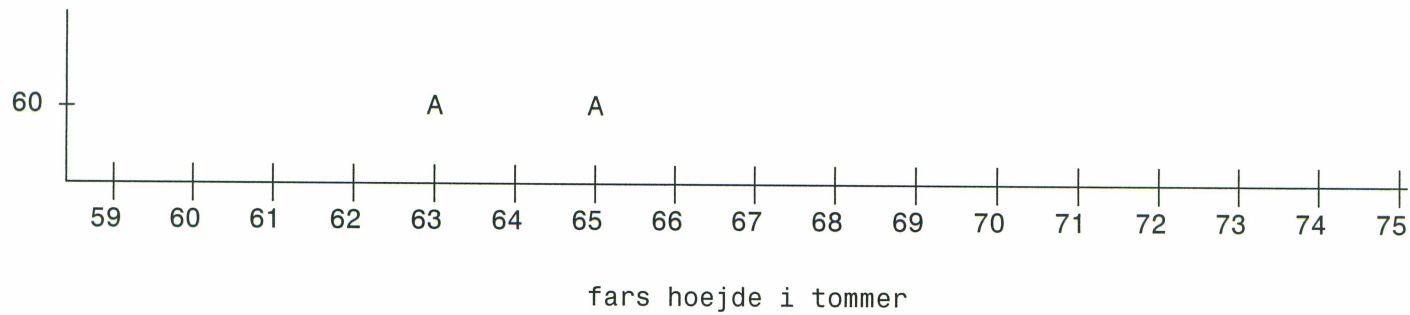
```

The MEANS Procedure

Variable	Label	N	Sum	USS
filius	sons hoejde i tommer	1078	74012.00	5089518.00
pater	fars hoejde i tommer	1078	72974.00	4947958.00

sons
hoe
jde
i
t
om
mer





NOTE: 11 obs hidden.

opgave 8.6

1)

Påstand:

$$E[f(X)] \approx f(E[X]), \quad \text{Var}(f(X)) \approx f'(E[X])^2 \text{Var}(X)$$

Beweis:

Tuf. sætn. 11.1.1 i Lindström (s. 536) er

$$f(x) \approx f(E[X]) + f'(E[X])(x - E[X])$$

Dvs.

$$E[f(X)] \approx f(E[X]) + f'(E[X])(E[X] - E[X]) = f(E[X])$$

$$E[(f(X))^2] \approx (f(E[X]))^2 + (f'(E[X]))^2 E[(X - E[X])^2]$$

$$= (f(E[X]))^2 + (f'(E[X]))^2 \text{Var}(X)$$

$$\text{Var}(f(X)) = E[(f(X))^2] - E[f(X)]^2 \approx (f'(E[X]))^2 \text{Var}(X)$$

□

2)

$$X \sim \text{po}(\lambda)$$

Et ønsket f opfylder $V(f(X)) \approx (f'(\lambda))^2 \lambda$

Sæt $f(x) = \frac{1}{\sqrt{x}} x$, da ser vi, at variansen approksimativt er uafh. af λ .

3)

$$\text{Def. : } f(x) = \frac{1}{2} \ln\left(\frac{1+x}{1-x}\right), \quad x \in]-1, 1[$$

$$f'(x) = \frac{1}{2} \frac{1-x}{1+x} \frac{2}{(1-x)^2} = \frac{1}{1-x^2}$$

$$Z(R) := \frac{1}{2} \ln \frac{1+R}{1-R}$$

Luf 1):

$$E[Z(R)] \approx \frac{1}{2} \ln \frac{1+E[R]}{1-E[R]} \approx \frac{1}{2} \ln \frac{1+p}{1-p}$$

$$\text{Var}(Z(R)) \approx \frac{1}{(1-E[R]^2)^2} \text{Var}(R) \approx \frac{1}{(1-p^2)^2} \frac{(1-p^2)^2}{n-1} = \frac{1}{n-1}$$

Dvs. $Z(R)$ har approksimativt konstant varians.

opgave 8.8

1)

$$L(\mu_1, \dots, \mu_k) = \prod_{i=1}^k \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi}\sigma_{0i}} e^{-\frac{1}{2}(x_{ij} - \mu_i)^2 / \sigma_{0i}^2}$$

$$= \prod_{i=1}^k (2\pi\sigma_{0i}^2)^{-\frac{n_i}{2}} e^{-\frac{1}{2} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2 / \sigma_{0i}^2}$$

Dus, $\hat{\mu}_i = \bar{X}_{i\cdot}$ $i=1, \dots, k$ og dermed at

$$(\hat{\mu}_1, \dots, \hat{\mu}_k) = (\bar{X}_{1\cdot}, \dots, \bar{X}_{k\cdot})$$

$\bar{X}_{i\cdot} = \frac{1}{n_i} (X_{i1} + \dots + X_{in_i})$ pga. at X 'erne er uafh.
