

2266

Bibel Mag



Technische Hogeschool Eindhoven

## *Onderafdeling der Wiskunde*

### *Multivariate Analyse*

door drs. A. J. Bosch  
en prof. dr. R. Doornbos

TECHNISCHE HOGESCHOOL EINDHOVEN

Afdeling Algemene Wetenschappen

Onderafdeling der Wiskunde

# MULTIVARIATE ANALYSE

door

**Drs. A.J. Bosch**

en

**Prof. Dr. R. Doornbos**

Datering: 70er jaren

# Inhoudsbeschrijving

## MULTIVARIATE ANALYSE

A.J. Bosch en R. Doornbos

Datering: 70er jaren

Onderwerp	blz
§0 Inleiding	0.1
§1 Enkele punten uit de matrixrekening	1.1 - 1.20
§2 Enkele begrippen uit de mathematische statistiek	2.1 - 2.6
§3 De multinormale verdeling	3.1 - 3.6
§4 Correlatie	4.1 - 4.7
§5 Meervoudige lineaire regressie	5.1 - 5.8
§6 De Wishart- en Hotelling's verdeling	6.1 - 6.5
§7 Enkele toetsen	7.1 - 7.8
§8 Componentenanalyse	8.1 - 8.6
§9 Factoranalyse	9.1 - 9.16
§10 Canonieke correlatie	10.1 - 10.5
§11 Discriminant analyse	11.1 - 11.9
§12 Clusteranalyse	12.1 - 12.5
A Normaal verdeelde, doch niet simultaan normaal verdeelde variabelen	A1 - A2
B Differentiatie van scalaire functies van de elementen van een matrix	B1 - B4
C Overzicht van het verband tussen één- en meerdimensionale analyse	C1 - C3
E Het Kronecker- of directproduct van matrices	E1 - E4

JdG, 24 Juli 2005

Inhoud

<u>Inleiding</u>	§ 0 Inleiding
	§ 1 Enkele punten uit de matrixrekening
<u>Basis</u>	§ 2 Enkele begrippen uit de mathematische statistiek
	§ 3 De multinormale verdeling
	§ 4 Correlatie
	§ 5 Meervoudige lineaire regressie
	§ 6 De Wishart- en Hotelling's verdeling
	§ 7 Enkele toetsen
<u>Multivariate</u>	§ 8 Componentenanalyse
<u>Analyse</u>	§ 9 Factoranalyse
	§ 10 Canonieke correlatie
	§ 11 Discriminant analyse
	§ 12 Clusteranalyse
	A Normaal verdeelde, doch niet simultaan normaal verdeelde variabelen
<u>Appendix</u>	B Differentiatie van scalaire functies van de elementen van een matrix
	C Overzicht van het verband tussen één- en meerdimensionale analyse
	E Het Kronecker- of directproduct van matrices.

Literatuur

- [1] Anderson, T.W., An introduction to multivariate statistical analysis, Wiley (1958).
- [2] Dhrymes, Econometrics, statistical foundations and applications, Harper and Row, London (1970).
- [3] Draper, Smith, Applied regression analysis, Wiley 1967).
- [4] Geer, J.P. v.d., Inleiding in de multivariate analyse, Van Loghum Slaterus, Arnhem (1967).
- [5] Graybill, Introduction to matrices with applications in statistics, Wadsworth Publish. Comp., Belmont, Calif. (1969).
- [6] Morrison, D.F., Multivariate Statistical Methods, McGraw-Hill Book-company (1967).
- [7] Rao, C.R., Linear statistical inference and its applications, Wiley (1965).

§ 0. Inleiding

In zekere zin omvat de multivariate analyse praktisch de gehele statistiek. Zelfs bij univariate problemen hebben we te maken met een steekproef en een steekproef is een speciaal geval van een vektor van stochastische variabelen. De term "multivariate analyse" wordt echter in engere zin gebruikt. Volgens M.G. Kendall: "Multivariate analysis is that branch of statistical analysis which is concerned with the relationships of sets of dependent variates".

In de multivariate analyse komen meestal de volgende onderwerpen ter sprake: de multinormale verdeling, enkelvoudige, partiële en multiple correlatie, faktor- en komponentenanalyse, kanonieke correlatie, discriminantanalyse, klassifikatie, clusteranalyse, toetsen van hypothesen.

