

# **EN INTRODUKTION TIL STATISTIK**

**BIND 2A**

**Knut Conradsen**


**4. UDGAVE**

**LYNGBY 1984**

**FORELÆSNINGSNOTE**

**imsot**

© Knut Conradsen 1978 , 1979, 1984

Trykt af  - DTH  
Bogbinder Hans Meyer

## FORORD

Efter at forskellige præliminære versioner er blevet afprøvet ved undervisning af polytekniske studerende og ved efteruddannelse af civilingeniører, foreligger hermed 2.den del af min fremstilling af den matematiske og anvendte statistik med specielt henblik på anvendelser inden for ingeniørvidenskaberne og dermed beslægtede felter. Af tekniske grunde udgives værket i 2 bind.

Fremstillingen kan nok i visse kapitler virke lidt for summarisk, blot refererende "tekniske" resultater; men det er selvsagt min hensigt så vidt muligt at ændre dette i kommende udgaver. Forslag til forbedring fra læsere er derfor meget velkomne.

På den anden side kan behandlingen af en række emner af pladsmæssige grunde kun blive af oversigtsmæssig karakter, og den interesserede læser må der henvises til litteraturen. De fleste steder er i selve teksten anført kilder og referencer til de behandlede emner; men ofte er supplerende læsning alene anført i referencelisterne.

Jeg vil gerne her benytte lejligheden til at takke kollegaer ved IMSOR for nyttige diskussioner vedrørende indhold og form, ligesom en række Højskoleinstitutter samt en række eksterne institutioner, virksomheder og personer takkes for at have stillet data til rådighed for eksempler. "Dataleverandøren" er som regel nævnt i forbindelse med de pågældende eksempler. Det gælder dog ikke eksempler vedrørende analyse af luftforurenende komponenter. De bygger næsten alle på data, der venligst er stillet til rådighed af Kraftværkernes Forureningsudvalg og Dansk Kedelforening.

Nævnes må også det store arbejde, som IMSORs sekretariat har haft ved maskinskrivningen og mangfoldiggørelsen af bogen. Det gælder især Kirsten Nielsen, Kirsten Studnitz og Bente Wilkenschildt samt Svend Aage Jensen og Jørgen Folkenberg.

Knut Conradsen

#### FORORD til anden udgave

I denne udgave er kapitel 9 blevet omarbejdet totalt. I de øvrige kapitler er kun foretaget rettelser af trykfejl og sket mindre redaktionelle og typografiske ændringer. Jeg har modtaget en række forslag til yderligere forbedringer af bogen, men har desværre ikke kunnet nå at inkorporere alle disse i denne udgave. Dette planlægges dog gjort i senere udgaver. Eventuelle forslag o.l. modtages med tak. Slutteligen må nævnes det store arbejde, som Birgit Pettersson og især Bente Wilkenschildt har haft med de typografiske ændringer og den omtalte omarbejdelse af bogen.

Knut Conradsen

#### FORORD til tredje udgave

I denne udgave er rettet en del trykfejl.

Knut Conradsen

#### FORORD til fjerde udgave

I denne udgave er foretaget nogle mindre ændringer.

Knut Conradsen

## INDHOLDSFORTEGNELSE

1	Resumé af den lineære algebra	1.1-1.70
1.1	Vektorrum	1.2
1.1.1	Definition af vektorrum	1.2
1.1.2	Direkte sum af vektorrum	1.6
1.2	Lineære afbildninger og matricer	1.8
1.2.1	Lineære afbildninger	1.8
1.2.2	Matricer	1.10
1.2.3	Lineære afbildningers matrixfremstillinger	1.12
1.2.4	Koordinattransformation	1.13
1.2.5	Rangen af en matrix	1.15
1.2.6	Determinanten af en matrix	1.17
1.2.7	Blokmatricer	1.19
1.3	Pseudoinverse eller generaliseret inverse matrix til en ikke-regulær matrix	1.22
1.4	Egenværdiproblemer. Kvadratiske former	1.35
1.4.1	Egenværdier og egenvektorer for symmetriske matricer	1.35
1.4.2	Singulær-værdi dekomposition af vilkårlig matrix. Q- og R-modus analyser	1.42
1.4.3	Kvadratiske former og positivt definitte matricer	1.46
1.4.4	Det generelle egenværdiproblem for symmetriske matricer	1.54
1.4.5	Sporet af en matrix	1.58
1.4.6	Differentiation af linearform og kvadratisk form	1.59
1.5	Tensor- eller Kronecker produkt af matricer	1.63
1.6	Indre produkter og normer	1.64
2	Flerdimensionale variable	2.1-2.69
2.1	Momenter af flerdimensionale variable	2.1
2.1.1	Middelværdi	2.1
2.1.2	Dispersionsmatricen	2.3
2.1.3	Kovarians	2.6

2.2	Den flerdimensionale normalfordeling	2.10
2.2.1	Definition og simple egenskaber	2.10
2.2.2	Uafhængighed og konturellipsoider	2.18
2.2.3	Betingede fordelinger	2.22
2.2.4	Reproduktivitetssætning og central grænseværdisætning	2.23
2.2.5	Estimation af parametre i en flerdimensional normalfordeling	2.25
2.2.6	Den todimensionale normale fordeling	2.27
2.3	Korrelation og regression	2.34
2.3.1	Den partielle korrelationskoefficient	2.35
2.3.2	Den multiple korrelationskoefficient	2.44
2.3.3	Regression	2.49
2.4	Spaltningssætningen	2.53
2.5	Wishart fordelingen og den generaliserede varians	2.59
2.6	Lidt om estimation af flerdimensionale parametre	2.64
3	Den generelle lineære model	3.1-3.60
3.1	Estimation i den generelle lineære model	3.1
3.1.1	Modelformulering	3.1
3.1.2	Estimation i det regulære tilfælde	3.5
3.1.3	Tilfældet $\underline{x}'\Sigma^{-1}\underline{x}$ singular	3.12
3.1.4	Estimation under bibetingelser	3.25
3.1.5	Konfidensintervaller for forudsagte værdier. Prediktionsinterval	3.32
3.2	Test i den generelle lineære model	3.40
3.2.1	Test for lavere dimension af modelrum	3.40
3.2.2	Successiv testning i den generelle lineære model	3.48
4	Regressionsanalyse	4.1-4.73
4.1	Lineær regressionsanalyse	4.1
4.1.1	Notation og model	4.2
4.1.2	Korrelation og regression	4.6
4.1.3	Analyse af forudsætninger	4.8
4.2	Regression efter ortogonale polynomier	4.14
4.2.1	Definition og modelformulering	4.14
4.2.2	Bestemmelse af ortogonale polynomier	4.19
4.3	Valg af "bedste" regressionsligning	4.27
4.3.1	Problemstillingen	4.27
4.3.2	Undersøgelse af samtlige regressioner	4.30
4.3.3	Backwards elimination	4.32
4.3.4	Forward selection	4.34
4.3.5	Stepwise regression	4.38
4.3.6	Nogle eksisterende programmer	4.41
4.3.7	Numerisk appendix	4.42

4.4	Andre regressionsmodeller og -løsninger	4.50
4.4.1	Ortogonal regression (lineær funktionel relation)	4.50
4.4.2	Ridge-regression	4.56
4.4.3	Ikke-lineær regression og kurvetilpasning	4.66
5	Variansanalyser	5.1-5.83
5.1	Indledning	5.1
5.2	Ensidet variansanalyse	5.3
5.2.1	Modeller	5.3
5.2.2	Analyse af den systematiske model	5.7
5.2.3	Analyse af den tilfældige model	5.11
5.2.4	Resumé af analyserne og et eksempel	5.14
5.3	Tosidet variansanalyse. Hierarkisk klassifikation og krydsklassifikation	5.18
5.3.1	Hierarkisk klassifikation og krydsklassifikation	5.18
5.3.2	Analyse af hierarkisk klassificerede data	5.24
5.3.3	Analyse af krydsklassificerede data	5.31
5.4	Variansanalysemodeller med 3 faktorer	5.39
5.5	Variansanalyser med flere faktorer	5.52
5.5.1	Estimation af parametre og beregning af kvadratafgivelsessummer	5.52
5.5.2	Beregning af forventede værdier af middeltkvadratafgivelsessummer	5.58
5.6	Variansanalyseprogrammet ANOVA	5.78
6	Tests i den flerdimensionale normale fordeling	6.1-6.44
6.1	Test for middelværdier	6.1
6.1.1	Hotelling's $T^2$ i enstikprøvesituationen	6.1
6.1.2	Hotelling's $T^2$ i tostikprøvesituationen	6.8
6.2	Den flerdimensionale generelle lineære model	6.13
6.3	Variansanalyser for flerdimensionale variable	6.28
6.3.1	Ensidet flerdimensional variansanalyse	6.28
6.3.2	Tosidet flerdimensional variansanalyse	6.30
6.4	Tests vedrørende dispersionsmatricer	6.39
6.4.1	Tests vedrørende en enkelt dispersionsmatrix	6.39
6.4.2	Test for, at flere dispersionsmatricer er ens	6.43
7	Diskriminantanalyse	7.1-7.51
7.1	Diskrimination mellem 2 populationer	7.1

7.1.1	Bayes- og minimaxløsninger .	7.1
7.1.2	Diskrimination mellem 2 normale populationer	7.4
7.1.3	Diskrimination med ukendte parametre	7.15
7.1.4	Test for bedste diskriminatorfunktion	7.19
7.1.5	Test for yderligere information	7.22
7.2	Diskrimination mellem flere populationer	7.24
7.2.1	Bayesløsning	7.24
7.2.2	Bayesløsning i tilfældet med flere normale fordelinger	7.26
7.2.3	Alternativ diskriminationsprocedure i tilfældet med flere populationer	7.32
7.3	Nogle standardprogrammer til beregning af lineære diskriminatorer	7.36
8	Principale komponenter, kanoniske variable og korrelationer samt faktoranalyse	8.1-8.64
8.1	Principale komponenter	8.2
8.1.1	Definition og simple egenskaber	8.2
8.1.2	Estimation og testning	8.7
8.2	Kanoniske variable og korrelationer	8.17
8.3	Faktoranalyse	8.20
8.3.1	Model og forudsætninger	8.21
8.3.2	Estimation af faktorer (faktorvægte)	8.24
8.3.3	Faktor rotation	8.27
8.3.4	Beregning af faktorværdier (factor scores)	8.33
8.3.5	Et case-study	8.39
8.3.6	Lidt om maximum likelihood faktoranalyse	8.48
8.3.7	Q-modus analyse	8.51
8.3.8	Nogle standardprogrammer	8.54
9	Statistisk analyse af tidsrækker	9.1-9.179
9.1	Fourier-transformationen, forskydningsoperatorer og lineære systemer	9.2
9.1.1	Fourier-transformationen	9.2
9.1.2	Forskydningsoperatorer	9.14
9.1.3	Tidsvarianter, lineære systemer	9.19
9.1.4	Sampling problemet	9.24
9.2	Stokastiske processer og deres momentfunktioner	9.29
9.2.1	Kort om tidsrækker og stokastiske processer	9.29
9.2.2	Den diskrete lineære proces. AR-, MA- og ARMA-processer	9.44
9.2.3	Den kontinuerte lineære proces	9.59
9.2.4	Estimation af autokovariansfunktionen	9.63
9.3	Den klassiske analyse	9.67
9.3.1	Udjævning og trend	9.67
9.3.2	Den klassiske dekomposition	9.80



9.4	Endimensional spektralanalyse	9.88
9.4.1	Spektret for en stokastisk proces	9.88
9.4.2	Estimation af (power-)spektre	9.100
9.5	Filtrering	9.116
9.6	Krydsspektralanalyse	9.132
9.6.1	Krydskovarians og krydskorrelation	9.132
9.6.2	Estimation af krydskovariansfunktion	9.136
9.6.3	Krydsspektret	9.137
9.6.4	Estimation af krydsspektret	9.141
9.6.5	Eksempel på krydsspektralanalyse	9.143
9.7	Box-Jenkins' metode	9.152
9.7.1	ARIMA-processer	9.152
9.7.2	Sæsonmodellen	9.156
9.7.3	Forudsigelser i ARIMA-processer	9.158
9.7.4	Identifikation af og estimation i en ARIMA-proces	9.164
9.7.5	Et eksempel	9.168

Index



## KAPITEL 1

Resumé af den lineære algebra

I dette kapitel skal vi give en oversigt over den lineære algebra med særlig henblik på dens anvendelse i statistikken. Kapitlet er ikke tænkt som en indføring i emnet, men kun som en genopfriskning af kendt stof. Der vil således ikke blive givet særligt mange eksempler indenfor de områder, der sædvanligt dækkes af kurser i algebra og geometri. Indenfor emner, der sædvanligvis ikke underkastes en så nøje gennemgang ved all round kurser, men som skønnes at have væsentlig betydning for en succesfuld anvendelse af algebraen i statistikken, vil der selvsagt blive givet flere eksempler og af og til også beviser.

I de senere år er man begyndt at involvere begreb som duale vektorrum i teorien for flerdimensional normal analyse. Uagtet de fordele dette medfører, har forfatteren dog valgt ikke at følge denne linie, hvorfor emnet selvsagt heller ikke berøres i dette oversigtskapitel.

I forbindelse med analyse af flerdimensionale statistiske problemer kommer man ofte ud for at skulle invertere matricer, der ikke nødvendigvis er regulære. Dette vil f.eks. være tilfældet, hvis man betragter et problem, der er givet på et ægte underrum i det betragtede  $n$ -dimensionale vektorrum. I stedet for at reducere betragtningerne til det relevante underrum fo-

retrækker mange (= de fleste) forfattere at give delvist algebraiske løsninger ved at indføre den såkaldte pseudo-inverse til en ikke-regulær matrix. For at lette læsningen af (tidsskrift-)litteraturen vil dette begreb derfor blive omtalt og søgt anskueliggjort geometrisk.

Her må det også præciseres, at anvendelsen af pseudo inverse matrixer giver en ganske bekvem måde at opstille løsninger til diverse matrixligninger i en algoritmisk form.

## 1.1 Vektorrum

Vi indleder med at give en oversigt over definitionen af og elementære egenskaber ved det helt fundamentale begreb, et (lineært) vektorrum.

### 1.1.1 Definition af vektorrum

Et vektorrum (over de reelle tal) er en mængde  $V$  forsynet med en kompositionsregel  $+ : V \times V \rightarrow V$  kaldet vektoraddition og en kompositionsregel  $\cdot : R \times V \rightarrow V$  kaldet skalarmultiplikation, som opfylder

- i)  $\forall \underline{u}, \underline{v} \in V : \underline{u} + \underline{v} = \underline{v} + \underline{u}$  (kommutative lov for vektoraddition)
- ii)  $\forall \underline{u}, \underline{v}, \underline{x} \in V : \underline{u} + (\underline{v} + \underline{x}) = (\underline{u} + \underline{v}) + \underline{x}$  (associative lov for vektoraddition)
- iii)  $\exists \underline{0} \in V \forall \underline{u} \in V : \underline{u} + \underline{0} = \underline{u}$  (eksistens af neutralt element)
- iv)  $\forall \underline{u} \in V \exists -\underline{u} \in V : \underline{u} + (-\underline{u}) = \underline{0}$  (eksistens af invers element)

- v)  $\forall \lambda \in \mathbb{R} \forall \underline{u}, \underline{v} \in V : \lambda(\underline{u} + \underline{v}) = \lambda \underline{u} + \lambda \underline{v}$  (distributive lov for skalarmultiplikation)
- vi)  $\forall \lambda_1, \lambda_2 \in \mathbb{R} \forall \underline{u} \in V : (\lambda_1 + \lambda_2) \underline{u} = \lambda_1 \underline{u} + \lambda_2 \underline{u}$  (distributive lov for skalarmultiplikation)
- vii)  $\forall \lambda_1, \lambda_2 \in \mathbb{R} \forall \underline{u} \in V : (\lambda_1 \lambda_2) \underline{u} = \lambda_1 (\lambda_2 \underline{u})$  (assosiative lov for skalarmultiplikation)
- viii)  $\forall \underline{u} \in V : 1 \underline{u} = \underline{u}$

Eksempel 1.1 Det eftervises let, at alle ordnede  $n$ -tupler

$$\underline{x} = \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix}$$

af reelle tal udgør et vektorrum, hvis kompositionerne defineres elementvis, i.e.

$$\begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n + y_n \end{bmatrix}$$

og

$$\lambda \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = \begin{bmatrix} \lambda x_1 \\ \cdot \\ \cdot \\ \cdot \\ \lambda x_n \end{bmatrix}$$

Dette vektorrum betegnes  $\mathbb{R}^n$

□

Et vektorrum  $U$  som er en delmængde af et vektorrum  $V$  kaldes et underrum i  $V$ . Betragter vi omvendt vektorer  $\underline{v}_1, \dots, \underline{v}_k \in V$ , kan vi definere

$$\text{span} \{ \underline{v}_1, \dots, \underline{v}_k \}$$

som det mindste underrum af  $V$ , der indeholder  $\{ \underline{v}_1, \dots, \underline{v}_k \}$ . Det vises let, at

$$\text{span} \{ \underline{v}_1, \dots, \underline{v}_k \} = \{ \sum_{i=1}^k \alpha_i \underline{v}_i \mid \alpha_i \in R, i = 1, \dots, k \}.$$

En vektor af formen  $\sum \alpha_i \underline{v}_i$  kaldes en linearkombination af vektorerne  $\underline{v}_i$ ,  $i = 1, \dots, k$ . Ovenstående resultat kan da udtrykkes, at  $\text{span} \{ \underline{v}_1, \dots, \underline{v}_k \}$  netop består af alle linearkombinationer af vektorerne  $\underline{v}_1, \dots, \underline{v}_k$ . Generelt defineres

$$\text{span} (U_1, \dots, U_p)$$

hvor  $U_i \subseteq V$ , som det mindste underrum af  $V$ , der indeholder alle  $U_i$ ,  $i = 1, \dots, p$ .

Ved et sideunderrum forstås en mængde af formen

$$\underline{v} + U = \{ \underline{v} + \underline{u} \mid \underline{u} \in U \},$$

hvor  $U$  er et underrum i  $V$ .

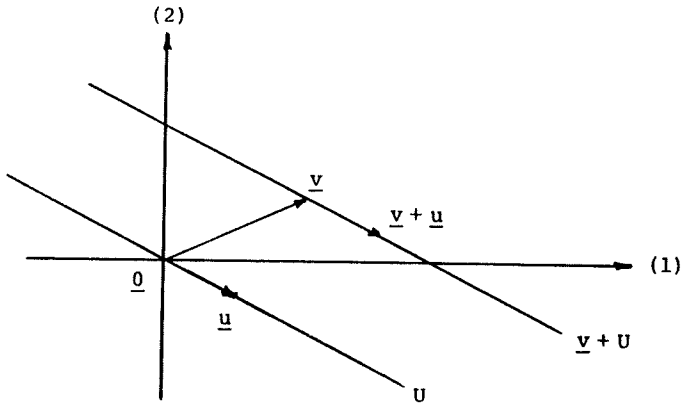
Situationen er skitseret i fig. 1.1.

Vektorer  $\underline{v}_1, \dots, \underline{v}_n$  siges at være lineært uafhængige, hvis relationen

$$\alpha_1 \underline{v}_1 + \dots + \alpha_n \underline{v}_n = 0$$

medfører, at

$$\alpha_1 = \dots = \alpha_n = 0.$$



Figur 1.1 Underrum og tilhørende sideunderrum i  $\mathbb{R}^2$ .

I modsat fald siges de at være lineært afhængige, og mindst en af dem kan da skrives som en linearkombination af de øvrige.

En basis for vektorrummet  $V$  er en mængde af lineært uafhængige vektorer, som udspænder hele  $V$ . En vilkårlig vektor har en entydig fremstilling som linearkombination af vektorer i en basis. Antallet af elementer i forskellige baser for et vektorrum er altid det samme. Hvis dette antal er endeligt kaldes det vektorrummets dimension og det skrives  $\dim(V)$ .

Eksempel 1.2  $\mathbb{R}^n$  har basen

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix},$$

og er således  $n$ -dimensionalt.

□

I en fremstilling

$$\underline{v} = \sum_{i=1}^n \alpha_i \underline{v}_i$$

hvor  $\{\underline{v}_1, \dots, \underline{v}_n\}$  er en basis for  $V$ , kaldes sættet  $\alpha_1, \dots, \alpha_n$   $\underline{v}$ 's koordinater med hensyn til basen  $\{\underline{v}_1, \dots, \underline{v}_n\}$ .

### 1.1.2 Direkte sum af vektorrum

Lad  $V$  være et (endelig dimensionalt) vektorrum og lad  $U_1, \dots, U_k$  være underrum af  $V$ . Vi siger da, at  $V$  er den direkte sum af underrummene  $U_1, \dots, U_k$ , og vi skriver

$$V = U_1 \oplus \dots \oplus U_k = \bigoplus_{i=1}^k U_i,$$

hvis en vilkårlig vektor  $\underline{v} \in V$  har netop én fremstilling af formen

$$\underline{v} = \underline{u}_1 + \dots + \underline{u}_k, \quad \underline{u}_1 \in U_1, \dots, \underline{u}_k \in U_k \quad (*)$$

Denne betingelse er ensbetydende med, at der for vektorer  $\underline{u}_i \in U_i$  gælder

$$\underline{u}_1 + \dots + \underline{u}_k = \underline{0} \Rightarrow \underline{u}_1 = \dots = \underline{u}_k = \underline{0}$$

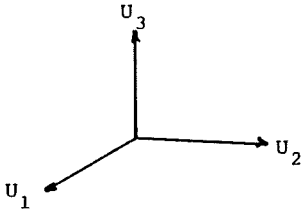
Dette er igen ensbetydende med, at

$$\dim(\text{span}(U_1, \dots, U_k)) = \sum_{i=1}^k \dim U_i = \dim V$$

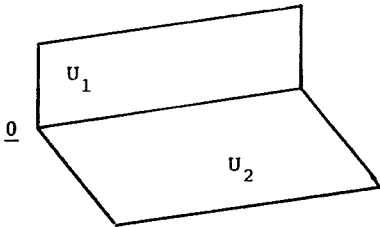
Dette er endelig ensbetydende med, at alle fællesmængder af nogle af  $U_i$ 'erne er  $\underline{0}$ . Det er selvfølgelig et gennemgående krav, at  $\text{span}(U_1, \dots, U_k) = V$ , d.v.s. at der overhovedet findes en fremstilling som (\*). Det er (\*)'s eventuelle entydighed, der bevirker, at vi eventuelt kalder "summen" direkte.



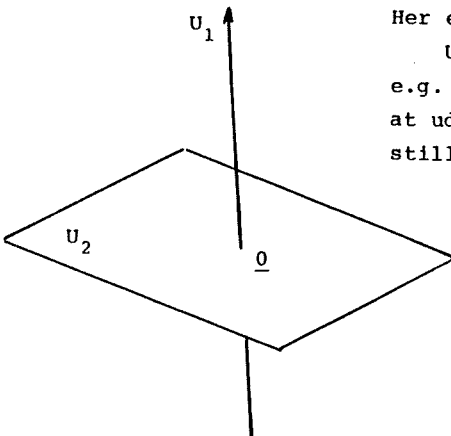
Vi skitserer nogle eksempler i nedenstående fig. 1.2



$U_1 \oplus U_2 \oplus U_3 = \mathbb{R}^3$   
 Summen er direkte, e.g.  
 fordi  $\dim U_1 + \dim U_2 +$   
 $\dim U_3 = 3$



$\mathbb{R}^3$  er ej direkte sum af  
 $U_1$  og  $U_2$ ; thi  $\dim U_1$   
 $+ \dim U_2 = 4$



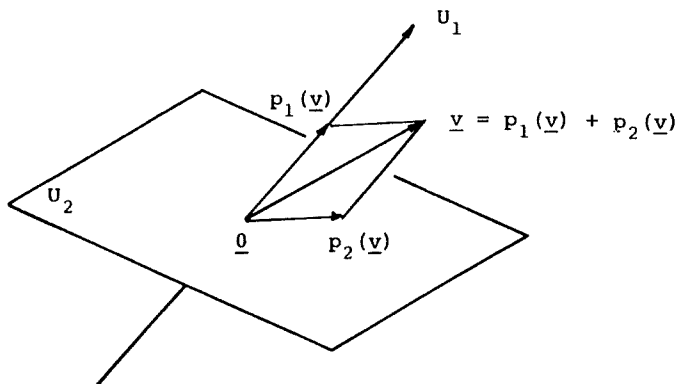
Her er

$U_1 \oplus U_2 = \mathbb{R}^3$   
 e.g. fordi  $U_1$  og  $U_2$  foruden  
 at udspænde  $\mathbb{R}^3$  også tilfreds-  
 stiller  $U_1 \cap U_2 = \underline{0}$

Er  $V$  spaltet i en direkte sum

$$V = U_1 \oplus \dots \oplus U_k$$

kalder vi en vilkårlig vektor  $\underline{v}$ 's komponent i  $U_i$  for  $\underline{v}$ 's projektion på  $U_i$  (efter retningen bestemt ved  $U_1, \dots, U_{i-1}, U_{i+1}, \dots, U_k$ )



Figur 1.3 Projektion af vektor.

og vi benævner den  $p_i(\underline{v})$

Afbildningen  $p_i$  er idempotent, i.e.  $p_i \circ p_i(\underline{v}) = p_i(\underline{v})$ ,  $\forall \underline{v}$  hvor fog betegner sammensætningen af  $f$  med  $g$ .

## 1.2 Lineære afbildninger og matricer

Vi indleder med et afsnit om lineære afbildninger

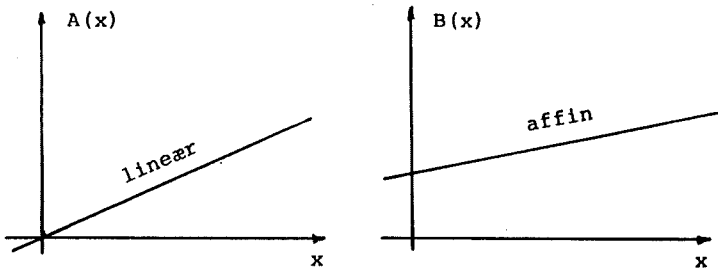
### 1.2.1 Lineære afbildninger

En afbildning  $A : U \rightarrow V$ , hvor  $U$  og  $V$  er vektorrum, siges at

være lineær, hvis

$$\forall \lambda_1, \lambda_2 \in \mathbb{R} \forall \underline{u}_1, \underline{u}_2 \in U : A(\lambda_1 \underline{u}_1 + \lambda_2 \underline{u}_2) = \lambda_1 A(\underline{u}_1) + \lambda_2 A(\underline{u}_2)$$

**Eksempel 1.3** En afbildning  $A : \mathbb{R} \rightarrow \mathbb{R}$  er lineær, netop hvis dens graf er en ret linie gennem  $(0,0)$ . Hvis grafen er en ret linie, der dog ikke går gennem  $(0,0)$  siges afbildningen at være affin.



Figur 1.4 Grafer for lineær og for affin afbildning  $\mathbb{R} \rightarrow \mathbb{R}$ .

Ved nulrummet  $N(A)$  for en lineær afbildning  $A : U \rightarrow V$  forstås underrummet

$$A^{-1}(\underline{0}) = \{ \underline{u} \mid A(\underline{u}) = \underline{0} \}$$

Der gælder følgende formel om sammenhængen mellem dimensionen af billedrum og nulrum

$$\dim N(A) + \dim A(U) = \dim U$$

Specielt gælder

$$\dim A(U) \leq \dim U$$

med lighedstegn netop, hvis  $A$  er injektiv (d.v.s. enentydig). Er  $A$  bijektiv (d.v.s. enentydig og "på"), ses umiddelbart, at  $\dim U = \dim V$ . Vi siger, at en sådan afbildning er en isomorfi, og at  $U$  og  $V$  er isomorfe. Det kan vises, at et vilkårligt  $n$ -dimensionalt (reelt) vektorrum er isomorft med  $\mathbb{R}^n$ . I den videre fremstilling vil vi derfor ofte identificere et  $n$ -dimensionalt vektorrum med  $\mathbb{R}^n$ .

Det kan vises, at de i forrige afsnit omtalte projektioner er lineære afbildninger.

### 1.2.2 Matricer

Ved en matrix  $\underline{A}$  forstås et rektangulært skema af tal som

$$\underline{A} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{m1} & \dots & a_{mn} \end{bmatrix} .$$

Ofte vil vi anvende den kortere skrivemåde

$$\underline{A} = (a_{ij}) .$$

Mere specifikt betegnes  $\underline{A}$  som en  $m \times n$  matrix, da der er  $m$  rækker og  $n$  søjler. Hvis  $m = 1$  kan matricen kaldes en rækkevektor og hvis  $n = 1$  en søjlevektor.

Den matrix man får frem ved at ombytte rækker og søjler kaldes den transponerede matrix til  $\underline{A}$  og betegnes  $\underline{A}'$ , i.e.

$$\underline{A}' = \begin{bmatrix} a_{11} & \dots & a_{m1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{1n} & \dots & a_{mn} \end{bmatrix}$$

En  $m \times n$  matrix kaldes kvadratisk, hvis  $n = m$ . En kvadratisk matrix for hvilken  $\underline{A} = \underline{A}'$  kaldes symmetrisk. Elementerne  $a_{ii}$ ,  $i = 1, \dots, n$  kaldes diagonalelementerne.

En særlig vigtig matrix er enhedsmatricen af orden  $n$

$$\underline{I}_n = \underline{I} = \begin{bmatrix} 1 & \dots & 0 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ 0 & \dots & 1 \end{bmatrix} .$$

En matrix der har nuller uden for diagonalen kaldes en diagonalmatrix. Vi anvender skrivemåden

$$\underline{A} = \text{diag}(\delta_1, \dots, \delta_n) = \begin{bmatrix} \delta_1 & \dots & 0 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ 0 & \dots & \delta_n \end{bmatrix} .$$

For givne  $n \times m$  matricer  $A$  og  $B$  definerer man matrix-summen

$$\underline{A} + \underline{B} = \begin{bmatrix} a_{11} + b_{11} & \dots & a_{1m} + b_{1m} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{n1} + b_{n1} & \dots & a_{nm} + b_{nm} \end{bmatrix} .$$

Skalarmultiplikationen defineres ved

$$c\underline{A} = \begin{bmatrix} ca_{11} & \dots & ca_{1m} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ ca_{n1} & \dots & ca_{nm} \end{bmatrix} ,$$

d.v.s. elementvis multiplikation.

For en  $m \times n$  matrix  $\underline{C}$  og en  $n \times p$  matrix  $\underline{D}$  defineres matrixproduktet  $\underline{P} = \underline{C}\underline{D}$  ved, at  $\underline{P}$  er en  $m \times p$  matrix med  $(i,j)$ 'te element

$$P_{ij} = \sum_{k=1}^n c_{ik} d_{kj}$$

Det kan bemærkes, at matrixproduktet ikke er kommutativt, i.e. at  $\underline{CD}$  ikke er lig  $\underline{DC}$ .

For transponeringen gælder nu følgende regler

$$(\underline{A} + \underline{B})' = \underline{A}' + \underline{B}'$$

$$(c\underline{A})' = c\underline{A}'$$

$$(\underline{CD})' = \underline{D}'\underline{C}'$$

### 1.2.3 Lineære afbildningers matrixfremstillinger

Det kan vises, at der til enhver lineær afbildning  $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$  svarer en  $m \times n$  matrix  $\underline{A}$ , således at

$$\forall \underline{x} \in \mathbb{R}^n : A(\underline{x}) = \underline{A}\underline{x}$$

Omvendt er et  $A$  defineret ved denne relation en lineær afbildning.  $\underline{A}$  bestemmes let som den matrix, der som søjler har koordinaterne for billedet af enhedsvektorerne i  $\mathbb{R}^n$ . E.g. haves

$$\underline{Ae}_2 = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} = \begin{bmatrix} a_{12} \\ \cdot \\ \cdot \\ a_{m2} \end{bmatrix} = \underline{a}_2$$

Er der ydermere givet en lineær afbildning  $B : \mathbb{R}^m \rightarrow \mathbb{R}^k$  med tilhørende matrix  $\underline{B}$  ( $k \times m$ ), da gælder, at  $B \circ A \leftrightarrow \underline{B}\underline{A}$  d.v.s.:

$$\forall \underline{x} \in \mathbb{R}^n \quad (B \circ A)(\underline{x}) = B(A(\underline{x})) = \underline{BAx}$$

Vi nævner her, at en  $n \times n$  matrix  $\underline{A}$  siges at være regulær, hvis den tilsvarende lineære afbildning er bijektiv. Dette er ensbetydende med eksistensen af en invers matrix, i.e. en matrix  $\underline{A}^{-1}$ , der tilfredsstiller

$$\underline{AA}^{-1} = \underline{A}^{-1}\underline{A} = \underline{I}$$

hvor  $\underline{I}$  er enhedsmatricen af  $n$ 'te orden.

En kvadratisk matrix, der svarer til en idempotent afbildning, kaldes også idempotent. Det ses let, at en matrix  $\underline{A}$  er idempotent, hvis og kun hvis

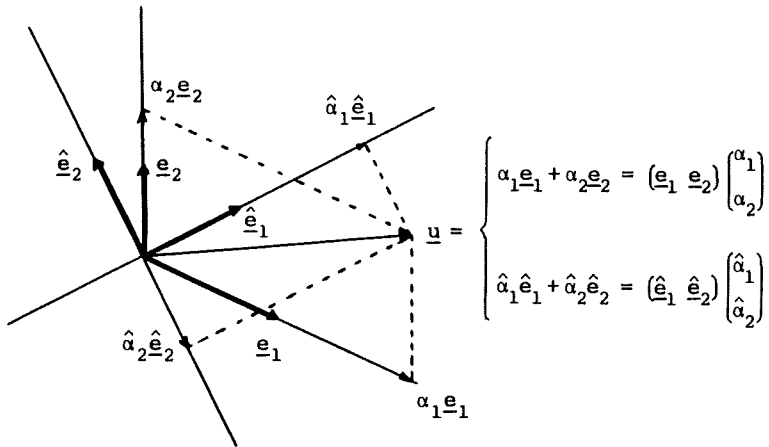
$$\underline{AA} = \underline{A}$$

Det kan bemærkes, at hvis en idempotent matrix er regulær, da er den lig enhedsmatricen, d.v.s. den tilsvarende afbildning er identiteten.

#### 1.2.4 Koordinationstransformation

I dette afsnit vil vi angive formler for en lineær afbildnings matrixfremstilling ved overgang fra et sæt baser til et andet.

Vi betragter først ændringen i koordinater ved overgang fra et koordinatsystem til et andet. Normalt vælger vi at undlade at skelne mellem en vektor  $\underline{u}$  og dens koordinatsæt. Det giver en enkel notation og vil ikke give anledning til forvekslinger. Når der er involveret flere koordinatsystemer, er det imidlertid åbenbart nødvendigt at foretage denne skelnen. Vi betragter i  $\mathbb{R}^n$  to koordinatsystemer  $(\underline{e}_1, \dots, \underline{e}_n)$  og  $(\hat{\underline{e}}_1, \dots, \hat{\underline{e}}_n)$ . Koordinaterne til en vektor  $\underline{u}$  i de to systemer betegnes henholdsvis  $(\alpha_1, \dots, \alpha_n)'$  og  $(\hat{\alpha}_1, \dots, \hat{\alpha}_n)'$ , jvf. nedenstående figur.



Figur 1.5 Skitse over koordinattransformationsproblemet.

Lad det "nye" system  $(\hat{\underline{e}}_1, \dots, \hat{\underline{e}}_n)$  være givet ved

$$(\hat{\underline{e}}_1, \dots, \hat{\underline{e}}_n) = (\underline{e}_1, \dots, \underline{e}_n) \underline{\underline{S}}$$

d.v.s.

$$\hat{\underline{e}}_i = s_{1i} \underline{e}_1 + \dots + s_{ni} \underline{e}_n, \quad i = 1, \dots, n.$$

Søjlerne i  $\underline{\underline{S}}$ -matricen er altså lig det "nye" systems "gamle" koordinater.  $\underline{\underline{S}}$  kaldes koordinattransformationsmatricen.

Bemærkning I mange fremstillinger anvendes udtrykket koordinattransformationsmatricen dog om matricen  $\underline{\underline{S}}^{-1}$ . Man må derfor altid gøre sig klart, hvilken matrix der er tale om.

Da

$$(\underline{e}_1 \ \dots \ \underline{e}_n) \begin{pmatrix} \alpha_1 \\ \cdot \\ \alpha \\ \cdot \\ \alpha_n \end{pmatrix} = (\hat{\underline{e}}_1 \ \dots \ \hat{\underline{e}}_n) \begin{pmatrix} \hat{\alpha}_1 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\alpha}_n \end{pmatrix},$$

(jvf. fig. 1.5), bliver sammenhængen mellem en vektors "gamle" koordinater og "nye" koordinater



$$\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \underline{\underline{S}} \begin{bmatrix} \hat{\alpha}_1 \\ \vdots \\ \hat{\alpha}_n \end{bmatrix} \iff \begin{bmatrix} \hat{\alpha}_1 \\ \vdots \\ \hat{\alpha}_n \end{bmatrix} = \underline{\underline{S}}^{-1} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} .$$

Vi betragter nu en lineær afbildning  $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , og lad  $A$ 's matrixfremstilling m.h.t. baserne  $(\underline{e}_1, \dots, \underline{e}_n)$  og  $(\underline{f}_1, \dots, \underline{f}_m)$  være

$$\underline{\underline{\beta}} = \underline{\underline{A}} \underline{\underline{\alpha}}$$

og fremstillingen m.h.t. baserne  $(\hat{e}_1, \dots, \hat{e}_n) = (\underline{e}_1, \dots, \underline{e}_n) \underline{\underline{S}}$  og  $(\hat{f}_1, \dots, \hat{f}_m) = (\underline{f}_1, \dots, \underline{f}_m) \underline{\underline{T}}$  være

$$\hat{\underline{\underline{\beta}}} = \hat{\underline{\underline{A}}} \hat{\underline{\underline{\alpha}}}$$

Da gælder

$$\hat{\underline{\underline{A}}} = \underline{\underline{S}}^{-1} \underline{\underline{A}} \underline{\underline{T}},$$

hvilket fås umiddelbart ved anvendelse af koordinattransformationsreglerne for koordinaterne.

Er der tale om afbildninger  $\mathbb{R}^n \rightarrow \mathbb{R}^n$  og anvendes samme koordinattransformation fås relationen

$$\hat{\underline{\underline{A}}} = \underline{\underline{S}}^{-1} \underline{\underline{A}} \underline{\underline{S}}.$$

Matricerne  $\underline{\underline{A}}$  og  $\hat{\underline{\underline{A}}} = \underline{\underline{S}}^{-1} \underline{\underline{A}} \underline{\underline{S}}$  kaldes da similære matricer.

### 1.2.5 Rangen af en matrix

Ved rangen af en lineær afbildning  $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$  forstås dimensionen af billedrummet, i.e.

$$\text{rg}(A) = \dim A(\mathbb{R}^n) .$$

Ved rangen af en matrix  $A$  forstås rangen af den tilsvarende lineære afbildning.

Det ses, at  $\text{rg}(\underline{A})$  netop er antallet af lineært uafhængige søjlevektorer i  $\underline{A}$ . Trivielt gælder derfor

$$\text{rg}(\underline{A}) \leq n .$$

Indføres den transponerede matrix  $\underline{A}'$  vises det let, at  $\text{rg}(\underline{A}) = \text{rg}(\underline{A}')$ , d.v.s. vi har

$$\text{rg}(\underline{A}) \leq \min(m, n) .$$

Hvis  $\underline{A}$  og  $\underline{B}$  er to  $m \times n$  matricer, gælder

$$\text{rg}(\underline{A} + \underline{B}) \leq \text{rg}(\underline{A}) + \text{rg}(\underline{B}) .$$

Denne relation er umiddelbar, når man erindrer, at der for de tilhørende afbildninger  $A$  og  $B$  gælder  $(A + B)(\mathbb{R}^n) \subseteq A(\mathbb{R}^n) \cup B(\mathbb{R}^n)$ .

Hvis  $\underline{A}$  er en  $(m \times n)$ -matrix og  $\underline{B}$  en  $(k \times m)$ -matrix, gælder

$$\text{rg}(\underline{B}\underline{A}) \leq \text{rg}(\underline{A}) .$$

Hvis  $\underline{B}$  er regulær  $(m \times m)$  gælder

$$\text{rg}(\underline{B}\underline{A}) = \text{rg}(\underline{A}) .$$

Disse relationer er umiddelbare følger af relationen  $\dim B(A(\mathbb{R}^n)) \leq \dim A(\mathbb{R}^n)$ , hvor vi har lighedstegn, hvis  $B$  er injektiv. Der gælder naturligvis de analoge relationer for en  $(n \times p)$ -matrix  $\underline{C}$ :

$$\text{rg}(\underline{A}\underline{C}) \leq \text{rg}(\underline{A})$$

med lighedstegn, hvis  $\underline{C}$  er en regulær  $(n \times n)$ -matrix. Af disse udledes specielt for regulære  $\underline{B}$  og  $\underline{C}$

$$\text{rg}(\underline{B}\underline{A}\underline{C}) = \text{rg}(\underline{A}) \dots$$

Til slut nævner vi, at en  $(n \times n)$ -matrix  $\underline{A}$  er regulær, netop hvis  $\text{rg}(\underline{A}) = n$ .

### 1.2.6 Determinanten af en matrix

Den abstrakte definition på determinanten af en kvadratisk  $p \times p$  matrix  $\underline{A}$  er

$$\det(\underline{A}) = \sum_{\text{alle } \sigma} \pm a_{1\sigma(1)} \cdots a_{p\sigma(p)},$$

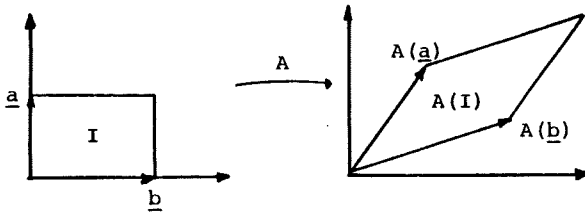
hvor  $\sigma$  er en permutation af tallene  $1, \dots, p$  og hvor  $+$  tegnet anvendes, hvis permutationen er lige (d.v.s. kan sammensættes af et lige antal ombytninger af naboer) og  $-$ , hvis den er ulige.

Vi skal ikke komme ind på baggrunden for denne definition, men uden bevis konstatere, at determinanten angiver volumenforholdet for den tilsvarende lineære afbildning, i.e. for en  $(n \times n)$ -matrix  $\underline{A}$

$$|\det(\underline{A})| = \frac{\text{vol}(A(I))}{\text{vol}(I)},$$

hvor  $I$  er en  $n$ -dimensional kasse og  $A(I)$  billedet af  $I$  (et  $n$ -dimensionalt parallelepipedum) ved den tilsvarende afbildning.

I 2 dimensioner er situationen skitseret i fig. 1.6



Figur 1.6 Et rektangel og dets billede ved en lineær afbildning.

For  $2 \times 2$  og  $3 \times 3$  matricer bliver definitionen på determinanten

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

$$\det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = aei + bfg + cdh - gec - hfa - idb .$$

For determinanter af højere orden (her  $n$ 'te) kan man udvikle determinanten efter den  $i$ 'te række, d.v.s.

$$\det(\underline{\underline{A}}) = \sum_{j=1}^n a_{ij} (-1)^{i+j} \det(\underline{\underline{A}}_{ij}),$$

hvor  $\underline{\underline{A}}_{ij}$  er den matrix, der fremkommer ved at slette den  $i$ 'te række og den  $j$ 'te søjle i  $\underline{\underline{A}}$ . Tallet

$$A_{ij} = (-1)^{i+j} \det(\underline{\underline{A}}_{ij})$$

kaldes også for elementet  $a_{ij}$ 's komplement (eng. cofactor). Der eksisterer selvfølgelig en helt analog udvikling efter søjler.

Når man explicit skal evaluere determinanten er der tre "handy" regler:

- i) ombytning af 2 rækker (søjler) i  $\underline{\underline{A}}$  multiplicerer  $\det(\underline{\underline{A}})$  med  $-1$ .
- ii) multipliceres en række (søjle) med en skalar  $\lambda$  multipliceres  $\det(\underline{\underline{A}})$  med denne skalar.
- iii) adderes et multiplum af en række (søjle) til en anden forbliver  $\det(\underline{\underline{A}})$  uændret.

Skal man bestemme rangen af en matrix, kan det være nyttigt at erindre, at rangen er det største tal  $r$  for hvilken matricen har en fra 0 forskellig underdeterminant af  $r$ 'te orden. Vi får så specielt, at en matrix  $\underline{\underline{A}}$  er regulær, netop hvis  $\det \underline{\underline{A}} \neq 0$ . Dette virker også intuitivt indlysende, når man betragter determinanten som volumenforholdet. Er dette 0, må afbildningen være "dimensionsreducerende" i en passende forstand.

For kvadratiske matricer  $\underline{\underline{A}}$  og  $\underline{\underline{B}}$  gælder

$$\det(\underline{\underline{AB}}) = \det(\underline{\underline{A}})\det(\underline{\underline{B}})$$

For en diagonalmatrix  $\underline{\underline{A}} = \text{diag}(\lambda_1, \dots, \lambda_n)$  gælder

$$\det(\underline{\underline{A}}) = \lambda_1 \dots \lambda_n$$

For en trekantmatrix  $\underline{\underline{C}}$  med diagonalelementer  $c_1, \dots, c_n$  gælder

$$\det(\underline{\underline{C}}) = c_1 \dots c_n$$

Ved hjælp af determinanter kan man direkte opskrive den inverse til en regulær matrix  $\underline{\underline{A}}$ . Der gælder nemlig, at

$$\underline{\underline{A}}^{-1} = \frac{1}{\det \underline{\underline{A}}} (A_{ij})',$$

dvs. den inverse til en regulær matrix  $\underline{\underline{A}}$  er den transponerede til den matrix, der fås ved at erstatte hvert element i  $\underline{\underline{A}}$  med sit komplement divideret med  $\det \underline{\underline{A}}$ . Det må dog her præciseres, at formelen ikke er direkte anvendelig ved inversion af store ma-

tricer grundet det store regnearbejde, der er forbundet med beregning af determinanter.

Noget tilsvarende gør sig gældende for Cramérs sætning om løsning af et lineært ligningssystem: Betragt den regulære matrix

$\underline{A} = (\underline{a}_1, \dots, \underline{a}_n)$ . Da er løsningen til ligningen

$$\underline{Ax} = \underline{b}$$

givet ved

$$x_i = \det(\underline{a}_1, \dots, \underline{a}_{i-1}, \underline{b}, \underline{a}_{i+1}, \dots, \underline{a}_n) / \det \underline{A}$$

### 1.2.7. Blokmatricer

Ved blokmatrix forstås en matrix af formen

$$\underline{B} = \begin{bmatrix} \underline{B}_{11} & \dots & \underline{B}_{1n} \\ \vdots & & \vdots \\ \underline{B}_{m1} & \dots & \underline{B}_{mn} \end{bmatrix}$$

hvor blokkene  $\underline{B}_{ij}$  er matricer af orden  $m_i \times n_j$ .

Ved addition og multiplikation kan man anvende de sædvanlige regneregler for matricer, og opfatte blokkene som elementer.

Vi finder således eksempelvis, at

$$\begin{bmatrix} \underline{A} & \underline{B} \\ \underline{C} & \underline{D} \end{bmatrix} \begin{bmatrix} \underline{R} \\ \underline{S} \end{bmatrix} = \begin{bmatrix} \underline{AR} + \underline{BS} \\ \underline{CR} + \underline{DS} \end{bmatrix},$$

naturligvis forudsat, at de involverede produkter eksisterer etc.

Vi nævner nu først et resultat om determinanter af "trekants" matricen.

Sætning 1.1 Lad den kvadratiske matrix  $\underline{A}$  være spaltet i blokmatricer

$$\underline{\underline{A}} = \begin{bmatrix} \underline{\underline{B}} & \underline{\underline{C}} \\ \underline{\underline{O}} & \underline{\underline{D}} \end{bmatrix}$$

hvor  $\underline{\underline{B}}$  og  $\underline{\underline{D}}$  er kvadratiske og  $\underline{\underline{O}}$  en matrix bestående af lutter 0'er. Da gælder

$$\det(\underline{\underline{A}}) = \det(\underline{\underline{B}})\det(\underline{\underline{D}})$$

Bevis Vi har, at

$$\begin{bmatrix} \underline{\underline{B}} & \underline{\underline{C}} \\ \underline{\underline{O}} & \underline{\underline{D}} \end{bmatrix} = \begin{bmatrix} \underline{\underline{I}} & \underline{\underline{O}} \\ \underline{\underline{O}} & \underline{\underline{D}} \end{bmatrix} \begin{bmatrix} \underline{\underline{B}} & \underline{\underline{C}} \\ \underline{\underline{O}} & \underline{\underline{I}} \end{bmatrix}$$

hvor  $\underline{\underline{I}}$ 'erne er enhedsmatricer og ikke nødvendigvis af samme orden. Hvis man udvikler den første matrix efter 1.ste række, ses, at den har samme determinant som den matrix, man får ved at slette 1.ste række og søjle. Gentages dette indtil den tiloversblevne underdeterminant er  $\underline{\underline{D}}$ , ses, at

$$\det \begin{bmatrix} \underline{\underline{I}} & \underline{\underline{O}} \\ \underline{\underline{O}} & \underline{\underline{D}} \end{bmatrix} = \det(\underline{\underline{D}})$$

Analogt fås, at den sidste matrix har determinanten  $\det(\underline{\underline{B}})$  og resultatet følger.

□

Den næste sætning udvider dette resultat.

Sætning 1.2 Lad matrixen  $\underline{\underline{\Sigma}}$  være spaltet i blokmatricer

$$\underline{\underline{\Sigma}} = \begin{bmatrix} \underline{\underline{\Sigma}}_{11} & \underline{\underline{\Sigma}}_{12} \\ \underline{\underline{\Sigma}}_{21} & \underline{\underline{\Sigma}}_{22} \end{bmatrix}$$

Da gælder

$$\det(\underline{\underline{\Sigma}}) = \det(\underline{\underline{\Sigma}}_{11} - \underline{\underline{\Sigma}}_{12}\underline{\underline{\Sigma}}_{22}^{-1}\underline{\underline{\Sigma}}_{21})\det(\underline{\underline{\Sigma}}_{22}),$$

forudsat, at  $\underline{\underline{\Sigma}}_{22}$  er regulær.

Bevis Da

$$\begin{bmatrix} \underline{\Sigma}_{11} & \underline{\Sigma}_{12} \\ \underline{\Sigma}_{21} & \underline{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \underline{I} & \underline{O} \\ -\underline{\Sigma}_{22}^{-1}\underline{\Sigma}_{21} & \underline{I} \end{bmatrix} = \begin{bmatrix} \underline{\Sigma}_{11} - \underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}\underline{\Sigma}_{21} & \underline{\Sigma}_{12} \\ \underline{O} & \underline{\Sigma}_{22} \end{bmatrix},$$

følger resultatet umiddelbart ved anvendelse af foregående sætning. □

Den sidste sætning giver et nyttigt resultat om inversion af matricer, der er spaltet i blokmatricer.

Sætning 1.3 For den symmetriske matrix

$$\underline{\Sigma} = \begin{bmatrix} \underline{\Sigma}_{11} & \underline{\Sigma}_{12} \\ \underline{\Sigma}_{21} & \underline{\Sigma}_{22} \end{bmatrix}$$

gælder

$$\underline{\Sigma}^{-1} = \begin{bmatrix} \underline{B}^{-1} & -\underline{B}^{-1}\underline{A}' \\ -\underline{A}\underline{B}^{-1} & \underline{\Sigma}_{22}^{-1} + \underline{A}\underline{B}^{-1}\underline{A}' \end{bmatrix},$$

hvor

$$\underline{A} = \underline{\Sigma}_{22}^{-1}\underline{\Sigma}_{21}$$

$$\underline{B} = \underline{\Sigma}_{11} - \underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}\underline{\Sigma}_{21}'$$

hvor det naturligvis forudsættes, at de anførte inverse eksisterer.

Bevis Resultatet følger umiddelbart ved multiplikation af  $\underline{\Sigma}$  og  $\underline{\Sigma}^{-1}$ . □



### 1.3 Pseudoinvers eller generaliseret invers matrix til en ikke-regulær matrix

Vi betragter en lineær afbildning

$$A : E \rightarrow F$$

hvor  $E$  er et  $n$ -dimensionalt og  $F$  et  $m$ -dimensionalt (euklidisk) vektorrum. Den til  $A$  svarende matrix benævnes som sædvanlig  $\underline{A}$ , og den har dimensionen  $m \times n$ . Vi sætter nulrummet for  $A$  lig  $U$ , d.v.s.

$$U = A^{-1}(0),$$

og benævner dets dimension  $r$ . Billedrummet

$$V = A(E)$$

har dimensionen  $s = n - r$ , jvr. p. 1.9.

Vi betragter nu et vilkårligt  $s$ -dimensionalt underrum  $U^* \subseteq E$ , som er komplementært til  $U$ , og et vilkårligt  $m - s$  dimensionalt underrum  $V^* \subseteq F$ , som er komplementært til  $V$ .

En vilkårlig vektor  $\underline{x} \in E$  kan nu skrives på formen

$$\underline{x} = \underline{u} + \underline{u}^*, \quad \underline{u} \in U \text{ og } \underline{u}^* \in U^*,$$

idet  $\underline{u}$  og  $\underline{u}^*$  er givet ved

$$\underline{u} = \underline{x} - P_{U^*}(\underline{x})$$

$$\underline{u}^* = P_{U^*}(\underline{x})$$

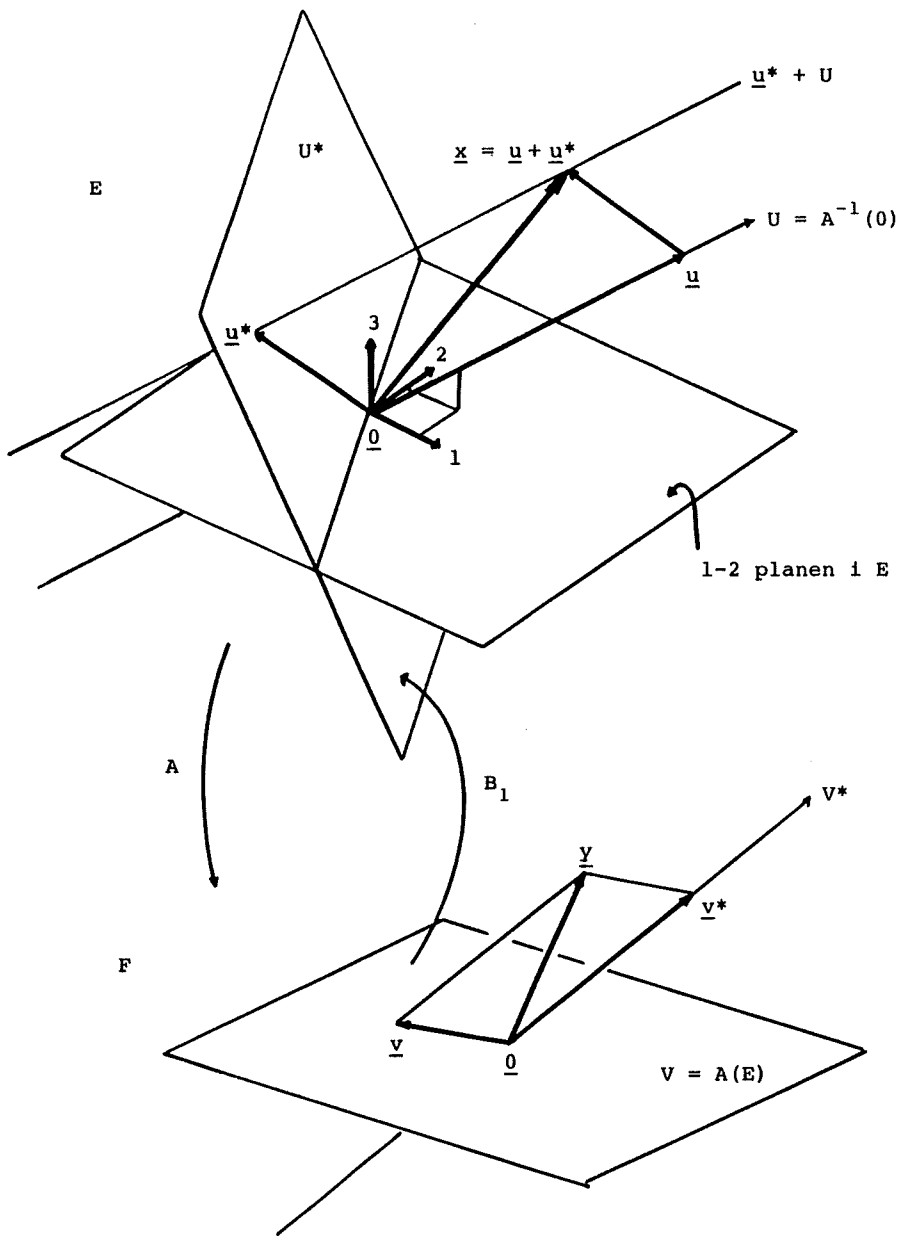
Her betegner  $P_{U^*}$  projektionen af  $E$  ind på  $U^*$  langs underrummet  $U$ . Tilsvarende kan ethvert  $\underline{y} \in F$  skrives

$$\underline{y} = (\underline{y} - P_V(\underline{y})) + P_V(\underline{y}) = \underline{y}^* + \underline{v}$$

hvor

$$P_V : F \rightarrow V$$

er projektionen af  $F$  ind på  $V$  langs  $V^*$ .



Figur 1.7 Skitse vedrørende pseudoinvers afbildning.

Da nu

$$A(\underline{x}) = A(\underline{u} + \underline{u}^*) = A(\underline{u}^*),$$

ses, at  $A$  er konstant på siderummene

$$\underline{u}^* + U = \{\underline{u}^* + \underline{u} \mid \underline{u} \in U\},$$

og det følger, at  $A$ 's restriktion til  $U^*$  er en bijektiv afbildning af  $U^*$  på  $V$ . Denne afbildning har derfor en invers

$$B_1 : V \rightarrow U^*$$

givet ved

$$B_1(\underline{v}) = \underline{u}^* \leftrightarrow A(\underline{u}^*) = \underline{v}$$

Vi er nu i stand til at formulere definitionen på den pseudo-inverse afbildning.

Definition 1.1 Ved en pseudoinvers eller generaliseret invers afbildning til afbildningen  $A$  forstås en afbildning

$$B = B_1 \circ p_V : F \rightarrow E,$$

hvor  $p_V$  og  $B_1$  er som nævnt foran.

Bemærkning 1 Den pseudoinverse er altså sammensætningen af projektionen ned på  $V$  langs  $V^*$  og den inverse til  $A$ 's restriktion til  $U^*$ .

Bemærkning 2 Den pseudoinverse er selvfølgelig på ingen måde entydig, idet vi får en for hvert valg af underrommene  $U^*$  og  $V^*$ .

Vi kan umiddelbart formulere nogle åbenbare egenskaber ved den pseudoinverse i følgende

Sætning 1.4 Den pseudoinverse  $B$  til  $A$  har følgende egenskaber

- i)  $\text{rg}(B) = \text{rg}(A) = s$
- ii)  $A \circ B = p_V : F \rightarrow V$
- iii)  $B \circ A = p_{U^*} : E \rightarrow U^*$

Det kan vises, at disse egenskaber også karakteriserer pseudoinverse afbildninger, idet der gælder

Sætning 1.5 Lad  $A : E \rightarrow F$  være lineær og med rang  $s$ . Antag, at  $B$  også har rang  $s$ , og at  $A \circ B$  og  $B \circ A$  begge er projektioner af rang  $s$ . Da er  $B$  en pseudoinvers af  $A$  som defineret ovenfor.

Bevis Forbigås (relativt simpel øvelse i lineær algebra).

Vi vil nu give en matrixfremstilling af de ovenfor anførte definitioner.

Definition 1.2 Lad  $\underline{A}$  være en  $(m \times n)$ -matrix af rang  $s$ . En  $(n \times m)$ -matrix  $\underline{B}$  af rang  $s$ , der tilfredsstiller

- i)  $\underline{A} \underline{B}$  idempotent med rang  $s$
- ii)  $\underline{B} \underline{A}$  idempotent med rang  $s$ ,

kaldes en pseudoinvers eller generaliseret invers til  $\underline{A}$ .

Ved hjælp af den pseudoinverse kan man karakterisere løsningsmængden til et system af lineære ligninger, idet vi har følgende

Sætning 1.6 Lad  $\underline{A}$  og  $\underline{B}$  være som i definition 1.2. Den generelle løsning til ligningen

$$\underline{A}\underline{x} = \underline{0}$$

er

$$(\underline{I} - \underline{B}\underline{A})\underline{z} \quad , \quad \underline{z} \in \mathbb{R}^n \quad ,$$

og den generelle løsning til ligningen (der forudsættes konsistent)

$$\underline{A}\underline{x} = \underline{y} \quad ,$$

er

$$\underline{B}\underline{y} + (\underline{I} - \underline{B}\underline{A})\underline{z} \quad , \quad \underline{z} \in \mathbb{R}^n \quad .$$

Bevis Vi betragter først den homogene ligning. En løsning  $\underline{x}$  er åbenbart et punkt i nulrummet  $N(\underline{A}) = \underline{A}^{-1}(\underline{0})$  for den lineære afbildning svarende til  $\underline{A}$ . Matricen  $\underline{B}\underline{A}$  svarer ifølge sætning 1.1 netop til projektionen ned på  $U^*$ . Følgelig svarer  $\underline{I} - \underline{B}\underline{A}$  til projektionen ned på nulrummet  $U = N(\underline{A})$ , hvorfor et vilkårligt  $\underline{x} \in N(\underline{A})$  kan skrives

$$\underline{x} = (\underline{I} - \underline{B}\underline{A})\underline{z} \quad , \quad \underline{z} \in \mathbb{R}^n \quad .$$

Dermed er påstanden angående den homogene ligning bevist.

Ligningen  $\underline{A}\underline{x} = \underline{y}$  har kun løsning (i.e. er kun konsistent), hvis  $\underline{y}$  ligger i billedrummet for  $\underline{A}$ . For et sådant  $\underline{y}$  gælder imidlertid, at

$$\underline{A}\underline{B}\underline{Y} = \underline{Y} \text{ ,}$$

ifølge sætning 1.4

Resultatet for den fuldstændige løsning følger umiddelbart.  $\square$

For at illustrere ovennævnte begrebet anfører vi nu

Eksempel 1.4 Vi betragter matricen

$$\underline{A} = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix} \text{ .}$$

$\underline{A}$  har åbenbart rangen 2.

Vi vil betragte den til  $\underline{A}$  svarende lineære afbildning

$$A : E \rightarrow F$$

hvor E og F er 3-dimensionale vektorrum med baserne  $\{\underline{e}_1, \underline{e}_2, \underline{e}_3\}$  og  $\{\underline{f}_1, \underline{f}_2, \underline{f}_3\}$ . Koordinaterne i disse baser benævnes med små x'er henholdsvis y'er, således at A har koordinatfremstillingen

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \text{ .}$$

Vi vil nu først bestemme nulrummet

$$U = N(A) = A^{-1}(\underline{0})$$

for A. Der gælder

$$\underline{x} \in U \Leftrightarrow \underline{Ax} = \underline{0}$$

$$\Leftrightarrow x_1 + x_2 + 2x_3 = 0 \wedge 2x_1 + x_2 + x_3 = 0$$

$$\Leftrightarrow x_1 = x_3 \wedge -3x_1 = x_2$$

$$\Leftrightarrow \underline{x}' = x_1(1, -3, 1) .$$

Nulrummet er altså

$$U = \left\{ t \cdot \begin{bmatrix} 1 \\ -3 \\ 1 \end{bmatrix} \mid t \in \mathbb{R} \right\} = \{ t \cdot \underline{u}_3 \mid t \in \mathbb{R} \}$$

Som komplementært underrum vælger vi at betragte det ortogonale komplement  $U^*$ . Dette har ligningen

$$(1, -3, 1)\underline{x} = 0 ,$$

eller

$$U^* = \{ \underline{x} \mid x_1 - 3x_2 + x_3 = 0 \} .$$

$U^*$  udspændes f.eks. af vektorerne

$$\underline{u}_1 = (1, 0, -1)' \text{ og } \underline{u}_2 = (3, 2, 3)' .$$

Vi betragter nu en ny basis for  $E$ , nemlig  $\{\underline{u}_1, \underline{u}_2, \underline{u}_3\}$ .

Koordinater i dette benævnes med små  $z$ 'er. Overgangen fra  $z$ -koordinater til  $x$ -koordinater er givet ved

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 1 \\ 0 & 2 & -3 \\ -1 & 3 & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}$$

eller

$$\underline{x} = \underline{Sz} .$$

$\underline{S}$  matrixens søjler er som bekendt  $\underline{u}$ 'ernes koordinater i  $\underline{e}$ -systemet.

$A$ 's billedrum  $V$  er 2-dimensional og udspændes af  $\underline{A}$ 's søjler. Vi kan e.g. vælge de to første, nemlig

$$\underline{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \quad \underline{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Som komplementært underrum  $V^*$  vælger vi  $V$ 's ortogonale komplement. Det frembringes af krydsproduktet af  $\underline{v}_1$  og  $\underline{v}_2$ :

$$\underline{v}_1 \times \underline{v}_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} = \underline{v}_3.$$

Vi betragter nu den nye basis  $\{\underline{v}_1, \underline{v}_2, \underline{v}_3\}$  for  $F$ . Koordinaterne heri benævnes med små  $w$ -er. Overgangen fra  $w$ -koordinater til  $y$ -koordinater er givet ved

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 1 & 1 \\ 2 & 1 & -1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix},$$

eller på kompakt form

$$\underline{y} = \underline{T}\underline{w}.$$

Vi vil nu bestemme koordinatudtryk for  $A$  i  $z$ - og  $w$ -koordinater. Da

$$\underline{y} = \underline{A}\underline{x}$$

fås

$$\underline{T}\underline{w} = \underline{A}\underline{S}\underline{z}$$

eller

$$\underline{w} = \underline{T}^{-1}\underline{A}\underline{S}\underline{z}.$$



Nu er

$$\underline{T}^{-1} = \begin{bmatrix} -1 & \frac{1}{2} & \frac{1}{2} \\ 2 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{bmatrix},$$

hvorfor

$$\begin{aligned} \underline{T}^{-1} \underline{AS} &= \begin{bmatrix} -1 & \frac{1}{2} & \frac{1}{2} \\ 2 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 1 \\ 0 & 2 & -3 \\ -1 & 3 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 & 0 \\ -3 & 11 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

Da  $\{\underline{u}_1, \underline{u}_2\}$  udspænder  $U^*$  og  $\{\underline{v}_1, \underline{v}_2\}$  udspænder  $V$ , har restriktionen

$$A : U^* \rightarrow V$$

koordinatfremstillingen

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ -3 & 11 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}.$$

Denne har den inverse afbildning

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 \\ \frac{3}{22} & \frac{1}{11} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}.$$

Opfattes punkterne som punkter i  $E$  og  $F$  og ikke blot som punkter i  $U^*$  og  $V$  fås

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ \frac{3}{22} & \frac{1}{11} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \quad (*)$$

Projektionen af  $F$  ned på  $V$  langs  $V^*$  har koordinatfremstillingen

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ 0 \end{bmatrix} \quad (**)$$

Dette er  $z$ - $w$  koordinatfremstillingen for den pseudoinverse  $B$  til afbildningen  $A$ . Vi ønsker imidlertid en beskrivelse i  $x$ - $y$  koordinater. Da

$$\underline{z} = \underline{S}^{-1}\underline{x} = \underline{C}\underline{w} = \underline{C}\underline{T}^{-1}\underline{y}$$

fås

$$\underline{x} = \underline{S}\underline{C}\underline{T}^{-1}\underline{y},$$

hvor  $\underline{C}$  er matricen i (\*).

Vi har derfor

$$\underline{B} = \underline{S}\underline{C}\underline{T}^{-1} = \begin{bmatrix} 1 & 3 & 1 \\ 0 & 2 & -3 \\ -1 & 3 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ \frac{3}{22} & \frac{1}{11} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -1 & \frac{1}{2} & \frac{1}{2} \\ 2 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{bmatrix} = \frac{1}{22} \begin{bmatrix} -8 & 7 & 7 \\ 2 & 1 & 1 \\ 14 & -4 & -4 \end{bmatrix}$$

Denne matrix er en pseudoinvers til  $\underline{A}$ .

□

Som det fremgår af foranstående eksempel er det ret byrdefuldt blot at gå frem efter definitionen ved udledningen af en pseudoinvers. Ofte vil man kunne bruge følgende

Sætning 1.7 Lad  $m \times n$  matricen  $\underline{\underline{A}}$  have rang  $s$  og lad

$$A = \begin{bmatrix} \underline{\underline{C}} & \underline{\underline{D}} \\ \underline{\underline{E}} & \underline{\underline{F}} \end{bmatrix},$$

hvor  $\underline{\underline{C}}$  er regulær og af dimension  $s \times s$ . En pseudoinvers til  $\underline{\underline{A}}$  er da

$$\underline{\underline{A}}^{-} = \begin{bmatrix} \underline{\underline{C}}^{-1} & \underline{\underline{0}} \\ \underline{\underline{0}} & \underline{\underline{0}} \end{bmatrix},$$

hvor  $0$ -matricerne har sådanne størrelser, at  $\underline{\underline{A}}^{-}$  har dimensionen  $n \times m$ .

Bevis Vi har

$$\underline{\underline{A}}\underline{\underline{A}}^{-} = \begin{bmatrix} \underline{\underline{C}} & \underline{\underline{D}} \\ \underline{\underline{E}} & \underline{\underline{F}} \end{bmatrix} \begin{bmatrix} \underline{\underline{C}}^{-1} & \underline{\underline{0}} \\ \underline{\underline{0}} & \underline{\underline{0}} \end{bmatrix} \begin{bmatrix} \underline{\underline{C}} & \underline{\underline{D}} \\ \underline{\underline{E}} & \underline{\underline{F}} \end{bmatrix} = \begin{bmatrix} \underline{\underline{C}} & \underline{\underline{D}} \\ \underline{\underline{E}} & \underline{\underline{E}}\underline{\underline{C}}^{-1}\underline{\underline{D}} \end{bmatrix}.$$

Da  $\text{rg}(\underline{\underline{A}}) = s$ , kan de sidste  $n-s$  søjler skrives som linearkombinationer af de første  $s$ , d.v.s. der eksisterer en matrix  $\underline{\underline{H}}$ , så

$$\begin{bmatrix} \underline{\underline{D}} \\ \underline{\underline{F}} \end{bmatrix} = \begin{bmatrix} \underline{\underline{C}} \\ \underline{\underline{E}} \end{bmatrix} \underline{\underline{H}}$$

eller

$$\begin{aligned} \underline{\underline{D}} &= \underline{\underline{C}}\underline{\underline{H}} \\ \underline{\underline{F}} &= \underline{\underline{E}}\underline{\underline{H}} \end{aligned}$$

Heraf fås

$$\underline{\underline{F}} = \underline{\underline{E}}\underline{\underline{C}}^{-1}\underline{\underline{D}}.$$

Indsættes dette i øverste formellinie, fås

$$\underline{\underline{A}} \underline{\underline{A}}^{-1} = \underline{\underline{A}}$$

Ved multiplikation til venstre med  $\underline{\underline{A}}^{-1}$  henholdsvis til højre med  $\underline{\underline{A}}^{-1}$  ses at  $\underline{\underline{A}}^{-1} \underline{\underline{A}}$  og  $\underline{\underline{A}} \underline{\underline{A}}^{-1}$  er idempotente. Sætningen følger nu af definitionen p. 1.26.

□

Vi illustrerer anvendelsen af sætningen i nedenstående

Eksempel 1.5 Vi betragter den i eksempel 1.4 givne matrix

$$\underline{\underline{A}} = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix}.$$

Da

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix},$$

kan vi som pseudoinvers anvende

$$\underline{\underline{A}}^{-1} = \begin{bmatrix} -1 & 1 & 0 \\ 2 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

□

Fordelen ved at anvende den i eksempel 1.4 beskrevne fremgangsmåde fremfor den langt enklere anførte i eksempel 1.5 er, at man der får en præcis geometrisk beskrivelse af forholdene.

Bemærkning Til slut må det anføres, at der i litteraturen optræder en række definitioner på pseudoinverse og generaliserede inverse, således at man ved en eventuel anvendelse må præcisere, hvad definitionen er. Her må især nævnes den såkaldte Moore-

Penrose inverse  $\underline{A}^+$  til en matrix  $\underline{A}$ . Den tilfredsstill

$$i) \quad \underline{A}\underline{A}^+\underline{A} = \underline{A}$$

$$ii) \quad \underline{A}^+\underline{A}\underline{A}^+ = \underline{A}^+$$

$$iii) \quad (\underline{A}\underline{A}^+)' = \underline{A}\underline{A}^+$$

$$iv) \quad (\underline{A}^+\underline{A})' = \underline{A}^+\underline{A}$$

Det ses trivielt at en Moore-Penrose invers virkelig er en generaliseret invers. De øvrige betingelser sikrer at man ved mindste kvadraters løsning til en inkonsistent ligning opnår løsninger med minimal norm. Dette skal vi ikke komme ind på her, men blot henvise den interesserede læser til litteraturen, e.g. Rao og Mitra (1971).

#### 1.4 Egenværdiproblemer. Kvadratiske former

Vi indleder med de grundlæggende definitioner og sætninger i

##### 1.4.1 Egenværdier og egenvektorer for symmetriske matricer

Selve den nedenfor anførte definition på en egenvektor og en egenværdi er gyldig for vilkårlige kvadratiske matricer, men vi vil overalt i det følgende forudsætte, at de involverede matricer er symmetriske med mindre andet explicit er nævnt.

En egenværdi  $\lambda$  for den symmetriske  $n \times n$  matrix  $\underline{A}$  er en løsning til ligningen

$$\det(\underline{A} - \lambda \underline{I}) = 0 .$$

Der er  $n$  (reelle) egenværdier (nogle eventuelt sammenfaldende). Hvis  $\lambda$  er en egenværdi, eksisterer der vektorer  $\underline{x} \neq 0$ , således at

$$\underline{A}\underline{x} = \lambda\underline{x},$$

d.v.s. der findes vektorer, der ved den til  $\underline{A}$  hørende lineære afbildning går over i et multiplum af sig selv. Sådanne vektorer kaldes egenvektorer svarende til egenværdien  $\lambda$ . Antallet af fra 0 forskellige egenværdier er lig  $\text{rg}(\underline{A})$ . Her skal enhver egenværdi medtages så ofte som dens multiplicitet angiver. Mere interessant er

Sætning 1.8 Hvis  $\lambda_i$  og  $\lambda_j$  er forskellige egenværdier, og hvis  $\underline{x}_i$  og  $\underline{x}_j$  er tilhørende egenvektorer, da er  $\underline{x}_i$  og  $\underline{x}_j$  ortogonale, d.v.s.  $\underline{x}_i'\underline{x}_j = 0$ .

Bevis Vi har

$$\underline{A}\underline{x}_i = \lambda_i\underline{x}_i$$

$$\underline{A}\underline{x}_j = \lambda_j\underline{x}_j .$$

Heraf fås umiddelbart

$$\underline{x}_j'\underline{A}\underline{x}_i = \lambda_i\underline{x}_j'\underline{x}_i$$

$$\underline{x}_i'\underline{A}\underline{x}_j = \lambda_j\underline{x}_i'\underline{x}_j .$$

Vi transponerer den første relation og får

$$\underline{x}_i'\underline{A}'\underline{x}_j = \lambda_i\underline{x}_i'\underline{x}_j .$$

Da  $\underline{A}$  er symmetrisk medfører dette, at

$$\lambda_i\underline{x}_i'\underline{x}_j = \lambda_j\underline{x}_i'\underline{x}_j ,$$

og da  $\lambda_i \neq \lambda_j$  er  $\underline{x}_i'\underline{x}_j = 0$  d.v.s.  $\underline{x}_i \perp \underline{x}_j$ .

Resultatet i sætning 1.8 kan suppleres med følgende sætning, der nævnes uden bevis.

Sætning 1.9 Hvis  $\lambda$  er en egenværdi med multiplicitet  $m$ , da udgør mængden af egenvektorer svarende til  $\lambda$  et  $m$ -dimensionalt underrum. Dette indebærer specielt, at der eksisterer  $m$  ortogonale egenvektorer svarende til  $\lambda$ .

Ved en kombineret af disse 2 sætninger bringes man let til at indse følgende

Corollar For en vilkårlig symmetrisk matrix  $\underline{A}$  eksisterer der en basis for  $\mathbb{R}^n$  bestående af indbyrdes ortogonale egenvektorer til  $\underline{A}$ .

Normeres vektorerne i en sådan basis bestående af ortogonale egenvektorer fås en ortonormal basis  $\{p_1, \dots, p_n\}$ . Sætter vi  $\underline{P}$  lig den  $n \times n$  matrix, hvis søjler netop er koordinaterne til disse vektorer, d.v.s.

$$\underline{P} = (p_1 \dots p_n)$$

fås

$$\underline{P}'\underline{P} = \underline{I}.$$

$\underline{P}$  er altså en ortogonal matrix, og

$$\underline{A}\underline{P} = \underline{P}\underline{\Lambda}$$

hvor  $\underline{\Lambda}$  er en diagonalmatrix med egenværdierne for  $\underline{A}$  (gentaget efter multiplicitet) i diagonalen. Ved hjælp af dette fås følgende

Sætning 1.10 Lad  $\underline{A}$  være en symmetrisk matrix. Da eksisterer der en ortogonal matrix  $\underline{P}$ , således at

$$\underline{P}' \underline{A} \underline{P} = \underline{\Lambda}$$

hvor  $\underline{\Lambda}$  er en diagonal matrix med  $\underline{A}$ 's egenverdier i diagonalen (gentaget så ofte multipliciteten angiver). Som  $\underline{P}$  kan man vælge en matrix, hvis søjler er ortonormerede egenvektorer til  $\underline{A}$ .

Bevis Trivielt følge af ovenstående relation. □

Sætning 1.11 Lad  $\underline{A}$  være en symmetrisk matrix med ikke-negative egenverdier. Da eksisterer der en regulær matrix  $\underline{B}$ , således at

$$\underline{B}' \underline{A} \underline{B} = \underline{E} ,$$

hvor  $\underline{E}$  er en diagonal matrix, der har 0'ler eller 1'ter i diagonalen. Antallet af 1'ter er lig  $\text{rg}(\underline{A})$ . Hvis  $\underline{A}$  har fuld rang, bliver  $\underline{E}$  specielt en enhedsmatrix.

Bevis Ved at multiplicere  $\underline{P}$  (til højre) med en diagonal matrix  $\underline{C}$ , der i diagonalen har elementerne

$$c_i = \begin{cases} \frac{1}{\sqrt{\lambda_i}} & \text{hvis } \lambda_i > 0 \\ 1 & \text{hvis } \lambda_i = 0 \end{cases} ,$$

fås sætningen umiddelbart med  $\underline{B} = \underline{P} \underline{C}$ . □

Relation i sætning 1.10 er ensbetydende med

$$\underline{A} = \underline{P} \underline{\Lambda} \underline{P}'$$



eller

$$\underline{A} = (\underline{p}_1 \dots \underline{p}_n) \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix} \begin{bmatrix} \underline{p}'_1 \\ \vdots \\ \underline{p}'_n \end{bmatrix},$$

d.v.s. vi har følgende opspaltning af matricen

$$\underline{A} = \lambda_1 \underline{p}_1 \underline{p}'_1 + \dots + \lambda_n \underline{p}_n \underline{p}'_n.$$

Denne opspaltning af den symmetriske matrix  $\underline{A}$  kaldes ofte dens spektral fremstilling, idet egenværdierne  $\{\lambda_1, \dots, \lambda_n\}$  betegnes matricens spektrum.

Med den åbenbare definition af  $\underline{\Delta}^h$ , nemlig som  $\text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$ , ser vi, at vi kan skrive

$$\underline{A} = (\underline{p} \underline{\Delta}^h) (\underline{p} \underline{\Delta}^h)' = \underline{G} \underline{G}'.$$

Det kan her indskydes, at hvis  $\underline{A}$  er positivt definit, findes der en fremstilling

$$\underline{A} = \underline{L} \underline{L}',$$

hvor  $\underline{L}$  er en nedre trekantmatrix. Denne fremstilling kaldes Cholesky faktoriseringen af  $\underline{A}$  (se f.eks. Wilkinson (1960)).

Endelig har vi

Sætning 1.12 Lad  $\underline{A}$  være en regulær, symmetrisk matrix. Da har  $\underline{A}$  og  $\underline{A}^{-1}$  de samme egenvektorer, men svarende til reciproke egenværdier.

1.40

Bevis Lad  $\lambda$  være en egenværdi for  $\underline{A}$  og  $\underline{x}$  en tilhørende egenvektor, d.v.s.

$$\underline{Ax} = \lambda \underline{x} .$$

Da  $\underline{A}$  er regulær, er dette ensbetydende med

$$\underline{A}^{-1} \underline{x} = \frac{1}{\lambda} \underline{x} ,$$

hvilket skulle vises. □

Endelig bemærkes, at

$$\det \underline{A} = \prod_1 \lambda_1 .$$

Eksempel 1.16 Ortogonal transformationer af planen. For at give en geometrisk forståelse af de transformationer, der reducerer en symmetrisk matrix til diagonalform, gennemgår vi de ortogonale transformationer af planen.

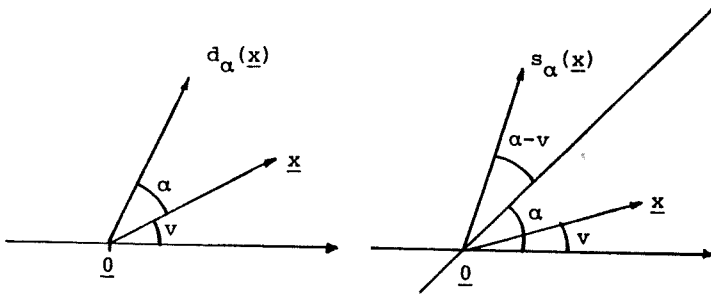
Ved at udnytte ortogonalitetsbetingelserne  $\underline{P}'\underline{P} = \underline{I}$  ses let, at de eneste ortogonale  $2 \times 2$ -matricer er matricer af formen

$$\begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix} \text{ og } \begin{bmatrix} \cos \alpha & \sin \alpha \\ \sin \alpha & -\cos \alpha \end{bmatrix} .$$

Vi vil nu vise, at disse svarer til drejninger (rotationer) om origo og spejlinger (refleksioner) i rette linier.