ser vi, at $\bar{X}_i \sim N(\mu_i, \sigma_{0i}^2/n_i)$

Pga. uafh. fås, at $(\bar{X}_{1\cdot}, \dots, \bar{X}_{k\cdot}) \sim N_k(\underline{\mu}, \underline{\Sigma})$

$$\text{hvor } \underline{\Sigma} = \begin{Bmatrix} \sigma_{01}^2/n_1 & 0 \\ 0 & \ddots \\ 0 & \dots & \sigma_{0k}^2/n_k \end{Bmatrix} \quad (\text{Jvf. prop. 9.4})$$

2)

Under H_0 bliv

$$\ell(\mu) = -\sum_{ij} \ln(\sqrt{2\pi}\sigma_{0i}^2) - \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^{n_i} \frac{(x_{ij} - \mu)^2}{\sigma_{0i}^2}$$

$$\frac{\partial \ell(\mu)}{\partial \mu}(\mu) = \sum_{i=1}^k \frac{1}{\sigma_{0i}^2} (x_{i\cdot} - n_i \mu)$$

$$\frac{\partial \ell}{\partial \mu}(\mu) = 0 \iff \mu = \frac{\sum_{i=1}^k \frac{x_{i\cdot}}{\sigma_{0i}^2}}{\sum_{i=1}^k \frac{n_i}{\sigma_{0i}^2}} = \frac{\sum_{i=1}^k \frac{n_i \bar{x}_{i\cdot}}{\sigma_{0i}^2}}{\sum_{i=1}^k \frac{n_i}{\sigma_{0i}^2}}$$

og dette er et maksimum (j.f. fortegnen på $\frac{d\ell}{d\mu}$)

$$\text{Altså er } \hat{\mu} = \frac{\sum_{i=1}^k \frac{n_i \bar{x}_i}{\sigma_{i0}^2}}{\sum_{i=1}^k \frac{n_i}{\sigma_{i0}^2}}$$

Pga. uafh. af x om $\bar{x}_i$ er uafh., dvs.

$$\sum_{i=1}^k \frac{n_i}{\sigma_{i0}^2} \bar{x}_i \sim N\left(\sum_{i=1}^k \frac{n_i}{\sigma_{i0}^2} \mu, \sum_{i=1}^k n_i / \sigma_{i0}^2\right) \text{ og dermed}$$

$$\hat{\mu} \sim N\left(\mu, \left(\sum_{i=1}^k n_i / \sigma_{i0}^2\right)^{-1}\right)$$

3)

$$Q_1(\underline{x})_{1.13} = e^{-\frac{1}{2} \left(\sum_{ij} (x_{ij} - \hat{\mu})^2 / \sigma_{01}^2 - \sum_{ij} (x_{ij} - \bar{x}_{i.})^2 / \sigma_{01}^2 \right)}$$

$$= e^{-\frac{1}{2} \sum_{ij} (-2x_{ij}(\hat{\mu} + \bar{x}_{i.}) + \hat{\mu}^2 - \bar{x}_{i.}^2) / \sigma_{01}^2}$$

$$= e^{-\frac{1}{2} \sum_i \frac{n_i}{\sigma_{01}^2} (\bar{x}_{i.} - \hat{\mu})^2}$$

$$-2 \ln Q_1(\underline{x}) = \sum_i \frac{n_i}{\sigma_{01}^2} (\bar{x}_{i.} - \hat{\mu})^2$$

Udgangspkt. i ovenstående er

$$M_1: \mathbf{X}_{ij} \sim N(\mu_i, \sigma_{i0}^2)$$

Hvilket er ækvivalent med (husk σ_{i0} er givet tal)

$$\mathbb{X}_{ij} := \mathbf{X}_{ij} / \sigma_{i0} \sim N(\mu_i / \sigma_{i0}, 1)$$

Vi kan bruge teorien om lin. normale modeller på $\mathbb{X}_{ij}$ ene

Underrummet L_1 hørende til middelværdistrukturen i M_1 har dimension k og en basis er givet ved k søjler, hvor i te søjle har 0'er på alle indgange undtagen indgangen fra $\sum_{j=1}^{i-1} n_j$ til $\sum_{j=1}^i n_j$ hvor der står $1/\sigma_{i0}$ jvf. s. 4.53; se vi $\hat{\mu}_i = \bar{\mathbf{X}}_i$.

I modellen $M_2: \mathbf{X}_{ij} \sim N(\mu, \sigma_{i0}^2)$ som er ækvivalent med $\mathbb{X}_{ij} \sim N(\mu / \sigma_{i0}, 1)$ er underrummet $L_2 \subset L_1$ og L_2 har dim. 1 og indgangen i en basisvektor for L_2 er fra $\sum_{j=1}^i n_j$ til $\sum_{j=1}^i n_j$ 1te indgang $1/\sigma_{i0}$

Ved at normere med $\sum n_i / \sigma_{i0}$ fås en orthonormalbasis og vi ser jvf. s. 4.52-53 at $\hat{\mu} = \frac{\sum \frac{\mathbf{X}_i \cdot}{\sigma_{i0}^2}}{\sum n_i / \sigma_{i0}^2} = \frac{\sum \frac{n \bar{\mathbf{X}}_i}{\sigma_{i0}^2}}{\sum n_i / \sigma_{i0}^2}$

$$\text{og } \sum_{i=1}^k \frac{n_i}{\sigma_{i0}^2} (\bar{\mathbf{X}}_i - \hat{\mu})^2 = \sum_{i,j} \left(\frac{\bar{\mathbf{X}}_i}{\sigma_{i0}} - \frac{\hat{\mu}}{\sigma_{i0}} \right)^2 = \|P_1(\mathbb{X}) - P_2(\mathbb{X})\|^2$$

$= -2 \ln Q(\mathbb{X})$ jvf. den foregående opg. besvarelse.

Jf. sætn. 4.1 $\hat{\mu} \perp -2 \ln Q(\mathbb{X})$ og $-2 \ln Q(\mathbb{X}) \sim \chi^2(k-1)$

*) P_1, P_2 betegner hh. projektorerne ned på L_1 og L_2

Meddelelse 17

Forelæsningerne

onsdag	11-12	Institut for Matematiske Fag, auditorium D1 (1-531-113)
fredag	9-11	Institut for Matematiske Fag, auditorium G1 (1-532-116)

Bemærk det ændrede tidspunkt og lokale for fredagsforelæsningen i forhold til efteråret.

Øvelserne

Hold	Ugedag	Tid	Lokale	Computerrum	Instruktor
2	mandag	9-12	1-531-011	1-531-015	Jan Pedersen
1	onsdag	8-11	1-531-011	1-531-015	Morten Fenger Grøn

De sidste forelæsninger i december

(Onsdag den 18. december) Afsnit 8.2 *Estimation* og Afsnit 8.3 *Test for hypotesen $p = 0$* .