Eigenlijk vallen ook onder de multivariate analyse:

regressie-, variantie- en covariantie-analyse, experimental design (statistische theorie van proefopzetten), tijdreeksen.

Doch deze onderwerpen zijn zo omvangrijk dat ze apart worden behandeld.

De multivariate analyse wordt zeer veel toegepast, o.a. in de psychologie en andere sociale wetenschappen, doch ook steeds meer op vele andere terreinen zoals econometrie, natuurkunde, biologie en landbouw. Eigenlijk overal waar men te maken heeft met zeer vele afhankelijke variabelen.

Voor het bestuderen van dit dictaat, kan men de volgende punten overslaan:

Uit §1 : 1.36; 1.38; 1.42; 1.43; 1.47; 1.55; 1.59; 1.60; 1.61; 1.74 t/m 1.83;  
1.96;

uit §2 : 2.17; 2.18; 2.45; 2.46.

§ 5 geheel; appendix E geheel; appendix C eventueel.

## § 1. Enkele punten uit de matrixrekening

Voor de behandeling van de multivariate analyse, lineaire modellen e.d. is de theorie der lineaire vectorruimten en de matrixrekening van fundamenteel belang. Met behulp van matrices worden de notaties veel overzichtelijker en voor vele problemen geeft een vectormodel een beter inzicht.

Vooral de orthogonale projectie blijkt een belangrijke rol te spelen.

### a. Definities, notaties

1.1.  $A_{n \times p} = \{a_{ij}\}$  is een *matrix* van de *orde*  $n \times p$  met  $n$  rijen,  $p$  kolommen en elementen  $a_{ij}$ ;  $(A)_{ij} = a_{ij}$ .

$A_{n \times n}$  is een *vierkante* matrix;  $A_{n \times n}$  is een *diagonaal*matrix als  $a_{ij} = 0$  voor  $i \neq j$ .

$A_{n \times n} = \text{diag}(a_{11}, \dots, a_{nn})$  d.w.z.  $A$  is een diagonaalmatrix met diagonaal-elementen  $a_{11}, \dots, a_{nn}$ . Zo ook  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ .

$I_{n \times n} =: I_n = \{\delta_{ij}\}$  is de *eenheidsmatrix*,  $\delta_{ij}$  het Kronecker-symbool.

$A = \alpha I$  is een *scalaire* matrix.

$A'$  is de *getransponeerde* matrix van  $A$  d.w.z.  $(A')_{ij} = a_{ji}$  voor alle  $i, j$ .

In het bijzonder is  $a'$  een *rijvector* en  $a$  de bijbehorende *kolomvector*.

$A_{n \times n}$  heet *symmetrisch* als  $A' = A$  (ook wel  $A^T$  i.p.v.  $A'$  geschreven).

$A_{n \times n}$  heet *antisymmetrisch* als  $A' = -A$ .

$u = (1, \dots, 1)'$ ;  $U = \{1\}$  (in de literatuur vaak  $J$ , matrix van enen).

$a_{i*}$  is de  $i$ -de rij van  $A$  als kolom geschreven,

$a'_{i*}$  is de  $i$ -de rij van  $A$  als rij geschreven,

$a_{*j}$  is de  $j$ -de kolom van  $A$  als kolom geschreven,

$a'_{*j}$  is de  $j$ -de kolom van  $A$  als rij geschreven.

Zo is  $(AB)_{ij} = a'_{i*} b_{*j}$ .

$A_{p \times q} = (a_{*1} \dots a_{*q})$ ;  $I_p = (\delta_{*1} \dots \delta_{*p})$ .

$\tilde{I}_p := (\delta_{*i_1} \dots \delta_{*i_p})$  d.w.z. de kolommen van  $I_p$  in andere volgorde.

Deze matrix heet een *permutatiematrix*.

Opmerking: Daar we gericht zijn op de statistiek, worden alle gegeven grootheden reëel ondersteld. Ook zullen we geen onderscheid maken in notatie tussen een lineaire afbeelding  $\mathcal{A}$  en de bijbehorende matrix  $A$ , daar we steeds met één basis werken die bovendien orthonormaal ondersteld is.