Vi gør det ved at bestemme koordinatudtryk for de lineære afbildninger  $d_\alpha$  og  $s_\alpha$ , der henholdsvis repræsenterer en drejning på vinklen  $\alpha$  af planen og en spejling i linien, der danner vinklen  $\alpha$  med 1. ste akse.

Afbildningerne er skitseret i figur 1.7



Figur 1.7 Drejning og spejling bestemt af vinklen  $\alpha$ .

Idet  $\underline{x} = r(\cos v, \sin v)'$ , hvor  $r$  kan sættes lig 1, har vi

$$\begin{aligned} d_{\alpha}(\underline{x}) &= \begin{bmatrix} \cos(\alpha + v) \\ \sin(\alpha + v) \end{bmatrix} = \begin{bmatrix} \cos \alpha \cos v - \sin \alpha \sin v \\ \sin \alpha \cos v + \cos \alpha \sin v \end{bmatrix} \\ &= \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} \cos v \\ \sin v \end{bmatrix}. \end{aligned}$$

Heraf ses, at  $d_{\alpha}$  har matrix fremstillingen

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \rightarrow \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Analogt fås

$$\begin{aligned} s_{\alpha}(\underline{x}) &= \begin{bmatrix} \cos(2\alpha - v) \\ \sin(2\alpha - v) \end{bmatrix} = \begin{bmatrix} \cos 2\alpha \cos v + \sin 2\alpha \sin v \\ \sin 2\alpha \cos v - \cos 2\alpha \sin v \end{bmatrix} \\ &= \begin{bmatrix} \cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & -\cos 2\alpha \end{bmatrix} \begin{bmatrix} \cos v \\ \sin v \end{bmatrix}, \end{aligned}$$

hvorfor  $s_{\alpha}$  har matrix fremstillingen

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \rightarrow \begin{bmatrix} \cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & -\cos 2\alpha \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Hermed er påstanden i indledningen bevist.

Det kan endvidere være nyttigt at erindre sig følgende relationer mellem drejninger og spejlinger af planen

$$s_{\frac{\pi}{4}} \circ d_{\alpha} = s_{\frac{\pi}{4} - \frac{\alpha}{2}}$$

$$s_{\alpha} = s_{\frac{\pi}{4}} \circ d_{\frac{\pi}{2} - 2\alpha}.$$

Den første relation følger af

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix} = \begin{bmatrix} \sin \alpha & \cos \alpha \\ \cos \alpha & \sin \alpha \end{bmatrix} = \begin{bmatrix} \cos(\frac{\pi}{4} - \alpha) & \sin(\frac{\pi}{4} - \alpha) \\ \sin(\frac{\pi}{4} - \alpha) & -\cos(\frac{\pi}{4} - \alpha) \end{bmatrix}.$$

Den sidste af de to relationer fås af den første ved at erstatte  $\alpha$  med  $\frac{\pi}{2} - 2\alpha$ . □

I det næste afsnit skal vi bl.a. betragte problemet med at generalisere spektralfremstillingen til en vilkårlig matrix.

#### 1.4.2 Singular-værdi dekomposition af vilkårlig matrix. Q- og R-modus analyser.

Vi giver først hovedresultatet, der også kendes under navnet Eckart-Young's sætning.

Sætning 1.13 Lad  $\underline{x}$  være en vilkårlig  $n \times p$  matrix med rang  $r$ . Da eksisterer der ortogonale matricer  $\underline{U}$  ( $p \times p$ ) og  $\underline{V}$  ( $n \times n$ ) samt positive tal  $\gamma_1, \dots, \gamma_r$ , således at

$$\underline{x} = \underline{V} \underline{\Gamma} \underline{U}' = \begin{bmatrix} \underline{v}_1 & \dots & \underline{v}_r \end{bmatrix} \begin{bmatrix} \gamma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \gamma_r \end{bmatrix} \begin{bmatrix} \underline{u}_1' \\ \vdots \\ \underline{u}_r' \end{bmatrix} = \gamma_1 \underline{v}_1 \underline{u}_1' + \dots + \gamma_r \underline{v}_r \underline{u}_r' ,$$

hvor  $\underline{\Gamma} = \text{diag}(\gamma_1, \dots, \gamma_r)$  og  $\underline{v}_1, \dots, \underline{v}_r$  er søjlerne i  $\underline{V}$  og  $\underline{u}_1, \dots, \underline{u}_r$  søjlerne i  $\underline{U}$ .

Bevis Forbigås. Se f.eks. Johnson (1963). □

Tallene  $\gamma_1, \dots, \gamma_r$  kaldes  $\underline{x}$ 's singulære værdier.

I det følgende vil vi nu undersøge sammenhængen mellem  $\underline{x}$ 's singulære værdier og egenværdiproblemerne for de symmetriske matrixer  $\underline{x}\underline{x}'$  ( $n \times n$ ) og  $\underline{x}'\underline{x}$  ( $p \times p$ ).

Vi anfører dog først

Sætning 1.14 For en vilkårlig (reel) matrix  $\underline{x}$  gælder, at  $\underline{x}'\underline{x}$  og  $\underline{x}\underline{x}'$  har ikke-negative egenværdier og

$$\text{rg}(\underline{x}'\underline{x}) = \text{rg}(\underline{x}\underline{x}') = \text{rg}(\underline{x})$$

Bevis Det er tilstrækkeligt at vise resultaterne for  $\underline{x}'\underline{x}$ .

Det er åbenbart, at  $\underline{x}'\underline{x}$  er symmetrisk, hvorfor der eksisterer en ortogonal matrix  $\underline{P}$ , så

$$\underline{P}' \underline{x}' \underline{x} \underline{P} = \underline{\Lambda}$$

d.v.s.

$$(\underline{x}\underline{P})' (\underline{x}\underline{P}) = \underline{\Lambda} ,$$

1.44

Sættes  $\underline{\underline{x}}\underline{\underline{P}} = \underline{\underline{B}} = (b_{ij})$ , fås  $\underline{\underline{B}}'\underline{\underline{B}} = \underline{\underline{\Lambda}}$ , d.v.s.

$$\lambda_i = \sum_j b_{ij}^2 > 0,$$

d.v.s.  $\underline{\underline{x}}'\underline{\underline{x}}$  har ikke-negative egenvektorer. Det ses endvidere, at

$$\begin{aligned} \text{rg}(\underline{\underline{x}}'\underline{\underline{x}}) &= \text{card} \{ \lambda_i \neq 0 \} \\ &= \text{card} \{ \text{søjler } \underline{\underline{b}}_j \text{ i } \underline{\underline{B}}, \text{ der er } \neq \underline{\underline{0}} \} \end{aligned}$$

Da  $\underline{\underline{b}}_i'\underline{\underline{b}}_j = 0$  for  $i \neq j$  (ifølge (\*)) er derfor

$$\text{rg}(\underline{\underline{x}}'\underline{\underline{x}}) = \text{rg}(\underline{\underline{B}}).$$

Da  $\underline{\underline{P}}$  er regulær, får vi af et resultat p. 1.16, at

$$\text{rg}(\underline{\underline{B}}) = \text{rg}(\underline{\underline{x}}\underline{\underline{P}}) = \text{rg}(\underline{\underline{x}}).$$

□

Vi anfører et lille corollar til sætningen.

Corollar Lad  $\underline{\underline{\Sigma}}$  være symmetrisk og positivt definit. Da er for en vilkårlig matrix  $\underline{\underline{x}}$

$$\text{rg}(\underline{\underline{x}}'\underline{\underline{\Sigma}}^{-1}\underline{\underline{x}}) = \text{rg } \underline{\underline{x}},$$

naturligvis forudsat, at de involverede produkter eksisterer.

Bevis Da  $\underline{\underline{\Sigma}}^{-1}$  også er regulær og positivt definit, eksisterer der en ortogonal matrix  $\underline{\underline{P}}$ , så

$$\underline{\underline{P}}'\underline{\underline{\Sigma}}^{-1}\underline{\underline{P}} = \underline{\underline{\Lambda}},$$

hvor  $\underline{\underline{\Lambda}}$  er en diagonalmatrix. Dette medfører

$$\underline{\Sigma}^{-1} = \underline{P}\underline{\Lambda}\underline{P}' = \underline{P}\underline{\Lambda}^h\underline{\Lambda}^h\underline{P}' = \underline{P}\underline{\Lambda}^h(\underline{P}\underline{\Lambda}^h)' = \underline{B}\underline{B}'.$$

Her betegner  $\underline{\Lambda}^h$  den diagonalmatrix, hvis diagonal elementer er kvadratroden af de tilsvarende elementer i  $\underline{\Lambda}$ . Det er tri-  
vielt, at  $\underline{B}$  er regulær. Denne relation indsættes, og vi får

$$\underline{x}'\underline{\Sigma}^{-1}\underline{x} = \underline{x}'\underline{B}\underline{B}'\underline{x} = (\underline{B}'\underline{x})'\underline{B}'\underline{x} ,$$

d.v.s.

$$\text{rg}(\underline{x}'\underline{\Sigma}^{-1}\underline{x}) = \text{rg}(\underline{B}'\underline{x}) = \text{rg} \underline{x} ,$$

hvilket skulle vises. □

Vi har nu med betegnelserne fra sætning 1.14.

#### Sætning 1.15

- i) Matricen  $\underline{x}\underline{x}'$  ( $n \times n$ ) har  $r$  positive egenværdier og  $n-r$  egenværdier lig 0. De positive egenværdier er  $\gamma_1^2, \dots, \gamma_r^2$ , hvor  $\gamma_1, \dots, \gamma_r$  er de singulære værdier for  $\underline{x}$ . De tilsvarende egenvektorer er  $\underline{v}_1, \dots, \underline{v}_r$ .
- ii) Tilsvarende har  $\underline{x}'\underline{x}$  ( $p \times p$ )  $r$  positive og  $(p-r)$  0-egen-  
værdier. De positive egenværdier er  $\gamma_1^2, \dots, \gamma_r^2$  og de  
tilsvarende egenvektorer er  $\underline{u}_1, \dots, \underline{u}_r$ .
- iii) De positive egenværdier for  $\underline{x}\underline{x}'$  og  $\underline{x}'\underline{x}$  er altså ens  
og relationen mellem de tilsvarende egenvektorer er  
( $m = 1, \dots, r$ )

$$\underline{v}_m = \frac{1}{\gamma_m} \underline{x} \underline{u}_m \text{ og } \underline{u}_m = \frac{1}{\gamma_m} \underline{x}' \underline{v}_m ,$$

eller på kompakt form

$$\underline{V} = \underline{X}\underline{U}\underline{\Gamma}^{-1} \text{ og } \underline{U} = \underline{X}'\underline{V}\underline{\Gamma}^{-1}$$

Bevis Følger ved anvendelse af Eckart-Young's sætning. □

Bemærkning 1 Analysen af matricen  $\underline{x}'\underline{x}$  kaldes R-modus analyse og analysen af  $\underline{xx}'$  Q-modus analyse, betegnelser, der stammer fra faktoranalysen, jvf. kapitel 8.

Bemærkning 2 Det fremgår af sætningen, at man kan opnå resultaterne fra en R-modus analyse ud fra en Q-modus analyse og omvendt. Ved en konkret anvendelse bør man derfor betragte den af matricerne  $\underline{x}'\underline{x}$  og  $\underline{xx}'$ , der er af mindst orden.

### 1.4.3 Kvadratiske former og positivt definitte matricer

I dette afsnit betragter vi fremdeles kun symmetriske matricer.

Ved den kvadratiske form svarende til den symmetriske matrix  $\underline{A}$  forstås afbildningen

$$\underline{x} \rightarrow \underline{x}'\underline{A}\underline{x} = \sum_{i=1}^n a_{ii}x_i^2 + 2 \sum_{i < j} a_{ij}x_i x_j .$$

Vi siger, at en symmetrisk matrix  $\underline{A}$  er positivt definit, respektive positivt semidefinit, hvis den tilhørende kvadratiske form er positiv, respektive ikke negativ, i vektorer forskellige fra 0-vektoren, i.e. såfremt

$$\forall \underline{x} \neq \underline{0} : \underline{x}'\underline{A}\underline{x} > 0 ,$$

respektive



$$\forall \underline{x} \neq 0 : \underline{x}'\underline{A}\underline{x} > 0.$$

Vi siger da også, at den kvadratiske form er positivt definit, respektive positivt semidefinit.

Vi har følgende

Sætning 1.16 Den symmetriske matrix  $\underline{A}$  er positivt definit, respektive semidefinit, hvis alle  $\underline{A}$ 's egenverdier er positive, respektive ikke negative.

Bevis Vi har med  $\underline{P}$  som i sætning 1.10

$$\begin{aligned} \underline{x}'\underline{A}\underline{x} &= \underline{x}'\underline{P}'\underline{P}\underline{A}\underline{P}\underline{P}'\underline{x} = (\underline{P}'\underline{x})'\underline{\Delta}(\underline{P}'\underline{x}) \\ &= \underline{y}'\underline{\Delta}\underline{y} = \lambda_1 y_1^2 + \dots + \lambda_n y_n^2 \end{aligned}$$

□

Et andet nyttigt resultat er

Sætning 1.17 En symmetrisk  $n \times n$  matrix  $\underline{A}$  er positivt definit, netop hvis alle hovedunderdeterminanter

$$d_i = \det \begin{bmatrix} a_{11} & \dots & a_{1i} \\ \vdots & & \vdots \\ a_{i1} & \dots & a_{ii} \end{bmatrix}, \quad i = 1, \dots, n,$$

er positive.

Bevis Forbigås

Vi anfører nu en meget vigtig sætning om ekstrema af kvadratiske former

Sætning 1.18 Sættes egenværdierne for den symmetriske matrix  $\underline{A}$  lig  $\lambda_1 \geq \dots \geq \lambda_n$  med tilhørende egenvektorer  $\underline{p}_1, \dots, \underline{p}_n$ , og defineres

$$R(\underline{x}) = \frac{\underline{x}' \underline{A} \underline{x}}{\underline{x}' \underline{x}} \quad ,$$

samt

$$M_k = \{ \underline{x} \mid \underline{x}' \underline{p}_i = 0, i = 1, \dots, k-1 \} \quad ,$$

da gælder, at

$$\sup_{\underline{x}} R(\underline{x}) = R(\underline{p}_1) = \lambda_1 \quad ,$$

$$\inf_{\underline{x}} R(\underline{x}) = R(\underline{p}_n) = \lambda_n \quad ,$$

$$\sup_{\substack{\underline{x} \\ \underline{x} \in M_k}} R(\underline{x}) = R(\underline{p}_k) = \lambda_k \quad .$$

Bevis En vilkårlig vektor  $\underline{x}$  kan skrives

$$\underline{x} = \alpha_1 \underline{p}_1 + \dots + \alpha_n \underline{p}_n \quad .$$

Hvis  $\underline{p}_i' \underline{x} = 0$ ,  $i = 1, \dots, k-1$ , fås  $\alpha_1 = \dots = \alpha_{k-1} = 0$ , d.v.s.

$$\underline{x} = \alpha_k \underline{p}_k + \dots + \alpha_n \underline{p}_n \quad .$$

Følgelig er

$$\underline{x}' \underline{A} \underline{x} = \alpha_k^2 \lambda_k + \dots + \alpha_n^2 \lambda_n \quad ,$$

og

$$R(\underline{x}) = \frac{\underline{x}'\underline{A}\underline{x}}{\underline{x}'\underline{x}} = \frac{\alpha_k^2 \lambda_k + \dots + \alpha_n^2 \lambda_n}{\alpha_k^2 + \dots + \alpha_n^2}$$

Det er klart, at dette udtryk er maksimalt for

$$(\alpha_k, \dots, \alpha_n) = (\alpha_k, 0, \dots, 0),$$

hvor det antager værdien  $\lambda_k$ . Resultatet med inf vises analogt.

□

Bemærkning 1 Sætningen giver for  $k = 1$ , at den enhedsvektor, i.e. den "retning", hvor den kvadratiske form antager sin største værdi, netop er egenvektoren svarende til den største egenværdi. Betragtes den kvadratiske form kun i enhedsvektorer, der er ortogonale på egenvektorer svarende til de  $k-1$  største egenværdier siger sætningen, at maksimum fås i den retning, der svarer til egenvektoren hørende til den  $k$ 'te største egenværdi.

Bemærkning 2  $R(\underline{x})$  kaldes også Rayleigh's kvotient.

Vi vil nu beskrive niveaukurverne for positivt definitte kvadratiske former.

Sætning 1.19 Lad  $\underline{A}$  være en positiv definit. Da er løsningsmængden til ligningen

$$\underline{x}'\underline{A}\underline{x} = c, \quad c > 0, \quad ,$$

en ellipsoide med hovedakser i egenvektorernes retninger. Første hovedakse svarer til den mindste egenværdi, anden til den næstmindste etc.

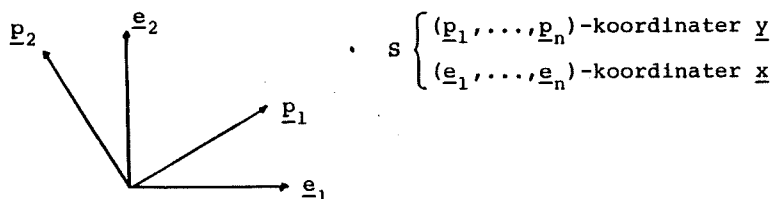
Bevis Vi betragter matricen  $\underline{P} = (\underline{p}_1 \dots \underline{p}_n)$ , hvis søjler er koordinaterne til orthonormerede egenvektorer til  $\underline{A}$ . Da gælder med  $\underline{y} = \underline{P}'\underline{x}$

$$\begin{aligned} \underline{x}'\underline{A}\underline{x} &= \underline{y}'\underline{\Delta}\underline{y} \\ &= \lambda_1 y_1^2 + \dots + \lambda_n y_n^2 \\ &= \frac{y_1^2}{(1/\sqrt{\lambda_1})^2} + \dots + \frac{y_n^2}{(1/\sqrt{\lambda_n})^2} \quad (*) \end{aligned}$$

Matrixligningen

$$\underline{y} = \underline{P}'\underline{x} \Leftrightarrow \underline{x} = \underline{P}\underline{y}$$

svarer til et basisskifte fra den oprindelige orthonormale basis  $\{\underline{e}_1, \dots, \underline{e}_n\}$  til den orthonormale basis  $\{\underline{p}_1, \dots, \underline{p}_n\}$ .



Figur 1.8 Skitse visende basisskift.

Lad nemlig  $S$  være et punkt, hvis  $\{\underline{e}_1, \dots, \underline{e}_n\}$  - koordinater kaldes  $\underline{x}$  og hvis  $\{\underline{p}_1, \dots, \underline{p}_n\}$  - koordinater kaldes  $\underline{y}$ . Da gælder

$$x_1 \underline{e}_1 + \dots + x_n \underline{e}_n = y_1 \underline{p}_1 + \dots + y_n \underline{p}_n \quad ,$$

eller

$$(\underline{e}_1 \dots \underline{e}_n) \underline{x} = (\underline{p}_1 \dots \underline{p}_n) \underline{y} \quad ,$$

d.v.s.

$$\underline{I} \underline{x} = \underline{P} \underline{y} \quad ,$$

hvor  $\underline{I}$  er en enhedsmatrix.

Udtrykket i (\*) viser derfor løsningsmængdens ligning i y-koordinater svarende til koordinatsystemet bestående af ortonormerede egenvektorer. Dette viser, at der er tale om en ellipsoide. Resten af sætningen følger nu ved at bemærke, at 1. ste hovedakse svarer til det  $y_1$ , for hvilken  $1/\sqrt{\lambda_1}$  er maksimal, d.v.s. for hvilken  $\lambda_1$  er minimal.

□

Bemærkning Hvis matricen kun er positiv semidefinit svarer løsningsmængden til ligningen til en elliptisk cylinder. Foretager vi nemlig et basisskifte til basen  $\{\underline{p}_1, \dots, \underline{p}_n\}$  bestående af ortonormale egenvektorer, hvor vi for simpelheds skyld antager, at  $\underline{p}_1, \dots, \underline{p}_r$  svarer til de fra 0 forskellige egenværdier, får vi

$$\underline{x}' \underline{A} \underline{x} = c \Leftrightarrow \lambda_1 y_1^2 + \dots + \lambda_r y_r^2 + 0 y_{r+1}^2 + \dots + 0 y_n^2 = c$$

$$\Leftrightarrow \lambda_1 y_1^2 + \dots + \lambda_r y_r^2 = c \quad .$$

Heraf følger påstanden. Betragter vi den kvadratiske forms restriktion til underrummet udspændt af egenvektorerne svarende til egenværdier  $> 0$ , bliver løsningsmængden en ellipsoide.

□

Eksempel 1.7 Vi betragter den symmetriske positivt definitte matrix

$$\underline{A} = \begin{bmatrix} 3 & \sqrt{2} \\ \sqrt{2} & 2 \end{bmatrix} .$$

Den til  $\underline{A}$  svarende kvadratiske form er

$$\underline{x}'\underline{A}\underline{x} = 3x_1^2 + 2x_2^2 + 2\sqrt{2}x_1x_2 ,$$

således at enhedsellipsen svarende til  $\underline{A}$  er løsningsmængden til ligningen

$$3x_1^2 + 2x_2^2 + 2\sqrt{2}x_1x_2 = 1 .$$

For at bestemme hovedakserne finder vi  $\underline{A}$ 's egenverdier. Vi finder

$$\det(\underline{A} - \lambda \underline{I}) = 0 \Leftrightarrow \lambda^2 - 5\lambda + 4 = 0$$

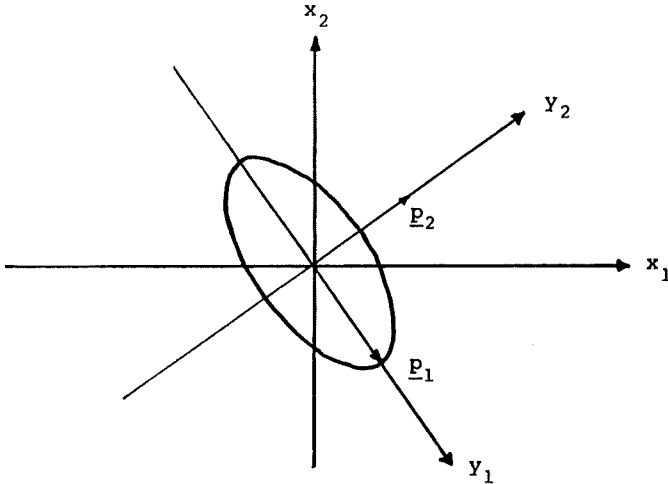
$$\Leftrightarrow \lambda = 1 \vee \lambda = 4 .$$

Egenvektorer svarende til  $\lambda = 1$  henholdsvis  $\lambda = 4$  ses at være af formen  $t(1, -\sqrt{2})$  henholdsvis  $t(1, \sqrt{2}/2)$ . Normeres disse fås

$$\underline{p}_1 = \begin{bmatrix} \frac{\sqrt{3}}{3} \\ -\frac{\sqrt{6}}{3} \end{bmatrix} , \quad \underline{p}_2 = \begin{bmatrix} \frac{\sqrt{6}}{3} \\ \frac{\sqrt{3}}{3} \end{bmatrix} .$$

Vælges basen  $\{\underline{p}_1, \underline{p}_2\}$ , bliver koordinatfremstillingen for den kvadratiske form

$$y + y_1^2 + 4y_2^2 ,$$



Figur 1.9 Ellipse bestemt af den i eksempel 1.7 anførte kvadratiske form.

Ellipsen får ligningen

$$\frac{y_1^2}{1^2} + \frac{y_2^2}{(\frac{1}{2})^2} = 1 \quad .$$

Den er skitseret i figur 1.9.

Da

$$\begin{aligned} \underline{e}_1 &= \begin{bmatrix} \frac{\sqrt{3}}{3} \\ -\frac{\sqrt{6}}{3} \end{bmatrix} \approx \begin{bmatrix} 0.577 \\ -0.820 \end{bmatrix} \\ &\approx \begin{bmatrix} \cos(-54.7^\circ) \\ \sin(-54.7^\circ) \end{bmatrix} \quad , \end{aligned}$$

svarer det nye koordinatsystem til en drejning af det gamle med vinklen  $-54.7^\circ$ .

□

#### 1.4.4 Det generelle egenværdiproblem for symmetriske matricer

I forbindelse med teorien for kanoniske korrelationer og i diskriminantanalysen får vi brug for et lidt mere generelt egenværdibegreb end det hidtil betragtede. Vi indfører begrebet i

Definition 1.3 Lad  $\underline{\underline{A}}$  og  $\underline{\underline{B}}$  være reelle  $m \times m$  symmetriske matricer og lad  $\underline{\underline{B}}$  være af fuld rang. Et tal  $\lambda$ , for hvilket

$$\det(\underline{\underline{A}} - \lambda \underline{\underline{B}}) = 0 \quad ,$$

benævnes en egenværdi af  $\underline{\underline{A}}$  m.h.t.  $\underline{\underline{B}}$ . For et sådant  $\lambda$  findes  $\underline{\underline{x}} \neq 0$  så

$$\underline{\underline{A}}\underline{\underline{x}} = \lambda \underline{\underline{B}}\underline{\underline{x}} \quad .$$

En sådan vektor  $\underline{\underline{x}}$  kaldes en egenvektor for  $\underline{\underline{A}}$  m.h.t.  $\underline{\underline{B}}$ .

Bemærkning De anførte begreber føres umiddelbart tilbage til egenværdi og egenvektor for den ikke-symmetriske matrix  $\underline{\underline{B}}^{-1}\underline{\underline{A}}$ .

Sætning 1.20 Vi betragter igen situationen i definition 1.3, og lad yderligere  $\underline{\underline{B}}$  være positivt definit. Der findes da  $m$  reelle egenværdier af  $\underline{\underline{A}}$  m.h.t.  $\underline{\underline{B}}$ . Hvis  $\underline{\underline{A}}$  er positivt semidefinit vil disse være ikke-negative, og hvis  $\underline{\underline{A}}$  er positivt definit, vil de være positive.

Bevis Ifølge sætning 1.11 findes en matrix  $\underline{\underline{I}}$  med

$$\underline{\underline{T}}'\underline{\underline{B}}\underline{\underline{T}} = \underline{\underline{I}}.$$

Vi sætter



$$\underline{\underline{D}} = \underline{\underline{T}}' \underline{\underline{A}} \underline{\underline{T}} .$$

$\underline{\underline{D}}$  er åbenbart symmetrisk, og da

$$\underline{\underline{x}}' \underline{\underline{D}} \underline{\underline{x}} = (\underline{\underline{T}} \underline{\underline{x}})' \underline{\underline{A}} (\underline{\underline{T}} \underline{\underline{x}}),$$

ses at  $\underline{\underline{D}}$  og  $\underline{\underline{A}}$  er positivt semidefinitte, resp. positivt definitte samtidigt.

Nu er

$$\begin{aligned} (\underline{\underline{D}} - \lambda \underline{\underline{I}}) \underline{\underline{v}} = 0 &\Leftrightarrow (\underline{\underline{T}}' \underline{\underline{A}} \underline{\underline{T}} - \lambda \underline{\underline{T}}' \underline{\underline{B}} \underline{\underline{T}}) \underline{\underline{v}} = 0 \\ &\Leftrightarrow (\underline{\underline{A}} - \lambda \underline{\underline{B}}) (\underline{\underline{T}} \underline{\underline{v}}) = 0 . \end{aligned}$$

Heraf følger, at  $\underline{\underline{D}}$ 's egenverdier er lig  $\underline{\underline{A}}$ 's egenverdier m.h.t.  $\underline{\underline{B}}$ , og at egenvektorerne for  $\underline{\underline{A}}$  m.h.t.  $\underline{\underline{B}}$  fås ved at anvende transformationen  $\underline{\underline{T}}$  på  $\underline{\underline{D}}$ 's egenvektorer. Resultatet vedrørende egenverdiernes fortegn følger trivielt.

□

**Sætning 1.21** Lad situationen være som ovenfor. Da eksisterer en basis for  $\mathbb{R}^m$  bestående af egenvektorer  $\underline{\underline{u}}_1, \dots, \underline{\underline{u}}_m$  af  $\underline{\underline{A}}$  m.h.t.  $\underline{\underline{B}}$ . Disse vektorer kan vælges konjugerede såvel m.h.t.  $\underline{\underline{A}}$  som m.h.t.  $\underline{\underline{B}}$ , d.v.s.

$$\underline{\underline{u}}_i' \underline{\underline{A}} \underline{\underline{u}}_j = \underline{\underline{u}}_i' \underline{\underline{B}} \underline{\underline{u}}_j = 0 .$$

**Bevis** Følger af beviset for ovenstående sætning og af corollar til sætning 1.9, når det erindres, at

$$0 = \underline{\underline{v}}_i' \underline{\underline{v}}_j = (\underline{\underline{v}}_i' \underline{\underline{T}}') \underline{\underline{T}}'^{-1} \underline{\underline{T}}^{-1} (\underline{\underline{T}} \underline{\underline{v}}_j) = \underline{\underline{u}}_i' \underline{\underline{B}} \underline{\underline{u}}_j ,$$

hvor  $\underline{v}_1, \dots, \underline{v}_m$  er en ortonormal basis for  $R^m$  bestående af egenvektorer til  $\underline{D}$ .

Endelig er

$$\underline{u}'_i \underline{A} \underline{u}_j = \lambda_j \underline{u}'_i \underline{B} \underline{u}_j = 0$$

□

Sætning 1.22 Lad  $\underline{A}$  være symmetrisk og lad  $\underline{B}$  være positivt definit. Da eksisterer en regulær matrix  $\underline{R}$  med

$$\underline{R}' \underline{A} \underline{R} = \underline{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n) \quad ,$$

og

$$\underline{R}' \underline{B} \underline{R} = \underline{I} \quad ,$$

hvor  $\lambda_1, \dots, \lambda_n$  er egenværdierne for  $\underline{A}$  m.h.t.  $\underline{B}$ . Benævnes den  $i$ 'te søjle i  $\underline{R}^{-1}$   $\underline{s}_i$  kan disse relationer skrives

$$\underline{A} = \lambda_1 \underline{s}_1 \underline{s}'_1 + \dots + \lambda_m \underline{s}_m \underline{s}'_m \quad ,$$

og

$$\underline{B} = \underline{s}_1 \underline{s}'_1 + \dots + \underline{s}_m \underline{s}'_m \quad .$$

Bevis Vi betragter det i beviset for sætning 1.20 anførte  $\underline{D} = \underline{T}' \underline{A} \underline{T}$ . Da  $\underline{D}$  er symmetrisk, findes ifølge sætning 1.10 en ortogonal matrix  $\underline{C}$  med

$$\underline{C}' \underline{D} \underline{C} = \underline{\Lambda} \quad ,$$

idet vi erindrer, at  $\underline{D}$ 's egenværdier netop er  $\underline{A}$ 's egenværdier m.h.t.  $\underline{B}$ .

Vælges  $\underline{R} = \underline{C}\underline{C}$ , fås

$$\underline{R}'\underline{B}\underline{R} = \underline{C}'\underline{T}'\underline{B}\underline{C} = \underline{C}'\underline{C} = \underline{I} \quad ,$$

og

$$\underline{B}'\underline{A}\underline{B} = \underline{C}'\underline{T}'\underline{A}\underline{C} = \underline{C}'\underline{D}\underline{C} = \underline{A} \quad .$$

□

Endelig anfører vi en analog til sætning 1.18, nemlig

Sætning 1.23 Lad  $\underline{A}$  være positivt semidefinit og  $\underline{B}$  positivt definit. Lad  $\underline{A}$ 's egenværdier m.h.t.  $\underline{B}$  være  $\lambda_1 \geq \dots \geq \lambda_m$  og lad  $\underline{v}_1, \dots, \underline{v}_m$  betegne en basis for  $R^m$  bestående af de tilsvarende egenvektorer med  $\underline{v}_i' \underline{B} \underline{v}_j = 0$   $i \neq j$ . Vi sætter

$$R(\underline{x}) = \frac{\underline{x}' \underline{A} \underline{x}}{\underline{x}' \underline{B} \underline{x}} \quad ,$$

samt

$$M_k = \{ \underline{x} \mid \underline{x}' \underline{B} \underline{v}_1 = \dots = \underline{x}' \underline{B} \underline{v}_{k-1} = 0 \} \quad ,$$

og har da

$$\sup_{\underline{x}} R(\underline{x}) = R(\underline{v}_1) = \lambda_1$$

$$\inf_{\underline{x}} R(\underline{x}) = R(\underline{v}_m) = \lambda_m$$

$$\sup_{\underline{x} \in M_k} R(\underline{x}) = R(\underline{v}_k) = \lambda_k \quad .$$

Bevis Uden tab af generalitet kan  $\underline{v}_i$ 'erne vælges så  $\underline{v}_i' \underline{B} \underline{v}_i = 1$ , og da en vilkårlig vektor  $\underline{x}$  kan skrives

$$\underline{x} = \alpha_1 \underline{v}_1 + \dots + \alpha_m \underline{v}_m \quad ,$$

fås

$$R(\underline{x}) = \frac{\sum \alpha_i^2 \underline{v}_i' \underline{A} \underline{v}_i}{\sum \alpha_i^2 \underline{v}_i' \underline{B} \underline{v}_i} = \frac{\sum \lambda_i \alpha_i^2}{\sum \alpha_i^2} \quad .$$

Heraf følger de to første påstande let. Hvis  $\underline{x} \in M_k$ , vil  $\underline{x}$  have fremstillingen

$$\underline{x} = \alpha_k \underline{v}_k + \dots + \alpha_m \underline{v}_m \quad ,$$

og

$$R(\underline{x}) = \frac{\lambda_k \alpha_k^2 + \dots + \lambda_m \alpha_m^2}{\alpha_k^2 + \dots + \alpha_m^2} \quad ,$$

hvoraf resultatet følger.

□

#### 1.4.5 Sporet af en matrix

Ved sporet af (den symmetriske) matrix  $\underline{A}$  forstås summen af diagonalelementerne, i.e.

$$\text{tr}(\underline{A}) = \sum_{i=1}^n a_{ii} \quad .$$

For matricer  $\underline{A}$  og  $\underline{B}$  gælder (kvadr. matr.)

$$\text{tr}(\underline{A}\underline{B}) = \text{tr}(\underline{B}\underline{A}) \quad . \quad (i)$$

Endvidere haves, at sporet er lig summen af egenverdierne, i.e.

$$\text{tr}(\underline{A}) = \sum_{i=1}^n \lambda_i \quad .$$

Dette er en triviell følge af (i) og sætning 1.10

For positivt semidefinitte matricer er sporet derfor et andet mål for "størrelsen" af en matrix. Er sporet stort, er i det mindste nogle egenverdier store. Til gengæld er dette mål ikke følsomt overfor, om enkelte egenverdier er 0, d.v.s. om matricen er udartet. Det er determinanten derimod, idet vi jo erindrer, at

$$\det(\underline{A}) = \prod_{i=1}^n \lambda_i .$$

Vi bemærker ydermere, at for en idempotent matrix A gælder

$$\text{tr}(\underline{A}) = \text{rg}(\underline{A}) .$$

Endvidere er

$$\text{tr}(\underline{B}\underline{B}^-) = \text{rg}(\underline{B}) ,$$

hvor B<sup>-</sup> er en vilkårlig pseudoinvers til B.

Til sidst nævner vi, at for en regulær matrix S gælder

$$\text{tr}(\underline{S}^{-1}\underline{B}\underline{S}) = \text{tr}(\underline{B}) .$$

#### 1.4.6 Differentiation af linearform og kvadratisk form

Lad  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . Vi anvender da følgende skrivemåde for vektoren af partielle afledede

$$\frac{\partial f(\underline{x})}{\partial \underline{x}} = \frac{\partial f}{\partial \underline{x}} = \begin{bmatrix} \frac{\partial f(\underline{x})}{\partial x_1} \\ \cdot \\ \cdot \\ \frac{\partial f(\underline{x})}{\partial x_n} \end{bmatrix} .$$

Der gælder nu følgende sætning om differentiation af visse former

Sætning 1.24 For en symmetrisk  $(n \times n)$ -matrix  $\underline{A}$  og en vilkårlig  $n$ -dimensional vektor  $\underline{b}$  gælder

$$(i) \quad \frac{\partial}{\partial \underline{x}} (\underline{b}' \underline{x}) = \underline{b}$$

$$(ii) \quad \frac{\partial}{\partial \underline{x}} (\underline{x}' \underline{x}) = 2 \underline{x}$$

$$(iii) \quad \frac{\partial}{\partial \underline{x}} (\underline{x}' \underline{A} \underline{x}) = 2 \underline{A} \underline{x} \quad .$$

Bevis Beviset for (i) og (ii) er trivielt. (iii) vises - bizart nok - bekvæmest ved hjælp af definitionen. For en vilkårlig vektor  $\underline{h}$  haves

$$(\underline{x} + \underline{h})' \underline{A} (\underline{x} + \underline{h}) = \underline{x}' \underline{A} \underline{x} + \underline{h}' \underline{A} \underline{h} + 2 \underline{h}' \underline{A} \underline{x}$$

Ved at vælge  $\underline{h} = (0, \dots, h_i, \dots, 0)'$  ses, at

$$\frac{\partial}{\partial x_i} (\underline{x}' \underline{A} \underline{x}) = 2 \sum_{j=1}^n a_{ij} x_j \quad ,$$

og heraf følger resultatet umiddelbart.

□

Vi belyser anvendelsen af sætningen i nedenstående

Eksempel 1.8 Vi vil finde minimum af funktionen

$$g(\underline{\theta}) = (\underline{y} - \underline{A}\underline{\theta})' \underline{B} (\underline{y} - \underline{A}\underline{\theta}) ,$$

hvor  $\underline{y}$ ,  $\underline{A}$  og  $\underline{B}$  er givne og  $\underline{B}$  endvidere positiv semidefinit

(og symmetrisk). Da  $g(\underline{\theta})$  er konveks (en paraboloid, eventuelt udartet), kan minimumspunktet findes ved at løse ligningen

$$\frac{\partial}{\partial \underline{\theta}} g(\underline{\theta}) = \underline{0} .$$

Vi foretager dog først en omskrivning af  $g$ . Vi har

$$\begin{aligned} g(\underline{\theta}) &= \underline{Y}' \underline{B} \underline{Y} - \underline{\theta}' \underline{A}' \underline{B} \underline{Y} + \underline{\theta}' \underline{A}' \underline{B} \underline{A} \underline{\theta} - \underline{Y}' \underline{B} \underline{A} \underline{\theta} \\ &= \underline{Y}' \underline{B} \underline{Y} - 2 \underline{Y}' \underline{B} \underline{A} \underline{\theta} + \underline{\theta}' \underline{A}' \underline{B} \underline{A} \underline{\theta} . \end{aligned}$$

Her er benyttet, at

$$\underline{\theta}' \underline{A}' \underline{B} \underline{Y} = \underline{Y}' \underline{B} \underline{A} \underline{\theta}$$

(begge  $1 \times 1$  matricer, d.v.s. en skalar, og hinandens transponerede). Heraf fås

$$\frac{\partial g}{\partial \underline{\theta}} = -2 \underline{A}' \underline{B} \underline{Y} + 2 \underline{A}' \underline{B} \underline{A} \underline{\theta} ,$$

og det ses, at

$$\frac{\partial g}{\partial \underline{\theta}} = \underline{0} \leftrightarrow \underline{A}' \underline{B} \underline{A} \underline{\theta} = \underline{A}' \underline{B} \underline{Y} .$$

Denne ligning har som sagt altid mindst en rod. Hvis  $\underline{A}' \underline{B} \underline{A}$  er regulær fås

$$\underline{\theta}_{\min} = (\underline{A}' \underline{B} \underline{A})^{-1} \underline{A}' \underline{B} \underline{Y} .$$

Hvis matricen er singulær, kan vi skrive

$$\underline{\theta}_{\min} = (\underline{A}' \underline{B} \underline{A})^{-} \underline{A}' \underline{B} \underline{Y} ,$$

hvor  $(\underline{A}' \underline{B} \underline{A})^{-}$  betegner en pseudoinvers til  $\underline{A}' \underline{B} \underline{A}$ .

□

1.62

Vi kan nu finde en alternativ beskrivelse af hovedakserne i en ellipsoide, idet vi har

Sætning 1.25 Lad  $\underline{A}$  være en positiv definit symmetrisk matrix. Hovedretningerne i ellipsoiden  $E_c$  med ligningen

$$\underline{x}'\underline{A}\underline{x} = c, \quad c > 0$$

er de retninger, hvor  $\underline{x}'\underline{x}$ ,  $\underline{x} \in E_c$ , har stationære punkter.

Bevis Vi kan antage, at  $x = 1$ . Vi skal da finde de stationære punkter for

$$f(\underline{x}) = \underline{x}'\underline{x}$$

under bibetingelsen

$$\underline{x}'\underline{A}\underline{x} = 1 \quad .$$

Vi anvender en Lagrange multiplikator teknik og definerer

$$\varphi(\underline{x}, \lambda) = \underline{x}'\underline{x} - \lambda(\underline{x}'\underline{A}\underline{x} - 1) \quad .$$

Ved differentiation fås

$$\frac{\partial \varphi}{\partial \underline{x}} = 2\underline{x} - 2\lambda\underline{A}\underline{x} \quad .$$

Hvis denne størrelse skal være  $\underline{0}$ , må

$$\underline{x} = \lambda\underline{A}\underline{x}$$

eller

$$\underline{A}\underline{x} = \frac{1}{\lambda} \underline{x} \quad ,$$



d.v.s.  $\underline{x}$  må være en egenvektor.

□

### 1.5 Tensor - eller Kronecker produkt af matricer

Dette produkt kan med fordel anvendes ved en behandling af den flerdimensionale generelle lineære model.

Definition 1.4 Lad  $\underline{A}$  være en  $m \times n$  matrix og  $\underline{B}$  en  $k \times l$  matrix. Ved tensor - eller Kronecker-produktet af  $\underline{A}$  og  $\underline{B}$  forstås matricen

$$\underline{A} \otimes \underline{B} = (a_{ij} \underline{B}) = \begin{bmatrix} a_{11} \underline{B} \dots a_{1n} \underline{B} \\ \vdots \\ a_{m1} \underline{B} \dots a_{mn} \underline{B} \end{bmatrix} \quad (*)$$

Dette begreb svarer til tensorproduktet af lineære afbildninger, som kan gives en koordinatuafhængig fremstilling (se f.eks. Bourbaki (1967)). Overføres dette til koordinatform kan enten anvendes (\*) eller, hvad der er helt equivalent hermed,  $\underline{A} \otimes \underline{B} = (\underline{A} b_{ij})$ . Dette svarer blot til at ændre koordinaternes rækkefølge, d.v.s. til at ombytte rækker og søjler i de involverede matricer.

Vi resumerer nu kort en række regneregler for tensorproduktet. Disse regneregler eftervises trivielt v.h.a. definitionen.

- |       |   |
|-------|---|
| (i)   | $\underline{0} \otimes \underline{A} = \underline{A} \otimes \underline{0} = \underline{0}$   |
| (ii)  | $(\underline{A}_1 + \underline{A}_2) \otimes \underline{B} = \underline{A}_1 \otimes \underline{B} + \underline{A}_2 \otimes \underline{B}$ |
| (iii) | $\underline{A} \otimes (\underline{B}_1 + \underline{B}_2) = \underline{A} \otimes \underline{B}_1 + \underline{A} \otimes \underline{B}_2$ |

$$(iv) \quad \alpha \underline{\underline{A}} \otimes \beta \underline{\underline{B}} = \alpha \beta \underline{\underline{A}} \otimes \underline{\underline{B}}$$

$$(v) \quad \underline{\underline{A}}_1 \underline{\underline{A}}_2 \otimes \underline{\underline{B}}_1 \underline{\underline{B}}_2 = (\underline{\underline{A}}_1 \otimes \underline{\underline{B}}_1) (\underline{\underline{A}}_2 \otimes \underline{\underline{B}}_2)$$

$$(vi) \quad (\underline{\underline{A}} \otimes \underline{\underline{B}})^{-1} = \underline{\underline{A}}^{-1} \otimes \underline{\underline{B}}^{-1}, \text{ hvis de inverse eksisterer}$$

$$(vii) \quad (\underline{\underline{A}} \otimes \underline{\underline{B}})^- = \underline{\underline{A}}^- \otimes \underline{\underline{B}}^-$$

$$(viii) \quad (\underline{\underline{A}} \otimes \underline{\underline{B}})' = \underline{\underline{A}}' \otimes \underline{\underline{B}}'$$

(ix) Lad  $\underline{\underline{A}}$ , symmetrisk og  $p \times p$ , have egenværdier  $\alpha_1, \dots, \alpha_p$  og egenvektorer  $\underline{x}_i$ , og lad  $\underline{\underline{B}}$ , symmetrisk og  $q \times q$ , have egenværdier  $\beta_1, \dots, \beta_q$  og egenvektorer  $\underline{y}_1, \dots, \underline{y}_q$ . Da vil  $\underline{\underline{A}} \otimes \underline{\underline{B}}$  have egenværdierne  $\alpha_i \beta_j$ ,  $i = 1, \dots, p$ ,  $j = 1, \dots, q$ , med tilsvarende egenvektorer

$$(\underline{x}_i \otimes \underline{y}_j \sim) \begin{bmatrix} x_{1i} y_j \\ \vdots \\ x_{pi} y_j \end{bmatrix}$$

$$(x) \quad \det (\underline{\underline{A}} \otimes \underline{\underline{B}}) = (\det \underline{\underline{A}})^q (\det \underline{\underline{B}})^p$$

## 1.6 Indre produkter og normer

For  $n$ -dimensionale vektorer har man som bekendt defineret det indre produkt eller skalarproduktet eller prikproduktet af  $\underline{x}$  og  $\underline{y}$  ved

$$\underline{x} \cdot \underline{y} = \underline{x}' \underline{y} = (x_1 \dots x_n) \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^n x_i y_i,$$

og man bemærker, at  $\underline{x}$  og  $\underline{y}$  er ortogonale, hvis og kun hvis

$$\underline{x} \cdot \underline{y} = \underline{x}'\underline{y} = 0.$$

Den tilhørende norm er

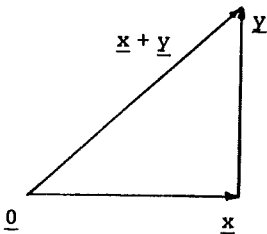
$$\|\underline{x}\| = (\underline{x} \cdot \underline{x})^{1/2} = (\underline{x}'\underline{x})^{1/2} = \sqrt{x_1^2 + \dots + x_n^2}$$

Vi bemærker, at  $\|\underline{x} - \underline{y}\|$  angiver den euklidiske afstand mellem punkterne  $\underline{x}$  og  $\underline{y}$ .

For ortogonale vektorer  $\underline{x}$  og  $\underline{y}$  (d.v.s.  $\underline{x} \perp \underline{y}$ ) har vi den pythagoræiske læresætning

$$\|\underline{x} + \underline{y}\|^2 = \|\underline{x}\|^2 + \|\underline{y}\|^2;$$

thi



$$\begin{aligned} \|\underline{x} + \underline{y}\|^2 &= (\underline{x} + \underline{y})'(\underline{x} + \underline{y}) = \\ &= \underline{x}'\underline{x} + \underline{x}'\underline{y} + \underline{y}'\underline{x} + \underline{y}'\underline{y} \\ &= \underline{x}'\underline{x} + \underline{y}'\underline{y} \\ &= \|\underline{x}\|^2 + \|\underline{y}\|^2. \end{aligned}$$

Endvidere bemærker vi, at projektionen  $p(\underline{x})$  (den ortogonale) af en vektor  $\underline{x}$  ned på et underrum  $U$  kan bestemmes ved hjælp af normen, idet vi har, at  $p(\underline{x})$  er givet ved

$$\|\underline{x} - p(\underline{x})\| = \min_{\underline{z} \in U} \|\underline{x} - \underline{z}\|$$

Bevis

Ifølge den pythagoræiske læresætning har vi, at

$$\begin{aligned} & \| \underline{x} - p(\underline{x}) \|^2 + \| \underline{z} - p(\underline{x}) \|^2 \\ &= \| \underline{x} - \underline{z} \|^2, \end{aligned}$$

d.v.s. minimumsværdien af  $\| \underline{x} - \underline{z} \|^2$  og dermed af  $\| \underline{x} - \underline{z} \|^2$  indtræffer for  $\underline{z} = p(\underline{x})$ .

□

Det er nu meget let at vise, at gyldigheden af ovenstående resultater blot beror på 4 grundlæggende egenskaber ved det indre produkt, nemlig, idet vi nu betegner det indre produkt af  $\underline{x}$  og  $\underline{y}$  med  $(\underline{x}|\underline{y})$ .

$$\text{IP 1: } (\underline{x}|\underline{y}) = (\underline{y}|\underline{x})$$

$$\text{IP 2: } (\underline{x} + \underline{y}|\underline{z}) = (\underline{x}|\underline{z}) + (\underline{y}|\underline{z})$$

$$\text{IP 3: } (k\underline{x}|\underline{y}) = k(\underline{x}|\underline{y})$$

$$\text{IP 4: } \underline{x} \neq \underline{0} \Rightarrow (\underline{x}|\underline{x}) > 0 .$$

For en vilkårlig bilinearform  $(\cdot|\cdot)$ , der tilfredsstiller ovenstående, kan man definere et ortogonalitetsbegreb ved

$$\underline{x} \perp \underline{y} \stackrel{\text{d}}{\iff} (\underline{x}|\underline{y}) = 0 .$$

For en vilkårlig positiv definit symmetrisk matrix  $\underline{A}$  kan vi definere et indre produkt ved

$$(\underline{x}|\underline{y})_{\underline{A}} = \underline{x}' \underline{A} \underline{y} .$$

Det er helt trivielt at eftervise, at IP 1-4 er opfyldte. For dette indre produkt og den deraf fødte norm givet ved

$$\| \underline{x} \|_{\underline{A}} = \sqrt{(\underline{x} | \underline{x})_{\underline{A}}} = \sqrt{\underline{x}' \underline{A} \underline{x}} ,$$

vil vi, hvor misforståelse er udelukket, anvende betegnelserne  $(\underline{x} | \underline{y})$  og  $\| \underline{x} \|$ .

Man bemærker, at mængden af punkter med konstant  $\underline{A}$ -norm lig 1 er mængden

$$\{ \underline{x} | \| \underline{x} \|^2 = 1 \} = \{ \underline{x} | \underline{x}' \underline{A} \underline{x} = 1 \} ,$$

d.v.s. punkterne på en ellipsoide.

Omvendt svarer der til enhver ikke udartet ellipsoide en symmetrisk positivt definit matrix  $\underline{A}$ , så

$$E = \{ \underline{x} | \underline{x}' \underline{A} \underline{x} = 1 \} = \{ \underline{x} | \| \underline{x} \|^2_{\underline{A}} = 1 \} .$$

Der er på denne måde tilvejebragt en forbindelse mellem mængden af mulige indre produkter og mængden af ellipsoider.

To vektorer  $\underline{x}$  og  $\underline{y}$  er ortogonale (med hensyn til  $\underline{A}$ ), hvis

$$\underline{x}' \underline{A} \underline{y} = 0 ,$$

d.v.s. hvis  $\underline{x}$  og  $\underline{y}$  er konjugerede retninger i ellipsoiden svarende til  $\underline{A}$ .

Man kan også indføre et vinkelbegreb v.h.a. definitionen

$$\cos(\angle \underline{a}, \underline{b}) = \frac{(\underline{a} | \underline{b})}{\| \underline{a} \| \| \underline{b} \|} .$$

Vi anfører nu et lemma, som skal anvendes i forbindelse med sætninger om uafhængighed af projektioner af normalt fordelte stokastiske variable.

Lemma Lad  $R^n$  være spaltet i en direkte sum

$$R^n = U_1 \oplus \dots \oplus U_k$$

af  $n_i$  dimensionale underrum, der er ortogonale m.h.t. den positivt definitte matrix  $\underline{\Sigma}^{-1}$ , d.v.s.

$$\underline{x} \perp \underline{y} \Leftrightarrow \underline{x}' \underline{\Sigma}^{-1} \underline{y} = 0 \quad .$$

Lad for  $i = 1, \dots, k$  projektionen  $p_i$  på  $U_i$  være givet ved matrixen  $\underline{C}_i$ . Da vil

$$\underline{C}_i \underline{\Sigma} \underline{C}_j' = 0$$

for alle  $i \neq j$ . Endvidere gælder

$$\underline{\Sigma}^{-1} \underline{C}_i = \underline{C}_i' \underline{\Sigma}^{-1} = \underline{C}_i' \underline{\Sigma} \underline{C}_i \quad .$$

Bevis Da  $p_i \circ p_i = p_i$ , gælder

$$\underline{C}_i \underline{C}_i = \underline{C}_i \quad ,$$

og da

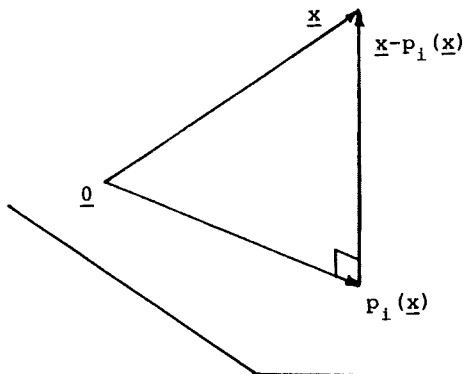
$$p_i(\underline{x}) \perp \underline{x} - p_i(\underline{x}) \quad ,$$

(se skitsen) vil

$$p_i(\underline{x})' \underline{\Sigma}^{-1} (\underline{x} - p_i(\underline{x})) = 0 \quad ,$$

d.v.s.

$$\underline{x} \underline{C}_i' \underline{\Sigma}^{-1} [\underline{x} - \underline{C}_i \underline{x}] = 0 \quad .$$



Dette gælder for alle  $\underline{x}$ , og derfor er

$$\underline{C}'_i \underline{\Sigma}^{-1} (\underline{I} - \underline{C}_i) = \underline{0} \quad ,$$

eller

$$\underline{C}'_i \underline{\Sigma}^{-1} = \underline{C}'_i \underline{\Sigma}^{-1} \underline{C}_i \quad .$$

Nu er højresiden klart symmetrisk, hvorfor vi får

$$\underline{C}'_i \underline{\Sigma}^{-1} = \underline{\Sigma}^{-1} \underline{C}_i \quad .$$

Ved at multiplicere foran og bagved med  $\underline{\Sigma}$  fås

$$\underline{\Sigma} \underline{C}'_i = \underline{C}_i \underline{\Sigma} \quad ,$$

hvorfor

$$\underline{C}_i \underline{\Sigma} \underline{C}'_i = \underline{C}_i \underline{C}_i \underline{\Sigma} = \underline{C}_i \underline{\Sigma} \quad .$$

Dette giver

$$\underline{C}_i \underline{\Sigma} \underline{C}'_j = \underline{C}_i \underline{\Sigma} \underline{C}'_i \underline{C}'_j = \underline{C}_i \underline{\Sigma} \underline{0} = \underline{0} \quad .$$

Det næstsidste lighedstegn følger af, at summen er direkte, hvorfor det for ethvert  $\underline{x}$  gælder

$$p_j(p_i(\underline{x})) = \underline{0} \quad ,$$

d.v.s.

$$\underline{C}_j \underline{C}'_i \underline{x} = \underline{0} \quad .$$

Da  $\underline{x}$  som nævnt er vilkårlig, medfører dette

$$\underline{C}_j \underline{C}'_i = \underline{0} \quad ,$$

1.70

eller

$$\underline{C} \begin{pmatrix} C \\ \vdots \\ C \\ \vdots \\ C \end{pmatrix} = \underline{0} \quad .$$

□

Referencer til kapitel 1

Bourbaki, N.: Algèbre. Chapitre 2: Algèbre Lineaire. Hermann, Paris 1967.

Dempster, A.P.: Elements of Continuous Multivariate Analysis. Addison-Wesley, Reading 1969.

Johnson, R.M.: On a theorem stated by Eckart and Young. Psychometrika, vol. 28 (1963), pp. 259-263.

Rao, G.R.: Linear Statistical Inference and Its Applications. John Wiley, New York 1965.

Rao, C.R., & Mitra, S.K.: Generalized Inverse of Matrices and Its Applications. John Wiley, New York 1971.

Wilkinson, J.H.: Error analysis of direct methods of matrix inversion. Journal of the Association of Computing Machinery, vol. 8 (1961), pp. 281-330.



## KAPITEL 2

Flerdimensionale variable

I dette kapitel gives indledningsvis en supplerung af de resultater om flerdimensionale stokastiske variable, der er anført i kapitel 0 i bind 1. Dernæst omtales specielt den flerdimensionale, normale fordeling og fordelinger afledt af denne. Endelig gives en kort beskrivelse af de særlige forhold, som estimation og testning af flere parametre simultant giver anledning til.

2.1 Momenter af flerdimensionale stokastiske variable

Vi indleder med

2.1.1 Middelværdi

Lad der være givet en stokastisk matrix, i.e. en matrix, hvis enkelte elementer er stokastiske variable:

$$\underline{X} = \begin{bmatrix} X_{11} & \dots & X_{1n} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ X_{k1} & \dots & X_{kn} \end{bmatrix} .$$

Vi definerer da middelværdien eller forventningsværdien af  $\underline{X}$  ved

$$E(\underline{X}) = \begin{bmatrix} E(X_{11}) \dots E(X_{1n}) \\ \vdots \\ E(X_{k1}) \dots E(X_{kn}) \end{bmatrix} = \begin{bmatrix} \mu_{11} \dots \mu_{1n} \\ \vdots \\ \mu_{k1} \dots \mu_{kn} \end{bmatrix} = \underline{\mu} .$$

Sætning 2.1 Lad  $\underline{A}$  være en  $k \times n$  matrix af konstanter. Da er

$$E(\underline{A} + \underline{X}) = \underline{A} + E(\underline{X}) .$$

Denne sætning er lige som den følgende en triviell følge af definitionen.

Sætning 2.2 Lad  $\underline{A}$  og  $\underline{B}$  være konstante matricer, således at  $\underline{AX}$  og  $\underline{XB}$  eksisterer. Da er

$$E(\underline{AX}) = \underline{A}E(\underline{X})$$

$$E(\underline{XB}) = E(\underline{X})\underline{B} .$$

Endelig har vi

Sætning 2.3 Lad  $\underline{X}$  og  $\underline{Y}$  være stokastiske matricer af samme orden. Da gælder

$$E(\underline{X} + \underline{Y}) = E(\underline{X}) + E(\underline{Y}) .$$

Bemærkning Vi har intetsteds nævnt, at vi naturligvis forudsætter, at de involverede forventningsværdier eksisterer. Dette forudsættes her og overalt i det følgende, hvor disse begreber måtte blive nævnt.

### 2.1.2 Dispersionsmatricen

Generaliseringen af variansen af en stokastisk variabel er begrebet dispersionsmatricen for en stokastisk vektor  $\underline{X}$ . Den defineres ved

$$D(\underline{X}) = \underline{\Sigma} = E\{(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})'\},$$

hvor

$$\underline{\mu} = E(\underline{X}).$$

Det skal bemærkes, at  $D(\underline{X})$  også ofte omtales som kovariansmatricen og da betegnes  $\text{Cov}(\underline{X})$ . Dette er dog en lidt uheldig vending, idet den vil kunne give anledning til forveksling med kovariansen mellem to (flerdimensionale) stokastiske variable. En anden hyppigt anvendt betegnelse er  $V(\underline{X})$ . Vi bemærker i øvrigt, at

$$(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})' = \begin{bmatrix} X_1 - \mu_1 \\ \vdots \\ X_n - \mu_n \end{bmatrix} (X_1 - \mu_1, \dots, X_n - \mu_n) =$$

$$\begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) & \dots & (X_1 - \mu_1)(X_n - \mu_n) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 & \dots & (X_2 - \mu_2)(X_n - \mu_n) \\ \vdots & \vdots & \ddots & \vdots \\ (X_n - \mu_n)(X_1 - \mu_1) & (X_n - \mu_n)(X_2 - \mu_2) & \dots & (X_n - \mu_n)^2 \end{bmatrix}$$

d.v.s. dispersionsmatricens  $(i, j)$ 'te element er  $\text{Cov}(X_i, X_j)$ ,  
eller

$$\underline{\Sigma} = D(\underline{X}) = \begin{bmatrix} V(X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & V(X_2) & \dots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \dots & V(X_n) \end{bmatrix}$$

Vi vil ofte anvende skrivemåden

$$\underline{\Sigma} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22}^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn}^2 \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{bmatrix},$$

idet vi om varianserne såvel anvender betegnelserne  $\sigma_1^2$  som  $\sigma_{11}$ . Vi har, at  $\underline{\Sigma}$  er symmetrisk. Mere interessant er

**Sætning 2.4** Dispersionsmatricen  $\underline{\Sigma}$  for en stokastisk vektor (d.v.s. flerdimensional stokastisk variabel) er positivt semi-definit.

Bevis For en vilkårlig vektor  $\underline{y}$  haves

$$\begin{aligned} \underline{y}' \underline{\Sigma} \underline{y} &= \underline{y}' E\{(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})'\} \underline{y} \\ &= E\{\underline{y}'(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})' \underline{y}\} \\ &= E\{[(\underline{X} - \underline{\mu})' \underline{y}]' [(\underline{X} - \underline{\mu})' \underline{y}]\} \\ &\geq 0, \end{aligned}$$

da størrelsen i de krøllede parenteser er  $\geq 0$ .

□

Der gælder sætninger ganske analoge til de fra de endimensionale stokastiske variable kendte.

**Sætning 2.5** Lad  $\underline{X}$  og  $\underline{Y}$  være uafhængige. Da er

$$D(\underline{X} + \underline{Y}) = D(\underline{X}) + D(\underline{Y}) .$$

Lad  $\underline{b}$  være en konstant. Da har vi

$$D(\underline{b} + \underline{X}) = D(\underline{X}) .$$

Hvis  $\underline{A}$  er en konstant matrix, således at  $\underline{AX}$  eksisterer, da gælder

$$D(\underline{AX}) = \underline{A}D(\underline{X})\underline{A}' .$$

**Bevis** Den første relation følger af

$$\begin{aligned} \text{Cov}(X_i + Y_i, X_j + Y_j) &= \text{Cov}(X_i, X_j) + \text{Cov}(X_i, Y_j) \\ &+ \text{Cov}(Y_i, X_j) + \text{Cov}(Y_i, Y_j) \\ &= \text{Cov}(X_i, X_j) + \text{Cov}(Y_i, Y_j) , \end{aligned}$$

idet  $\text{Cov}(Y_i, X_j) = 0$ , da  $X_j$  og  $Y_i$  er uafhængige.

Den næste relation er triviell. Den sidste følger af

$$\begin{aligned} D(\underline{AX}) &= E\{(\underline{AX} - \underline{A}\underline{\mu})(\underline{AX} - \underline{A}\underline{\mu})'\} \\ &= E\{\underline{A}[\underline{X} - \underline{\mu}][\underline{X} - \underline{\mu}]'\underline{A}'\} \\ &= \underline{A}E\{[\underline{X} - \underline{\mu}][\underline{X} - \underline{\mu}]'\}\underline{A}' \\ &= \underline{A}D(\underline{X})\underline{A}' \\ &= \underline{A}\underline{E}\underline{A}' \end{aligned}$$

□

Sætter vi

$$\underline{V} = \text{diag}\left(\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_n}\right) = \begin{bmatrix} \sigma_1^{-1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_n^{-1} \end{bmatrix}$$

og "skalærer" vi  $\underline{X}$  med  $\underline{V}$ , fås

$$D(\underline{V}\underline{X}) = \underline{V}\underline{E}\underline{V}' = \begin{bmatrix} 1 & \frac{\sigma_{12}}{\sigma_1 \sigma_2} & \dots & \frac{\sigma_{1n}}{\sigma_1 \sigma_n} \\ \frac{\sigma_{12}}{\sigma_1 \sigma_2} & 1 & \dots & \frac{\sigma_{2n}}{\sigma_2 \sigma_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1n}}{\sigma_1 \sigma_n} & \frac{\sigma_{2n}}{\sigma_2 \sigma_n} & \dots & 1 \end{bmatrix} .$$

Vi ser, at elementerne netop er korrelationskoefficienterne mellem  $X_j$  komponenter, hvorfor denne matrix også benævnes korrelationsmatrixen for  $\underline{X}$ , og vi skriver

$$R(\underline{X}) = \begin{bmatrix} 1 & \dots & \rho_{1n} \\ \vdots & \ddots & \vdots \\ \rho_{1n} & \dots & 1 \end{bmatrix} ,$$

hvor altså

$$\rho_{ij} = \text{Cor}(X_i, X_j) = \frac{\text{Cov}(X_i, X_j)}{\sqrt{V(X_i)V(X_j)}} .$$

### 2.1.3. Kovarians

Lad der være givet to stokastiske variable

$$\underline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix} \text{ og } \underline{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_q \end{bmatrix}$$

med middelværdier  $\underline{\mu}$  og  $\underline{\nu}$ . Vi definerer da kovariansen mellem  $\underline{X}$  og  $\underline{Y}$  ved

$$C(\underline{X}, \underline{Y}) = E[(\underline{X} - \underline{\mu})(\underline{Y} - \underline{\nu})'] = \begin{bmatrix} \text{Cov}(X_1, Y_1) & \dots & \text{Cov}(X_1, Y_q) \\ \vdots & & \vdots \\ \text{Cov}(X_p, Y_1) & \dots & \text{Cov}(X_p, Y_q) \end{bmatrix} .$$

Der gælder da

$$C(\underline{X}, \underline{X}) = D(\underline{X})$$

og

$$C(\underline{X}, \underline{Y}) = [C(\underline{Y}, \underline{X})]' .$$

Mere dyb er

Sætning 2.6 Lad  $\underline{X}$  og  $\underline{Y}$  være som ovenfor, og lad  $\underline{A}$  og  $\underline{B}$  være henholdsvis  $n \times p$  og  $m \times q$  matricer af konstanter. Da er

$$C(\underline{AX}, \underline{BY}) = \underline{A}C(\underline{X}, \underline{Y})\underline{B}' .$$

Er  $\underline{U}$  en  $p$ -dimensional og  $\underline{V}$  en  $q$ -dimensional stokastisk variabel gælder endvidere

$$C(\underline{X} + \underline{U}, \underline{Y}) = C(\underline{X}, \underline{Y}) + C(\underline{U}, \underline{Y})$$

$$C(\underline{X}, \underline{Y} + \underline{V}) = C(\underline{X}, \underline{Y}) + C(\underline{X}, \underline{V}) .$$

Endelig gælder,

$$D(\underline{X} + \underline{U}) = D(\underline{X}) + D(\underline{U}) + C(\underline{X}, \underline{U}) + C(\underline{U}, \underline{X}) .$$

Bevis Vi har ifølge definitionen

$$\begin{aligned} C(\underline{A}\underline{X}, \underline{B}\underline{Y}) &= E[(\underline{A}\underline{X} - \underline{A}\underline{\mu})(\underline{B}\underline{Y} - \underline{B}\underline{\nu})'] \\ &= E[\underline{A}(\underline{X} - \underline{\mu})(\underline{Y} - \underline{\nu})'\underline{B}'] \\ &= \underline{A}E[(\underline{X} - \underline{\mu})(\underline{Y} - \underline{\nu})']\underline{B}' \\ &= \underline{A}C(\underline{X}, \underline{Y})\underline{B}' . \end{aligned}$$

Hermed er den første påstand godtgjort. Tilsvarende fås - idet vi sætter  $E(\underline{U}) = \underline{\delta}$  -

$$\begin{aligned} C(\underline{X} + \underline{U}, \underline{Y}) &= E[(\underline{X} + \underline{U} - \underline{\mu} - \underline{\delta})(\underline{Y} - \underline{\nu})'] \\ &= E[(\underline{X} - \underline{\mu})(\underline{Y} - \underline{\nu})' + (\underline{U} - \underline{\delta})(\underline{Y} - \underline{\nu})'] \\ &= E[(\underline{X} - \underline{\mu})(\underline{Y} - \underline{\nu})'] + E[(\underline{U} - \underline{\delta})(\underline{Y} - \underline{\nu})'] \\ &= C(\underline{X}, \underline{Y}) + C(\underline{U}, \underline{Y}) , \end{aligned}$$

og den tilsvarende relation med  $\underline{Y} + \underline{V}$  vises helt analogt.

Endelig har vi

$$\begin{aligned} D(\underline{X} + \underline{U}) &= C(\underline{X} + \underline{U}, \underline{X} + \underline{U}) \\ &= C(\underline{X}, \underline{X}) + C(\underline{X}, \underline{U}) + C(\underline{U}, \underline{X}) + C(\underline{U}, \underline{U}) . \end{aligned}$$

□

Hvis  $C(\underline{X}, \underline{Y}) = \underline{0}$  siges  $\underline{X}$  og  $\underline{Y}$  at være ukorrelerede. Det svarer



til at alle komponenter i  $\underline{X}$  er ukorrelerede med samtlige komponenter i  $\underline{Y}$ .

Ved behandlingen af den flerdimensionale generelle lineære model får vi brug for følgende

Sætning 2.7 Lad  $X_1, \dots, X_n$  være uafhængige,  $p$ -dimensionale stokastiske variable med samme dispersionsmatrix  $\underline{\Sigma} = (\sigma_{ij})$ . Vi sætter

$$\underline{X} = \begin{bmatrix} X_1' \\ \vdots \\ X_n' \end{bmatrix} = \begin{bmatrix} X_{11} \dots X_{p1} \\ \vdots \\ X_{1n} \dots X_{pn} \end{bmatrix}$$

(Bemærk, at variabel-index er det første index og gentagelsesnr.-index er det andet). Defineres

$$vc(\underline{X}) = \begin{bmatrix} X_{11} \\ \vdots \\ X_{1n} \\ \vdots \\ X_{p1} \\ \vdots \\ X_{pn} \end{bmatrix}$$

d.v.s. som vektoren bestående af søjlerne i  $\underline{X}$  (vc = vector of columns) fås

$$D(vc(\underline{X})) = \underline{\Sigma} \otimes \underline{I}_n,$$

hvor  $\underline{I}_n$  er enhedsmatricen af  $n$ 'te orden.

Bevis Følger trivielt af definitionen på tensorprodukt og af definitionen på dispersionsmatricen.

□

## 2.2 Den flerdimensionale normalfordeling

Den flerdimensionale normalfordeling spiller samme vigtige rolle i teorien for flerdimensionale variable, som normalfordelingen gør i det endimensionale tilfælde. Vi indleder med

### 2.2.1 Definition og simple egenskaber

Lad  $X_1, \dots, X_p$  være indbyrdes uafhængige,  $N(0,1)$  - fordelte variable. Vi siger da, at

$$\underline{X} = \begin{bmatrix} X_1 \\ \cdot \\ \cdot \\ \cdot \\ X_p \end{bmatrix},$$

er standardiseret (normeret) p-dimensionalt normalt fordelt, og vi skriver

$$\underline{X} \in \underline{N}(\underline{0}, \underline{I}) = N_p(\underline{0}, \underline{I}),$$

hvor den sidste betegnelse anvendes, hvis der kan opstå tvivl om dimensionen.

Vi bemærker, at

$$E(\underline{X}) = \underline{0}, D(\underline{X}) = \underline{I}.$$

Vi definerer den flerdimensionale normalfordeling med generelle parametre i

Definition 2.1 Vi siger, at den  $p$ -dimensionale stokastiske variabel  $\underline{X}$  er normalt fordelt med parametre  $\underline{\mu}$  og  $\underline{\Sigma}$ , hvis  $\underline{X}$  har samme fordeling som

$$\underline{\mu} + \underline{A}\underline{U},$$

hvor  $\underline{A}$  tilfredsstiller

$$\underline{A}\underline{A}' = \underline{\Sigma},$$

og hvor  $\underline{U}$  er standardiseret  $p$ -dimensionalt normalt fordelt. Vi skriver

$$\underline{X} \in N(\underline{\mu}, \underline{\Sigma}) = N_p(\underline{\mu}, \underline{\Sigma}),$$

hvor den sidste betegnelse igen anvendes, hvis der kan være tvivl om dimensionen.

Bemærkning Definitionen er kun lovlig, hvis man viser, at  $\underline{A}\underline{A}' = \underline{B}\underline{B}'$  medfører

$$E(\underline{\mu} + \underline{A}\underline{U}) = E(\underline{\mu} + \underline{B}\underline{V}),$$

hvor  $\underline{U}$  og  $\underline{V}$  er standardiseret normalt fordelte og ikke nødvendigvis af samme dimension. Relationen er gyldig, men dette vil vi dog ikke komme nærmere ind på her. Af sætning 1.10 fås umiddelbart, at der for en vilkårlig positivt semidefinit matrix  $\underline{\Sigma}$  eksisterer en matrix  $\underline{A}$  med  $\underline{A}\underline{A}' = \underline{\Sigma}$ , således at betegnelsen  $N(\underline{\mu}, \underline{\Sigma})$  har mening for en vilkårlig positivt semidefinit  $p \times p$  matrix  $\underline{\Sigma}$  og en vilkårlig  $p$ -dimensional vektor  $\underline{\mu}$ .

Man bemærker trivielt, at

$$\underline{X} \in N(\underline{\mu}, \underline{\Sigma}) \rightarrow \text{i) } E(\underline{X}) = \underline{\mu} \wedge \text{ii) } D(\underline{X}) = \underline{\Sigma}$$

d.v.s. fordelingen er parametriseret ved dens middelværdi og

dispersionsmatrix.

Hvis  $\underline{\Sigma}$  har fuld rang har fordelingen en tæthed, som vi anfører i

Sætning 2.8 Lad  $\underline{X} \in N_p(\underline{\mu}, \underline{\Sigma})$ , og lad  $\text{rg}(\underline{\Sigma}) = p$ . Da har  $\underline{X}$  tætheden

$$\begin{aligned} f(\underline{x}) &= \frac{1}{\sqrt{2\pi}^p} \frac{1}{\sqrt{\det \underline{\Sigma}}} \exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu})' \underline{\Sigma}^{-1}(\underline{x} - \underline{\mu})\right] \\ &= \frac{1}{\sqrt{2\pi}^p} \frac{1}{\sqrt{\det \underline{\Sigma}}} \exp\left[-\frac{1}{2} \|\underline{x} - \underline{\mu}\|^2\right], \end{aligned}$$

hvor den anførte norm er den ved  $\underline{\Sigma}^{-1}$  bestemte, jvf. p. 1.67.

Bevis Lad  $\underline{U} \in N_p(0, \underline{I})$ . Da har  $\underline{U}$  tætheden

$$\begin{aligned} h(\underline{u}) &= \prod_{i=1}^p \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}u_i^2) = \frac{1}{\sqrt{2\pi}^p} \exp(-\frac{1}{2} \sum_{i=1}^p u_i^2) \\ &= \frac{1}{\sqrt{2\pi}^p} \exp(-\frac{1}{2}\underline{u}'\underline{u}). \end{aligned}$$

Vi betragter dernæst transformationen fra  $R^p \rightarrow R^p$  givet ved

$$\underline{u} \rightarrow \underline{x} = \underline{\mu} + \underline{A}\underline{u}$$

hvor  $\underline{A}\underline{A}' = \underline{\Sigma}$ . Af sætning 1.14 følger, at  $\underline{A}$  er regulær. Vi får

$$\underline{u} = \underline{A}^{-1}(\underline{x} - \underline{\mu}),$$

hvorfor

$$\underline{u}'\underline{u} = (\underline{x} - \underline{\mu})' \underline{A}^{-1}' \underline{A}^{-1} (\underline{x} - \underline{\mu})$$

$$= (\underline{x} - \underline{\mu})' \underline{\Sigma}^{-1} (\underline{x} - \underline{\mu}).$$

Da endvidere

$$\det(\underline{\Sigma}) = \det(\underline{A}\underline{A}') = \det(\underline{A})^2,$$

d.v.s.

$$\det(\underline{A}^{-1}) = \frac{1}{\sqrt{\det \underline{\Sigma}}}$$

følger resultatet af sætning 0.8 i bind I.

□

Vi bemærker at den inverse dispersionsmatrix  $\underline{\Sigma}^{-1}$  ofte kaldes præcisionen for den normale fordeling.

Hvis  $\underline{\Sigma}$  ikke er regulær, er fordelingen udartet og har ingen tæthed. Vi indfører da begrebet den affine støtte i

Definition 2.2 Lad  $\underline{X} \in N_p(\underline{\mu}, \underline{\Sigma})$ . Ved den (affine) støtte for  $\underline{X}$  forstås det mindste (side-) underrum af  $R^p$ , hvor  $\underline{X}$  er defineret med sandsynlighed 1.

Bemærkning Hvis man indskrænker betragtningerne til den affine støtte, er  $\underline{X}$  regulært fordelt og har da en tæthed som angivet i sætning 2.8.

Vi har forskellige muligheder for at bestemme støtten for en p-dimensional normalfordeling. Der er først

Sætning 2.9 Lad  $\underline{X} \in N_p(\underline{\mu}, \underline{\Sigma})$ , og lad  $A$  være en  $p \times m$  matrix, så  $\underline{A}\underline{A}' = \underline{\Sigma}$ . Vi sætter  $V$  lig  $\underline{A}$ 's billedrum, i.e.

2.14

$$V = \{ \underline{v} \in \mathbb{R}^p \mid \exists \underline{u} \in \mathbb{R}^m : \underline{v} = \underline{A}\underline{u} \} .$$

Da er den (affine) støtte for  $\underline{X}$  sideunderrummet

$$\underline{\mu} + V = \{ \underline{\mu} + \underline{v} \mid \underline{v} \in V \} .$$

Bevis Forbigås.

Endvidere gælder

Sætning 2.10 Lad  $\underline{X}$  være som i foregående sætning. Da er under rummet  $V$  lig den direkte sum af egenrummene svarende til de fra 0 forskellige egenverdier i  $\underline{\Sigma}$ .

Bevis Forbigås.

Endelig har vi

Sætning 2.11 Lad  $\underline{X}$  være som i de foregående sætninger. Da er underrummet  $V$  lig det ortogonale komplement til nulrummet for  $\underline{\Sigma}$ , i.e.

$$V = \{ \underline{v} \mid \underline{\Sigma}\underline{v} = \underline{0} \}^\perp$$

Bevis Forbigås.

Vi illustrerer de 3 sætninger i

Eksempel 2.1 Vi betragter

$$\underline{X} \in N\left(\begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 2 \\ 2 & 5 & 3 \\ 2 & 3 & 5 \end{bmatrix}\right) = N(\underline{\mu}, \underline{\Sigma}).$$

Da

$$\det\left(\begin{bmatrix} 1 & 2 & 2 \\ 2 & 5 & 3 \\ 2 & 3 & 5 \end{bmatrix}\right) = 0,$$

er  $\underline{X}$  singulært fordelt, og vi vil bestemme dens affine støtte.

Vi søger først en matrix  $\underline{A}$ , så  $\underline{A}\underline{A}' = \underline{\Sigma}$ . Til den ende bestemmes først  $\underline{\Sigma}$ 's egenverdier og -vektorer (normerede). De er

$$\lambda_1 = 9 \wedge \underline{p}_1 = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix},$$

$$\lambda_2 = 2 \wedge \underline{p}_2 = \begin{bmatrix} 0 \\ \sqrt{2}/2 \\ -\sqrt{2}/2 \end{bmatrix},$$

$$\lambda_3 = 0 \wedge \underline{p}_3 = \begin{bmatrix} 2\sqrt{2}/3 \\ -\sqrt{2}/6 \\ -\sqrt{2}/6 \end{bmatrix}.$$

Følgelig er

$$\underline{\Sigma} = \begin{bmatrix} \frac{1}{3} & 0 & \frac{2\sqrt{2}}{3} \\ \frac{2}{3} & \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{6} \\ \frac{2}{3} & -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{6} \end{bmatrix} \begin{bmatrix} 9 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ 0 & \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{2\sqrt{2}}{3} & -\frac{\sqrt{2}}{6} & -\frac{\sqrt{2}}{6} \end{bmatrix}.$$

Heraf fås, at vi som  $\underline{A}$ -matrix kan vælge

$$\underline{A} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 2 & -1 & 0 \end{bmatrix}.$$

Opfattes  $\underline{A}$  som matrix for en lineær afbildning  $\mathbb{R}^3 \rightarrow \mathbb{R}^3$  fås, at billedrummet er

$$\begin{aligned} V &= \{ \underline{A}\underline{u} \mid \underline{u} \in \mathbb{R}^3 \} \\ &= \{ u_1 \underline{p}_1 + u_2 \underline{p}_2 \mid u_1 \in \mathbb{R} \wedge u_2 \in \mathbb{R} \}. \end{aligned}$$

Det ses umiddelbart, at dette også er den direkte sum af egenrummene svarende til de fra 0 forskellige egenværdier.

Nulrummet for  $\underline{A}$  er givet ved

$$\underline{A}\underline{u} = \underline{0} \Leftrightarrow \underline{u} = t \cdot \underline{p}_3.$$

Heraf fås så igen den samme beskrivelse af  $V$ .

Den affine støtte for  $\underline{V}$  er altså sideunderrummet

$$\underline{\mu} + V = \left\{ \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} + u_1 \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix} + u_2 \begin{bmatrix} 0 \\ \sqrt{2}/2 \\ -\sqrt{2}/2 \end{bmatrix} \mid u_1, u_2 \in \mathbb{R} \right\}.$$

□

Bemærkning Af eksemplet fremgår beviserne for sætningerne 2.9-2.11 næsten fuldstændigt.

Vi formulerer nu en triviell, men nyttig sætning.