(Fredag den 20. december) Forelæsningen dækkede Afsnit 8.4-6 og 8.9. Eksempel 8.2 blev brugt til at illustrere anvendelsen af teknikkerne fra Afsnit 8.3 og 8.4: Test af $p = 0$, konfidensinterval for korrelationskoefficient, test af identitet af to korrelationskoefficienter, fælles estimat for korrelationskoefficient baseret på $k = 2$ uafhængige stikprøver, konfidensinterval for korrelationskoefficient baseret på $k = 2$ uafhængige stikprøver.

Forelæsningerne i denne uge

Som det fremgår af programmet for de sidste forelæsninger i december, lærte I før jul at håndtere korrelation i én, to og flere stikprøver fra den todimensionale normalfordeling. Men vi mangler ~~at~~ om middelværdien i den todimensionale normalfordeling: Test af simpel hypotese om middelværdien i én stikprøve fra en normalfordeling, herunder konfidensinterval for middelværdierne i en todimensional normalfordeling, sammenligning af middelværdierne i to todimensionale normalfordelinger baseret på to uafhængige stikprøver.

(Onsdag den 5. februar) Preben Blæsild vil forelæse over så meget han kan nå af Afsnit 8.7 og Afsnit 8.8.

(Fredag den 7. februar) Jeg fortsætter hen mod at afslutte kapitlet om den todimensionalfordeling. Anneks til Kapitel 8 om beregninger i SAS mangler jeg at omtale.

Øvelserne i denne uge

(Uge 6: 3.-9. februar) Eksamenssættet til eksamen i Statistik α fra fredag den 31. januar 2003.

Øvelserne i de næste uger

(Uge 7: 10.-16. februar) 8.1, 8.3, 8.5, 8.6, 8.8.

(Uge 7: 17.-23. februar) (foreløbigt program:) 8.9, 8.10, 8.11, 8.12, 8.13, 8.14

Opgave 1

Data nedenfor viser antal studenter, der for nylig blev optaget på 2. dels studierne ved Berkeley universitet, opdelt efter køn. For det naturvidenskabelige fakultet var resultatet

<i>Naturvidenskab</i>	<i>Mænd</i>	<i>Kvinder</i>	<i>Total</i>
<i>Optaget</i>	198	52	250
<i>Afvist</i>	242	61	303
<i>Total</i>	440	113	553

mens for det humanistiske fakultet viste opgørelsen

<i>Humaniora</i>	<i>Mænd</i>	<i>Kvinder</i>	<i>Total</i>
<i>Optaget</i>	16	79	95
<i>Afvist</i>	109	431	540
<i>Total</i>	125	510	635

- (1) Vis, inden for hvert fakultet, at det kan antages, at sandsynligheden for optagelse er den samme for mandlige og kvindelige studerende.
- (2) Undersøg, om sandsynligheden for optagelse kan antages at være den samme ved det naturvidenskabelige og humanistiske fakultet.
- (3) Kan det antages, at sandsynligheden for optagelse ved det naturvidenskabelige fakultet er 0.5?

Uden opdeling efter fakultet ville opgørelsen i stedet se ud som følger

	<i>Mænd</i>	<i>Kvinder</i>	<i>Total</i>
<i>Optaget</i>	214	131	345
<i>Afvist</i>	351	492	843
<i>Total</i>	565	623	1188

I denne tabel er 38% af mændene og 21% af kvinderne blevet optaget. Det ser således ud til, at kvinderne er blevet diskrimineret, på trods af konklusionen i spørgsmål (1).

- (4) Kan du give en forklaring på dette tilsyneladende paradoks?

Opgave 2

Ved EM i kvindehåndbold i december 2002 blev der i mellemrunden spillet $n = 18$ kampe. Antallet af mål i disse kampe var:

51 56 48 63 55 46 36 52 53 56 47 68 55 58 48 50 49 47

- (1) Vis ved at betragte Fishers dispersionsindeks, at det kan antages, at antallet af mål i de 18 kampe er Poisson fordelt.
(Ved beregningerne kan det benyttes, at summen og kvadratsummen af antallet af mål er henholdsvis $S = 938$ og $USS = 49732$.)

I den indledende runde blev der lavet 1219 mål i 24 kampe.

- (2) Vis, at det kan antages, at middelværdien af antallet af mål per kamp er den samme i de to runder.
- (3) Angiv estimat og 95% konfidensinterval for middelværdien af antallet af mål per kamp i de ialt 42 kampe i den indledende runde og i mellemrunden.

Opgave 3

Knoglemarv				Blodprøve		
1.53	2.87	3.50	4.03	2.74	3.72	5.02
1.96	3.12	3.58	4.05	3.05	3.72	5.22
2.05	3.15	3.66	4.29	3.10	3.73	5.39
2.21	3.15	3.71	4.38	3.28	3.80	5.43
2.23	3.15	3.82	4.47	3.33	3.83	5.48
2.48	3.26	3.89	4.50	3.44	3.88	5.74
2.56	3.29	3.91	4.60	3.48	4.08	6.52
2.61	3.44	3.97	4.64	3.52	4.31	7.03
2.66	3.44	3.99	5.14	3.65	4.38	
2.70	3.46	4.01	5.16	3.67	4.57	
2.76	3.47	4.02		3.69	4.59	

Tabel 1: DNMT1 målt på 43 prøver af knoglemarv og på 30 blodprøver.

Data til denne opgave er en del af et større materiale som er stillet til rådighed af Anni Aggerholm, Århus Amtssygehus. Undersøgelsen drejer sig om sammenhængen mellem nogle bestemte gener og nogle undergrupper af blodsygdommen *akut myeloid leukæmi* (AML).