1.2.  $|A| = \det A$  is de *determinant* van  $A$ ;  $\text{abs}|A|$  is de absolute waarde van  $\det A$ .  
 $|A|_{ij}$  is de *cofactor* van  $a_{ij}$ , dat is  $(-1)^{i+j}$   $\times$  de onderdeterminant van  $a_{ij}$ .  
 $A^{-1}$  is de *inverse* van  $A_{n \times n}$  zo deze bestaat.  
 $(A)^{ij} := (A^{-1})_{ij}$ , ook wel als  $a^{ij}$  genoteerd.

1.3.  $\text{sp } A_{n \times n} := \sum_1^n a_{ii}$  heet het *spoor* van  $A$  (Engels: trace).

1.4.  $r(A)$  is de *rang* van  $A$ , dat is het maximum aantal onafhankelijke rijen van  $A$ .

$A_{n \times p}$  heet van *volle rang* als  $r(A) = \min(n, p)$ .

$A_{n \times p}$  heet van *volle rijrang* als  $r(A) = n$  (dus  $n \leq p$ ).

$A_{n \times p}$  heet van *volle kolomrang* als  $r(A) = p$  (dus  $p \leq n$ ).

$A \in \mathcal{M}_{p,q}^r$  d.w.z.  $A$  is een matrix van de orde  $p \times q$  met rang  $r$ .

1.5. Het *direct- of Kronecker product* van twee willekeurige matrices  $A$  en  $B$  is  $A \otimes B := \{a_{ij} B\}$ , d.w.z. elk element van  $A$  wordt met  $B$  vermenigvuldigd.

1.6.  $\vec{A} := (a'_{*1} \dots a'_{*q})'$ , d.w.z. de kolommen van  $A$  achtereenvolgens onder elkaar genoteerd (ook wel *vec*  $A$  genoemd). Als  $A$  van de orde  $p \times q$  is, is  $\vec{A}$  dus een vector met  $pq$  componenten.

1.7. We definiëren voor elke matrix  $A$  een "*pseudokwadraat*", nl.  $A^2 := A'A$ .  
Analoog  $a^2 := a'a$ ;  $A^{-2} := (A^2)^{-1}$ .

Als het geen verwarring schept, schrijven we vaak  $A^2$  i.p.v.  $A^2$  (bijv. als  $A$  niet vierkant is bestaat het gewone kwadraat  $A^2$  niet).

In de multivariate-statistiek komt het pseudodkwadraat echter veel vaker voor, het gewone vrijwel niet. Zoals we zullen zien biedt het nog meer voordelen.

1.8. Het *inwendig product* van de vectoren  $a$  en  $b$  wordt genoteerd als  $(a,b)$ . Daar we steeds (vanaf nu stilzwijgend) een orthonormale basis veronderstellen, is  $(a,b) = a'b = b'a = \sum_i a_i b_i$ . Opmerking: vectoren onderstrepen we niet. Analoog voor matrices  $A$  en  $B$  van dezelfde orde:  
 $(A,B) := \text{sp}(A'B)$ . Eenvoudig is na te gaan dat dit inderdaad een inproduct definiëert.

1.9. Via het inproduct worden gedefiniëerd:

De *norm* van een vector  $a$  is  $\|a\| := \sqrt{(a,a)} = \sqrt{a^2}$ ; analoog

de *norm* van een matrix  $A$  is  $\|A\| := \sqrt{(A,A)} = \sqrt{\text{sp}A^2}$ . (Euclid.- of Frobeniusnorm)

De *hoek*  $\psi$  tussen 2 vectoren  $a$  en  $b$  (beide  $\neq 0$ ):  $\cos \psi := \frac{(a,b)}{\|a\| \|b\|}$ ;

$a$  en  $b$  heten *orthogonaal*, notatie  $a \perp b$ , als  $(a,b) = 0$ .

Een reële vectorruimte met inproduct wordt *Euclidische ruimte*  $E_n$  genoemd.

$\lim_{n \rightarrow \infty} A_n = A$ , d.w.z.  $\|A_n - A\| < \epsilon$  voor  $n > N$ .