Sætning 2.12 Lad  $\underline{X} \in N(\underline{\mu}, \underline{V})$ . Da gælder

$$\underline{A}\underline{X} + \underline{b} \in N(\underline{A}\underline{\mu} + \underline{b}, \underline{A}\underline{V}\underline{A}'),$$



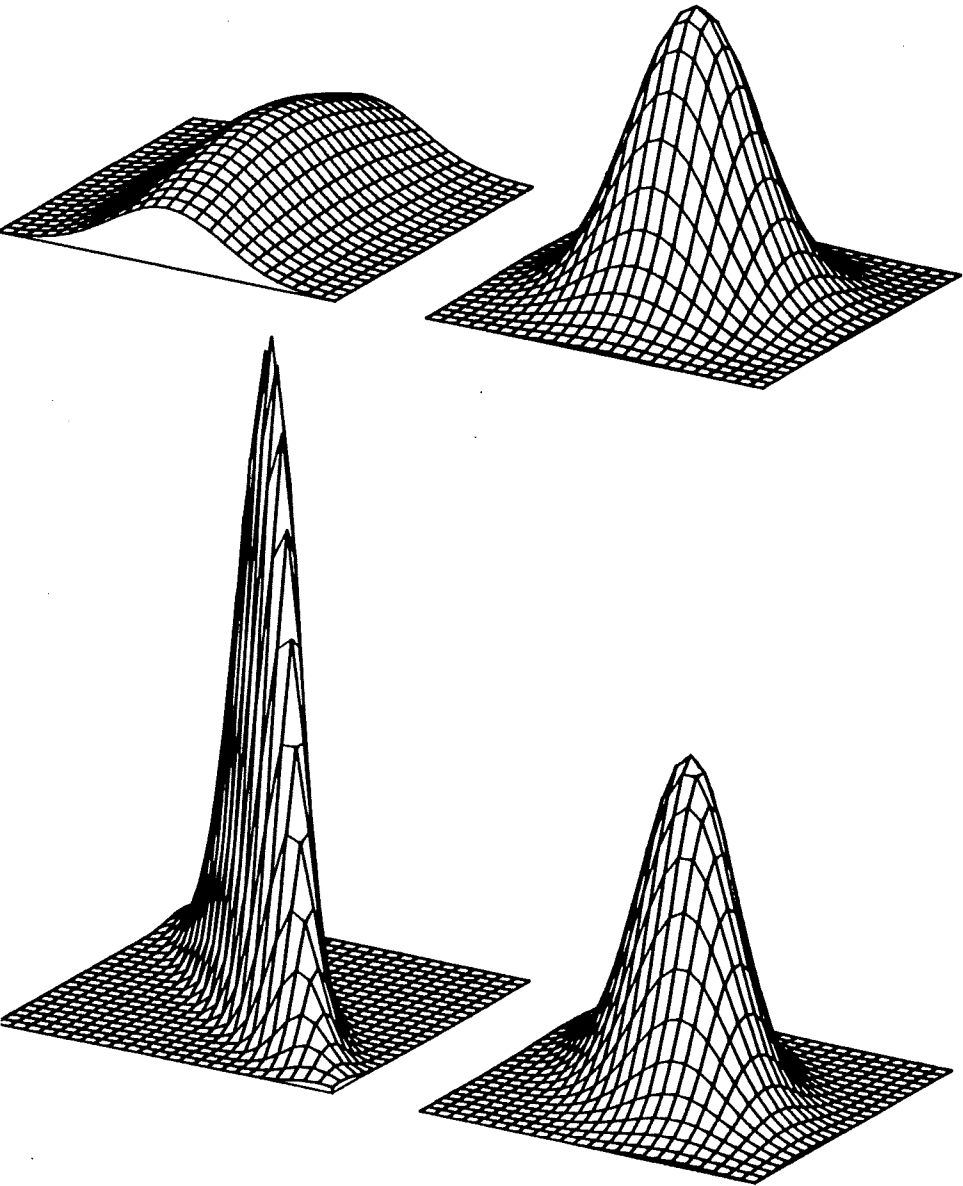


Fig. 2.1 Tæthedsfunktioner for todimensionale normalfordelinger med dispersionsmatricer  
 $\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$ ,  $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ ,  $\begin{pmatrix} 1 & -0.9 \\ -0.9 & 1 \end{pmatrix}$  og  $\begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}$ .

hvor det naturligvis forudsættes, at de anførte matrixprodukter m.v. eksisterer.

Bevis Trivial følge af definitionen.

### 2.2.2 Uafhængighed og konturellipsoider

I dette afsnit skal vi dels give betingelser for uafhængighed af normalt fordelte stokastiske variable, og dels godtgøre, at niveaukurverne for tæthedsfunktioner er ellipsoider. Vi har først

Sætning 2.13 Lad

$$\underline{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \in N \left( \begin{bmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \end{bmatrix}, \begin{bmatrix} \underline{\Sigma}_{11} & \underline{\Sigma}_{12} \\ \underline{\Sigma}_{21} & \underline{\Sigma}_{22} \end{bmatrix} \right).$$

Da gælder, at

$$\underline{X}_i \in N(\underline{\mu}_i, \underline{\Sigma}_{ii}),$$

og

$$\underline{X}_1, \underline{X}_2 \text{ stokastisk uafhængige} \Leftrightarrow \underline{\Sigma}_{12} = \underline{\Sigma}'_{21} = \underline{0},$$

hvor  $\underline{0}$  er en matrix bestående af lutter nuller.

Bevis Første udsagn fås direkte af den foregående sætning. Det andet fås ved at godtgøre, at betingelsen  $\underline{\Sigma}_{12} = \underline{0}$  sikrer, at fordelingen bliver en produktfordeling.

□

Af sætningen følger specielt, at komponenterne i en vektor  $\underline{X} \in N(\underline{\mu}, \underline{\Sigma})$  er stokastisk uafhængige, netop hvis  $\underline{\Sigma}$  er en diagonalmatrix. Vi skal nu vise, at uafhængigheden blot er et spørgsmål om at vælge et passende koordinatsystem.

Lad  $\underline{X} \in N(\underline{\mu}, \underline{\Sigma})$  og lad  $\underline{\Sigma}$  have de ortonormerede egenvektorer  $\underline{p}_1, \dots, \underline{p}_n$ . Vi betragter nu koordinatsystemet, der har begyndelsespunkt i  $\underline{\mu}$  og vektorerne  $\underline{p}_1, \dots, \underline{p}_n$  som basisvektorer. Koordinaterne i dette system benævnes  $\underline{y}$ .

Sættes

$$\underline{P} = (\underline{p}_1 \dots \underline{p}_n),$$

har vi følgende sammenhæng mellem de oprindelige koordinater  $\underline{x}$  og de nye koordinater  $\underline{y}$  for et vilkårligt punkt  $\in \mathbb{R}^n$ .

$$\underline{y} = \underline{P}'(\underline{x} - \underline{\mu}) \Leftrightarrow \underline{x} = \underline{P}\underline{y} + \underline{\mu},$$

jev. p. 1.14.

NB. Ovenstående relation er en relation mellem koordinater for en fast vektor betragtet i 2 koordinatsystemer.

Lader vi i overensstemmelse hermed  $\underline{y}$  være de nye koordinater for  $\underline{X}$  har vi

Sætning 2.14 Lad  $\underline{X} \in N(\underline{\mu}, \underline{\Sigma})$  og lad  $\underline{y}$  være som ovenfor. Da gælder

$$\underline{y} \in N(\underline{0}, \underline{\Delta}),$$

hvor  $\underline{\Delta}$  er en diagonalmatrix med  $\underline{\Sigma}$ 's egenverdier i diagonalen.

Bevis Følger af sætning 2.12 og sætning 1.10.

Bemærkning Ved at foretage en translation og en drejning (eller spejling) af det oprindelige koordinatsystem har vi opnået, at dispersionsmatricen er en diagonalmatrix, i.e. at komponenterne i den stokastiske vektor er ukorrelerede og dermed uafhængige.

Ved at foretage en reskalering af akserne, kan vi endda opnå, at dispersionsmatricen får lutter nuller eller ettaller i diagonalen. Betragter vi nemlig basisvektorerne

$$c_1 \underline{e}_1, \dots, c_n \underline{e}_n,$$

hvor

$$c_i = \begin{cases} \frac{1}{\sqrt{\lambda_i}} & \text{hvis } \lambda_i > 0 \\ 1 & \text{hvis } \lambda_i = 0 \end{cases},$$

jev. p. 1.38, og kaldes koordinaterne i dette system  $\underline{z}$ , fås sammenhængen

$$\underline{z} = \underline{\underline{C}}' \underline{\underline{P}}' (\underline{x} - \underline{\mu}) = (\underline{\underline{PC}})' (\underline{x} - \underline{\mu}),$$

hvor  $\underline{\underline{C}} = \text{diag}(c_1, \dots, c_n)$ .

Sætter vi  $\underline{z}$ -koordinaterne for  $\underline{x}$  lig  $\underline{z}$  fås derfor

$$\underline{z} \in N(\underline{0}, \underline{\underline{E}}),$$

hvor

$$\underline{\underline{E}} = (\underline{\underline{PC}})' \underline{\underline{\Sigma}} \underline{\underline{PC}} = \underline{\underline{C}}' \underline{\underline{P}}' \underline{\underline{\Sigma}} \underline{\underline{PC}} = \underline{\underline{C}}' \underline{\underline{\Lambda}} \underline{\underline{C}}$$

har nuller eller ettaller i diagonalen.

Overgangen til disse nye baser har en tæt tilknytning til niveaukurverne for tæthedsfunktion for den normale fordeling.

Tætheden for et  $\underline{X} \in N(\underline{\mu}, \underline{\Sigma})$  er som tidligere anført

$$\begin{aligned} f(\underline{x}) &= k \cdot \exp(-\frac{1}{2}(\underline{x} - \underline{\mu})' \underline{\Sigma}^{-1}(\underline{x} - \underline{\mu})) \\ &= k \cdot \exp(-\frac{1}{2}(\|\underline{x} - \underline{\mu}\|^2)). \end{aligned}$$

Derfor er

$$f(\underline{x}) = k_1 \Leftrightarrow (\underline{x} - \underline{\mu})' \underline{\Sigma}^{-1}(\underline{x} - \underline{\mu}) = c,$$

hvor  $k_1$  og  $c$  er konstanter. Da  $\underline{\Sigma}^{-1}$ , er positivt definit er niveaukurverne

$$E_c = \{\underline{x} | f(\underline{x}) = k_1\}$$

derfor ellipsoider, jvf. p. 1.49. Af sætning 1.19 fremgår endvidere, at hovedakserne i disse ellipsoider netop er egenvektorer til  $\underline{\Sigma}^{-1}$ , men ifølge sætning 1.12 er disse også egenvektorer til  $\underline{\Sigma}$ . I de nye koordinater bliver tæthederne derfor

$$g(\underline{y}) = k \cdot \exp(-\frac{1}{2} \sum \frac{1}{\lambda_i} y_i^2),$$

hvor  $\lambda_i$  er den  $i$ 'te egenværdi til  $\underline{\Sigma}$ , og

$$h(\underline{z}) = k_1 \exp(-\frac{1}{2} \sum z_i^2).$$

Ellipsoiderne  $E_1$  kaldes ofte kontur-ellipsoider.

Ved hjælp af ovenstående betragtninger får vi

Sætning 2.15 Lad  $\underline{P}$  og  $\underline{C}$  være som ovenfor. Da gælder

$$(\underline{X} - \underline{\mu})' (\underline{PC}) (\underline{PC})' (\underline{X} - \underline{\mu}) \in \chi^2(\text{rg } \underline{\Sigma}).$$

Hvis specielt  $\underline{\Sigma}$  har fuld rang  $p$  gælder

$$(\underline{X} - \underline{\mu})' \underline{\Sigma}^{-1} (\underline{X} - \underline{\mu}) = \|\underline{X} - \underline{\mu}\|^2 \in \chi^2(p).$$

Bevis

$$\begin{aligned}
 & (\underline{X} - \underline{\mu})' (\underline{PC}) (\underline{PC})' (\underline{X} - \underline{\mu}) \\
 &= \underline{Z}' \underline{Z} = \sum \delta_i z_i^2,
 \end{aligned}$$

hvor  $\delta_i = 1$  hvis  $\lambda_i \neq 0$  og lig 0 ellers.

Da de ikke udartede komponenter i  $\underline{Z}$  er stokastisk uafhængige og  $N(0,1)$ -fordelte følger resultatet umiddelbart. Den sidste bemærkning følger af, at

$$\underline{PC} (\underline{PC})' = \underline{PCC}' \underline{P}' = \underline{P} \underline{\Lambda}^{-1} \underline{P}' = \underline{\Sigma}^{-1}$$

□

Bemærkning Resultatet i sætningen er, at sandsynligheden for at få et udfald inden for en kontur-ellipsoide kan beregnes ved hjælp af en  $\chi^2$ -fordeling.

Disse begreber vil blive eksemplificeret i eksempel 2.3, hvor vi betragter den todimensionale fordeling.

2.2.3 Betingede fordelinger

Vi betragter i dette afsnit en spaltning af en stokastisk variabel  $\underline{X} \in N_p(\underline{\mu}, \underline{\Sigma})$ , nemlig

$$\underline{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}; \quad \underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}; \quad \underline{\Sigma} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Vi har da

Sætning 2.16 Hvis  $X_2$  er regulært fordelt, i.e. hvis  $\Sigma_{22}$  har fuld rang, da er den betingede fordeling af  $X_1$  givet  $X_2 = x_2$  igen en normal fordeling, og der gælder

$$E(\underline{X}_1 | \underline{X}_2 = \underline{x}_2) = \underline{\mu}_1 + \underline{\Sigma}_{12} \underline{\Sigma}_{22}^{-1} (\underline{x}_2 - \underline{\mu}_2)$$

$$D(\underline{X}_1 | \underline{X}_2 = \underline{x}_2) = \underline{\Sigma}_{11} - \underline{\Sigma}_{12} \underline{\Sigma}_{22}^{-1} \underline{\Sigma}_{21}$$

Hvis  $\underline{\Sigma}_{22}$  ikke har fuld rang, er den betingede fordeling stadig normal, og  $\underline{\Sigma}_{22}^{-1}$  skal i ovenstående formler blot erstattes med en generaliseret invers  $\underline{\Sigma}_{22}^-$ .

Bevis Beviset er teknisk og forbigås, jvf. dog afsnit 2.2.5.

Bemærkning Det fremgår, at den betingede varians er uafhængig af  $\underline{x}_2$ . Dette resultat gælder selvsagt ikke for vilkårlige fordelinger, men er specielt for den normale fordeling. Endvidere fremgår, at den betingede middelværdi er en affin funktion af  $\underline{x}_2$ , jvf. betragtningerne i afsnit 2.3.3.

Vi skal ikke her diskutere den nærmere betydning af sætningen, men henviser til eksemplerne i afsnit 2.2.5.

#### 2.2.4 Reproduktivitetssætning og central grænseværdisætning

I analogi til reproduktivitetssætningen for den endimensionale normale fordeling har vi

Sætning 2.17 (Reproduktivitetssætning) Lad  $\underline{X}_1, \dots, \underline{X}_k$  være uafhængige, og lad  $\underline{X}_i \in N(\underline{\mu}_i, \underline{\Sigma}_i)$ .

Da er

$$\sum_{i=1}^k \underline{X}_i \in N\left(\sum_{i=1}^k \underline{\mu}_i, \sum_{i=1}^k \underline{\Sigma}_i\right).$$

Bevis Forbigås.

Som i det endimensionale tilfælde gælder der også centrale grænseværdisætninger, i.e. at summer af uafhængige flerdimensionale stokastiske variable under generelle forudsætninger er asymptotisk normalt fordelte. Vi anfører en analog til Lindeberg-Levy's sætning.

Sætning 2.18 (Central grænseværdisætning). Lad de uafhængige og identisk fordelte variable  $\underline{X}_1, \dots, \underline{X}_n, \dots$  have endeligt første og andet moment

$$\underline{\mu} = E(\underline{X}_1), \quad \underline{\Sigma} = D(\underline{X}_1).$$

Da vil - med  $\bar{\underline{X}}_n = \frac{1}{n}(\underline{X}_1 + \dots + \underline{X}_n)$  -

$$\sqrt{n}(\bar{\underline{X}}_n - \underline{\mu})$$

have en  $N(\underline{0}, \underline{\Sigma})$ -fordeling som grænsefordeling, og vi siger, at  $\bar{\underline{X}}_n$  er asymptotisk  $N(\underline{\mu}, \frac{1}{n}\underline{\Sigma})$ -fordelt.

Bevis Såvel denne som den foregående sætning kan vises ud fra de tilsvarende endimensionale sætninger ved dels at bruge en sætning, som karakteriserer den flerdimensionale fordeling (en flerdimensional variabel er normalt fordelt, netop hvis enhver linearkombination af dens komponenter er (endimensionalt) normalt fordelt), og dels en sætning, som karakteriserer en flerdimensional grænsefordeling ved grænsefordelinger af linearkombinationer af komponenterne (koordinaterne). Dette ligger dog ud over denne fremstillings rammer, og for nærmere detaljer må den interesserede læser henvises til litteraturen, e.g. Rao (1967), afsnit 2c.5.



### 2.2.5 Estimation af parametrene i en flerdimensional normal fordeling

Vi betragter en række observationer  $\underline{X}_1, \dots, \underline{X}_n$ , der antages uafhængige og identisk  $N_p(\underline{\mu}, \underline{\Sigma})$  - fordelte. Vi antager, at der er flere observationer end dimensionen angiver, d.v.s. at  $n > p$ . Vi skal i dette afsnit anføre skøn over parametrene  $\underline{\mu}$  og  $\underline{\Sigma}$ .

Vi indfører betegnelserne

$$\underline{X}_i = \begin{bmatrix} X_{1i} \\ \vdots \\ X_{pi} \end{bmatrix}$$

$$\bar{\underline{X}} = \frac{1}{n} \sum_{i=1}^n \underline{X}_i = \begin{bmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_p \end{bmatrix}$$

$$\underline{S} = \frac{1}{n-1} \sum_{i=1}^n (\underline{X}_i - \bar{\underline{X}})(\underline{X}_i - \bar{\underline{X}})' = \frac{1}{n-1} \sum_{i=1}^n \underline{X}_i \underline{X}_i' - \frac{n}{n-1} \bar{\underline{X}} \bar{\underline{X}}'$$

Betragter vi datamatricen

$$\underline{X} = \begin{bmatrix} \underline{X}'_1 \\ \vdots \\ \underline{X}'_n \end{bmatrix} = \begin{bmatrix} X_{11} & \dots & X_{p1} \\ \vdots & & \vdots \\ X_{1n} & \dots & X_{pn} \end{bmatrix},$$

hvor den  $i$ 'te række netop er den  $i$ 'te observation, kan vi også skrive

$$(n-1)\underline{\underline{S}} = \sum_{i=1}^n (\underline{X}_i - \underline{\bar{X}})(\underline{X}_i - \underline{\bar{X}})' = \underline{\underline{X}}'\underline{\underline{X}} - n\underline{\bar{X}}\underline{\bar{X}}'.$$

Med disse betegnelser er vi i stand til at formulere

**Sætning 2.19** Lad situationen være som ovenfor beskrevet. Da er maximum likelihood estimatorene for  $\underline{\mu}$  og  $\underline{\Sigma}$  lig

$$\hat{\underline{\mu}} = \underline{\bar{X}}$$

$$\hat{\underline{\Sigma}} = \frac{n-1}{n}\underline{\underline{S}} = \frac{1}{n} \sum_{i=1}^n (\underline{X}_i - \underline{\bar{X}})(\underline{X}_i - \underline{\bar{X}})' .$$

$\hat{\underline{\mu}}$  er et centralt skøn over  $\underline{\mu}$ , og  $\hat{\underline{\Sigma}}$  er et centralt skøn over  $\underline{\Sigma}$ .

**Bevis** Forbigås, se f.eks. Anderson (1958), kapitel 3.

**Bemærkning** Da den empiriske dispersionsmatrix  $\underline{\underline{S}}$  er et centralt skøn over  $\underline{\Sigma}$ , og da den kun afviger fra maximum likelihood estimatoren med faktoren  $n/n-1$ , foretrækkes  $\underline{\underline{S}}$  ofte som skøn. Man vil da også hyppigt anvende betegnelsen  $\hat{\underline{\Sigma}}$  om  $\underline{\underline{S}}$ . Man må derfor i konkrete fremstillinger være opmærksom på, hvad et udtryk som  $\hat{\underline{\Sigma}}$  præcist dækker over.

Fordelingen af  $\hat{\underline{\mu}}$  fås trivielt ved hjælp af sætning 2.17. Der gælder

$$\hat{\underline{\mu}} = \underline{\bar{X}} \in N_p(\underline{\mu}, \frac{1}{n}\underline{\Sigma}).$$

Fordelingen af  $\underline{\underline{S}}$  er mere kompliceret. Den er anført i afsnit 2.5.

Vi anfører et eksempel på estimation af parametrene i næste afsnit.

### 2.2.6 Den todimensionale normale fordeling

Vi specialiserer nu de tidligere resultater til to dimensioner.

Lad  $\underline{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$  være normalt fordelt  $(\underline{\mu}, \underline{\Sigma})$ , hvor

$$\underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

og

$$\underline{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

Da

$$\det(\underline{\Sigma}) = \sigma_1^2 \sigma_2^2 - \sigma_{12}^2$$

er, såfremt  $\det(\underline{\Sigma}) \neq 0$ ,

$$\underline{\Sigma}^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} \begin{bmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{bmatrix}.$$

Indføres korrelationskoefficienten  $\rho$

$$\rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2},$$

fås

$$\underline{\Sigma}^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix},$$

og tætheden bliver

$$f(x_1, x_2) = \frac{1}{2\pi} \frac{1}{\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2} \frac{1}{1-\rho^2} \left[ \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 - 2\rho \frac{x_1 - \mu_1}{\sigma_1} \frac{x_2 - \mu_2}{\sigma_2} + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 \right]\right\}.$$

Grafen er skitseret i fig. 2.2

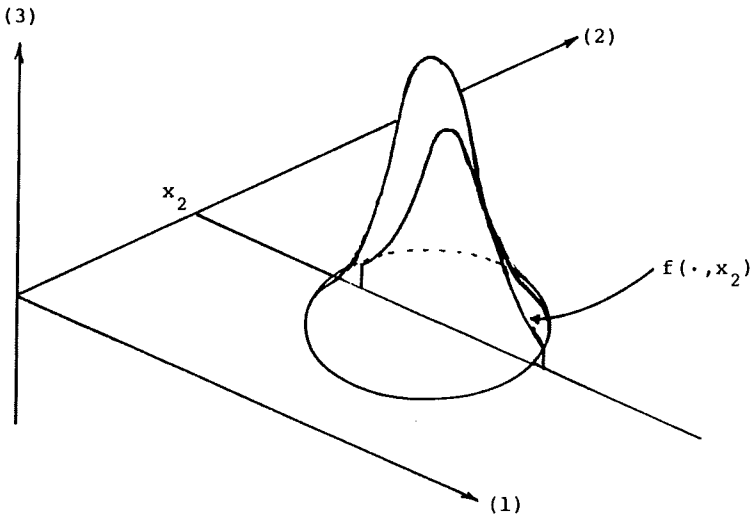


Fig. 2.2 Tæthed for to-dimensional normal fordeling.

Det ses umiddelbart, at der er tale om en produktfordeling, i.e. at  $X_1$  og  $X_2$  er stokastisk uafhængige, netop hvis  $\rho = 0$ , d.v.s. netop hvis  $\underline{\Sigma}$  er en diagonalmatrix.

Den betingede tæthed for  $X_1$  givet  $X_2 = x_2$  er proportional med skæringskurven mellem planen gennem  $(0, x_2, 0)$  parallel med (1)-(3) planen. Kaldes tætheden  $g$  har vi altså

$$g(\cdot) = cf(\cdot, x_2) ,$$

hvor  $c$  er en normeringskonstant. Vi har

$$\begin{aligned} g(x_1) &= k_1 \cdot \exp\left[-\frac{1}{2} \frac{1}{1-\rho^2} \left\{ \left[ \frac{x_1 - \mu_1}{\sigma_1} \right]^2 - 2\rho \frac{x_1 - \mu_1}{\sigma_1} \frac{x_2 - \mu_2}{\sigma_2} \right\} \right] \\ &= k_2 \exp\left[-\frac{1}{2} \frac{1}{1-\rho^2} \left[ \frac{x_1 - \mu_1}{\sigma_1} - \rho \frac{x_2 - \mu_2}{\sigma_2} \right]^2 \right] \\ &= k_3 \exp\left[-\frac{1}{2} \frac{1}{\sigma_1^2 (1-\rho^2)} (x_1 - \mu_1 - \rho \frac{\sigma_1}{\sigma_2} (x_2 - \mu_2))^2 \right] \\ &= k_3 \left[ \exp - \frac{1}{2\gamma^2} (x_1 - \xi_1)^2 \right] . \end{aligned}$$

Vi har her ikke holdt regnskab med led der kun indeholder  $x_2$ . De er gået ud som diverse konstanter. Af slutresultatet aflæses, at den betingede fordeling er normal, og at

$$k_3 = \frac{1}{\sqrt{2\pi} \sigma_1 \sqrt{1-\rho^2}} ,$$

samt endelig, at

$$E(X_1 | X_2 = x_2) = \xi_1 = \mu_1 + \rho \frac{\sigma_1}{\sigma_2} (x_2 - \mu_2)$$

og

$$V(X_1 | X_2 = x_2) = \gamma^2 = \sigma_1^2(1-\rho^2) .$$

Vi har bevist resultatet i sætning 2.16 i tilfældet  $n = 2$ . Det bemærkes, at den betingede middelværdi afhænger lineært (rettere affint) af  $x_2$ , og at den betingede varians er uafhængig af  $x_2$ . Endvidere er

$$V(X_1 | X_2 = x_2) \leq V(X_1) ,$$

og den kvadrerede korrelationskoefficient angiver variansreduktionen, d.v.s. den brøkdel af  $X_1$ 's varians, der forklares af  $X_2$ , idet

$$\rho^2 = \frac{V(X_1) - V(X_1 | X_2 = x_2)}{V(X_1)} .$$

I det følgende eksempel betragter vi et numerisk eksempel, der også involverer et estimationsproblem.

Eksempel 2.2 I nedenstående tabel er anført sammenhørende målinger af luftens indhold af svævestøv målt i  $\mu\text{g}/\text{m}^3$ . Ved målingerne er anvendt to forskellige måleprincipper, nemlig et sværtningssprincip (med et såkaldt OECD-apparat) og et vejningsprincip (med en såkaldt High Volume Sampler). Grunden til, at der kan optræde store afvigelser er bl.a., at målinger ved sværtningssprincippet er meget følsomme overfor svævestøvs eventuelle afvigelse fra "normalstøv". Således vil et stort indhold af kalkstøv i luften kunne bevirke, at målingerne bliver systematisk for små.

Metode	I	2	5	15	16	16	19	26	24	16	36
	II	2	12	4	21	41	14	31	29	31	8
	I	39	42	44	40	42	42	50	51	58	64
	II	30	44	26	60	34	34	14	41	58	47

Vi opfatter ovenstående data som realiserede udfald af uafhængige, identisk fordelte stokastiske variable

$$\begin{bmatrix} X_1 \\ Y_1 \end{bmatrix}, \dots, \begin{bmatrix} X_{20} \\ Y_{20} \end{bmatrix}.$$

Vi vil undersøge, om fordelingen kan antages at være normal med parametre  $(\underline{\mu}, \underline{\Sigma})$ . Såfremt fordelingen er normal, finder vi skønnene

$$\underline{\mu} = \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix} = \begin{bmatrix} \bar{X} \\ \bar{Y} \end{bmatrix} = \begin{bmatrix} 32.35 \\ 29.05 \end{bmatrix},$$

og

$$\hat{\underline{\Sigma}} = \begin{bmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} \\ \hat{\sigma}_{12} & \hat{\sigma}_2^2 \end{bmatrix} = \begin{bmatrix} S_1^2 & S_{12} \\ S_{12} & S_2^2 \end{bmatrix} = \begin{bmatrix} 311 & 182 \\ 182 & 279 \end{bmatrix},$$

hvor  $\hat{\underline{\Sigma}}$  er det centrale skøn over  $\underline{\Sigma}$ . Specielt er

$$S_{12} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}).$$

Vi vil altså undersøge, om observationerne kan tænkes at komme fra en normal fordeling med parametre  $(\hat{\underline{\mu}}, \hat{\underline{\Sigma}})$ . Til den ende bestemmer vi først konturellipserne. Egenværdierne og egenvektorerne for  $\hat{\underline{\Sigma}}$  er

$$\hat{\lambda}_1 = 477.613 \text{ og } \hat{E}_1 = \begin{bmatrix} 0.736 \\ 0.678 \end{bmatrix}$$

og

$$\hat{\lambda}_2 = 112.676 \text{ og } \hat{E}_2 = \begin{bmatrix} -0.678 \\ 0.736 \end{bmatrix}.$$

Vælger vi koordinatsystemet med centrum i  $\hat{\underline{\mu}}$  og med  $\underline{p}_1$  og  $\underline{p}_2$  som basisvektorer, får konturellipserne ligninger af formen

$$\frac{z_1^2}{\hat{\lambda}_1} + \frac{z_2^2}{\hat{\lambda}_2} = c ,$$

eller

$$\frac{z_1^2}{477.613} + \frac{z_2^2}{112.676} = c ,$$

hvor de nye koordinater  $\underline{z}$  er givet ved

$$\underline{Pz} = (\underline{p}_1 \underline{p}_2) \underline{z} = \underline{x} - \hat{\underline{\mu}} .$$

I figur 2.2 er der indtegnet observationerne og 3 konturellipser svarende til c-værdierne  $c_1 = \chi^2(2)_{0.40}$ ,  $c_2 = \chi^2(2)_{0.80} = 3.22$  og  $c_3 = \chi^2(2)_{0.95} = 5.99$ .

Dette bevirker, ifølge sætning 2.16, at der i normalfordelingen med parametre  $(\hat{\underline{\mu}}, \hat{\underline{\Sigma}})$  er sandsynlighederne 40%, 80% og 95% for at få observationer indenfor den inderste, den mellemste og den yderste ellipse. For områderne mellem ellipserne, resp. uden for disse, fås da sandsynlighederne 40%, 40%, 15% og 5%. Disse tal kan sammenlignes med de tilsvarende observerede relative hyppigheder, som er 40%, 30%, 30% og 0%. Overensstemmelsen er, om ikke overvældende, så dog acceptabel.

Hvis man ønsker et mere præcist resultat, kan man udføre et  $\chi^2$ -test. Det vil da være rimeligt at foretage en yderligere inddeling af planen efter egenvektorernes retninger. I det aktuelle tilfælde får vi da  $4 \times 4$  områder med estimerede sandsynligheder på 10%, 10%, 3.75% og 1.25%. Man kan da beregne den sædvanlige  $\chi^2$ -teststørrelse



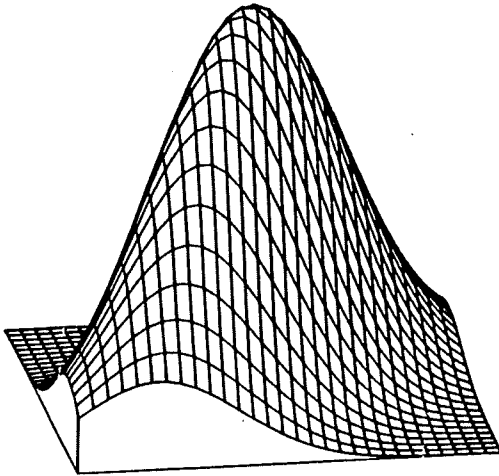
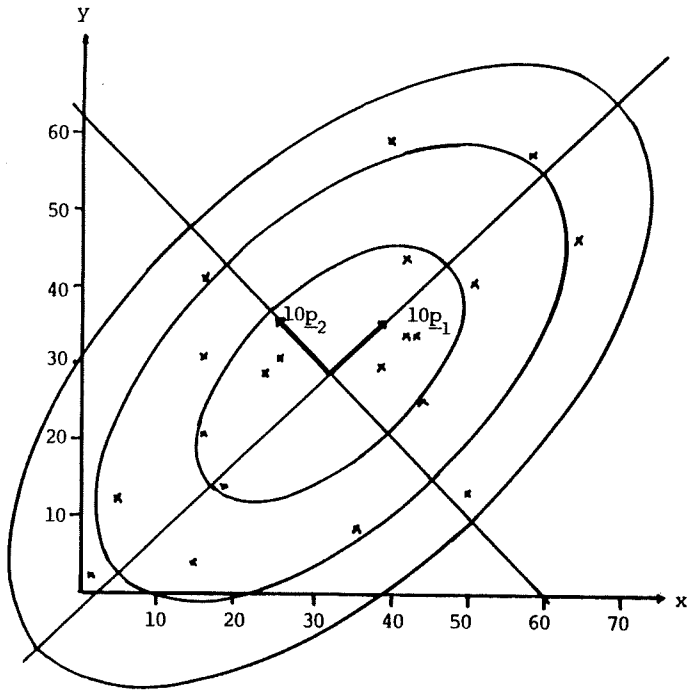


Fig. 2.3 Estimerede konturellipser og estimeret tæthedsfunktion svarende til data i eksempel 2.2.

$$\Sigma \frac{(\text{obs.}-\text{forv.})^2}{\text{forv.}}$$

og sammenligne med en  $\chi^2(n-6)$ -fordeling (vi har estimeret 5 parametre). I det konkrete tilfælde er der måske lige lovligt få observationer til at kunne foretage denne analyse.

Korrelationskoefficienten estimeres til

$$\hat{\rho} = \frac{182}{\sqrt{311 \cdot 279}} = 0.62 ,$$

hvorfor de betingede varianser skønnes til

$$\hat{V}(X|Y = y) = 311(1 - \hat{\rho}^2) = 192$$

$$\hat{V}(Y|X = x) = 279(1 - \hat{\rho}^2) = 172.$$

Vi ser, at de betingede varianser er blevet reduceret med 38% svarende til, at  $\hat{\rho}^2 = 0.38$ . Dette at den betingede varians af f.eks. en OECD-måling for fastholdt High Volume Sampler måling er væsentlig mindre end den ubetingede varians virker vel også ganske rimeligt. Ved vi, at svævestøvsindholdet ved hjælp af en High Volume Sampler er bestemt til e.g.  $2 \mu\text{g}/\text{m}^3$ , vil vi vel ikke forvente at få resultater med OECD-apparatet, der afviger væsentligt fra dette, svarende til, at vi får en beskedent betinget varians. Kendes High Volume Sampler resultatet ikke, ja da må vi forvente, at OECD-målingen vil kunne ligge i hele det naturlige variationsområde, svarende til en større, ubetinget varians.

□

### 2.3 Korrelation og regression

I dette afsnit skal vi nøjere diskutere parametrene betydning i en flerdimensional normal fordeling. Vi vil først forsøge

at generalisere de egenskaber ved korrelationskoefficienten, som fremgik af foregående afsnit.

### 2.3.1 Den partielle korrelationskoefficient

Udgangspunktet er formelen for de betingede fordelinger i en flerdimensional normal fordeling. Lad  $\underline{X} \in N_p(\underline{\mu}, \underline{\Sigma})$ , og lad de variable være spaltet som følger

$$\underline{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}; \quad \underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}; \quad \underline{\Sigma} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

hvor  $X_1$  består af de  $m$  første pladser i  $\underline{X}$  og analogt med de øvrige. Da er den betingede dispersion af  $X_1$  for givet  $X_2 = \underline{x}_2$  som anført i sætning 2.16 lig

$$D(X_1 | X_2 = \underline{x}_2) = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}.$$

Ved den partielle korrelationskoefficient mellem  $X_i$  og  $X_j$ ,  $i, j \leq m$ , for givet  $X_2 = \underline{x}_2$  forstås da korrelationen i den betingede fordeling af  $X_1$  givet  $X_2 = \underline{x}_2$ . Den betegnes  $\rho_{ij|m+1, \dots, p}$ .

Sætter vi

$$\underline{\Sigma} = \begin{bmatrix} \sigma_1^2 & \dots & \sigma_{p1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \sigma_{1p} & \dots & \sigma_p^2 \end{bmatrix}$$

og

$$\underline{\Sigma}_{11} - \underline{\Sigma}_{12} \underline{\Sigma}_{22}^{-1} \underline{\Sigma}_{21} = \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \vdots & & \vdots \\ a_{1m} & \dots & a_{mm} \end{bmatrix},$$

har vi altså

$$\rho_{ij|m+1, \dots, n} = \frac{a_{ij}}{\sqrt{a_{ii}} \sqrt{a_{jj}}}. \quad (*)$$

Er specielt  $\underline{X}$  tredimensional fås med

$$\underline{\Sigma} = \left[ \begin{array}{cc|c} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \hline \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{array} \right],$$

at

$$\begin{aligned} \underline{\Sigma}_{11} - \underline{\Sigma}_{12} \underline{\Sigma}_{22}^{-1} \underline{\Sigma}_{21} &= \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} - \frac{1}{\sigma_3^2} \begin{bmatrix} \rho_{13}^2\sigma_1^2\sigma_3^2 & \rho_{13}\rho_{23}\sigma_1\sigma_2\sigma_3^2 \\ \rho_{13}\rho_{23}\sigma_1\sigma_2\sigma_3^2 & \rho_{23}^2\sigma_2^2\sigma_3^2 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1^2(1-\rho_{13}^2) & \sigma_1\sigma_2(\rho_{12}-\rho_{13}\rho_{23}) \\ \sigma_1\sigma_2(\rho_{12}-\rho_{13}\rho_{23}) & \sigma_2^2(1-\rho_{23}^2) \end{bmatrix}. \end{aligned}$$

Heraf følger, at den partielle korrelationskoefficient mellem  $X_1$  og  $X_2$  givet  $X_3$  er

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

For en  $p$ -dimensional vektor  $\underline{X}$  finder vi derfor

$$\rho_{ij|k} = \frac{\rho_{ij} - \rho_{ik}\rho_{jk}}{\sqrt{(1-\rho_{ik}^2)(1-\rho_{jk}^2)}} \quad (**)$$

Da man kan finde betingede fordelinger for givet  $X_{m+1}, \dots, X_p$  ved successive betingninger, kan man derfor få partielle korrelationskoefficienter af højere orden ved successiv anvendelse af (\*\*). Eksempelvis findes

$$\rho_{ij|k\ell} = \frac{\rho_{ij|k} - \rho_{i\ell|k}\rho_{j\ell|k}}{\sqrt{(1-\rho_{i\ell|k}^2)(1-\rho_{j\ell|k}^2)}},$$

hvor vi først har betinget med  $X_k$  og dernæst med  $X_\ell$ .

I afsnit 2.2.6 så vi, at korrelationskoefficienten er et mål for den variansreduktion vi opnår ved at betinge med den ene variabel. Da de partielle korrelationskoefficienter blot er korrelationer i betingede fordelinger, kan vi her anvende den samme tolkning. F.eks. har vi, at  $\rho_{ij|k\ell}^2$  angiver den brøkdelen af  $X_i$ 's varians for fastholdt  $X_k = x_k$  og  $X_\ell = x_\ell$ , der forklares af  $X_j$ . Det må her indskærpes, at disse tolkninger er snævert knyttede til forudsætningen om normalitet. I det generelle tilfælde, vil de betingede varianser afhænge af de værdier, hvormed der betinges (i.e. afhænge af  $x_k$  og  $x_\ell$ ).

Ved estimation af partielle korrelationer estimerer man blot dispersionsmatricen og beregner derefter de partielle korrelationer som anført. Hvis estimatet af dispersionsmatricen er en maksimum likelihood estimator, da vil de skøn over de partielle korrelationer, man får frem på denne måde også være maksimum likelihood estimatører (jvf. sætning 10 p. 2.28 i bind I).

Vi vil nu give en illustration af begreberne i

Eksempel 2.3 (Data stammer fra Tokkesdal Pedersen og Skjøth

(1976)). I nedenstående tabel er anført korrelationskoefficienter mellem 3- og 28-døgnstyrken for Portland Cement og cementens indhold af mineralerne  $C_3S$  (Alit, tricalciumsilikat,  $Ca_3SiO_5$ ) og  $C_3A$  (Aluminat, tricalciumaluminat,  $Ca_3Al_2O_6$ ), samt finheden (BLAINE). Korrelationerne er estimeret på basis af 51 sammenhørende observationer.

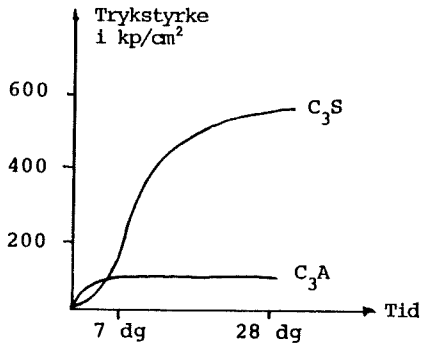
	$C_3S$	$C_3A$	BLAINE	Styrke 3	Styrke 28
$C_3S$	1	-0.309	0.091	0.158	0.344
$C_3A$	-0.309	1	0.192	0.120	-0.166
BLAINE	0.091	0.192	1	0.745	0.320
Styrke 3	0.158	0.120	0.745	1	0.464
Styrke 28	0.344	-0.168	0.320	0.464	1

Tabel 2.1 Korrelationsmatrix for 5 cementvariable.

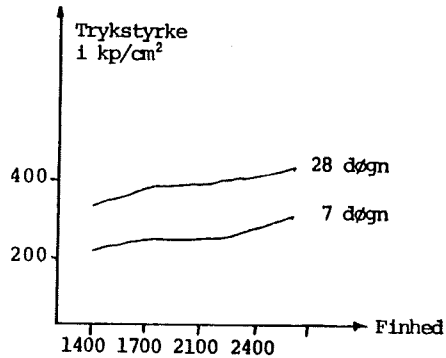
Det skal yderligere oplyses, at  $C_3S$  udgør ca. 35-60% af almindelige portlandklinker og  $C_3A$  ca. 5-18% af klinkerne. BLAINE-tallet er et mål for den specifikke overflade, således at et stort BLAINE tal svarer til en meget finmalet cement.

Vi vil især interessere os for sammenhængen mellem  $C_3A$  indholdet i klinkerne og styrkerne. Det er almindeligt anerkendt jvf. nedenstående figur, at et stort  $C_3A$ -indhold giver en større 3-døgnstyrke, hvilket også er i overensstemmelse med  $\hat{\rho}_{C_3A, \text{Strk } 3} = 0.120$ .

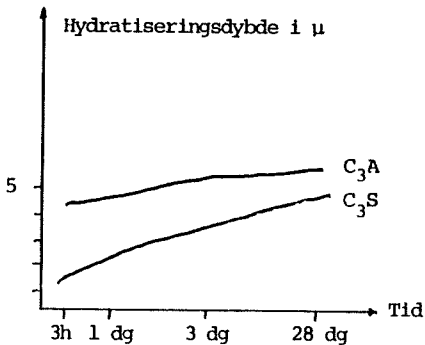
Problemet er, om denne større 3-døgnstyrke for cementer med højt  $C_3A$ -indhold alene skyldes  $C_3A$ 's større hydratiseringsgrad (jo hurtigere vandet går i forbindelse med cementen, jo hurtigere vil den selvsagt reagere og få øget styrke).  $C_3A$ 's



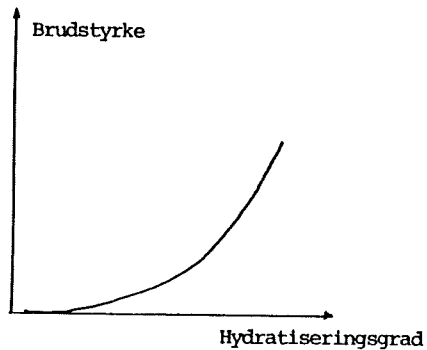
Figur 2.4 Trykbrudstyrker ved stuetemperatur af pastaer af  $\text{C}_3\text{S}$  og  $\text{C}_3\text{A}$  lagret til forskellig tid. (efter Knudsen (1975)).



Figur 2.5 Trykstyrker for forskellige finheder af cementen. (efter Knudsen (1975)).



Figur 2.6 Hydratiseringsgrad for cementmineraller i forhold til tid (efter Knudsen (1975)).



Figur 2.7 Sammenhæng mellem hydratiseringsgrad målt ved såkaldt korrigeret glødetab og brudstyrke (efter Knudsen (1975)).

langt større hydratisering efter 3 døgn fremgår af figur 2.6, og hydratiseringsgradens indflydelse på styrken er skitseret i figur 2.7.

Ser vi på korrelationsmatricen fremgår imidlertid også, at  $C_3A$ -indholdet er positivt korreleret med BLAINE-tallet, d.v.s. at cementer med meget  $C_3A$  som regel vil være ganske fintmalede, og som det fremgår af figur 2.5 skulle dette også medvirke til en øgning af styrkerne.

Endelig bemærker vi, at 28-døgnstyrken er svagt negativt korreleret med  $C_3A$ -indholdet. Dette virker ikke urimeligt, når vi betragter  $C_3S$ 's og  $C_3A$ 's styrkeudvikling med tiden, som f.eks. anført i figur 2.4, selv om den øgede finhed (for en  $C_3A$ -rig cement) også skulle afspejles i 28 døgnstyrkerne, jvf. fig. 2.5.

For at adskille disse forskellige egenskaber ved  $C_3A$  fra de effekter, der opstår ved at  $C_3A$ -rig cement åbenbart er lettere at male og derfor som regel optræder i en lidt mere fintmalt form, finder vi de betingede korrelationer for fastholdt BLAINE-tal. De bliver

	$C_3S$	$C_3A$	Styrke 3	Styrke 28
$C_3S$	1	-0.333	0.137	0.333
$C_3A$	-0.333	1	-0.035	-0.246
Styrke 3	0.137	-0.035	1	0.358
Styrke 28	0.333	-0.246	0.358	1

Tabel 2. Korrelationsmatrix for 4 cementvariable for fastholdt BLAINE-tal.

Vi ser, at den partielle korrelationskoefficient mellem styrke 3 og  $C_3A$  for givet finhed er negativ (hvorimod den ubetingede korrelationskoefficient var positiv). Dette indebæ-



rer, at vi for fastholdt finhed må forvente, at cementer med højt  $C_3A$  indhold vil tendere til at have lavere styrker. Dette kunne indicere, at de store 3-døgnstyrker for cementer med højt  $C_3A$ -indhold måske snarere skyldes, at disse cementer har højt BLAINE-tal (d.v.s. at de åbenbart males noget lettere) end, at  $C_3A$  hydratiserer hurtigt!

Vi ser en tilsvarende effekt på korrelationen mellem  $C_3A$  og Styrke 28. Her er den ubetingede korrelation  $-0.168$  og den partielle for fastholdt BLAINE-tal formindsket til  $-0.246$ .

□

Bemærkning Ovenstående eksempel viser, at man må være meget varsom med fortolkninger af korrelationskoefficienter. Det vil f.eks. være direkte misvisende at udtale, at et højt  $C_3A$  indhold sikrer en stor 3-døgnstyrke. For det første kan man ikke slutte noget om eventuelle årsagssammenhænge ved at vurdere korrelationer. Det man kan udtale i første omgang er, at der er en tendens til, at højt  $C_3A$ -indhold og høj 3-døgnstyrke optræder samtidigt. Dette kan meget vel skyldes at de begge afhænger af en ukendt 3. die faktor uden, at der er tale om nogen direkte sammenhæng mellem de to variable. For det andet ser vi også, at en overgang til partielle korrelationer kan give et fortegnsskifte svarende til en effekt, der er modsat den, man får ved den direkte analyse. Dette skyldes en samvariation med den tredje faktor (her BLAINE), som forstyrrer billedet.

I mange situationer vil det være ønskeligt at kunne teste, hvorvidt en korrelationskoefficient kan antages at være 0. Man kan da benytte

Sætning 2.20 Lad  $R = R_{ij|m+1\dots p}$  være den empiriske partielle korrelationskoefficient mellem  $X_i$  og  $X_j$  for givet  $X_{m+1}, \dots, X_p$ . Den forudsættes beregnet ud fra det centrale

skøn over dispersionsmatricen, og på grundlag af  $n$  observationer. Da er

$$\frac{R}{\sqrt{1-R^2}} \sqrt{n-2-(p-m)} \in t(n-2-(p-m)),$$

hvis  $\rho_{ij|m+1\dots p} = 0$ .

2

Bevis Forbigås.

Bemærkning Tallet  $(p-m)$  angiver det antal variable, der fastholdes. Antallet af frihedsgrader er derfor blot lig antallet af observationer minus 2 minus antal fastholdte variable. Sætningen er også gyldig, hvis  $p-m = 0$ , d.v.s. hvis der er tale om en ubetinget korrelationskoefficient.

Vi fortsætter eksempel 2.3 i

Eksempel 2.4 Lad os undersøge, om den fundne værdi af  $r_{24|3}$  er signifikant forskellig fra 0. Vi finder med  $r_{24|3} = R$ :

$$\begin{aligned} \frac{R}{\sqrt{1-R^2}} \sqrt{n-2-(p-m)} &= \frac{-0.035}{\sqrt{1-0.035^2}} \cdot \sqrt{51-2-(5-4)} \\ &= -0.243 = t(48)_{40\%}. \end{aligned}$$

En hypotese om, at  $\rho_{24|3}$  er 0 vil derfor blive accepteret ved et test på niveau  $\alpha$  for  $\alpha < 80\%$ .

□

Hvis vi ønsker at teste andre værdier for  $\rho$  eller at bestemme konfidensintervaller, kan vi benytte

Sætning 2.21 Lad situationen være som i foregående sætning.  
Vi betragter hypotesen

$$H_0 : \rho_{ij|m+1, \dots, p} = \rho_0$$

mod

$$H_1 : \rho_{ij|m+1, \dots, n} \neq \rho_0 .$$

Vi sætter

$$z = \frac{1}{2} \log \frac{1+R_{ij|m+1, \dots, p}}{1-R_{ij|m+1, \dots, p}}$$

og

$$z_0 = \frac{1}{2} \log \frac{1+\rho_0}{1-\rho_0} .$$

Under  $H_0$  vil da

$$(z - z_0) \cdot \sqrt{n-(p-m)-3} \text{ app. } \in N(0,1) .$$

Bevis Forbigås.

Eksempel 2.5 Lad os eksempelvis bestemme et 95% konfidensinterval for  $\rho_{24|3}$  i eksempel 2.4 . Nu er

$$P\{-1.96 < (z - z) \cdot \sqrt{51-(5-4)-3} < 1.96\} \approx 95\%$$

$$\Leftrightarrow P\{-1.96 - 6.86 z < -6.86 z < 1.96 - 6.86 z\} \approx 95\%$$

$$\Leftrightarrow P\{z - 0.29 < z < z + 0.29\} \approx 95\% .$$

Sammenhængen mellem  $z$  og  $\rho_{24|3} = \rho$  er

$$z = \frac{1}{2} \log \frac{1+\rho}{1-\rho} \Leftrightarrow \rho = \frac{e^{2z} - 1}{e^{2z} + 1}$$

Den observerede værdi af Z er

$$z = \frac{1}{2} \log \frac{1-0.035}{1+0.035} = -0.03501 .$$

Grænserne for z bliver

$$[-0.3250, 0.2549] .$$

De tilsvarende grænser for  $\rho_{25|4}$  er

$$\left[ \frac{e^{-0.6500} - 1}{e^{-0.6500} + 1}, \frac{e^{0.5098} - 1}{e^{0.5098} + 1} \right] = [-0.31, 0.25] .$$

□

### 2.3.2 Den multiple korrelationskoefficient

De partielle korrelationskoefficienter er en mulig udvidelse af korrelationen mellem 2 variable. Disse partielle størrelser sigter mere på at beskrive samvariationen mellem to variable. Vi vil nu forlade dette sigte og i stedet betragte formelen (p. 2.30)

$$\rho^2 = \frac{V(X_1) - V(X_1|X_2 = x_2)}{V(X_1)} ,$$

som giver korrelationskoefficientens variationsreducerende betydning. Dette ønsker vi nu at generalisere. Vi betragter igen opspaltningen af den p-dimensionale normalt fordelte vektor  $\underline{X}$  i en m-dimensional vektor  $\underline{X}_1$  og en (p-m)-dimensional vektor  $\underline{X}_2$ , og den deraf affødte opspaltning af parametrene, d.v.s.

$$\underline{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}; \underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}; \underline{\Sigma} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Vi definerer da den multiple korrelationskoefficient mellem  $X_1$ ,  $i = 1, \dots, m$  og  $X_2$  som den maksimal korrelation mellem  $X_1$  og en linearkombination af  $X_2$ 's elementer.

Den betegnes  $\rho_i |_{m+1, \dots, p}$ .

Det kan vises, at den "optimale" linearkombination af  $X_2$ 's elementer er

$$\beta'_i X_2 = (\Sigma_{12} \Sigma_{22}^{-1})_i X_2,$$

hvor  $\beta'_i$  er den  $i$ 'te række i matricen  $\Sigma_{12} \Sigma_{22}^{-1}$ . Det er denne matrix, der forekommer i udtrykket for den betingede middelværdi af  $X_1$  for givet  $X_2$ . Denne er jo

$$E(X_1 | X_2 = \underline{x}_2) = \Sigma_{12} \Sigma_{22}^{-1} (\underline{x}_2 - \underline{\mu}_2) = \begin{bmatrix} \beta'_1 \\ \vdots \\ \beta'_m \end{bmatrix} (\underline{x}_2 - \underline{\mu}_2).$$

Det kan også vises, at

$$\inf_{\alpha} V(X_1 - \alpha' X_2) = V(X_1 - \beta'_i X_2),$$

d.v.s. den betragtede linearkombination minimaliserer variansen af  $X_1 - \alpha' X_2$ .

Vi har nu følgende vigtige

**Sætning 2.22** Vi betragter den ovenstående situation. Lad  $\sigma_i$  være den  $i$ 'te søjle i  $\Sigma_{21}$ , d.v.s.  $\sigma'_i$  den  $i$ 'te række i  $\Sigma_{12}$ .

Da gælder

$$\rho_{i|m+1, \dots, p} = \frac{\sqrt{\sigma_{i2}^{-1} \Sigma_{22}^{-1} \sigma_{i1}}}{\sqrt{\sigma_{i1}}} .$$

Sættes

$$\underline{\Sigma}_i = \begin{bmatrix} \sigma_{i1} & \sigma_{i2} \\ \sigma_{i1} & \Sigma_{22} \end{bmatrix} ,$$

er

$$1 - \rho_{i|m+1, \dots, p}^2 = \frac{\det \underline{\Sigma}_i}{\sigma_{i1} \det \Sigma_{22}} = \frac{V(X_i | \underline{X}_2)}{V(X_i)} .$$

Bevis Beviserne for påstandene inden sætningen er ret simple. Der skal blot indføres en Lagrange multiplikator og benyttes, at dispersionmatricen er positivt semidefinit.

Postulaterne i sætningen følger ved anvendelse af formlen for den betingede dispersionsstruktur (p. 2.23) anvende på  $\underline{\Sigma}_i$  ved hjælp af matrixformlerne i afsnit 1.2.7.

□

Bemærkning 1 Med sætningen har vi fået en række karakteristika for den multiple korrelationskoefficient, og da

$$\rho_{i|m+1, \dots, p}^2 = \frac{V(X_i) - V(X_i | \underline{X}_2)}{V(X_i)} ,$$

ser vi, at vi har fået generaliseret variansreduktionsegenskaben. Det er iøvrigt vigtigt at bemærke, at det af determinantformlen fremgår, at man kan beregne den multiple korrelationskoefficient ud fra korrelationsmatricen ved de samme formler,

som gælder ved beregning ud fra dispersionsmatricen.

Angående estimation af multiple korrelationskoefficienter gælder fuldstændig samme bemærkning som anført p. 2.37 angående partielle koefficienter.

I nedenstående eksempel fortsættes eksempel 2.4.

Eksempel 2.6 For at få et indtryk af, i hvor høj grad mineralindholdene  $C_3A$  og  $C_3S$  i eksempel 2.4 kan forklare variationer i f.eks. 3-døgnsstyrken kan vi beregne den multiple korrelationskoefficient mellem Styrke 3 og  $(C_3S, C_3A)$ . Vi finder

$$1 - \hat{\rho}_3^2|_{12} = \frac{\det \begin{bmatrix} 1 & 0.158 & 0.120 \\ 0.158 & 1 & -0.309 \\ 0.120 & -0.309 & 1 \end{bmatrix}}{1 \cdot \det \begin{bmatrix} 1 & -0.309 \\ -0.309 & 1 \end{bmatrix}}$$

hvor nummereringen af de variable svarer til den i eksempel 2.4 anvendte. Vi finder

$$\hat{\rho}_3^2|_{12} = 1 - 0.9435 = 0.0565 .$$

Data tyder derfor på, at kun knap 6% af variationen i cementstyrken (af prøver indsamlet på den måde de konkrete data er indsamlet) kan forklares ved variationer i  $C_3S$  og  $C_3A$  indholdet alene.

□

Hvis den multiple korrelationskoefficient er 0, (i.e. hvis  $\sigma_i = 0$ ) er det ikke vanskeligt at bestemme fordelingen af  $\hat{\rho}_i^2|_{m+1, \dots, p}$ . Vi giver resultatet i en lidt ændret form i

Sætning 2.23 Lad  $R = \hat{\rho}_{i|m+1, \dots, p}$  være den empiriske multiple korrelationskoefficient mellem  $X_i$  og  $\underline{X}_2 = (X_{m+1}, \dots, X_p)$  baseret på  $n$  observationer. Da er

$$\frac{R^2}{1-R^2} \cdot \frac{n-(p-m)-1}{p-m} \in F(p-m, n-(p-m)-1) ,$$

såfremt  $\rho_{i|m+1, \dots, p} = 0$ .

Bevis Forbigås.

Dette kan bruges ved test af hypotesen

$$H_0 : \rho_{i|m+1, \dots, p} = 0 \text{ mod } H_1 : \rho_{i|m+1, \dots, p} \neq 0 .$$

Vi forkaster hypotesen for store værdier af teststørrelsen. Vi illustrerer dette i

Eksempel 2.7 Vi betragter situationen fra eksempel 2.6 og vil undersøge, om det kan antages, at den multiple korrelation mellem  $X_1$  og  $(X_2, X_3)$  er 0. Vi finder teststørrelsen

$$\frac{R^2}{1-R^2} \cdot \frac{51-(3-1)-1}{3-1} = \frac{0.0565}{0.9435} \cdot \frac{48}{2} = 1.44 .$$

Da

$$F(2, 48)_{0.90} = 2.42 ,$$

vil vi i det mindste acceptere en hypotese om, at  $\rho_{3|12} = 0$  på alle niveauer  $\alpha < 10\%$ . Med det foreliggende materiale kan det altså ikke afvises, at  $\rho_{3|12} = 0$ . Dette udelukker selvsagt ikke, at den er  $\neq 0$  (hvad den givetvis er); vi kan blot ikke med sikkerhed konstatere det med de foreliggende data,



fordi  $\rho_{3|12}$  formentlig er ret lille.

□

Vi skal ikke komme ind på spørgsmål angående test for andre værdier af  $\rho_{i|m+1, \dots, n}$ .

### 2.3.3 Regression

Vi skal ikke her give nogen særligt dybtgående gennemgang af den såkaldte regressionssteori, der ikke må forveksles med den i følgende afsnit omtalte (lineære) regressionsanalyse.

Lad  $\begin{pmatrix} Y \\ X \end{pmatrix}$  være en stokastisk vektor. Ved regressionen af Y på X forstås da funktionen, der er givet ved

$$g(\underline{x}) = E(Y|X = \underline{x}) ,$$

d.v.s. den betingede middelværdi som funktion af den betingede variabel.

Lad specielt  $\begin{pmatrix} Y \\ X \end{pmatrix}$  være normalt fordelt med parametre

$$\underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \text{ og } \underline{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_1' \\ \sigma_1 & \Sigma_{22} \end{bmatrix} .$$

Da viser sætning 2.17, at

$$g(\underline{x}) = E(Y|X = \underline{x}) = \mu_1 + \sigma_1' \Sigma_{22}^{-1} (\underline{x} - \underline{\mu}_2) ,$$

d.v.s. regressionen er lineær (affin).

Vi specialiserer nu til 2 dimensioner.

Lad  $\begin{pmatrix} Y \\ X \end{pmatrix}$  være normalt fordelt med parametre

$$\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \text{ og } \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} .$$

Da er regressionen af Y på X givet ved

$$E(Y|X = x) = \mu_y + \rho \frac{\sigma_y}{\sigma_x} (x - \mu_x) ,$$

og regressionen af X på Y ved

$$E(X|Y = y) = \mu_x + \rho \frac{\sigma_x}{\sigma_y} (y - \mu_y) .$$

Lad os antage, at der foreligger målinger  $\begin{bmatrix} X_1 \\ Y_1 \end{bmatrix}, \dots, \begin{bmatrix} X_n \\ Y_n \end{bmatrix}$ .

Maksimum likelihood estimatorer for hældningskoefficienterne fås da ved blot at indsætte maksimum likelihood estimatorer for de indgående parametre. Da

$$\hat{\rho} = \frac{\Sigma (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\Sigma (X_i - \bar{X})^2 \Sigma (Y_i - \bar{Y})^2}} = \frac{SAP_{xy}}{\sqrt{SAK_x SAK_y}} ,$$

$$\hat{\sigma}_x^2 = \frac{1}{n} \Sigma (X_i - \bar{X})^2 ,$$

$$\hat{\sigma}_y^2 = \frac{1}{n} \Sigma (Y_i - \bar{Y})^2 ,$$

ses, at f.eks. skønnet over hældningskoefficienten i udtrykket for regressionen af Y på X bliver

$$\hat{\rho} \frac{\hat{\sigma}_y}{\hat{\sigma}_x} = \frac{SAP_{xy}}{SAK_x} .$$

Dette giver den empiriske regressionsligning

$$\hat{E}(Y|X = x) = \bar{Y} + \frac{SAP_{xy}}{SAK_x} (x - \bar{X}),$$

d.v.s. præcis samme resultat, som vi opnåede i den endimensionale lineære regressionsanalyse, jvf. p. 5.47 i bind 1. Her var forudsætningerne dog ganske anderledes, idet det blev antaget, at værdierne af den uafhængige variable (her  $X$ , i bind 1 t) var deterministiske værdier, hvor vi i denne fremstilling regner med, at de er realiserede udfald af en normalt fordelt variabel, der er korreleret med den afhængige variabel. Hvad angår estimationen er det altså underordnet, hvilken af de to modeller man arbejder med; men tolkningen af de færdige resultater afhænger selvsagt heraf.

Vi fortsætter nu betragtningerne i eksempel 2.8.

Eksempel 2.8 Vi vil i dette eksempel bestemme de lineære overgange fra en måling ved den ene af de to i eksempel 2.2 anførte målemetoder til den anden.

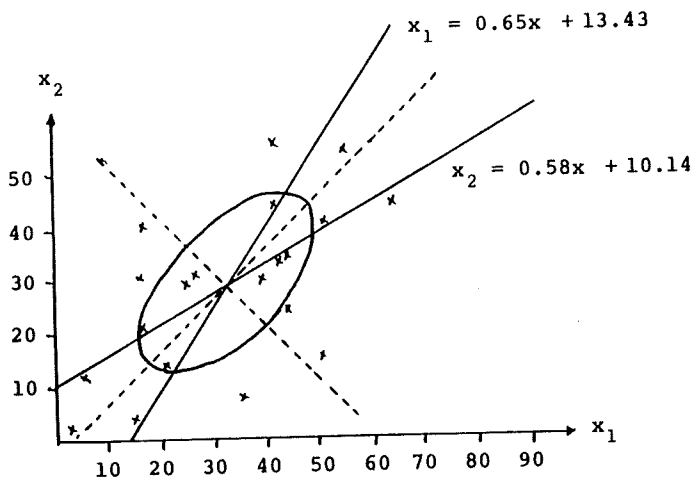
Vi finder regressionerne

$$\begin{aligned}\hat{E}(X_1|X_2 = x_2) &= \bar{x}_1 + \hat{\rho} \frac{s_1}{s_2} (x_2 - \bar{x}_2) \\ &= 0.65x_2 + 13.43\end{aligned}$$

og

$$\begin{aligned}\hat{E}(X_2|X_1 = x_1) &= \bar{x}_2 + \hat{\rho} \frac{s_2}{s_1} (x_1 - \bar{x}_1) \\ &= 0.58x_1 + 10.14.\end{aligned}$$

Disse linier er angivet i nedenstående figur.



Hvis vi vil undersøge, om der overhovedet er en sammenhæng mellem  $X_1$  og  $X_2$  kan vi undersøge korrelationskoefficienten. Den er fundet til

$$\hat{\rho} = \frac{182}{\sqrt{311 \cdot 279}} = 0.617 ,$$

d.v.s.

$$\hat{\rho}^2 = 0.380 .$$

Teststørrelsen for test af hypotesen  $\rho = 0$  er, jvf. p. 2.42, med  $p = m = 2$

$$t = \frac{0.617}{\sqrt{1-0.380}} \sqrt{20-2} = 3.32 > t(18)_{99.5} .$$

Ved test på niveau  $\alpha > 1\%$  vil hypotesen derfor blive forkastet, og vi vil antage, at  $\rho \neq 0$ , d.v.s. at der er en lineær sammenhæng mellem måleresultater ved de to metoder, og den er estimeret ved de to regressioner. Man kan så på sædvanlig vis finde skøn over usikkerheder m.v.

I figuren er også indtegnet en konturellipse og dens hovedakser. Det kan vises, at første hovedakse er den linie, man får frem ved at minimalisere punkternes ortogonale afstandes kvadratsum, hvor regressionsligningerne jo findes ved at minimalisere de lodrette henholdsvis vandrette afstande. Første hovedakse kaldes derfor også den ortogonale regression. Dette begreb vender vi tilbage til i kapitel 4.

□

#### 2.4 Spaltningsætningen

I dette afsnit betragter vi en stokastisk variabel  $\underline{X} \in N(\underline{\mu}, \underline{\Sigma})$ , hvor  $\underline{\Sigma}$  er regulær af orden  $n$ . Vi betragter det indre produkt defineret ved  $\underline{\Sigma}^{-1}$  og den derudfra affødte norm, nemlig

$$(\underline{x} | \underline{y}) = \underline{x}' \underline{\Sigma}^{-1} \underline{y}$$

og

$$\| \underline{x} \| = \sqrt{(\underline{x} | \underline{x})} = \sqrt{\underline{x}' \underline{\Sigma}^{-1} \underline{x}}$$

Lad nu underrummene  $U_1, \dots, U_k$  være ortogonale (m.h.t. dette indre produkt), således at

$$R = U_1 \oplus \dots \oplus U_k .$$

Vi sætter  $\dim U_i = n_i$  og betegner projektionen på  $U_i$  med  $p_i$ . Den tilsvarende matrix benævnes  $\underline{C}_i$ .

Lad betegnelserne være som ovenfor. Da gælder

Sætning 2.24 (Spaltningsætningen) Sættes

$$Y_i = p_i (\underline{X} - \underline{\mu}), \quad i = 1, \dots, k$$

og

$$K_i = \| Y_i \|^2 = \| p_i (\underline{X} - \underline{\mu}) \|^2, \quad i = 1, \dots, k,$$

da gælder, at

$$\underline{X} - \underline{\mu} = \sum_{i=1}^k Y_i$$

og

$$\| \underline{X} - \underline{\mu} \|^2 = \sum_{i=1}^k K_i.$$

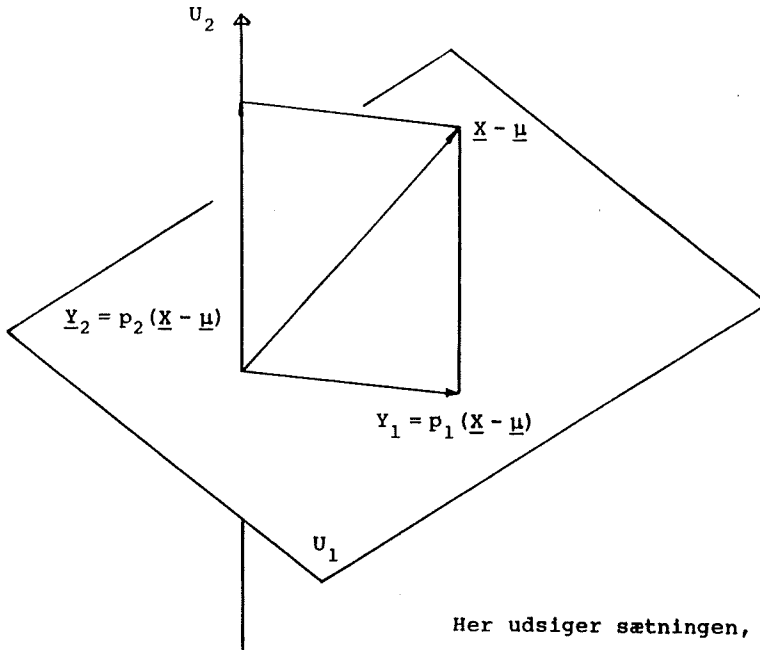
Endvidere er  $Y_1, \dots, Y_k$  stokastisk uafhængige og normalt fordelte, og  $K_1, \dots, K_k$  er stokastisk uafhængige  $\chi^2(n_i)$ -fordelte variable.

Bevis Vi har, at  $Y_i = \underline{C}_i (\underline{X} - \underline{\mu})$ , hvorfor

$$\underline{Y} = \begin{bmatrix} Y_1 \\ \cdot \\ \cdot \\ Y_k \end{bmatrix} = \begin{bmatrix} \underline{C}_1 \\ \cdot \\ \cdot \\ \underline{C}_k \end{bmatrix} (\underline{X} - \underline{\mu}).$$

Heraf fås

$$D(\underline{Y}) = \begin{bmatrix} \underline{C}_1 \\ \cdot \\ \cdot \\ \underline{C}_k \end{bmatrix} \cdot \Sigma \cdot (\underline{C}_1', \dots, \underline{C}_k') = (\underline{C}_i \Sigma \underline{C}_j')_{(i,j)}.$$



Her udsiger sætningen, at  $Y_1$  og  $Y_2$  er uafhængige, og at  $\|Y_1\|^2 \in \chi^2(2)$  og  $\|Y_2\|^2 \in \chi^2(1)$ .

Nu er for  $i \neq j$  ifølge lemma p. 1.68

$$\underline{C}_i \underline{\Sigma} \underline{C}_j' = \underline{0} .$$

Heraf følger at  $\underline{Y}$ 's komponenter er stokastisk uafhængige (fordi  $\underline{Y}$  er normalt fordelt).

Vi skal dernæst bestemme fordelingen af  $\|p_i(\underline{X} - \underline{\mu})\|^2$ . Vi har, at  $\underline{X}$  kan skrives

$$\underline{X} = \underline{\mu} + \underline{AZ} ,$$

hvor  $\underline{Z} \in N(0, \underline{\Sigma})$  og  $\underline{A}\underline{A}' = \underline{\Sigma}$ . Heraf følger

$$\begin{aligned} \|\underline{P}_i(\underline{X} - \underline{\mu})\|^2 &= \|\underline{P}_i(\underline{A}\underline{Z})\|^2 = \|\underline{C}_i\underline{A}\underline{Z}\|^2 \\ &= \underline{Z}'\underline{A}'\underline{C}_i'\underline{\Sigma}^{-1}\underline{C}_i\underline{A}\underline{Z} = \underline{Z}'\underline{D}_i\underline{Z} . \end{aligned}$$

Nu er

$$\begin{aligned} \underline{D}_i\underline{D}_i &= \underline{A}'\underline{C}_i'\underline{\Sigma}^{-1}\underline{C}_i\underline{A}\underline{A}'\underline{C}_i'\underline{\Sigma}^{-1}\underline{C}_i\underline{A} \\ &= \underline{A}'\underline{C}_i'\underline{C}_i'\underline{\Sigma}^{-1}\underline{\Sigma}\underline{C}_i'\underline{\Sigma}^{-1}\underline{C}_i\underline{A} \\ &= \underline{A}'\underline{C}_i'\underline{\Sigma}^{-1}\underline{C}_i\underline{A} \\ &= \underline{D}_i , \end{aligned}$$

d.v.s.  $\underline{D}_i$  er idempotent. Vi har i ovenstående gentagne gange benyttet lemma p. 1.68. Det er åbenbart, at  $\text{rg}(\underline{D}_i) = n_i$ . Da

$$\begin{aligned} \underline{D}_i &= \underline{A}'\underline{C}_i'\underline{A}^{-1}\underline{A}^{-1}\underline{C}_i\underline{A} \\ &= (\underline{A}^{-1}\underline{C}_i\underline{A})'(\underline{A}^{-1}\underline{C}_i\underline{A}) , \end{aligned}$$

er  $\underline{D}_i$  positiv semidefinit (jvf. sætning 1. p. 1. ). Der eksisterer derfor en ortogonal matrix  $\underline{P}'$  (sætning 1. ) så

$$\underline{P}'\underline{D}_i\underline{P} = \underline{\Delta}_i \text{ eller } \underline{D}_i = \underline{P}\underline{\Delta}_i\underline{P}' ,$$

hvor  $\underline{\Delta}_i$  er en diagonalmatrix med rang  $n_i$ . Da  $\underline{D}_i$  er idempotent, fås

$$\underline{P}\underline{\Delta}_i\underline{P}' = \underline{P}\underline{\Delta}_i\underline{P}'\underline{P}\underline{\Delta}_i\underline{P}' = \underline{P}\underline{\Delta}_i^2\underline{P}' ,$$

eller  $\underline{\Delta}_i = \underline{\Delta}_i^2$ . Derfor har  $\underline{\Delta}_i$   $n_i$  1-taller og  $n-n_i$  0'ler i diagonalen. Derfor er



$$\begin{aligned}\underline{Z}' \underline{D}_{\underline{I}} \underline{Z} &= \underline{Z}' \underline{P} \underline{\Lambda} \underline{P}' \underline{Z} = (\underline{P}' \underline{Z})' \underline{\Lambda} (\underline{P}' \underline{Z}) \\ &= \underline{V}' \underline{\Lambda} \underline{V} \\ &= \underbrace{V_1^2 + \dots + V_n^2}_{n_1 \text{ led} \neq 0}.\end{aligned}$$

Da  $\underline{V} \in N(\underline{0}, \underline{P}' \underline{P}) = N(\underline{0}, \underline{I})$  ses det, at

$$\underline{Z}' \underline{D}_{\underline{I}} \underline{Z} = \| P_1(\underline{X} - \underline{\mu}) \|^2 \in \chi^2(n_1).$$

Q.E.D.

**Eksempel 2.8** Lad  $X_1, \dots, X_n$  være uafhængige  $N(\mu, \sigma^2)$  - fordelte. Da er

$$\underline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \in N(\underline{\mu}, \sigma^2 \underline{I}).$$

Vi betragter underrummet  $U_1$  givet ved

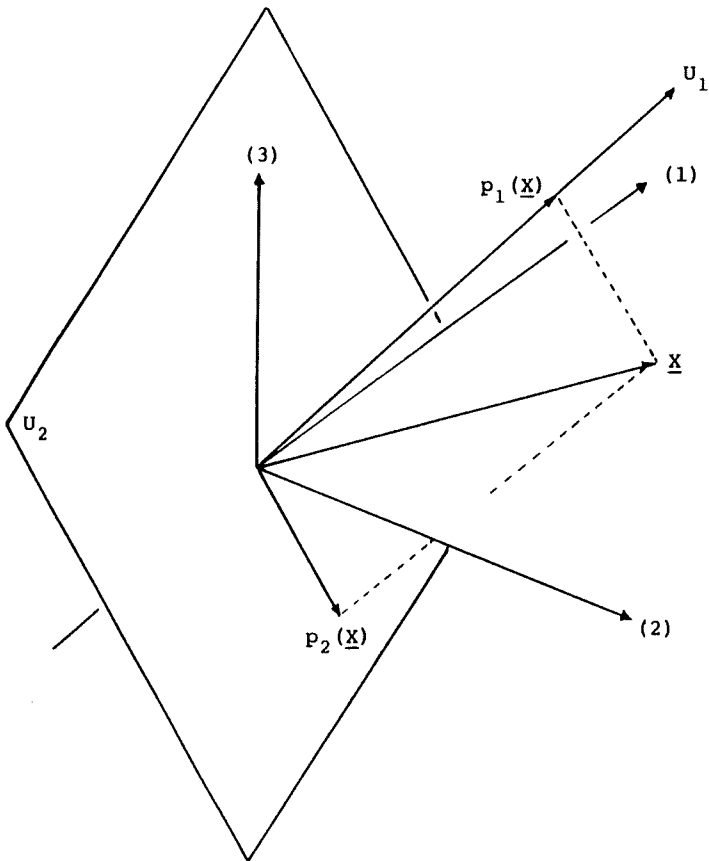
$$\underline{x} \in U_1 \leftrightarrow x_1 = \dots = x_n,$$

og det på  $U_1$  vinkelrette (m.h.t.  $\sigma^2 \underline{I}$ )  $U_2$ . (dette ortogonalitetsbegreb falder sammen med det sædvanlige). Nu viser identiteten

$$\Sigma (x_i - y)^2 = \Sigma (x_i - \bar{x})^2 + n(\bar{x} - y)^2,$$

at projektionen på  $U_1$  er givet ved

$$P_1(\underline{x}) = \begin{bmatrix} \bar{x} \\ \vdots \\ \bar{x} \end{bmatrix},$$



hvorfor

$$p_2(\underline{x}) = \underline{x} - p_1(\underline{x}) = \begin{bmatrix} x_1 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{bmatrix} .$$

Da  $\dim U_1 = 1$  og  $\dim U_2 = n-1$  får vi af spaltningsætningen, at

$$p_1(\underline{x} - \underline{\mu}) \text{ og } \| p_2(\underline{x} - \underline{\mu}) \|^2$$

er stokastisk uafhængige.  $p_1(\underline{X} - \underline{\mu})$  er normalt fordelt og  $\|p_2(\underline{X} - \underline{\mu})\|^2$  er  $\chi^2(n-1)$ -fordelt. Da

$$p_1(\underline{X} - \underline{\mu}) = \begin{bmatrix} \bar{X} - \mu \\ \vdots \\ \bar{X} - \mu \end{bmatrix},$$

og

$$\|p_2(\underline{X} - \underline{\mu})\|^2 = \frac{1}{\sigma^2} \sum_1 (X_i - \bar{X})^2,$$

genfinder vi her et resultat om fordelingen af  $\bar{X}$  og  $(n-1)S^2 = \frac{1}{\sigma^2} \sum_1 (X_i - \bar{X})^2$ .

□

## 2.5 Wishart fordelingen og den generaliserede varians

I det endimensionale tilfælde udledes en række stikprøvefordelinger fra det normale tilfælde. Den vigtigste af disse er  $\chi^2$ -fordelingen, der svarer til summen af kvadrerede normalt fordelte variable. Den flerdimensionale analog er Wishart-fordelingen.

Vi giver definitionen ved hjælp af tætheden i

**Definition 2.3** Lad  $\underline{V}$  være en kontinuert fordelt stokastisk  $p \times p$ -matrix, der er symmetrisk og positivt semidefinit med sandsynlighed 1. Da siges  $\underline{V}$  at følge en Wishart fordeling med parametre  $(n, \underline{\Sigma})$ , ( $n \geq p$ ), hvis tætheden for  $\underline{V}$  er

$$f(\underline{v}) = c \cdot [\det(\underline{v})]^{1/2(n-p-1)} \exp(-\frac{1}{2} \text{tr}(\underline{v} \cdot \underline{\Sigma}^{-1})),$$

for  $\underline{V}$  positivt definit og 0 ellers. Her betegner  $\underline{\Sigma}$  en positivt definit  $p \times p$ -matrix, og  $c$  er konstanten givet ved

$$\frac{1}{c} = 2^{\frac{1}{2}np} n^{p(p-1)/4} (\det \underline{\Sigma})^{\frac{1}{2}n} \prod_{i=1}^p \Gamma(\frac{1}{2}(n+1-i)) .$$

Vi skriver kort

$$\underline{V} \in W(n, \underline{\Sigma}) = W_p(n, \underline{\Sigma}) ,$$

hvor den sidste betegnelse anvendes, hvis der kan opstå tvivl om dimensionen.

Vi anfører nu først en bemærkning om middelværdi og varians af komponenterne i en Wishart fordeling

Lad  $\underline{V} = (V_{ij})$  være Wishart fordelt  $W(n, \underline{\Sigma})$ , hvor  $\underline{\Sigma} = (\sigma_{ij})$ . Da gælder

$$\begin{aligned} E(V_{ij}) &= n\sigma_{ij} \\ V(V_{ij}) &= n(\sigma_{ij}^2 + \sigma_{ii}\sigma_{jj}) \\ \text{Cov}(V_{ij}, V_{kl}) &= n(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) . \end{aligned}$$

Bevis Forbigås.

Analogien med  $\chi^2$ -fordelingen fremgår af

Sætning 2.25 Lad  $\underline{X}_i \in N_p(0, \underline{\Sigma})$ ,  $i = 1, \dots, n$ , være uafhængige og regulært fordelte. Da gælder for  $n \geq p$

$$\underline{V} = \sum_{i=1}^n \underline{X}_i \underline{X}_i' \in W(n, \underline{\Sigma}) .$$

Bevis Forbigås.

Bemærkning Hvis  $n \leq p$  har  $\underline{V}$  som defineret i sætningen ingen tæthedsfunktion. Vi vælger dog alligevel at sige, at  $\underline{V}$  er Wishart fordelt med parametre  $(n, \underline{\Sigma})$ .

Tilsvarende bemærkninger gør sig gældende hvis  $\underline{\Sigma}$  er singular. Med denne vedtagelse gælder sætningen uden restriktionen  $n \geq p$ .

En næsten triviell følge af ovenstående er nu

Sætning 2.26 Lad  $\underline{V}_1, \dots, \underline{V}_k$  være uafhængige stokastiske  $p \times p$ -matricer, der er  $W(n_1, \underline{\Sigma})$ -fordelte. Da gælder

$$\underline{V} = \underline{V}_1 + \dots + \underline{V}_k \in W(n_1 + \dots + n_k, \underline{\Sigma}) .$$

En af hovedsætningerne i teorien for stikprøvefunktioner af normalt fordelte stokastiske variable er, at  $\bar{X}$  og  $S^2$  er uafhængige og at  $S^2$  er  $\sigma^2 \chi^2/f$ -fordelt med 1 frihedsgrad mindre end der er observationer. Denne sætning har sin flerdimensionale analog i

Sætning 2.27 Lad  $\underline{X}_i \in N_p(\underline{\mu}, \underline{\Sigma})$ ,  $i = 1, \dots, n$ , være stokastisk uafhængige. Vi sætter

$$\bar{\underline{X}} = \frac{1}{n} \sum_{i=1}^n \underline{X}_i ,$$

$$\underline{S} = \frac{1}{n-1} \sum_{i=1}^n (\underline{X}_i - \bar{\underline{X}})(\underline{X}_i - \bar{\underline{X}})' .$$

Da er

$$\bar{\underline{X}} \in N_p\left(\underline{\mu}, \frac{1}{n} \underline{\Sigma}\right)$$

og

$$\underline{S} \in W(n-1, \frac{1}{n-1} \underline{\Sigma}) .$$

Endvidere er  $\bar{\underline{X}}$  og  $\underline{S}$  stokastisk uafhængige.