Ved en kompliceret genteknologisk teknik måles ekspressionsniveauet for DNA methyltransferase 1, 3A og 3B på prøver af enten blod eller knoglemarv. Disse målinger betegnes henholdsvis DNMT1, DNMT3A og DNMT3B.

Ideelt skulle alle målinger foretages på knoglemarv, men man har ikke prøver af knoglemarv for alle patienter, så første skridt i undersøgelsen bliver at undersøge om målinger på blodprøver måler det samme som målinger på knoglemarv. I Tabel 1 er gengivet målingerne af DNMT1 på 43 prøver af knoglemarv og på 30 blodprøver. Alle prøver er fra forskellige AML patienter, som alle er multiplex negative, det vil sige at de ikke har nogle bestemte genetiske forandringer.

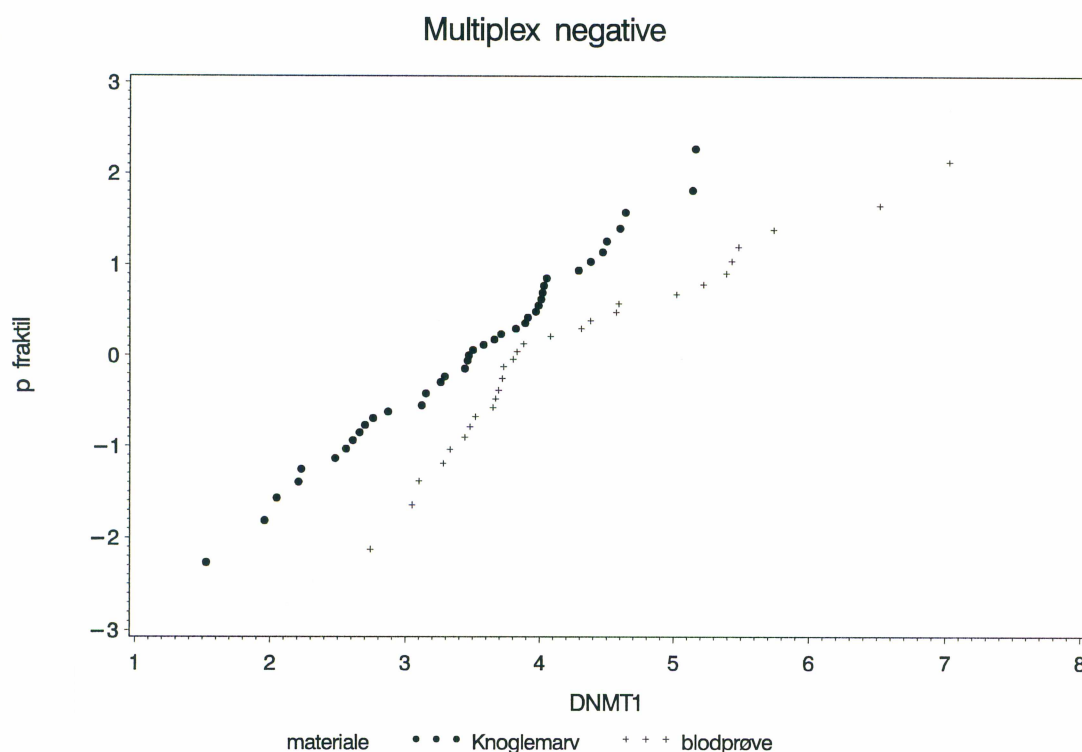
I Tabel 2 er gengivet nogle beregnede størrelser, som kan bruges til besvarelse af nogle af spørgsmålene i det følgende. Desuden er i Figur 1 gengivet fraktildiagrammer for data i Tabel 1.

- (1) Undersøg, om det kan antages, at variansen på DNMT1 er den samme for knoglemarv som for blodprøver.

- (2) Undersøg, om det kan antages, at middelværdien på DNMT1 er den samme for knoglemarv som for blodprøver.

	<i>Knoglemarv</i>	<i>Blodprøve</i>	<i>I alt</i>
<i>n</i>	43	30	73
<i>S</i>	148.87	127.39	276.26
<i>USS</i>	545.9693	573.0173	1118.9866

Tabel 2: Standardberegninger på DNMT1 målingerne i Tabel 1.

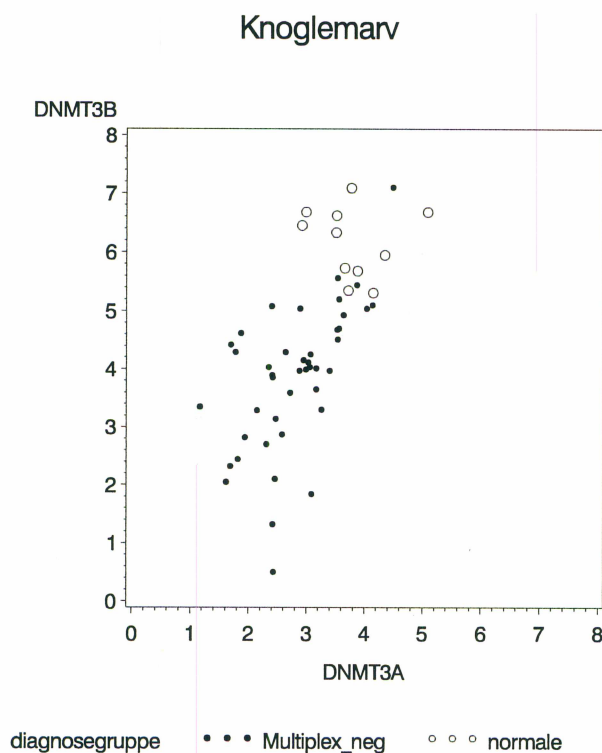


Figur 1: Fraktildiagrammer for DNMT1 målt på knoglemarv og på blodprøver.