Zij  $a \in V$ ,  $B \subset V$ ;  $d[a,B] := \inf_b \{\|a-b\|; b \in B\}$  heet de *afstand* van  $a$  tot  $B$ .

Opmerking: dat deze norm gekozen is hangt samen met een metriek, waarbij de methode der kleinste kwadraten wordt toegepast. Er zijn andere normen mogelijk.

1.10.  $A_{n \times n}$  heet *regulier* als  $|A| \neq 0$ , anders *singulier*.

1.11.  $A_{n \times n}$  heet *idempotent* als  $AA = A$  (dus ook als  $A^2 = A$ ).

1.12. Een reguliere matrix  $A$  heet *orthogonaal* als  $A^2 = I$  oftewel  $A' = A^{-1}$ .

1.13.  $x'Ax$  is een *kwadratische vorm* in  $x$ . Hierbij wordt  $A$  symmetrisch ondersteld (zie 1.56).

1.14. Een matrix  $A$  heet *positief definitief* (afgekort p.d. of  $A > 0$ ) als  $A$  symmetrisch is en  $\forall x \neq 0 (Ax,x) = x'Ax > 0$ .

Is  $x'Ax \geq 0$  voor alle  $x$  en  $\exists x \neq 0$  met  $x'Ax = 0$ , dan heet  $A$  *semipositief definitief*. Is  $x'Ax \geq 0$  voor alle  $x$ , dan is  $A$  *niet negatief definitief* ( $A \geq 0$ ).

$A > B$  d.w.z.  $A - B > 0$ .



- 1.15. Zij  $V$  een reële vectorruimte. Als  $\mathcal{A}x = \lambda x$  voor  $x \neq 0 \in V$  en  $\lambda$  reëel, dan heet  $x$  *eigenvector* en  $\lambda$  de bijbehorende *eigenwaarde* van de afbeelding  $\mathcal{A}$ ; de eigenwaarden van  $\mathcal{A}$  zijn de reële wortels van de *karakteristieke vergelijking*  $|A - \lambda I| = 0$ . De eigenwaarden van de matrix  $A$  zijn alle wortels, die dus complex kunnen zijn.
- 1.16. Een matrix  $A$  heet *defect* als er géén reguliere matrix  $B$  bestaat zodat  $B^{-1}AB = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ ;  $B$  en  $\Lambda$  mogen complexe elementen bevatten. Zijn  $B$  en  $\Lambda$  reëel, dan heet  $A$  *reëel diagonaliseerbaar*.
- 1.17. Het stelsel lineaire vergelijkingen  $Ax = b$  heet *consistent* als er minstens een oplossing is.
- 1.18.  $N(A) := \{x \in V \mid Ax = 0\}$  heet de *nulruimte* of *kern* van  $A$ .
- 1.19.  $\langle A \rangle = \langle a_{*1} \dots a_{*q} \rangle$  is de *kolommenruimte* van  $A_{p \times q}$ , dat is de lineaire deelruimte in  $\mathbb{R}^p$  opgespannen door de kolommen van  $A$ .  
Analoog is  $\langle A' \rangle$  de *rijenruimte* van  $A$ , dat is de lineaire deelruimte in  $\mathbb{R}^q$  opgespannen door de rijen van  $A$ .
- 1.20.  $E_\lambda(\mathcal{A}) := \{x \in V \mid \mathcal{A}x = \lambda x\}$  is de *eigenruimte* van  $\mathcal{A}$  behorende bij de eigenwaarde  $\lambda$ ;  $\dim E_\lambda$  heet de *meetkundige multipliciteit* van  $\lambda$ .
- 1.21. Zij  $D \subset V$  een lineaire deelruimte van  $V$  en  $a \in V$ .  
 $S = a + D$  heet een *lineaire variëteit* en  $\dim S := \dim D$ .
- 1.22. Zij  $D$  een lineaire deelruimte van de Euclidische ruimte  $E_n$ , dan is  $D^\perp := \{x \in E_n \mid x \perp D\}$  het *orthogonale complement* (*orthoplement*) van  $D$  in  $E_n$ .
- 1.23. Het *spectrum*  $\sigma(A)$  van  $A$  is de verzameling van alle eigenwaarden van  $A$ .

b. Eigenschappen

1.30. Een belangrijk hulpmiddel is het *partitioneren* van matrices:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} E & F \\ G & H \end{pmatrix} = \begin{pmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{pmatrix}.$$

De afmetingen zijn zodanig dat alle produkten bestaan.