Bevis Forbigås.

Vi går nu over til at betragte nogle resultater vedrørende marginale fordelinger. Vi har

Sætning 2.28 Lad  $\underline{V}$  være Wishart fordelt med parametre  $(n, \underline{\Sigma})$ . Vi betragter spaltningen

$$\underline{V} = \begin{bmatrix} \underline{V}_{11} & \underline{V}_{12} \\ \underline{V}_{21} & \underline{V}_{22} \end{bmatrix} \text{ og } \underline{\Sigma} = \begin{bmatrix} \underline{\Sigma}_{11} & \underline{\Sigma}_{12} \\ \underline{\Sigma}_{21} & \underline{\Sigma}_{22} \end{bmatrix} .$$

Det gælder da, at

$$\underline{V}_{ii} \in W(n, \underline{\Sigma}_{ii}) .$$

Endvidere gælder

Sætning 2.29 Vi betragter igen ovenstående situation. Hvis  $\underline{\Sigma}_{12}$  og  $\underline{\Sigma}_{21}$  er  $\underline{0}$ -matricer, da er  $\underline{V}_{11}$  og  $\underline{V}_{22}$  stokastisk uafhængige.

Bevis for sætningerne. De følger umiddelbart ved at betragte de tilsvarende spaltninger af normalt fordelte vektorer, som frembringer Wishart fordelingerne. □

Da den flerdimensionale normalfordeling kan defineres på en koordinatfri måde, i.e. uafhængigt af et givent koordinatsystem, er det ikke overraskende, at noget tilsvarende gør sig gældende for Wishart-fordelingen. Da skift fra koordinater i et koordinatsystem til koordinater i andet foregår ved hjælp af matrixmultiplikationer, har vi følgende

Sætning 2.30 Lad  $\underline{V} \in W_p(n, \underline{\Sigma})$  og lad  $\underline{A}$  være en vilkårlig, fast  $r \times p$ -matrix. Da er

$$\underline{A}\underline{V}\underline{A}' \in W_r(n, \underline{A}\underline{\Sigma}\underline{A}') .$$

Bevis Som antydnet ovenfor skal man blot betragte de normalt fordelte vektorer, der frembringer  $V$  og transformere dem. Resultatet følger da umiddelbart.

□

Vi afslutter nu kapitlet med at indføre en anden generalisering af den endimensionale varians til det flerdimensionale tilfælde end dispersionsmatricen.

Definition 2.4 Lad den  $p$ -dimensionale vektor  $\underline{X}$  have dispersionsmatricen  $\underline{\Sigma}$ . Ved den generaliserede varians af  $\underline{X}$  forstås determinanten af dispersionsmatricen, i.e.

$$\text{gen.var.}(\underline{X}) = \det(\underline{\Sigma}) .$$

Bemærkning I afsnit 1.2.6 godtgjordes, at determinanten af en matrix svarer til volumenforholdet ved den tilsvarende lineære afbildning, d.v.s. den er et intuitivt rimeligt mål for "størrelsen" af en matrix.

Hvis der foreligger observationer  $\underline{X}_1, \dots, \underline{X}_n$ , defineres den empiriske generaliserede varians på åbenbar måde ud fra den empiriske dispersionsmatrix

$$\underline{\hat{\Sigma}} = \frac{1}{n-1} \sum_{i=1}^n (\underline{X}_i - \bar{\underline{X}})(\underline{X}_i - \bar{\underline{X}})' ,$$

nemlig som dennes determinant.

I det normale tilfælde kan vi finde fordelingen af den empiriske generaliserede varians, idet vi har

Sætning 2.31 Lad  $\underline{X}_i \in N_p(\underline{\mu}, \underline{\Sigma})$ ,  $i = 1, \dots, n$ , være stokastisk uafhængige. Da følger den empiriske generaliserede varians samme fordeling som

$$\frac{\det \underline{\underline{S}}}{(n-1)^p} \cdot z_1 \dots z_p ,$$

hvor  $z_1, \dots, z_p$  er stokastisk uafhængige og  $z_i \in \chi^2(n-1)$ .

Bevis Forbigås.

For  $p = 1$  og  $2$  er det muligt at finde tætheden af den empiriske generaliserede varians. For større værdier af  $p$  involverer denne tæthed dog integraler, som det ikke umiddelbart er muligt at udtrykke ved kendte funktioner. For  $n \rightarrow \infty$  har vi dog

Sætning 2.33 Lad  $\underline{\underline{S}}$  være som ovenfor (i det normale tilfælde). Da gælder, at

$$\sqrt{n-1} \left( \frac{\det(\underline{\underline{S}})}{\det(\underline{\underline{\Sigma}})} - 1 \right) \text{ asymptotisk } \in N(0, 2p) .$$

Bevis Forbigås.

## 2.6 Lidt om estimation af flerdimensionale parametre

I dette afsnit skal vi anføre udvidelser til tilfældet med flerdimensionale parametre af de i bind I, kapitel 2, anførte resultater om nedre grænser for variansen af en central estimator (Cramer-Rao's ulighed) og af sætningen om den asymptotiske normalitet af maximum likelihood estimatorer.

Sætningerne synes teknisk meget komplicerede, men kvintessensen af de regulatitetsbetingelser, der anføres, er, at differentiation med hensyn til parametrene kan ombyttes med integration efter  $\underline{x}$ 'erne.



Vi må først indføre begrebet informationsmatricen for en flerdimensional parameter.

**Definition 2.5** Lad  $X_1, \dots, X_n$  være indbyrdes uafhængige, identisk fordelte stokastiske variable med frekvensfunktion  $f(\underline{x}, \underline{\theta})$ , hvor  $\underline{\theta} \in \Omega \subseteq R^k$ . Ved informationsmatricen  $\underline{i}(\underline{\theta})$  forstås da matricen med  $(i, j)$ 'te element

$$\underline{i}(\underline{\theta})_{i,j} = E_{\underline{\theta}} \left\{ \frac{\partial \log f(\underline{X}_1, \underline{\theta})}{\partial \theta_i} \frac{\partial \log f(\underline{X}_1, \underline{\theta})}{\partial \theta_j} \right\} .$$

**Lemma** Under passende regularitetsbetingelser, fx. de i efterfølgende sætninger anførte, gælder

$$\underline{i}(\underline{\theta})_{i,j} = -E_{\underline{\theta}} \left\{ \frac{\partial^2 \log f(\underline{X}_1, \underline{\theta})}{\partial \theta_i \partial \theta_j} \right\} .$$

**Bevis** Forbigås.

**Sætning 2.34 (Cramer-Rao's ulighed).** Lad  $X_1, \dots, X_n$  være indbyrdes uafhængige, identisk fordelte stokastiske variable med frekvensfunktion  $f(\underline{x}, \underline{\theta})$ , hvor  $\underline{\theta}$  er en ukendt parameter,  $\underline{\theta} \in \Omega \subseteq R^k$ . Lad endvidere for alle  $i, j$  og for alle  $\underline{\theta} \in \overset{\circ}{\Omega}$  (det indre af  $\Omega$ )

$$1) \quad \frac{\partial f(\underline{x}, \underline{\theta})}{\partial \theta_i} \text{ eksistere for alle } \underline{x} ,$$

$$2) \quad E_{\underline{\theta}} \left\{ \frac{\partial \log f(\underline{X}_1, \underline{\theta})}{\partial \theta_i} \right\} = 0 ,$$

$$3) \quad E_{\underline{\theta}} \left\{ \left( \frac{\partial \log f(\underline{X}_1, \underline{\theta})}{\partial \theta_i} \right)^2 \right\} < \infty ,$$

$$4) \quad \det \underline{i}(\underline{\theta}) \neq 0 .$$

Lad endvidere  $\underline{d}(\underline{X}_1, \dots, \underline{X}_n) = (d_1, \dots, d_k)'$  være et centralt skøn over  $\underline{\theta}$ , der tilfredsstiller

$$5) \quad E_{\underline{\theta}} \left\{ d_i \cdot \frac{\partial}{\partial \theta_j} \log \prod_v f(\underline{X}_v, \underline{\theta}) \right\} = \delta_{ij} ,$$

$$6) \quad E_{\underline{\theta}} \{ d_i^2 \} < \infty .$$

Da er

$$D(\underline{d}(\underline{X}_1, \dots, \underline{X}_n)) \geq \frac{1}{n} \underline{I}^{-1}(\underline{\theta}) ,$$

d.v.s. matricen

$$D(\underline{d}(\underline{X}_1, \dots, \underline{X}_n)) - \frac{1}{n} \underline{I}^{-1}(\underline{\theta})$$

er positiv semidefinit.

**Bevis** Forbigås, se fx. Witting & Nölle (1970).

**Bemærkning 1** Som allerede anført indledningsvis, er de anførte betingelser ikke synderligt restriktive. Hvis man antager, at fordelingen  $f$  har en støtte, der er uafhængig af  $\underline{\theta}$ , sikrer 1)-3) og 5)-6) blot, at differentiation med hensyn til  $\underline{\theta}$  og integration med hensyn til  $\underline{x}$  kan ombyttes. Eksempelvis har vi i det kontinuerte tilfælde med  $\Pi f(\underline{X}_v; \underline{\theta}) = p(\underline{X}_1, \dots, \underline{X}_n; \underline{\theta}) = p$

$$\begin{aligned} E_{\underline{\theta}} \left\{ d_i \frac{\partial \log p}{\partial \theta_j} \right\} &= \int d_i \frac{\partial \log p}{\partial \theta_j} p \, d\underline{z} \\ &= \int d_i \frac{1}{p} \frac{\partial p}{\partial \theta_j} p \, d\underline{z} \\ &= \frac{\partial}{\partial \theta_j} \int d_i p \, d\underline{z} \end{aligned}$$

$$= \frac{\partial}{\partial \theta_j} \theta_i$$

$$= \delta_{ij} ,$$

hvor  $\delta_{ij}$  altså er Kroneckers  $\delta$  ( $= 0$  for  $i \neq j$ , og  $= 1$  for  $i = j$ ). Integrationen foregår med hensyn til  $\underline{z}$  givet ved  $\underline{z}' = (\underline{x}'_1, \dots, \underline{x}'_n)$ .

**Bemærkning 2** Ud fra sætningen kan også formuleres resultater om en nedre grænse for den generaliserede varians af  $\underline{d}$  etc.

Vi anfører dernæst resultatet om den asymptotiske fordeling af maximum likelihood estimatorer.

**Sætning 2.35** Lad  $X_1, \dots, X_n$  være indbyrdes uafhængige, identisk fordelte stokastiske variable med frekvensfunktion  $f(\underline{x}, \underline{\theta})$ ,  $\underline{\theta} \in \Omega \subseteq \mathbb{R}^k$ . Lad endvidere (for  $\underline{\theta} \in \hat{\Omega}$  og for alle  $i, j$ )

$$1) \quad \underline{\theta}_1 \neq \underline{\theta}_2 \Rightarrow f(\cdot, \underline{\theta}_1) \neq f(\cdot, \underline{\theta}_2),$$

$$2) \quad \frac{\partial^2 f(\underline{x}, \underline{\theta})}{\partial \theta_i \partial \theta_j} \text{ eksistere og være kontinuert,}$$

$$3) \quad E_{\underline{\theta}} \left\{ \frac{1}{f(X_1, \underline{\theta})} \frac{\partial f(X_1, \underline{\theta})}{\partial \theta_i} \right\} = 0 ,$$

$$4) \quad E_{\underline{\theta}} \left\{ \frac{1}{f(X_1, \underline{\theta})} \frac{\partial^2 f(X_1, \underline{\theta})}{\partial \theta_i \partial \theta_j} \right\} = 0 ,$$

5) eksistere en omegn  $U$  omkring  $\underline{\theta}$  og en funktion  $M(\underline{x}, \underline{\theta})$  med  $E_{\underline{\theta}}\{M(X_1, \underline{\theta})\} < \infty$ , så

$$\left| \frac{\partial^2 \log f(\underline{x}, \underline{\theta})}{\partial \theta_i \partial \theta_j} \right|_{\underline{\theta}=\underline{\theta}^*} \leq M(\underline{x}, \underline{\theta}) , \quad \forall \underline{\theta}^* \in U ,$$

6)  $\det \underline{j}(\underline{\theta}) \neq 0$ .

Kaldes  $\underline{\theta}$ 's maximum likelihood estimator  $\hat{\underline{\theta}}_n$ , og er denne konsistent, da vil  $\sqrt{n}(\hat{\underline{\theta}}_n - \underline{\theta})$  være asymptotisk normalt fordelt med parametre  $(\underline{0}, \underline{j}^{-1}(\underline{\theta}))$ .

Bevis Forbigås, se fx. Witting & Nölle (1970).

Bemærkning 1 Der kan anføres betragtninger om betingelserne 1)-6), som er ganske analoge til de efter sætning 2.34 anførte.

Bemærkning 2 Det kræves i sætningen, at  $\hat{\underline{\theta}}_n$  er konsistent, d.v.s. konvergerer i sandsynlighed mod den sande parameter-værdi  $\underline{x}$ . En sådan betingelse er ofte ganske let at eftervise direkte. Man kan selvfølgelig også anføre yderligere stramme-de regularitetsbetingelser, som automatisk vil sikre denne konsistens. De svarer helt til betingelsen 5), men vedrører blot de tredje afledede i stedet for de anden, cf. Cox & Hinkley (1974).

Bemærkning 3 Sætningen bevarer også sin gyldighed med forholdsvis simple modifikationer (bl.a. af parametrene) i tilfældet med uafhængige, men ikke nødvendigvis identisk fordelte variable  $X_1, \dots, X_n$ .

#### Referencer til kapitel 2

Anderson, T. W.: An Introduction to Multivariate Statistical analysis. John Wiley & Sons, New York 1958.

Cox, D. R. & Hinkley, D. V.: Theoretical Statistics. Chapman and Hall, London (1974).

- Knudsen, Jan Gunder: En Statistisk Analyse af Cementstyrke.  
Eksamensprojekt, IMSOR 1975.
- Pedersen, Steen Tokkesdal & Skjøth, Poul: Statistisk Analyse  
af Data fra Cementfabrikation. Eksamensprojekt, IMSOR 1976.
- Rao, C. R.: Linear Statistical Inference and Its Applications.  
John Wiley & Sons, New York 1965.
- Witting, H. & G. Nölle: Angewandte Mathematische Statistik.  
B. G. Teubner, Stuttgart 1970.



## KAPITEL 3

Den generelle lineære model

I dette kapitel skal vi formulere en model, som naturligt generaliserer de fra bind 1 kendte varians- og regressionsanalyser. Sætninger og definitioner vil i vidt omfang blive tolket geometrisk for at give en mere intuitiv forståelse af sammenhængene.

3.1 Estimation i den generelle lineære model

Vi giver først en beskrivelse af modellen i

3.1.1 Modelformulering

Vi betragter en  $n$ -dimensional stokastisk variabel  $\underline{y} \in N(\underline{\mu}, \sigma^2 \underline{\Sigma})$ , hvor  $\underline{\Sigma}$  er kendt. Vi betragter normen givet ved  $\underline{\Sigma}^{-1}$ , i.e.

$$\| \underline{x} \|^2 = \underline{x}' \underline{\Sigma}^{-1} \underline{x} .$$

Den ved den inverse dispersionsmatrix  $(\sigma^2 \underline{\Sigma})^{-1}$  definerede norm er givet ved

$$\| \underline{x} \|^2_{\sigma^2} = \frac{1}{\sigma^2} \underline{x}' \underline{\Sigma}^{-1} \underline{x} = \frac{1}{\sigma^2} \| \underline{x} \|^2 .$$

De to normer er altså proportionale, og de giver derfor anledning til samme orthogonalitetsbegreb.

Vi vil nu betragte en række problemer i forbindelse med estimation og testning af middelværdien  $\underline{\mu}$  i tilfælde, hvor  $\underline{\mu}$  er en kendt lineær funktion af ukendte parametre, i.e.

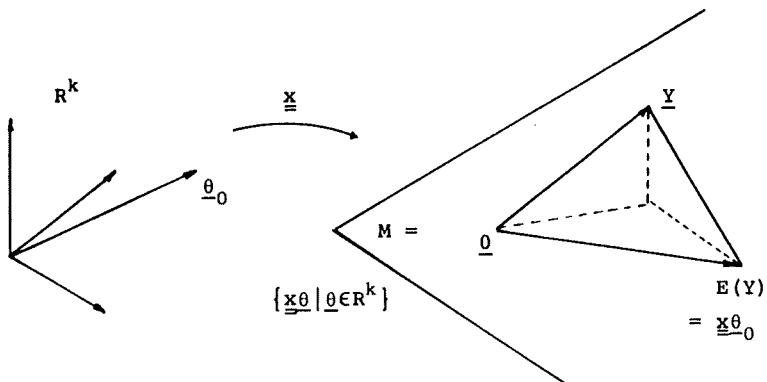
$$\underline{\mu} = \underline{x}\underline{\theta},$$

eller

$$\begin{bmatrix} \mu_1 \\ \cdot \\ \cdot \\ \cdot \\ \mu_n \end{bmatrix} = \begin{bmatrix} x_{11} & \dots & x_{1k} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_{n1} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \cdot \\ \cdot \\ \cdot \\ \theta_k \end{bmatrix},$$

hvor  $\underline{x}$  altså er kendt.

Geometrisk kan dette udtrykkes, at vi antager, at forventningsværdien af den stokastiske vektor  $\underline{Y}$  ligger i et under-rum  $M$  af  $R^n$ .  $M$  er billedet af  $R^k$  ved den til  $\underline{x}$  svarende lineære afbildning. Dimensionen af  $M$  er  $\text{rg}(\underline{x}) \leq k$ . Situationen er anskueliggjort i nedenstående figur.



Figur 3.1 Geometrisk skitse af generel lineær model.



Vi kalder en sådan model, hvor den ukendte middelværdi  $\underline{\mu}$  er en (kendt) lineær funktion af den ukendte parameter  $\underline{\theta}$  for en (generel) lineær model. Dette gælder også uden forudsætningen om, at  $\underline{Y}$  er normalt fordelt.

Eksempel 3.1 Vi betragter en sædvanlig endimensional regressionsanalysemodel, d.v.s. der foreligger observationer

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = 1, \dots, n,$$

hvor  $E(\varepsilon_i) = 0$ . Denne model kan skrives

$$\begin{bmatrix} Y_1 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix},$$

eller

$$\underline{Y} = \underline{X}\underline{\theta} + \underline{\varepsilon},$$

d.v.s. modellen er lineær i ovenstående betydning.

□

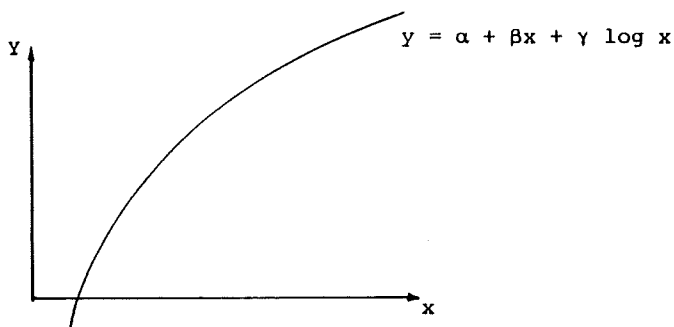
Et andet eksempel er

Eksempel 3.2 Vi betragter nu en situation, hvor

$$Y_i = \alpha + \beta x_i + \gamma \log x_i + \varepsilon_i, \quad i = 1, \dots, n$$

stadig med  $E(\varepsilon_i) = 0$ .

3.4



Også i denne situation får vi en lineær model frem, nemlig

$$\begin{bmatrix} Y_1 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & \log x_1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & x_n & \log x_n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_n \end{bmatrix} .$$

Vi bemærker, at vendingen lineær ikke går på, at  $E(Y|X) = \alpha + \beta x + \gamma \log x$  skal være lineær i den uafhængige variabel  $x$ , men at  $E(Y|x)$  opfattet som funktion af den ukendte parameter  $(\alpha, \beta, \gamma)$ ' skal være lineær. Havde vi haft en model som

$$Y_i = \alpha + \beta \log(\gamma x_i + \delta) + \epsilon_i ,$$

hvor  $\alpha, \beta, \gamma$  og  $\delta$  er de ukendte parametre, da ville det ikke være muligt at skrive

$$\underline{Y} = \underline{X} \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{bmatrix} + \underline{\epsilon}$$

med en kendt  $\underline{X}$ -matrix, og der vil derfor ikke være tale om en lineær model.

□

### 3.1.2 Estimation i det regulære tilfælde

Vi formulerer først resultatet vedrørende estimation af  $\underline{\theta}$  i

**Sætning 3.1** Lad  $\underline{x}$  og  $\underline{\theta}$  være som i foregående afsnit, og lad  $\underline{y} \in N_n(\underline{x}\underline{\theta}, \sigma^2\underline{\Sigma})$ , hvor  $\underline{\Sigma}$  er positivt definit. Da er maksimum likelihood estimatoren  $\hat{\underline{\theta}}$  for  $\underline{\theta}$  givet ved, at  $\underline{x}\hat{\underline{\theta}}$  er projektionen (m.h.t.  $\underline{\Sigma}$ ) ned på  $M$ , d.v.s.  $\hat{\underline{\theta}}$  er løsning til den såkaldte normalligning

$$(\underline{x}'\underline{\Sigma}^{-1}\underline{x})\hat{\underline{\theta}} = \underline{x}'\underline{\Sigma}^{-1}\underline{y} .$$

Hvis  $\underline{x}$  har fuld rang  $k$ , er

$$\hat{\underline{\theta}} = (\underline{x}'\underline{\Sigma}^{-1}\underline{x})^{-1}\underline{x}'\underline{\Sigma}^{-1}\underline{y} ,$$

og som linearkombination af normalt fordelte variable er  $\hat{\underline{\theta}}$  selv normalt fordelt med parametre

$$E(\hat{\underline{\theta}}) = \underline{\theta}$$

$$D(\hat{\underline{\theta}}) = \sigma^2(\underline{x}'\underline{\Sigma}^{-1}\underline{x})^{-1} .$$

Det bemærkes specielt, at  $\hat{\underline{\theta}}$  er central for  $\underline{\theta}$ .

**Bevis** Hvis  $\underline{y} \in N(\underline{x}\underline{\theta}, \sigma^2\underline{\Sigma})$ , hvor  $\underline{\Sigma}$  er regulær, er tætheden for  $\underline{y}$

$$f(\underline{y}) = \frac{1}{\sqrt{2\pi}^n} \frac{1}{\sigma^n} \frac{1}{\sqrt{\det\underline{\Sigma}}} \exp\left[-\frac{1}{2\sigma^2} (\underline{y} - \underline{x}\underline{\theta})'\underline{\Sigma}^{-1}(\underline{y} - \underline{x}\underline{\theta})\right]$$

$$= k \cdot \frac{1}{\sigma^n} \exp\left[-\frac{1}{2\sigma^2} \|\underline{y} - \underline{x}\underline{\theta}\|^2\right] .$$

Vi har likelihood-funktionen

$$L(\underline{\theta}) = k \cdot \frac{1}{\sigma^n} \exp\left[ -\frac{1}{2\sigma^2} \|\underline{Y} - \underline{x}\underline{\theta}\|^2 \right],$$

d.v.s.

$$\log L(\underline{\theta}) = k_1 - \frac{1}{2\sigma^2} \|\underline{Y} - \underline{x}\underline{\theta}\|^2.$$

Det er nu åbenbart, at maksimalisering af likelihoodfunktionen er ensbetydende med minimalisering af den kvadratiske afstand mellem et vilkårligt punkt i  $M$  og observationen, i.e. ensbetydende med minimalisering af

$$\|\underline{Y} - \underline{x}\underline{\theta}\|^2 = (\underline{Y} - \underline{x}\underline{\theta})' \underline{\Sigma}^{-1} (\underline{Y} - \underline{x}\underline{\theta}).$$

Ifølge resultatet p. 1.65 er den værdi af  $\underline{x}\underline{\theta}$ , som giver minimum, lig den ortogonale (m.h.t.  $\underline{\Sigma}^{-1}$ ) projektion af  $\underline{Y}$  ned på  $M$ . Ifølge eksempel 1.8 p. 1.60 er det optimale  $\underline{\theta}$  løsning til ligningen

$$(\underline{x}' \underline{\Sigma}^{-1} \underline{x}) \underline{\theta} = \underline{x}' \underline{\Sigma}^{-1} \underline{Y}.$$

Hvis  $\underline{x}' \underline{\Sigma}^{-1} \underline{x}$  har fuld rang  $k$ , i.e. hvis  $\underline{x}$  har rang  $k$  (jvf. p. 1.44) er derfor

$$\underline{\theta}_{\text{opt.}} = (\underline{x}' \underline{\Sigma}^{-1} \underline{x})^{-1} \underline{x}' \underline{\Sigma}^{-1} \underline{Y}.$$

Vi har nu vist den første halvdel af sætningen.

Af sætning 2.2 fås, at

$$E(\hat{\underline{\theta}}) = (\underline{x}' \underline{\Sigma}^{-1} \underline{x})^{-1} \underline{x}' \underline{\Sigma}^{-1} \underline{x} \underline{\theta} = \underline{\theta},$$

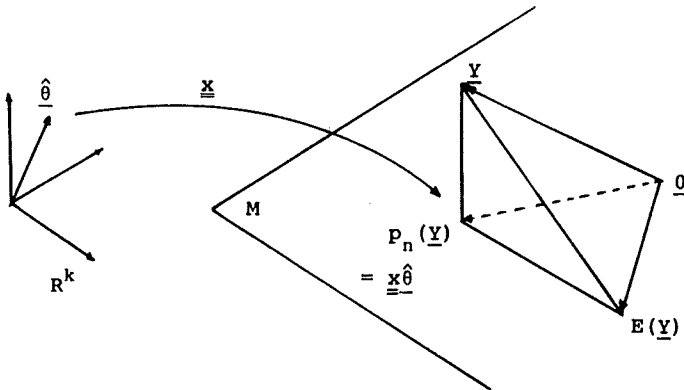
og ved hjælp af sætning 2.5 fås

$$D(\hat{\underline{\theta}}) = (\underline{x}' \underline{\Sigma}^{-1} \underline{x})^{-1} \underline{x}' \underline{\Sigma}^{-1} (\sigma^2 \underline{\Sigma}) \underline{\Sigma}^{-1} \underline{x} (\underline{x}' \underline{\Sigma}^{-1} \underline{x})^{-1}$$

$$= \sigma^2 (\underline{\mathbf{x}}' \underline{\Sigma}^{-1} \underline{\mathbf{x}})^{-1} ,$$

Q.E.D.

Vi har illustreret problematikken i nedenstående fig. 3.2



Figur 3.2 Geometrisk skitse af estimationsproblematikken i den generelle lineære model.

Bemærkning Vi bemærker, at  $\underline{\theta}$  er estimeret ved at minimalisere den kvadrerede afstand ned til  $M$ .  $\hat{\theta}$  er derfor også et mindste kvadraters skøn over  $\underline{\theta}$ . Hvis man ikke opretholder fordelingsforudsætningen vil man derfor alligevel ofte anvende den i sætning 3.1 anførte estimator  $\hat{\theta}$  som skøn over  $\underline{\theta}$ . Det kan iøvrigt vises, at mindste kvadraters skønnet  $\hat{\theta}$  har den mindste generaliserede varians blandt alle de estimators, der er lineære funktioner af observationerne (den såkaldte Gauss-Markov sætning), se f.eks. Kendall & Stuart (1967).

Da  $\sigma^2$  oftest er ukendt, vil vi nu udlede estimators for den. Vi har

Sætning 3.2 Lad situationen være som ovenfor. Maximum likelihood skønnet for  $\sigma^2$  er

$$\frac{1}{n} \| \underline{Y} - \underline{X}\hat{\underline{\theta}} \|^2 .$$

Det centrale skøn over  $\sigma^2$  er

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n - \text{rg}\underline{X}} \| \underline{Y} - \underline{X}\hat{\underline{\theta}} \|^2 \\ &= \frac{1}{n - \text{rg}\underline{X}} (\underline{Y} - \underline{X}\hat{\underline{\theta}})' \underline{\Sigma}^{-1} (\underline{Y} - \underline{X}\hat{\underline{\theta}}) \end{aligned}$$

hvor  $\hat{\underline{\theta}}$  er maksimum likelihood skønnet over  $E(\underline{Y})$ . Der gælder, at

$$\hat{\sigma}^2 \in \sigma^2 \chi^2(n - \text{rg}\underline{X}) / (n - \text{rg}\underline{X}) ,$$

og  $\hat{\sigma}^2$  er uafhængig af maksimum likelihood skønnet over middelværdien og dermed uafhængig af  $\hat{\underline{\theta}}$ .

Bevis Likelihoodfunktionen bliver

$$L(\underline{\theta}, \sigma^2) = k \cdot \frac{1}{\sigma^n} \exp\left[ -\frac{1}{2} \frac{1}{\sigma^2} \| \underline{Y} - \underline{X}\underline{\theta} \|^2 \right] ,$$

og

$$\log L(\underline{\theta}, \sigma^2) = k_1 - \frac{n}{2} \log \sigma^2 - \frac{1}{2} \frac{1}{\sigma^2} \| \underline{Y} - \underline{X}\underline{\theta} \|^2 .$$

Nu er

$$\begin{aligned} \frac{\partial}{\partial \sigma^2} \log L &= -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \| \underline{Y} - \underline{X}\underline{\theta} \|^2 \\ &= -\frac{n}{2} \frac{1}{\sigma^4} \left( \sigma^2 - \frac{1}{n} \| \underline{Y} - \underline{X}\underline{\theta} \|^2 \right) . \end{aligned}$$

Ved differentiation med hensyn til  $\underline{\theta}$  fås blot det sædvanlige normalligningssystem. Vi har derfor, at maksimum likelihood skønnene  $(\hat{\underline{\theta}}, \hat{\sigma}^2)$  for  $(\underline{\theta}, \sigma^2)$  er løsningen til

$$\underline{\mathbf{x}}' \underline{\Sigma}^{-1} \underline{\mathbf{x}} \hat{\underline{\theta}} = \underline{\mathbf{x}}' \underline{\Sigma}^{-1} \underline{\mathbf{y}}$$

$$\hat{\sigma}^2 = \frac{1}{n} \|\underline{\mathbf{y}} - \underline{\mathbf{x}} \hat{\underline{\theta}}\|^2 - \frac{1}{n} (\underline{\mathbf{y}} - \underline{\mathbf{x}} \hat{\underline{\theta}})' \underline{\Sigma}^{-1} (\underline{\mathbf{y}} - \underline{\mathbf{x}} \hat{\underline{\theta}}) .$$

Betragter vi opspaltningen af  $R^n$  som direkte sum af  $M$  og  $M^\perp$ , hvor  $M^\perp$  er det ortogonale (m.h.t.  $\underline{\Sigma}^{-1}$ ) komplement til  $M$ , fås, at

$$P_M(\underline{\mathbf{y}} - \underline{\mathbf{x}} \hat{\underline{\theta}}) = \underline{\mathbf{x}} \hat{\underline{\theta}} - \underline{\mathbf{x}} \underline{\theta}$$

og

$$\underline{\mathbf{y}} - \underline{\mathbf{x}} \hat{\underline{\theta}}$$

er stokastisk uafhængige, og at

$$\begin{aligned} \|\underline{\mathbf{y}} - \underline{\mathbf{x}} \hat{\underline{\theta}}\|^2 &\in \sigma^2 \chi^2(\dim M^\perp) \\ &= \sigma^2 \chi^2(n - \text{rg} \underline{\mathbf{x}}) . \end{aligned}$$

Heraf fås specielt

$$E(\hat{\sigma}^2) = \frac{1}{n} (n - \text{rg} \underline{\mathbf{x}}) \sigma^2 ,$$

d.v.s. maksimum likelihood estimatoren for  $\sigma^2$  er ikke central. Ønsker vi et centralt skøn, kan vi anvende

$$\frac{1}{n - \text{rg} \underline{\mathbf{x}}} \|\underline{\mathbf{y}} - \underline{\mathbf{x}} \hat{\underline{\theta}}\|^2 .$$

Da vi som oftest vil bruge det centrale skøn over  $\sigma^2$ , vil vi knytte betegnelsen  $\hat{\sigma}^2$  til dette.

Q.E.D.

Bemærkning Hvis  $\underline{\Sigma}$  specielt er enhedsmatricen, bliver  $\|\underline{y}\|^2 = \sum y_i^2$ . Derfor fås i dette tilfælde

$$\hat{\sigma}^2 = \frac{1}{n - \text{rg}\underline{\Sigma}} \sum_{i=1}^n (y_i - \hat{E}(y_i))^2 ,$$

hvor  $\hat{E}(y_i) = (\underline{x}\hat{\theta})_i$ . Størrelsen  $y_i - \hat{E}(y_i)$  er lig den  $i$ 'te observations afvigelse fra den estimerede model, og den kaldes det  $i$ 'te residual. Vi har altså i tilfældet  $\underline{\Sigma} = \underline{I}$ , at variansskønnet er proportionalt med summen af de kvadrerede residualer, som benævnes  $\text{SAK}_{\text{res}}$ . Vi vedtager, at vi vil anvende denne betegnelse generelt for den kvadrerede afstand mellem observationen og den estimerede model, d.v.s.

$$\text{SAK}_{\text{res}} = \|\underline{y} - \underline{x}\hat{\theta}\|^2 = (\underline{y} - \underline{x}\hat{\theta})' \underline{\Sigma}^{-1} (\underline{y} - \underline{x}\hat{\theta}) .$$

Inden vi går videre, giver vi et lille eksempel til illustration.

Eksempel 3.3 Ved fremstilling af et syntetisk produkt benyttes hovedsageligt to råmaterialer A og B. Kvaliteten af det færdige produkt kan beskrives ved en stokastisk variabel, der følger den normale fordeling med middelværdi  $\mu$  og varians  $\sigma^2$ . Det vides desuden, at middelværdien afhænger lineært af tilsætningen af A og B, d.v.s.

$$\mu = x_A \theta_A + x_B \theta_B ,$$

hvor  $x_A$  angiver den benyttede mængde af materiale A og  $x_B$  tilsvarende angiver den benyttede mængde af materiale B.  $\sigma^2$  forudsættes uafhængig af råvaretilsætningen.

Til bestemmelse af  $\theta_A$  og  $\theta_B$  udførtes nu 3 forsøg efter følgende plan



Forsøg	Indhold af A	Indhold af B
1	100%	0%
2	0%	100%
3	50%	50%

De enkelte forsøg antages stokastisk uafhængige. Den simultane fordeling af forsøgsresultaterne  $Y_1, Y_2, Y_3$  er da en tredimensionel normal fordeling med middelværdi

$$\underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \theta_A \\ \theta_B \end{bmatrix} = \underline{\underline{x}} \underline{\theta} ,$$

og dispersionsmatrix  $\sigma^2 I$ .

Vi har

$$\underline{\underline{x}}' \underline{\underline{x}} = \begin{bmatrix} \frac{5}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{5}{4} \end{bmatrix} \quad \text{og} \quad (\underline{\underline{x}}' \underline{\underline{x}})^{-1} = \begin{bmatrix} 5/6 & -1/6 \\ -1/6 & 5/6 \end{bmatrix} ,$$

og

$$\underline{\underline{x}}' \underline{y} = \begin{bmatrix} y_1 + \frac{1}{2}y_3 \\ y_2 + \frac{1}{2}y_3 \end{bmatrix} ,$$

hvorfor

$$\begin{bmatrix} \hat{\theta}_A \\ \hat{\theta}_B \end{bmatrix} = \begin{bmatrix} 5/6 & -1/6 \\ -1/6 & 5/6 \end{bmatrix} \begin{bmatrix} y_1 + \frac{1}{2}y_3 \\ y_2 + \frac{1}{2}y_3 \end{bmatrix} = \begin{bmatrix} \frac{5}{6}y_1 - \frac{1}{6}y_2 + \frac{1}{3}y_3 \\ -\frac{1}{6}y_1 + \frac{5}{6}y_2 + \frac{1}{3}y_3 \end{bmatrix} .$$

I det aktuelle tilfælde observeredes

3.12

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} 10.11 \\ 0.81 \\ 5.24 \end{bmatrix},$$

således at

$$\begin{bmatrix} \hat{\theta}_A \\ \hat{\theta}_B \end{bmatrix} = \begin{bmatrix} 10.037 \\ 0.735 \end{bmatrix}.$$

Heraf findes let

$$\hat{E}(\underline{Y}) = \underline{X}\hat{\theta} = \begin{bmatrix} 10.037 \\ 0.735 \\ 5.386 \end{bmatrix},$$

og

$$\underline{Y} - \hat{E}(\underline{Y}) = \underline{Y} - \underline{X}\hat{\theta} = \begin{bmatrix} 0.07 \\ 0.07 \\ -0.15 \end{bmatrix}.$$

Dette giver residualkvadratafvigelseessummen

$$(\underline{Y} - \underline{X}\hat{\theta})'(\underline{Y} - \underline{X}\hat{\theta}) = 0.07^2 + 0.07^2 + 0.15^2 = 0.0338,$$

hvorfor et centralt skøn over  $\sigma^2$  er

$$\frac{1}{3-2} 0.0338 = 0.0338.$$

### 3.1.3 Tilfældet $\underline{X}'\underline{\Sigma}^{-1}\underline{X}$ singular

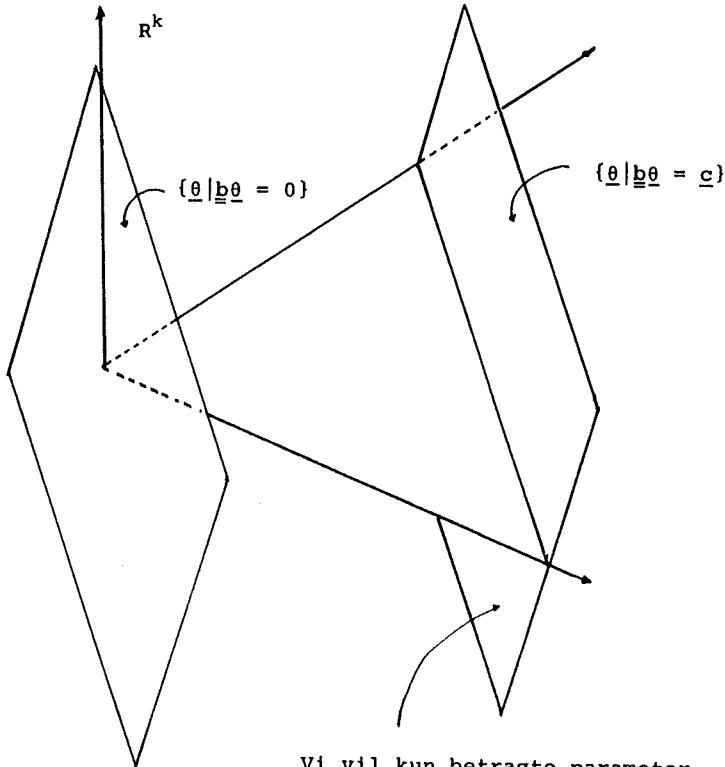
Hvis  $\text{rg}(\underline{X}) = p < k$  er  $\underline{X}'\underline{\Sigma}^{-1}\underline{X}$  singular, og vi kan derfor ikke umiddelbart løse ligningen

$$(\underline{X}'\underline{\Sigma}^{-1}\underline{X})\hat{\theta} = \underline{X}'\underline{\Sigma}^{-1}\underline{Y}.$$

Hvis vi kan finde en pseudoinvers til  $\underline{\underline{x}}'\underline{\underline{\Sigma}}^{-1}\underline{\underline{x}}$  kan vi umiddelbart skrive

$$\hat{\underline{\theta}} = (\underline{\underline{x}}'\underline{\underline{\Sigma}}^{-1}\underline{\underline{x}})^{-1}\underline{\underline{x}}'\underline{\underline{\Sigma}}^{-1}\underline{\underline{y}} .$$

Fra tid til anden er det dog muligt at benytte et lille trick ved bestemmelsen af den pseudoinverse. Grunden til singulariteten er, at vi har for mange parametre. Det ville derfor være rimeligt at kræve, at  $\underline{\theta}$  kun kan variere frit i et sideunderrum i  $\mathbb{R}^k$ . Et sådant kan eg. bestemmes ved, at  $\underline{\theta}$  skal tilfredsstille de lineære ligninger (bånd)



Vi vil kun betragte parameter-værdier  $\underline{\theta}$ , der ligger i dette sideunderrum i  $\mathbb{R}^k$ .

$$\underline{b}\underline{\theta} = \underline{c}$$

eller

$$\begin{bmatrix} b_{11} & \dots & b_{1k} \\ \vdots & & \vdots \\ b_{m1} & \dots & b_{mk} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_k \end{bmatrix} = \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix} .$$

Hvis der overhovedet eksisterer  $\underline{\theta}$ 'er, der tilfredsstiller dette ligningssystem, da udgør de et sideunderrum af dimension  $k - \text{rg}(\underline{b})$ .

Da  $\text{rg}(\underline{x}) = p$ , og da vi har  $k$   $\theta$ -komponenter, vil det være rimeligt at "fjerne"  $k-p$  af disse, d.v.s. at kræve, at  $k - \text{rg}(\underline{b}) = p$  eller  $k = p + \text{rg}(\underline{b})$ .

Hvis nu

$$\text{rg} \begin{bmatrix} \underline{x} \\ \underline{b} \end{bmatrix} = \text{rg} \begin{bmatrix} x_{11} & \dots & x_{1k} \\ \vdots & & \vdots \\ x_{n1} & \dots & x_{nk} \\ b_{11} & \dots & b_{1k} \\ \vdots & & \vdots \\ b_{m1} & \dots & b_{mk} \end{bmatrix} = k ,$$

kan vi betragte "modellen"

$$\begin{bmatrix} \underline{y} \\ \underline{c} \end{bmatrix} = \begin{bmatrix} \underline{x} \\ \underline{b} \end{bmatrix} \underline{\theta} + \begin{bmatrix} \underline{\varepsilon} \\ \underline{0} \end{bmatrix} .$$

Vi sætter

$$D = \begin{bmatrix} \underline{y}^{-1} & \underline{0}_{n,m} \\ \underline{0}_{m,n} & \underline{I}_{m,m} \end{bmatrix} = \begin{bmatrix} \underline{y}^{-1} & \underline{0} \\ \underline{0} & \underline{I} \end{bmatrix} ,$$

hvor den forkortede skrivemåde ikke vil give anledning til vanskeligheder.

Beregner vi nu helt sædvanligt

$$\begin{aligned}\hat{\theta} &= \{[\underline{\mathbf{x}}' \underline{\mathbf{b}}'] \underline{\mathbf{D}} \begin{bmatrix} \underline{\mathbf{x}} \\ \underline{\mathbf{b}} \end{bmatrix}\}^{-1} \{[\underline{\mathbf{x}}' \underline{\mathbf{b}}'] \underline{\mathbf{D}} \begin{bmatrix} \underline{\mathbf{y}} \\ \underline{\mathbf{c}} \end{bmatrix}\} \\ &= (\underline{\mathbf{x}}' \underline{\Sigma}^{-1} \underline{\mathbf{x}} + \underline{\mathbf{b}}' \underline{\mathbf{b}})^{-1} (\underline{\mathbf{x}}' \underline{\Sigma}^{-1} \underline{\mathbf{y}} + \underline{\mathbf{b}}' \underline{\mathbf{c}}),\end{aligned}$$

har vi fået en størrelse, der minimaliserer

$$\begin{aligned}g(\theta) &= \left\{ \begin{bmatrix} \underline{\mathbf{y}} \\ \underline{\mathbf{c}} \end{bmatrix} - \begin{bmatrix} \underline{\mathbf{x}} \\ \underline{\mathbf{b}} \end{bmatrix} \theta \right\}' \underline{\mathbf{D}} \left\{ \begin{bmatrix} \underline{\mathbf{y}} \\ \underline{\mathbf{c}} \end{bmatrix} - \begin{bmatrix} \underline{\mathbf{x}} \\ \underline{\mathbf{b}} \end{bmatrix} \theta \right\} \\ &= \begin{bmatrix} \underline{\mathbf{y}} - \underline{\mathbf{x}}\theta \\ \underline{\mathbf{0}} \end{bmatrix}' \begin{bmatrix} \underline{\Sigma}^{-1} & \underline{\mathbf{0}} \\ \underline{\mathbf{0}} & \underline{\mathbf{I}} \end{bmatrix} \begin{bmatrix} \underline{\mathbf{y}} - \underline{\mathbf{x}}\theta \\ \underline{\mathbf{0}} \end{bmatrix} \\ &= (\underline{\mathbf{y}} - \underline{\mathbf{x}}\theta)' \underline{\Sigma}^{-1} (\underline{\mathbf{y}} - \underline{\mathbf{x}}\theta) \\ &= \|\underline{\mathbf{y}} - \underline{\mathbf{x}}\theta\|^2.\end{aligned}$$

Da dette netop er denne størrelse vi skal bestemme for at finde ML-skønnene, ser vi, at

$$\hat{\theta} = (\underline{\mathbf{x}}' \underline{\Sigma}^{-1} \underline{\mathbf{x}} + \underline{\mathbf{b}}' \underline{\mathbf{b}})^{-1} (\underline{\mathbf{x}}' \underline{\Sigma}^{-1} \underline{\mathbf{y}} + \underline{\mathbf{b}}' \underline{\mathbf{c}})$$

virkelig er maksimum likelihood estimatoren for  $\theta$ . Det eneste der kræves, er altså, at vi kan finde en matrix  $\underline{\mathbf{b}}$  så  $\begin{bmatrix} \underline{\mathbf{x}} \\ \underline{\mathbf{b}} \end{bmatrix}$  får fuld rang, og dette svarer til, at vi indskrænker  $\theta$ 's variationsområde.

Dispersionsmatricen for  $\hat{\theta}$  bliver

$$D(\hat{\theta}) = \sigma^2 \{ \underline{\mathbf{x}}' \underline{\Sigma}^{-1} \underline{\mathbf{x}} + \underline{\mathbf{b}}' \underline{\mathbf{b}} \}^{-1} \underline{\mathbf{x}}' \underline{\Sigma}^{-1} \underline{\mathbf{x}} \{ \underline{\mathbf{x}}' \underline{\Sigma}^{-1} \underline{\mathbf{x}} + \underline{\mathbf{b}}' \underline{\mathbf{b}} \}^{-1}.$$

Dette udtryk fås ved umiddelbar anvendelse af sætning 2.5.

Det centrale skøn over  $\sigma^2$  bliver som før

$$\frac{1}{n - r_{\underline{\mathbf{x}}}} \|\underline{\mathbf{y}} - \underline{\mathbf{x}}\hat{\theta}\|^2$$

Her er  $n\text{-rg}\underline{x} = n\text{-k} + \text{rg}\underline{b}$ .

Vi giver nu først et lille teoretisk

Eksempel 3.4 Vi betragter en meget simpel ensidet variansanalyse med to grupper med to observationer i hver. Vi kan forestille os, at man vil undersøge en katalysators effekt på udbyttet af en proces, og at man derfor foretager 4 forsøg, to med katalysatoren i niveau A og to med katalysatoren i niveau B. Vi har derfor målinger

$$\begin{aligned} \text{niveau A: } & Y_{11}, Y_{12} \\ \text{niveau B: } & Y_{21}, Y_{22} . \end{aligned}$$

Hvis vi regner med, at observationerne er stokastisk uafhængige og har middelværdier

$$E(Y_{11}) = E(Y_{12}) = \theta_1$$

$$E(Y_{21}) = E(Y_{22}) = \theta_2 ,$$

kan vi skrive modellen som

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + \underline{\varepsilon} = \underline{x}\underline{\theta} + \underline{\varepsilon} .$$

Man finder let

$$\underline{x}'\underline{x} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} ,$$

og

$$\hat{\underline{\theta}} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} Y_{11} + Y_{12} \\ Y_{21} + Y_{22} \end{bmatrix} = \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix} ,$$

hvilket er de sædvanligt kendte estimatorer. Anvender vi i stedet parametriseringen

$$E(Y_{11}) = E(Y_{12}) = \mu + \alpha_1$$

$$E(Y_{21}) = E(Y_{22}) = \mu + \alpha_2 ,$$

d.v.s. vi udtrykker effekten af en katalysator som et niveau plus den specifikke effekt af den pågældende katalysator - da fås

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} + \underline{\epsilon} = \underline{x}\underline{\alpha} + \underline{\epsilon} .$$

Det ses let, at  $\underline{x}$  har rangen 2 (summen af de to sidste søjler er lig den første). Vi vil derfor forsøge at indføre et lineært bånd mellem parametrene. Vi forsøger med

$$\alpha_1 + \alpha_2 = 0 \text{ d.v.s.: } (0 \ 1 \ 1) \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} = 0 .$$

Vi opstiller nu formelt modellen

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \underline{\epsilon} \\ 0 \end{bmatrix} ,$$

eller

$$\begin{bmatrix} \underline{Y} \\ 0 \end{bmatrix} = \begin{bmatrix} \underline{x} \\ 0 \quad 1 \quad \underline{1} \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \underline{\epsilon} \\ 0 \end{bmatrix} .$$

Vi har herefter

$$\begin{bmatrix} \underline{x} \\ 0 \\ 1 \\ 1 \end{bmatrix}' \begin{bmatrix} \underline{x} \\ 0 \\ 1 \\ 1 \end{bmatrix} = \underline{x}'\underline{x} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 3 & 1 \\ 2 & 1 & 3 \end{bmatrix}.$$

Den inverse til denne matrix er

$$\begin{bmatrix} +\frac{1}{2} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & +\frac{1}{2} & 0 \\ -\frac{1}{4} & 0 & +\frac{1}{2} \end{bmatrix}.$$

Da

$$\begin{bmatrix} \underline{x} \\ 0 \\ 1 \\ 1 \end{bmatrix}' \begin{bmatrix} \underline{y} \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & & & & 0 \end{bmatrix} \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ 0 \end{bmatrix} = \begin{bmatrix} \Sigma Y_{ij} \\ Y_{11} + Y_{12} \\ Y_{21} + Y_{22} \end{bmatrix},$$

fås

$$\begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{2} & 0 \\ -\frac{1}{4} & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \Sigma Y_{ij} \\ Y_{11} + Y_{12} \\ Y_{21} + Y_{22} \end{bmatrix} = \begin{bmatrix} \bar{Y} \\ \bar{Y}_1 - \bar{Y} \\ \bar{Y}_2 - \bar{Y} \end{bmatrix},$$

d.v.s. netop de estimatorer vi er vant til. (NB. Vi ved på forhånd, at vi vil komme frem til disse, jvf. p. 3.15).

□

Vi giver nu et lidt mere praktisk betonet eksempel på estimation af parametre i tilfælde, hvor  $\underline{x}'\underline{\Sigma}^{-1}\underline{x}$  er singular.

Eksempel 3.5 Til fremstilling af enzymer kan anvendes 2 principielt forskellige bakterietyper. Den ene udskiller gennem



sit stofskifte syre under dyrkning (syredanner), den anden udskiller neutrale stofskifteprodukter. For at regulere pH-værdien i det substrat, bakterierne dyrkes på, kan der tilsættes en såkaldt pH-buffer. Man ved, at pH-bufferen i sig selv ikke har nogen effekt på enzymudbyttet, men udelukkende virker gennem en vekselvirkning med surhedsgraden og bakteriernes stofskifteprodukter.

For en "neutral" bakteriestamme, som lever på et substrat uden pH-buffer, kendes middelenzymudbyttet (normudbyttet). For at estimere de ovennævnte vekselvirkningseffekter er ved 7 forsøg målt forskellen mellem normudbyttet og det faktiske enzymudbytte som nedenfor angivet.

		pH-buffer	
		tilsat	ikke tilsat
bakterie stamme	syredanner	0,-2	-19, -15
	neutral	-6, 0,-2	

Figur 3.3 Forskelle mellem normudbytte og det faktiske enzymudbytte ved forskellige forsøgsomstændigheder.

Vi vil først formulere en matematisk model, der kan beskrive ovenstående eksperiment.

Der foreligger observationer

$$Y_{11v}, v = 1,2$$

$$Y_{12v}, v = 1,2$$

$$Y_{21v}, v = 1,2,3 .$$

Disse antages at have forventningsværdierne

$$E(y_{11v}) = \mu_1 + \theta_{11}$$

$$E(y_{12v}) = \mu_1 + \theta_{12}$$

$$E(y_{21v}) = \theta_{21}$$

hvor  $\mu_1$  er effekten af at anvende syredannende bakterier og  $\theta_{ij}$  angiver vekselvirkningerne mellem tilstedeværelsen af pH-buffer og bakteriestammetyper.

Endvidere antages observationerne at være stokastiske uafhængige og at have samme ukendte varians  $\sigma^2$ .

Vi kan nu formulere modellen som en generel lineær model. Vi har

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{213} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \theta_{11} \\ \theta_{12} \\ \theta_{21} \end{bmatrix} + \underline{\epsilon}$$

hvor fejlen  $\underline{\epsilon} \in N_7(0, \sigma^2 \underline{I})$ .

Vi finder

$$\underline{X}'\underline{X} = \begin{bmatrix} 4 & 2 & 2 & 0 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

og

$H_0: \theta_{21} = 0$

variation sak $\theta$	
M-F	1.33
O-M	28.674
O-F	50.0

$$F = \frac{21.33}{4} / \frac{28.674}{4} = 2.97 < F$$

0.

$$\underline{\underline{x}}' \underline{\underline{y}} = \begin{bmatrix} y_{1..} \\ y_{11.} \\ y_{12.} \\ y_{21.} \end{bmatrix},$$

hvor et punktum på en indeksplads angiver, at der er summeret over det pågældende indeks.

Da  $\underline{\underline{x}}' \underline{\underline{x}}$  åbenbart kun har rangen 3, kan vi ikke umiddelbart invertere den. I stedet kan vi finde en pseudoinvers. Vi benytter sætning 1.7 p. 1.33 og får

$$(\underline{\underline{x}}' \underline{\underline{x}})^- = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{3} \end{bmatrix},$$

hvorfor estimatorerne for parametrene bliver - med dette specielle valg af pseudoinvers -

$$\hat{\underline{\underline{\theta}}} = (\underline{\underline{x}}' \underline{\underline{x}})^- \underline{\underline{x}}' \underline{\underline{y}} = \begin{bmatrix} 0 \\ \bar{y}_{11.} \\ \bar{y}_{12.} \\ \bar{y}_{21.} \end{bmatrix},$$

hvor f.eks.

$$\bar{y}_{13.} = \frac{1}{3} \sum_{v=1}^3 y_{13v}.$$

Da

$$\underline{\underline{I}} - (\underline{\underline{x}}' \underline{\underline{x}})^- \underline{\underline{x}}' \underline{\underline{x}} = \begin{bmatrix} +1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

er

$$(\underline{I} - (\underline{x}'\underline{x})^{-1}\underline{x}'\underline{z})\underline{z} = \begin{bmatrix} z_1 \\ -z_1 \\ -z_1 \\ 0 \end{bmatrix}$$

Ifølge sætning 1.6 er den fuldstændige løsning til normal-ligningerne derfor alle vektorer af formen

$$\hat{\underline{\theta}} + \begin{bmatrix} t \\ -t \\ -t \\ 0 \end{bmatrix} = \begin{bmatrix} t \\ \bar{y}_{11} \cdot -t \\ \bar{y}_{12} \cdot -t \\ \bar{y}_{21} \cdot -t \end{bmatrix}, \quad t \in \mathbb{R}.$$

En vilkårlig maksimum likelihood estimator for  $\underline{\theta}$  er altså af denne form.

Den observerede værdi af  $\hat{\underline{\theta}}$  er

$$\hat{\underline{\theta}}_{\text{obs}} = \begin{bmatrix} 0 \\ -1 \\ -17 \\ -2\frac{2}{3} \end{bmatrix}.$$

Det er åbenbart, at denne estimator ikke er særligt tilfredsstillende, idet  $\hat{\mu}_1$  f.eks. altid vil være 0. For at få estimators, der bedre svarer til vore forestillinger om den fysiske virkelighed, må vi lægge nogle bånd på parametrene. Det forekommer rimeligt at kræve

$$\theta_{11} + \theta_{12} = 0,$$

d.v.s.

$$(0 \ 1 \ 1 \ 0) \begin{bmatrix} \mu_1 \\ \theta_{11} \\ \theta_{12} \\ \theta_{21} \end{bmatrix} = 0 ,$$

eller

$$\underline{\underline{b}}\underline{\underline{\theta}} = 0 .$$

Det er åbenbart, at

$$\text{rg} \left( \begin{bmatrix} \underline{\underline{x}} \\ \underline{\underline{b}} \end{bmatrix} \right) = 4 ,$$

hvorfor vi kan anvende resultatet fra p. 3.15. Vi finder

$$\underline{\underline{x}}'\underline{\underline{x}} + \underline{\underline{b}}'\underline{\underline{b}} = \begin{bmatrix} 4 & 2 & 2 & 0 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 2 & 2 & 0 \\ 2 & 3 & 1 & 0 \\ 2 & 1 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} .$$

Da

$$\begin{bmatrix} 4 & 2 & 2 \\ 2 & 3 & 1 \\ 2 & 1 & 3 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{2} & 0 \\ -\frac{1}{4} & 0 & \frac{1}{2} \end{bmatrix} ,$$

fås

$$(\underline{\underline{x}}'\underline{\underline{x}} + \underline{\underline{b}}'\underline{\underline{b}})^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & -\frac{1}{4} & 0 \\ -\frac{1}{4} & \frac{1}{2} & 0 & 0 \\ -\frac{1}{4} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{3} \end{bmatrix} .$$

Vi finder nu

$$\hat{\theta} = (\underline{\underline{x}}'\underline{\underline{x}} + \underline{\underline{b}}'\underline{\underline{b}})^{-1}\underline{\underline{x}}'\underline{\underline{y}} = \begin{bmatrix} \bar{y}_{1..} \\ \bar{y}_{11.} - \bar{y}_{1..} \\ \bar{y}_{12.} - \bar{y}_{1..} \\ \bar{y}_{21.} \end{bmatrix} .$$

Den observerede værdi er

$$\begin{bmatrix} -9 \\ +8 \\ -8 \\ -2\frac{2}{3} \end{bmatrix} (= \begin{bmatrix} \text{syredanneeffekt} \\ \text{vex.v. buffer \& syre} \\ \text{vex.v. (-buffer) \& syre} \\ \text{vex.v. buffer \& neutral} \end{bmatrix} ) .$$

Vi finder dernæst dispersionsmatricen for  $\hat{\theta}$ . Der gælder

$$D(\hat{\theta}) = \sigma^2 (\underline{\underline{x}}'\underline{\underline{x}} + \underline{\underline{b}}'\underline{\underline{b}})^{-1}\underline{\underline{x}}'\underline{\underline{x}}(\underline{\underline{x}}'\underline{\underline{x}} + \underline{\underline{b}}'\underline{\underline{b}})^{-1}$$

$$= \sigma^2 \begin{bmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & -\frac{1}{4} & 0 \\ 0 & -\frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{3} \end{bmatrix} ,$$

d.v.s. estimatorerne er ikke uafhængige.

For at estimere  $\sigma^2$  finder vi residualvektoren. Da

$$\underline{\underline{x}}\hat{\theta} = \begin{bmatrix} \hat{\mu}_1 + \hat{\theta}_{11} \\ \hat{\mu}_1 + \hat{\theta}_{11} \\ \hat{\mu}_1 + \hat{\theta}_{12} \\ \hat{\mu}_1 + \hat{\theta}_{12} \\ \hat{\theta}_{21} \\ \hat{\theta}_{21} \\ \hat{\theta}_{21} \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ -17 \\ -17 \\ -2\frac{2}{3} \\ -2\frac{2}{3} \\ -2\frac{2}{3} \end{bmatrix} ,$$

er residualvektoren

$$\underline{y} - \underline{\hat{x}\theta} = \begin{bmatrix} +1 \\ -1 \\ -2 \\ +2 \\ -3\frac{1}{3} \\ +2\frac{2}{3} \\ \frac{2}{3} \end{bmatrix} .$$

Dermed er

$$\| \underline{y} - \underline{\hat{x}\theta} \|^2 = (\underline{y} - \underline{\hat{x}\theta})' (\underline{y} - \underline{\hat{x}\theta}) = 1^2 + \dots + \left(\frac{2}{3}\right)^2 = 28\frac{2}{3} .$$

Et centralt skøn over  $\sigma^2$  er derfor

$$s^2 = \frac{1}{7-3} 28\frac{2}{3} = 7\frac{1}{6} .$$

□

#### 3.1.4 Estimation under bibetingelser

Vi skal nu beskæftige os med et problem, der meget minder om det i foregående afsnit behandlede, nemlig estimation af parametre, der skal tilfredsstille lineære bånd som

$$\underline{H}'\theta = \underline{\xi} .$$

Dette vil f.eks. være tilfældet, hvor man estimerer vinkler i en bestemt trekant. De skal tilfredsstille

$$(1 \ 1 \ 1) \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = 180^\circ ,$$

og det vil derfor være rimeligt også at kræve dette opfyldt af estimerterne.

Hovedresultatet om estimation af  $\underline{\theta}$  kan udtrykkes i

Sætning 3.3 Lad  $E(\underline{Y}) = \underline{X}\underline{\theta}$ , hvor  $\underline{Y}$  er en  $n$ -dimensional vektor,  $\underline{X}$  en  $n \times k$  matrix af kendte koefficienter, og  $\underline{\theta}$  den  $k$ -dimensionale vektor af ukendte parametre, der yderligere forudsættes at tilfredsstille de  $s$  lineære restriktioner

$$\underline{H}'\underline{\theta} = \underline{\xi} ,$$

hvor  $\underline{H}$  er en  $k \times s$  matrix, og  $\underline{\xi}$  en  $s$ -dimensional vektor. Endelig forudsættes, at  $D(\underline{Y}) = \sigma^2 \underline{\Sigma}$ , hvor  $\underline{\Sigma}$  er en kendt matrix. Mindste kvadraters estimator  $\underline{\hat{\theta}}$  for  $\underline{\theta}$  under bibetingelsen  $\underline{H}'\underline{\theta} = \underline{\xi}$  er da løsning til ligningssystemet

$$\begin{bmatrix} \underline{X}'\underline{\Sigma}^{-1}\underline{X} & \underline{H}' \\ \underline{H}' & \underline{0} \end{bmatrix} \begin{bmatrix} \underline{\theta} \\ \underline{\lambda} \end{bmatrix} = \begin{bmatrix} \underline{X}'\underline{\Sigma}^{-1}\underline{Y} \\ \underline{\xi} \end{bmatrix} .$$

Bevis Vi skal finde

$$\min_{\underline{H}'\underline{\theta}=\underline{\xi}} (\underline{Y} - \underline{X}\underline{\theta})' \underline{\Sigma}^{-1} (\underline{Y} - \underline{X}\underline{\theta}) .$$

Vi indfører Lagrangemultiplikatorer  $\underline{\lambda}$  og sætter

$$F(\underline{\theta}, \underline{\lambda}) = \frac{1}{2} (\underline{Y} - \underline{X}\underline{\theta})' \underline{\Sigma}^{-1} (\underline{Y} - \underline{X}\underline{\theta}) + \underline{\lambda}' (\underline{H}'\underline{\theta} - \underline{\xi}) .$$

Da bliver

$$\frac{\partial F}{\partial \underline{\theta}} = - \underline{X}'\underline{\Sigma}^{-1}\underline{Y} + \underline{X}'\underline{\Sigma}^{-1}\underline{X}\underline{\theta} + \underline{H}\underline{\lambda}$$

$$\frac{\partial F}{\partial \underline{\lambda}} = \underline{H}'\underline{\theta} - \underline{\xi} .$$



Disse to afledede er 0 i ethvert ekstremum for  $(\underline{Y} - \underline{x}\theta)' \underline{\Sigma}^{-1} (\underline{Y} - \underline{x}\theta)$  under bibetingelsen  $\underline{H}'\theta = \underline{E}$ . Udnyttes dette fås umiddelbart påstanden i sætningen.

Q.E.D.

Vi betragter dernæst problemet med estimation af  $\sigma^2$  i

**Sætning 3.4** Lad en pseudoinvers til koefficientmatricen i sætning 3.3 være

$$\begin{bmatrix} \underline{x}' \underline{\Sigma}^{-1} \underline{x} & \underline{H} \\ \underline{H}' & \underline{0} \end{bmatrix}^{-} = \begin{bmatrix} \underline{C}_1 & \underline{C}_2 \\ \underline{C}_3 & \underline{C}_4 \end{bmatrix} .$$

Da er

$$D(\underline{\tilde{\theta}}) = \sigma^2 \underline{C}_1 ,$$

og et centralt skøn for  $\sigma^2$  er

$$\hat{\sigma}^2 = \frac{1}{f} (\underline{Y}' \underline{\Sigma}^{-1} \underline{Y} - \underline{\tilde{\theta}}' \underline{x}' \underline{\Sigma}^{-1} \underline{Y} - \underline{E}' \underline{\tilde{\chi}}) ,$$

hvor  $(\underline{\tilde{\theta}}', \underline{\tilde{\chi}})'$  er løsning til ligningssystemet i sætning 3.3, og

$$f = n - \text{rg}(\underline{x}' \underline{H}) + \text{rg}(\underline{H}) .$$

**Bevis** Ved at indføre den pseudoinverse fås, at

$$\underline{\tilde{\theta}} = \underline{C}_1 \underline{x}' \underline{\Sigma}^{-1} \underline{Y} + \underline{C}_2 \underline{E}$$

$$\underline{\tilde{\chi}} = \underline{C}_3 \underline{x}' \underline{\Sigma}^{-1} \underline{Y} + \underline{C}_4 \underline{E} .$$

Heraf fås umiddelbart

$$D(\underline{\tilde{\theta}}) = \sigma^2 \underline{C}_1 \underline{x}' \underline{\Sigma}^{-1} \underline{\Sigma} \underline{\Sigma}^{-1} \underline{x} \underline{C}_1'$$

$$\begin{aligned}
 &= \sigma^2 \underline{\underline{C}}_1 \underline{\underline{X}}' \underline{\underline{\Sigma}}^{-1} \underline{\underline{X}} \underline{\underline{C}}_1' \\
 &= \sigma^2 \underline{\underline{C}}_1 .
 \end{aligned}$$

Det sidste lighedstegn følger ved at benytte egenskaber ved pseudoinverse matricer.

Ved direkte indsætning ses, at

$$(\underline{\underline{Y}} - \underline{\underline{X}}\tilde{\underline{\theta}})' \underline{\underline{\Sigma}}^{-1} (\underline{\underline{Y}} - \underline{\underline{X}}\tilde{\underline{\theta}}) = (\underline{\underline{Y}}' \underline{\underline{\Sigma}}^{-1} \underline{\underline{Y}} - \tilde{\underline{\theta}}' \underline{\underline{X}}' \underline{\underline{\Sigma}}^{-1} \underline{\underline{Y}} - \underline{\underline{E}}' \tilde{\underline{\lambda}}) ,$$

så vi mangler blot at vise, at antallet af frihedsgrader er det i sætningen anførte. Betragter vi løsningerne til ligningen

$$\underline{\underline{H}}' \underline{\underline{\theta}} = \underline{\underline{E}} ,$$

kan disse skrives på formen

$$\underline{\underline{\theta}} = \underline{\underline{\theta}}_0 + \underline{\underline{B}}\underline{\underline{\beta}} ,$$

hvor  $\underline{\underline{\theta}}_0$  er en partikulær løsning, og  $\underline{\underline{B}}$  er en  $(k \times s)$  matrix ( $\text{rg}(\underline{\underline{H}}) = k-s$ ) med

$$\underline{\underline{H}}' \underline{\underline{B}} = \underline{\underline{0}} .$$

Endelig er  $\underline{\underline{\beta}}$  en  $s$ -dimensional vektor af "frie", nye parametre. Betragtes

$$\underline{\underline{Z}} = \underline{\underline{Y}} - \underline{\underline{X}}\underline{\underline{\theta}}_0 ,$$

fås

$$\begin{aligned}
 \mathbb{E}(\underline{\underline{Z}}) &= \underline{\underline{X}}\underline{\underline{\theta}} - \underline{\underline{X}}\underline{\underline{\theta}}_0 = \underline{\underline{X}}(\underline{\underline{\theta}} - \underline{\underline{\theta}}_0) \\
 &= \underline{\underline{X}}(\underline{\underline{\theta}} - \underline{\underline{\theta}}_0) \\
 &= \underline{\underline{X}}\underline{\underline{B}}\underline{\underline{\beta}} .
 \end{aligned}$$

Vi kan nu blot betragte modellen

$$\underline{z} = \underline{x}\underline{B}\underline{\beta} + \underline{\varepsilon} ,$$

hvor  $\underline{\varepsilon}$  er fejlvektoren og løse denne. Dette giver de i det tidligere anførte estimater.

Med

$$\hat{\underline{\beta}} = (\underline{B}'\underline{x}'\underline{\Sigma}^{-1}\underline{x}\underline{B})^{-1}\underline{B}'\underline{x}'\underline{\Sigma}^{-1}\underline{y} ,$$

bliver

$$\tilde{\underline{\theta}} = \underline{\theta}_0 + \underline{B}\hat{\underline{\beta}} ,$$

hvorfor

$$\begin{aligned} \underline{y} - \underline{x}\tilde{\underline{\theta}} &= \underline{z} + \underline{x}\underline{\theta}_0 - \underline{x}\underline{\theta}_0 - \underline{x}\underline{B}\hat{\underline{\beta}} \\ &= \underline{z} - \underline{x}\underline{B}\hat{\underline{\beta}} , \end{aligned}$$

og fra den almindelige teori fås at antallet af frihedsgrader bliver  $n - \text{rg}(\underline{x}\underline{B})$ . Nu er

$$\begin{aligned} \text{rg}(\underline{x}\underline{B}) &= \dim\{\underline{x}\underline{B}\underline{\beta} \mid \underline{\beta} \in \mathbb{R}^s\} \\ &= \dim\{\underline{x}\underline{y} \mid \underline{H}'\underline{y} = 0, \underline{y} \in \mathbb{R}^m\} \\ &= \text{rg}\left(\begin{array}{c} \underline{x} \\ \underline{H}' \end{array}\right) - \text{rg}(\underline{H}) . \end{aligned}$$

Det sidste lighedstegn fås af relationen

$$\dim S_1^* + \dim S_2^* = \text{rg}\left(\begin{array}{c} \underline{x} \\ \underline{H}' \end{array}\right) ,$$

hvor

$$S_1^* = \left\{ \begin{pmatrix} \underline{X} \\ \underline{H}' \end{pmatrix} \underline{Y} \mid \underline{Y} \in N(H) \right\}$$

$$S_2^* = \left\{ \begin{pmatrix} \underline{X} \\ \underline{H}' \end{pmatrix} \underline{Y} \mid \underline{Y} \in N(H)^\perp \right\}$$

(idet man bemærker, at  $\dim S_2^* = \text{rg } H$ ).

Q.E.D.

Vi vil nu give et simpelt, illustrativt eksempel.

Eksempel 3.6 Lad os antage, at vi har 3 uafhængige målinger af vinklerne i en trekant (f.eks. ude i terræn), og at disse er fundet til

$$v_1 = 52^\circ, 54^\circ$$

$$v_2 = 74^\circ, 74^\circ$$

$$v_3 = 48^\circ, 46^\circ .$$

Vi antager endvidere, at usikkerheden på disse 3 bestemmelser er ens og kan udtrykkes ved en varians  $\sigma^2$ .

Vi formulerer ovenstående som en sædvanlig lineær model med bibetingelse, i.e.

$$\begin{bmatrix} v_{11} \\ v_{12} \\ v_{21} \\ v_{22} \\ v_{31} \\ v_{32} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix}$$

$$(1, 1, 1) \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = 180 ,$$

$$D(\underline{\epsilon}) = \sigma^2 \underline{\underline{I}} .$$

Vi finder

$$\begin{bmatrix} \underline{\underline{X}}' \underline{\underline{\Sigma}}^{-1} \underline{\underline{X}} & \underline{\underline{H}} \\ \underline{\underline{H}}' & \underline{\underline{0}} \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & | & 1 \\ 0 & 2 & 0 & | & 1 \\ \hline 0 & 0 & 2 & | & 1 \\ 1 & 1 & 1 & | & 0 \end{bmatrix} .$$

En (pseudo)invers til denne er

$$\begin{bmatrix} \underline{\underline{C}}_1 & \underline{\underline{C}}_2 \\ \underline{\underline{C}}_3 & \underline{\underline{C}}_4 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 2 & -1 & -1 & | & 2 \\ -1 & 2 & -1 & | & 2 \\ \hline -1 & -1 & 2 & | & 2 \\ 2 & 2 & 2 & | & -4 \end{bmatrix} .$$

Vi finder derfor

$$\begin{aligned} \underline{\underline{\hat{\theta}}} &= \frac{1}{6} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} 106 \\ 148 \\ 94 \end{bmatrix} + \frac{1}{6} \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} 180 \\ &= \begin{bmatrix} -5 \\ 16 \\ -11 \end{bmatrix} + \begin{bmatrix} 60 \\ 60 \\ 60 \end{bmatrix} \\ &= \begin{bmatrix} 55 \\ 76 \\ 49 \end{bmatrix} . \end{aligned}$$

Vi bemærker trivielt, at summen af koordinaterne er 180.

Dispersionsmatricen bliver

$$D(\tilde{\theta}) = \sigma^2 \underline{C}_1 = \frac{\sigma^2}{6} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} .$$

Skønnet over  $\sigma^2$  bliver

$$\sigma^2 = \frac{1}{6-3+1} (20992 - 21684 - (-720)) = 7 = 2.6^2 ,$$

idet

$$\tilde{\lambda} = \frac{1}{6} [2, 2, 2] \begin{bmatrix} 106 \\ 148 \\ 94 \end{bmatrix} + \left(-\frac{4}{6}\right) 180 = 116 - 120 = -4$$

□

Bemærkning Som det er antydnet i ovenstående eksempel nyder denne teori speciel anvendelse inden for geodæsien og landmålingen. Vi skal dog ikke komme ind på nærmere detaljer her.

### 3.1.5 Konfidensintervaller for forudsagte værdier. Prediktionsinterval

Vi betragter modellen ( $n > k$ )

$$\begin{bmatrix} Y_1 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix} = \begin{bmatrix} x_{11} \cdots x_{1k} \\ \cdot \\ \cdot \\ \cdot \\ x_{n1} \cdots x_{nk} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \cdot \\ \cdot \\ \cdot \\ \theta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_1 \end{bmatrix} ,$$

hvor

$$\underline{\epsilon} \in N(\underline{0}, \sigma^2 \underline{\Sigma}) .$$

Vi vil her benævne  $Y$ 'erne de afhængige variable og  $x$ 'erne de uafhængige variable.

Her er som vanligt  $\sigma^2$  ukendt og  $\Sigma$  kendt. Vi har estimatoren

$$\hat{\underline{\theta}} = (\underline{x}'\underline{\Sigma}^{-1}\underline{x})^{-1}\underline{x}'\underline{\Sigma}^{-1}\underline{Y}$$

for  $\underline{\theta}$ , og  $\sigma^2$  estimeres ved

$$\begin{aligned}\hat{\sigma}^2 &= s^2 = \frac{1}{n-k} \|\underline{Y} - \underline{x}\hat{\underline{\theta}}\|^2 \\ &= \frac{1}{n-k} (\underline{Y} - \underline{x}\hat{\underline{\theta}})' \underline{\Sigma}^{-1} (\underline{Y} - \underline{x}\hat{\underline{\theta}}) .\end{aligned}$$

Hvis vi ønsker at forudsige forventningsværdien af en kommende observation  $Y$  af den afhængige variabel, svarende til værdierne

$$(z_1, \dots, z_k) = \underline{z}'$$

af de uafhængige variable, er det åbenbart, at vi vil anvende

$$\underline{Z} = (z_1, \dots, z_k) \begin{bmatrix} \hat{\theta}_1 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\theta}_k \end{bmatrix} = \underline{z}' \hat{\underline{\theta}}$$

som et "bedste" gæt.

Vi har, at  $E(\underline{Z}) = E(\underline{Y})$ , og

$$\begin{aligned}V(\underline{Z}) &= \underline{z}' D(\hat{\underline{\theta}}) \underline{z} \\ &= \sigma^2 \underline{z}' (\underline{x}' \underline{\Sigma}^{-1} \underline{x})^{-1} \underline{z} \\ &= \sigma^2 c ,\end{aligned}$$

hvor altså

$$c = (z_1, \dots, z_k) (\underline{\underline{x}}' \underline{\underline{\Sigma}}^{-1} \underline{\underline{x}})^{-1} \begin{bmatrix} z_1 \\ \cdot \\ \cdot \\ \cdot \\ z_k \end{bmatrix} .$$

Vi får derfor umiddelbart

$$\frac{Z - E(Y)}{\sigma \sqrt{C}} \in N(0,1) ,$$

og dermed

$$\frac{Z - E(Y)}{S \sqrt{C}} \in t(n-k) .$$

Vi er nu i stand til at formulere og bevise

Sætning 3.5 Lad situationen være som ovenfor. Da er  $(1-\alpha)$ -konfidensintervallet for en kommende observation  $Y$ 's forventningsværdi

$$[z - t(n-k)_{1-\alpha/2} s \sqrt{C} , z + t(n-k)_{1-\alpha/2} s \sqrt{C}] .$$

Bevis Af ovenstående betragtninger følger umiddelbart

$$1-\alpha = P\{z - t(n-k)_{1-\alpha/2} s \sqrt{C} \leq E(Y) \leq z + t(n-k)_{1-\alpha/2} s \sqrt{C}\} ,$$

og vi har derfor sætningen.

Q.E.D.

Nu er man ofte mere interesseret i et konfidensinterval for kommende observationer end for observationernes forventningsværdi. Vi vælger at betragte det lidt mere generelle problem, at bestemme et konfidensinterval for gennemsnittet  $\bar{y}_q$  af  $q$



observationer taget under betingelsen  $(z_1, \dots, z_k)$ . Vi har, hvis  $Y_{iq} \in N(E(Y), c_1 \sigma^2)$ , at

$$\bar{Y}_q \in N(E(Y), \frac{c_1}{q} \sigma^2) .$$

Hvis vi nu forudsætter, at de kommende observationer er uafhængige af dem, vi har fået, er

$$Z - \bar{Y}_q \in N(0, \sigma^2 (c + \frac{c_1}{q})) ,$$

d.v.s.

$$\frac{Z - \bar{Y}_q}{S \sqrt{c + \frac{c_1}{q}}} \in t(n-k) .$$

Heraf udledes som før

**Sætning 3.6** Lad os forudsætte, at  $q$  kommende observationer under betingelserne  $(z_1, \dots, z_k)$  har en varians  $c_1 \sigma^2$ , og at de er indbyrdes uafhængige og uafhængige af de tidligere observationer. Da er  $(1-\alpha)$ -konfidensintervallet for gennemsnittet af de  $q$  målinger lig intervallet

$$\left[ z - t(n-k)_{1-\alpha/2} s \sqrt{c + \frac{c_1}{q}} , z + t(n-k)_{1-\alpha/2} s \sqrt{c + \frac{c_1}{q}} \right] .$$

**Bemærkning 1** Ovenstående interval er et konfidensinterval for en observation, og ikke, som vi er vant til, for en parameter. Man taler derfor hyppigt om et prediktionsinterval for at skelne de to situationer fra hinanden.

Bemærkning 2 Vi ser, at overgangen til intervallet for  $\bar{Y}_q$  i stedet for intervallet for  $E(\bar{Y}_q) = E(Y)$  blot består i, at leddet under rodtegnet er øget med variansen på  $\bar{Y}_q/\sigma$ .