Svaret var det samme for de to andre variable, DNMT3A og DNMT3B, så i det følgende betragtes kun målinger på knoglemarv. Man er interesseret i hvordan sammenhængen er mellem DNMT1, DNMT3A og DNMT3B i forskellige grupper af patienter og hos normale personer. I Figur 2 er vist sammenhørende værdier af DNMT3A og DNMT3B i de to diagnosegrupper multiplex negative og normale, og i Tabel 3 er nogle beregnede størrelser på de to variable DNMT3A og DNMT3B i de to diagnosegrupper.

I besvarelsen af følgende spørgsmål må det antages, at DNMT3A og DNMT3B er todimensionalt normalfordelt inden for hver af diagnosegrupperne.

- (3) Estimer korrelationskoefficienten mellem DNMT3A og DNMT3B i diagnosegrupperne og test for begge grupper, om det kan antages at korrelationen er 0.
- (4) Lav et 95% konfidensinterval for korrelationskoefficienten mellem DNMT3A og DNMT3B i gruppen af multiplex negative.



Figur 2: Sammenhørende værdier af DNMT3A og DNMT3B i de to diagnosegrupper multiplex negative og normale.

	<i>Multiplex negative</i>		<i>Normale</i>	
	DNMT3A	DNMT3B	DNMT3A	DNMT3B
<i>n</i>	43		11	
<i>SSD</i>	24.02285	65.63779	3.71016	3.56720
<i>SPD</i>	23.55552		-0.53560	

Tabel 3: Standardberegninger for DNMT3A og DNMT3B i diagnosegrupperne multiplex negative og normale.

Opgave 4

Fra en fabrikation af ståltråd udtages løbende stikprøver og trækstyrken måles. Fabrikanten garanterer en strækstyrke på mindst 420 N.

Af 69 stikprøver var 6 under specifikationsgrænsen på 420 N.

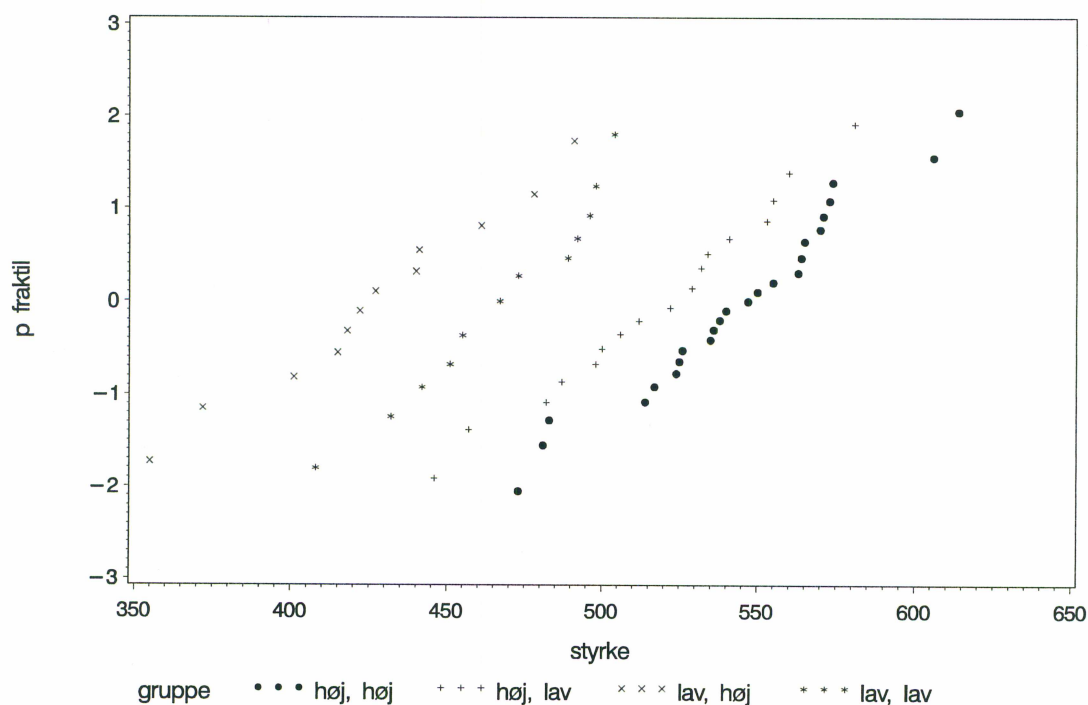
- (1) Angiv skøn og konfidensinterval for hvor stor en andel af produktionen, der ikke opfylder specifikationen.

Heldigvis havde man registreret detaljer om produktionsbetingelserne for de 69 stikprøver, og de viste at to faktorer, som man forventede ville have betydning for strækstyrken, havde varieret under produktionen. Det drejede sig om ståltemperaturen før hærkning og temperaturen i det oliebad, hvori ståltråden blev hærket. De to faktorer vil blive benævnt som henholdsvis *stål* og *hærkning*. Det viser sig at de to faktorer har været på to niveauer, som for begge faktorer betegnes med *høj* og *lav*. Hvis man opdeler strækstyrkerne efter niveauerne af stål og hærkning, får man resultatet i Tabel 4. Data er også illustreret gennem fraktildiagrammet i Figur 3.

Niveauerne af stål og hærkning							
<i>høj lav</i>		<i>høj høj</i>		<i>lav lav</i>		<i>lav høj</i>	
446	529	473	550	408	467	355	427
457	529	481	555	432	473	372	440
482	532	483	563	442	489	401	441
487	534	514	564	451	492	415	461
498	541	517	564	455	496	418	478
500	553	524	565	455	498	422	491
506	555	525	570	467	504		
512	560	526	571				
522	581	535	573				
		536	574				
		538	606				
		540	614				
		547					

Tabel 4: 69 trækstyrker opdelt efter de fire forskellige kombinationer af niveauerne af stål og hærkning.

I de følgende spørgsmål kan SAS udskrifterne benyttes. Data er indlæst i datasættet *staaal-traad* og værdierne af styrke, stål og hærkning er i variablene *styrke*, *staaal* og *haerdning*.