In het bijzonder is:  $|AB| = \sum_j a_{*j} b'_{j*}$ ;  $Ax = b$  geeft  $\sum_j a_{*j} x_j = b$ .

$$A'A = A^2 = \sum_1^n a_{i*} a'_{i*}. \quad (A \in \mathbb{M}_{p,n})$$

$$\text{Analoog: } AA' = (A')^2 = \sum_1^p a_{*j} a'_{*j}.$$

1.31. In het algemeen is  $AB \neq BA$  (dus ook  $A'A \neq AA'$ ). Is echter  $B = \beta I$ , dan is  $AB = BA$ ; tevens is  $B' = B$ .

1.32.  $sp$  is een *lineaire operator*, nl.  $sp(\alpha A + \beta B) = \alpha sp A + \beta sp B$ .  
Tevens geldt  $sp(AB) = sp(BA)$ . Bewijs:

$$sp(AB) = \sum_i (AB)_{ii} = \sum_{i,k} a_{ik} b_{ki} = \sum_{k,i} b_{ki} a_{ik} = \sum_k (BA)_{kk} = sp(BA).$$

In het bijzonder geldt:  $sp(ABC) = sp(CAB) = sp(BCA)$

$$sp(ab') = sp(b'a) = b'a; \quad sp A = sp A';$$

$$\|A\|^2 = sp A^2 = sp \sum_i a_{i*} a'_{i*} = sp \sum_i a'_{i*} a_{i*} = \sum_i a_{i*}^2 = \sum_{ij} a_{ij}^2.$$

Oftewel:

De norm van een matrix is de wortel uit de som van de kwadraten der elementen. Bovendien is  $sp A = \sum_i \lambda_i$  (som der eigenwaarden van A).

1.33.  $(AB)^{-1} = B^{-1} A^{-1}$  (A, B n x n matrices en regulier)

$$|A| = \sum_i a_{ij} |A|_{ij} = \sum_j a_{ij} |A|_{ij}; \quad |A^{-1}| = 1/|A|; \quad |AB| = |A||B|;$$

$$a^{ij} = (A^{-1})_{ij} = |A|_{ji} / |A|.$$

Bovendien is  $|A| = \prod \lambda_i$  (product der eigenwaarden van A).

Daar  $|A - \lambda I| = \prod (\lambda_i - \lambda)$  is, neem  $\lambda = 1$ :  $|A - I| = \prod (\lambda_i - 1) = |A - I|$ .

1.34.  $\begin{vmatrix} A & B \\ 0 & C \end{vmatrix} = |A||C|$  (A, C vierkante matrices) nl.

$$\begin{vmatrix} A & B \\ 0 & C \end{vmatrix} = \begin{vmatrix} I & 0 \\ 0 & C \end{vmatrix} \begin{vmatrix} A & B \\ 0 & I \end{vmatrix} = |C||A|.$$

1.35. Is  $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$  ( $A_{11}, A_{22}$  vierkant,  $A_{22}$  regulier), dan geldt:

$$|A| = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \begin{vmatrix} I_{k \times k} & 0 \\ -A_{22}^{-1} A_{21} & I_{r \times r} \end{vmatrix} = \begin{vmatrix} A_{11 \cdot 2} & A_{12} \\ 0 & A_{22} \end{vmatrix} = |A_{11 \cdot 2}| |A_{22}|$$

waarin  $A_{11 \cdot 2} := A_{11} - A_{12} A_{22}^{-1} A_{21}$ .

Analoog geldt  $|A| = |A_{22 \cdot 1}| |A_{11}|$  (nu alleen  $A_{11}$  regulier ondersteld).

Deze enigszins vreemde eigenschap wordt gebruikt bij het afleiden van toetsingsgrootheden via het likelihood-ratio-criterium.