Eksempel 3.7 Vi betragter følgende samhørende målinger af en uafhængig variabel  $x$  og en afhængig variabel  $y$ :

x	0	1	2	3	4	5	6
y	0.4	0.3	1.5	1.3	1.9	4.2	8

Vi antager, at  $y$ 'erne er realiserede udfald af uafhængige stokastiske variable  $Y_1, \dots, Y_7$ , der er normalt fordelte med middelværdier

$$E(Y|x) = \beta x^2 ,$$

og varianser

$$V(Y|0) = \sigma^2, V(Y|x) = x^2\sigma^2, x > 0.$$

Vi søger et konfidensinterval for en kommende måling svarende til  $x = 10$ . Denne måling kaldes  $Y$ , og vi har

$$E(Y) = 100\beta$$

$$V(Y) = 100\sigma^2 .$$

Vi omformulerer nu problemet til matrix form:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ Y_7 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 4 \\ 9 \\ 16 \\ 25 \\ 36 \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_7 \end{bmatrix} = \underline{\underline{\mathbf{x}}}\beta + \underline{\underline{\varepsilon}},$$

hvor

$$D(\underline{\underline{\varepsilon}}) = \sigma^2 \begin{bmatrix} 1 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 4 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 9 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 16 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 25 & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 36 \end{bmatrix} = \sigma^2 \underline{\underline{\Sigma}}.$$

Vi har, at

$$\begin{aligned} \underline{\underline{\mathbf{x}}}' \underline{\underline{\Sigma}}^{-1} \underline{\underline{\mathbf{x}}} &= (0, 1, 4, 9, 16, 25, 36) \text{diag}\left(1, \frac{1}{4}, \dots, \frac{1}{36}\right) \\ &= 91. \end{aligned}$$

$$\underline{\underline{\mathbf{x}}}' \underline{\underline{\Sigma}}^{-1} \underline{\underline{\mathbf{Y}}} = 0.3 + 1.5 + 1.3 + 1.9 + 4.2 + 8.0 = 17.2.$$

Altså er

$$\hat{\beta} = \frac{17.2}{91} = 0.1890,$$

og

3.38

$$P_M(\underline{Y}) = \begin{bmatrix} 0 \\ 1 \\ 4 \\ 9 \\ 16 \\ 25 \\ 36 \end{bmatrix} \cdot 0.1890 = \begin{bmatrix} 0 \\ 0.1890 \\ 0.7560 \\ 1.7010 \\ 3.0240 \\ 4.7250 \\ 6.8040 \end{bmatrix} .$$

Residualerne er

$$\underline{Y} - P_M(\underline{Y}) = \begin{bmatrix} 0.4000 \\ 0.1110 \\ 0.7440 \\ -0.4010 \\ -1.1240 \\ -0.5250 \\ 1.1960 \end{bmatrix} ,$$

hvorfor

$$\| \underline{Y} - P_M(\underline{Y}) \|^2 = (0.4000 \dots 1.1960) \begin{bmatrix} 1/1 & & & & & & & \\ & 1/1 & & & & & & \\ & & \ddots & & & & & \\ & & & \ddots & & & & \\ & & & & \ddots & & & \\ & & & & & \ddots & & \\ & & & & & & 1/36 & \\ & & & & & & & 1.1960 \end{bmatrix} \begin{bmatrix} 0.4000 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 1.1960 \end{bmatrix}$$

$$= 1.5265358 ,$$

d.v.s.

$$\hat{\sigma}^2 = s^2 = \frac{1}{7-1} 1.5265358 = 0.2544 = 0.5044^2 .$$

Konstanterne  $c$  og  $c_1$  er lig med

$$c = 100 \cdot \frac{1}{91} \cdot 100 = 109.89$$

$$c_1 = 10^2 = 100 .$$

Forudsigelsen til  $x$  lig 10 er

$$z = 100 \hat{\beta} = 18.90$$

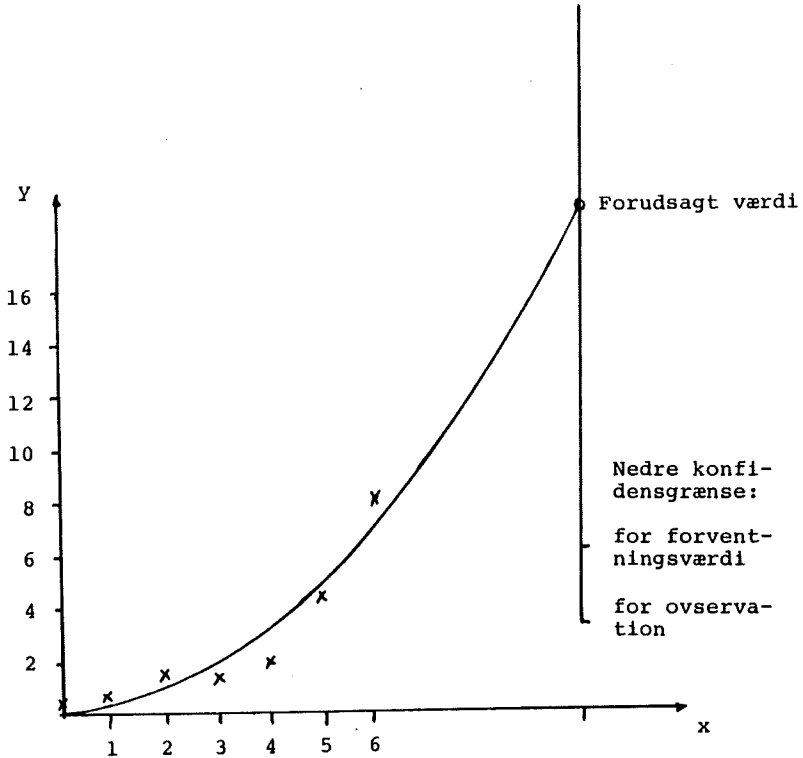
Konfidensintervallet for forventningsværdien i  $x = 10$  er derfor givet ved

$$\begin{aligned} & 18.90 \pm t(6)_{0.975} 0.5044 \sqrt{109.89} \\ & = 18.90 \pm 2.447 \cdot 0.5044 \sqrt{109.89} \\ & = 18.90 \pm 12.9 . \end{aligned}$$

Intervallet for den næste observation er

$$\begin{aligned} & 18.90 \pm t(6)_{0.975} \cdot 0.5044 \sqrt{109.89 + 100} \\ & = 18.90 \pm 17.88 , \end{aligned}$$

d.v.s. et væsentligt bredere interval end for den forventede værdi. Forklaringen er simpelthen, at vi har en varians på  $10^2 \sigma^2 = 100 \sigma^2$  i  $x = 10$ . Vi afbilder observationerne og det estimerede polynomium i nedenstående graf. Endvidere er de to konfidensintervallet angivet.



□

### 3.2 Test i den generelle lineære model

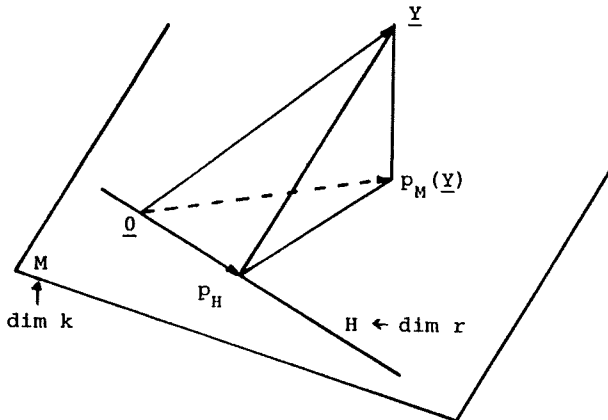
I dette afsnit skal vi dels undersøge om middelværdivektoren kan antages at ligge i et ægte underrum af "modelrummet", og dels undersøge om middelværdivektoren successivt kan antages at ligge i underrum af aftagende dimensioner. Først

#### 3.2.1 Test for lavere dimension af modelrum

Lad  $\underline{y} \in N_n(\underline{\mu}, \sigma^2 \underline{\Sigma})$ , hvor  $\underline{\Sigma}$  er regulær og kendt. Vi forudsætter, at  $\underline{\mu} \in M$ , et  $k$ -dimensionalt underrum, og vi vil teste hypotesen

$$H_0 : \underline{\mu} \in H \text{ mod } H_1 : \underline{\mu} \in M \setminus H ,$$

hvor  $H$  er et  $r$ -dimensionalt underrum af  $M$ . Vi betragter i det følgende den ved  $\underline{\Sigma}^{-1}$  definerede norm. Maksimum likelihood estimatoren for  $\underline{\mu}$  er da  $p_M(\underline{Y})$  - projektionen på  $M$  - og hvis  $H_0$  er sand, da er maksimum likelihood estimatoren  $p_H(\underline{Y})$ ,  $\underline{Y}$ 's projektion på  $H$ . ML estimatorerne for  $\sigma^2$  er i de to tilfælde  $\frac{1}{n} \| \underline{Y} - p_M(\underline{Y}) \|^2$  henholdsvis  $\frac{1}{n} \| \underline{Y} - p_H(\underline{Y}) \|^2$ .



Likelihoodfunktionen er

$$\begin{aligned} L(\underline{\mu}, \sigma^2) &= \frac{1}{\sqrt{2\pi}^n} \frac{1}{\sigma^n} \frac{1}{\sqrt{\det \underline{\Sigma}}} \exp\left(-\frac{1}{2\sigma^2} (\underline{Y} - \underline{\mu})' \underline{\Sigma}^{-1} (\underline{Y} - \underline{\mu})\right) \\ &= k \cdot \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2} \| \underline{Y} - \underline{\mu} \|^2\right) . \end{aligned}$$

Med disse betegnelser har vi

Sætning 3.7 Lad situationen være som ovenfor. Da er kvotienttestet på niveau  $\alpha$  af

$$H_0 : \underline{\mu} \in H \text{ mod } H_1 : \underline{\mu} \in M \setminus H,$$

ækvivalent med testet givet ved det kritiske område

$$C_\alpha = \{ (y_1, \dots, y_n) \mid \frac{\| p_M(\underline{Y}) - p_H(\underline{Y}) \|^2 / (k-r)}{\| \underline{Y} - p_M(\underline{Y}) \|^2 / (n-k)} > F(k-r, n-k)_{1-\alpha} \} .$$

Bevis Kvotientteststørrelsen er

$$\begin{aligned} Q &= \frac{\sup_{H_0} L(\underline{\mu}, \sigma^2)}{\sup L(\underline{\mu}, \sigma^2)} = \frac{L(p_H(\underline{Y}), \hat{\sigma}_H^2)}{L(p_M(\underline{Y}), \hat{\sigma}_M^2)} \\ &= \left[ \frac{\| \underline{Y} - p_M(\underline{Y}) \|^2}{\| \underline{Y} - p_H(\underline{Y}) \|^2} \right]^{\frac{n}{2}} \frac{\exp(-\frac{n}{2})}{\exp(-\frac{n}{2})} = \left[ \frac{\| \underline{Y} - p_M(\underline{Y}) \|^2}{\| \underline{Y} - p_H(\underline{Y}) \|^2} \right]^{\frac{n}{2}} . \end{aligned}$$

Heraf fås

$$Q < q \Leftrightarrow \frac{\| \underline{Y} - p_M(\underline{Y}) \|^2}{\| \underline{Y} - p_H(\underline{Y}) \|^2} < k_1 .$$

Da vi forkaster hypotesen for små værdier af  $Q$  ses det, at vi netop forkaster, når længden af kateten  $\underline{Y} - p_M(\underline{Y})$  er meget mindre end længden af hypotenusen. Da - ifølge Pythagoras' læresætning -

$$\| \underline{Y} - p_H(\underline{Y}) \|^2 = \| \underline{Y} - p_M(\underline{Y}) \|^2 + \| p_H(\underline{Y}) - p_M(\underline{Y}) \|^2 ,$$

ser vi, at man lige så godt kan sammenligne kateterne i.e. benytte,



$$Q < q \Leftrightarrow \frac{\|P_M(\underline{y}) - P_H(\underline{y})\|^2 / (k-r)}{\|\underline{y} - P_M(\underline{y})\|^2 / (n-k)} > c \quad (*)$$

Under  $H_0$  er såvel tæller som nævner  $\sigma^2 \chi^2(f)/f$  fordelte med henholdsvis  $k-r$  og  $n-k$  frihedsgrader, og de er ydermere uafhængige (ifølge spaltningssætningen). Kvotienten vil derfor være F-fordelt under  $H_0$ , og sætningen følger.

Grunden til, at vi i (\*) har divideret de pågældende normer med dimensionen af det relevante underrum er selvfølgelig, at vi ønsker at teststørrelsen skal være F-fordelt under  $H_0$ , og ikke blot proportional med en F-fordeling.

Q.E.D.

Man samler sædvanligt beregningerne i et variansanalyseeskema.

Variation	SAK	Frihedsgrader = dimension
Af model fra hypotese	$\ P_M(\underline{y}) - P_H(\underline{y})\ ^2$	$k-r$
Af observationer fra model	$\ \underline{y} - P_M(\underline{y})\ ^2$	$n-k$
Af observationer fra hypotese	$\ \underline{y} - P_H(\underline{y})\ ^2$	$n-r$

Bemærkning Hyppigt vil man være i den situation, at underrummene  $M$  og  $H$  er parametriseret, i.e.

$$\underline{\mu} \in M \Leftrightarrow \exists \underline{\theta} \in R^k (\underline{\mu} = \underline{x}_0 \underline{\theta})$$

$$\underline{\mu} \in H \Leftrightarrow \exists \underline{\gamma} \in R^r (\underline{\mu} = \underline{x}_0 \underline{\gamma}) ,$$

## 3.44

hvor  $\underline{x}$  og  $x_0$  er  $n \times k$  henholdsvis  $n \times r$  ( $r \leq k$ ) matricer. Vi har da, at  $p_M(\underline{Y}) = \underline{x}\hat{\theta}$  og  $p_H(\underline{Y}) = \underline{x}_0\hat{\gamma}$  beregnes ved at løse ligningerne

$$\begin{aligned}(\underline{x}'\underline{\Sigma}^{-1}\underline{x})\hat{\theta} &= \underline{x}'\underline{\Sigma}^{-1}\underline{Y} \\ (\underline{x}_0'\underline{\Sigma}^{-1}\underline{x}_0)\hat{\gamma} &= \underline{x}_0'\underline{\Sigma}^{-1}\underline{Y}\end{aligned}$$

med hensyn til  $\hat{\theta}$  og  $\hat{\gamma}$ .

Vi betragter nu igen modellen fra p. 3.10.

Eksempel 3.8 Vi har modellen

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + \underline{\varepsilon} .$$

De observerede data var  $\underline{y}' = (10.11, 0.81, 5.24)$ .

Vi ønsker at teste hypotesen

$$H_0 : \theta_2 = 0 \text{ mod } H_1 : \theta_2 \neq 0 .$$

Vi omformulerer hypotesen til

$$H_0 : E(\underline{Y}) = \begin{bmatrix} 1 \\ 0 \\ \frac{1}{2} \end{bmatrix} \theta_1 = \begin{bmatrix} 1 \\ 0 \\ \frac{1}{2} \end{bmatrix} \gamma .$$

Estimatoren for  $\gamma$  er

$$\hat{\gamma} = [(1 \ 0 \ \frac{1}{2}) \begin{bmatrix} 1 \\ 0 \\ \frac{1}{2} \end{bmatrix}]^{-1} [(1 \ 0 \ \frac{1}{2}) \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}] = \frac{4}{5}Y_1 + \frac{2}{5}Y_3 .$$

Den observerede værdi er  $\hat{y} = 10.184$ . Heraf fås

$$\underline{x}_0 \hat{y} = \begin{bmatrix} 1 \\ 0 \\ \frac{1}{2} \end{bmatrix} 10.184 = \begin{bmatrix} 10.184 \\ 0 \\ 5.092 \end{bmatrix} ,$$

og

$$\| \underline{y} - \underline{x}_0 \hat{y} \|^2 = (\underline{y} - \underline{x}_0 \hat{y})' (\underline{y} - \underline{x}_0 \hat{y}) = 0.6835 .$$

Da vi havde (p. 3.12)

$$\| \underline{y} - \underline{x} \hat{\theta} \|^2 = (\underline{y} - \underline{x} \hat{\theta})' (\underline{y} - \underline{x} \hat{\theta}) = 0.0338 ,$$

får vi

$$\| \underline{x} \hat{\theta} - \underline{x}_0 \hat{y} \|^2 = 0.6835 - 0.0338 = 0.6497 .$$

Følgelig er teststørrelsen

$$\frac{\| \underline{x} \hat{\theta} - \underline{x}_0 \hat{y} \|^2 / (2-1)}{\| \underline{y} - \underline{x} \hat{\theta} \|^2 / (3-2)} = 19.2 < F(1,1)_{0.90} ,$$

og vi accepterer altså hypotesen, i hvert fald for  $\alpha < 10\%$ .

Forklaring på frihedsgraderne:

$$\text{rg } \underline{x} = \text{rg} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} = 2 = k$$

$$\text{rg } \underline{x}_0 = \text{rg} \begin{bmatrix} 1 \\ 0 \\ \frac{1}{2} \end{bmatrix} = 1 = r$$

$$n = 3 \quad .$$

□

Vi betragter dernæst en fortsættelse til eksempel 3.5 p 3.18.

Eksempel 3.9 Det forekommer ud fra problemstillingen rimeligt at antage, at parameteren  $\theta_{21} = 0$ . Vi vil derfor teste hypotesen

$$H_0 : \theta_{21} = 0 \text{ mod } H_1 : \theta_{21} \neq 0 .$$

Hypoteserummet H er derfor givet ved, at

$$E(\underline{y}) = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \theta_{11} \\ \theta_{12} \end{bmatrix} = \begin{bmatrix} \mu_1 + \theta_{11} \\ \mu_1 + \theta_{11} \\ \mu_1 + \theta_{12} \\ \mu_1 + \theta_{12} \\ 0 \\ 0 \\ 0 \end{bmatrix} .$$

Vi finder nu

$$\underline{x}'_1 \underline{x}_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix} ,$$

og

$$\underline{x}'_1 \underline{y} = \begin{bmatrix} y_{1..} \\ y_{11.} \\ y_{12.} \end{bmatrix} .$$

Vi ser at  $\underline{x}'_1 \underline{x}_1$  er singular, og vi indfører det lineære bånd

$$\underline{b}'\underline{\theta} = (0 \ 1 \ 1) \begin{bmatrix} \mu_1 \\ \theta_{11} \\ \theta_{12} \end{bmatrix} = \theta_{11} + \theta_{12} = 0 .$$

Da

$$\underline{b}'\underline{b} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} ,$$

fås

$$\underline{x}'\underline{x} + \underline{b}'\underline{b} = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 3 & 1 \\ 2 & 1 & 3 \end{bmatrix} .$$

Denne matrix er inverteret p. 3.18. Vi finder derfor estimatoren under  $H_0$  til

$$\hat{\theta}_1 = \begin{bmatrix} 1/2 & -1/4 & -1/4 \\ -1/4 & 1/2 & 0 \\ -1/4 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} Y_{1..} \\ Y_{11.} \\ Y_{12.} \end{bmatrix} = \begin{bmatrix} \bar{Y}_{1..} \\ \bar{Y}_{11.} - \bar{Y}_1 \\ Y_{12.} - \bar{Y}_1 \end{bmatrix} .$$

Den observerede værdi er  $(-9, +8, -8)'$ . Den nye residualvektor er

$$Y - \underline{x}_1 \hat{\theta}_1 = (1, -1, -2, +2, -6, 0, -2)' .$$

Normen af denne vektor er 50, og antallet af frihedsgrader er  $7-2 = 5$ . Vi finder derfor, at

$$\begin{aligned} \| P_M(\underline{Y}) - P_H(\underline{y}) \|^2 &= \| y - P_H(\underline{y}) \|^2 - \| y - P_M(\underline{y}) \|^2 \\ &= 50 - 28\frac{2}{3} = 21\frac{1}{3} . \end{aligned}$$

Vi kan nu samle beregningerne i nedenstående variansanalyse-skema.

Variation	SAK	f	S <sup>2</sup>	Test
M - H	$21\frac{1}{3}$	3-2=1	$21\frac{1}{3}$	2.97
O - M	$28\frac{2}{3}$	7-3=4	$7\frac{1}{6}$	
O - H	50	7-2=5		

Da den observerede værdi af teststørrelsen  $2.97 < F(1,4)_{.90}$  vil vi acceptere hypotesen, og vi vil derfor gå ud fra, at  $H_0$  er sand.

□

### 3.2.2 Successiv testning i den generelle lineære model

Vi vil i dette afsnit illustrere fremgangsmåden ved den test-procedure man bør følge, når man successivt ønsker at undersøge om middelværdivektoren for ens observationer ligger i underrum  $H_1$  med

$$H_0 \supseteq H_1 \supseteq H_2 \supseteq \dots \supseteq H_m, \quad m \leq k.$$

Som udgangspunkt vælger vi at betragte følgende tal over udbyttet ved penicillingæring under anvendelse af 2 forskellige sukkerarter laktose og rørsukker ved koncentrationerne 2%, 4%, 6% og 8%. (i g./100 ml.)

		Faktor B: koncentration			
		2%	4%	6%	8%
Faktor A:	Laktose	0.606	0.660	0.984	0.908
	Rørsukker	0.761	0.933	1.072	0.979

Tallene stammer fra Davies (1967) (p.314). Udbyttet er udtrykt ved logaritmen til mycelievægten efter en uges vækst.

Vi er nu interesserede i at undersøge de to faktorer A og B's indflydelse på det færdige resultat (udbytte). Vi antager, at observationerne er stokastisk uafhængige og normalt fordelte. Kaldes de

$$L : Y_{11}, Y_{12}, Y_{13}, Y_{14}$$

og

$$R : Y_{21}, Y_{22}, Y_{23}, Y_{24} ,$$

vil vi endvidere antage, at

$$E(Y_{ij}) = \alpha_i + \beta_i x_j + \gamma_i x_j^2$$

hvor  $x_j$  angiver  $j$ 'te sukkerkoncentration. Vi foretager en skalaændring af sukkerkoncentrationen

$$2\% \leftrightarrow -3$$

$$4\% \leftrightarrow -1$$

$$6\% \leftrightarrow 1$$

$$8\% \leftrightarrow 3 ,$$

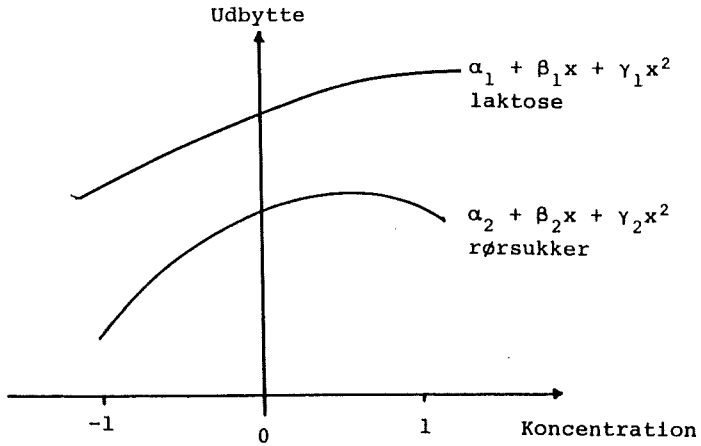
eller - mere stringent - definerer  $x$  ved

$$x_j = \frac{x_j' - 5\%}{1\%} .$$

Vi får da følgende udtryk for middelværdierne

$$E(Y_{ij}) = \alpha_i + \beta_i x_j + \gamma_i x_j^2 .$$

Vi antager altså, at udbyttet inden for de givne grænser kan beskrives ved andengradspolynomier.



Man kan nu f.eks. successivt undersøge

- 1) om  $\gamma_1 = \gamma_2 = 0$ , d.v.s. om en beskrivelse ved affine funktioner er tilstrækkelig,
  - 2) - hvis dette accepteres - om  $\beta_1 = \beta_2 = \beta$ , d.v.s. om margineffekten ved at øge koncentrationen er den samme for de to sukkerarter,
  - 3) - hvis dette accepteres - om  $\alpha_1 = \alpha_2 = \alpha$ , d.v.s. om de to sukkerarter er ens hvad angår udbyttet, og hvis dette accepteres,
  - 4) om  $\beta = 0$ , d.v.s. om koncentrationen overhovedet har nogen indflydelse.
- i) Vi skriver først modellen op i matrixform



$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{24} \end{bmatrix} = \begin{bmatrix} 1 & -3 & 9 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 3 & 9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -3 & 9 \\ 0 & 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 3 & 9 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \gamma_1 \\ \alpha_2 \\ \beta_2 \\ \gamma_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \end{bmatrix},$$

eller

$$\underline{Y} = \underline{X}\theta + \underline{\epsilon}.$$

Vi finder

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ -3 & -1 & 1 & 3 & 0 & 0 & 0 & 0 \\ 9 & 1 & 1 & 9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & -3 & -1 & 1 & 3 \\ 0 & 0 & 0 & 0 & 9 & 1 & 1 & 9 \end{bmatrix} \begin{bmatrix} 1 & -3 & 9 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 3 & 9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -3 & 9 \\ 0 & 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 3 & 9 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 20 & 0 & 0 & 0 \\ 0 & 20 & 0 & 0 & 0 & 0 \\ 20 & 0 & 164 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 20 \\ 0 & 0 & 0 & 0 & 20 & 0 \\ 0 & 0 & 0 & 20 & 0 & 164 \end{bmatrix}.$$

Da

$$\begin{bmatrix} 4 & 0 & 20 \\ 0 & 20 & 0 \\ 20 & 0 & 164 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{41}{64} & 0 & -\frac{5}{64} \\ 0 & \frac{1}{20} & 0 \\ -\frac{5}{64} & 0 & \frac{1}{64} \end{bmatrix},$$

er

$$(\underline{\underline{x}}' \underline{\underline{x}})^{-1} = \begin{bmatrix} \frac{41}{64} & 0 & -\frac{5}{64} & 0 & 0 & 0 \\ 0 & \frac{1}{20} & 0 & 0 & 0 & 0 \\ -\frac{5}{64} & 0 & \frac{1}{64} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{41}{64} & 0 & -\frac{5}{64} \\ 0 & 0 & 0 & 0 & \frac{1}{20} & 0 \\ 0 & 0 & 0 & -\frac{5}{64} & 0 & \frac{1}{64} \end{bmatrix}.$$

Følgelig er

$$\hat{\underline{\underline{\theta}}} = \begin{bmatrix} -\frac{1}{16}y_{11} + \frac{9}{16}y_{12} + \frac{9}{16}y_{13} - \frac{1}{16}y_{14} \\ -\frac{3}{20}y_{11} - \frac{1}{20}y_{12} + \frac{1}{20}y_{13} + \frac{3}{20}y_{14} \\ \frac{1}{16}y_{11} - \frac{1}{16}y_{12} - \frac{1}{16}y_{13} + \frac{1}{16}y_{14} \\ -\frac{1}{16}y_{21} + \frac{9}{16}y_{22} + \frac{9}{16}y_{23} - \frac{1}{16}y_{24} \\ \frac{3}{20}y_{21} - \frac{1}{20}y_{22} - \frac{1}{20}y_{23} + \frac{3}{20}y_{24} \\ \frac{1}{16}y_{21} - \frac{1}{16}y_{22} - \frac{1}{16}y_{23} + \frac{1}{16}y_{24} \end{bmatrix} = \begin{bmatrix} 0.830 \\ 0.062 \\ -0.008 \\ 1.019 \\ 0.040 \\ -0.017 \end{bmatrix}.$$

Modellen svarer til et 6-dimensionalt underrom  $M$  i  $\mathbb{R}^8$  ( $\text{rg } \underline{\underline{x}} = 6$ ), og vi har - idet vi regner med normen svarende til  $\underline{\underline{I}}$  - at projektionen på  $M$  er

$$P_M(\underline{\underline{y}}) = \underline{\underline{x}} \hat{\underline{\underline{\theta}}} = \begin{bmatrix} 1 & -3 & 9 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 3 & 9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -3 & 9 \\ 0 & 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 3 & 9 \end{bmatrix} \begin{bmatrix} 0.830 \\ 0.062 \\ -0.008 \\ 1.019 \\ 0.040 \\ -0.017 \end{bmatrix} = \begin{bmatrix} 0.572 \\ 0.760 \\ 0.884 \\ 0.944 \\ 0.746 \\ 0.962 \\ 1.042 \\ 0.986 \end{bmatrix}.$$

Vi har derfor residualerne

$$\underline{Y} - P_M(\underline{Y}) = \begin{bmatrix} 0.034 \\ -0.100 \\ 0.100 \\ -0.036 \\ 0.015 \\ -0.029 \\ 0.030 \\ -0.007 \end{bmatrix} .$$

Den kvadrerede længde af denne vektor er

$$\| \underline{Y} - P_M(\underline{Y}) \|^2 = 0.034^2 + \dots + (-0.007)^2 = 0.024467 .$$

Som skøn over  $\sigma^2$  kan derfor anvendes

$$\hat{\sigma}^2 = \frac{1}{8-6} 0.024467 = 0.0122335 .$$

ii) Hvis hypotesen  $\underline{\mu} \in H_1$ , d.v.s.  $\gamma_1 = \gamma_2 = 0$ , eller

$$\underline{Y} = \begin{bmatrix} 1 & -3 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \alpha_2 \\ \beta_2 \end{bmatrix} + \underline{\varepsilon}_1 = \underline{x}_1 \underline{\delta}_1 + \underline{\varepsilon}_1 ,$$

er sand, får vi estimerterne

$$\hat{\underline{\delta}}_1 = (\underline{x}_1' \underline{x}_1)^{-1} \underline{x}_1' \underline{Y} = \begin{bmatrix} \frac{1}{4}y_{11} + \frac{1}{4}y_{12} + \frac{1}{4}y_{13} + \frac{1}{4}y_{14} \\ -\frac{3}{20}y_{11} - \frac{1}{20}y_{12} + \frac{1}{20}y_{13} + \frac{3}{20}y_{14} \\ \frac{1}{4}y_{21} + \frac{1}{4}y_{22} + \frac{1}{4}y_{23} + \frac{1}{4}y_{24} \\ -\frac{3}{20}y_{21} - \frac{1}{20}y_{22} + \frac{1}{20}y_{23} + \frac{3}{20}y_{24} \end{bmatrix} = \begin{bmatrix} 0.790 \\ 0.062 \\ 0.936 \\ 0.040 \end{bmatrix} .$$

Residualerne bliver

$$\underline{Y} - p_{H_1}(\underline{Y}) = \underline{Y} - \underline{x}_1 \hat{\delta}_1 = \begin{bmatrix} 0.002 \\ -0.068 \\ 0.132 \\ -0.068 \\ -0.055 \\ 0.037 \\ 0.096 \\ -0.077 \end{bmatrix} .$$

Den kvadrerede længde af denne vektor er

$$\| \underline{Y} - p_{H_1}(\underline{Y}) \|^2 = 0.002^2 + \dots + (-0.077)^2 = 0.046215 .$$

iii) Hvis  $\underline{\mu} \in H_2$ , d.v.s.  $\beta_1 = \beta_2 = \beta$ , er modellen

$$\underline{Y} = \begin{bmatrix} 1 & 0 & -3 \\ 1 & 0 & -1 \\ 1 & 0 & 1 \\ 1 & 0 & 3 \\ 0 & 1 & -3 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta \end{bmatrix} + \underline{\varepsilon}_2 = \underline{x}_2 \underline{\delta}_2 + \underline{\varepsilon}_2 .$$

Estimaterne bliver

$$\hat{\delta}_2 = (\underline{x}_2' \underline{x}_2)^{-1} \underline{x}_2' \underline{Y} = \begin{bmatrix} 0.790 \\ 0.936 \\ 0.051 \end{bmatrix} ,$$

og residualerne

$$Y - P_{H_2}(Y) = \begin{bmatrix} -0.031 \\ -0.079 \\ 0.143 \\ -0.035 \\ -0.022 \\ 0.048 \\ 0.085 \\ -0.110 \end{bmatrix} .$$

Den kvadrerede norm af residualvektoren er

$$\|Y - P_{H_2}(Y)\|^2 = (-0.031)^2 + \dots + (-0.110)^2 = 0.050989 .$$

- iv) Hvis  $\underline{\mu} \in H_3$ , d.v.s.  $\beta_1 = \beta_2 = \beta$  og  $\alpha_1 = \alpha_2 = \alpha$ , er modellen

$$Y = \begin{bmatrix} 1 & -3 \\ 1 & -1 \\ 1 & 1 \\ 1 & 3 \\ 1 & -3 \\ 1 & -1 \\ 1 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \underline{\varepsilon}_3 = \underline{x}_3 \underline{\delta}_3 + \underline{\varepsilon}_3 .$$

Vi finder

$$\hat{\underline{\delta}}_3 = (\underline{x}_3' \underline{x}_3)^{-1} \underline{x}_3' Y = \begin{pmatrix} 0.863 \\ 0.051 \end{pmatrix} ,$$

og

$$\underline{Y} - p_{H_3}(\underline{Y}) = \begin{bmatrix} -0.104 \\ -0.152 \\ 0.070 \\ -0.108 \\ 0.051 \\ 0.121 \\ 0.158 \\ -0.037 \end{bmatrix},$$

hvorfor

$$\|\underline{Y} - p_{H_3}(\underline{Y})\|^2 = 0.094059.$$

v) Endelig betragtes tilfældet  $\underline{\mu} \in H_4$ , d.v.s.  $\beta = 0$ , eller

$$\underline{Y} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \alpha = \underline{x}_4 \delta_4 + \underline{\varepsilon}_4.$$

Vi finder da

$$\hat{\delta}_4 = \hat{\alpha} = (\underline{x}_4' \underline{x}_4)^{-1} \underline{x}_4' \underline{Y}' = 0.863,$$

hvorfor

$$\underline{Y} - p_{H_4}(\underline{Y}) = \begin{bmatrix} -0.250 \\ -0.203 \\ 0.121 \\ 0.045 \\ -0.102 \\ 0.070 \\ 0.209 \\ 0.116 \end{bmatrix},$$

og

$$\| \underline{Y} - p_{H_4}(\underline{Y}) \|^2 = 0.196365 .$$

Idet vi sætter  $\text{rg}(\underline{x}_i) = r_i$  og  $\text{rg}(\underline{x}) = k$  kan vi sammenfatte testningen i et variansanalysekema som

Variation	SAK	Frihedsgrader = dimension
$H_4 - H_3$	$\  p_{H_4}(\underline{Y}) - p_{H_3}(\underline{Y}) \ ^2$	$r_3 - r_4 = 2 - 1 = 1$
$H_3 - H_2$	$\  p_{H_3}(\underline{Y}) - p_{H_2}(\underline{Y}) \ ^2$	$r_2 - r_3 = 3 - 2 = 1$
$H_2 - H_1$	$\  p_{H_2}(\underline{Y}) - p_{H_1}(\underline{Y}) \ ^2$	$r_1 - r_2 = 4 - 3 = 1$
$H_1 - M$	$\  p_{H_1}(\underline{Y}) - p_M(\underline{Y}) \ ^2$	$k - r_1 = 6 - 4 = 2$
$M - \text{obs.}$	$\  p_M(\underline{Y}) - \underline{Y} \ ^2$	$n - k = 8 - 6 = 2$
$H_4 - \text{obs.}$	$\  p_{H_4}(\underline{Y}) - \underline{Y} \ ^2$	$n - r_4 = 8 - 1 = 7$

Dette skema er en simpel udvidelse af skemaet p. 3.43. Vi kan anvende spaltningssætningen og får - under de forskellige hypoteser - at SAK'erne er uafhængige og er fordelt som  $\sigma^2 \chi^2$  med de anførte frihedsgrader.

Hvis en hypotese  $H_i$  er accepteret bliver teststørrelsen for test af  $H_{i+1}$

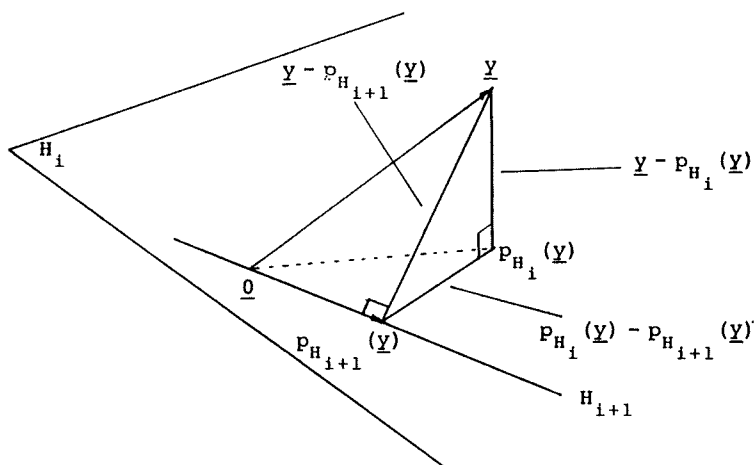
$$\frac{\| p_{H_{i+1}}(\underline{Y}) - p_{H_i}(\underline{Y}) \|^2 / (r_i - r_{i+1})}{\| p_{H_i}(\underline{Y}) - \underline{Y} \|^2 / (n - r_i)} .$$

Under hypotesen er denne størrelse  $F(r_i - r_{i+1}, n - r_i)$  - for-

delt (ifølge spaltningssætningen), og vi forkaster - stadig ifølge teorien fra forrige afsnit - for store værdier af  $Z$ , d.v.s. for

$$Z > F(r_i - r_{i+1}, n - r_i)_{1-\alpha} .$$

Inden vi konkret går i gang med testningen, vil det være på sin plads at anføre nogle beregningsformler. Vi betragter overgangen fra  $H_i$  til  $H_{i+1} \subset H_i$ .



Ved hjælp af Pythagoras sætning ser vi nu, at der findes to alternative beregningsmåder for

$$z = \| p_{H_{i+1}}(\underline{y}) - p_{H_i}(\underline{y}) \|^2 ,$$

nemlig

$$z = \| p_{H_i}(\underline{y}) \|^2 - \| p_{H_{i+1}}(\underline{y}) \|^2 \quad (1)$$



og

$$z = \| \underline{y} - P_{H_{i+1}}(\underline{y}) \|^2 - \| \underline{y} - P_{H_i}(\underline{y}) \|^2 . \quad (ii)$$

Af disse må den første foretrækkes af numeriske grunde, men hvis man alligevel har beregnet residualernes kvadratsummer er det åbenbart lettest at bruge (ii).

Variansanalyseeskemaet bliver i vores tilfælde

Variation	SAK	f	Teststørrelse
$H_4 - H_3$	0.102306	1	$\frac{0.102306/1}{0.094059/6} = 5.44$
$H_3 - H_2$	0.043070	1	$\frac{0.043070/1}{0.050981/5} = 4.22$
$H_2 - H_1$	0.004774	1	$\frac{0.004774/1}{0.046215/4} = 0.41$
$H_1 - M$	0.021748	2	$\frac{0.021748/2}{0.024467/2} = 0.89$
$M - \text{obs}$	0.024467	2	
$\text{Obs} - H_4$	0.196365	7	

Da

$$4.22 \approx F(1,5)_{.91} ,$$

og

$$5.44 \approx F(1,6)_{.94} ,$$

vil man ikke - ved test på e.g. niveau  $\alpha = 5\%$  - afvise nogen af hypoteserne  $H_1$ ,  $H_2$ ,  $H_3$  eller  $H_4$ .

Bemærk: at vi selvfølgelig ikke ville teste e.g.  $H_2$ , hvis vi havde fået forkastet  $H_1$ , idet  $H_2$  jo er en delhypotese af  $H_1$ .

Konklusionen må derfor blive, at vi - indtil evt. nye undersøgelser afkræfter dette - vil arbejde med den model, at udbyttet  $Y$  ved penicillingæringen er uafhængig af sukkerart og den koncentration ( $2\% \leq \text{konc.} \leq 8\%$ ) ved hvilken gæringen sker. Vi har med

$$E(Y) = \alpha \quad \text{og} \quad V(Y) = \sigma^2 ,$$

at

$$\hat{\alpha} = 0.863 ,$$

og

$$\hat{\sigma}^2 = \frac{0.196365}{7} = 0.028052 \approx 0.17^2 .$$

Endvidere er

$$V(\hat{\alpha}) = \frac{\sigma^2}{8} \approx \frac{\hat{\sigma}^2}{8} = 0.0035 \approx 0.059^2 .$$

### Referencer til kapitel 3

Davies, O.L. (ed.): Design and Analysis of Industrial Experiments (sec. ed.). Oliver and Boyd, London 1967.

Kendall, M.G., & Stuart, A.: The Advanced Theory of Statistics, vol. 2. Charles Griffin & Co., London 1967.

Rao, C.R.: Linear Statistical Inference and Its Applications. John Wiley & Sons, New York 1965.

## KAPITEL 4

### Regressionsanalyse

I dette kapitel giver vi en oversigt over regressionsanalysen. Det meste fremtræder som specialtilfælde af den generelle lineære model, men da en række anvendelser ofte er knyttet til regressionsituationer, vil vi beskrive resultaterne i dette sprog.

Der er anført et lille afsnit om ortogonal regression (ikke at forveksle med regression efter ortogonale polynomier). Rent statistisk hører dette sådan set hjemme under afsnittet om principale komponenter og faktoranalysen, og hvad angår beregningsmetoder, henvises også til dette afsnit. Jeg har dog ud fra et "curve-fitting" synspunkt fundet det formålstjenligt også at nævne begrebet her i dette kapitel.

#### 4.1 Lineær regressionsanalyse

I dette afsnit vil den lineære regressionsanalyse blive analyseret v.h.a. teorien for den generelle lineære model. Vi indleder med

4.1.1 Notation og model

I den almindelige regressionsanalyse arbejder vi med modellen

$$E(Y) = \alpha + \beta_1 x_1 + \dots + \beta_k x_k ,$$

hvor  $x$ 'erne er kendte størrelser og  $\beta$ 'erne (og  $\alpha$ ) er ukendte parametre. Hvis vi har givet  $n$  observationer af  $Y$  kan vi mere præcist skrive modellen

$$\begin{bmatrix} Y_1 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{k1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 1 & x_{1n} & \dots & x_{kn} \end{bmatrix} \cdot \begin{bmatrix} \alpha \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_n \end{bmatrix} ,$$

eller

$$\underline{Y} = \underline{X}\underline{\beta} + \underline{\epsilon} .$$

Vi forudsætter som sædvanligt, at

$$D(\underline{\epsilon}) = \sigma^2 \underline{\Sigma} ,$$

hvor  $\underline{\Sigma}$  er kendt og  $\sigma^2$  (oftest) ukendt.

Estimatorerne findes på sædvanlig vis ved at løse normal-  
ligningerne

$$\underline{X}' \underline{\Sigma}^{-1} \underline{X} \underline{\beta} = \underline{X}' \underline{\Sigma}^{-1} \underline{Y} ,$$

eller, hvis  $\underline{\Sigma} = \underline{I}$

$$\underline{X}' \underline{X} \hat{\underline{\beta}} = \underline{X}' \underline{Y} .$$

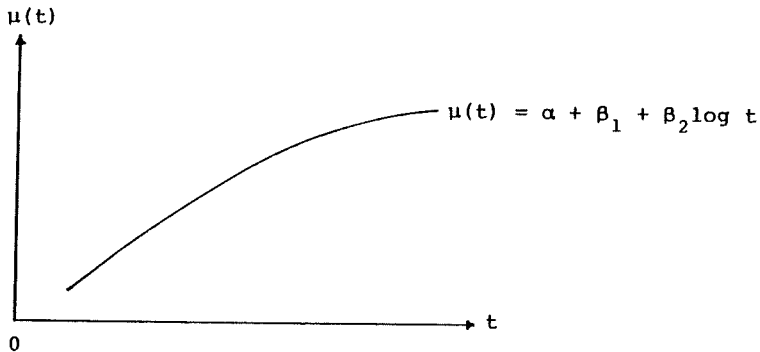
I det første tilfælde taler vi om en vægtet regressionsanalyse.

Inden vi går videre, vil det nok være hensigtsmæssigt endnu engang at præcisere, hvad der menes med ordet lineær i vendingen lineær regressionsanalyse.

Meningen er - som i den almindelige lineære model - at der er tale om linearitet i parametrene. Vi kan sagtens lave regression efter e.g. tiden og logaritmen til tiden. Modellen er da blot

$$E(Y) = \alpha + \beta_1 t + \beta_2 \log t ,$$

jvf. eksempel 3.2.



Med  $n$  observationer bliver modellen på matrixform

$$\begin{bmatrix} Y_1 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & t_1 & \log t_1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & t_n & \log t_n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_n \end{bmatrix} .$$

En anden banal ting, det kan være nyttigt at pointere, er, at man kan "tvinge" regressionsfladen gennem 0 ved at stryge  $\alpha$ 'et og 1'ste søjle i  $\underline{x}$ -matricen, i.e. anvende modellen

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} x_{11} \cdots x_{k1} \\ \vdots \\ x_{1n} \cdots x_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix} .$$

Det kan være nyttigt at bemærke, at man kan anvende følgende trick, når man ønsker, at regressionsfladen skal gå gennem  $\underline{0}$ . Vi forudsætter, at  $\underline{\underline{\epsilon}} = \underline{\underline{I}}$ .

Vi betragter observationerne  $Y_1, \dots, Y_n$  og de tilsvarende værdier af de uafhængige variable  $1, x_{i1}, \dots, x_{ik}$ ,  $i = 1, \dots, n$ . Hvis vi hertil føjer  $-Y_1, \dots, -Y_n$  og  $1, -x_{i1}, \dots, -x_{ik}$ ,  $i = 1, \dots, n$ , og skriver den sædvanlige model op, fås

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \\ -Y_1 \\ \vdots \\ -Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} \cdots x_{k1} \\ \vdots \\ 1 & x_{1n} \cdots x_{kn} \\ 1 & -x_{11} \cdots -x_{k1} \\ \vdots \\ 1 & -x_{1n} \cdots -x_{kn} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \underline{\underline{\epsilon}} ,$$

eller mere kompakt skrevet

$$\begin{bmatrix} \underline{Y} \\ -\underline{Y} \end{bmatrix} = \begin{bmatrix} \underline{1} & \underline{X} \\ \underline{1} & -\underline{X} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \underline{\underline{\epsilon}} ,$$

hvor vi har anvendt en  $i$  forhold til notationen p. 4.2 lidt afvigende definition af  $\underline{X}$ -matricen og  $\underline{\beta}$ .

Normalligningerne bliver

$$\begin{bmatrix} \underline{1}' & \underline{1}' \\ \underline{X}' & -\underline{X}' \end{bmatrix} \begin{bmatrix} \underline{1} & \underline{X} \\ \underline{1} & -\underline{X} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \underline{1}' & \underline{1}' \\ \underline{X}' & -\underline{X}' \end{bmatrix} \begin{bmatrix} \underline{Y} \\ -\underline{Y} \end{bmatrix} ,$$

eller

$$\begin{bmatrix} 2n & 0 \\ 0 & 2\underline{\underline{x}}'\underline{\underline{x}} \end{bmatrix} \begin{bmatrix} \alpha \\ \underline{\underline{\beta}} \end{bmatrix} = \begin{bmatrix} 0 \\ 2\underline{\underline{x}}'\underline{\underline{y}} \end{bmatrix} .$$

Skrives disse ligninger ud, fås

$$\begin{aligned} 2n\alpha &= 0 \\ 2\underline{\underline{x}}'\underline{\underline{x}}\underline{\underline{\beta}} &= 2\underline{\underline{x}}'\underline{\underline{y}} , \end{aligned}$$

eller

$$\begin{aligned} \alpha &= 0 \\ \underline{\underline{x}}'\underline{\underline{x}}\underline{\underline{\beta}} &= \underline{\underline{x}}'\underline{\underline{y}} . \end{aligned}$$

Med andre ord får vi på denne måde bestemt estimatorerne til koefficienterne i en regressionsflade, der er tvunget gennem 0.

Grunden til, at ovenstående ofte er anvendeligt, er, at man i en mængde standardprogrammer ikke kan tvinge fladen gennem 0. Ved hjælp af ovenstående lille trick kan problemet så omgås.

Output fra et sådant program må dog tolkes med varsomhed, idet alle SAK'er er dobbelt så store, som de bør være. E.g. vil residualkvadratafvigelsessummen være beregnet som

$$\begin{aligned} & \left( \begin{bmatrix} \underline{\underline{y}} \\ -\underline{\underline{y}} \end{bmatrix} - \begin{bmatrix} \underline{\underline{x}} \hat{\underline{\underline{\beta}}} \\ -\underline{\underline{x}} \hat{\underline{\underline{\beta}}} \end{bmatrix} \right)' \left( \begin{bmatrix} \underline{\underline{y}} \\ -\underline{\underline{y}} \end{bmatrix} - \begin{bmatrix} \underline{\underline{x}} \hat{\underline{\underline{\beta}}} \\ -\underline{\underline{x}} \hat{\underline{\underline{\beta}}} \end{bmatrix} \right) \\ &= ((\underline{\underline{y}} - \underline{\underline{x}}\hat{\underline{\underline{\beta}}})', [-\underline{\underline{y}} + \underline{\underline{x}}\hat{\underline{\underline{\beta}}}]') \begin{bmatrix} \underline{\underline{y}} - \underline{\underline{x}}\hat{\underline{\underline{\beta}}} \\ -\underline{\underline{y}} + \underline{\underline{x}}\hat{\underline{\underline{\beta}}} \end{bmatrix} \\ &= 2[\underline{\underline{y}} - \underline{\underline{x}}\hat{\underline{\underline{\beta}}}]'[\underline{\underline{y}} - \underline{\underline{x}}\hat{\underline{\underline{\beta}}}] , \end{aligned}$$

i.e det dobbelte af den korrekte residualkvadratafvigelsessum.

De angivne frihedsgrader vil heller ikke være korrekte. Man må da selv opstille en almindelig lineær model og finde de korrekte frihedsgrader ved dimensionsbetragtninger.

### 4.1.2 Korrelation og regression

I sætning 2.23 p. 2.48 er anført et resultat, der kan benyttes ved et test for, om den multiple korrelationskoefficient mellem normalt fordelte variable er 0. Vi skal nu vise, at dette resultat har en sammenhæng med et test i en regressionsmodel.

Vi antager, at vi har den sædvanlige model p. 4.2, og vi antager, at  $\underline{\Sigma} = \underline{I}$ .

Vi kan nu uden videre anvende teorien fra kapitel 3 til at teste diverse hypoteser om parametrene  $\alpha, \beta_1, \dots, \beta_k$ .

Ved formelle regninger kan vi estimere den multiple korrelationskoefficient mellem  $Y$  og  $x_1, \dots, x_k$  ved hjælp af de udtryk, som er anført i afsnit 2.3.2.

Det kan vises, at vi får

$$R^2 = \frac{\|\underline{y} - p_0(\underline{y})\|^2 - \|\underline{y} - p_M(\underline{y})\|^2}{\|\underline{y} - p_0(\underline{y})\|^2},$$

hvor

$$p_0(\underline{y}) = \begin{bmatrix} \bar{y} \\ \cdot \\ \cdot \\ \cdot \\ \bar{y} \end{bmatrix} (= \underline{x} \cdot \hat{\underline{\beta}}),$$

og

$$p_M(\underline{y}) = \underline{x}\underline{\hat{\beta}} = \hat{E}(\underline{y}).$$

Disse resultater er ikke særligt overraskende. Vi erindrer, at den multiple korrelationskoefficient bl.a. kan fås frem ved at finde den linearkombination af  $\underline{x}$ , der minimaliserer varian-



sen af  $(Y - \underline{\alpha}'\underline{X})$ , og dette svarer jo præcis til at opskrive betingelsen for mindste kvadraters estimater.

Sætter vi

$$SAK_{\text{tot}} = \|\underline{Y} - P_0(\underline{Y})\|^2 = \sum_i (Y_i - \bar{Y})^2,$$

og

$$SAK_{\text{res}} = \|\underline{Y} - \underline{\hat{\beta}}\|^2 = \sum_i (Y_i - \hat{E}(Y_i))^2,$$

kan vi skrive

$$R^2 = \frac{SAK_{\text{tot}} - SAK_{\text{res}}}{SAK_{\text{tot}}},$$

d.v.s. den multiple korrelationskoefficient kan også her udtrykkes som den del af den totale variation i Y'erne, som vi har fået forklaret ved hjælp af de uafhængige variable.

En lignende refortolkning af de partielle korrelationer er naturligvis også mulig.

Vi ser i øvrigt, at hvis vi formelt opskriver det p. 2.48 anførte test for  $\rho_{Y|x_1, \dots, x_p} = 0$ , får vi

$$\begin{aligned} \frac{R^2}{1-R^2} \frac{n-k-1}{k} &= \frac{\|\underline{Y} - P_0(\underline{Y})\|^2 - \|\underline{Y} - P_M(\underline{Y})\|^2}{\|\underline{Y} - P_M(\underline{Y})\|^2} \frac{n-k-1}{k} \\ &= \frac{\|P_M(\underline{Y}) - P_0(\underline{Y})\|^2 / k}{\|\underline{Y} - P_M(\underline{Y})\|^2 / (n-k-1)} \\ &= \frac{(SAK_{\text{tot}} - SAK_{\text{res}}) / k}{SAK_{\text{res}} / (n-k-1)} \end{aligned}$$

Ifølge den almindelige teori (p. 3.42) er dette netop teststørrelsen for hypotesen

$$\begin{bmatrix} \alpha \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} = \begin{bmatrix} \alpha \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix},$$

og teststørrelsens fordeling er en  $F(k, n-k-1)$ -fordeling - præcis det samme, som vi fik p. 3.42.

Det er, hvad angår de numeriske forhold ved testningen, altså underordnet, om vi opfatter  $x$ 'erne som realiserede udfald af en  $k$ -dimensionalt normalt fordelt stokastisk variabel eller som faste, deterministiske størrelser, vi selv kan bestemme.

Dette punkt kan derfor holdes uden for den diskussion af forudsætningerne, vi betragter i næste afsnit.

#### 4.1.3 Analyse af forudsætninger

Hvis man for samhørende  $x$ -værdier

$$x_{1i}, \dots, x_{pi}$$

har flere observationer af  $Y$ , er det muligt at foretage de sædvanlige tests for fordelingstype (histogram, fraktildiagram,  $\chi^2$ -test etc.) og for homogenitet af varianser (Bartlett's test m.fl.). Endvidere kan vi lave runtest for tilfældighed etc. etc.

Nu er situationen den, at vi kun meget sjældent har (mere end højst et par) gentagelser for de forskellige værdier af den uafhængige variabel. Derfor er det ikke muligt at lave disse undersøgelser af forudsætninger. Man betragter i stedet resi-

dualerne

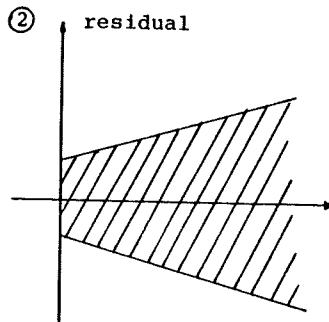
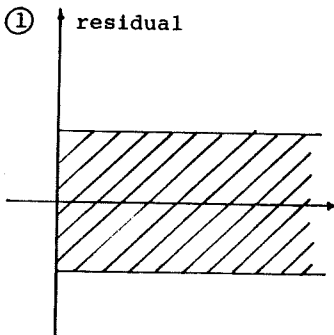
$$E_i = Y_i - \hat{E}(Y_i) = Y_i - \hat{\alpha} - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \dots - \hat{\beta}_k x_{ki} .$$

Disse vil, såfremt modellen holder, være approximativt uafhængige og  $N(0, \sigma^2)$ -fordelt.

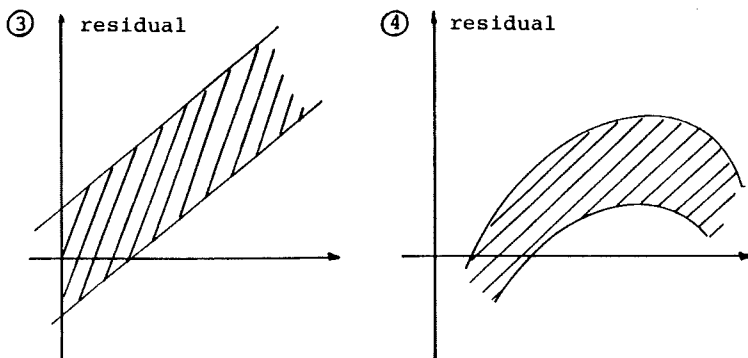
Hvis man afbilder residualerne på forskellige måder og derved får noget frem, der ikke ser ud til at være (eller kunne være) realiserede udfald af indbyrdes uafhængige  $N(0, \sigma^2)$ -fordelte stokastiske variable, har vi fået en indikation for, at der er noget i vejen med modellen.

Man vil vel oftest starte med en sædvanlig fordelingsanalyse af residualerne, i.e. lave runtest, tegne histogram, fraktildiagram etc.

Dernæst kan man afbilde residualerne mod forskellige størrelser (tiden, uafhængige variable etc.). Vi giver følgende 4 skitser til illustration af hyppigt forekommende residualplots.



## 4.10



Vi vil nu give en kort beskrivelse af hvad årsagen til plots af disse udseender kan være. Vi konstaterer først, at 1 altid må anses for acceptabelt (jvf. dog p. 4.12).

i) Plot af residualer mod tiden

- ② Variansen vokser med tiden. Lav en vægtet analyse.
- ③ Mangler led af formen  $\beta \cdot \text{tid}$
- ④ Mangler led af formen  $\beta_1 \cdot \text{tid} + \beta_2 \cdot \text{tid}^2$

ii) Plot af residualer mod  $\hat{E}(Y_i)$ 

- ② Variansen vokser med  $E(Y_i)$ . Lav en vægtet analyse eller transformer  $Y$ 'erne (e.g. med logaritme eller lign.)
- ③ Manglende konstantled (regressionen muligvis fejlagtigt tvunget gennem 0). Fejl i analyse.
- ④ Dårlig model. Forsøg med transformation af  $Y$ 'erne.

iii) Plot mod uafhængig variabel  $x_i$

- ② Variansen vokser med  $x_i$ . Lav en vægtet analyse eller transformer  $Y$ 'erne
- ③ Fejl i beregningerne
- ④ Mangler kvadratisk led i  $x_i$

Ovenstående er ikke tænkt som en udtømmende beskrivelse af hvorledes man analyserer residualplots, men nærmest som en vejledning i, efter hvilke retningslinier en sådan analyse kan foregå.

Bemærkning Man ser ofte, at der laves residualplots af typpen residual mod afhængig variabel, i.e.

$$Y_i - \hat{E}(Y_i) \text{ mod } Y_i,$$

og der udtrykkes da ofte undren over, at billedet er som i ③. Deri ligger imidlertid intet abnormt. Det kan nemlig vises, at

$$\text{Cor}(Y_i, Y_i - \hat{E}(Y_i)) = 1 - R^2,$$

d.v.s. de er positivt korrelerede. Hvis den multiple korrelationskoefficient blot er lidt mindre end 1, vil man derfor få et billede som ③. Kun hvis regressionsfladen går gennem samtlige punkter, i.e.  $R^2 = 1$ , vil man få et billede som ①.

#### 4.12

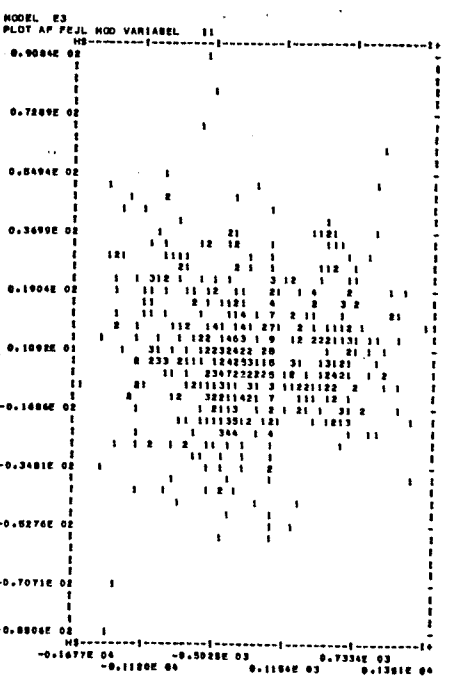
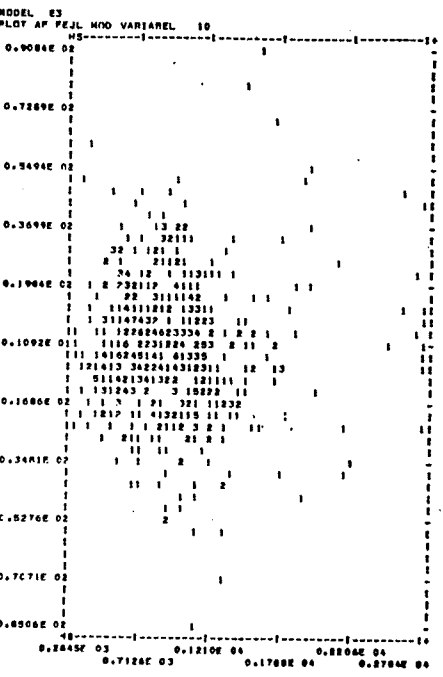
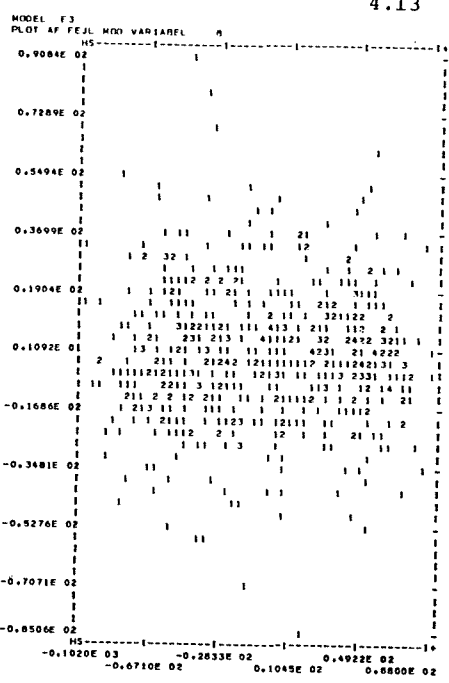
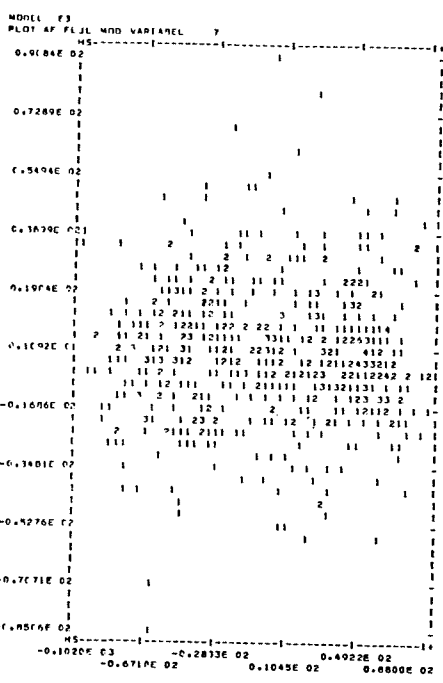
I praksis vil man ofte få sit residualplot printet ud på en linieskriver. Da kan plottene se ud som anført p. 4.13. De 4 plot er taget fra Spliid (1973) p. 14-15 i appendix C.

Ved tolkningen af disse plot må man være opmærksom på, at der ikke altid er lige mange observationer for hver værdi af den uafhængige variable.

Dette er eksempelvis tilfældet ved det plot, der afbilder fejlen (residualet) mod variabel 10.

Der er 7 observationer svarende til  $x_{10} \sim 0.2704E04$  og 35 observationer svarende til  $x_{10} \sim 0.7126E03$ . Variationsbredden for residualerne er nogenlunde den samme i de to tilfælde. Hvis residualerne svarende til de to værdier af  $x_{10}$  havde samme varians, ville man imidlertid forvente at variationsbredden for de mange observationer var den største.

Dette vil med andre ord sige, at hvis man har flest observationer omkring tyngdepunktet for en uafhængig variabel, skal et residualplot snarere være ellipseformet end af formen (1) for at være tilfredsstillende.



## 4.2 Regression efter ortogonale polynomier

Når man skal foretage en regressionsanalyse efter polynomier kan man ofte opnå ganske væsentlige beregningsmæssige besparelser ved at indføre de såkaldte ortogonale polynomier. Dette giver i sidste instans samme udtryk for estimater af middelværdien som funktion af den uafhængige variabel, men altså med et væsentligt mindre regnearbejde.

### 4.2.1 Definition og modelformulering

Vi antager, at der er givet en polynomial regressionsmodel, i.e., at

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \xi_0(t_1) & \xi_1(t_1) & \dots & \xi_k(t_1) \\ \vdots & \vdots & & \vdots \\ \xi_0(t_n) & \xi_1(t_n) & & \xi_k(t_n) \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix} .$$

Her betegner  $\xi_i$ ,  $i = 0, 1, \dots, k$ , kendte polynomier af  $i$ 'te grad i  $t$ . Vi forudsætter, at

$$\begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix} \in N(\underline{0}, \sigma^2 \underline{I}) .$$

I denne model kan vi på helt sædvanlig måde estimere og teste hypoteser angående parametrene  $(\alpha, \beta_1, \dots, \beta_k)$ .

Som anført indledningsvis kan man her med stor fordel betragte såkaldte ortogonale polynomier  $\xi_i$ , idet beregningsarbejdet da reduceres væsentligt.

Vi indfører disse polynomier i



Definition 4.1 Ved et sæt ortogonale polynomier svarende til værdierne  $t_1, \dots, t_n$  forstås polynomier  $\xi_0, \xi_1, \dots$ , hvor  $\xi_i$  er af  $i$ 'te grad, som tilfredsstiller

$$\sum_{j=1}^n \xi_i(t_j) = 0, \quad i = 1, 2, \dots \quad (1)$$

$$\sum_{j=1}^n \xi_\mu(t_j) \xi_\gamma(t_j) = 0, \quad \mu \neq \gamma. \quad (2)$$

Bemærkning Det påpeges, at  $\xi_0$  er en konstant, hvorfor (1) selvfølgelig ikke finder anvendelse på  $\xi_0$ . Af notationsmæssige grunde sættes  $\xi_i(t_j) = \xi_{ij}, \forall i, j$ . Vi skal senere vende tilbage til problemet med bestemmelse af ortogonale polynomier.

Forudsætter vi nu, at polynomierne i selve modellen er ortogonale, fås med

$$\underline{\xi} = \begin{bmatrix} \xi_0 \dots \xi_{k1} \\ \vdots \\ \xi_0 \dots \xi_{kn} \end{bmatrix} = \begin{bmatrix} \xi_0(t_1) \xi_1(t_1) \dots \xi_k(t_1) \\ \vdots \\ \xi_0(t_n) \xi_1(t_n) \dots \xi_k(t_n) \end{bmatrix},$$

at

$$\underline{\xi}' \underline{\xi} = \begin{bmatrix} n\xi_0^2 & 0 & \dots & 0 \\ 0 & \Sigma \xi_{1i}^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \dots & & \Sigma \xi_{ki}^2 \end{bmatrix},$$

d.v.s.  $\underline{\xi}' \underline{\xi}$  er en diagonalmatrix. Vi finder derfor

$$\hat{\underline{\beta}} = (\underline{\underline{\Xi}}' \underline{\underline{\Xi}})^{-1} \underline{\underline{\Xi}}' \underline{\underline{y}} = \begin{bmatrix} \bar{y}/\epsilon_0 \\ \Sigma \epsilon_{1i} y_i / \Sigma \epsilon_{1i}^2 \\ \vdots \\ \Sigma \epsilon_{ki} y_i / \Sigma \epsilon_{ki}^2 \end{bmatrix},$$

og

$$D(\hat{\underline{\beta}}) = \sigma^2 \begin{bmatrix} 1/n\epsilon_0^2 & 0 & \dots & 0 \\ 0 & 1/\Sigma \epsilon_{1i}^2 & & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ 0 & \dots & & 1/\Sigma \epsilon_{ki}^2 \end{bmatrix}.$$

Vi har derfor, at estimatorene for parametrene er ukorreleerede, og da vi arbejder i en normal model altså også stokastisk uafhængige.

Vi finder, at residualkvadratafgivelsessummen er

$$\begin{aligned} \text{SAK}_{\text{res}} &= \|\underline{\underline{y}} - \underline{\underline{\Xi}}\hat{\underline{\beta}}\|^2 \\ &= (\underline{\underline{y}} - \underline{\underline{\Xi}}\hat{\underline{\beta}})' (\underline{\underline{y}} - \underline{\underline{\Xi}}\hat{\underline{\beta}}) \\ &= \underline{\underline{y}}' \underline{\underline{y}} - \hat{\underline{\beta}}' \underline{\underline{\Xi}}' \underline{\underline{\Xi}} \hat{\underline{\beta}} \\ &= \Sigma y_i^2 - (\hat{\alpha}^2 n \epsilon_0^2 + \hat{\beta}_1^2 \Sigma \epsilon_{1i}^2 + \dots + \hat{\beta}_k^2 \Sigma \epsilon_{ki}^2) \\ &= \Sigma (y_i - \bar{y})^2 - (\hat{\beta}_1^2 \Sigma \epsilon_{1i}^2 + \dots + \hat{\beta}_k^2 \Sigma \epsilon_{ki}^2). \end{aligned}$$

Heraf fås nu umiddelbart følgende

Sætning 4.1 Vi har følgende spaltning af den totale variation

$$\Sigma (Y_i - \bar{Y})^2 = \hat{\beta}_1^2 \Sigma \xi_{1i}^2 + \dots + \hat{\beta}_k^2 \Sigma \xi_{ki}^2 + \Sigma (Y_i - \bar{Y} - \hat{\beta}_1 \xi_{1i}(t_1) - \dots - \hat{\beta}_k \xi_{ki}(t_1))^2 ,$$

eller - med en letfattelig notation -

$$SAK_{\text{tot}} = SAK_{1.\text{grad}} + \dots + SAK_{k.\text{grad}} + SAK_{\text{res}} ,$$

d.v.s. den er spaltet i led svarende til hvert polynomium plus residualkvadratafgivelsessummen. Frihedsgraderne er  $n-1$  henholdsvis  $1, \dots, 1$  og  $n-k-1$ .

Bevis Triviel følge af ovenstående.

Ved hjælp af spaltningssætningen fås ydermere

Sætning 4.2 De i foregående sætning anførte kvadratsummer er stokastisk uafhængige med forventede værdier

$$\begin{aligned} E(SAK_{i.\text{grad}}) &= E(\hat{\beta}_i^2 \Sigma \xi_{ij}(t_j)^2) \\ &= \sigma^2 + \beta_i^2 \Sigma_j \xi_{ij}(t_j)^2 , \quad i = 1, \dots, k . \end{aligned}$$

og

$$E(SAK_{\text{res}}) = E[\Sigma_j (Y_j - \bar{Y} - \dots - \hat{\beta}_k \xi_{kj}(t_j))^2] = (n-k-1)\sigma^2 .$$

Endelig er

$$\frac{1}{\sigma^2} SAK_{\text{res}} \in \chi^2(n-k-1) ,$$

og - såfremt  $\beta_i = 0$  -

$$\frac{1}{\sigma^2} SAK_{i.\text{grad}} \in \chi^2(1) .$$

Bevis Umiddelbart.

Sætningerne indeholder de nødvendige resultater for at kunne opstille tests for hypoteserne

$$H_{0i} : \beta_i = 0 \text{ mod } H_{1i} : \beta_i \neq 0 .$$

Vi samler resultaterne i et variansanalysekema

Variation	SAK	f	E(SAK/f)
Lineær	SAK <sub>1.grad</sub>	1	$\sigma^2 + \beta_1^2 \sum_j E_1(t_j)^2$
Kvadratisk	SAK <sub>2.grad</sub>	1	$\sigma^2 + \beta_2^2 \sum_j E_2(t_j)^2$
Kubisk	SAK <sub>3.grad</sub>	1	$\sigma^2 + \beta_3^2 \sum_j E_3(t_j)^2$
⋮			
⋮			
k'te ordens	SAK <sub>k.grad</sub>	1	$\sigma^2 + \beta_k^2 \sum_j E_k(t_j)^2$
Residual	SAK <sub>res</sub>	n-k-1	$\sigma^2$
Total	SAK <sub>tot</sub>	n-1	

Bemærkning Den store fordel ved at anvende ortogonale polynomier i regressionsanalysen er, at man uden at ændre nogle af de tidligere beregninger kan indføre polynomier af (p+1)'te orden, (p+2)'den orden etc. Ved bestemmelse af ordenen for det beskrivende polynomium vil man som regel fortsætte en (estimation og) testning jndtil 2 på hinanden følgende  $\beta_i$ 'er er lig 0, idet bidrag der skyldes lige ordens led og ulige ordens led er væsensforskellige. Dette er dog naturligvis en regel, der skal følges med varsomhed. Hvis man eg. har en på de fysiske forhold baseret formodning om, at led af 5'te orden er af betyd-

ning, vil man selvfølgelig ikke slutte analysen, selv om 3. og 4. ordens koefficienterne ikke afviger signifikant fra 0.

#### 4.2.2 Bestemmelse af ortogonale polynomier

Det ses umiddelbart, at multiplikation med en konstant ikke ændrer ved ortogonalitetsbetingelserne (1) og (2). Vi vælger derfor at sætte

$$\xi_0(t) = \xi_0 = 1 .$$

Polynomiet af 1. ste grad er

$$\xi_1(t) = t + a ,$$

idet vi uden videre kan sætte koefficienten til  $t$  lig 1. Af (1) fås

$$0 = \sum_{j=1}^n \xi_1(t_j) = \sum_{j=1}^n (t_j + a) = \sum_{j=1}^n t_j + na ,$$

eller

$$a = -\frac{1}{n} \sum_{j=1}^n t_j = -\bar{t} ,$$

d.v.s.

$$\xi_1(t) = t - \bar{t} .$$

Vi kan dernæst bestemme  $\xi_2$  som en linearkombination af 1,  $\xi_1$  og  $\xi_1^2$ , d.v.s.:

$$\xi_2(t) = a_{02} + a_{12}(t - \bar{t}) + a_{22}(t - \bar{t})^2 .$$

Af (1) fås

$$0 = \sum_{j=1}^n \xi_2(t_j) = na_{02} + a_{12} \sum_j (t_j - \bar{t}) + a_{22} \sum_j (t_j - \bar{t})^2$$

$$\frac{a_{02}}{a_{22}} = -\frac{1}{n} \sum_j (t_j - \bar{t})^2$$

Af (2) fås

$$0 = \sum_{j=1}^n \xi_1(t_j) \xi_2(t_j)$$

$$= a_{02} \sum_j (t_j - \bar{t}) + a_{12} \sum_j (t_j - \bar{t})^2 + a_{22} \sum_j (t_j - \bar{t})^3$$

$$= a_{12} \sum_j (t_j - \bar{t})^2 + a_{22} \sum_j (t_j - \bar{t})^3$$

Heraf fås

$$\frac{a_{12}}{a_{22}} = -\frac{\sum_j (t_j - \bar{t})^3}{\sum_j (t_j - \bar{t})^2}$$

$\xi_3, \xi_4$  etc. bestemmes på ganske analog vis.

Beregningerne bliver særligt simple, hvis  $t_1$ 'erne er ækvivalente. Da sætter vi

$$a_{12}^2 = \frac{36}{14} a_{22}^2$$

Handwritten calculations and notes:

- $\frac{4+1+1+9}{3} = \frac{15}{3} = 5$
- $\frac{14}{3}$
- $a_{12} = 14 + a_{22}$
- $a_{22} = \frac{14}{3}$
- $\frac{8+1+27}{14}$

$$u_j = \frac{t_j - (t_1 - w)}{w},$$

hvor  $w = t_2 - t_1 = t_{i+1} - t_i$ . Vi har så

$$u_i = i, \quad i = 1, \dots, n.$$

Svarende til værdierne  $1, \dots, n$  har vi da polynomierne givet ved

$$\xi_0(t) = 1 \quad (4)$$

$$\xi_1(t) = t - \frac{n+1}{2} \quad (5)$$

$$\xi_{i+1}(t) = \xi_1(t)\xi_i(t) - \frac{i^2(n^2-i^2)}{4(4i^2-1)} \xi_{i-1}(t). \quad (6)$$

I omstående tabel p. 4.22 har vi givet værdier af ortogonale polynomier  $\xi_1, \dots, \xi_k$ ,  $k \leq 5, i \quad t = 1, \dots, n$  for  $n = 1, \dots, 8$ .

For at undgå brudne tal og store værdier har man valgt at anføre polynomier, hvor koefficienten til højstegradsleddet er et tal  $\lambda$ , der fremgår af tabellen. Endvidere er der angivet størrelserne

$$D = \sum_{j=1}^n \xi_i(j)^2 = \sum_{j=1}^n \xi_{ij}^2.$$

272 > 14/2

36 0.2  
14

n	3		4		5			6			7			8										
	$\epsilon_1$	$\epsilon_2$	$\epsilon_1$	$\epsilon_2$	$\epsilon_1$	$\epsilon_2$	$\epsilon_3$	$\epsilon_4$	$\epsilon_1$	$\epsilon_2$	$\epsilon_3$	$\epsilon_4$	$\epsilon_5$	$\epsilon_1$	$\epsilon_2$	$\epsilon_3$	$\epsilon_4$	$\epsilon_5$						
1	-1	1	-3	1	-1	1	-2	2	-1	1	-1	-1	-3	5	-1	3	-1	-7	7	-7				
2	0	-2	-1	-1	3	-1	2	-4	-3	-1	7	-3	5	-2	0	1	-7	4	-5	1	5	-13	23	
3	1	1	1	-1	-3	9	-2	0	-1	-4	4	2	10	-1	-3	1	1	-5	-3	-3	7	-3	-17	
			3	1	1	1	-1	-2	1	-4	-4	2	10	0	-4	0	6	0	-1	-5	3	9	-15	
					2	2	1	1	3	-1	-7	-3	-5	1	-3	-1	1	5	1	-5	-3	9	15	
									5	5	5	1	1	2	0	-1	-7	-4	3	-3	-7	-3	17	
														3	5	1	3	1	5	1	-5	-13	-23	
D	2	6	20	4	20	10	14	10	70	70	84	180	28	252	28	84	6	154	84	168	168	264	616	2184
$\lambda$	1	3	2	1	$\frac{10}{3}$	1	1	$\frac{5}{6}$	$\frac{35}{12}$	2	$\frac{3}{2}$	$\frac{5}{3}$	$\frac{7}{12}$	$\frac{21}{10}$	1	1	$\frac{1}{6}$	$\frac{7}{12}$	2	1	$\frac{2}{3}$	$\frac{7}{12}$	$\frac{10}{10}$	

Tabel 4.1 Værdier af ortogonale polynomier.



Vi giver nu et illustrativt

Eksempel 4.1 I nedenstående tabel er der anført sammenhørende værdier mellem reaktionstemperaturen og udbyttet af en proces (i en fast tid).

<u>Temperatur</u>	<u>Udbytte</u>
200° F	0.75 oz
210° F	1.00 oz
220° F	1.35 oz
230° F	1.80 oz
240° F	2.60 oz
250° F	3.60 oz
260° F	5.45 oz

Vi vil forsøge at beskrive udbyttet som en funktion af temperaturen ved hjælp af et polynomium. Vi antager, at forudsætningerne for at foretage en regressionsanalyse er til stede.

Vi transformerer først temperaturerne  $\tau_i$ ,  $i = 1, \dots, 7$  v.h.a. følgende relation

$$t_i = \frac{\tau_i - (200 - 10)}{10} = \frac{\tau_i - 190}{10}$$

Vi får da værdierne  $t_1, \dots, t_7 = 1, \dots, 7$ .

Vi anfører beregningerne i følgende skema

-1  
-2  
\* 3  
-4  
-5  
-6  
-7

$t_j$	$\xi_1$	$\xi_2$	$\xi_3$	$\xi_4$	$\xi_5$	$y_j$
1	-3	5	-1	3	-1	0.75
2	-2	0	1	-7	4	1.00
3	-1	-3	1	1	-5	1.35
4	0	-4	0	6	0	1.80
5	1	-3	-1	1	5	2.60
6	2	0	-1	-7	-4	3.60
7	3	5	1	3	1	5.45
$\Sigma \xi_{ij}^2$	28	84	6	154	84	$16.55 = \Sigma y_j$
$\Sigma \xi_{ij} y_j$	20.55	11.95	0.85	1.15	0.55	$56.0475 = \Sigma y_j^2$
$\lambda$	1	1	1/6	7/12	7/20	

Heraf fås

$$\begin{aligned} \Sigma (y_i - \bar{y})^2 &= 56.0475 - \frac{16.55^2}{7} \\ &= 56.0475 - 39.1289 \\ &= 16.9186 \end{aligned}$$

$$\hat{\alpha} = \frac{16.55}{7} = 2.36$$

$$\hat{\beta}_1 = \frac{20.55}{28} = 0.7339 \quad \text{SAK}_{1.\text{grad}} = \frac{20.55^2}{28} = 15.0822$$

$$\hat{\beta}_2 = \frac{11.95}{84} = 0.1423 \quad \text{SAK}_{2.\text{grad}} = \frac{11.95^2}{84} = 1.7000$$

$$\hat{\beta}_3 = \frac{0.85}{6} = 0.1417 \quad \text{SAK}_{3.\text{grad}} = \frac{0.85^2}{6} = 0.1204$$

$$\hat{\beta}_4 = \frac{1.15}{154} = 0.0075 \quad \text{SAK}_{4.\text{grad}} = \frac{1.15^2}{154} = 0.0086$$

$$\hat{\beta}_5 = \frac{0.55}{84} = 0.0065 \quad \text{SAK}_{5.\text{grad}} = \frac{0.55^2}{84} = 0.0036$$

Vi samler resultaterne i nedenstående skema.

Vi ser, at leddene af 1., 2. og 3. grad er signifikante og de to følgende ikke signifikante, så vi bruger et polynomium af 3. die grad til beskrivelse.

Variation	SAK	f	S <sup>2</sup>	Test	F-fraktion
Total	16.9186	6			
1.grad	15.0822	1	15.0822		
Residual 1	1.8364	5	0.3673	41.06	99.8%
2.grad	1.7000	1	1.7000		
Residual 2	0.1364	4	0.0341	49.85	99.7%
3.grad	0.1204	1	0.1204		
Residual 3	0.0160	3	0.0053	22.72	98%
4.grad	0.0086	1	0.0086		
Residual 4	0.0074	2	0.0037	2.32	75%
5.grad	0.0036	1	0.0036		
Residual 5	0.0038	1	0.0038	0.95	<50%

Af rekursionsformlen (5) og (6) fås - da  $n = 7$  -

$$E_1(t) = t - 4$$

$$E_2(t) = (t - 4)^2 - \frac{48}{12}$$

$$= t^2 - 8t + 12$$

$$E_3(t) = (t - 4)(t^2 - 8t + 12) - \frac{4 \cdot 45}{4 \cdot 15} (t - 4)$$

$$= t^3 - 12t^2 + 41t - 36 .$$

Da  $\lambda_1 = 1$ ,  $\lambda_2 = 1$  og  $\lambda_3 = 1/6$  får vi følgende estimerede polynomium

## 4.26

$$\begin{aligned}\hat{\mu}(t) &= 2.36 + 1 \cdot \hat{\beta}_1 \mathcal{E}_1(t) + 1 \cdot \hat{\beta}_2 \mathcal{E}_2(t) + 1/6 \hat{\beta}_3 \mathcal{E}_3(t) \\ &= 0.0236 t^3 - 0.1409 t^2 + 0.5631 t + 0.2818 .\end{aligned}$$

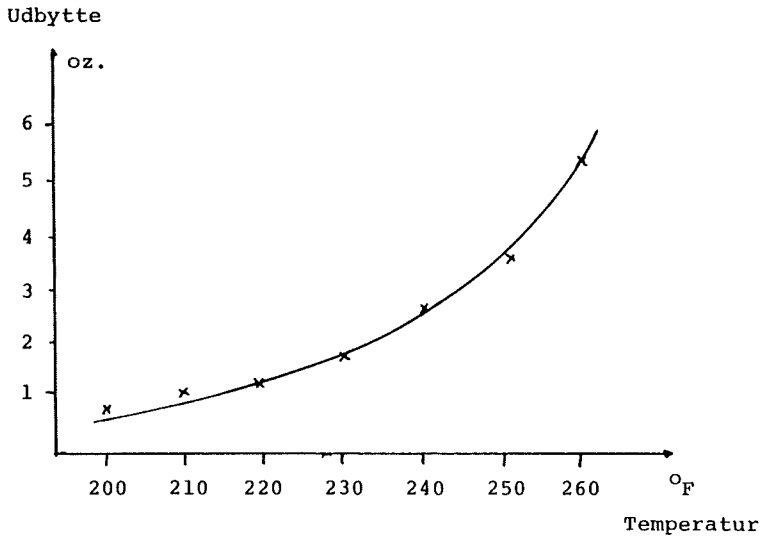
Da

$$t_i = \frac{\tau_i - 190}{10} ,$$

får vi et udtryk, hvori de oprindelige temperaturer indgår, ved blot at indsætte denne relation i udtrykket for  $\hat{\mu}(t)$ . Det giver

$$g(\tau) = 0.000024\tau^3 - 0.014861\tau^2 + 3.147610\tau_i - 223.154400 .$$

Dette estimerede polynomium er sammen med de oprindelige data anført i nedenstående figur.