Figur 3: Fraktildiagrammer for trækstyrken for de fire forskellige kombinationer af niveauerne af stål og hærkning.

- (2) Undersøg, om det kan antages, at variansen på strækstyrken er den samme for de fire grupper, der er defineret ved de fire kombinationer af stål og hærkning. Standardberegninger er givet i tabellen nedenfor.

staal*haerdning	ni	Si	USSi	Si2/ni	SSDi	fi	Estimeret varians
høj lav	18	9324	4851464	4829832.00	21632.00	17	1272.4706
høj høj	25	13608	7436404	7407106.56	29297.44	24	1220.7267
lav lav	14	6529	3055031	3044845.79	10185.21	13	783.4780
lav høj	12	5121	2202739	2185386.75	17352.25	11	1577.4773
	69	34582	17545638	17467171.10	78466.90	65	1207.1831

- (3) Vis, at der er vekselvirkning mellem faktorerne stål og hærkning.
- (4) Hvilken indstilling af faktorerne stål og hærkning giver den største strækstyrke? Angiv 95% konfidensinterval for middelværdien af strækstyrken ved denne indstilling.
- (5) Angiv et 99% prediktionsinterval for en ny observation hvis både stål og hærkning er på højt niveau.

```
PROC GLM DATA=staaltraad;
CLASS staal haerdning;
MODEL styrke= staal*haerdning/SS1 SOLUTION;
RUN;QUIT;
```

giver følgende udskrift

The SAS System
 The GLM Procedure

Dependent Variable: styrke

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	135073.6464	45024.5488	37.30	<.0001
Error	65	78466.9043	1207.1831		
Corrected Total	68	213540.5507			

R-Square	Coeff Var	Root MSE	styrke Mean
0.632543	6.932431	34.74454	501.1884

Source	DF	Type I SS	Mean Square	F Value	Pr > F
staal*haerdning	3	135073.6464	45024.5488	37.30	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	466.3571429 B	9.28586923	50.22	<.0001
staal*haerdning høj høj	77.9628571 B	11.59804695	6.72	<.0001
staal*haerdning høj lav	51.6428571 B	12.38115897	4.17	<.0001
staal*haerdning lav høj	-39.6071429 B	13.66842941	-2.90	0.0051
staal*haerdning lav lav	0.0000000 B	.	.	.

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

```

PROC GLM DATA=staaltraad;
CLASS staal haerdning;
MODEL styrke= staal haerdning/SS1;
RUN;QUIT;

```

giver følgende udskrift

The SAS System
The GLM Procedure

Dependent Variable: styrke

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	117710.1427	58855.0713	40.53	<.0001
Error	66	95830.4081	1451.9759		
Corrected Total	68	213540.5507			

R-Square	Coeff Var	Root MSE	styrke Mean
0.551231	7.602890	38.10480	501.1884

Source	DF	Type I SS	Mean Square	F Value	Pr > F
staal	1	117687.6348	117687.6348	81.05	<.0001
haerdning	1	22.5079	22.5079	0.02	0.9013

STAT 1, TØ nr. 16

Stat α, V02/03

opgave 1

1)

Jf a)-d) s. 327 er #optagne hh. mænd og kvinder på hh. nat. og hum. fakultet binomialfordelt (#ansøgere indeholder ikke information i forhold til den aktuelle problemstilling, derfor er Poissonmodellen jf. bm. s. 415m irrelevant)

For at 'vise' det enskede skal vi teste om vi kan reducere til to binomialford., en for hvert fakultet:

Nat:

Føventide # (jf. (6.47) s. 352)

	$\overset{M}{198,92}$	$\overset{K}{51,08}$
O	198,92	51,08

A	241,08	61,92
---	--------	-------

≥ 5 derfor fæls

$$E(X_{nat}) = 1 - F_{\chi^2(1)}(-2 \ln Q(X_{nat})) \geq 1 - 0,2 = 0,8 \quad (6.49 \text{ s. } 35)$$

Huv jf. (6.48 s. 352):

$$\begin{aligned} -2 \ln Q(X_{nat}) &= 2 \cdot (198 \cdot \ln(\frac{198}{198,92}) + 52 \cdot \ln(\frac{52}{51,08}) \\ &+ 242 \cdot \ln(\frac{242}{241,08}) + 61 \cdot \ln(\frac{61}{61,92})) = 0,03798 \end{aligned}$$

Hmm :

Forventede #

	M	K
O	18,70	76,30
A	106,30	433,70

≥ 5 derfor fås

$$E(X_{\text{hum}}) = 1 - F_{\chi^2(1)}(-2 \ln Q(X_{\text{hum}})) \geq 1 - 0,6 = 0,4$$

$$\text{Hvor } -2 \ln Q(X_{\text{hum}}) = 2 \left(16 \ln \left(\frac{16}{18,7} \right) + 79 \ln \left(\frac{79}{76,3} \right) + 109 \ln \left(\frac{109}{106,3} \right) + 431 \ln \left(\frac{431}{433,7} \right) \right) = 0,58937$$

Da testst. er større end 0,05 kan vi ikke afvise, at ss. for optagelse er den samme for mænd som for kvinder indenfor hvert fakultet.

2)

	N	H	Σ
O	250	95	345
A	303	540	843
Σ	553	635	1188

Jvf. 1) kan vi opfatte # optagne som værende binomial fordelt for hh. Nat. og Hum. fakultet.