1.36. Zij  $A \in \mathcal{M}_{p,q}^r$  en gepartitioneerd als in 1.35 met  $r(A_{11}) = r < \min(p,q)$ .  
Dan is  $A_{22 \cdot 1} = 0$ .

Bewijs:

Daar  $r(A) = r(A_{11})$  is  $(A_{21} A_{22}) = K(A_{11} A_{12})$ ; oftewel  $A_{21} = K A_{11}$  dus

$K = A_{21} A_{11}^{-1}$  en  $A_{22} = K A_{12} = A_{21} A_{11}^{-1} A_{12}$  (dus  $A_{22 \cdot 1} = 0$ ).

1.37. Is A orthogonaal, dan is  $(Ax, Ay) = (x, y)$  nl.  $(Ax, Ay) = x' A' Ay = x' y = (x, y)$ .

In het bijzonder is  $\|Ax\| = \|x\|$ .

$A^{-1}$  is eveneens orthogonaal daar  $(A^{-1})' = (A')^{-1} = (A^{-1})^{-1}$  (zie 1.51).

1.38.  $\tilde{I}_p^2 = I_p$ , dus orthogonaal, nl.  $(\tilde{I}^2)_{ij} = \delta_{*i_i} \delta_{*i_j} = \delta_{ij}$ .

Zo is  $A \tilde{I}_q$  een matrix met dezelfde kolommen als van A, alleen gepermuteerd.  
En  $\tilde{I}_p A$  " " " " rijen " " " " " " " "

Wil men de kolommen van A als volgt hebben:  $\tilde{A} = (a_{*3} \ a_{*4} \ a_{*1} \ a_{*2})$  dan moet men navermenigvuldigen met  $\tilde{I}_4 = (\delta_{*3} \ \delta_{*4} \ \delta_{*1} \ \delta_{*2})$ .

1.39. Zij  $A \in L(V,W)$ , d.w.z. een lineaire afbeelding van  $V$  in  $W$ , dan geldt:

$$\dim N(A) + r(A) = \dim V.$$

Hieruit volgt:

Zij  $A \in \mathcal{M}_{p,q}^r$  en  $Ax = b$  consistent, dan is dim. oplossingsvariëteit  $S$

$$\dim S = q - r. \quad (\text{Zie Wsk 20.})$$

1.40. Zij  $D$  een lineaire deelruimte  $\subset E_n$ , dan geldt:

$$\dim D + \dim D^\perp = n \quad (\text{Zie Wsk 30.})$$

1.41.  $N(A') = \langle A \rangle^\perp$ . (Zie Wsk 30.)

1.42. Zij  $A \in \mathcal{M}_{p,q}^{q-m} \Rightarrow \exists D \in \mathcal{M}_{q,m}^m$  met  $AD = 0$  nl.

$\dim N(A) = m$  dus  $N(A) = \langle d_{*1} \dots d_{*m} \rangle$  en  $A(d_{*1} \dots d_{*m}) =: AD = 0$ .

Tevens geldt:

Stel  $B \in \mathcal{M}_{m,q}^m$ ;  $BD$  regulier  $\Leftrightarrow \langle B' \rangle \cap \langle A' \rangle = \{0\}$ ,

oftewel rijen van  $B$  onafhankelijk rijen van  $A$ .

Bewijs:

$\Rightarrow$  Stel  $\exists b_{i*} \in \langle A' \rangle$ , d.w.z.  $\exists t$  met  $A't = b_{i*}$  dus  $D'b_{i*} = D'A't = 0$ .

Nu is  $BD$ , dus  $D'B'$  regulier en  $D'b_{i*} = (D'B')_{*i} = 0$  strijdig.

$\Leftarrow$  Nu is  $\begin{pmatrix} A \\ B \end{pmatrix} \in \mathcal{M}_{p+m,q}^q$ . Volgens 1.80 is er een linkerinverse (VW)

zodat  $VA + WB = I_q$ . Dus  $VAD + WBD = D$  oftewel  $WBD = D$ .

Analoog is er een linkerinverse  $E$  zodat  $ED = EWBD = I_m$ .

Nu is  $m = r(EWBD) \leq r(BD) \leq m$  dus  $r(BD) = m$  oftewel

$BD$  regulier.