Figur 4.2 Sammenhængen mellem temperatur og udbytte ved den i eksempel 4.2 angivne proces.

□

### 4.3 Valg af den "bedste" regressionsligning

I dette afsnit vil vi beskæftige os med problemet at vælge en passende (lille) mængde af uafhængige variable, som giver mulighed for en rimelig beskrivelse af vort datamateriale.

#### 4.3.1 Problemstillingen

Hvis man e.g. er i den (kedelige) situation ikke at kunne formulere en på de fysiske forhold begrundet model for det fænomen, man studerer, vil man ofte som en sidste udvej ty til blot at registrere alle de størrelser, man overhovedet mener kan være af betydning for ens måleværdier. Hvis man så laver regression efter f.eks. polynomier i disse uafhængige variable (ud fra en Taylor-approximations-synsvinkel), vil man meget hurtigt komme op på et enormt antal led i ens regressionsligning. Har man 10 "grund"-variable  $x_1, \dots, x_{10}$ , vil der i et almindeligt 2.dengrads polynomium i disse findes 66 led. Går vi op til 3.die grad får vi i størrelsesordenen 150 led. Udtryk der indeholder så mange led vil - hvis det overhovedet er muligt at estimere alle parametrene - være meget tunge at arbejde med. Hvis man e.g. ønsker at bestemme optimale produktionsbetingelser for en kemisk proces, kan man estimere responsfladen og finde maksimum for denne. Dette vil være en endog overordentlig vanskelig opgave, hvis der er mange variable involveret. Man vil derfor søge at finde et væsentligt mindre antal led, der giver en tilpas "god" beskrivelse af variationen i materialet. (Jævnfør også afsnittet om ridge regression).

Det er dog vigtigt her at gøre sig klart, at et udtryk man kommer frem til ved de metoder, vi skal omtale, må benyttes med varsomhed. Der vil (formentlig) være tale om et udtryk, der beskriver de foreliggende data udmærket. Om metoden er velegnet til at forudsige fremtidige observationer, vil afhænge af, om udtrykket også afspejler de fysiske forhold tilstrækkeligt godt. En måde at få opklaret dette problem på, er i første

omgang alene at basere estimationen på e.g. halvdelen af data-materialet, og så sammenligne den resterende halvdel med den estimerede model. Hvis overensstemmelsen her er tilstrækkeligt stor, har man fået en indikation for, at modellen ikke er uanvendelig som forudsigelsesmodel.

Vi vil bruge et enkelt gennemgående eksempel som illustration af de metoder, vi vil omtale. For at det skal være muligt at overskue (og eventuelt kontrollere) de enkelte beregninger, har vi kun udtaget en meget lille del af det oprindelige materiale. Man skal derfor ikke vurdere metodernes egnethed v.h.a. eksemplet, men kun anvende det som illustration af principperne og gangen i disse. Data er nogle sammenhørende målinger af kvaliteten  $y$  af et fødeadditiv (målt ved viskositeten) og nogle produktionsparametre  $x_1$ ,  $x_2$  og  $x_3$  (tryk, temperatur og neutraliseringsgrad). For at lette beregningerne er data kodede, i.e. der er subtraheret nogle konstanter og divideret med passende tal. Vi har følgende målinger

$y$	$x_1$	$x_2$	$x_3$
4.9	0	0	2
3.0	1	0	1
0.2	1	1	0
2.9	1	2	2
6.4	2	1	2

Erfaringen har vist, at det inden for et passende lille variationsområde af produktionsparametrene er rimeligt at regne med, at kvaliteten afhænger lineært af disse. Vi anvender derfor modellen

$$E(Y|\underline{x}) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 ,$$

eller skrevet på matrixform

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 2 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 2 & 2 \\ 1 & 2 & 1 & 2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix} ,$$

$$\underline{\epsilon} \in N(\underline{0}, \sigma^2 \underline{I}) .$$

Vi har i appendix (p. 4.42) anført samtlige  $2^3$  regressionsanalyser med  $y$  som afhængig variabel og en eller flere af  $x$ 'erne som uafhængige variable. Der er følgende mulige modeller

$$\begin{aligned} M &: E(Y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \\ H_{12} &: E(Y) = \alpha + \beta_1 x_1 + \beta_2 x_2 \\ H_{13} &: E(Y) = \alpha + \beta_1 x_1 + \beta_3 x_3 \\ H_{23} &: E(Y) = \alpha + \beta_2 x_2 + \beta_3 x_3 \\ H_1 &: E(Y) = \alpha + \beta_1 x_1 \\ H_2 &: E(Y) = \alpha + \beta_2 x_2 \\ H_3 &: E(Y) = \alpha + \beta_3 x_3 \\ H_0 &: E(Y) = \alpha \end{aligned} .$$

Vi anfører for hver af disse 8 modeller estimatorerne for  $\alpha$  og  $\beta$ 'erne, vi finder projektionen af observationsvektoren på det til modellen svarende underrum, vi bestemmer residualvektoren, residualvektorens kvadrerede længde (residualkvadratafvigelsessummen), variansskøn og den multiple korrelationskoefficient. Dernæst anfører vi variansanalysekemaer for de mulige sekvenser af successive testninger af hypoteser om, at middelværdivektoren tilhører stadig mindre (lavere dimension) underrum i kæder som

$$M \supseteq H_{12} \supseteq H_2 \supseteq H_0 .$$

Ovenstående kæde af underrum svarer til successiv testning af hypoteserne

$$\beta_3 = 0, \quad \beta_1 = 0, \quad \beta_2 = 0 .$$

Der er 6 (= 3!) mulige skemaer af denne art. Endelig anføres nogle partielle korrelationsmatricer. Sætter vi  $y = x_4$  defineres dispersionsmatricen (den empiriske) som bekendt ved, at det  $(i, j)$ 'te element er

$$S_{ij} = \frac{1}{n-1} \sum_{\mu} (x_{i\mu} - \bar{x}_i)(x_{j\mu} - \bar{x}_j) .$$

Korrelationsmatricens  $(i, j)$ 'te element er da

$$r_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}} .$$

Ved hjælp af formlen p. 2.36 i afsnit 2. beregnes dernæst de partielle korrelationer for givet  $x_3$  og for givet  $x_2, x_3$ .

Vi er nu rustede til at omtale nogle af de mest anvendte måder til at udvælge enkelte uafhængige variable til beskrivelse af variationen i den afhængige variable.

#### 4.3.2 Undersøgelse af samtlige regressioner

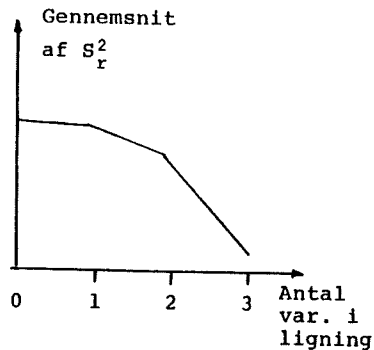
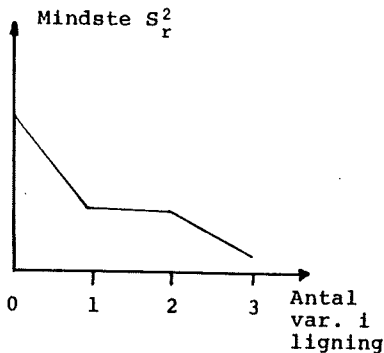
Denne metode kan selvsagt kun anvendes, hvis der er rimeligt få variable.

Vi samler resultatet fra appendicet i følgende skema



Model	Multiple $R^2$	Residualv. $S_r^2$	Gennemsn. af $S_r^2$
$H_0 : E(Y) = \alpha$	0	5.47	5.47
$H_1 : E(Y) = \alpha + \beta_1 x_1$	5.1%	6.91	5.35
$H_2 : E(Y) = \alpha + \beta_2 x_2$	3.8%	7.01	
$H_3 : E(Y) = \alpha + \beta_3 x_3$	70.8%	2.13	
$H_{12} : E(Y) = \alpha + \beta_1 x_1 + \beta_2 x_2$	15.3%	9.26	4.68
$H_{13} : E(Y) = \alpha + \beta_1 x_1 + \beta_3 x_3$	76.0%	2.63	
$H_{23} : E(Y) = \alpha + \beta_2 x_2 + \beta_3 x_3$	80.4%	2.14	
$M : E(Y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$	97.1%	0.634	0.634

Et blik på de multiple korrelationskoefficienter viser, at der ikke er vundet så afgørende meget ved at gå fra 1 variabel ( $x_3$ ) op til 2 variable. Det afgørende spring sker først ved at gå op til fulde 3 variable. Overvejelser i disse baner fører til, at man vil foretrække alene at anvende  $x_3$  d.v.s. modellen  $E(Y|x) = \alpha + \beta_3 x_3$ . Denne afgørelse bliver yderligere bestyrket ved at se på residualvariansen  $S_r^2$ . Vi ser da, at  $S_r^2$  for den "bedste" ligning i en variabel er mindre end for den "bedste" ligning i 2 variable, hvilket kraftigt indikerer, at vi bør nøjes med en variabel (eller tage alle tre). Ser vi foruden på de mindste  $S_r^2$  også på de gennemsnitlige værdier og afbilder dem efter antallet af indgående variable, får vi grafer som



Dette antyder også, at antallet af variable i en ligning bør være enten 1 eller 3 (der opnås ingen væsentlige forbedringer ved at gå fra 1 til 2).

Ser man alene på grafen med de gennemsnitlige værdier er det ikke indlysende, om man overhovedet bør inddrage nogen uafhængig variabel. Man kan derfor teste om  $\beta_3$  i modellen  $H_3$  ( $E(y|x) = \alpha + \beta_3 x_3$ ) kan antages at være 0. Teststørrelsen er

$$\frac{\|P_{H_0}(Y) - P_{H_3}(Y)\|^2 / 1}{\|Y - P_{H_3}(Y)\|^2 / 3} = \frac{21.868 - 6.38}{6.38/3} \approx 7.28 .$$

Vi vil derfor få forkastet  $\beta_3 = 0$  på alle niveauer større end 8%.

Som en sammenfatning af disse (løse) betragtninger vil vi konkludere, at vi anvender modellen  $H_3$ :

$$E(Y|x) = \alpha + \beta_3 x_3 \approx 0.4 + 2.2x_3 .$$

(Her betegner  $\approx$  estimeret til). Skønnet over usikkerheden (variansen) på målingerne er (estimeret med 3 frihedsgrader)

$$s^2 = 2.13 .$$

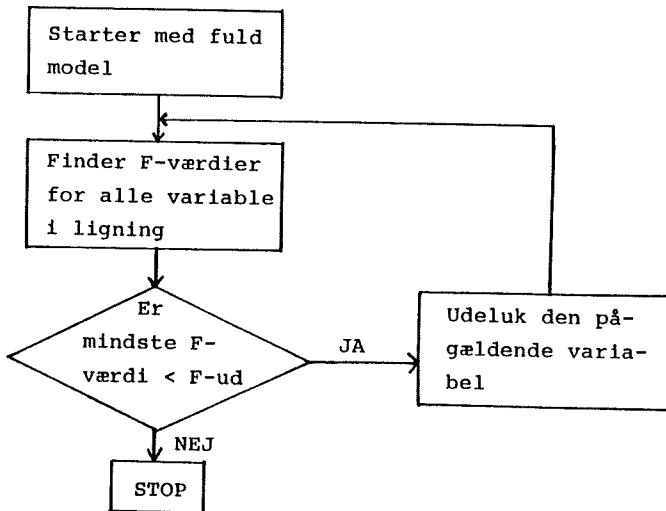
Bemærkning Det må her tilføjes, at ideen med at se på gennemsnittene af residualvarianserne forekommer noget tvivlsom. Den er medtaget fordi metoden nyder en vis udbredelse - i det mindste i litteraturen.

#### 4.3.3 Backwards elimination

Denne metode er langt mere økonomisk, hvad angår regnetid, end den foregående. Man starter her med den fuldstændige model M og undersøger så hvilken af koefficienterne, der har den mindste F-værdi for et test af hypotesen, at koefficienten er 0.

Denne variable udelukkes, og proceduren gentages så med de resterende  $k-1$  variable etc.

Man kan så stoppe proceduren, når ingen af de tilbageværende variable har en F-værdi, der er mindre end  $1-\alpha$  fraktilen i den relevante F-fordeling.



Figur 4.3 Flow diagram for Backwards-elimination procedure i trinvis regressionsanalyse.

Vi kan anskueliggøre fremgangsmåden ved hjælp af vort eksempel. Vi samler data i omstående skema.

Af skemaet fremgår, at vi også her vil ende med modellen  $H_3$ :  $E(y) = \alpha + \beta_3 x_3$ , når vi arbejder med et  $\alpha$ , der er større end 8%.

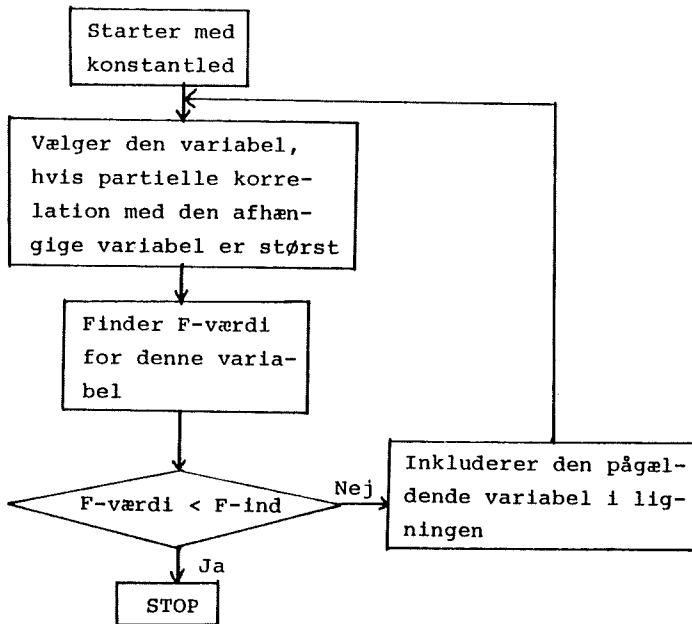
Trin	F-værdi for test af $\beta_1 = 0$	Fraktil i F-fordel.
Model : $E(Y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$		
1	$\beta_1 : \frac{3.652/1}{0.634/1} = 5.76$	$\approx F(1,1)_{0.71}$
	$\beta_2 : \frac{4.621/1}{0.634/1} = 7.29$	$\approx F(1,1)_{0.72}$
	$\beta_3 : \frac{17.879/1}{0.634/1} = 28.20$	$\approx F(1,1)_{0.86}$
Udelader $x_1$ : model nu $E(Y) = \alpha + \beta_2 x_2 + \beta_3 x_3$		
2	$\beta_2 : \frac{2.095/1}{4.285/2} = 0.98$	$\approx F(1,2)_{0.55}$
	$\beta_3 : \frac{16.757/1}{4.285/2} = 7.82$	$\approx F(1,2)_{0.88}$
Udelader $x_2$ : model nu $E(Y) = \alpha + \beta_3 x_3$		
3	$\beta_3 : \frac{15.488/1}{6.38/3} = 7.28$	$\approx F(1,3)_{0.92}$

Ulempen ved denne metode er, at vi skal løse den fuldstændige regressionsmodel, hvilket kan være besværligt, hvis der er mange uafhængige variable.

Dette problem bliver der taget højde for ved den næste procedure.

#### 4.3.4 Forward selection

Ved denne procedure starter man med kun at have konstantleddet i ligningen. Dernæst vælger man den uafhængige variable, der er mest korreleret med den afhængige variable. Man udfører et F-test for, om dennes koefficient er signifikant  $\neq 0$ . Hvis ja, drages den med ind i modellen.



Figur 4.4 Flowdiagram for Forward-selection procedure i trinvis regressionsanalyse.

Blandt de resterende uafhængige variable udvælges nu den, der har den største (numerisk) partielle korrelationskoefficient med den afhængige variable givet de(n) variable, der allerede er i ligningen. Man udfører et F-test for om denne nye variable har bidraget til reduktion af residualvariansen, i.e. om koefficienten til den er  $\neq 0$ . Hvis ja, fortsættes som før, hvis nej, standses analysen.

I vort eksempel bliver gangen

① Af korrelationsmatricen (p. 4.48) ses, at  $x_3$  har den højeste korrelationskoefficient med  $y$ , nemlig 0.8416. Vi tester, om  $\beta_3$  i modellen  $E(Y) = \alpha + \beta_3 x_3$  kan antages at være 0. Vi har teststørrelsen (se p. 4.47).

$$\frac{15.488/1}{6.38/3} = 7.28 \approx F(1,3)_{0.92} .$$

Hvis vi regner med  $\alpha = 10\%$  fortsættes (da vi da forkaster  $\beta_3 = 0$ ).

② Af den partielle korrelationsmatrix givet  $x_3$  (p. 4.49) ses, at den variable, der, givet  $x_3$  er med i ligningen, har den største partielle korrelationskoefficient med  $y$ 'erne er  $x_2$  ( $\rho_{x_2 y | x_3} = -0.5728$ ). Vi inddrager  $x_2$  og undersøger, om  $\beta_2$  i modellen

$$E(y) = \alpha + \beta_2 x_2 + \beta_3 x_3$$

kan antages at være 0. Vi har teststørrelsen (se p. 4.48)

$$\frac{2.095/1}{4.2855/2} = 0.98 \approx F(1,2)_{0.55} .$$

Da vi regnede med  $\alpha = 10\%$ , er denne størrelse altså ikke signifikant forskellig fra 0, og vi slutter analysen her og uden at medtage  $x_2$ . Den resulterende model er

$$E(Y) = \alpha + \beta_3 x_3 ,$$

hvor  $\alpha$  og  $\beta$  estimeres som tidligere. Vi bemærker her specielt, at  $x_1$  slet ikke har været med i ligningen.

Bemærkning 1 Hvis vi havde opereret med  $\alpha = 50\%$ , ville vi have fortsat analysen og betragtet de partielle korrelationer givet  $x_2$  og  $x_3$ . Ifølge matricen p. 4.49 er den partielle korrelationskoefficient mellem  $y$  og  $x_1$  givet  $x_2$  og  $x_3$  er i ligningen

$$\rho_{x_1 y | x_2 x_3} = 0.8956 .$$

Nu er  $x_1$  den eneste tiloversblevne variable, så derfor er den trivielt den af de tiloversblevne, der har de største partielle korrelationer med  $y$ . Vi drager  $x_1$  med ind i ligningen og undersøger, om  $\beta_1$  i modellen  $E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$  er signifikant forskellig fra 0. Teststørrelsen er (p. 4.48)

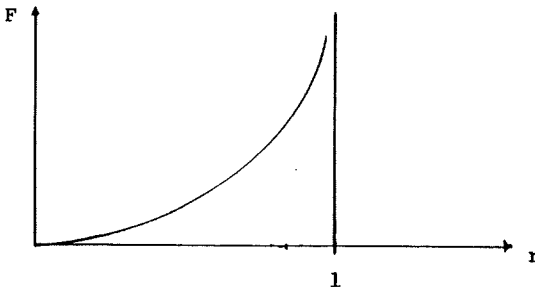
$$\frac{3.652/1}{0.634/1} = 5.76 \approx F(1,1)_{0.71} .$$

I dette tilfælde har vi set, at ligningen blev udvidet væsentligt ved at ændre  $\alpha$ . Det er vigtigt at gøre sig klar, at ændringer i  $\alpha$  kan have drastiske ændringer i den resulterende model til følge.

Bemærkning 2 Det i hvert trin at vælge den variable, der har den største partielle korrelation med den afhængige variable, er ensbetydende med at vælge de variable, der har den største F-værdi i de partielle F-tests. Dette resultat følger af, at sammenhængen mellem den partielle korrelationskoefficient og F-teststørrelsen er af formen

$$F = g(r) = \frac{r^2}{1-r^2} \cdot f ,$$

hvor  $f$  er frihedsgradsantallet for nævneren (jvf. p. 4.7). Denne sammenhæng er åbenbart voksende



Hvis vi e.g. i trin 2 vil beregne F-teststørrelsen ud fra korrelationsmatricen, får vi

$$F = \frac{(-0.5728)^2}{1 - (-0.5728)^2} \cdot 2 = 0.98 .$$

Dette ses endvidere, at omtalte kriterium er ensbetydende med i hvert enkelt trin hele tiden at indtage den variable, der bevirker den største reduktion i residualkvadratafvigelsessummen.

Bemærkning 3 I mange af de eksisterende standardregressionsprogrammer er det ikke muligt at specificere en  $\alpha$ -værdi. Man må i stedet angive et fast tal som grænse for de F-teststørrelser, man vil acceptere respektive forkaste. Man må da ved at se i en tabel over F-fraktiler finde en passende værdi. Ønsker man e.g.  $\alpha \approx 5\%$ , ser man, at man bør vælge værdien 4, idet

$$F(1, n)_{.95} \approx 4 ,$$

for rimeligt store værdier af  $n$ .

"Forward selection" metoden udmærker sig frem for "Backward elimination" metoden ved, at vi ikke behøver at regne den totale ligning ud. Den største ulempe ved metoden er nok, at der ikke tages hensyn til, at nogle variable kan være blevet overflødiggjort ved at andre senere er kommet ind. Hvis vi f.eks. tænker os, at  $x_1 = ax_2 + bx_3$  (ca.), og at  $x_1$  er valgt som den mest betydende variable. Hvis vi så senere i analysen også inddrager  $x_2$  og  $x_3$ , er det klart, at vi ikke længere har "brug" for  $x_1$ . Den bør derfor udelukkes. Det sker ved den sidste metode, vi omtaler:

#### 4.3.5 Stepwise regression

Navnet er dårligt, idet man med lige så fuld ret ville kunne



benævne de to foregående metoder med dette navn. Der er også mange forfattere, der bruger navnet stepwise regression som en fællesbetegnelse for en række forskellige procedurer. I nærværende fremstilling tænker vi helt specifikt på følgende metode.

Udvælgelsen af den variable, der skal ind i ligningen foregår som ved forward selection proceduren; men i hvert enkelt trin undersøger man så hver af de variable, der er i ligningen, som om de var de sidst tilføjede variable. Man beregner så på denne måde en F-teststørrelse for alle variable, der er i ligningen. Hvis nogle af disse er mindre end  $1-\alpha$ -fraktilen i den relevante F-fordeling, udelukkes de pågældende variable.

Hvis vi ser på vort standardeksempel, får vi følgende trin ( $\alpha_{\text{ind}} = 50\%$ ,  $\alpha_{\text{ud}} = 40\%$ ).

①  $x_3$  tages ind som ved forward selection proceduren, og vi tester om  $\beta_3$  er signifikant forskellig fra 0. Teststørrelsen og konklusion bliver som før.

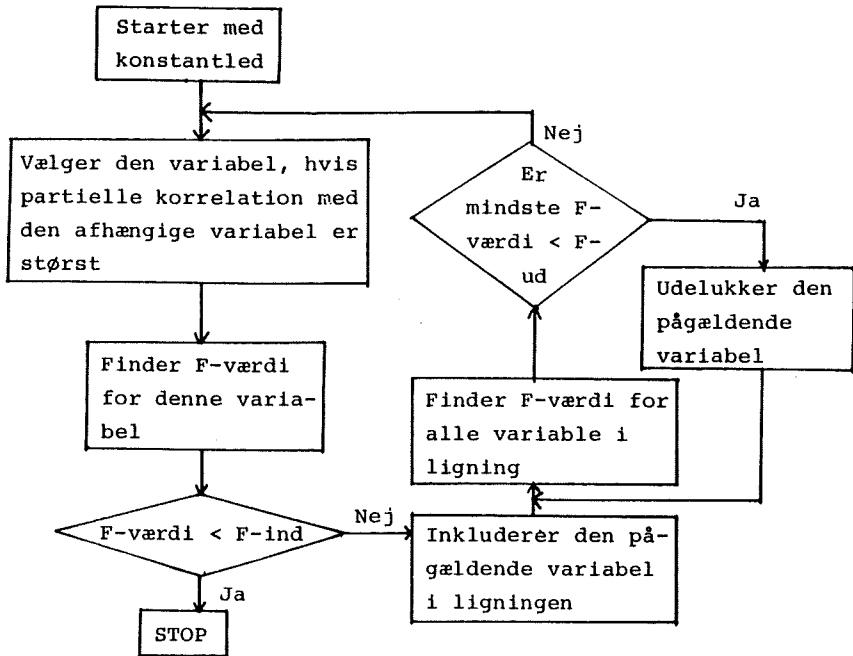
② Vi tager nu  $x_2$  ind. Vi danner det partielle F-test for  $\beta_2$  (i modellen  $E(Y) = \alpha + \beta_2 x_2 + \beta_3 x_3$ ):

$$x_2 : \text{F-værdi} = \frac{2.095/1}{4.285/2} = 0.98 \approx F(1,2)_{0.55} .$$

Dernæst laver vi et partielt F-test for  $\beta_3$  (i modellen  $E(Y) = \alpha + \beta_2 x_2 + \beta_3 x_3$ ). Ved hjælp af skemaet p. 4.48 fås, at

$$x_3 : \text{F-værdi} = \frac{16.757/1}{4.285/2} = 7.82 \approx F(1,2)_{0.88} .$$

③ Vi fjerner igen  $x_2$  fra ligningen da  $0.55 < 0.60$ . Forskellen i dette trin mellem forward selection proceduren og stepwise proceduren er, at vi også beregner en F-værdi for  $x_3$  og dermed åbner mulighed for at  $x_3$  igen elimineres fra ligningen. Dette var ikke muligt ved den rene forward selection procedure.



Figur 4.5 Flowdiagram for Stepwise-Regression procedure i trinvis regressionsanalyse.

④ Den eneste tilbageværende variable er  $x_1$ . Den har en partiel F-værdi på

$$x_1 : F\text{-værdi} = \frac{1.125/1}{5.255/2} = 0.43 < F(1,2)_{0.50} ,$$

så den kommer overhovedet ikke ind i ligningen.

Analysen stopper, og vi har modellen

$$E(Y) = \alpha + \beta_3 x_3 .$$

Bemærkning 1 Grunden til, at vi undersøgte den partielle F-værdi under 2, men ikke under 4 er, at  $x_1$  overhovedet ikke kommer ind i ligningen, da

$$0.43 < F(1,2)_{0.50} = F_{1-\alpha_{ind}}$$

Derimod kom  $x_2$  ind i ligningen, da

$$0.98 \approx F(1,2)_{0.55} > F_{1-\alpha_{ind}}$$

Bemærkning 2 Som under afsnittet om forward selection proceduren kan vi konstatere, at man ofte bliver tvunget til at arbejde med faste F-værdier i stedet for  $1-\alpha$  fraktiler. Når man ikke anvender samme niveau ved afgørelsen af, om man vil tage flere variable ind, som man bruger ved undersøgelse af, om nogle variable skal ud, vil man ofte gøre den sidste ca. halvt så stor som den første, i.e.

$$F\text{-ud af ligning} = \frac{1}{2} F\text{-ind i ligning} .$$

(Dette er det modsatte af det, vi har anvendt i eksemplet).

#### 4.3.6 Nogle eksisterende programmer

Vi skal ingenlunde give en oversigt over, hvad der findes af regressionsanalyseprogrammer, men blot nævne 3, der alle er let tilgængelige på NEUCC.

Der er for det første SSP-programmet STEPR, som kalder en del subrutiner fra SSP biblioteket. Det er et rent forward selection program. Det er ikke umiddelbart muligt at tvinge en regressionsflade gennem 0. Der er mulighed for at få tabel over residualer. De af STEPR kaldte rutiner fra SSP-biblioteket er lagt ind under WATFIV-systemet.

#### 4.42

BMD02R er et "stepwise" program. Det er med dette system helt trivielt at tvinge en regressionsflade gennem 0. Det er endvidere muligt at få såvel tabeller over residualer som plots af residualer mod input-variable. Endelig kan man få udskrevet tabel over analysens forskellige trin.

Det af H. Spliid udarbejdede program REGR, er et "stepwise" program, der er kombineret med en backwards elimination procedure. Det vil derfor være normalt her at anvende mindre F-værdier for ind- og udtagning af variable under den trinvis regression. Man får da en ligning indeholdende temmeligt mange variable, og på denne kan man anvende en backwards elimination procedure. Denne metode er nok en af de mere velegnede til at finde gode "prediktorer" blandt et stort antal variable. I den nyeste version af REGR er det dels muligt at tvinge regressionsfladen gennem 0, og dels er det muligt at udføre en simpel vægtet regression. Der findes særlige residualplotrutiner til programmet, og endvidere findes der en version af REGR under WATFIV systemet.

En meget omfattende samling regressionsrutiner findes i SAS-systemet. Foruden Forward og Backwards proceduren findes en stepwise procedure, en såkaldt "maximum  $R^2$ -improvement" procedure, en procedure der udfører samtlige regressioner m.fl.

For alle de nævnte programmer gælder, at man på forhånd har mulighed for at angive visse variable, som man altid ønsker medtaget i regressionsligningen.

#### 4.3.7 Numerisk appendix

Vi anfører i dette appendix en udregning af de størrelser, der er anvendt i de tidligere afsnit. Det skulle ikke være nødvendigt at gennemgå alle disse udregninger, men de er anført, for at man ved hjælp af disse skal kunne checke sin forståelse af de forskellige principper.

A. DATA:

Y	x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>
4.9	0	0	2
3.0	1	0	1
0.2	1	1	0
2.9	1	2	2
6.4	2	1	2

B. Grundmodel:  $E(Y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$  eller

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 2 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 2 & 2 \\ 1 & 2 & 1 & 2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}$$

$$\underline{\epsilon} \in N(0, \sigma^2 \underline{I})$$

C. Estimatorer i submodeller

1) Model M:  $E(Y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} -0.175 \\ 1.450 \\ -1.400 \\ 2.375 \end{bmatrix}; P_M(Y) = \begin{bmatrix} 4.575 \\ 3.650 \\ -0.125 \\ 3.225 \\ 6.075 \end{bmatrix}; Y - P_M(Y) = \begin{bmatrix} +0.325 \\ -0.650 \\ +0.325 \\ -0.325 \\ +0.325 \end{bmatrix}$$

$$\frac{1}{5-4} \| Y - P_M(Y) \|^2 = \frac{0.633750}{1} = 0.633750$$

$$R^2 = \frac{21.868 - 0.633750}{21.868} = 97.1\%$$

ii) Model  $H_{12}$ :  $E(Y) = \alpha + \beta_1 x_1 + \beta_2 x_2$

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 3.026 \\ 1.243 \\ -0.987 \end{bmatrix}; P_{H_{12}}(Y) = \begin{bmatrix} 3.026 \\ 4.269 \\ 3.282 \\ 2.295 \\ 4.525 \end{bmatrix}; Y - P_{H_{12}}(Y) = \begin{bmatrix} +1.874 \\ -1.269 \\ -3.082 \\ +0.605 \\ -1.875 \end{bmatrix}$$

$$\frac{1}{5-3} \|Y - P_{H_{12}}(Y)\|^2 = \frac{18.512611}{2} = 9.2563$$

$$R^2 = \frac{21.868 - 18.512611}{21.868} = 15.3\%$$

iii) Model  $H_{13}$ :  $E(Y) = \alpha + \beta_1 x_1 + \beta_3 x_3$

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} -0.350 \\ 0.750 \\ 2.200 \\ \end{bmatrix}; P_{H_{13}}(Y) = \begin{bmatrix} 4.05 \\ 2.60 \\ 0.40 \\ 4.80 \\ 5.55 \end{bmatrix}; Y - P_{H_{13}}(Y) = \begin{bmatrix} +0.85 \\ +0.40 \\ -0.20 \\ -1.90 \\ +0.85 \end{bmatrix}$$

$$\frac{1}{5-3} \|Y - P_{H_{13}}(Y)\|^2 = \frac{5.2250}{2} = 2.6275$$

$$R^2 = \frac{21.868 - 5.2550}{21.868} = 76.0\%$$

iv) Model  $H_{23}$ :  $E(Y) = \alpha + \beta_2 x_2 + \beta_3 x_3$

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 0.945 \\ -0.872 \\ 2.309 \end{bmatrix}; P_{H_{23}}(Y) = \begin{bmatrix} 5.563 \\ 3.254 \\ 0.073 \\ 3.819 \\ 4.691 \end{bmatrix}; Y - P_{H_{23}}(Y) = \begin{bmatrix} -0.663 \\ -0.254 \\ +0.127 \\ -0.919 \\ +1.709 \end{bmatrix}$$

$$\frac{1}{5-3} \|Y - P_{H_{23}}(Y)\|^2 = \frac{4.285456}{2} = 2.1427$$

$$R^2 = \frac{21.868 - 4.2855}{21.868} = 80.4\%$$

v) Model  $H_1$ :  $E(Y) = \alpha + \beta_1 x_1$

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 2.73 \\ 0.75 \end{bmatrix}; P_{H_1}(Y) = \begin{bmatrix} 2.73 \\ 3.48 \\ 3.48 \\ 3.48 \\ 4.23 \end{bmatrix}; Y - P_{H_1}(Y) = \begin{bmatrix} +2.17 \\ -0.48 \\ -3.28 \\ -0.58 \\ +2.17 \end{bmatrix}$$

$$\frac{1}{5-2} \|Y - P_{H_1}(Y)\|^2 = \frac{20.7430}{3} = 6.9143$$

$$R^2 = \frac{21.868 - 20.743}{21.868} = 5.1\%$$

vi) Model  $H_2$ :  $E(Y) = \alpha + \beta_2 x_2$

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 3.914 \\ -0.543 \end{bmatrix}; P_{H_2}(Y) = \begin{bmatrix} 3.914 \\ 3.914 \\ 3.371 \\ 2.828 \\ 3.371 \end{bmatrix}; Y - P_{H_2}(Y) = \begin{bmatrix} +0.986 \\ -0.914 \\ -3.171 \\ +0.072 \\ +3.029 \end{bmatrix}$$

$$\frac{1}{5-2} \|Y - P_{H_2}(Y)\|^2 = \frac{21.042858}{3} = 7.0143$$

$$R^2 = \frac{21.868 - 21.043}{21.868} = 3.8\%$$

vii) Model  $H_3$ :  $E(Y) = \alpha + \beta_3 x_3$

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 2.2 \end{bmatrix}; P_{H_3}(Y) = \begin{bmatrix} 4.8 \\ 2.6 \\ 0.4 \\ 4.8 \\ 4.8 \end{bmatrix}; Y - P_{H_3}(Y) = \begin{bmatrix} +0.1 \\ +0.4 \\ -0.2 \\ -1.9 \\ +1.6 \end{bmatrix}$$

$$\frac{1}{5-2} \| \underline{Y} - P_{H_3}(\underline{Y}) \|^2 = \frac{6.38}{3} = 2.1267$$

$$R^2 = \frac{21.868 - 6.38}{21.868} = 70.8\%$$

viii) Model  $H_0: E(Y) = \alpha$

$$\hat{\alpha} = 3.48$$

$$P_{H_0}(\underline{Y}) = \begin{bmatrix} 3.48 \\ 3.48 \\ 3.48 \\ 3.48 \\ 3.48 \end{bmatrix}; \quad \underline{Y} - P_{H_0}(\underline{Y}) = \begin{bmatrix} +1.42 \\ -0.48 \\ -3.28 \\ -0.58 \\ +2.92 \end{bmatrix}$$

$$\frac{1}{5-1} \| \underline{Y} - P_{H_0}(\underline{Y}) \|^2 = \frac{21.8680}{4} = 5.4670$$

#### D. Successive testninger

1)  $H \supseteq H_{12} \supseteq H_1 \supseteq H_0$  d.v.s.:  $\beta_3 = 0, \beta_2 = 0, \beta_1 = 0$

Variation	SAK	fr.gr.
$H_0 - H_1 (\beta_1=0)$	$21.868 - 20.7430 = 1.125$	1
$H_1 - H_{12} (\beta_2=0)$	$20.7430 - 18.5126 = 2.230$	1
$H - M_{12} (\beta_3=0)$	$18.5126 - 0.6338 = 17.879$	1
M-Obs	$0.6338 = 0.634$	1
$H_0 - \text{Obs}$	21.868	4



2)  $M \supseteq H_{12} \supseteq H_2 \supseteq H_0$  d.v.s.:  $\beta_3 = 0, \beta_1 = 0, \beta_2 = 0$

Variation	SAK	fr.gr.
$H_0 - H_2$ ( $\beta_2=0$ )	$21.8680 - 21.0429 = 0.825$	1
$H_2 - H_{12}$ ( $\beta_1=0$ )	$21.0429 - 18.5126 = 2.530$	1
$H_{12} - M$ ( $\beta_3=0$ )	$18.5126 - 0.6338 = 17.879$	1
M-Obs	$0.6338 = 0.634$	1
$H_0$ -Obs	21.868	4

3)  $M \supseteq H_{13} \supseteq H_1 \supseteq H_0$  d.v.s.:  $\beta_2 = 0, \beta_3 = 0, \beta_1 = 0$

Variation	SAK	fr.gr.
$H_0 - H_1$ ( $\beta_1=0$ )	$21.8680 - 20.7430 = 1.125$	1
$H_1 - H_{13}$ ( $\beta_3=0$ )	$20.7430 - 5.2550 = 15.488$	1
$H_{13} - M$ ( $\beta_2=0$ )	$5.2550 - 0.6338 = 4.621$	1
M-Obs	$0.6338 = 0.634$	1
$H_0$ -Obs	21.868	4

4)  $M \supseteq H_{13} \supseteq H_3 \supseteq H_0$  d.v.s.:  $\beta_2 = 0, \beta_1 = 0, \beta_3 = 0$

Variation	SAK	fr.gr.
$H_0 - H_3$ ( $\beta_3=0$ )	$21.8680 - 6.38 = 15.488$	1
$H_3 - H_{13}$ ( $\beta_1=0$ )	$6.38 - 5.2550 = 1.125$	1
$H_{13} - M$ ( $\beta_2=0$ )	$5.2550 - 0.6338 = 4.621$	1
M-Obs	$0.6338 = 0.634$	1
$H_0$ -Obs	21.868	4

- 5)  $M \supseteq H_{23} \supseteq H_2 \supseteq H_0$  d.v.s.:  $\beta_1 = 0, \beta_3 = 0, \beta_2 = 0$

Variation	SAK	fr.gr.
$H_0 - H_2$ ( $\beta_2=0$ )	$21.8680 - 21.0429 = 0.825$	1
$H_2 - H_{23}$ ( $\beta_3=0$ )	$21.0429 - 4.2855 = 16.757$	1
$H_{23} - M$ ( $\beta_1=0$ )	$4.2855 - 0.6338 = 3.652$	1
M-Obs	$0.6338 = 0.634$	1
$H_0$ -Obs	21.868	4

- 6)  $M \supset H_{23} \supset H_3 \supset H_0$  d.v.s.:  $\beta_1 = 0, \beta_2 = 0, \beta_3 = 0$

Variation	SAK	fr.gr.
$H_0 - H_3$ ( $\beta_3=0$ )	$21.868 - 6.38 = 15.488$	1
$H_3 - H_{23}$ ( $\beta_2=0$ )	$6.38 - 4.2855 = 2.095$	1
$H_{23} - M$ ( $\beta_1=0$ )	$4.2855 - 0.6338 = 3.652$	1
M-Obs	$0.6338 = 0.634$	1
$H_0$ -Obs	21.868	4

E. Dispersions- og korrelationsmatrix for data

$$\text{Dispersionsmatrix} = \frac{1}{5-1} \begin{pmatrix} 2 & 1 & 0 & 1.50 \\ 1 & 2.8 & 0.4 & -1.52 \\ 0 & 0.4 & 3.2 & 7.04 \\ 1.50 & -1.52 & 7.04 & 21.868 \end{pmatrix} \begin{matrix} x_1 \\ x_2 \\ x_3 \\ y \end{matrix}$$

$x_1 \quad x_2 \quad x_3 \quad y$

$$\text{korrelationsmatrix} = \begin{pmatrix} 1 & 0.4225 & 0 & 0.2268 \\ 0.4225 & 1 & 0.1336 & -0.1942 \\ 0 & 0.1336 & 1 & 0.8416 \\ 0.2268 & -0.1942 & 0.8416 & 1 \end{pmatrix} \begin{matrix} x_1 \\ x_2 \\ x_3 \\ y \end{matrix}$$

$x_1 \quad x_2 \quad x_3 \quad y$

F. Partielle korrelationer givet  $x_3$ :

$$\begin{pmatrix} 1 & 0.4225 & 0.2268 \\ 0.4225 & 1 & -0.1942 \\ 0.2268 & -0.1942 & 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 0.1336 \\ 0.8416 \end{pmatrix} [1]^{-1} [0. \ 0.1336 \ 0.8416] =$$

$$= \begin{pmatrix} 1 & 0.4225 & 0.2268 \\ 0.4225 & 0.9822 & -0.3066 \\ 0.2268 & -0.3066 & 0.2917 \end{pmatrix},$$

d.v.s. korrelationsmatricen er

$$\begin{pmatrix} 1 & 0.4263 & 0.4199 \\ 0.4263 & 1 & -0.5728 \\ 0.4199 & -0.5728 & 1 \end{pmatrix} \begin{matrix} x_1 \\ x_2 \\ y \end{matrix}.$$

$x_1 \qquad x_2 \qquad y$

G. Partielle korrelationer givet  $x_2, x_3$ :

Først beregnet ud fra ovenstående partielle korrelationsmatrix:

$$\begin{pmatrix} 1 & 0.4199 \\ 0.4199 & 1 \end{pmatrix} - \begin{pmatrix} 0.4263 \\ 0.5728 \end{pmatrix} [1]^{-1} [0.4263 \ -0.5728] =$$

$$\begin{pmatrix} 0.8183 & 0.6641 \\ 0.6641 & 0.6719 \end{pmatrix},$$

hvorfor korrelationsmatricen bliver

$$\begin{pmatrix} 1 & 0.8956 \\ 0.8956 & 1 \end{pmatrix} \begin{matrix} x_1 \\ y \end{matrix}.$$

Som kontrol beregnes den ud fra den oprindelige kovariansmatrix:

$$\begin{aligned}
& \begin{pmatrix} 2 & 1.50 \\ 1.50 & 21.868 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ -1.52 & 7.04 \end{pmatrix} \begin{pmatrix} 2.8 & 0.4 \\ 0.4 & 3.2 \end{pmatrix}^{-1} \begin{pmatrix} 1 & -1.52 \\ 0 & 7.04 \end{pmatrix} \\
&= \begin{pmatrix} 2 & 1.50 \\ 1.50 & 21.868 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ -1.52 & 7.04 \end{pmatrix} \begin{pmatrix} 0.3636 & -0.0455 \\ -0.0455 & 0.3182 \end{pmatrix} \begin{pmatrix} 1 & -1.52 \\ 0 & 7.04 \end{pmatrix} \\
&= \begin{pmatrix} 1.6363 & 2.3727 \\ 2.3727 & 4.2855 \end{pmatrix},
\end{aligned}$$

hvorfor den partielle korrelationsmatrix er

$$\begin{pmatrix} 1 & 0.8960 \\ 0.8960 & 1 \end{pmatrix} \begin{matrix} x_1 \\ y \end{matrix}$$

$x_1$              $y$

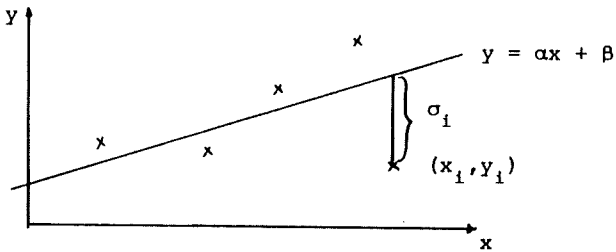
Afvigelserne i elementerne uden for diagonalen skyldes afrundingsfejl.

#### 4.4 Andre regressionsmodeller og -løsninger

I dette afsnit skal vi dels betragte nogle mere "curve-fitting" orienterede problemer, dels nogle ulineære regressionsmodeller og endelig et alternativ til mindste kvadraters metode, nemlig den såkaldte ridge-regression.

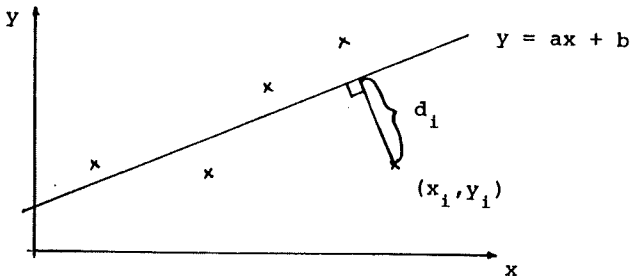
##### 4.4.1 Ortoqonal regression (lineær funktionel relation)

Ved den almindelige mindste kvadraters estimation af en regressionsflade minimaliserer man kvadratsummen af de lodrette afstande mellem regressionsfladen og de observerede punkter.



Almindelig regression:  $a$  og  $\beta$  bestemmes ved at minimalisere  $\sum \sigma_1^2$ .

Ofte vil man dog være i den situation, at det vil være mere rimeligt at minimalisere de ortogonale afstande, og så taler vi om ortogonal regression (ikke at forveksle med regression efter ortogonale polynomier).



Ortogonal regression:  $a$  og  $b$  bestemmes ved at minimalisere  $\sum d_1^2$ .

Lad der f.eks. foreligge variable  $\mu_1, \dots, \mu_p$ , som tilfredsstiller en lineær relation

$$\alpha_0 + \alpha_1 \mu_1 + \dots + \alpha_p \mu_p = 0, \quad (*)$$

d.v.s. de variable ligger i en hyperplan med ovenstående ligning. Vi er interesserede i at bestemme denne plan, i.e. i at bestemme  $\alpha_0, \dots, \alpha_p$ . Antag, at det ikke er muligt at observere værdierne  $\mu_1, \dots, \mu_p$ , men kun størrelser

$$X_{ji} = \mu_{ji} + Z_{ji}, \quad j=1, \dots, p, \quad i=1, \dots, n,$$

hvor  $Z_{ji}$ 'erne er stokastiske variable med middelværdi 0, og hvor  $\mu_{1i}, \dots, \mu_{pi}$ ,  $i=1, \dots, n$ , tilfredsstiller (\*).

Estimationen af parametrene  $\alpha_i$  på basis af et sådant sæt observationer kaldes i litteraturen ofte estimation af en lineær funktionel relation.

Det vil her være intuitivt rimeligt netop at anvende den hyperplan, der fås ved at minimalisere de ortogonale afstande ned til denne. Hvis  $Z_{ji}$ 'erne er normalt fordelte med samme varians, kan det vises (se f.eks. Kendall & Stuart (1967) p. 392), at den plan giver maksimum likelihood skønnene over  $\alpha$ 'erne.

Vi formulerer løsningen til problemet i

Sætning 4.3 Lad der være givet  $n$  punkter  $\underline{x}_1, \dots, \underline{x}_n \in \mathbb{R}^p$ . Koefficienterne i den hyperplan

$$\alpha_0 + \alpha_1 x_1 + \dots + \alpha_p x_p = 0,$$

som minimaliserer kvadratsummen af de vinkelrette afstande fra punkterne til planen, er for  $\alpha_1, \dots, \alpha_p$ 's vedkommende koordinaterne til en normeret egenvektor for den empiriske dispersionsmatrix for  $\underline{x}$ 'erne svarende til den mindste egenværdi. Den sidste koefficient er givet ved

$$\alpha_0 = -\alpha_1 \bar{x}_1 - \dots - \alpha_p \bar{x}_p.$$

Bevis Vi forudsætter altså, at vi har observationerne

$$\begin{pmatrix} x_{11} \\ \vdots \\ x_{p1} \end{pmatrix}, \dots, \begin{pmatrix} x_{1n} \\ \vdots \\ x_{pn} \end{pmatrix} .$$

Afstanden fra et punkt med koordinater  $\underline{x} = (x_1, \dots, x_p)'$  ned til hyperplanen (\*) vises let at være

$$\left| \frac{\alpha_0 + \alpha_1 x_1 + \dots + \alpha_p x_p}{\sqrt{\alpha_1^2 + \dots + \alpha_p^2}} \right| .$$

Vi skal derfor bestemme  $\alpha_0, \dots, \alpha_p$  således, at

$$f(\underline{\alpha}) = \sum_{i=1}^n \frac{(\alpha_0 + \alpha_1 x_{1i} + \dots + \alpha_p x_{pi})^2}{\alpha_1^2 + \dots + \alpha_p^2}$$

minimaliseres. Indføres  $x_{0i}$  ved  $x_{0i} = 1, \forall i$ , kan vi skrive

$$f(\underline{\alpha}) = \sum_{i=1}^n \left( \sum_{j=0}^p \alpha_j x_{ji} \right)^2 / \sum_{j=1}^p \alpha_j^2 .$$

Dette minimaliseringsproblem omformuleres bekvemt til, at vi skal finde minimum af

$$g(\underline{\alpha}) = \sum_{i=1}^n \left( \sum_{j=0}^p \alpha_j x_{ji} \right)^2$$

under bibetingelsen

$$\sum_{j=1}^p \alpha_j^2 = 1 .$$

Indføres en Lagrange-multiplikator  $\lambda$ , ses, at vi skal finde det globale minimum af

$$\varphi(\underline{\alpha}, \lambda) = \sum_{i=1}^n \left( \sum_{j=0}^p \alpha_j x_{ji} \right)^2 - \lambda \left( \sum_{j=1}^p \alpha_j^2 - 1 \right) .$$

Vi finder for  $v = 1, \dots, p$

$$\frac{\partial \varphi}{\partial \alpha_v} = 2 \sum_{i=1}^n \sum_{j=0}^p \alpha_j x_{ji} x_{vi} - 2\lambda \alpha_v ,$$

og for  $v = 0$

$$\frac{\partial \varphi}{\partial \alpha_0} = 2 \sum_{i=1}^n \sum_{j=0}^p \alpha_j x_{ji} x_{0i} = 2 \sum_{i=1}^n \sum_{j=0}^p \alpha_j x_{ji} .$$

Vi skal sætte disse partielle afledede lig 0. Den sidste ligning bliver da

$$\sum_{i=1}^n \sum_{j=0}^p \alpha_j x_{ji} = 0 ,$$

eller

$$\alpha_0 = -\alpha_1 \bar{x}_1 - \dots - \alpha_p \bar{x}_p .$$

Indsættes dette i de første ligninger, kan disse omformes til

$$\sum_{i=1}^n \sum_{j=1}^p \alpha_j (x_{ji} - \bar{x}_j) (x_{vi} - \bar{x}_v) - \lambda \alpha_v = 0 .$$

Kalder vi den empiriske dispersionsmatrix for observationerne for

$$\hat{\Gamma} = (\hat{Y}_{jv}) ,$$



ser vi, at ligningerne omformes til

$$\sum_{j=1}^p \alpha_j \hat{Y}_{jv} - \frac{\lambda}{n-1} \alpha_v = 0, \quad v=1, \dots, p.$$

Sætter vi

$$\begin{pmatrix} \alpha_1 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_p \end{pmatrix} = \underline{\alpha}_0,$$

kan ovenstående ligningssystem skrives

$$\underline{\hat{\Gamma}} \underline{\alpha} = \frac{\lambda}{n-1} \underline{\alpha},$$

d.v.s.  $\underline{\alpha}$  er en egenvektor til  $\underline{\hat{\Gamma}}$  svarende til egenværdien  $\lambda/n-1$ .

Spørgsmålet er nu, hvilken af de  $p$  egenværdier for  $\underline{\hat{\Gamma}}$  skal vi ved en konkret udregning vælge. Ved nogle manipulationer med de oprindelige ligninger (jvf. Kendall & Stuart (1967) p. 394) bringes man til at indse, at vi skal vælge den mindste egenværdi.

Vi har nu godtgjort sætningen.

□

Bemærkning Det resultat, der er anført i sætningen har snæver tilknytning til de resultater der gennemgås i kapitel 8 om principale komponenter.

#### 4.4.2 Ridge-regression

Ved analysen af regressionsmodeller, især modeller med mange uafhængige variable, løber man ofte ind i nogle stabilitetsproblemer. Det viser sig således ofte at udelukkelse af enkelte observationer kan forårsage voldsomme ændringer i størrelsen af de enkelte koefficienter, ja sågar bevirke et fortegnsskift. Disse problemer kan man søge undgået ved hjælp af forskellige former for trinvisse regressionsprocedurer - dog ikke altid med lige stort held. I stedet for at udelukke enkelte variable helt og så fokusere fuldstændigt på andre, kan man prøve at benytte lidt af den information, der ligger i alle de enkelte variable. Dette har Hoerl & Kennard (1970) gjort med den såkaldte ridge regressionsanalyse - en metode der har vist sig at give forbedrede prediktions resultater.

Vi betragter den sædvanlige model

$$Y = \underline{X}\beta + \underline{\epsilon} ,$$

hvor  $\underline{X}$  er en kendt  $n \times p$  matrix,  $\beta$  den ukendte parametervektor og  $\underline{\epsilon}$  fejlvektoren.

Vi forudsætter, at

$$E(\underline{\epsilon}) = \underline{0}$$

$$D(\underline{\epsilon}) = \sigma^2 \underline{I}_n .$$

Den ordinære LS-estimator bliver - idet vi forudsætter at  $\underline{X}$  har fuldt rang -

$$\hat{\underline{\beta}} = (\underline{X}'\underline{X})^{-1}\underline{X}'Y .$$

Vi forudsætter yderligere, at de uafhængige variable er skaleret, så  $\underline{X}'\underline{X}$  er på korrelationsform (i.e. de enkelte uafhængige variable er reduceret med deres gennemsnit og divideret med deres spredning). Denne normering skal tjene til at gøre skøn-

nene numerisk stabile, og det er altid anbefalelsesværdigt i en praktisk situation at foretage denne operation.

Hvis  $\underline{\underline{x}}'\underline{\underline{x}}$  i denne form er nær ved en enhedsmatrix, i.e. hvis de uafhængige variable nærmest er ortogonale, er mindste kvadraters skønnet  $\hat{\underline{\underline{\beta}}}$  udmærket. Hvis de uafhængige variable derimod er stærkt korrelerede (d.v.s. der er tale om en stor grad af multicollinearitet), bliver de skøn, man får frem, som ovenfor nævnt meget ustabile.

Lad os kort undersøge nogle egenskaber ved  $\hat{\underline{\underline{\beta}}}$ , som ikke er gennemgået i det tidligere. Ifølge den generelle teori haves

$$D(\hat{\underline{\underline{\beta}}}) = \sigma^2 (\underline{\underline{x}}'\underline{\underline{x}})^{-1} .$$

Sætter vi  $L$  lig afstanden fra  $\hat{\underline{\underline{\beta}}}$  til  $\underline{\underline{\beta}}$ , d.v.s.

$$L^2 = (\hat{\underline{\underline{\beta}}} - \underline{\underline{\beta}})' (\hat{\underline{\underline{\beta}}} - \underline{\underline{\beta}}) ,$$

fås

$$E(L^2) = \sum_{i=1}^p E[(\hat{\beta}_i - \beta_i)^2] = \sum_{i=1}^p V(\hat{\beta}_i) = \sigma^2 \text{tr}(\underline{\underline{x}}'\underline{\underline{x}})^{-1} .$$

Da

$$L^2 = \hat{\underline{\underline{\beta}}}'\hat{\underline{\underline{\beta}}} - 2\underline{\underline{\beta}}'\hat{\underline{\underline{\beta}}} + \underline{\underline{\beta}}'\underline{\underline{\beta}} ,$$

fås, at den forventede værdi af den kvadrerede længde af  $\hat{\underline{\underline{\beta}}}$  er

$$E(\hat{\underline{\underline{\beta}}}'\hat{\underline{\underline{\beta}}}) = \sigma^2 \text{tr}(\underline{\underline{x}}'\underline{\underline{x}})^{-1} + \underline{\underline{\beta}}'\underline{\underline{\beta}} .$$

Kaldes egenverdierne for  $\underline{\underline{x}}'\underline{\underline{x}}$  for

$$\lambda_1 \geq \dots \geq \lambda_p ,$$

fås altså (ifølge sætning 1.12 og resultatet p. 1.58)

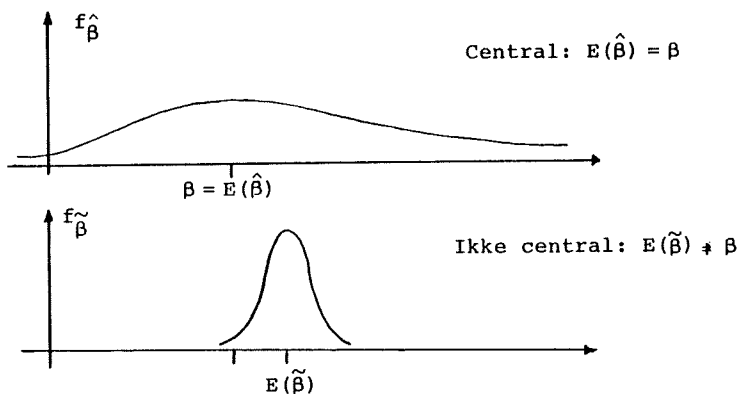
$$E(L^2) = \sigma^2 \left( \frac{1}{\lambda_1} + \dots + \frac{1}{\lambda_p} \right) > \frac{\sigma^2}{\lambda_p} ,$$

og

$$E(\hat{\underline{\beta}}' \hat{\underline{\beta}}) = \sigma^2 \left( \frac{1}{\lambda_1} + \dots + \frac{1}{\lambda_p} \right) + \underline{\beta}' \underline{\beta} > \frac{\sigma^2}{\lambda_p} + \underline{\beta}' \underline{\beta} .$$

Hvis de uafhængige variable er stærkt korrelerede, vil egenværdierne for  $\underline{x}'\underline{x}$  være meget forskellige, og den mindste derfor meget lille ( $\ll 1$ ). Ifølge ovenstående relationer vil den kvadrerede afstand mellem  $\underline{\beta}$  og  $\hat{\underline{\beta}}$  i dette tilfælde derfor have en stor forventningsværdi, og den kvadrerede længde af  $\hat{\underline{\beta}}$  vil have en forventningsværdi langt større end den kvadrerede længde af  $\underline{\beta}$ .

Denne tendens til "oppustning" af  $\hat{\underline{\beta}}$  skyldes kravet om centralitet. Spørgsmålet er, om man ved at undlade dette krav kan få målinger der i en vis forstand ligger "nærmere" ved  $\underline{\beta}$ . Problemet er skitseret i nedenstående figur.



Som det relevante kriterium kunne man anvende den såkaldte MSE = mean squared error. Den er (i det endimensionale tilfæl-

de)

$$\text{MSE} = E\{(\tilde{\beta} - \beta)^2\} = v(\tilde{\beta}) + \{E(\tilde{\beta}) - \beta\}^2 ,$$

d.v.s. lig variansen på estimatoren plus kvadratet på den såkaldte skævhed (eng. = bias). Det vil åbenbart være hensigtsmæssigt at tillade en lille skævhed, hvis dette kan give en stor reduktion i variansen og dermed i MSE. Dette opnås netop med en ridge estimator.

Definition 4.2 Ved en ridge estimator til  $\beta$  i modellen

$$Y = \underline{x}\beta + \varepsilon$$

forstås en estimator  $\hat{\beta}_k^* = \hat{\beta}^*$ , der er løsning til

$$(\underline{x}'\underline{x} + k \cdot \underline{I})\hat{\beta}^* = \underline{x}'Y ,$$

d.v.s.

$$\hat{\beta}^* = (\underline{x}'\underline{x} + k \cdot \underline{I})^{-1}\underline{x}'Y .$$

Her er  $k$  en konstant  $\in [0,1]$  .

Vi skal nu blot citere en række egenskaber for  $\hat{\beta}^*$ . Disse egenskaber skal bl.a. bruges ved fastlæggelsen af  $k$ , der ikke er givet på forhånd.

Vi har

Sætning 4.4 Lad situationen være som i definitionen. Vi sætter  $\underline{x}'Y = g$  og betegner residualkvadratafvigelsessummen for en vilkårlig estimator  $\tilde{\beta}$  for

$$H(\tilde{\beta}) = (Y - \underline{x}\tilde{\beta})'(Y - \underline{x}\tilde{\beta}) .$$

4.60

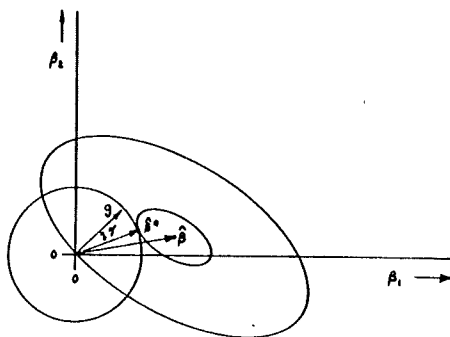
Da gælder, at gradienten af  $H$  i  $\tilde{\beta} = \underline{0}$  er proportional med og modsat rettet  $\underline{g}$ , og  $\hat{\beta}_k^*$  kan fastlægges ved, at den for fastholdt længde minimaliserer  $H(\tilde{\beta})$ , d.v.s.

$$\min_{\|\tilde{\beta}\| = \|\hat{\beta}_k^*\|} H(\tilde{\beta}) = H(\hat{\beta}_k^*) .$$

Endvidere er  $H(\hat{\beta}_k^*)$  en voksende funktion af  $k$ . Længden af  $\hat{\beta}_k^*$  er en aftagende funktion af  $k$ , og vinklen  $\gamma$  mellem  $\hat{\beta}_k^*$  og  $\underline{g}$  er en aftagende funktion af  $k$ .

Bevis Ikke særligt kompliceret, men forbigås. Læseren henvises til Hoerl og Kennard (1970) og Marquardt (1970).

Nedenstående instruktive figur er taget fra Marquardt & Snee (1975)



Den skitserer situationen geometrisk i tilfældet  $p=2$ . Punktet  $\hat{\beta}$  i centrum af ellipserne er LS-løsningen. Ellipserne er niveaukurver for  $H$ . Cirklen med centrum i origo er tangent til

den lille ellipse. Vi bemærker, at  $\hat{\beta}^*$  er den korteste vektor, der giver en residualkvadratafgivelsessum så lille som H's værdi på den lille ellipse. Endvidere bemærkes, at  $\hat{\beta}^*$  altid ligger mellem  $\hat{\beta}$  og  $g$ .

Andre hovedegenskaber ved ridge skønnet fremgår af

Sætning 4.5 Lad situationen være som ovenfor. Da er  $\hat{\beta}_k^* = \hat{\beta}^*$  en lineær transformation af  $\hat{\beta}$ , idet nemlig

$$\hat{\beta}^* = \underline{z}_k \hat{\beta} = (\underline{x}'\underline{x} + k\underline{I})^{-1} (\underline{x}'\underline{x}) \hat{\beta} .$$

$\hat{\beta}^*$  er ikke central, idet

$$E(\hat{\beta}^*) = \underline{z}_k \underline{\beta} .$$

Dispersionsmatricen for  $\hat{\beta}^*$  er

$$D(\hat{\beta}^*) = \sigma^2 [\underline{x}'\underline{x} + k\underline{I}]^{-1} (\underline{x}'\underline{x}) [\underline{x}'\underline{x} + k\underline{I}]^{-1} ,$$

og den forventede kvadratiske afstand til  $\underline{\beta}$  er

$$E[(\hat{\beta}^* - \underline{\beta})'(\hat{\beta}^* - \underline{\beta})] = \text{tr}(D(\hat{\beta}^*)) + \underline{\beta}' (\underline{z}_k - \underline{I})' (\underline{z}_k - \underline{I}) \underline{\beta} .$$

I det sidste udtryk er det første led lig variansen på den kvadrerede længde af  $\hat{\beta}^*$ , og det sidste led lig kvadratet på skævheden.

Bevis Forbigås. Elementær øvelse i matrixmanipulationer.

Af sætningen følger

Corollar Hvis  $\underline{\beta}'\underline{\beta}$  er begrænset, eksisterer der et  $k > 0$ , således at forventningen af den kvadratiske afstand mellem  $\underline{\beta}$  og

$\hat{\beta}^*$  er strengt mindre end forventningen af den kvadratiske afstand mellem  $\underline{\beta}$  og  $\hat{\beta}$ .

Bevis Dette følger ved at bemærke, at  $\text{tr}(D(\hat{\beta}^*))$  er en aftagende funktion af  $k$ , hvorimod  $\underline{\beta}'(\underline{z}_k - \underline{I})'(\underline{z}_k - \underline{I})\underline{\beta}$  er voksende. Da  $k \rightarrow 0 \Rightarrow \hat{\beta}^* \rightarrow \hat{\beta}$  ses resultatet nu umiddelbart.

Q.E.D.

Det eneste problem, der nu tilbagestår, er bestemmelsen af et rimeligt  $k$ . Her bruger man det såkaldte ridge trace, der begrundes med ovenstående corollar.

Definition 4.3 Ved et ridge trace forstås en afbildning af de enkelte koefficienter i ridge estimatet som en funktion af  $k$ .

Man benytter som nævnt ridge tracet ved fastlæggelsen af  $k$ . Filosofien bag dette er en følsomhedsanalyse-betragtning. Man ser på ens ridge trace, hvilke koefficienter der er følsomme over for ændringer i data. Man vælger så den mindste værdi af  $k$ , der giver et stabilt forløb af koefficienterne. Disse betragtninger vil vi anskueliggøre i

Eksempel 4.2 (Marquardt & Snee (1975)). I eksemplet betragtes sammenhængen mellem ASTM(\*) og gas chromatograf destillation af benzin. Standardmetoden at måle benzins flygtighed (i.e. den brøkdeler, der er fordampet ved forskellige temperaturer) er en ASTM destillation. Gas chromatograf destillationen er langt nøjagtigere, men specifikationer for benzin er angivet i ASTM-termer. Med henblik på at kunne benytte gas chromatograf målinger til on-line kontrol af flygtigheden har man

-----  
\* (American Standard for Testing of Materials).



følgelig brug for en model til at forudsige ASTM destillationen af en blanding ud fra gas chromatograf destillationen.

I dette eksempel vil vi kun se på en enkelt ASTM temperatur, nemlig 158°F. Den afhængige variabel  $y$  angiver altså den brøkdelen af benzinen, der ved en ASTM destillation er fordampet ved en temperatur på 158°F. De afhængige variable  $x_1, \dots, x_{15}$  angiver de brøkdelen, der er fordampet i de respektive temperaturintervaller ved gas chromatograf metoden. Den anvendte model er

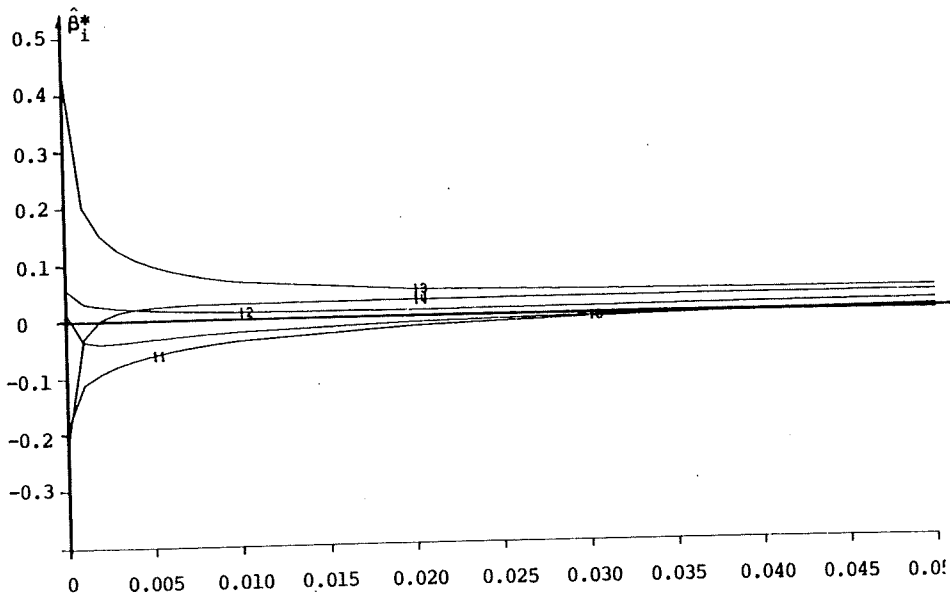
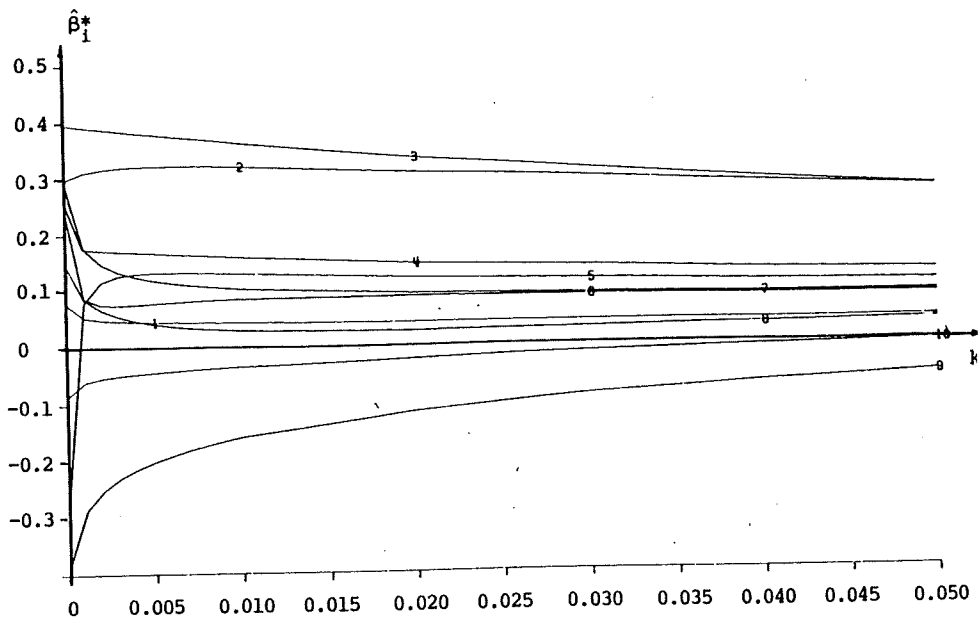
$$E(y) = \beta_1 x_1 + \dots + \beta_{15} x_{15} .$$

De uafhængige variable summerer til 1. Derfor er der ikke medtaget et konstantled.

Hovedbrugen af modellen skulle dels være forudsigelser af benzin ASTM destillationer, for hvilke forudsigelsesstandardafvigelsen skulle være  $\leq 1.5\%$ , og dels som input til lineær programmeringsberegninger af optimale blandeprocedurer. Tidligere analyser ved hjælp af mindste kvadraters metode og stepwise regressionsprocedurer havde givet koefficienter, som var uacceptable fra en fysisk synsvinkel.

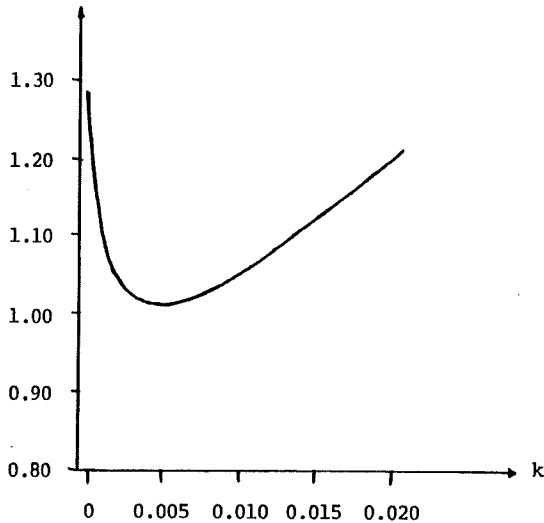
Der forelå 59 sammenhørende værdier af de 16 variable. Disse blev delt i to dele - en på 29 sæt sammenhørende værdier, der blev brugt til estimation af modellen, og en på 30, der blev brugt til at finde forudsigelsesfejl. I figuren p. 4.64 er vist ridge tracet. Man bemærker, at systemet stabiliseres for værdier omkring  $k = 0.005 - 0.01$ . I det aktuelle tilfælde vælges  $k = 0.006$ .

4.64



Ridge trace for data fra eksempel 4.2. Koefficienterne er anført i en relativ skala ( $\bar{Y}$  er sat lig 0).

Prediktionsstandardafvigelse

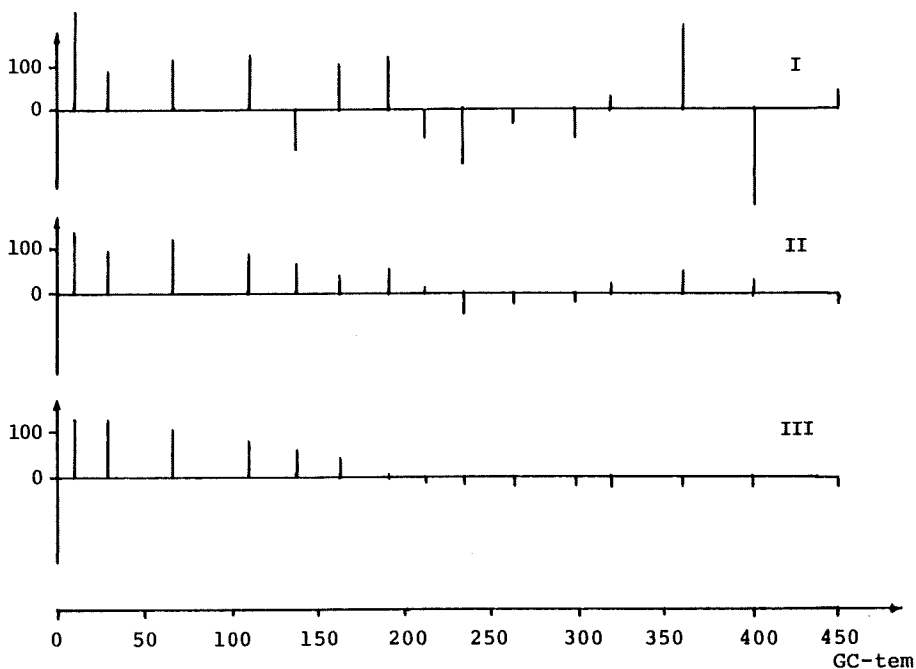


Prediktionsstandardafvigelse mod k.

I ovenstående figur vises den prediktionsstandardafvigelse, man fandt for de forskellige værdier af  $k$ . Man bemærker, at minimum indtræffer for  $k = 0.006$ , den fra ridge tracet valgte værdi. Mindste kvadraters analysen har en prediktions-standardafvigelse på 1.28 og ridge modellen på 1.01. I nedenstående figur er koefficienterne i de to modeller sammenlignet med koefficienter, der er resultatet af teoretiske overvejelser. Disse sidste kan selvsagt ikke anses for at være de "sande", men man forventer selvfølgelig en udpræget lighed.

Der er en mindre uoverensstemmelse mellem den her anførte koefficient til  $x_1$  og den i ridge tracet angivne; men det er ikke muligt ud fra kilden at afgøre, hvilken der er den korrekte.

## Koefficienter



I: Mindste kvadrater    II: Ridge    III: Teoretiske

Sammenligning af koefficienter.

Man ser, at ridge modellens koefficienter udviser et langt mere "roligt" forløb med stigende GC-temperatur, og de ligger tættere ved de "teoretiske" værdier.

□

#### 4.4.3 Ikke-lineær regression og kurvetilpasning

Ofte vil man skulle analysere regressionssituationer, der giver anledning til ikke-lineære normalligninger eller likelihoodligninger.

Man kan da selvfølgelig direkte anvende et generelt program til maksimalisering af ikke-lineære funktioner eller en iterativ metode til at løse de ikke-lineære ligninger. Dispersionsmatrixen for de herved fremkomne estimater kan da skønnes ved hjælp af den reciproke informationsmatrix, jvf. afsnit 2.6.

Vi giver et par eksempler.

Eksempel 4.3 Vi betragter nogle data, der vedrører konservering af jerngenstande fra jernalderen. Data stammer fra Eva Salomonsen (1977). På Nationalmuseets konserveringsanstalt har man i 63 år anvendt Rosenbergs glødemetode til fjernelse af klorider fra jern. For at undersøge effektiviteten af denne metode er 295 jerngenstande, konserveret i årene 1913-1974, blevet undersøgt, og antallet af defekte genstande, i.e. genstande hvor der er konstateret en fortsat nedbrydning, er opgjort for hver årgang. Tallene er anført nedenfor.

Periode	Antal undersøgte	Antal defekte	Antal defekte i % af undersøgte
1913	52	14	26.9
1921-24	34	11	32.4
1933-34	53	10	18.9
1940-43	47	13	27.7
1953-54	56	4	7.1
1961-64	46	4	8.7
1972-74	7	0	0
I alt	295	56	19.0

Antal defekte, glødede jerngenstande i forhold til det totale antal undersøgte for hver årgang.

Som det fremgår af tabellen, vokser defektprocenten med tiden, og denne vækst ønskes modelleret. En rimelig model vil vel være at sætte

$X_i$  = antal defekte for "alder"  $t_i$

$n_i$  = antal undersøgte for "alder"  $t_i$

$p_i$  = sandsynligheden for, at en genstand med "alder"  $t_i$  er defekt ,

og så postulere, at

$$X_i \in B(n_i, p_i) .$$

Som "alder" vælges selvsagt den tid, der er hengået siden glødebehandlingen. For de perioder, der udstrækker sig over flere år, er glødetidspunktet sat til midten af det betragtede tidsinterval.

Det tilbagestående problem er at fastlægge defektprocenten  $p_t$ 's afhængighed af tiden. Her er en meget benyttet model den logistiske kurve:

$$p_i = p(t_i) = \frac{1}{1 + \exp(-\alpha - \beta t_i)} .$$

Da kurven har asymptoter  $p = 0$  og  $p = 1$  og er stadigt voksende, tilfredsstiller den selvfølgelig de mest umiddelbare krav, man må stille. Defineres den såkaldte logit

$$\text{logit } p_i = \log \frac{p_i}{1 - p_i} ,$$

ses, at

$$\text{logit } p_i = \alpha + \beta t_i ,$$

d.v.s. at modellen er lineær i disse logits. Modellen er historisk især benyttet i forbindelse med bioassays, navnlig af Berkson.

Likelihoodfunktionen er

$$L(\alpha, \beta) = \prod_{i=1}^n \binom{n_i}{x_i} \left\{ \frac{1}{1 + \exp(-\alpha - \beta t_i)} \right\}^{x_i} \left\{ \frac{\exp(-\alpha - \beta t_i)}{1 + \exp(-\alpha - \beta t_i)} \right\}^{n_i - x_i}$$

og dermed

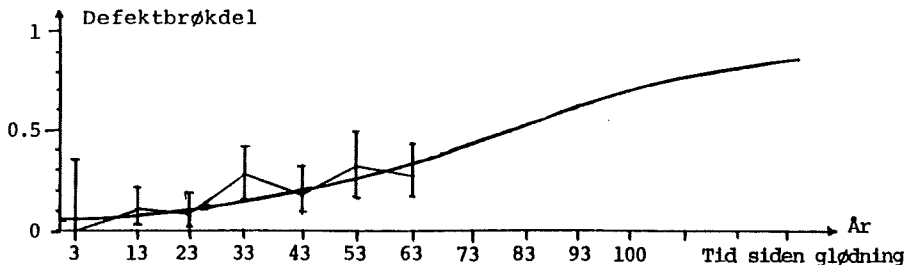
$$\begin{aligned} \log L(\alpha, \beta) &= \\ &= c - \sum_i x_i \log(1 + \exp(-\alpha - \beta t_i)) - \sum_i (n_i - x_i) (\alpha + \beta t_i) \\ &\quad - \sum_i (n_i - x_i) \log(1 + \exp(-\alpha - \beta t_i)) \\ &= c - \sum_i n_i \log(1 + \exp(-\alpha - \beta t_i)) - \sum_i (n_i - x_i) (\alpha + \beta t_i) . \end{aligned}$$

Man kan nu enten differentiere dette udtryk med hensyn til  $\alpha, \beta$  og sætte differentialkoefficienterne lig 0, eller man kan maksimalisere  $\log L(\alpha, \beta)$  ved hjælp af et generelt maksimaliseringsprogram. Ved at gøre det første er fundet

$$\hat{\alpha} = -2.99984$$

$$\hat{\beta} = 0.3813 .$$

Den resulterende logistiske kurve er indtegnet i nedenstående figur. Endvidere er der angivet et 95%-konfidensinterval omkring de enkelte observationer.



Der synes at være en rimelig overensstemmelse, men alligevel skal man nok være varsom med - skønt figuren ellers nok kunne friste - at foretage ekstrapolationer hundreder af år ud i fremtiden.

□

Ofte vil man være i en situation, hvor man ønsker at tilpasse en given datamængde med en passende glat kurve; men hvor der ikke synes at være mulighed for (eller være uforholdsmæssig besværligt) at opstille en "universel" lov, der kan dække hele det betragtede område. Man kunne da tænke sig at foretage en stykkevis approximation med forskellige funktioner. Her er en meget velegnet klasse de såkaldte spline-funktioner, der er opkaldt efter et særligt "tegneapparat".