Vi teste homogenitet af nat. og hum. fnd. :

Forventede # (jvf. (6.47) s. 352)

	N	H
O	160,59	184,41
A	392,41	450,59

≥ 5 derfor fås

$$E(X) = 1 - F_{\chi^2(1)}(-2 \ln Q(x)) < 0,05$$

(6.49 s.35)

Hvor (jvf. 6.48 s.352):

$$\begin{aligned} -2 \ln Q(x) &= 2 \left(250 \cdot \ln \left(\frac{250}{160,59} \right) + 95 \cdot \ln \left(\frac{95}{184,41} \right) + 303 \cdot \ln \left(\frac{303}{392,41} \right) \right. \\ &\quad \left. + 540 \cdot \ln \left(\frac{540}{450,59} \right) \right) = 134,08 \end{aligned}$$

Dus, hypotesen om homogenitet forkastes, altså er ss. for optagelse ikke den samme på nat. og hum.

3)

Under hypotesen om ss. for optagelse = 0,5 fås forventede # (6.16 s.334)

$$\begin{array}{l} O \quad 276,5 \\ A \quad 276,5 \end{array} \quad \geq 5 \quad \text{derfor fås}$$

$$E(X) = 1 - F_{\chi^2(1)}(-2 \ln Q(x)) < 0,05 \quad (6.18 s.335)$$

$$\begin{aligned} \text{Hvor } -2 \ln Q(x) &= 2 \cdot \left(250 \cdot \ln \left(\frac{250}{276,5} \right) + 303 \cdot \ln \left(\frac{303}{276,5} \right) \right) \\ &= 5,0874 \end{aligned}$$

Dus hypotesen forkastes, dvs. ss. for at blive optaget på nat. er ikke 0,5.

4)

Juf. (6.12 s. 332) er estimationen for ss for optagelse på hh. nat. og hum. $\frac{250}{553} = 0,45$ og

$\frac{95}{635} = 0,15$, da andelen af mænd som

søgr nat. er $\frac{440}{553} = 0,80$ og andelen af

kvinder som søgr hum er $\frac{510}{635} = 0,80$ er

forklaringen på 'paradokset', at mændene søg ind hvor det er 'let' at blive optaget, mens kvinderne søg ind hvor det er 'svært'.

opgave 2

1)

$$t = \frac{s^2}{\bar{x}} = \frac{\frac{1}{n-1} (VSS - \frac{s^2}{n})}{\frac{1}{n} s} = 0,96150$$

(7.14) s. 399

Da

$$\chi^2_{0,025}(17)/(17) = 0,4450$$

$$\chi^2_{0,975}(17)/(17) = 1,7759$$

accepteres modellen (obs. poisson-fordelt) jf.
(7.15) s. 399.

2)

Jf. 1) har vi modellen

$$M_0: X_{ij} \sim \text{po}(\lambda_i) \quad i = \text{runde}, j = \text{kamp}$$

Vi ønsker, at teste

$$H_0: \lambda_i = m_i \lambda, \quad m_i = \# \text{ kampe i runde } i.$$

De forventede antal i hh. den indl. runde og
melleml. runden er (7.29 s. 404)

$$\frac{24}{18+24} (938+1219) = 1232,57, \quad \frac{18}{18+24} (938+1219) = 924,43$$

$$E(X) = 1 - F_{\chi^2(1)}(-2 \ln Q(X)) \geq 1 - 0,5 = 0,5 \quad (7.31 \text{ s. 404})$$

$$\text{hvor } -2 \ln Q(X) = 2 \left(938 \ln \left(\frac{938}{924,43} \right) + 1219 \ln \left(\frac{1219}{1232,57} \right) \right) = 0,34818$$

(jf. 7.30 s. 404)

Des. hypotesen forkastes ikke, og vi kan derfor antage at middelværdi for #mål/kamp, λ er ens i den indl- og mellemrunden.

3)

$$\hat{\lambda} = \frac{938 + 1219}{18 + 24} = 51,36$$

(7.28)
s. 404

Konfidensint. for λ (mf. 7.35 + 7.36 s. 405):

$$[49,23 ; 53,57]$$

$$\text{idet } x_1 = 2157, m_1 = 42, u_{0,975} = 1,960$$

opgave 3

1)

Som det fremgår af fraktildiagrammene i Figur 1 kan vi formode, at data er normalfordelt.

$$M_b: x_{ij} \sim N(\mu_i, \sigma_i^2) \quad i = \{\text{knoglerne}, \text{blodprøve3}, j = 1, \dots, n_i\}$$

Fraktildiagrammene peger på at de to fordelinger har samme varians (ens hældning) men forskellig middelværdi (forskellig afskæring).

$$H_{0b}: \sigma_k^2 = \sigma_b^2$$

$$\hat{\sigma}_k^2 \leftarrow s_k^2 = \frac{1}{n_k - 1} (USS_k - \frac{s_k^2}{n_k}) = 0,72780$$

$$\hat{\sigma}_b^2 \leftarrow s_b^2 = \frac{1}{n_b - 1} (USS_b - \frac{s_b^2}{n_b}) = 1,10610$$

$$F = \frac{s_b^2}{s_k^2} = 1,51978$$

$$E(x) = 2(1 - F_{\underset{29}{n_b-1}, \underset{42}{n_k-1}}(F)) \geq 2(1 - 0,9) = 0,2 > 0,05$$

Dvs. H_0 forkastes ikke, altså kan vi reducere til

$$M_i: x_{ij} \sim N(\mu_i, \sigma^2) \quad (i, j \text{ som før})$$

$$\sigma^2 \leftarrow s_1^2 = \frac{f_b s_b^2 + f_k s_k^2}{f_b + f_k} = 0,88232$$

$$\mu_k \leftarrow \frac{s_k}{n_k} = 3,46209$$

$$\mu_b \leftarrow \frac{s_b}{n_b} = 4,24633$$

2)

$$H_{0H}: \mu_k = \mu_b$$

$$t(x) = \frac{\bar{x}_b - \bar{x}_k}{\sqrt{s_1^2 \left(\frac{1}{n_k} + \frac{1}{n_b} \right)}} = 3,50970, \quad f_1 = f_k + f_b = 71$$

$$E(x) = 2(1 - F_{t(f_1)}(|t(x)|)) \leq 2(1 - 0,9995) < 0,05$$

Jus. hypotesen forkastes, altså bestemmer M_1 data bedst.