Disse indføres i

Definition 4.3 Lad der være givet et interval  $[a, b]$  og punkter  $x_1, \dots, x_n$ , der alle ligger i intervallet. Lad der endvidere svare en værdi  $y_i$  til hvert  $x_i$ . Ved en spline af orden  $2m-1$  med knuder  $x_1, \dots, x_n$  forstås da en funktion  $\varphi$ , der tilfredsstiller

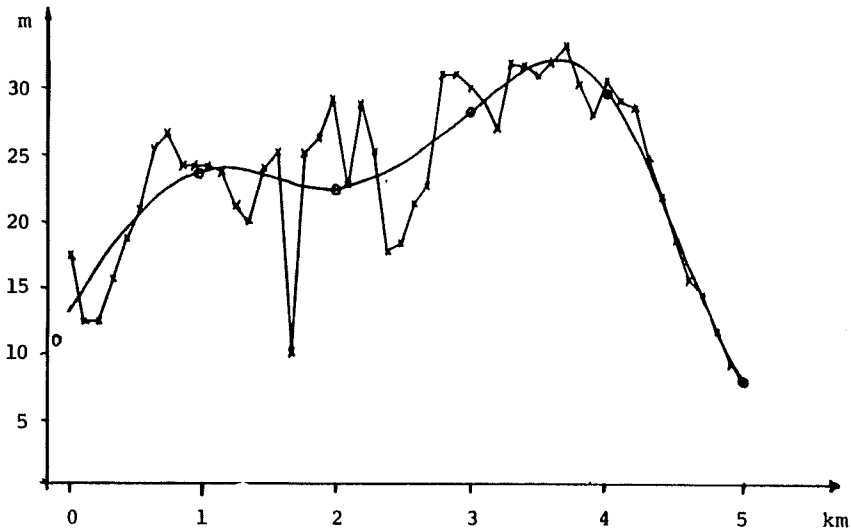
- 1)  $\varphi$  er et polynomium af grad  $2m-1$  i  $[x_i, x_{i+1}]$
- 2)  $\varphi$  er et polynomium af grad  $m-1$  i  $[a, x_1]$  og  $[x_n, b]$
- 3)  $\varphi, \varphi', \dots, \varphi^{(2m-2)}$  er kontinuerte i  $x_1, \dots, x_n$
- 4)  $\varphi(x_j) = y_j$

Bemærkning En spline af orden  $2m-1$  er altså sammensat "glat" af  $(2m-1)$ 'te grads polynomier. Det kan vises, at den således fremkomne kurve meget minder om det, man ville få ved at sømme to søm ned i hver knude og dernæst tvinge en meget elastisk stålstang igennem disse (en såkaldt tegne-spline). Dette skal vi ikke komme yderligere ind på her, men blot henvise læseren til den righoldige litteratur om emnet.



I Harwell subroutine biblioteket (1973) findes en række programmer til at approximere givne data med kubiske spline-funktioner, dels på basis af på forhånd specificerede knuder og dels ved "selv" at vælge knuder.

**Eksempel 4.4** I nedenstående figur er anført koten for en række punkter på en linie af længden 5 km i Dyrehaven. Data er stillet til rådighed af Poul Frederiksen, Institut for Landmåling og Fotogrammetri. Det er ved forskellige projekter inter-



Koter og approximerede kubiske spline.

essant at beregne et udtryk for en variation omkring en passende valgt "glad" "trendkurve" (trend = tendens). Det forekommer derfor oplagt at indlægge en kubisk spline-funktion. Dette er gjort ved hjælp af Harwell programmet VB05B, der på basis af observationer  $(x_1, y_1), \dots, (x_m, y_m)$  minimaliserer en størrelse

$$F = \sum_{i=1}^m w_i^2 \{y_i - S(x_i)\}^2 ,$$

hvor  $w_i$ 'er er brugerspecificerede vægte, og  $S$  er en kubisk spline-funktion med knuder  $j$ ,  $j = 1, \dots, n$ , hvis abscisser er specificeret af brugeren.

Den resulterende spline og dens knuder er ligeledes anført. Man bemærker den overordentlig nydelige tilpasning mellem de meget uregelmæssige observationer og spline-funktionen.

□

Referencer til kapitel 4

- Barr, A.J., Goodnight, J.H., Sall, J.P., & Helwig, J.T.:  
A Users Guide to SAS 76. SAS Institute, Raleigh 1976.
- Dixon, W.J. (ed.): BMD-Biomedical Computer Programs, University of California Press, Los Angeles 1972.
- Draper, N.R., & Smith, H.: Applied Regression Analysis.  
John Wiley & Sons, New York 1966.
- Handscomb, D.C. (ed.): Methods of Numerical Approximation.  
Pergamon Press, Oxford 1966.
- Harwell Subroutine Library. Atomic Energy Research Establishment, Harwell 1973.
- Hoerl, A.E., & Kennard, R.W.: Ridge regression. Biased estimation for nonorthogonal problems. Technometrics vol. 12, No. 1 (1970), pp. 55-67.
- Kendall, M.G., & Stuart, A.: The Advanced Theory of Statistics, vol. 2. Charles Griffin & Co., London 1967.
- Marquardt, D.W.: Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. Technometrics, vol. 12, No. 3 (1970), pp. 591-612.
- Marquardt, D.W., & Snee, R.D.: Ridge regression in practice. The American Statistician, vol. 29 (1975), pp. 3-20.
- Salomonsen, Eva: Fjernelse af klorider fra forhistorisk jern. Konservatorskolen, København 1977.
- Spliid, H.: User Guide for Stepwise Regression Program REGGO. IMSOR, Lyngby 1974.
- System/360 Scientific Subroutine Package (SSP). The International Business Machines Corporation 1970.



## KAPITEL 5

Variansanalyser5.1 Indledning

Variansanalyser er en fællesbetegnelse for en lang række statistiske metoder, der bygger på en model, der kan udtrykkes:

$$\begin{aligned} \text{Observeret værdi} &= \sum_i (\text{parametre, som repræsenterer} \\ &\quad \text{anviselige effekter} \\ &+ \sum_i (\text{stokastiske variable, som repræ-} \\ &\quad \text{senterer anviselige effekter}) \\ &+ \text{stokastisk variabel, som repræ-} \\ &\quad \text{senterer residuale effekter.} \end{aligned}$$

Udtrykt formelt:

$$X = \alpha_1 + \dots + \alpha_m + A_1 + \dots + A_n + Z .$$

Her betyder anviselig effekt en effekt, der skyldes faktorer, som vi tager hensyn til i vort eksperiment. Hvis man e.g. forestiller sig, at den observerede værdi  $X$  er udbyttet af en kemisk reaktion, vil tilstedeværelsen af en bestemt katalysator være en faktor, som vi registrerer, og som vi regner med, har en indflydelse på udbyttet. Det vil formentlig være mest

## 5.2

rimeligt at regne denne effekt som en konstant. Hvis udbyttet også afhænger af de meteorologiske forhold, vil det være rimeligt at knytte en stokastisk effekt til udtrykket. Ved at forudsætte med at indføre effekter er det oftest muligt at få fjernet det meste af residualvariationen. Man må derfor vælge at standse "indtagningen" af faktorer, når det skønnes, at residualvariationen ikke mindskes tilstrækkeligt herved.

Vi forudsætter

- 1) at forventningsværdien af hver residualvariabel  $Z$  er 0,
- 2) at de er stokastisk uafhængige,
- 3) at de har samme varians, og
- 4) at de er normalt fordelte.

Af disse betingelser er 3) den mest afgørende, idet analyserne af resultaterne hele tiden sker ved at måle forskelle mellem middelværdier relativt til residualvariansen (eller rette-  
re spredningen).

I de tilfælde, hvor vi kun har deterministiske komponenter  $\alpha_1, \dots, \alpha_m$ , er det klart, at vi har en generel, lineær model, og alle resultater om estimatorer og testning kan umiddelbart afledes som specialtilfælde af det tidligere gennemgåede. I de øvrige tilfælde må vi udlede resultaterne særskilt. Af tidsmæssige grunde vil vi dog ofte kun anføre en heuristisk begrundelse for de postulerede resultater.

Hvis alle  $A_i$ 'erne er 0, d.v.s., hvis vi har modellen

$$X = \alpha_1 + \dots + \alpha_m + Z ,$$

taler vi om en model med systematisk variation. I modellen

$$X = (\mu +) A_1 + \dots + A_n + Z ,$$

(hvor  $\mu$  er et niveaumål) taler vi om tilfældig variation. De øvrige kaldes blandede modeller. Man taler også om type I og type II variansanalyser. Type II-modellen (tilfældig variation) kaldes også en varianskomponentmodel og type I-modellen (systematisk variation) en ren parametrisk model.

Vi går nu over til at betragte den simpleste model, hvor vi kun har én faktor.

## 5.2 Ensidedet variansanalyse

En del af det følgende stof vil være en repetition af det i bind I gennemgåede om den ensidede variansanalyse, men dette er gjort for at få en sammenhængende beskrivelse af teorien.

### 5.2.1 Modeller

Vi har givet observationer

$$\begin{array}{c} X_{11}, \dots, X_{1n_1} \\ \vdots \\ X_{k1}, \dots, X_{kn_k} \end{array}$$

svarende til  $k$  niveauer af en enkelt faktor. (Hvis man e.g. måler temperaturen i noget smeltet tin med 5 forskellige termometre, har vi én faktor, nemlig termometer, og denne har 5 "niveauer").

Vi har de mulige modeller

$$\text{I) } X_{ij} = \mu + \alpha_i + Z_{ij}, \quad i=1, \dots, k, \quad j=1, \dots, n_i$$

$$\text{II) } X_{ij} = \mu + A_i + Z_{ij}, \quad i=1, \dots, k, \quad j=1, \dots, n_i .$$

## 5.4

Her betegner  $\mu$  et generelt niveau, og  $\alpha_i$  henholdsvis  $A_i$  angiver effekten af, at vor måling er foretaget med faktoren i niveau  $i$ .

Vi forudsætter

$$I) \quad \sum_i n_i \alpha_i = 0$$

$$Z_{ij} \in N(0, \sigma^2), \text{ stokastisk uafhængige for alle } i, j.$$

$$II) \quad \left. \begin{array}{l} A_i \in N(0, \sigma_a^2) \\ Z_{ij} \in N(0, \sigma^2) \end{array} \right\} \text{ stokastisk uafhængige for alle } i, j.$$

Vi vil nu kort belyse forskellen på de to modeller.

I tilfælde I har vi, at

$$\text{Cov}(X_{i\nu}, X_{i\mu}) = \text{Cov}(Z_{i\nu}, Z_{i\mu}) = 0, \quad \mu \neq \nu,$$

og derfor er observationerne i samme gruppe ukorrelerede og dermed uafhængige (grundet normaliteten). I tilfælde II gælder derimod, stadig for  $\mu \neq \nu$ ,

$$\begin{aligned} \text{Cov}(X_{i\nu}, X_{i\mu}) &= \text{Cov}(A_i + Z_{i\nu}, A_i + Z_{i\mu}) \\ &= \text{Cov}(A_i, A_i) \\ &= \sigma_a^2, \end{aligned}$$

d.v.s. målinger fra samme gruppe er ikke uafhængige.

Ser vi alene på den enkelte observation gælder

$$I) \quad V(X_{ij}) = V(Z_{ij}) = \sigma^2$$



$$\text{II)} \quad V(X_{ij}) = V(A_i) + V(Z_{ij}) = \sigma^2 + \sigma_a^2 .$$

Vi har altså, at variansen i en type II model er lig summen af residualvariansen  $\sigma^2$  og variansen på den tilfældige komponent  $A_i$ . Heraf fremgår også, hvorfor type II-modellen kaldes en varianskomponentmodel.

Den væsentlige forskel mellem de to modeller er, at vi i den ene situation vælger at tolke forskelle som udtryk for deterministiske afvigelser og i den anden som udtryk for stokastiske. Hvilket valg, der er det rimelige, afhænger selvfølgelig af forsøgsomstændighederne og af, hvad resultatet skal bruges til, i.e. af, hvorledes en gentagelse af experimentet ville være. Dette forhold vil vi belyse nedenfor i et eksempel.

Først vil vi dog lige præcisere det "tekniske" mål for den statistiske analyse. Det vil i første omgang være: at estimere de involverede parametre, i.e.

$$\alpha_1, \dots, \alpha_k ,$$

respektive

$$\sigma_a^2 ,$$

og i anden omgang: at teste hypoteserne

$$H_I : \alpha_1 = \dots = \alpha_k = 0 \text{ mod } K_I : \exists \alpha_i \neq 0 ,$$

respektive

$$H_{II} : \sigma_a^2 = 0 \text{ mod } K_{II} : \sigma_a^2 > 0 .$$

Accept af disse tests giver begge, at niveauet af faktoren ingen betydning har. Den eneste variation, der er i materialet, er residualvariationen.

## 5.6

Lad os søge at klargøre forskellen på de to modeller ved hjælp af et

Eksempel 5.1 I nedenstående tabel er anført resultatet af en række målinger af trækstyrke af noget gummi (målt i pounds per square inch.)

A	B	C	D
3210	3225	3220	3545
3000	3320	3410	3600
3315	3165	3320	3580
	3145	3370	3485

Hvis betegnelserne A, B, C, D angiver, hvilken af 4 forskellige produktionsmetoder, der er anvendt ved fremstillingen af de gummitråde, der danner basis for de pågældende målinger, vil det være mest rimeligt at anvende den systematiske model

$$X_{ij} = \mu + \alpha_i + Z_{ij} ,$$

hvor  $\alpha_i$  angiver effekten af den  $i$ 'te metode.

Hvis vi derimod blot har udtaget 4 forskellige kasser fra en og samme fabriks lager, og dernæst målt styrken af gummi fra disse kasser med henblik på at få et skøn over styrken af fremtidige gummiprøver, vil det være mere rimeligt at anvende modellen

$$X_{ij} = \mu + A_i + Z_{ij} ,$$

hvor  $A_i$  altså angiver den tilfældige afgivelse, vi har fra middelkvaliteten i denne tilfældigt valgte kasse. Hvis vi her havde anvendt en systematisk model, kunne vi (formelt) estimere

et  $\alpha_i$ . Et sådant  $\alpha_i$  vil imidlertid ikke være af nogen synderlig interesse. Hvis man en anden dag tilfældigt udvælger sig en kasse, har man ikke fået nogen oplysning om kvaliteten i denne kasse, selv om man kender  $\alpha_1, \dots, \alpha_4$ . Hvis man derimod anvender model II og anfører et estimat for  $\sigma_a^2$ , kan vi regne med, at kvaliteten af gummiet i den foreliggende kasse "nok er  $\mu$  + et tilfældigt bidrag fra en  $N(0, \sigma_a^2)$ -fordeling" (plus et tilfældigt bidrag fra residualen). (Vi må lige præcisere, at det i den systematiske model selvfølgelig er overordentligt interessant at få estimeret parametrene  $\alpha_i$ , idet vi jo så kan angive et skøn over trækstyrken for gummi, produceret efter de 4 forskellige metoder).

Vi ser her et eksempel på, hvor det er afgørende at gøre sig klar hvad en gentagelse af eksperimentet vil være. Hvis en gentagelse her i tilfælde 2 blot ville være tilfældigt at udvælge gummiprøver fra en af de allerede udvalgte 4 kasser, ja da vil vi selvfølgelig skulle anvende den systematiske model. Det er jo da interessant om vi vælger fra en kasse, som vi på forhånd har fundet er "god", eller om vi vælger fra en kasse som er "dårlig". Den en gang erkendte forskel mellem de fire kasser vil vi selvfølgelig ikke se bort fra, og blot kalde tilfældig, hvis vi skal udtage prøver fra de selvsamme kasser.

□

### 5.2.2 Analyse af den systematiske model

Fra det elementære kursus vides, at variansanalyseeskemaet bliver (hypotese  $\alpha_1 = \dots = \alpha_k = 0$ )

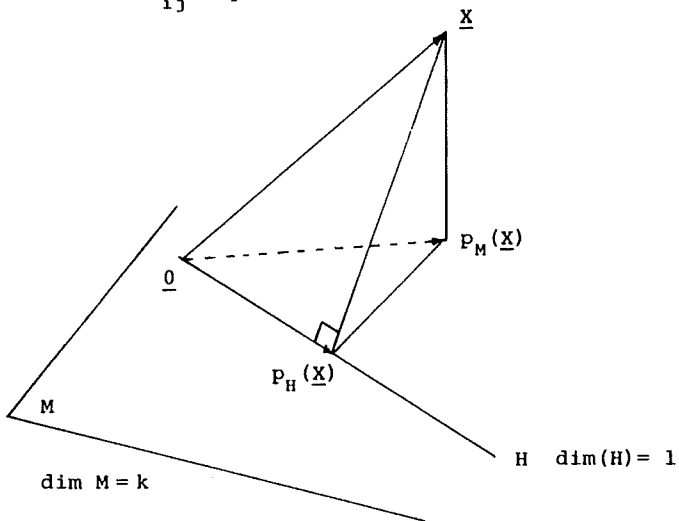
Variation	SAK	Fr.g.
Hypotese-model	$\sum_i n_i (\bar{X}_i - \bar{X})^2$	$k - 1$
Model-observationer	$\sum_{ij} (X_{ij} - \bar{X}_i)^2$	$N - k$
Hypotese-observationer	$\sum_{ij} (X_{ij} - \bar{X})^2$	$N - 1$

hvor

$$N = \sum_i n_i$$

$$\bar{X}_i = \frac{1}{n_i} \sum_j X_{ij}$$

$$\bar{X} = \frac{1}{N} \sum_{ij} X_{ij}$$



Udtrykt i sproget fra den generelle model har vi, at  $M$  er det underrum af  $R^N$ , hvor de  $n_1$  første koordinater er ens, de  $n_2$  næste koordinater er ens, ..., de  $n_k$  sidste koordinater er ens.  $H$  er det underrum af  $M$ , hvor samtlige koordinater er ens. Det er klart, at  $\dim M = k$  og  $\dim H = 1$ . Vi har da med den sædvanlige norm

$$\| \underline{X} - p_M(\underline{X}) \|^2 = \sum_{ij} (X_{ij} - \bar{X}_i)^2$$

$$\| p_H(\underline{X}) - p_M(\underline{X}) \|^2 = \sum_i n_i (\bar{X}_i - \bar{X})^2 .$$

Spaltningssætningen giver umiddelbart, at disse er stokastisk uafhængige og  $\sigma^2 \chi^2(f)$ -fordelte under  $H_0$ . Hermed er den i bind 1 postulerede uafhængighed af disse kvadratafvigelsessummer godtgjort. Vi vil nu bestemme forventningsværdierne af disse størrelser - også uden betingelsen, at  $H_0$  er rigtig.

For nemheds skyld anvender vi betegnelserne

$$SAK_1 = \| \underline{X} - p_M(\underline{X}) \|^2 = \sum_{ij} (X_{ij} - \bar{X}_i)^2 ,$$

og

$$SAK_2 = \| p_M(\underline{X}) - p_H(\underline{X}) \|^2 = \sum_i n_i (\bar{X}_i - \bar{X})^2 .$$

Vi har altid, at

$$SAK_1 \in \sigma^2 \chi^2(N - k) .$$

Da

$$\begin{aligned} SAK_2 &= \sum_i n_i (\bar{X}_i - \bar{X})^2 = \sum_i n_i (\bar{Z}_i - \bar{Z} + \alpha_i)^2 \\ &= \sum_i n_i (\bar{Z}_i - \bar{Z})^2 + \sum_i n_i \alpha_i^2 + 2 \sum_i n_i \alpha_i (\bar{Z}_i - \bar{Z}) , \end{aligned}$$

## 5.10

er

$$\begin{aligned}
 E(\text{SAK}_2) &= E(\sum_i n_i (\bar{z}_i - \bar{z})^2) + \sum_i n_i \alpha_i^2 + 2 \sum_i n_i \alpha_i \cdot 0 \\
 &= \sigma^2(k-1) + \sum_i n_i \alpha_i^2 .
 \end{aligned}$$

Leddet  $\sigma^2(k-1)$  skyldes, at størrelserne  $Z_{ij}$  er  $\in N(0, \sigma^2)$  og derfor tilfredsstillter hypotesen om fuldstændig homogenitet. En ensidet variansanalyse på  $Z_{ij}$ 'erne giver da (v.h.a. spaltningssætningen)

$$\sum_i n_i (\bar{z}_i - \bar{z})^2 \in \sigma^2 \chi^2(k-1) ,$$

og heraf følger resultatet. Vi kan derfor supplere variansanalysekemaet p. 5.8 med en søjle med de forventede værdier af  $\text{SAK}_i/f_i$ 'erne:

$S^2 = \frac{\text{SAK}}{f}$	$E(S^2)$
$S_2^2 = \frac{\text{SAK}_2}{f_2}$	$\sigma^2 + \frac{1}{k-1} \sum_i n_i \alpha_i^2$
$S_1^2 = \frac{\text{SAK}_1}{f_1}$	$\sigma^2$

Vi bemærker, at disse middelværdier smukt anskueliggør, hvorfor et test af hypotesen

$$H_I : \alpha_1 = \dots = \alpha_k = 0 \text{ mod } K_I : \exists i : \alpha_i \neq 0 .$$

må have et kritisk område af formen

$$\{ \underline{x} \mid \frac{s_2^2}{s_1^2} > c \} .$$

### 5.2.3 Analyse af den tilfældige model

Vi er her som nævnt interesseret i test af hypotesen

$$H_0 : \sigma_a^2 = 0 \quad \text{mod} \quad H_1 : \sigma_a^2 > 0 .$$

Hvis hypotesen er sand, er  $A_i$  lig 0 med sandsynligheden 1, d.v.s. vi har, at

$$X_{ij} = \mu + Z_{ij} ,$$

hvor  $Z_{ij} \in N(0, \sigma^2)$ . Der er da ingen særlig gruppeeffekt - resultaterne afviger kun fra hinanden ved den tilfældige fejl  $Z_{ij}$ . Vi analyserer nu de samme kvadratafgivelsessummer, som vi havde i den systematiske model. Vi har

$$\begin{aligned} \text{SAK}_1 &= \sum_{ij} (X_{ij} - \bar{X}_i)^2 \\ &= \sum_{ij} (Z_{ij} - \bar{Z}_i)^2 \in \sigma^2 \chi^2(N - k) . \end{aligned}$$

Under  $H_0$  er

$$\begin{aligned} \text{SAK}_2 &= \sum_i n_i (\bar{X}_i - \bar{X})^2 \\ &= \sum_i n_i (\bar{Z}_i - \bar{Z})^2 \in \sigma^2 \chi^2(k - 1) , \end{aligned}$$

og denne størrelse er uafhængig af  $\text{SAK}_1$  (brug e.g. spaltnings-sætningen på de variable  $Z_{ij}$ ). Generelt har vi

$$\begin{aligned}
 \text{SAK}_2 &= \sum n_i (\bar{X}_i - \bar{X})^2 \\
 &= \sum n_i (\bar{Z}_i - \bar{Z} + A_i - \bar{A})^2 \\
 &= \sum n_i (\bar{Z}_i - \bar{Z})^2 + \sum n_i (A_i - \bar{A})^2 \\
 &\quad + 2 \sum n_i (\bar{Z}_i - \bar{Z})(A_i - \bar{A}) ,
 \end{aligned}$$

hvor

$$\bar{A} = \frac{1}{N} \sum_i n_i A_i .$$

Heraf fås

$$E(\text{SAK}_2) = \sigma^2(k-1) + E(\sum_i n_i (A_i - \bar{A})^2) + 0 .$$

Nu er

$$\sum_i n_i (A_i - \bar{A})^2 = \sum_i n_i A_i^2 - N\bar{A}^2 ,$$

og dermed

$$\begin{aligned}
 E(\sum_i n_i (A_i - \bar{A})^2) &= \sum_i n_i E(A_i^2) - NE(\bar{A}^2) \\
 &= \sum_i n_i \sigma_a^2 - N \frac{1}{N^2} \sum_i n_i^2 \sigma_a^2 \\
 &= \sigma_a^2 (N - \frac{1}{N} \sum_i n_i^2) .
 \end{aligned}$$

Det andet lighedstegn følger let, når man erindrer, at  $E(A_i^2) = V(A_i) + E(\bar{A}^2) = V(\bar{A}) + E(\bar{A}^2)$ , da disse stokastiske variable har forventningsværdierne 0.



Samles ovenstående, fås

$$E(\text{SAK}_2 / (k - 1)) = \sigma^2 + \frac{1}{k-1} (N - \frac{1}{N} \sum n_i^2) \sigma_a^2 .$$

Er alle  $n_i$ 'er lig  $n$ , fås specielt

$$E(S_2^2) = E(\text{SAK}_2 / (k - 1)) = \sigma^2 + n\sigma_a^2 .$$

Da  $E(S_1^2) = \sigma^2$ , og da  $S_2^2/S_1^2 \in F(k - 1, N - k)$  under  $H_0$ , ses, at også i den tilfældige model er det kritiske område af formen

$$\{ \underline{x} \mid \frac{S_2^2}{S_1^2} > F(k - 1, N - k)_{1-\alpha} \} ,$$

denne gang er det blot hypotesen  $\sigma_a^2 = 0$ , vi undersøger.

Vi har altså fået den samme teststørrelse og det samme kritiske område som under analysen af den parametriske model. Dette kunne måske forlede læseren til at tro, at de to situationer i virkeligheden ikke adskiller sig fra hinanden. At dette ikke er tilfældet, kan bl.a. ses af, at teststørrelsernes fordeling uden  $H_0$  ikke er ens.

I det parametriske tilfælde er

$$\frac{S_2^2}{S_1^2} \in F(k - 1, N - k; \frac{1}{\sigma^2} \sum n_i \alpha_i^2) ,$$

altså ikke-centralt F-fordelt. Dette resultat er, om end en ikke helt triviel følge af, så dog ikke uforståeligt ved sammenligning med definitionen på ikke-centrale  $\chi^2$ - og F-fordelinger.

Hvis vi indskrænker os til det balancerede tilfælde (i.e. hvor alle  $n_i$ 'er er ens), har vi i den tilfældige model

5.14

$$\begin{aligned} \text{SAK}_2 &= n \sum_i [(\bar{z}_i + A_i) - (\bar{z} + \bar{A})]^2 \\ &= n \sum_i (V_i - \bar{V})^2, \end{aligned}$$

hvor

$$V_i = \bar{z}_i + A_i \in N(0, \frac{1}{n} \sigma^2 + \sigma_a^2).$$

Derfor er

$$\text{SAK}_2 \in (\sigma^2 + n\sigma_a^2) \chi^2(k - 1).$$

Da det endvidere indses, at  $\text{SAK}_2$  og  $\text{SAK}_1$  fremdeles er uafhængige, har vi altså

$$\begin{aligned} \frac{S_2^2}{S_1^2} &\in \frac{1}{\sigma^2} (\sigma^2 + n\sigma_a^2) F(k - 1, N - k) \\ &= (1 + n\sigma_a^2/\sigma^2) F(k - 1, N - k), \end{aligned}$$

d.v.s. centralt F-fordelt med skalaparameter  $1 + n\sigma_a^2/\sigma^2$ ,  
d.v.s. et helt andet resultat, end vi havde i den parametriske  
model.

#### 5.2.4 Resumé af analyserne og et eksempel

Vi opstiller nu et variansanalysekema med de forventede værdier af  $S^2$ -størrelserne indsat.

Variation	SAK	D.F.	$S^2$	$E(S^2)$	
				Systematisk model	Tilfældig model
Mellem grup.	$SAK_2 = \sum_i n_i (\bar{X}_i - \bar{X})^2$	$k-1$	$S_2^2$	$\sigma^2 + \frac{1}{k-1} \sum_i n_i \alpha_i^2$	$\sigma^2 + \frac{1}{k-1} [N - \frac{1}{N} \sum_i n_i^2] \sigma_a^2$
Inden f.grup.	$SAK_1 = \sum_{ij} \sum (X_{ij} - \bar{X}_i)^2$	$N-k$	$S_1^2$	$\sigma^2$	$\sigma^2$
Total	$SAK_0 = \sum_{ij} \sum (X_{ij} - \bar{X})^2$	$N-1$			

Det ses, at et centralt skøn over  $\sigma_a^2$  er

$$\hat{\sigma}_a^2 = (S_2^2 - S_1^2) / \left( \frac{1}{k-1} (N - \frac{1}{N} \sum_i n_i^2) \right) .$$

Dette skøn kan antage negative værdier, skønt det er et estimat over en ikke-negativ størrelse, nemlig  $\sigma_a^2$ . Det er i sådanne tilfælde sædvanen at sætte  $\hat{\sigma}_a^2 = 0$ .

De kritiske områder for test af

$$\alpha_1 = \dots = \alpha_k = 0$$

mod alle alternativer, respektive for test af

$$\sigma_a^2 = 0 \quad \text{mod} \quad \sigma_a^2 > 0 ,$$

er ens og lig

$$\{ (x_{11}, \dots, x_{kn_k}) \mid s_2^2/s_1^2 > F(k-1, N-k)_{1-\alpha} \}$$

ved test på niveau  $\alpha$ . Fordelingerne af  $S_2^2/S_1^2$  i de to tilfælde er forskellige, når  $H_0$  ej er opfyldt.

## 5.16

Eksempel 5.2 Vi betragter de p. 5.6 anførte data. Af hensyn til beregningerne koder vi data v.h.a. relationen

$$x'_{ij} = \frac{x_{ij} - 3000}{5} .$$

Vi udfører nu de sædvanlige beregninger til en ensidet variansanalyse og får variansanalysekemaet

Variation	SAK	Fr. gr.	S <sup>2</sup>	E(S <sup>2</sup> )	
				systematisk model	tilfældig model
Mellem grup.	12961.98	3	4320.66	$(\sigma^2 + \frac{1}{3} \sum \alpha_i^2) \frac{1}{25}$	$(\sigma^2 + 3.73 \sigma_a^2) \frac{1}{25}$
Inden f.grup.	3911.75	11	355.61		
Total	16873.73	14			

Da

$$\frac{4320.66}{355.61} = 12.15 \approx F(3, 11) .9991 ,$$

forkastes hypotesen

$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0 .$$

(hvis vi arbejder med den parametriske model), henholdsvis hypotesen

$$\sigma_a^2 = 0$$

(hvis vi arbejder med den tilfældige model).

Arbejder vi med den systematiske model  $X_{ij} = \mu + \alpha_i + Z_{ij}$ , må vi konkludere, at det ikke kan antages, at metoderne er ens. Vi har estimerterne

$$\begin{aligned}\hat{\mu} &= 3327.33 \\ \hat{\alpha}_1 &= -152.33 \\ \hat{\alpha}_2 &= -113.58 \\ \hat{\alpha}_3 &= 2.67 \\ \hat{\alpha}_4 &= 225.17 ,\end{aligned}$$

og

$$\hat{\sigma}^2 = 25 \cdot 355.61 = 94.28^2 .$$

Arbejder vi med den tilfældige model  $X_{ij} = \mu + A_i + Z_{ij}$ , har vi, at vi må antage, at der er tilfældige fluktuationer imellem de forskellige kasser. Disse udtrykkes ved estimerterne

$$\begin{aligned}\hat{\mu} &= 3327.33 \\ \hat{\sigma}^2 &= 25 \cdot 355.61 = 94.28^2 \\ \hat{\sigma}_a^2 &= 26551.67 = 162.95^2\end{aligned}$$

□

Vi har i dette afsnit gjort grundigt rede for, hvorledes sammenhængen er mellem analysen af en tilfældig model og en tilsvarende systematisk. Vi vil derfor i de følgende afsnit nøjes med at præcisere de relevante resultater uden at give et bevis for disse.

### 5.3 Tosidet variansanalyse. Hierarkisk klassifikation og krydsklassifikation

#### 5.3.1 Hierarkisk klassifikation og krydsklassifikation

Vi skal nu beskæftige os med analysen af forsøg, hvor man undersøger virkningen af 2 faktorer A og B. Alt efter formålet med en sådan undersøgelse får vi 2 typer af modeller, nemlig den såkaldte krydsklassifikation og den såkaldte hierarkiske klassifikation. Lad os give et lille eksempel, inden vi anfører de stringente definitioner.

Eksempel 5.3 Ved den afregning, en kornproducent modtager for noget leveret korn, indgår vandindholdet i % som en meget afgørende faktor. Lad os antage, at en bonde har 3 siloer med korn, som han ønsker at sælge. Kornhandleren tager e.g. 3 stikprøver fra 4 tilfældigt valgte steder i hver silo (ikke nødvendigvis de samme steder i hver silo) og måler vandindholdet i kornet. Vi har altså målinger

$$X_{ijv}, \quad \begin{array}{ll} i = 1, 2, 3 & \text{(silonr.)} \\ j = 1, 2, 3, 4 & \text{(målested nr.)} \\ v = 1, 2, 3 & \text{(gentagelsesnr.)} \end{array} .$$

Som en model til beskrivelse af måleresultaterne kan vi vælge

$$X_{ijv} = \mu + \alpha_i + T_{j(i)} + Z_{v(ij)} ,$$

$i = 1, 2, 3$  (silonr.),  $j = 1, 2, 3, 4$  (målestednr.),  
 $v = 1, 2, 3$  (gentagelsesnr.). Her betegner  $\mu$  et generelt niveau,  $\alpha_i$  den  $i$ 'te silos afvigelse fra niveauet og  $T_{j(i)}$  den (tilfældige) afvigelse fra niveauet i den  $i$ 'te silo, som der er det  $j$ 'te målested i denne silo.  $Z_{v(ij)}$  er residualen. Vi skal senere kommentere den specielle anvendelse af parenteser ved indiceringen.

Lad os nu betragte en anden situation, hvor bonden er interesseret i, hvorledes vandindholdet i kornet fordeler sig på 4 specielt valgte steder i siloerne (e.g. ved bunden,  $\frac{1}{3}$  oppe,  $\frac{2}{3}$  oppe og ved toppen). Hvis han derfor lader udtage 3 prøver fra hvert af stederne i hver silo og måler vandindholdet, har vi igen målinger

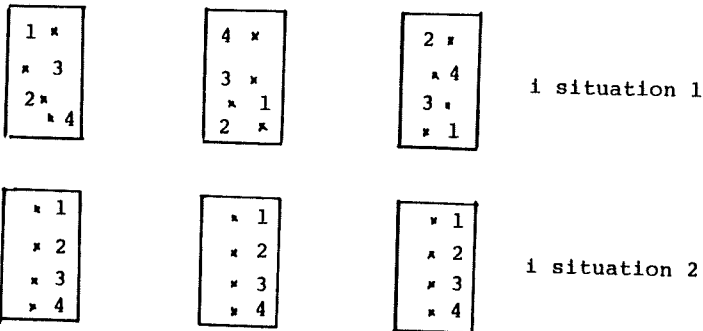
$$X_{ijv}, \quad i = 1, 2, 3 \\ j = 1, 2, 3, 4 \quad \text{og} \\ v = 1, 2, 3 .$$

Her vil vi vælge modellen

$$X_{ijv} = \mu + \alpha_i + \beta_j + \theta_{ij} + Z_{v(ij)} .$$

Her er  $\mu$  og  $\alpha_i$  fremdeles niveau og den  $i$ ,te silos afvigelse fra dette.  $\beta_j$  angiver målesteds afvigelse fra niveauet og  $\theta_{ij}$  vekselvirkningen mellem målested og silo.  $Z_{v(ij)}$  er residualt.

#### Målesteder



Vi ser, at det ikke ville være rimeligt at anvende denne model i det første tilfælde; thi dér var der ingen sammenhæng mellem målestederne i de forskellige siloer. Der havde de tre prøver fra målested nr. 1 i silo 1 intet til fælles med de tre

prøver fra målested nr. 1 i silo 2 (andet end, at de tilfældigvis er taget som de første i hver silo; det første målested kan derimod være i bunden i den ene silo i midten i den anden etc.). I den anden situation stammer de begge fra e.g. toppen af de respektive siloer, og det vil derfor være rimeligt at indføre en stedparameter  $\beta_j$ . Det er disse parametre, som undersøgelse nr. 2 sigter mod at få belyst (estimeret), således at man kan komme med udsagn som, at kornet i toppen af en silo indeholder gennemsnitligt  $\frac{1}{2}\%$  mindre vand end det ved bunden af siloen etc. I undersøgelse nr. 1 er det af hovedinteresse at få estimeret  $\mu$  og  $\alpha_i$ , således at man kan udtale sig om det gennemsnitlige vandindhold i hver af de 3 siloer. Hvis man er interesseret i, hvor meget vandindholdet varierer i de enkelte siloer, kan man estimere  $V(T_{j(i)})$ .

Vi ser altså, at vi med de  $3 \times 4 \times 3$  målinger har 2 mulige modeller. Hvilken der skal anvendes, afhænger som sagt af, på hvilken måde data er indsamlet og med hvilket formål.

□

For at kunne skelne de to situationer fra hinanden vælger vi den anførte skrivemåde med at anbringe  $i$  i en parentes efter  $j$ , hvis  $j$  er underordnet i d.v.s. hvis det er sådan, at sted  $j$  i den  $i_1$ 'te silo intet har tilfælles med sted  $j$  i den  $i_2$ 'te silo. Analogt gælder selvfølgelig for en indicering som  $v(ij)$  etc.

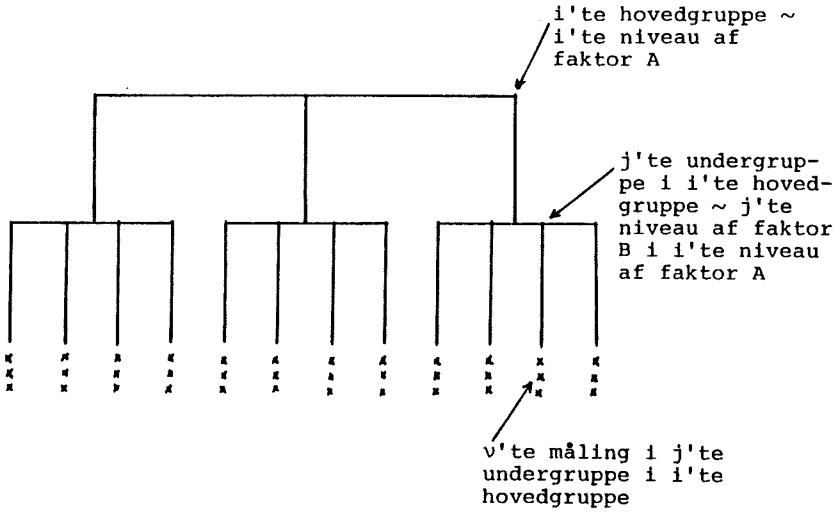
Efter dette eksempel er vi nu i stand til at formulere de to typer af modeller.

Overalt i resten af dette afsnit betragter vi observationer

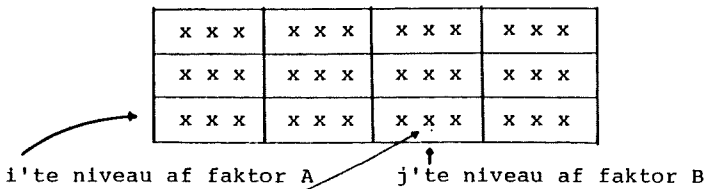
$$\begin{aligned} X_{ijv}, \quad i &= 1, \dots, k \text{ svarende til faktor A} \\ j &= 1, \dots, m \text{ svarende til faktor B} \\ v &= 1, \dots, n \text{ svarende til gentagelser.} \end{aligned}$$

Vi taler om en hierarkisk klassifikation, hvis data "hænger" sammen på følgende måde





Her er faktor B underordnet A, og vi skriver  $A \supset B$ .  
 Hvis sammenhængen er



v'te måling taget i i'te niveau af faktor A og j'te niveau af faktor B.

taler vi om en krydsklassifikation. Vi skriver  $A \times B$ .

I den hierarkiske klassifikation har vi følgende modeller

$$\text{I: } X_{ijv} = \mu + a_i + b(a)_{j(i)} + z_{v(ij)}$$

$$\text{II: } X_{ijv} = \mu + A_i + B(A)_{j(i)} + z_{v(ij)}$$

$$\text{III: } X_{ijv} = \mu + a_i + B(a)_{j(i)} + z_{v(ij)}$$

Vi har her valgt at følge den tidligere omtalte regel for indiceret: de overordnede index er anført i en parentes.

Endvidere har vi angivet de enkelte parametre med samme bogstaver som de faktorer, de repræsenterer. Skal parameteren opfattes deterministisk, er valgt et lille bogstav; skal den opfattes stokastisk er valgt et stort. Optræder der flere bogstaver i en parameter, skal den opfattes stokastisk, blot eet af bogstaverne er stort.

Bemærk, at når vi har valgt en betegnelse som  $b(a)_{j(i)}$  angives ikke dermed et produkt. Der er blot tale om en parameter  $\theta_{j(i)}$ , som vi af mnemotekniske grunde kalder  $b(a)_{j(i)}$ .

Vi forudsætter, at

$$\sum_i a_i = 0$$

$$\sum_j b(a)_{j(i)} = 0 \quad \forall i$$

$$z_{v(ij)} \in N(0, \sigma^2)$$

$$A_i \in N(0, \sigma_A^2)$$

$$B(A)_{j(i)} \text{ (resp. } B(a)_{j(i)}) \in N(0, \sigma_{B(A)}^2)$$

} Stokastisk  
uafhængige

Vi bemærker, at variansen svarende til  $B(A)_{j(i)}$  og  $B(a)_{j(i)}$  er indiceret  $B(A)$ , for at angive, at  $B$  er underordnet  $A$ .

Vi kalder I en systematisk eller rent parametrisk model, II en

stokastisk eller en varianskomponentmodel og III en blandet model.

Man kan formelt også betragte blandede modeller af formen

$$X_{ijv} = \mu + A_i + b(a)_{j(i)} + Z_{v(ij)} ,$$

men dette forekommer ikke særlig rimeligt ("vekselvirkningen" er deterministisk og hovedeffekten stokastisk).

I den krydsede klassifikation har vi modellerne

$$\text{I: } X_{ijv} = \mu + a_i + b_j + ab_{ij} + Z_{v(ij)}$$

$$\text{II: } X_{ijv} = \mu + A_i + B_j + AB_{ij} + Z_{v(ij)}$$

$$\text{III: } X_{ijv} = \mu + a_i + b_j + AB_{ij} + Z_{v(ij)}$$

$$\text{IV: } X_{ijv} = \mu + a_i + B_j + aB_{ij} + Z_{v(ij)}$$

Her er I en systematisk (eller rent parametrisk) model, II en tilfældig (eller en varianskomponent) model, og III og IV blandede modeller. Der findes naturligvis også en til IV analog model, hvor  $a_i$  er erstattet med  $A_i$  og  $B_j$  med  $b_j$ . Som i det hierarkiske tilfælde vil vi ikke betragte modeller, hvor vekselvirkningen er systematisk, og hovedvirkningen er tilfældig. I tilfældet med siloerne ville dette f.eks. svare til, at vi regner med, at vandindholdet i den  $i$ 'te silo er "tilfældigt", f.eks. svarende til en tilfældig stikprøveudtagning blandt alle siloer i et givet område. I denne situation ville det virke barokt at påstå, at den afvigelse vi opnår fra den gennemsnitlige værdi ved at måle et tilfældigt valgt sted i en tilfældigt valgt silo, naturligt beskrives ved en deterministisk parameterværdi.

Vi har følgende bånd og forudsætninger om fordelingerne

$$\sum_i a_i = 0$$

$$\sum_j b_j = 0$$

$$\sum_j a b_{ij} = 0 \quad \forall_i$$

$$\sum_i a b_{ij} = 0 \quad \forall_j$$

$$A_i \in N(0, \sigma_A^2)$$

$$B_j \in N(0, \sigma_B^2)$$

$$a b_{ij}, AB_{ij} \in N(0, \sigma_{AB}^2)$$

$$Z_{p(ij)} \in N(0, \sigma^2)$$

} stokastisk uafhængige

Vi vil nu gå over til den statistiske analyse af de to typer af modeller. Begrundelserne for metoderne er analoge til dem, vi anvendte under den ensidede variansanalyse, så de forbigås, og vi nøjles med at anføre resultaterne.

### 5.3.2 Analyse af hierarkisk klassificerede data

Grundlaget for de statistiske analyser kan findes i følgende variansanalysetabel med forventningsværdier over middelvadrat-afvigelsessummer.

Vi har anvendt betegnelserne

$$\bar{X}_{ij\cdot} = \frac{1}{n} \sum_v X_{ijv}$$

$$\bar{X}_{i..} = \frac{1}{mn} \sum_j \sum_v X_{ijv}$$

$$\bar{X} = \frac{1}{kmn} \sum_i \sum_j \sum_v X_{ijv} .$$

Variation	SAK	Fr.gr.
Mellem hovedgrupper	$SAK_A = nm \sum_i (\bar{X}_{i..} - \bar{X})^2$	$f_A = k-1$
Mellem undergrupper inden for hovedgrupper	$SAK_{B(A)} = n \sum_{ij} \sum (\bar{X}_{ij.} - \bar{X}_{i..})^2$	$f_{B(A)} = k(m-1)$
Residual	$SAK_{Res} = \sum_{ijv} \sum \sum (X_{ijv} - \bar{X}_{ij.})^2$	$f_{Res} = km(n-1)$
Total	$SAK_0 = \sum_{ijv} \sum \sum (X_{ijv} - \bar{X})^2$	$kmn-1$

$$SAK_0 = \sum_{i,j,v} \sum (X_{ijv} - \bar{X})^2$$

	Systematisk model $X_{ijv} = \mu + a_i$ $+b(a)_{j(i)} + Z_{v(ij)}$	Tilfældig model $X_{ijv} = \mu + A_i$ $+B(A)_{j(i)} + Z_{v(ij)}$	Blandet model $X_{ijv} = \mu + a_i$ $+B(a)_{j(i)} + Z_{v(ij)}$
$E\left(\frac{SAK_A}{f_A}\right)$	$\sigma^2 + mn \frac{1}{k-1} \sum_i a_i^2$	$\sigma^2 + n\sigma_{B(A)}^2 + mn\sigma_A^2$	$\sigma^2 + n\sigma_{B(A)}^2$ $+ mn \frac{1}{k-1} \sum_i a_i^2$
$E\left(\frac{SAK_{B(A)}}{f_{B(A)}}\right)$	$\sigma^2 + \frac{n}{k(m-1)} \sum_{i,j} b(a)_{j(i)}^2$	$\sigma^2 + n\sigma_{B(A)}^2$	$\sigma^2 + n\sigma_{B(A)}^2$
$E\left(\frac{SAK_{Res}}{f_{Res}}\right)$	$\sigma^2$	$\sigma^2$	$\sigma^2$

Hvis vi e.g. i den tilfældige model ønsker at teste hypotesen

$$H_0: \sigma_A^2 = 0 \quad \text{mod} \quad H_1: \sigma_A^2 > 0,$$

bliver det kritiske område

$$\left\{ (x_{111}, \dots, x_{kmn}) \mid \frac{sak_A / (k-1)}{sak_{B(A)} / k(m-1)} > F(k-1, k(m-1))_{1-\alpha} \right\} .$$

At

$$\frac{SAK_A / (k-1)}{SAK_{B(A)} / k(m-1)} \in F(k-1, k(m-1)) ,$$

hvis  $\sigma_A^2 = 0$ , indses ved analoge betragtninger til de p. 5.13 gjorde. At de kritiske værdier er store værdier af brøken, indses let ved at betragte de forventede værdier.

Hvis vi i den systematiske model ønsker at teste den tilsvarende hypotese

$$H_0: a_1 = \dots = a_k = 0 \text{ mod } H_1: \exists i, j: a_i \neq a_j ,$$

bliver det kritiske område

$$\left\{ (x_{111}, \dots, x_{kmn}) \mid \frac{sak_A / (k-1)}{sak_{Res} / km(n-1)} > F(k-1, km(n-1))_{1-\alpha} \right\} .$$

Her bemærker vi en forskel mellem analysen af den systematiske model og den tilfældige.

Ved hjælp af tabellen over de forventede værdier konstruerer man nu let de kritiske områder for de øvrige hypoteser, man måtte ønske at teste.

Inden vi giver et eksempel, skal vi anføre nogle estimatorer:

$$\begin{aligned} \hat{\mu} &= \bar{X} \\ \hat{a}_i &= \bar{X}_{i..} - \bar{X} \\ b(a)_{j(i)} &= \bar{X}_{ij.} - \bar{X}_{i..} . \end{aligned}$$

Estimatorerne for de forskellige varianser fås som summer og differenser af de forskellige kvadratafvigelsessummer. I den tilfældige model haves e.g., at

$$\hat{\sigma}_A^2 = \frac{1}{mn} (SAK_A / f_A - SAK_{B(A)} / f_{B(A)})$$

er et centralt skøn over  $\sigma_A^2$ .

Vi anfører nu et eksempel, hvor data er taget fra J.R. Johnson (1957), her citeret fra Johnson & Leone (1967).

**Eksempel 5.4** Ved en undersøgelse af 60 mm morterer har man testet 5 partier ammunition. Hvert parti blev testet 2 forskellige dage på grund af eksperimentets størrelse. I nedenstående tabel er der anført observeret rækkevidde (i yards) for granater af typen HE, M49A2 affyret med standardladning ved en elevation på 45%.

Det er klart, at vi her har en hierarkisk model. Hovedgrupperne er ammunitionspartierne, og undergrupperne er dagnumrene (grunden til, at vi ikke har en krydsklassifikation er, at dagene er forskellige. Hvis vi havde prøvet samtlige partier e.g. 24/10-52 og igen samtlige partier den 31/10-52, havde vi haft en krydsklassifikation).

Vi vælger modellen

$$X_{ijv} = \mu + a_i + B(a)_{j(i)} + Z_{v(ij)} ,$$

hvor

$$i = 1, \dots, 5 \quad (\text{parti})$$

$$j = 1, 2 \quad (\text{dag inden for parti})$$

$$v = 1, \dots, 19 \quad (\text{gentagelse inden for parti og dag})$$

og

Parti	MA-1-53		MA-1-82		MA-1-363		MA-1-604		MA-1-613	
Dato	24/10	31/10	17/12	12/1	19-20/8	8-9/9	30/12	10/1	20/1	31/1
	52	52	52	53	53	53	54	55	55	55
	1890	2057	1925	2112	1988	1932	1967	1980	2000	2110
	1863	2028	1902	2083	1876	1862	2021	2025	1769	1983
	1927	1964	2043	2096	1874	1863	2014	1983	1885	2098
	1830	1955	1957	2078	1914	1927	2019	1862	2004	2084
	1803	1976	1946	2031	1882	1907	2002	2041	1904	2015
	1951	1996	1940	2084	1774	1763	2128	2001	1865	1978
Række-	1995	2057	1916	2017	1872	1841	1949	1970	1927	2098
vidde	1967	1979	1967	2035	1822	1914	1904	2053	1972	2124
i	1934	2010	1958	1978	1891	1837	2029	1978	1886	2077
yards	1897	2037	1879	2045	1855	1911	1989	1940	2019	2036
	1869	2013	1995	2002	1809	1866	2052	1980	1884	2060
	1847	1990	1980	2078	1894	1797	2042	1921	1990	2141
	1882	2015	1934	2118	1870	1983	1835	2002	1884	2075
	1965	1975	1990	2017	1910	1873	1970	1969	1938	2074
	1954	1934	1906	2107	1970	1907	2038	1849	1950	2003
	1973	2074	1985	2005	1980	1923	2000	2030	1999	2077
	1870	2071	2000	2094	1885	1962	2041	2006	1987	2110
	1894	1993	1951	1993	1775	1859	2033	1999	1823	2061
	1927	1943	1979	2020	1871	1993	1996	2006	1951	2099

$$a_1 + \dots + a_5 = 0$$

$$\left. \begin{array}{l} B(a)_{j(i)} \in N(0, \sigma_{B(A)}^2) \\ Z_{v(ij)} \in N(0, \sigma^2) \end{array} \right\} \text{stokastisk uafhængige .}$$

Grunden til, at vi har valgt en blandet model er, at den  $j$ 'te dags afvigelse fra det pågældende partigennemsnit er sammensat af en række tilfældige bidrag, hvoraf det væsentligste formentlig er de meteorologiske forhold. Endvidere er vi næppe særskilt interesserede i at analysere forholdet de angivne dage,



men nok snarere interesserede i at arbejde med en model, der er egnet til at forudsige, hvad man kan forvente ved en gentagelse af eksperimentet på helt andre dage. Det ville derfor ikke være rimeligt at anvende en rent systematisk model.

Det primære må nu være at undersøge, om alle  $a_i$ 'erne kan antages at være 0, i.e. om det kan antages, at ammunitionspartierne er ens. Sekundært vil man være interesseret i at undersøge, om  $\sigma_{B(A)}^2$  kan antages at være 0. Hvis dette er tilfældet, indebærer det bl.a., at man ved fremtidige undersøgelser af samme art f.eks. ikke behøver at tage hensyn til, at forsøgene strækker sig over flere dage.

Vi samler beregningerne i følgende variansanalysekema. (Vi skal senere vende tilbage til beregningsspørgsmålet).

Variation	SAK	f	S <sup>2</sup>	E(S <sup>2</sup> )
Mellem partier	369254	4	92314	$\sigma^2 + 19\sigma_{B(A)}^2 + 38 \cdot \frac{1}{4} \sum a_i^2$
Mellem dage inden for partier	369941	5	73988	$\sigma^2 + 19\sigma_{B(A)}^2$
Inden for dage	517065	180	2873	$\sigma^2$
Total	1256260	189		

Da

$$\frac{92314}{73988} = 1.25 = F(4, 5)_{0.59} ,$$

og

$$\frac{73988}{2873} = 25.75 > F(5, 180)_{0.9995} ,$$

ser vi, at hypotesen alle  $\alpha_i = 0$  accepteres på alle niveauer mindre end 41%, hvorimod hypotesen  $\sigma_t^2 = 0$  forkastes meget kraftigt. (At det er de anførte brøker, der er de relevante, ses let ved hjælp af søjlen over de forventede værdier af  $S^2$ -størrelserne).

Vor konklusion må derfor blive, at der ikke skønnes at være nogen (signifikant) forskel mellem ammunitionspartierne. Derimod er der en tydelig dag-effekt. Den estimerede varians for denne er

$$\hat{\sigma}_{B(A)}^2 = \frac{1}{19} (73988 - 2873) = 3742.89 = 61.18^2 .$$

Endelig har vi følgende estimat over forsøgsfejls varians:

$$\hat{\sigma}^2 = 2873 = 53.60^2 .$$

Et skøn over variansen på en enkelt måling er derfor lig

$$\hat{\sigma}^2 + \hat{\sigma}_{B(A)}^2 = 6616 = 81.3^2 .$$

Bemærkning Hvis vi - noget søgt - havde anvendt en rent systematisk model, havde vi ved test af hypotesen om alle  $\alpha_{j(1)} = 0$  fået teststørrelsen

$$\frac{73988}{2873} = 25.75 > F(5, 180)_{0.9995} ,$$

og ved test af hypotesen alle  $a_i = 0$  teststørrelsen

$$\frac{92314}{2873} = 32.13 > F(4, 180)_{0.9995} .$$

I denne situation må vor konklusion derfor blive, at såvel  $a_{j(i)}$ 'er som  $a_i$ 'er (NB!) må antages at være forskellige fra 0, altså tilsyneladende et resultat, der er i modstrid med det tidligere opnåede. Deri ligger dog ingen egentlig modstrid; thi modellerne er forskellige, og vi tolker afvigelser (variationer) forskelligt i de to situationer.

□

### 5.3.3 Analyse af krydsklassificerede data

Variansanalysekemaet er her det fra den almindelige tosidede variansanalyse (d.v.s. den systematiske model) velkendte. Foruden dette anfører vi også en oversigt over de forventede værdier af  $SAK_i/f_i$ . (Vi benævner som sædvanlig index  $i$  som rækkeindex og index  $j$  som søjleindex):

Variation	SAK	Frihedsgr.
Mellem rækker (faktor A)	$SAK_A = mn \sum_i (\bar{X}_{i..} - \bar{X})^2$	k-1
Mellem søjler (faktor B)	$SAK_B = kn \sum_j (\bar{X}_{.j.} - \bar{X})^2$	m-1
Vekselvirkning (AB)	$SAK_{AB} = n \sum_{ij} (\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X})^2$	(k-1)(m-1)
Residual	$SAK_{Res} = \sum_{ijv} (X_{ijv} - \bar{X}_{ij.})^2$	km(n-1)
Total	$\sum_i \sum_j \sum_v (X_{ijv} - \bar{X})^2$	kmn-1

	Systematisk model $X_{ijv} = \mu + a_i + b_j + ab_{ij} + Z_{v(ij)}$	Tilfældig model $X_{ijv} = \mu + A_i + B_j + AB_{ij} + Z_{v(ij)}$
$E\left(\frac{SAK_A}{f_A}\right)$	$\sigma^2 + mn \frac{1}{k-1} \sum_i a_i^2$	$\sigma^2 + n\sigma_{AB}^2 + mn\sigma_A^2$
$E\left(\frac{SAK_B}{f_B}\right)$	$\sigma^2 + kn \frac{1}{m-1} \sum_j b_j^2$	$\sigma^2 + n\sigma_{AB}^2 + kn\sigma_B^2$
$E\left(\frac{SAK_{AB}}{f_{AB}}\right)$	$\sigma^2 + n \frac{1}{(k-1)(m-1)} \sum_{i,j} ab_{ij}^2$	$\sigma^2 + n\sigma_{AB}^2$
$E\left(\frac{SAK_{Res}}{f_{Res}}\right)$	$\sigma^2$	$\sigma^2$

	Blandet model $X_{ijv} = \mu + a_i + b_j + AB_{ij} + Z_{v(ij)}$	Blandet model $X_{ijv} = \mu + a_i + B_j + aB_{ij} + Z_{v(ij)}$
$E\left(\frac{SAK_A}{f_A}\right)$	$\sigma^2 + n\sigma_{AB}^2 + mn \frac{1}{k-1} \sum_i a_i^2$	$\sigma^2 + n\sigma_{AB}^2 + mn \frac{1}{k-1} \sum_i a_i^2$
$E\left(\frac{SAK_B}{f_B}\right)$	$\sigma^2 + n\sigma_{AB}^2 + kn \frac{1}{m-1} \sum_j b_j^2$	$\sigma^2 + n\sigma_{AB}^2 + kn\sigma_B^2$
$E\left(\frac{SAK_{AB}}{f_{AB}}\right)$	$\sigma^2 + n\sigma_{AB}^2$	$\sigma^2 + n\sigma_{AB}^2$
$E\left(\frac{SAK_{Res}}{f_{Res}}\right)$	$\sigma^2$	$\sigma^2$

Estimatorerne for de systematiske parametre er

$$\hat{\mu} = \bar{x}$$

$$\hat{a}_i = \bar{x}_{i..} - \bar{x}$$

$$\hat{b}_j = \bar{x}_{.j.} - \bar{x}$$

$$\hat{ab}_{ij} = \bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x}$$

Varianserne estimeres som vanligt ved passende summer og differenser mellem  $SAK_1/f_1$ 'er. Teststørrelserne for de forskellige hypoteser er kvotienter mellem  $SAK_1/f_1$ 'er. Af de forventede værdier ses umiddelbart, hvilke der kan komme på tale.

Vi illustrerer anvendelserne af disse skemaer ved hjælp af følgende

Eksempel 5.5 (Data stammer fra reference [6]).

På en klædefabrik har man undersøgt 4 spolemaskiners produktion. Nærmere betegnet har man til 3 forskellige tidspunkter taget 4 tråde af fast længde og vejnet disse. Resultaterne fremgår af følgende tabel

		Maskine (faktor B)			
		1	2	3	4
Tidspunkt (faktor A)	α	7.50	7.81	7.50	8.05
		7.70	7.94	7.77	7.53
		7.93	7.23	7.96	8.16
		7.52	7.94	7.83	7.76
	β	8.17	8.29	8.46	8.38
		8.09	8.54	8.33	8.47
		8.11	8.45	8.27	8.38
		7.96	8.43	8.24	8.60
	γ	8.01	8.01	8.37	8.16
		8.17	7.92	8.27	7.96
		8.05	8.27	8.07	8.08
		7.91	7.92	8.28	8.52

Man er åbenbart interesseret i at få beskrevet variationen mellem maskinerne. Afhængigt af, hvorledes tidspunkterne er valgte, er der 2 mulige modeller.

Hvis tidspunkterne er systematisk valgt (α f.eks. ½ time efter arbejdstids start, β = ½ time før frokost og γ = ½ time før fyraften), vil vi anvende modellen

$$I: \quad X_{ij} = \mu + a_i + b_j + ab_{ij} + Z_{v(ij)} \quad \begin{array}{l} i=1,2,3 \\ j=1,2,3,4 \\ v=1,2,3,4, \end{array}$$

hvor  $\mu$  angiver et niveau,  $a_i$  angiver tidseffekten,  $b_j$  maskineffekten og  $ab_{ij}$  vekselvirkningen.

Hvis tidspunkterne er tilfældigt valgte, bør vi anvende modellen

$$II: \quad X_{ijv} = \mu + A_i + b_j + Ab_{ij} + Z_{v(ij)} ,$$

idet vi nu tolker tidseffekten og vekselvirkningen som stokastiske variable.

Variansanalyseeskemaet bliver

Variation	SAK	Fr.gr.	S <sup>2</sup>	E(S <sup>2</sup> )	
				I	II
Mellem tider (A)	2.6264	2	1.3132	$\sigma^2 + 8\sum_i a_i^2$	$\sigma^2 + 4\sigma_{AB}^2 + 16\sigma_A^2$
Mellem mask. (B)	0.3907	3	0.13023	$\sigma^2 + 4\sum_j b_j^2$	$\sigma^2 + 4\sigma_{AB}^2 + 4\sum_j b_j^2$
Vekselvirkn. (AB)	0.1918	6	0.03197	$\sigma^2 + \frac{2}{3} \sum_{i,j} ab_{ij}^2$	$\sigma^2 + 4\sigma_{AB}^2$
Residual (Res)	1.2768	36	0.03547	$\sigma^2$	$\sigma^2$
Total	4.4857	47			

Teststørrelsen for, om  $ab_{ij} = 0$  i den parametriske model, respektive, om  $\sigma_{AB}^2 = 0$  i den blandede model, er

$$\frac{0.03197}{0.03547} = 0.90 \approx F(6, 36)_{0.50} .$$

Vi vil derfor acceptere, at  $ab_{ij} = 0$  henholdsvis  $\sigma_{AB}^2 = 0$  ved test på alle niveauer mindre end 50%. Vi antager derfor dette, og får som et forbedret skøn over  $\sigma^2$

$$\hat{\sigma}^2 = \frac{0.1918 + 1.2768}{6 + 36} = 0.0350 .$$

Vi ser nu, at teststørrelserne for test om  $a_i = 0$  og  $b_j = 0$  henholdsvis  $\sigma_a^2 = 0$  og  $\beta_j = 0$  bliver ens i de to modeller, nemlig

$$\frac{S_{\text{tid}}^2}{0.0350} = 37.52 = F(2,42)_{0.9995}$$

$$\frac{S_{\text{maskine}}^2}{0.0350} = 3.72 = F(3,42)_{0.98} .$$

Vi vil åbenbart forkaste hypotesen om, at  $a_i$ 'erne = 0 respektive  $\sigma_A^2 = 0$ . Hypotesen om, at  $b_j$ 'erne er lig 0, er mere tvivlsom.

Hvis tidspunkterne er tilfældigt valgte (d.v.s. vi regner med model II), bliver konklusionen, at det er tvivlsomt, om der er forskel på maskinerne. Der spores derimod en betydelig variation mellem målinger taget til forskellige tidspunkter. Variansen på disse variationer estimeres ved

$$\hat{\sigma}_A^2 = \frac{1}{16}(1.3132 - 0.0350) = 0.0799 = 0.2826^2 .$$

Niveauet estimeres til

$$\hat{\mu} = 8.068 ,$$

og "forsøgsfejls" varians til

$$\hat{\sigma}^2 = 0.0350 = 0.187^2 ,$$

der er estimeret med 42 frihedsgrader.

Er tidspunkterne systematisk valgte, må vi fremdeles konkludere, at det er tvivlsomt, om der er forskel på maskinerne. Derimod er der en tydelig tidseffekt, der udtrykkes ved estimaterne

$$\hat{a}_1 = -0.310$$

$$\hat{a}_2 = 0.255$$

$$\hat{a}_3 = 0.055 .$$



Niveauet estimeres til

$$\hat{\mu} = 8.068 ,$$

og forsøgsfejls varians (med 42 frihedsgrader) til

$$\hat{\sigma}^2 = 0.0350 = 0.187^2 .$$

□

Ved sammenligning af variansanalyseeskemaerne for den hierarkiske model og krydsklassifikationsmodellen ses, at SAK for variationen mellem hovedgrupper i den hierarkiske model er den samme som SAK for variationen mellem rækker i krydsmodellen. Endvidere er residual-SAK'erne og de totale SAK'er ens. Heraf fås umiddelbart, at

$$SAK_{B(A)} = SAK_B + SAK_{AB}$$

eller udtrykt mere direkte

$$n \sum_i \sum_j (\bar{X}_{ij} - \bar{X}_{i.})^2 = kn \sum_j (\bar{X}_{.j} - \bar{X})^2 + n \sum_i \sum_j (\bar{X}_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X})^2$$

(Denne relation vises også let direkte).

Dette er væsentligt, fordi der findes let tilgængelige standardprogrammer til beregning af kvadratsummerne i en almindelig (krydsklassificeret) tosidet variansanalyse. Man kan derfor køre sin hierarkiske model som en krydsklassificeret og så få den korrekte variansanalysetabel ved blot at addere to kvadratsummer. Vi bemærker, at denne additionsregel også er gyldig for frihedsgraderne.

Regnes eksempel 5.4 som en krydset tosidet variansanalyse, fås følgende tabel

Variation	SAK	
Mellem rækker	369254	4
Mellem søjler	197000	} 5 1 4
Vekselvirkning	172941	
Residual	517065	
Total	1256260	189

Det ses, at vi ved addition af SAK (mellem søjler) og SAK (vekselvirkning) netop får det oprindelige variansanalysekema.

Et andet meget afgørende forhold er spørgsmålet med at bestemme den "fejlvariens", som effekterne skal måles relativt til.

I den krydsede model med deterministiske parametre skal vi bruge residualkvadratafvigelsessummen. Hvis vi kun har en observation pr. celle er denne lig 0. Under forudsætning af, at der ikke er nogen vekselvirkning (i.e. at alle  $a_{ij} = 0$ ) kan vi bruge vekselvirkningskvadratafvigelsessummen.

Har vi derimod en model med tilfældig vekselvirkning, kan vi altid bruge vekselvirkningskvadratafvigelsessummen. En sådan model kunne f.eks. svare til et forsøg som følgende.

Eksempel 5.6 Ved en analyse af forskellige diæters indflydelse på personers vægttab for forskellige aldersgrupper har man opnået følgende måleresultater:

		B: Alder	
		Person på 20 år ...	Person på 50 år
A: Diæt	1	$X_{111}, \dots, X_{115}$	$\dots X_{141}, \dots, X_{145}$
	$\vdots$	$\vdots$	$\vdots$
	$\vdots$	$\vdots$	$\vdots$
	$\vdots$	$\vdots$	$\vdots$
	5	$X_{511}, \dots, X_{515}$	$\dots X_{541}, \dots, X_{545}$

Ved bestemmelsen af vægttabet har man taget 5 uafhængige målinger af vægten.

Her har vi en model

$$X_{ijv} = \mu + a_i + b_j + AB_{ij} + Z_{v(ij)} ,$$

hvor

$$\sum a_i = \sum b_j = 0$$

$$\left. \begin{array}{l} AB_{ij} \in N(0, \sigma_{AB}^2) \\ Z_{v(ij)} \in N(0, \sigma^2) \end{array} \right\} \text{ uafhængige .}$$

Vi regner begge hovedeffekterne deterministiske, den ene svarer til en bestemt diæt, den anden til en bestemt aldersgruppe. Derimod fortolker vi afvigelserne fra denne model som tilfældige. De kan e.g. skyldes det konkrete valg af forsøgspersoner etc.

Ved testningen af, om  $a_i$ 'er eller  $b_j$ 'er kan antages at være lig 0, er det klart, at vi ikke vil måle disse i forhold til  $\sigma^2$ , som er et udtryk for veje-usikkerheden; men derimod i forhold til den "biologiske varians"  $\sigma_{AB}^2$ , et resultat man også opnår ved anvendelse af den foran anførte teori.

□

#### 5.4 Variansanalysemodeller med 3 faktorer

I modeller, hvori der indgår mere end 2 faktorer, får vi mulighed for mere indviklede strukturer, hvor nogle faktorer er krydsede og andre er hierarkisk ordnede. Inden vi giver generelle formler, vil vi betragte nogle af de vigtigste muligheder med 3 faktorer A, B og C.

5.40

Vi kalder den almindelige observation

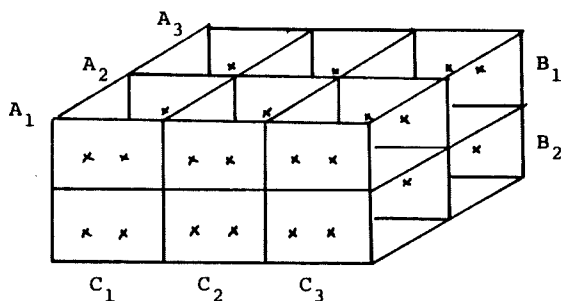
$X_{ijrv}$ ,  $i = 1, \dots, k$  angiver A-niveau  
 $j = 1, \dots, m$  angiver B-niveau  
 $r = 1, \dots, p$  angiver C-niveau  
 $v = 1, \dots, n$  angiver gentagelsesnr.

Det grundlæggende variansanalysekema for samtlige variansanalyser med de 3 faktorer er anført p. 5.42.

Vi betragter nu de enkelte strukturer.

Struktur 1 Alle faktorer krydsklassificerede, d.v.s. symbolsk  $A \times B$ ,  $A \times C$ ,  $B \times C$ .

Denne struktur kan grafisk anskueliggøres som i følgende skitse.



Med denne datastruktur har vi for det første den rent systematiske model

$$i) \quad X_{ijrv} = \mu + a_i + b_j + c_r + ab_{ij} + ac_{ir} + bc_{jr} + abc_{ijr} + Z_{v(ijr)}$$

og den rent tilfældige model

$$ii) \quad X_{ijrv} = \mu + A_i + B_j + C_r + AB_{ij} + AC_{ir} + BC_{jr} + ABC_{ijr} + Z_{v(ijr)} .$$

Endvidere forekommer der de blandede modeller, i.e. modeller, hvor visse af komponenterne er konstanter og visse er stokastiske variable.

Parametrene og de stokastiske variable skal tilfredsstille

$$\sum_i a_i = 0, \quad \sum_j b_j = 0, \quad \sum_r c_r = 0$$

$$\sum_i ab_{ij} = 0, \quad \sum_j ab_{ij} = 0; \quad \sum_i ac_{ir} = 0, \quad \sum_r ac_{ir} = 0;$$

$$\sum_j bc_{jr} = 0, \quad \sum_r bc_{jr} = 0 .$$

$$\sum_i abc_{ijr} = 0, \quad \sum_j abc_{ijr} = 0, \quad \sum_r abc_{ijr} = 0 .$$

$$A_i \in N(0, \sigma_A^2), \quad B_j \in N(0, \sigma_B^2), \quad C_r \in N(0, \sigma_C^2)$$

$$AB_{ij} \in N(0, \sigma_{AB}^2), \quad AC_{ir} \in N(0, \sigma_{AC}^2), \quad BC_{jr} \in N(0, \sigma_{BC}^2)$$

$$ABC_{ijr} \in N(0, \sigma_{ABC}^2)$$

$$Z_{v(ijr)} \in N(0, \sigma^2) .$$

De stokastiske variable forudsættes endvidere at være uafhængige.

Variations- årsag	Kvadratafvigelsessum	Frihedsgrader
A	$SAK_A = mpn \sum_{i=1}^k (\bar{X}_{i..} - \bar{X})^2$	k-1
B	$SAK_B = kpn \sum_{j=1}^m (\bar{X}_{.j.} - \bar{X})^2$	m-1
C	$SAK_C = kmn \sum_{r=1}^p (\bar{X}_{...r} - \bar{X})^2$	p-1
AB	$SAK_{AB} = pn \sum_{i=1}^k \sum_{j=1}^m (\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X})^2$	(k-1)(m-1)
AC	$SAK_{AC} = mn \sum_{i=1}^k \sum_{r=1}^p (\bar{X}_{i.r} - \bar{X}_{i..} - \bar{X}_{...r} + \bar{X})^2$	(k-1)(p-1)
BC	$SAK_{BC} = kn \sum_{j=1}^m \sum_{r=1}^p (\bar{X}_{.j.r} - \bar{X}_{.j.} - \bar{X}_{...r} + \bar{X})^2$	(m-1)(p-1)
ABC	$SAK_{ABC} = n \sum_{i=1}^k \sum_{j=1}^m \sum_{r=1}^p (\bar{X}_{ijr} - \bar{X}_{ij.} - \bar{X}_{i.r} - \bar{X}_{.j.r} + \bar{X}_{i..} + \bar{X}_{.j.} + \bar{X}_{...r} - \bar{X})^2$	(k-1)(m-1)(p-1)
Residual	$SAK_{Res} = \sum_{i=1}^k \sum_{j=1}^m \sum_{r=1}^p \sum_{v=1}^n (X_{ijrv} - \bar{X}_{ijr})^2$	kmp(n-1)
Total	$SAK_{Tot} = \sum_{i=1}^k \sum_{j=1}^m \sum_{r=1}^p \sum_{v=1}^n (X_{ijrv} - \bar{X})^2$	kmpn-1

Variansanalysekema for den fuldstændigt krydsede model med 3 faktorer.

Vi har nu det nedenfor anførte variansanalysekema.

Variation	SAK	f	E(SAK/f)
A	$SAK_A$	k-1	$\sigma^2 + (n\sigma_{ABC}^2 + mn\sigma_{AC}^2 + np\sigma_{AB}^2) + nmp\sigma_A^2$
B	$SAK_B$	m-1	$\sigma^2 + (n\sigma_{ABC}^2 + nk\sigma_{BC}^2 + np\sigma_{AB}^2) + nkpc\sigma_B^2$
C	$SAK_C$	p-1	$\sigma^2 + (n\sigma_{ABC}^2 + nk\sigma_{BC}^2 + nm\sigma_{AC}^2) + nkmc\sigma_C^2$
AB	$SAK_{AB}$	(k-1)(m-1)	$\sigma^2 + (n\sigma_{ABC}^2) + np\sigma_{AB}^2$
AC	$SAK_{AC}$	(k-1)(p-1)	$\sigma^2 + (n\sigma_{ABC}^2) + nm\sigma_{AC}^2$
BC	$SAK_{BC}$	(m-1)(p-1)	$\sigma^2 + (n\sigma_{ABC}^2) + nk\sigma_{BC}^2$
ABC	$SAK_{ABC}$	(k-1)(m-1)(p-1)	$\sigma^2 + n\sigma_{ABC}^2$
Res	$SAK_{Res}$	(kmp(n-1))	$\sigma^2$
Total	$SAK_{Tot}$	kmpn-1	

Her skal forventningsværdierne af  $SAK_i/f_i$  tolkes på en særlig måde. Hvis de pågældende effekter er systematiske, skal vi læse

$$\sigma_{ABC}^2 = \frac{1}{(k-1)(m-1)(p-1)} \sum_i \sum_j \sum_r abc_{ijr}^2$$

$$\sigma_{BC}^2 = \frac{1}{(m-1)(p-1)} \sum_j \sum_r bc_{jr}^2$$

$$\sigma_{AC}^2 = \frac{1}{(k-1)(p-1)} \sum_i \sum_r ac_{ir}^2$$

$$\sigma_{AB}^2 = \frac{1}{(k-1)(m-1)} \sum_i \sum_j ab_{ij}^2$$

$$\sigma_C^2 = \frac{1}{p-1} \sum_r c_r^2$$

$$\sigma_B^2 = \frac{1}{m-1} \sum_j b_j^2$$

$$\sigma_A^2 = \frac{1}{k-1} \sum_i a_i^2 .$$

De led i søjlen med forventningsværdierne, som er anbragt i parenteser, skal kun medtages, hvis de er udtryk for en tilfældig (stokastisk) effekt.

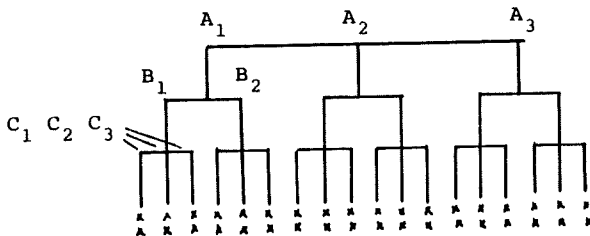
Hvis vi e.g. har en blandet model

$$X_{ijrv} = \mu + a_i + b_j + c_r + ab_{ij} + ac_{ir} + BC_{jr} + ABC_{ijr} + Z_{v(ijr)} ,$$

bliver eksempelvis

$$E(SAK_{C/f_C}) = \sigma^2 + n\sigma_{ABC}^2 + nk\sigma_{BC}^2 + nkm \frac{1}{p-1} \sum_r c_r^2 .$$

Struktur 2 Den rent hierarkiske struktur, i.e.  $A \supset B$ ,  $B \supset C$ ,  $A \supset C$ . Strukturen er anskueliggjort i følgende skitse.





Den rent systematiske model bliver

$$X_{ijrv} = \mu + a_i + b(a)_{j(i)} + c(ab)_{r(ij)} + Z_{v(ij)r}$$

Analogt skrives den tilfældige model

$$X_{ijrv} = \mu + A_i + B(A)_{j(i)} + C(AB)_{r(ij)} + Z_{v(ij)r}$$

og helt tilsvarende for de blandede modeller.

Variansanalyseeskemaet bliver

Variation	SAK	f	E(SAK/f)
A	$SAK_A$	$k-1$	$\sigma^2 + (n\sigma_{C(AB)}^2 + np\sigma_{B(A)}^2) + npm\sigma_A^2$
B(A)	$SAK_B + SAK_{AB}$	$k(m-1)$	$\sigma^2 + (n\sigma_{C(AB)}^2) + np\sigma_{B(A)}^2$
C(AB)	$SAK_C + SAK_{AC} + SAK_{BC}$ $+ SAK_{ABC}$	$km(p-1)$	$\sigma^2 + n\sigma_{C(AB)}^2$
Res	$SAK_{Res}$	$kmp(n-1)$	$\sigma^2$
Total	$SAK_{Tot}$	$kmpn-1$	

Hvis en effekt er systematisk, læses varianskomponenten efter nedenstående retningslinier

$$\sigma_{C(AB)}^2 = \frac{1}{km(p-1)} \sum_i \sum_j \sum_r c(ab)_{r(ij)}^2$$

$$\sigma_{B(A)}^2 = \frac{1}{k(m-1)} \sum_i \sum_j b(a)_{j(i)}^2$$

$$\sigma_A^2 = \frac{1}{k-1} \sum_i a_i^2$$

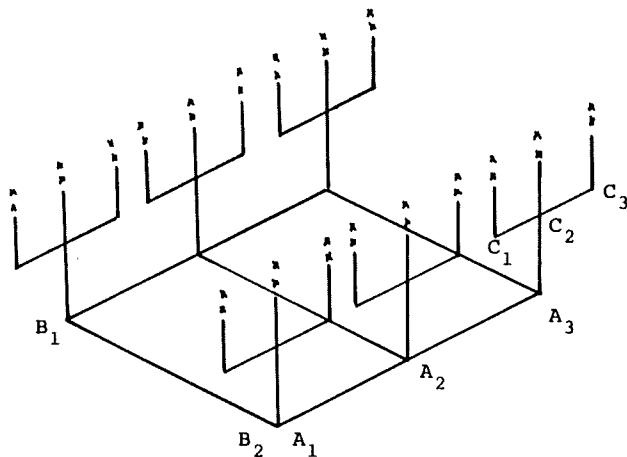
Angående parenteserne henvises til kommentaren p. 5.44.

NB! I de tilfældige modeller gælder de sædvanlige (i.e. de p. 5.40 nævnte) forudsætninger om de stokastiske komponenter. Der er imidlertid ændrede bånd på de systematiske, nemlig blot

$$\sum_r c(ab)_{r(ij)} = 0; \quad \sum_j b(a)_{j(i)} = 0; \quad \sum_i a_i = 0.$$

d.v.s. det er kun summer over index, der ikke står i en parentes, der er 0.

Struktur 3 De to faktorer krydsede og den tredje underordnet disse, d.v.s.  $A \times B$ ,  $A \supset C$ ,  $B \supset C$ . Strukturen er anskueliggjort grafisk nedenfor.



Den systematiske model er

$$X_{ijrv} = \mu + a_i + b_j + ab_{ij} + c(ab)_{r(ij)} + Z_{v(ijr)} ,$$

og den tilfældige

$$X_{ijrv} = \mu + A_i + B_j + AB_{ij} + C(AB)_{r(ij)} + Z_{v(ijr)} .$$

De blandede modeller dannes på vanlig måde. De ændrede bånd på parametrene er

$$\sum_r c(ab)_{r(ij)} = 0$$

$$\sum_i ab_{ij} = 0, \sum_j ab_{ij} = 0$$

$$\sum_i a_i = 0, \sum_j b_j = 0 .$$

Variansanalyseeskemaet er anført p. 5.48. Den deterministiske tolkning af varianskomponenterne er

$$\sigma_{C(AB)}^2 = \frac{1}{km(p-1)} \sum_i \sum_j \sum_r c(ab)_{r(ij)}^2$$

$$\sigma_{AB}^2 = \frac{1}{(k-1)(m-1)} \sum_i \sum_j ab_{ij}^2$$

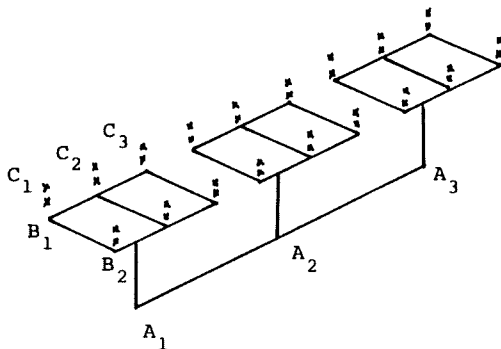
$$\sigma_B^2 = \frac{1}{m-1} \sum_j b_j^2$$

$$\sigma_A^2 = \frac{1}{k-1} \sum_i a_i^2 .$$

Variation	SAK	f	E(SAK/f)
A	$SAK_A$	$k-1$	$\sigma^2 + (n\sigma_{C(AB)}^2 + np\sigma_{AB}^2) + npm\sigma_A^2$
B	$SAK_B$	$m-1$	$\sigma^2 + (n\sigma_{C(AB)}^2 + np\sigma_{AB}^2) + npk\sigma_B^2$
AB	$SAK_{AB}$	$(k-1)(m-1)$	$\sigma^2 + (n\sigma_{C(AB)}^2) + np\sigma_{AB}^2$
C(AB)	$SAK_C + SAK_{AC}$ $+ SAK_{BC} + SAK_{ABC}$	$km(p-1)$	$\sigma^2 + n\sigma_{C(AB)}^2$
Res	$SAK_{Res}$	$kmp(n-1)$	$\sigma^2$
Total	$SAK_{Tot}$	$kmpn-1$	

Angående parenteserne gælder de sædvanlige bemærkninger (p.5.44)

**Struktur 4** To faktorer krydsklassificerede og begge underordnet den tredje faktor. Symbolsk  $A \supset B$ ,  $A \supset C$  og  $B \times C$ . Strukturen er anskueliggjort grafisk nedenfor.