Altså kan man ikke umiddelbart bruge målinger fra knoglerne og blod i flæng.

3)

Jf. oplysningerne i opg. og figuren på side 5, kan vi antage at startmodellen er

$$H_0: X_{ij} \sim N_2(\mu_i, \Sigma_i) \quad i = \{\text{multi.}, \text{norm. 3}\}, j = 1, \dots, n_i$$

Figuren peger på at korrelationen er 0

$$r_m \leftarrow \frac{SPD}{\sqrt{SSD_x SSD_y}} = 0,01494 = r_m$$

$$r_n \leftarrow -0,04047 = r_n$$

$$H_{0n}: r_n = 0$$

$$T = \sqrt{n_n - 2} \frac{r_n}{\sqrt{1 - r_n^2}} = -0,12131$$

$$E(x) = 2(1 - F_{t(n_n - 2)}(|T|)) \geq 2(1 - 0,6) = 0,8 > 0,05$$

$$H_{0m}: \rho_m = 0$$

$$T = \sqrt{n_m - 2} \frac{r_m}{\sqrt{1 - r_m^2}} = 0,09567$$

$$E(x) = 2(1 - F_{t(n_m - 2)}(|T|)) \geq 2(1 - 0,6) = 0,8 > 0,05$$

Altså kan vi ikke afvise at korrelationen er 0

4)

Jvf. s. 8.7-8.10 i Bnd 2 af notation er et 95% konf. int. for ξ_m :

$$\left] z - u_{0,975} \frac{1}{\sqrt{n_m - 3}}, z + u_{0,975} \frac{1}{\sqrt{n_m - 3}} \right[= \left] -0,29496, 0,32484 \right[$$

$$z = \hat{\xi}_m = \frac{1}{2} \ln \frac{1 + r_m}{1 - r_m} = 0,01494$$

$$u_{0,975} = 1,960$$

lignethet (jvf. s. 8.10) giver et 95% konf. int. for ρ_m :

$$\left] -0,28670; 0,31388 \right[$$

$$\text{idet } \rho = \frac{e^{2\xi} - 1}{e^{2\xi} + 1}$$

opgave 4

1)

Juf. a) - d) s. 327 er antal stikprøver som ikke opfylder specifikationen binomialfordelt
 $X_i \sim b(n, p)$, $n = 69$

$$p \leftarrow \frac{6}{69} \approx 0,08696 \quad (6.12 \text{ s. } 332)$$

For at normalapprox., som bruges til at angive et konf. int. s. 338, kan anvendes skal vi if. bm. 6.9 helst have: $np(1-p) > 9$, men
 $n\hat{p}(1-\hat{p}) = 5,48 \leq 9$

Af (6.28) + (6.29) s. 338 fås ($u_{0,975} = 1,960$)
at 95% konf. int. for p er:

$$]0,04046; 0,17702[$$

2)

Juf. oplysningerne i opg. herunder Figur 3 kan vi antage flg. startmodel:

$$M_0: X_{ijk} \sim N(\mu_{ij}, \sigma_{ij}^2) \quad i = \text{stål nr.}, j = \text{hærd nr.}, k = 1, \dots, n_{ij}$$

Fraktildiagrammet peger på at varianserne er ens, og at trædstyrkerne er forskellige.

$$H_{02} : \sigma_{ij}^2 = \sigma^2$$

$$-2 \ln Q(x) = f_1 \ln s^2 - \sum_{ij} f_{ij} \ln s_{ij}^2 = 1,57385$$

$$C = 1 - \frac{1}{3(k-1)} \left(\sum \frac{1}{f_{ij}} - \frac{1}{f_1} \right) = 1,02810$$

$$B_a = \frac{-2 \ln Q(x)}{C} = 1,47247$$

$$E(x) = 1 - F_{\chi^2(3)}(B_a) \geq 1 - 0,4 = 0,6 > 0,05$$

Juf. s. 113

Dvs. vi har ikke grund til at forkaste hypotesen.

Vi reducerer til

$$M_i : x_{ijk} \sim N(\mu_{ij}, \sigma^2)$$

3)

$$H_0 : \mu_{ij} = \alpha_i + \beta_j \quad (\text{altre hypotese om ingen vekselvirk.})$$

$$F = \frac{\frac{95830,4081 - 78466,9043}{66 - 65}}{1207,1931} = 14,38349 \quad ((4.12) \text{ s. 209})$$

$$E(x) = 1 - F_{F(1,65)}(F(x)) \leq 1 - 0,9995 = 0,0005 < 0,05$$

Dvs. hypotesen forkastes, og vi kan konkludere, at der er vekselvirkning mellem faktorene.

4)

J.f. estimation for x_{ij} s. 8

gives et højt niveau stål og hærdering den bedste strækstyrke; skønnet for strækstyrken er $466,35714 + 77,96286 = 544,32$

Konfidensint. (95%) er (j.f. b.m. s. 160)

$[521,15870; 567,48130]$

idet $t_{0,995}(65) \approx 1,997$, $\text{Std. err.} = 11,59804695$

5)

Analogt med overvejelserne s. 147 fås, at et 99% pred.int. er

$$[\hat{\mu} - \sqrt{s^2(1+\frac{1}{n})} t_{0,995}(f_1); \hat{\mu} + \sqrt{s^2(1+\frac{1}{n})} t_{0,995}(f_1)]$$

$$= [422,16605; 666,47395]$$

$\hat{\mu} = 544,32$, $s^2 = 1207,1831$, $n = 25$, $f_1 = 65$, $t_{0,995}(f_1) = 3,4475$
(j.f. def. s. 180 på t-fød.)