Vi har den systematiske model

$$X_{ijrv} = \mu + a_i + b(a)_{j(i)} + c(a)_{r(i)} + bc(a)_{jr(i)} + Z_{v(ijr)} .$$

Den tilfældige model er

$$X_{ijrv} = \mu + A_i + B(A)_{j(i)} + C(A)_{r(i)} + BC(A)_{jr(i)} + Z_{v(ijr)} .$$

Båndene på parametrene i den systematiske (og de blandede) modeller er

$$\sum_j bc(a)_{jr(i)} = 0, \quad \sum_r bc(a)_{jr(i)} = 0$$

$$\sum_r c(a)_{r(i)} = 0, \quad \sum_j b(a)_{j(i)} = 0$$

$$\sum_i a_i = 0 .$$

Variansanalyseeskemaet bliver

Variation	SAK	f	E(SAK/f)
A	$SAK_A$	$k-1$	$\sigma^2 + (n\sigma_{BC(A)}^2 + mn\sigma_{C(A)}^2 + pn\sigma_{B(A)}^2) + mpn\sigma_A^2$
B(A)	$SAK_B + SAK_{AB}$	$k(m-1)$	$\sigma^2 + (n\sigma_{BC(A)}^2) + pn\sigma_{B(A)}^2$
C(A)	$SAK_C + SAK_{AC}$	$k(p-1)$	$\sigma^2 + (n\sigma_{BC(A)}^2) + mn\sigma_{C(A)}^2$
BC(A)	$SAK_{BC} + SAK_{ABC}$	$k(m-1)(p-1)$	$\sigma^2 + n\sigma_{BC(A)}^2$
Res	$SAK_{Res}$	$kmp(n-1)$	$\sigma^2$
Total	$SAK_{Tot}$	$kmpn-1$	

Varianskomponenterne skal, såfremt de tilsvarende effekter er systematiske, læses som

$$\sigma_{BC(A)}^2 = \frac{1}{k(m-1)(p-1)} \sum_i \sum_j \sum_r bc(a)_{jr(i)}^2$$

$$\sigma_{C(A)}^2 = \frac{1}{k(p-1)} \sum_i \sum_r c(a)_{r(i)}^2$$

$$\sigma_{B(A)}^2 = \frac{1}{k(m-1)} \sum_i \sum_j b(a)_{j(i)}^2$$

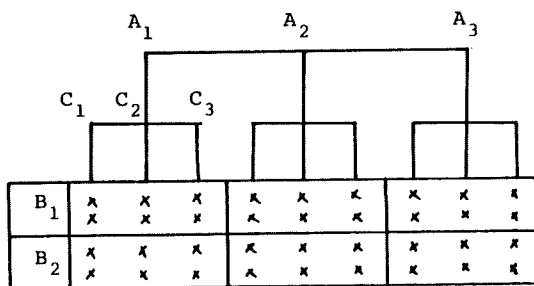
$$\sigma_A^2 = \frac{1}{k-1} \sum_i a_i^2 .$$

Vedrørende parenteserne henvises fremdeles til bemærkningen p. 5.44.

Struktur 5 De to faktorer krydsede og den tredje underordnet den ene og krydset med den anden, d.v.s.

$$A \times B, B \times C, A \supset C .$$

Strukturen er søgt anskueliggjort i nedenstående skitse.



Den systematiske model er

$$X_{ijrv} = \mu + a_i + b_j + ab_{ij} + c(a)_{r(i)} + bc(a)_{jr(i)} + Z_{v(ijr)}$$

og den tilfældige

$$X_{ijrv} = \mu + A_i + B_j + AB_{ij} + C(A)_{r(i)} + BC(A)_{jr(i)} + Z_{v(ijr)}$$

Parameterbåndene er

$$\sum_i a_i = \sum_j b_j = \sum_i ab_{ij} = \sum_j ab_{ij} = 0$$

$$\sum_r c(a)_{r(i)} = \sum_r bc(a)_{jr(i)} = \sum_j bc(a)_{jr(i)} = 0$$

Variansanalyseeskemaet bliver

Variation	SAK	f	E(S <sup>2</sup> )
A	SAK <sub>A</sub>	k-1	$\sigma^2 + (n\sigma_{BC(A)}^2 + pn\sigma_{AB}^2 + mn\sigma_{C(A)}^2) + mpn\sigma_A^2$
B	SAK <sub>B</sub>	m-1	$\sigma^2 + (n\sigma_{BC(A)}^2 + pn\sigma_{AB}^2) + kp\sigma_B^2$
AB	SAK <sub>AB</sub>	(k-1)(m-1)	$\sigma^2 + (n\sigma_{BC(A)}^2) + pn\sigma_{AB}^2$
C(A)	SAK <sub>C</sub> + SAK <sub>AC</sub>	k(p-1)	$\sigma^2 + (n\sigma_{BC(A)}^2) + mn\sigma_{C(A)}^2$
BC(A)	SAK <sub>BC</sub> + SAK <sub>ABC</sub>	k(m-1)(p-1)	$\sigma^2 + n\sigma_{BC(A)}^2$
Res	SAK <sub>Res</sub>	kmp(n-1)	$\sigma^2$
Tot	SAK <sub>Tot</sub>	kmpn-1	

Den deterministiske parameterfortolkning er

$$\sigma_A^2 = \frac{1}{k-1} \sum_i a_i^2$$

$$\sigma_B^2 = \frac{1}{m-1} \sum_j b_j^2$$

$$\sigma_{AB}^2 = \frac{1}{(k-1)(m-1)} \sum_{i,j} ab_{ij}^2$$

$$\sigma_{C(A)}^2 = \frac{1}{k(p-1)} \sum_{i,r} c(a)_{r(i)}^2$$

$$\sigma_{BC(A)}^2 = \frac{1}{k(m-1)(p-1)} \sum_{i,j,r} bc(a)_{jr(i)}^2$$

med den sædvanlige (p. 5.44) bemærkning om parenteserne.

## 5.5 Variansanalyser med flere faktorer

Vi indleder med et afsnit om

### 5.5.1 Estimation af parametre og beregning af kvadratafvi- gelsessummer

Vi har set, at variansanalysemodellerne med to og tre faktorer bygger på de fuldstændigt krydsede modeller. Helt analogt bygger de forskellige modeller med  $k$  faktorer på analysen af den fuldstændigt krydsede model med  $k$  faktorer.

Vi skal nu give en oversigt over fremgangsmåden. Faktorerne be-  
nævnes

$$F_1, \dots, F_k,$$

men for at undgå en ekstra indicering vil vi ofte skrive

$$A, B, \dots, G, H.$$

Antallet af niveauer for  $F_i$  sættes lig  $n_i$ . Svarende til disse niveauer forudsættes, at der foreligger observationer



$$\begin{aligned}
 X_{i_1 \dots i_k v}, \quad & i_1 = 1, \dots, n_1 \quad (\sim \text{faktor } F_1 = A) \\
 & \vdots \\
 & i_k = 1, \dots, n_k \quad (\sim \text{faktor } F_k = H) \\
 & v = 1, \dots, n,
 \end{aligned}$$

hvor  $v$  angiver et gentagelsesindex.

I den deterministiske eller rent systematiske model antages, at middelværdierne i cellerne er spaltet som følger

$$\begin{aligned}
 E(X_{i_1 \dots i_k v}) &= \mu_{i_1 \dots i_k} \\
 &= \mu + a_{i_1} + \dots + h_{i_k} \\
 &\quad + ab_{i_1 i_2} + \dots + gh_{i_{k-1} i_k} \\
 &\quad \vdots \\
 &\quad + a_{i_1 \dots i_{k-1}} + \dots + b_{i_2 \dots i_k} \\
 &\quad + a_{i_1 \dots i_k}.
 \end{aligned}$$

Her har vi også anvendt skrivemåden  $a, b, \dots, g, h$  i stedet for  $f_1, f_2, \dots, f_{k-1}, f_k$ .

For en vilkårlig af effekterne og af vekselvirkningerne gælder, at summen over et vilkårligt index er 0.

Estimatorerne for parametrene bestemmes på sædvanlig måde. Man finder følgende skøn, hvor vi samtidig har angivet den eller de faktorer, hvis effekter vi estimerer

$$\begin{aligned}
 A : \hat{a}_{i_1} &= \bar{X}_{i_1 \dots} - \bar{X} \\
 &\vdots \\
 H : \hat{h}_{i_k} &= \bar{X}_{\dots i_k} - \bar{X} \\
 AB : \hat{ab}_{i_1 i_2} &= \bar{X}_{i_1 i_2 \dots} - \bar{X}_{i_1 \dots} - \bar{X}_{\dots i_2 \dots} + \bar{X} \\
 &\vdots
 \end{aligned}$$

$$\text{GH} : \widehat{gh}_{i_{k-1}i_k} = \bar{X}_{\dots i_{k-1}i_k} - \bar{X}_{\dots i_{k-1}} - \bar{X}_{\dots i_k} + \bar{X}$$

$$\text{ABC} : \widehat{abc}_{i_1 i_2 i_3}$$

$$= \bar{X}_{i_1 i_2 i_3 \dots} - \bar{X}_{i_1 i_2 \dots} - \bar{X}_{i_1 \dots i_3 \dots} - \bar{X}_{\dots i_2 i_3 \dots}$$

$$+ \bar{X}_{i_1 \dots} + \bar{X}_{\dots i_2 \dots} + \bar{X}_{\dots i_3 \dots} - \bar{X}$$

⋮

$$\text{AB---G} : \widehat{ab---g}_{i_1 \dots i_{k-1}}$$

$$= \bar{X}_{i_1 \dots i_{k-1}} - \bar{X}_{i_1 \dots i_{k-2}} - \bar{X}_{i_1 \dots i_{k-3} i_{k-1}} - \dots$$

$$+ (-1)^{k-3} \left[ \bar{X}_{i_1 i_2 \dots} + \dots + \bar{X}_{\dots i_{k-2} i_{k-1}} \right]$$

$$+ (-1)^{k-2} \left[ \bar{X}_{i_1 \dots} + \dots + \bar{X}_{\dots i_{k-1}} \right] + (-1)^{k-1} \bar{X}$$

⋮

$$\text{AB---H} : \widehat{ab---h}_{i_1 \dots i_k}$$

$$= \bar{X}_{i_1 \dots i_k} - \bar{X}_{i_1 \dots i_{k-1}} - \dots + (-1)^k \bar{X}$$

Et punktum på en indeksplads for X'ernes vedkommende betyder, at vi har summeret over det pågældende index. For at undgå forvekslinger med punktummer i en opremsning som a---h, er disse punktummer her erstattet med - .

Skønnet fremgår altså som

gennemsnittet over de indices, der svarer til faktorer, der ikke indgår i effekten,

- alle de gennemsnit, der fremkommer ved at tage gennemsnit over yderligere et index,
- + alle de gennemsnit, der fremkommer ved at tage gennemsnit over yderligere 2 indices (i forhold til det første gennemsnit),
- 
- .
- .
- .
- +  $(-1)^m$  \* totale gennemsnit.

For en vilkårlig faktorkombination  $F_p \dots F_r$  defineres nu den tilsvarende vekselvirkningskvadratafvigelsessum

$$SAK_{F_p \dots F_r} = \Sigma (\widehat{f_p \dots f_r})^2_{i_p \dots i_r} ,$$

hvor summen udstrækkes over samtliche k indices  $i_1, \dots, i_k$ .

Af hensyn til eventuelle beregninger kan det bemærkes, at kvadratsummerne kan skrives som summer og differenser af kvadraterne på de led, der indgår i de enkelte vekselvirkningsestimater.

Eksempelvis er

$$\begin{aligned} SAK_{F_1 F_2 F_3} &= \Sigma (\widehat{f_1 f_2 f_3})^2_{i_1 i_2 i_3} \\ &= \Sigma \bar{X}_{i_1 i_2 i_3}^2 \dots - \Sigma \bar{X}_{i_1 i_2}^2 \dots - \Sigma \bar{X}_{i_1 i_3}^2 \dots \\ &\quad - \Sigma \bar{X}_{i_2 i_3}^2 \dots + \Sigma \bar{X}_{i_1}^2 \dots + \Sigma \bar{X}_{i_2}^2 \dots \\ &\quad + \Sigma \bar{X}_{i_3}^2 \dots - \Sigma \bar{X}^2 . \end{aligned}$$

Her må det stadig erindres, at der summeres over samtlige indices. E.g. er

$$\Sigma \bar{X}^2 = n_1 \dots n_k n \bar{X}^2 .$$

Ved beregninger bør man af numeriske årsager heller ikke beregne de enkelte kvadratsummer direkte, men beregne kvadrater på summer og så foretage de nødvendige divisioner bagefter. E.g. er

$$\sum \bar{X}_{i_1 i_2 \dots}^2 = \frac{1}{n_3 \cdot n_4 \dots n_k n} \sum_{i_1, i_2} \left( \sum_{i_3, \dots, i_k, v} X_{i_1 i_2 i_3 \dots i_k v} \right)^2$$

Antallet af frihedsgrader for kvadratsummerne er

$$DF(SAK_{F_p \dots F_r}) = (n_p - 1) \dots (n_r - 1) .$$

Vi samler disse resultater i det grundlæggende variansanalysekema for den fuldstændigt balancerede k-faktor variansanalysemodel. Skemaet er anført nedenfor.

Variationsårsag	SAK	Frihedsgrader
$F_1$	$SAK_{F_1}$	$n_1 - 1$
$\vdots$	$\vdots$	$\vdots$
$F_k$	$SAK_{F_k}$	$n_k - 1$
$F_1 F_2$	$SAK_{F_1 F_2}$	$(n_1 - 1)(n_2 - 1)$
$\vdots$	$\vdots$	$\vdots$
$F_1 \dots F_k$	$SAK_{F_1 \dots F_k}$	$(n_1 - 1) \dots (n_k - 1)$
Residual	$SAK_{Res}$	$n_1 \dots n_k (n - 1)$
Total	$SAK_{Tot}$	$n_1 \dots n_k (n - 1)$

Grundlæggende variansanalysekema for balanceret k-faktorforsøg.

Forekommer der en hierarkisk ordning af visse faktorer eller en blanding af krydsning og hierarkisk ordning af faktorerne i modellen, kan såvel estimatorer som kvadratafvigelsessummer beregnes relativt enkelt ud fra den fuldstændigt krydsede model.

Lad os e.g. antage, at faktorer, der for simpelheds skyld benævnes  $B_1, \dots, B_r$ , er underordnede faktorer  $C_1, \dots, C_s$ .

Da finder vi for effekter svarende til  $B_1 \dots B_r (C_1 \dots C_s)$

$$\begin{aligned}
 & \overbrace{\{b_1 \dots b_r (c_1 \dots c_s)\}}_{i_1 \dots i_r (j_1 \dots j_s)} \\
 &= \overbrace{\{b_1 \dots b_r\}}_{i_1 \dots i_r} + \overbrace{\{b_1 \dots b_r c_1\}}_{i_1 \dots i_r j_1} + \dots + \\
 &+ \overbrace{\{b_1 \dots b_r c_2 \dots c_s\}}_{i_1 \dots i_r j_2 \dots j_s} + \overbrace{\{b_1 \dots b_r c_1 \dots c_s\}}_{i_1 \dots i_r j_1 \dots j_s} \\
 &= \sum_{\{l_1, \dots, l_t\} \subseteq \{1, \dots, s\}} \overbrace{\{b_1 \dots b_r c_{l_1} \dots c_{l_t}\}}_{i_1 \dots i_r j_{l_1} \dots j_{l_t}}.
 \end{aligned}$$

Formlen synes meget kompliceret; men dens indhold er blot, at man beregner estimator i den fuldstændigt krydsede model, og derefter finder man estimator (i den hierarkiske model) for effekter svarende til  $B_1 \dots B_r (C_1 \dots C_s)$  ved at addere estimator for alle de effekter (i den krydsede model), der indeholder samtlige  $B_1, \dots, B_r$  og fra 0 til alle  $C_1, \dots, C_s$ .

Vi bemærker i øvrigt, at de parameterbånd, der kommer på tale her, er, at summer over et vilkårligt index, der ikke står inde i en parentes, giver 0. Det ses umiddelbart, at også estimatorerne tilfredsstiller dette krav.

Skal man beregne kvadratafvigelsessummen svarende til ovennævnte effekt, sker det efter fuldstændigt samme regel, i.e.

$$\begin{aligned}
 \text{SAK}_{B_1 \dots B_r}(C_1 \dots C_s) &= \text{SAK}_{B_1 \dots B_r} + \text{SAK}_{B_1 \dots B_r} C_1 + \dots + \\
 &\quad \text{SAK}_{B_1 \dots B_r} C_2 \dots C_s + \text{SAK}_{B_1 \dots B_r} C_1 \dots C_s \\
 &= \sum_{\{\ell_1 \dots \ell_t\} \subseteq \{1, \dots, s\}} \text{SAK}_{B_1 \dots B_r} C_{\ell_1} \dots C_{\ell_t}
 \end{aligned}$$

Antallet af frihedsgrader bliver summen af frihedsgraderne. Sættes antallet af niveauer for  $C_v$  lig  $n_v$  og antallet af niveauer for  $B_v$  lig  $m_v$ , fås

$$\text{DF}(\text{SAK}_{B_1 \dots B_r}(C_1 \dots C_s)) = n_1 \dots n_s (m_1 - 1) \dots (m_r - 1) .$$

### 5.5.2 Beregning af forventede værdier af middelvadratafvi- gelsessummer

Ved testningen i en variansanalyse er det som sagt meget væsentligt at kunne beregne forventede værdier af middelvadratafvi-  
gelsessummer.

Vi betragter en variansanalyse med faktorer

$$\dots F_1 \dots ; A ; D_1, \dots, D_t ; B_1, \dots, B_r ; C_1, \dots, C_s ; R ,$$

hvor  $R$  svarer til gentagelse. Antal niveauer for  $F_i = n_i$ .

Hvis der er tale om varianskomponentmodeller eller blandinger mellem varianskomponentmodeller og rent systematiske modeller, gør vi de sædvanlige forudsætninger, nemlig at alle involverede stokastiske variable er

- |                          |                              |
|--------------------------|------------------------------|
| 1) stokastisk uafhængige | 3) middelværdi 0 og med      |
| 2) normalt fordelte med  | 4) en varians, der evt. er 0 |

Endvidere betragter vi i første omgang kun modeller, hvor det er således, at hvis en faktor er stokastisk, da er alle vekselvirkninger, hvori denne indgår, også stokastiske.

Der beregnes først et skema

Effekt	Faktor ..... $F_i$ .....	R
⋮ A	⋮ $\delta_{F_i} A$	⋮ n
⋮ $D_1 \dots D_t$	⋮ $\delta_{F_i} D_1 \dots D_t$	⋮ n
⋮ $B_1 \dots B_r (C_1 \dots C_s)$	⋮ $\delta_{F_i} B_1 \dots B_r (C_1 \dots C_s)$	⋮ n
⋮	⋮	⋮
$R(\dots A \dots C_s)$	..... 1 .....	1

Her er

$$\delta_{F_i} A = \begin{cases} n_i, & \text{hvis } F_i \neq A \\ 1, & \text{hvis } F_i = A \text{ og } F_i \text{ stokastisk} \\ 0, & \text{hvis } F_i = A \text{ og } F_i \text{ deterministisk} \end{cases}$$

$$\delta_{F_i} D_1 \dots D_t = \begin{cases} n_i, & \text{hvis } F_i \neq D_j \quad \forall j = 1, \dots, t \\ 1, & \text{hvis } \exists j: F_i = D_j, \text{ og blot en af} \\ & \text{D'erne er stokastisk} \\ 0, & \text{hvis } \exists j: F_i = D_j, \text{ og alle D'er} \\ & \text{er deterministiske} \end{cases}$$

$$\delta_{F_i} B_1 \dots B_r (C_1 \dots C_s) = \begin{cases} n_i, & \text{hvis } \forall j: F_i \neq B_j \wedge \forall k: F_i \neq C_k \\ 1, & \text{hvis } \exists k: F_i = C_k \\ 1, & \text{hvis } \exists j: F_i = B_j \text{ og blot en af} \\ & \text{B'erne eller C'erne er stokasti-} \\ & \text{ske} \\ 0, & \text{hvis } \exists j: F_i = B_j \text{ og alle B'er og} \\ & \text{C'er er deterministiske} \end{cases}$$

Ved den praktiske udfyldning af skemaet går man mest hensigtsmæssigt frem som følger. Vi betragter rækken svarende til en bestemt effekt.

- ① Vi udsøger nu de faktorer, som ikke indgår i effekten. I de respektive søjler skrives antallet af niveauer for den tilsvarende faktor.
- ② Dernæst udsøges de faktorer, som indgår i en parentes i effekten. I de respektive søjler skrives 1.
- ③ På de resterende pladser skrives 1, hvis effekten er stokastisk, og 0, hvis den er deterministisk.

Den forventede værdi af middeltkvadratafvigelsessummen svarende til en bestemt effekt er af formen

$$E(\text{SAK}/f) = \sigma^2 + \dots + \alpha_1 \varphi_A^2 + \dots + \alpha_2 \varphi_{D_1 \dots D_t}^2 \\ + \dots + \alpha_3 \varphi_{B_1 \dots B_r (C_1 \dots C_s)}^2$$

Her er  $\alpha_i$  konstanter, der bestemmes i det følgende, og

$$\varphi_{\text{effekt}}^2 = \begin{cases} \sigma_{\text{effekt}}^2, & \text{hvis effekten er stokastisk} \\ \frac{1}{f_{\text{effekt}}} \sum (\text{parameter}_{\text{effekt}})^2, & \text{hvis effekten er deterministisk.} \end{cases}$$

$\alpha_i$ 'erne fremkommer nu ved



- i at fjerne alle de søjler, der svarer til faktorer, der indgår i effekten,
- ii opsøge alle de rækker, der svarer til effekter, som indeholder alle de faktorer, der indgår i den bestemte effekt, vi undersøger.
- iii I hver af de betragtede rækker danner vi nu produktet af de led, der efter sletningen i **i** er tilbage.

De derved fremkomne størrelser er koefficienterne  $\alpha_i$ .

Efter således at have bestemt alle forventningsværdier, danner man F-tests for de relevante hypoteser på vanlig måde ved at beregne kvotienter mellem SAK/f'er.

Ofte vil det dog ikke være muligt at danne en teststørrelse umiddelbart, idet der muligvis ikke findes nogen SAK/f, der har en forventningsværdi, som direkte muliggør en testning af den interessante parameter.

Man vil da danne en linearkombination af flere SAK/f'er, som har den ønskede forventningsværdi, f. eks.

$$MS = a_1 MS_1 + \dots + a_k MS_k,$$

hvor  $MS_i = SAK_i / f_i$ . Hvis  $E(MS_i) = \sigma_i^2$ , fås

$$E(MS) = a_1 \sigma_1^2 + \dots + a_k \sigma_k^2$$

og

$$\begin{aligned} V(MS) &= a_1^2 V(MS_1) + \dots + a_k^2 V(MS_k) \\ &= 2 \left[ a_1^2 \frac{\sigma_1^4}{F_1} + \dots + a_k^2 \frac{\sigma_k^4}{F_k} \right]. \end{aligned}$$

Vi har her benyttet, at  $V(\chi^2(f)) = 2f$ , d.v.s.

$$V(\sigma^2 \chi^2(f)/f) = 2 \frac{\sigma^4}{f}.$$

Hvis nu MS var  $\sigma^2 \chi^2(f)/f$ -fordelt, ville

$$f = \frac{2 E^2(MS)}{V(MS)}.$$

Da nu  $\hat{\theta}_i^2 = MS_i$ , finder vi estimatet

$$f = \frac{[a_1 MS_1 + \dots + a_k MS_k]^2}{\frac{MS_1^2}{a_1^2 \frac{1}{f_1}} + \dots + \frac{MS_k^2}{a_k^2 \frac{1}{f_k}}}.$$

Vi foretager nu testningen, som om MS var  $\sigma^2 \chi^2(\hat{f})/\hat{f}$ -fordelt, og danner de relevante F-størrelser m.v.

Indholdet i ovenstående metode er, at vi opfatter MS som følgende den  $\sigma^2 \chi^2(f)/f$ -fordeling, der har samme middelværdi og varians som MS. Hvis alle  $a_i$ 'er er positive, vides dette at være en fremragende approximation til den eksakte fordeling. Hvis nogle  $a_i$ 'er er negative, kender forfatteren ingen resultater angående kvaliteten af approximationen.

Inden vi giver et illustrativt eksempel, skal vi gøre opmærksom på, at der optræder en række variansanalysemodeller i litteraturen, der nok minder meget om de her omtalte, men som alligevel afviger på væsentlige punkter.

Vi præciserer, at metodernes gyldighed forudsætter, at der er tale om fuldstændige, balancerede faktorforsøg, hvor faktorer er krydsede eller hierarkiske, og alle involverede, stokastiske variable er forudsat uafhængige.

Hvis man ønsker at betragte afhængige vekselvirkningsvariable, nemlig f. eks.  $T_{ij}$ 'er, der tilfredsstiller

$$\sum_i T_{ij} = 0$$

i modellen

$$X_{ijv} = \mu + \alpha_i + B_j + T_{ij} + Z_{v(ij)},$$

må man ændre på reglerne for udfyldelse af "skemaet" p. 5.59.  
Fase (3) ændres til

(3a). På de resterende pladser skrives 1, hvis faktoren er stokastisk, og 0, hvis den er deterministisk.

Ændringerne i de præcise formler for  $\delta$ 'erne er åbenbare. Vendingen "blot en af D'erne (eller B'erne eller C'erne) er stokastiske" erstattes med "og  $F_i$  er stokastisk". Tilsvarende erstattes "og alle D'er (og B'er og C'er) er deterministiske" med "og  $F_i$  er deterministisk". Den præcise forudsætning, vi nu må gøre, er, at summen af vekselvirkningsvariable over indices, der hører til deterministiske faktorer, er 0, d.v.s. de er ikke længere uafhængige.

I de tilfælde, hvor der såvel er deterministiske som stokastiske faktorer, er de F-tests, man kommer frem til, kun approximative (Scheffé (1959) p. 270). Hvis kun en hovedeffekt er stokastisk, kan man dog opnå eksakte tests ved hjælp af Hotellings  $T^2$ -størrelse, jvf. Scheffé (1959) p. 270 og 288.

Hvis der ønskes taget hensyn til, at en faktor muligvis kun kan forekomme i endeligt mange niveauer, e.g. H, og vi i vores forsøg tilfældigt har udvalgt h blandt disse, erstattes 1-taller i det grundlæggende skema med størrelsen

$$1 - \frac{h}{H}.$$

Dette gælder dog ikke 1-taller affødt ved, at et index er hierarkisk underordnet andre. Sådanne 1-taller bevares. Der kan henvises til Johnson & Leone (1964) p. II. 144-5 eller Bennett & Franklin (1954) p. 414.

Det forekommer forfatteren, at begrebet afhængige vekselvirkningsvariable giver en lidt mærkelig ad hoc forklaring af den datagenerende proces, men disse modeller har vundet stort indpas i litteraturen. Således fremføres de af Bennett & Franklin (1954) p. 414, Hicks (1973) p. 177, Johnson & Leone (1964) p. II 144 og Scheffé (1959) p. 284, og af ovenstående nævner kun Johnson & Leone (1964) p. II 145 tilfældet med uafhængige vekselvirkninger, skønt disse modeller ofte må foretrækkes.

Det må dog indskærpes, at det er væsentligt at gøre sig klart, hvilken model man opererer med. Fremgangsmåden ved testningen kan være forskellig ved de to modeller (og selvfølgelig bliver tolkningen af resultater altid forskellig).

Beviser for, at skemaerne til brug for beregninger af forventede værdier af SAK/f'er er korrekt, må føres induktivt. Et sådant bevis er skitseret hos Cornfield & Tukey (1959). Også Wilk & Kempthorne (1955) skitserer beviser.

Slutteligen vil det vel være rimeligt kort at komme med nogle korte bemærkninger vedrørende variansanalysernes robusthed over for afvigelser fra forudsætningerne.

Den grundlæggende antagelse er, at residualstørrelserne  $Z_{ij\dots v}$  er normalt fordelte, uafhængige og har samme varians  $\sigma^2$ . Det viser sig imidlertid, at normalitetsantagelsen ikke er særligt kritisk. Styrkefunktionen for testene er relativt ufølsom over for selv relativt store afvigelser i fordelingerne form.

Sådanne afvigelser kan imidlertid få "katastrofale" konsekvenser for et eventuelt test for, at varianserne er ens. Således er Bartlett's test meget følsomt over for afvigelser fra normalitet. Der findes dog andre ikke helt så følsomme tests (se bind 1 p. 5.82).

Til gengæld viser det sig imidlertid, at kravet om varianshomo-genitet ikke er helt så afgørende, som man kunne forestille sig, hvis forsøget er balanceret, i.e. hvis der er lige mange observationer i hver celle.

For en mere dybtgående diskussion må henvises til litteraturen, e.g. Scheffé (1959) p. 331.

Vi giver nu et illustrativt eksempel.

Eksempel 5.6 Nedenstående data (Bennett & Franklin (1954)) giver summerne af tre gentagne bestemmelser af 700%-elasticitetsmodulet for emner fremstillet af gummi, stammende fra 4 forskellige ladninger rågummi.

For hver ladning har man efter en omhyggelig blandingsprocedure udvalgt to stikprøver af gummi fra hver af 4 baller rågummi.

Hver stikprøve er dernæst delt i to sub-stikprøver. Disse blev dernæst brugt til at fremstille standardblandinger indeholdende henholdsvis 0.5% og 3.0% stearinsyre. Prøver fra hver af disse blandinger blev dernæst hærde i 60 eller 120 minutter, hvorefter 700%-elasticitetsmodulet er bestemt.

Rækkefølgen, hvori disse operationer udføres, blev omhyggeligt randomiseret, således at et tidstrend ikke skulle kunne influere på de effekter, man var interesseret i at undersøge.

		C <sub>1</sub>				C <sub>2</sub>				C <sub>3</sub>				C <sub>4</sub>			
		B <sub>1</sub>		B <sub>2</sub>		B <sub>1</sub>		B <sub>2</sub>		B <sub>1</sub>		B <sub>2</sub>		B <sub>1</sub>		B <sub>2</sub>	
		A <sub>1</sub>	A <sub>2</sub>	A <sub>1</sub>	A <sub>2</sub>	A <sub>1</sub>	A <sub>2</sub>	A <sub>1</sub>	A <sub>2</sub>	A <sub>1</sub>	A <sub>2</sub>	A <sub>1</sub>	A <sub>2</sub>	A <sub>1</sub>	A <sub>2</sub>	A <sub>1</sub>	A <sub>2</sub>
E <sub>1</sub>	D <sub>1</sub>	3810	6340	4080	6660	3004	5470	2833	6220	4210	7640	3910	6510	3810	8100	4570	8310
	D <sub>2</sub>	6230	9050	6370	8650	5720	8390	5620	7300	6190	9720	6860	9960	8530	10210	7450	10650
E <sub>2</sub>	D <sub>1</sub>	4350	6450	3800	6920	3380	6330	3308	6280	3310	6040	3302	6670	4380	8010	4860	8100
	D <sub>2</sub>	6250	9090	6700	9000	6060	8770	5820	9180	5660	8540	5690	8720	7130	10170	7350	9430
E <sub>3</sub>	D <sub>1</sub>	3690	6510	3590	6510	2758	5740	2945	6950	3445	6690	3273	5980	4160	7370	4670	7380
	D <sub>2</sub>	6630	8680	6360	8440	5630	8200	5170	8310	6180	8060	5700	8210	7380	9560	7610	9400
E <sub>4</sub>	D <sub>1</sub>	4390	7470	4700	7210	3810	6830	3320	6020	2684	6010	3080	5630	5270	8010	4640	7540
	D <sub>2</sub>	7170	9370	8800	6280	9240	6170	8680	5790	5790	7980	5180	7960	7350	10510	7590	10420

Data til eksempel 5.6

Vi ser, at der i alt optræder 5 faktorer, nemlig

$$\begin{aligned} A_m &= \text{stearinsyreniveau, } m = 1,2 \\ B_k &= \text{stikprøve-nr., } k = 1,2 \\ C_i &= \text{ladnings-nr., } i = 1, \dots, 4 \\ D_\ell &= \text{hærdningsniveau, } \ell = 1,2 \\ E_j &= \text{balle-nr., } j = 1, \dots, 4 . \end{aligned}$$

Der foreligger sådan set også en gentagelseseseffekt, men da vi kun har opgivet summerne af de tre gentagelsesmålinger og ikke også deres kvadratsummer, må vi se bort fra dem i dette eksempel.

Ved fastlæggelsen af en model vil det nu være hensigtsmæssigt at gå frem som følger.

i) Vi skal først afgøre, hvilke af ovenstående faktorer, der er krydsede og hvilke, der er nestede. Vi starter systematisk fra oven

$$\begin{aligned} &A \times B, A \times C, A \times D, A \times E, \\ &B \subset C, B \times D, B \subset E, \\ &C \times D, C \supset E, \\ &D \times E. \end{aligned}$$

ii) Ved fastlæggelse af middelværdistrukturen inddrager vi nu faktorerne 1 efter 1, og det giver, idet vi noterer hvert enkelt trin (vi har ikke taget stilling til hvilke faktorer, der skal opfattes deterministiske og hvilke, der skal opfattes stokastiske. Vi anvender derfor de samme symboler, som vi har anvendt for faktorerne, i.e. her: store bogstaver),

$$\begin{aligned} &A \\ &+ B(EC) + AB(EC) \\ &+ C + AC \\ &+ D + AD + BD(EC) + ABD(EC) + CD + ACD \\ &+ E(C) + AE(C) + DE(C) + ADE(C) . \end{aligned}$$

Vi bemærker, at når en ny faktor inddrages, skal man a) huske at angive alle faktorer, der er overordnet den pågældende faktor, i en parentes bag faktoren. Eksempelvis er den faktor, der inddrages som nr. 2, faktoren B. Ved at checke i listen ovenfor ser vi, at B er underordnet såvel C som E. De skal derfor altid optræde i en parentes efter B. Derefter skal man b) anføre de led, der fremkommer ved at "parre" den sidst inddragne faktor med alle de led, der allerede forekommer. Herunder gøres igen brug af de under i) anførte relationer mellem faktorerne, idet man samtidigt erindrer, at hvis en faktorkombination allerede forekommer, skal den selvfølgelig ikke med igen, ligesom et "bogstav" udelades uden for en parentes, hvis det ved "parringen" optræder såvel inden i som uden for en parentes. E.g. har vi ved inddragelsen af E(C) parrene

$$\begin{aligned} & AC \ \& \ E(C) \ , \\ & BD(EC) \ \& \ E(C) \ . \end{aligned}$$

Her giver det første anledning til

$$AE(C) \ ,$$

fordi A og E er krydsede. Det andet par giver ikke anledning til noget nyt led, da alle bogstaver i E(C) allerede forekommer i BD(EC).

iii) Vi skal nu afgøre, hvilke af de anførte effekter, der er tilfældige og hvilke, der er systematiske.

Faktorerne, der angiver stearinsyreniveauet og hårdningsniveauet (i.e. A og D) må anses for at være systematiske. Tilsvarende må balle-nr. og stikprøve-nr. (faktorerne E og B) anses for at være tilfældige. Det er lidt mere tvivlsomt, hvad ladningsnr. skal sættes til. Hvis der f. eks. er tale om 4 leverancer fra hvert af de 4 største producentlande, ville det nok være rimeligt at regne faktoren for systematisk. Hvis der imidlertid er tale om tilfældige ladninger indkøbt på en auktion, vil man regne den for tilfældig. Der foreligger ikke yderligere informatio



ner, så vi må træffe et valg. Vi anser den sidste mulighed for at være den mest rimelige, således at vi regner C for tilfældig.

Nu regnes alle de effekter, hvori der kun indgår systematiske faktorer, for systematiske og resten for tilfældige.

Dette giver anledning til den endelige middelværdiopspaltning

$$\begin{aligned}
 X_{mki\ell jv} = & \mu + a_m + B(EC)_{k(ji)} + aB(EC)_{mk(ji)} \\
 & + C_i + aC_{mi} + d_\ell + ad_{m\ell} + Bd(EC)_{k\ell(ji)} \\
 & + aBd(EC)_{mk\ell(ji)} + Cd_{i\ell} + aCd_{mi\ell} + E(C)_{j(i)} \\
 & + aE(C)_{mj(i)} + dE(C)_{\ell j(i)} + adE(C)_{m\ell j(i)} \\
 & + Z_{v(mki\ell j)}
 \end{aligned}$$

hvor gentagelsesindex  $v$  her kun antager 1 værdi.

Her gælder så de sædvanlige forudsætninger om de enkelte led.

iv) For at bestemme de forventede værdier af de SAK'er, der svarer til ovennævnte effekter, udfylder vi nu det p. 5.58 anførte skema.

Tallene i parentes under de enkelte faktorer i tabellen p. 5.70 angiver antallet af niveauer for den pågældende faktor. Et  $d$  i en parentes bag en effekt angiver, at effekten er deterministisk, og et  $t$ , at den er tilfældig. Søjlen med R svarer til gentagelseseffekten. (Vi har ingen muligheder for at finde den, men det udelukker selvsagt ikke, at den kan være til stede.)

Ved hjælp af de p. 5.61 anførte regler finder vi dernæst følgende skema (p. 5.71) over de forventede værdier af middelværdi-afvigelsessummerne.

I skemaet er anvendt betegnelsen  $\phi_{AD}^2$  i stedet for  $\sigma_{AD}^2$  for at angive, at effekten AD er systematisk. Analogt med effekterne A og D. Som fodtegn anvender vi de under ii) anførte betegnelser for effekterne.

Effekt	Faktor					R (3)
	C (4)	E (4)	B (2)	D (2)	A (2)	
C (t)	1	4	2	2	2	3
D (d)	4	4	2	0	2	3
A (d)	4	4	2	2	0	3
E(C) (t)	1	1	2	2	2	3
CD (t)	1	4	2	1	2	3
AC (t)	1	4	2	2	1	3
AD (d)	4	4	2	0	0	3
B(CE) (t)	1	1	1	2	2	3
DE(C) (t)	1	1	2	1	2	3
AE(C) (t)	1	1	2	2	1	3
ACD (t)	1	4	2	1	1	3
BD(CE) (t)	1	1	1	1	2	3
AB(CE) (t)	1	1	1	2	1	3
ADE(C) (t)	1	1	2	1	1	3
ABD(CE) (t)	1	1	1	1	1	3
R(ABDCE) (t)	1	1	1	1	1	1

Tabel til bestemmelse af forventede værdier af kvadrat-  
afvigelsessummer.

Ved hjælp af et standardprogram (ANOVA se p. 5.78) kan vi finde kvadratsummer svarende til en fuldstændigt krydset 5-faktor variansanalyse. Resultatet af en sådan kørsel er vist p. 5.72. Vi finder nu variansanalysekemaet svarende til den her foreliggende struktur ved at addere de forskellige SAK-værdier efter de regler, der er anført p. 5.58.

Dette er gjort i skemaet p. 5.73. Nu må det imidlertid erindres, at vi har foretaget beregningerne på summerne af de gentagne bestemmelser og ikke på gennemsnittene. Dette bevirker selvsagt, at de anførte SAK'er er ganget med en faktor 3 i forhold til det de bør være. Derfor optræder der en SAK = SAK'/3-søjle. (Vi har udeladt  $\Sigma$ . Der er de  $9/3 = 3$  for store).

Effekt	E(SAK/f)
C	$96\sigma_C^2 + 24\sigma_{E(C)}^2 + 48\sigma_{CD}^2 + 48\sigma_{AC}^2 + 12\sigma_{B(CE)}^2 + 12\sigma_{DE(C)}^2 + 12\sigma_{AE(C)}^2 + 24\sigma_{ACD}^2 + 6\sigma_{BD(CE)}^2 + 6\sigma_{AB(CE)}^2 + 6\sigma_{ADE(C)}^2 + 3\sigma_{ABD(CE)}^2 + \sigma^2$
D	$192\phi_D^2 + 48\sigma_{CD}^2 + 12\sigma_{DE(C)}^2 + 24\sigma_{ACD}^2 + 6\sigma_{BD(CE)}^2 + 6\sigma_{ADE(C)}^2 + 3\sigma_{ABD(CE)}^2 + \sigma^2$
A	$192\phi_A^2 + 48\sigma_{AC}^2 + 12\sigma_{AE(C)}^2 + 24\sigma_{ACD}^2 + 6\sigma_{AB(CE)}^2 + 6\sigma_{ADE(C)}^2 + 3\sigma_{ABD(CE)}^2 + \sigma^2$
E(C)	$24\sigma_{E(C)}^2 + 12\sigma_{B(CE)}^2 + 12\sigma_{DE(C)}^2 + 12\sigma_{AE(C)}^2 + 6\sigma_{BD(CE)}^2 + 6\sigma_{AB(CE)}^2 + 6\sigma_{ADE(C)}^2 + 3\sigma_{ABD(CE)}^2 + \sigma^2$
CD	$48\sigma_{CD}^2 + 12\sigma_{DE(C)}^2 + 24\sigma_{ACD}^2 + 6\sigma_{BD(CE)}^2 + 6\sigma_{ADE(C)}^2 + 3\sigma_{ABD(CE)}^2 + \sigma^2$
AC	$48\sigma_{AC}^2 + 12\sigma_{AE(C)}^2 + 24\sigma_{ACD}^2 + 6\sigma_{AB(CE)}^2 + 6\sigma_{ADE(C)}^2 + 3\sigma_{ABD(CE)}^2 + \sigma^2$
AD	$96\phi_{AD}^2 + 24\sigma_{ACD}^2 + 6\sigma_{ADE(C)}^2 + 3\sigma_{ABD(CE)}^2 + \sigma^2$
B(CE)	$12\sigma_{B(CE)}^2 + 6\sigma_{BD(CE)}^2 + 6\sigma_{AB(CE)}^2 + 3\sigma_{ABD(CE)}^2 + \sigma^2$
DE(C)	$12\sigma_{DE(C)}^2 + 6\sigma_{BD(CE)}^2 + 6\sigma_{ADE(C)}^2 + 3\sigma_{ABD(CE)}^2 + \sigma^2$
AE(C)	$12\sigma_{AE(C)}^2 + 6\sigma_{AB(CE)}^2 + 6\sigma_{ADE(C)}^2 + 3\sigma_{ABD(CE)}^2 + \sigma^2$
ACD	$24\sigma_{ACD}^2 + 6\sigma_{ADE(C)}^2 + 3\sigma_{ABD(CE)}^2 + \sigma^2$
BD(CE)	$6\sigma_{BD(CE)}^2 + 3\sigma_{ABD(CE)}^2 + \sigma^2$
AB(CE)	$6\sigma_{AB(CE)}^2 + 3\sigma_{ABD(CE)}^2 + \sigma^2$
ADE(C)	$6\sigma_{ADE(C)}^2 + 3\sigma_{ABD(CE)}^2 + \sigma^2$
ABD(CE)	$3\sigma_{ABD(CE)}^2 + \sigma^2$

Source of variation	Sums of squares	Degrees of freedom
A	247 331 000	1
B	80 501	1
AB	27 907	1
C	47 874 590	3
AC	1 010 358	3
BC	61 802	3
ABC	164 078	3
D	191 805 600	1
AD	1 802 150	1
BD	113 407	1
ABD	5 126	1
CD	282 293	3
ACD	349 272	3
BCD	356 120	3
ABCD	526 729	3
E	2 658 565	3
AE	356 797	3
BE	415 995	3
ABE	236 395	3
CE	11 951 750	9
ACE	1 207 328	9
BCE	708 651	9
ABCE	1 066 008	9
DE	251 008	3
ADE	357 499	3
BDE	174 139	3
ABDE	142 968	3
CDE	884 268	9
ACDE	1 408 724	9
BCDE	1 827 469	9
ABCDE	1 278 840	9
Total	516 714 200	127

Kvadratafvigelsesummer fundet med ANOVA. Grundet afrundingsfejl summerer de anførte SAK'er ikke op til den givne total.

Variationskilde	Faktor kombination	SAK'	SAK = SAK' / 3	f	SAK / f
C	C	47 874 590	15 958 197	3	5 319 399
D	D	191 805 600	63 935 200	1	63 935 200
A	A	247 331 000	82 443 667	1	82 443 667
E(C)	E + CE	14 610 315	4 870 105	12	405 842
CD	CD	282 293	94 098	3	31 366
AC	AC	1 010 358	336 786	3	112 262
AD	AD	1 802 150	600 717	1	600 717
B(CE)	B + BC + BE + BCE	1 266 949	422 316	16	26 395
DE(C)	DE + CDE	1 135 277	378 426	12	31 536
AE(C)	AE + ACE	1 564 125	521 375	12	43 448
ACD	ACD	349 272	116 424	3	38 808
BD(CE)	BD + BCD + BDE + BCDE	2 471 134	823 711	16	51 482
AB(CE)	AB + ABC + ABE + ABCE	1 494 387	498 129	16	31 133
ADE(C)	ADE + ACDE	1 766 223	588 741	12	49 062
ABD(CE)	ABD + ABDE + ABCD + ABCDE	1 953 663	651 221	16	40 701
Total		516 714 200	172 238 067	127	1 356 205

Ved hjælp af variansanalyseeskemaet og skemaet over de forventede værdier af SAK/f kan vi nu teste en række hypoteser om effekterne og varianskomponenterne.

Eksempelvis har vi, at

$$\frac{49062}{40701} = 1.21 < F(12,16)_{0.90} = 1.99 ,$$

hvorfor vi i hvert fald for alle  $\alpha > 10\%$  vil antage hypotesen, at

$$\sigma_{ADE(C)}^2 = 0 .$$

Tilsvarende ses, at vi kan antage

$$\sigma_{AB(CE)}^2 = 0$$

$$\sigma_{BD(CE)}^2 = 0 .$$

Når vi nu vil undersøge, om det kan antages, at  $\sigma_{ACD}^2 = 0$ , kan vi enten betragte teststørrelsen

$$\frac{SAK_{ACD}/3}{SAK_{ADE(C)}/12} ,$$

som er  $F(3,12)$ -fordelt under  $H_0$ , eller vi kan benytte os af, at vi har fået accepteret, at  $\sigma_{ADE(C)}^2 = 0$ , og så betragte

$$\frac{SAK_{ACD}/3}{\text{skøn over } \{\sigma^2 + 3\sigma_{ABD(CE)}^2\}} .$$

Som nævner kan vi da, stadig under forudsætning af at

$$\sigma_{AB(CE)}^2 = 0 ,$$

$$\sigma_{BD(CE)}^2 = 0 ,$$

og

$$\sigma_{ADE(C)}^2 = 0,$$

bruge

$$S_1^2 = (SAK_{ABD(CE)} + SAK_{ADE(C)} + SAK_{AB(CE)} + SAK_{BD(CE)}) / (16+12+16+16) .$$

Teststørrelsen bliver da  $F(3,60)$ -fordelt under  $H_0$ . Fordelen ved denne fremgangsmåde er åbenbart, at vi får flere frihedsgrader i nævneren, d.v.s. vores skøn over den varians, vi skal sammenligne med, bliver præcisere. Ulempen er selvsagt, at selv om vi har fået accepteret, at  $\sigma_{AB(CE)}^2 = 0$  etc., er dette jo ikke nødvendigvis rigtigt, således at teststørrelsen ikke har den fordeling, som vi regner med.

For at løse problemet med, om man skal "poole" disse varianser, kan man anvende en håndregel som f. eks. kun at poole varianser, hvis  $F$ -teststørrelsen er mindre end 2. Dette er selvsagt en vilkårlig regel, men den nyder en vis udbredelse i praksis (cf. Johnson & Leone vol. II).

Anvendes denne fremgangsmåde, fås teststørrelsen

$$\frac{38808}{2561802/60} = 0.91 \approx F(3,60)_{50\%} ,$$

og det ses, at også hypotesen

$$\sigma_{ACD}^2 = 0$$

accepteres.

Fortsættes på denne måde, ses, at vi må antage

$$\exists m, l : ad_{ml} \neq 0$$

$$\sigma_{AC}^2 \neq 0$$

$$\sigma_{E(C)}^2 \neq 0 .$$

Ved en undersøgelse af, om alle  $a_m$ 'er er 0, ses, at vi kan anvende teststørrelsen

$$\frac{SAK_A/1}{SAK_{AC}/3} = \frac{82443667}{112262} = 734.4 ,$$

som skal sammenlignes med en  $F(1,3)$ -fraktil. Det fremgår, at vi på alle rimelige niveauer må antage, at der eksisterer  $a_m$ 'er forskellige fra 0.

Når vi skal undersøge  $d_\ell$ 'erne, kan vi anvende teststørrelsen

$$\frac{SAK_D/1}{S_1^2} = \frac{63935200}{2561802/60} = 1497.4 ,$$

og det ses igen, at vi på alle rimelige niveauer får forkastet hypotesen, at alle  $d_\ell$ 'er er lig 0.

Endelig mangler vi at undersøge  $\sigma_C^2$ . Hvis vi går ud fra alle de hypoteser, vi indtil nu har fået bekræftet, ses, at

$$E(SAK_C/f_C) = 96 \sigma_C^2 + 24 \sigma_{E(C)}^2 + 48 \sigma_{AC}^2 + 3 \sigma_{ABD(CE)}^2 + \sigma^2 ,$$

og

$$E(SAK_{E(C)}/f_{E(C)}) = 24 \sigma_{E(C)}^2 + 3 \sigma_{ABD(CE)}^2 + \sigma^2$$

$$E(SAK_{AC}/f_{AC}) = 48 \sigma_{AC}^2 + 3 \sigma_{ABD(CE)}^2 + \sigma^2 ,$$

Det er derfor ikke umiddelbart muligt at danne en teststørrelse, så vi må bruge et approximativt test, som angivet p. 5.60.

Vi finder, at

$$E\{SAK_{E(C)}/f_{E(C)} + SAK_{AC}/f_{AC} - S_1^2\}$$



$$= E\{S_{E(C)}^2 + S_{AC}^2 - S_1^2\}$$

$$= 24 \sigma_{E(C)}^2 + 48 \sigma_{AC}^2 + 3 \sigma_{ABD(CE)}^2 + \sigma^2 .$$

Vi kan derfor bruge størrelsen i den krøllede parentes som nævner ved et approximativt F-test. Idet

$$S_1^2 = 2\,561\,802/60 = 42697 ,$$

estimerer vi antallet af frihedsgrader til (jvf. p. 5.62)

$$\hat{f} = \frac{[405\,842 + 112\,262 - 42\,697]^2}{\frac{405\,842^2}{12} + \frac{112\,262^2}{3} + \frac{42\,697^2}{60}}$$

$$= 12.59 ,$$

og selve variansskønnet er

$$S^2 = 405\,842 + 112\,262 - 42\,697 = 475\,407 .$$

Teststørrelsen bliver derfor

$$\frac{5\,319\,399}{475\,407} = 11.19 ,$$

som skal sammenlignes med en  $F(3,12.6)$ -fordeling. Det ses, at resultatet af dette approximative F-test bliver, at vi må antage

$$\sigma_C^2 \neq 0 .$$

Som resultat af denne analyse må vi derfor antage, at der er en effekt af ladningsnr., af hårdningsniveau, af stearinsyre niveau og af ballenr. inden for ladningsniveau. Endvidere er der vekselvirkning mellem stearinsyre niveau og hårdningsniveau og mellem stearinsyre niveau og ladningsnr.

En viderefortolkning af disse resultater og en angivelse af estimer overlades til læseren.

□

## 5.6 Variansanalyseprogrammet ANOVA

Vi skal nu kort omtale, hvorledes man nemt løser variansanalyseproblemer ved hjælp af SSP-SAMPLE-programmet ANOVA.

ANOVA kalder nogle SSP subrutiner, der alle er lagt ind under WATFIV compileren, således at det er ret hurtigt og nemt at køre programmet.

Vi bringer en udskrift af programmet.

```

DIMENSION X(3000)
DIMENSION HEAD(6),LEVEL(6),ISTEP(6),KOUNT(6),LASTS(6)
DIMENSION SUMSQ(63),NDF(63),SMEAN(63)
DIMENSION FMT(15)
DATA X/3000*0./
1  FORMAT(14,A2,17,A4,3X,11((A1,I4)/(A1,I4,A1,I4,A1,I4,A1,I4,A1,I4))
2  FORMAT(126H ANALYSIS OF VARIANCE,*,*,A4,A2//)
3  FORMAT(18H LEVELS OF FACTORS/(3X,A1,7X,I4))
4  FORMAT(1H0//11H GRAND MEAN F20.5//)
5  FORMAT(10H SOURCE OF 1BX,7HSUMS OF 10X,10H DEGREES OF 9X,4HMEAN/10H VA
RIATION 1BX,7HSQUARES 11X,7H FREEDOM 10X,7HSQUARES/)
6  FORMAT(1H (SA1,F20.5,10X,16,F20.5)
7  FORMAT(6H TOTAL 10X,F20.5,10X,16)
8  FORMAT(19F4,0)
100 READ (5,1) PR,PR1,K,RLANK,(HEAD(I),LEVEL(I),I=1,K)
WRITE (6,2) PR,PR1
WRITE (6,3) (HEAD(I),LEVEL(I),I=1,K)
N=LEVEL(1)
DO 102 I=2,K
102 N=N+LEVEL(I)
READ (5,8) (X(I),I=1,N)
CALL AVDAT (K,LEVEL,N,X,L,ISTEP,KOUNT)
CALL AVCAL (K,L,FVFL,X,L,ISTEP,LASTS)
CALL MEAND (K,L,FVFL,X,GMEAN,SUMSQ,NDF,SMEAN,ISTEP,KOUNT,LASTS)
WRITE (6,4) GMEAN
WRITE (6,5)
LL=2**K-1
ISTEP(1)=1
DO 105 I=2,K
105 ISTEP(I)=0
DO 110 I=1,LL
110 FMT(I)=BLANK
NN=0
SUM=0.0
NN=NN+1
120 NN=NN+1
L=0
DO 140 I=1,K
FMT(1)=BLANK
IF(ISTEP(I)) 130, 140, 130
130 L=L+1
FMT(L)=HEAD(I)
140 CONTINUE
WRITE (6,6) (FMT(I),I=1,15),SUMSQ(NN),NDF(NN),SMEAN(NN)
SUM=SUM+SUMSQ(NN)
IF(NN-LL) 145, 170, 170
145 DO 160 I=1,K
IF(ISTEP(I)) 147, 150, 147
147 ISTEP(I)=0
GO TO 160
150 ISTEP(I)=1
GO TO 120
160 CONTINUE
170 NN=N-1
WRITE(6,7) SUM,N
GO TO 100
END

```

I den anførte version beregner dette program kvadratafgivelsessummer og frihedsgrader i en variansanalyse med indtil 6 faktorer og indtil 3000 data (mere specifikt kræves  $\prod_{i=1}^6$  [niveau for faktor  $i$ ]+1] < 3000). Hvis disse tal er for små og skal ændres, henvises til SSP-manualen p. 424.

Om input

i) Data. Hvis vi e.g. har data af formen

$$\begin{aligned} X_{ijv} &, i = 1, \dots, k \\ & j = 1, \dots, m \\ & v = 1, \dots, n, \end{aligned}$$

skal de hules i rækkefølgen

$$X_{111}, X_{211}, \dots, X_{k11}, X_{121}, X_{221}, \dots, X_{k21}, X_{131}, \dots, X_{kmn},$$

d.v.s., vi skal lade de første indices variere hurtigst. Data hules efter format-statement nr. 8, i udskriften altså efter

FORMAT (19 F 4.0) .

Dette kort må følgelig ændres, hvis man har data, der er hullet efter et andet format.

ii) Styrekort. Styrekortet hules som følger

Kolonne	Indhold
1-6	navn på analyse (gerne alfamerisk)
7-8	antal faktorer
9-15	blanke
{ 16	navn på første faktor
{ 17-20	antal niveauer for første faktor
{ 21	navn på anden faktor
{ 22-25	antal niveauer for anden faktor
{ 26	navn på tredje faktor
{ 27-30	antal niveauer for tredje faktor
.	.
.	.
.	.

Som output fås nu kvadratafgivelsessummer, frihedsgrader og middelkvadrater for samtlige faktorer. Alt efter hvilken model, man

arbejder med, må man så selv addere de relevante kvadratsummer.

Det vil vel her især være på sin plads at bemærke, at hvis man har gentagelser i ens målinger, må man - selv om man ikke regner med en gentagelseeffekt - indføje en sidste faktor, gentagelsesfaktoren, der eksempelvis kan benævnes R. Man får da residualkvadratafgivelseessummen ved at addere de kvadratsummer, hvori bogstavet R indgår. Hvis man e.g. har faktorerne A, B og C foruden R, bliver

$$\begin{aligned} \text{SAK}_{\text{Res}} = & \text{SAK}_R + \text{SAK}_{AR} + \text{SAK}_{BR} + \text{SAK}_{CR} + \text{SAK}_{ABR} + \text{SAK}_{ACR} \\ & + \text{SAK}_{BCR} + \text{SAK}_{ABCR} , \end{aligned}$$

og analogt med frihedsgraderne.

Eksempel 5.7. Vi betragter de i eksemplet p. 5.27 anførte data. Vi har åbenbart 2 faktorer, henholdsvis 5 og 2 niveauer og endvidere "gentagelsesfaktoren" med 19 niveauer. Vi vælger at indlæse data søjlevis, d.v.s. først data fra den 24/10 1952, dernæst fra den 31/10 1952 etc.

Programmet vil da opfatte materialet på følgende vis

$X_{vji}$ , $v = 1, \dots, 19$	svarende til, at første faktor er gentagelsesfaktoren
$j = 1, 2$	svarende til, at anden faktor er dagnr.-faktoren
$i = 1, \dots, 5$	svarende til, at tredje faktor er partnr.-faktoren.

Benævnes disse faktorer henholdsvis A, B og C, skal vi - foruden \$ ENTRY kortet - anvende følgende input materiale (vi har kaldt eksemplet HIERAR)

Anvendes det p. 5.77 anførte Format (19 F 4.0), vil datakortene blive indlæst som

1890.0 , 1863.0 , ... 2099.0 .

Det resulterende output er

ANALYSIS OF VARIANCE.....HIERAR

LEVELS OF FACTORS

A	19
B	2
C	5

GRAND MEAN            1966.51500

SOURCE OF VARIATION	SUMS OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES
A	63177.03000	18	3509.83400
B	196999.50000	1	196999.50000
AB	60065.04000	18	3336.94600
C	369253.50000	4	92313.37000
AC	235088.00000	72	3265.11100
BC	172941.10000	4	43235.29000
ABC	158734.80000	72	2204.65000
TOTAL	1256258.00000	189	

For nu på basis af dette at konstruere den relevante variansanalysetabel bemærker vi, at modellen p. 5.27 med de her forekommende symboler kan skrives

$$X_{vji} = \mu + C_i + B(C)_{j(i)} + Z_{v(ji)} ,$$

hvor  $B(C)$  er stokastisk og  $C$  deterministisk.

Variansanalysekemaet bliver derfor

Variation	SAK	Frihedsgrad
Mellem partier (C)	$SAK_C$	4
Mellem dage inden for partier (B(C))	$SAK_B + SAK_{BC}$	1+4
Inden for dage (Res.)	$SAK_A + SAK_{AB} + SAK_{AC} + SAK_{ABC}$	18+18+72+72
Total	$SAK_{Tot.}$	189

Indsættes værdierne fra outputtet, ses, at vi netop får varians-analyseskemaet p. 5.29 frem.

□

#### Referencer til kapitel 5

Bennett, C.A., & Franklin, N.L.: Statistical Analysis in Chemistry and the Chemical Industry. John Wiley & Sons, New York 1954.

Cornfield, J., & Tukey, J.W.: Average Values of Mean Squares in Factorials. The Annals of Mathematical Statistics, vol. 27 (1956), pp. 907-949.

Hicks, C.R.: Fundamental Concepts in the Design of Experiments (sec. ed.). Holt, Rinehart and Winston, New York 1973.

Johnson, J.R.: An Application of the Design of Experiments in the Surveillance of Ammunition. Proceedings of the First Conference on the Design of Experiments in Army Research, Development and Testing, 1957.

Johnson, N.L., & Leone, F.C.: Statistics and Experimental Design in Engineering and the Physical Sciences, vol. I + II. John Wiley & Sons, New York 1964.

Opgaver i anvendt statistik. Institut for Matematisk Statistik,  
Københavns Universitet 1968.

Scheffé, H.: The Analysis of Variance. John Wiley & Sons, New  
York 1959.

System/360 Scientific Subroutine Package (SSP). The Internatio-  
nal Business Machines Corporation 1970.

Wilk, M.B., & Kempthorne, O.: Fixed, mixed, and random models  
in the analysis of variance. Journal of the American Sta-  
tistical Association, vol. 50 (1955), pp. 1144-1167.





## INDEX

- a posteriori fordeling, 7.2  
 a priori fordeling, 7.2  
 affin afbildning, 1.9  
 affin støtte, 2.13  
 algebra, 1.1 f  
 aliasing, 9.27  
 alignment, 9.143  
 Andersons U, 6.19  
 ANOVA, 5.69, 5.78 f  
 AR-proces,  
   kontinuert, 9.61 f  
   diskret, 9.39, 9.51 f,  
     9.92, 9.97  
 ARMA-proces  
   kontinuert, 9.61  
   diskret, 9.57 f  
 ARIMA-proces, 9.152 f  
   multiplikativ sæsonmodel,  
     9.156 f  
 associativ lov, 1.2, 1.3  
 autokorrelationsfunktion, 9.31  
   AR-proces, diskret, 9.52  
   MA-proces, 9.49, 9.165  
 Autokovariansfunktion, 9.31 f  
   ARMA-proces, 9.58  
   estimation, 9.63 f  
   hvid støj, 9.44, 9.59  
   lineær proces, diskret, 9.47  
   lineær proces, kontinuert,  
     9.60  
   stationær proces, 9.32  
 autoregressiv proces, se  
 AR-proces  
 B, 9.15  
 backwards elimination, 4.32 f,  
 4.42  
 bagudrettet forskydningsopera-  
 tor, 9.15  
 balanceret variansanalyse,  
 5.13  
 Bartlett vindue, 9.5, 9.108,  
 9.113, 9.118  
 basis, 1.5  
 Bayes løsning, se diskriminant-  
 analyse  
 beregningsformler, 3.58  
 beslutningsfunktion, 7.2  
 betinget fordeling, 2.22 f,  
 2.49 f  
 betinget middelværdi givet pro-  
 ces, 9.160  
 bibetingelser, 3.25  
 blokmatrix, 1.19 f  
 Box-Jenkins' metode, 9.152 f  
 båndpas filter, 9.117  
 "canonical variables, the first  
 two", 7.34  
 central grænseværdisætning, 2.24  
 Cholesky faktoriseringsform, 1.39  
 cofactor, 1.18  
 cospektrum, 9.138, 9.141  
 Cramér-Rao's ulighed, 2.65  
 Cramér's sætning, 1.20  
 cyklisk svingning, 9.81 f  
 $\delta$ , se Dirac's  $\delta$   
 datavindue, 9.24  
 definit, 1.46  
 determinant, 1.17 f, 1.20,  
 1.21, 1.40  
 deterministisk proces, 9.43  
 diag, 1.11  
 diagonalelement, 1.11  
 differensligningsform, 9.154  
 differensoperator, 9.15  
 differentiation af  
   kvadratisk form, 1.60  
   linearform, 1.60  
 digital filtrering af tidsræk-  
 ke, 9.120 f  
 dim, 1.5  
 dimension, 1.5  
 Dirac's  $\delta$ , 9.9 f  
 direkte sum, 1.6 f, 1.68  
 discriminant score, 7.25  
 diskriminantanalyse, 7.1 f  
   Bayes løsning, 7.1 f,  
     7.24 f, 7.26 f

- diskriminantanalyse fortsat  
   flere normale populationer, 7.26 f, 7.32 f  
   flere populationer, 7.24 f  
   minimax løsning, 7.1 f  
   to normale populationer, 7.4 f  
   to populationer, 7.1 f  
   ukendte parametre, med, 7.15 f, 7.31, 7.32 f  
 diskriminantfunktion, diskriminator, 7.7, 7.32  
 diskriminantværdi, 7.25  
 dispersionsmatrix, 7.3 f  
 dispersionsmatrix, estimation af, se estimation  
 dispersionsmatrix, test for, se test  
 distributiv lov, 1.3  
 drejning, 1.40  
  
 Eckart-Youngs sætning, 1.42  
 effekt, 5.1  
 egenvektor, 1.36 f, 1.45  
 egenvektor m.h.t. matrix, 1.54  
 egenværdi, 1.35 f, 1.45, 1.58, 1.59  
 egenværdiproblem, generelle, 1.54 f  
 egenværdi m.h.t. matrix, 1.54  
 ellipsoide, 1.49  
 elliptisk cylinder, 1.51  
 empirisk dispersionsmatrix, 2.26  
 empirisk generaliseret varians, 2.63  
 empirisk partiel korrelation, 2.41  
 enhedsmatrix, 1.11  
 enhedsoperator, 9.16  
 ensemble, 9.30  
 equimax, 8.55  
 ergodeproblem, 9.64  
 estimation af/  
   AR-proces, 9.166  
   ARIMA-proces, 9.164 f  
   dispersionsmatrix, 2.25, 2.61, 6.2, 6.8, 6.16 f, 7.15  
   egenværdi i dispersionsmatrix, 8.7  
   faktor vægte, 8.24 f  
   faktor værdi, 8.33 f  
 flerdimensional generel lineær model, 6.15 f  
  
 flerdimensionale parametre, 2.64 f  
 flerdimensional variansanalyse, 6.29, 6.33  
 generel lineær model, 3.5 f  
 kanoniske korrelationer, 8.20  
 kanoniske variable, 8.20  
 kovariansfunktion, 9.63 f  
 krydskovariansfunktion, 9.136 f  
 krydsspektrum, 9.141 f  
 MA-proces, 9.166  
 multipel korrelationskoefficient, 2.47  
 normal fordeling, 2.25  
 partiel korrelationskoefficient, 2.37  
 principal komponent, 8.7  
 spektrum, 9.100 f  
 euklidisk afstand, 1.65  
 eksponentiel udjævning, 9.73 f, 9.125 f  
  
 F, 9.15  
 factor loading, 8.21  
 factor score, 8.21  
 faktor, 5.1  
 faktor vægt, 8.21  
 faktor værdi, 8.21  
 faktor analyse, 8.20 f  
   estimation af vægte, 8.24 f  
   maximum likelihood analyse, 8.48 f  
   principal faktorenløsning, 8.27  
   rotation, 8.27 f  
   Q-modus analyse, 8.51 f  
   test for model, 8.50 f  
 faktorer, fælles, 8.21  
 faktorer, unikke, 8.22  
 fasefunktion, 9.20  
 fasespektrum, 9.138  
 filtrering, 9.116 f  
 flerdimensional generel lineær model, se generel lineær model, flerdimensional  
 flerdimensional normal fordeling, se normal fordeling, flerdimensional  
 flerdimensional variansanalyse, se variansanalyse, flerdimensional  
 foldning, 9.27

- foldningsintegral, 9.14  
 forskydningsoperator, 9.14 f  
 forstærkning, 9.20  
 forudsigelse, 9.36, 9.38, 9.41, 9.43  
 i ARIMA-proces, 9.158 f  
 opdatering af, 9.163  
 v.h.a. eksponentiel udjævning, 9.75  
 forudsigelsesligning, 9.160  
 forventningsværdi, 2.2 f  
 forward selection, 4.34 f  
 Fourier transformation, diskret, 9.3 f, 9.29  
 kontinuert, 9.2 f  
 frekvensresponsfunktion, 9.20, 9.121  
 funktionel relation, 4.52  
 fysisk realiserbarhed, 9.24  
 gain function, 9.20  
 Gauss-Markov's sætning, 3.7, 6.16  
 generaliseret varians, 2.59, 2.63  
 generel lineær model, 3.1 f  
 generel lineær model, flerdimensional, 6.13 f  
 geodæsi, 3.32  
 glidende gennemsnits proces, se MA-proces  
 glidende gennemsnits udjævning, 9.68 f, 9.124 f  
 Gompertz' trend, 9.79  
 Hamming vindue, 9.7, 9.108, 9.113, 9.118  
 Hamming tapering, 9.122, 9.146  
 Hanning vindue, 9.6, 9.108, 9.113, 9.118, 9.143  
 Hanning tapering, 9.122, 9.125, 9.129  
 Heavyside funktion, 9.10  
 hierarkisk klassifikation, se variansanalyse, hierarkisk klassifikation  
 Hotellings  $T^2$ ,  
 enstikprøvesituation, 6.1 f  
 tostikprøvesituation, 6.8 f, 6.43, 7.16  
 hvid støj, 9.44, 9.159, 9.91  
 høj-pas filter, 9.117, 9.123, 9.124, 9.125, 9.126  
 $\underline{I}$ , 1.11  
 $\underline{I}_n$ , 1.11  
 idempotent afbildning, 1.8  
 idempotent matrix, 1.13, 1.59  
 impulsmoduleret signal, 9.26  
 impulsresponsfunktion, 9.20, 9.118  
 indeterministisk proces, 9.43  
 indre produkt, 1.64 f  
 informationsmatrix, 2.65  
 invers form, 9.155  
 invers matrix, 1.13, 1.19, 1.21, 1.39, 1.64  
 inverst element, 1.2  
 invertibel proces, 9.47, 9.49, 9.51, 9.57  
 isomorfi, 1.10  
 kanonisk korrelationskoefficient, 8.17 f  
 kanonisk variabel, 8.17 f  
 karakteristisk ligning, 9.23, 9.49, 9.52, 9.62  
 klassifikation, se diskrimination  
 knude, 4.70  
 koherensspektrum, 9.139, 9.140  
 kommunalitet, 8.23  
 kommutativ lov, 1.2, 9.16  
 komplement, 1.18  
 konfidensinterval for  
 forudsagt værdi, 3.32 f, 9.163  
 korrelationskoefficient, 2.43  
 partiel korrelationskoefficient, 2.43  
 konfidensområde for  
 middelværdi, 6.6, 6.10  
 spektrum, 9.113  
 konjugerede retninger, 1.67  
 konjugerede vektorer, 1.55  
 konturellipsoide, 2.18 f, 2.21  
 koordinater, 1.6  
 koordinattransformation, 1.13 f  
 koordinattransformationsmatrix, 1.14  
 korrelationskoefficient, 2.27, 2.30  
 korrelationsmatrix, 2.3  
 korrespondanceanalyse, 8.54  
 kovarians, 2.6 f  
 kovariansmatrix, 2.3 f

- Kroneckerprodukt, 1.63  
 krydsamplitudespektrum, 9.138  
 krydsklassifikation, se vari-  
 ansanalyse, krydsklassifi-  
 kation  
 krydskorrelationsfunktion,  
 9.134 f  
 krydskovariansfunktion,  
 9.133 f  
 estimation, 9.136  
 krydsspektralanalyse, 9.132 f  
 krydsspektrum, 9.137 f  
 estimation, 9.141 f  
 kvadratisk form, 1.46, 1.60  
 kvadratisk matrix, 1.11  
 kvadraturspektrum, 9.138  
 kvotienttest, 3.42
- $E$ , 9.15  
 $\Lambda$ , se Wilk's  $\Lambda$   
 lag, 9.15  
 lagvindue, 9.105  
 landmåling, 3.32  
 lav-pas filter, 9.117, 9.123,  
 9.124, 9.125, 9.126  
 linearkombination, 1.4  
 lineær afbildning, 1.8 f,  
 1.12 f  
 lineær afhængighed, 1.5  
 lineær funktionel relation,  
 4.52  
 lineær ligning, løsning af,  
 1.19, 1.27  
 lineær proces, 9.44 f, 9.59 f  
 lineær regressionsanalyse,  
 4.1 f  
 lineær uafhængighed, 1.4  
 lineært bånd, 3.13, 3.25  
 lineært system, tidsinvari-  
 ant, 9.19 f, 9.44, 9.60,  
 9.94, 9.135, 9.140  
 Little Jiffy, 8.50  
 logaritmisk parabolisk trend,  
 9.79  
 logistisk kurve, 4.68  
 logistisk trend, 9.79  
 logit, 4.68
- MA-proces, 9.48 f, 9.61,  
 9.91  
 Mahalanobis afstand, 7.16  
 Mahalanobis afstand, genera-  
 liseret, 7.37  
 matrix, 1.10 f  
 matrix, regulær, 1.13
- matrixprodukt, 1.12  
 matrixsum, 1.11  
 maximum likelihood estima-  
 tion, 2.68  
 MDISC, 7.36 f  
 middelkvadratafvigelsessum,  
 forventet værdi af, 5.10,  
 5.13, 5.15, 5.25, 5.32,  
 5.43, 5.45, 5.48, 5.49,  
 5.51, 5.58 f, 5.60 f, 5.62  
 middelværdi, 2.1 f  
 middelværdifunktion, 9.70  
 modificeret exponentielt trend,  
 9.79  
 Moore-Penrose invers, 1.35  
 multicollinearitet, 4.57  
 multipel korrelationskoeffici-  
 ent, 2.44 f, 4.6
- $N_p(\underline{\mu}, \underline{\Sigma})$ , 2.11  
 $V$ , 9.15  
 $V_s$ , 9.156  
 neutralt element, 1.2  
 norm, 1.65  
 normal fordeling, flerdimensi-  
 onal, 2.10 f  
 normal fordeling, todimensio-  
 nal, 2.27 f  
 normalligninger, 3.5  
 nulrum, 1.9  
 Nyqvist-frekvens, 9.28
- oblimin rotation, 8.55  
 oblique rotation, 8.33  
 opdatering af forudsigelse,  
 9.163  
 ortogonal matrix, 1.37  
 ortogonal regression, 2.53,  
 4.50 f  
 ortogonal transformation, 1.40  
 ortogonale polynomier, 4.14 f,  
 9.76  
 ortogonale vektorer, 1.36,  
 1.65, 1.67  
 ortonormal basis, 1.37
- parabolisk trend, 9.79  
 partiel autokorrelationsfunk-  
 tion, 9.55 f  
 AR-proces, 9.21, 9.165  
 ARMA-proces, 9.59  
 MA-proces, 9.56  
 partiel korrelationskoeffici-  
 ent, 2.35 f, 4.7

- Parzen vindue, 9.8, 9.108,  
 9.113, 9.118, 9.143  
 positiv definit, 1.46 f  
 positiv semidefinit, 1.46 f  
 power spectrum, se spektrum  
 prediktion, 4.27, 4.56, 4.65,  
 se også forudsigelse  
 prediktionsinterval, 3.32 f,  
 9.163  
 prewhitening, 9.67, 9.123  
 principale komponenter, 4.55,  
 8.2 f  
 principale koordinater, 8.54  
 prikprodukt, 1.64 f  
 "probability associated with  
 largest discriminant func-  
 tion", 7.37  
 projektion, 1.8, 1.65  
 præcision, 2.13  
 pseudoinvers afbildning, 1.25 f  
 pseudoinvers matrix, 1.2,  
 1.22 f, 1.59, 1.64  
 pythagoræiske læresætning,  
 1.65
- Q-modus, 1.42, 1.46, 8.51  
 quartimax rotation, 8.28
- $R^2$ , 4.6  
 $R^n$ , 1.3, 1.5, 1.10  
 R-modus, 1.42, 1.46  
 random shock, 9.45  
 random shock form, 9.154  
 rang af afbildning, 1.15  
 rang af matrix, 1.16, 1.43,  
 1.44  
 Rayleigh's kvotient, 1.49  
 regression, 2.49 f  
 regressionsanalyse, 4.1 f  
 efter ortogonale polynomi-  
 er, 4.14 f  
 flerdimensional, 6.15,  
 6.23  
 ikke lineær, 4.66 f
- ortogonal regression, 4.50  
 ridge regression, 4.56 f  
 regressionsligning, valg af  
 bedste, 4.27 f  
 backwards elimination,  
 4.32 f, 4.42  
 forward selection, 4.34 f,  
 4.41
- max  $R^2$  improvement, 4.42  
 samtlige regressioner,  
 4.30 f, 4.42  
 stepwise regression, 4.38 f,  
 4.42  
 regulær matrix, 1.13, 1.16,  
 1.17  
 rektangulært vindue, 9.8,  
 9.108, 9.113, 9.118  
 reproduktivitetssætning for  
 normal fordeling, 2.23  
 Wishart fordeling, 2.61  
 residual, 3.10, 4.9, 9.81 f  
 residualplot, 4.9  
 retlinet trend, 9.79  
 rg, 1.16  
 ridge estimator, 4.59  
 ridge regression, 4.56 f  
 ridge trace, 4.62, 4.64  
 rækkevektor, 1.10
- S, 9.16  
 $S^2$ , forventet værdi, se middel-  
 kvadratafvigelsessum  
 SAK, 3.43, 6.30, 6.33  
 beregning, 3.58, 5.52 f  
 frihedsgrader, 5.56, 5.58  
 forventet værdi, se middel-  
 kvadratafvigelsessum  
 sampling-problem, 9.24 f  
 semidefinit, 1.46  
 sideunderrum, 1.4  
 similaritetsmål, 8.54  
 similære matricer, 1.15  
 singular værdi, 1.43  
 singularværdi dekomposition,  
 1.4 f  
 sinusproces, 9.35 f, 9.96  
 forstyrret, 9.38 f, 9.97  
 skjult, 9.37 f, 9.97  
 skalainvarians, 8.48  
 skalamultiplikation, 1.2, 1.11  
 skalarprodukt, 1.64 f  
 Slutski's sætning, 9.128  
 spaltningsætning, 2.53 f  
 span, 1.4  
 spejling, 1.40  
 spektraldekomposition for  
 matrix, 1.39  
 spektralfordeling, 9.90  
 spektralrepræsentation, 9.44  
 spektraltæthed, 9.88  
 spektralvindue, 9.24, 9.108,  
 9.113

- spektrum, 9.88 f  
 AR-proces, 9.92, 9.97  
 estimation, 9.100 f  
 hvid støj, 9.91  
 integreret, 9.89  
 konfidensområde, 9.113  
 lineær proces, 9.95  
 MA-proces, 9.91  
 udglattet, 9.105  
 spektrum for matrix, 1.39  
 spline funktion, 4.70 f  
 spor af matrix, 1.58 f  
 stabilt system, 9.22  
 stationaritet, 9.32 f, 9.47,  
 9.49, 9.51, 9.57, 9.62  
 stepwise regression, 4.38 f  
 stikprøvespektrum, 9.100  
 stokastisk matrix, 2.1  
 stokastisk proces, 9.29 f  
 fordeling af, 9.30  
 støtte, 2.13  
 successiv testning, 3.48 f  
 summationsoperator, 9.16  
 symmetrisk matrix, 1.11, 1.37  
 sæsondifferensoperator, 9.156  
 sæsonindex, 9.82  
 sæsonvariation, 9.81 f, 9.156 f  
 søjlevektor, 1.10
- $T^2$ , se Hotelling's  $T^2$   
 tabsfunktion, 7.2, 7.24  
 tendens, se trend  
 tensorprodukt, 1.63  
 test for/i  
 bedste diskriminantfunktion,  
 7.19  
 diagonalstruktur i dispersi-  
 onsmatrix, 6.40, 8.17  
 egenværdi i dispersionsma-  
 trix, 8.9  
 egenværdi i korrelations-  
 matrix, 8.9  
 ens dispersionsmatricer,  
 6.43  
 faktormodel, 8.50  
 flerdimensional general  
 lineær model, 6.18 f  
 flerdimensional variansana-  
 lyse, 6.28 f  
 forudsætninger i regr.  
 analyse, 4.8 f  
 general linear model, 3.40 f  
 korrelationskoefficient,  
 2.42 f  
 middelværdi, 6.1 f, 6.8 f,  
 6.43, 7.16  
 multipel korrelationskoeffi-  
 cient, 2.48  
 partiel korrelationskoeffi-  
 cient, 2.42 f  
 proportional dispersion,  
 6.41  
 uafhængighed, 6.40, 8.17  
 yderligere information i  
 diskran., 7.22  
 tidsinvariant lineært system,  
 se lineært system  
 tidsrække, 9.2  
 tidsrækkeanalyse, 9.1 f  
 tog af  $\delta$ -funktioner, 9.12  
 tr, 1.59  
 transponering, 1.10, 1.12,  
 1.64  
 trend, 9.76 f, 9.81
- U, 6.19  
 $U(p, q, r)$ , 6.19  
 uafhængighed, 2.18 f  
 udjævning, 9.67 f  
 ukorreleret, 2.8  
 underrum, 1.4
- variens, biologisk, 5.39  
 variansanalyse, 5.1 f,  
 6.28 f  
 blandet model, 5.23 f  
 ensidet, 5.3 f  
 ensidet flerdimensional,  
 6.28 f, 7.34, 7.38  
 variansanalyse fortsat  
 flere faktorer, 5.52 f  
 hierarkisk klassifikation,  
 5.18 f, 5.24 f, 5.44 f,  
 5.56 f, 5.67  
 krydsklassifikation,  
 5.18 f, 5.31 f, 5.40 f,  
 5.52 f, 5.67  
 robusthed, 5.63 f  
 systematisk model, 5.3,  
 5.7 f, 5.22 f  
 tilfældig model, 5.3, 5.5,  
 5.11 f, 5.22 f  
 tosidet, 5.18 f  
 tosidet flerdimensional,  
 6.30 f  
 trefaktor, 5.39 f  
 type I, se systematisk

- type II, se tilfældig
- variaskomponent, se tilfældig
- vekselvirkning, 5.31, 5.38, 5.53 f
- variaskfunktion, 9.31
- variate difference method, 9.77
- variation
  - inden for grupper, 5.15, 6.29, 7.32
  - mellem grupper, 5.15, 6.29, 7.32
  - mellem hovedgrupper, 5.25
  - mellem undergrupper i.f. hvd.gr., 5.25
  - spaltning af totale, 3.42, 3.57, 4.18, 5.15, 5.25, 5.31, 5.42, 5.43, 5.44, 5.48, 5.49, 5.51, 5.56, 6.29, 6.32
  - systematisk, 5.2
  - tilfældig, 5.3
- variationsbredde, 4.12
- varimax rotation, 8.28, 8.55
- VC, 2.9
- vektoraddition, 1.2
- vektorrum, 1.2 f
- vinkel, 1.67
- vægtet regression, 4.2
- vægtfunktion, 9.20, 9.121
  
- $W(n, \underline{\Sigma})$ , 2.60
- Wilk's  $\Lambda$ , 6.19, 7.34
- Wishart fordeling, 2.59 f
- Wolds dekomposition, 9.44
  
- Yule-Walker ligninger, 9.52