1.43. Zij  $y_{q \times 1} \neq 0$ ,  $z_{p \times 1}$  gegeven vectoren. Dan is er een  $X_{p \times q}$  met  $Xy = z$ ,  
nl. Stel  $y_k \neq 0$ . Neem  $x_{*j} = 0$  voor  $j \neq k$  en  $x_{*k} = z/y_k$ , dan is

$$Xy = \sum_j x_{*j} y_j = z.$$

De rang van een matrix

Nu volgen enkele eigenschappen betreffende de rang van een produkt van matrices. Dit is belangrijk o.a. om te kijken of een produkt een inverse heeft en om de dimensie van de beeldruimte te bepalen.

1.44. a)  $r(A') = r(A)$  (zie Wiskunde 10).

b)  $r(AB) \leq \min[r(A), r(B)]$ , nl.  $r(AB) = r(Ab_{*1}, \dots, Ab_{*s}) = r(\sum a_{*i} b_{i1}, \dots, \sum a_{*i} b_{is}) \leq r(A)$ .

Analoog:  $r(AB) = r(B'A') \leq r(B') = r(B)$ . Dus  $r(ab') = 1$  als  $a, b \neq 0$ .

c)  $A \in \mathcal{M}_{p,q}^q$ :  $r(AB) = r(B)$ , nl.:  $r(B) \geq r(AB) = r(B'A') \geq r[B'A'A(A'A)^{-1}] = r(B') = r(B)$ ;  $A'A$  is regulier volgens 1.45 nl.  $r(A'A) = r(A) = q$ .

$B \in \mathcal{M}_{q,s}^q$ :  $r(AB) = r(A)$ , nl.:  $r(AB) = r(B'A') = r(A') = r(A)$ .

Bijzonder geval: de rang verandert niet door voor- of navermenigvuldiging met reguliere matrix.

1.45.  $r(A^2) = r(A)$  (analoog  $r(AA') = r(A)$ ; in het algemeen is  $r(AB) \neq r(BA)$ ).

Bewijs:

$Ax = 0$  geeft  $A^2x = 0$  dus  $N(A) \subset N(A^2)$ ;  $A^2x = 0$  geeft  $x'A'Ax = (Ax)^2 = 0$  dus  $Ax = 0$  oftewel  $N(A^2) \subset N(A)$ ;  $\dim N(A^2) = \dim N(A)$ , dus  $r(A^2) = r(A)$ .

1.46. Zij  $r(A) = r$ , dan bevat  $A$  een reguliere  $r \times r$  ondermatrix (zie Wsk 10).

1.47. Als  $ABA = A \Rightarrow r(AB) = r(BA) = r(A)$ .

Bewijs:

$r(A) \leq r(AB)$  en  $r(AB) \leq r(A)$  dus  $r(AB) = r(A)$ .

1.48.  $\langle AB \rangle \subset \langle A \rangle$ , nl.  $\langle AB \rangle = \langle Ab_{*1} \dots Ab_{*q} \rangle \subset \langle A \rangle$ .

Is  $B$  regulier of van volle rijrang (zie 1.44) dan is  $\langle AB \rangle = \langle A \rangle$ .

Verder is  $\langle X'X \rangle = \langle X' \rangle$  (met 1.45).

1.49.  $Ax = b$  consistent  $\Leftrightarrow b \in \langle A \rangle$ . Ga zelf na (Wsk 10).

Gevolg: Is  $A \in \mathcal{M}_{pq}^p$  dan is  $Ax = b$  consistent voor alle  $b_{p \times 1}$ .

1.50.  $X'Xb = X'y$  is consistent daar  $X'y \in \langle X' \rangle = \langle X'X \rangle$ .

Symmetrische en positief definitie matrices

Nu volgen enkele eigenschappen betreffende symmetrische en positief definitie matrices. Deze spelen een belangrijke rol o.a. omdat de variantie-covariantie matrix (semi)positief definit is.

1.51.  $(A+B)' = A' + B'$ ;  $(AB)' = B'A'$ .

Bewijs:

$$(AB)'_{ij} = (AB)_{ji} = \sum_k a_{jk} b_{ki} = \sum_k b'_{ik} a'_{kj} = (B'A')_{ij}.$$

$$(A^{-1})' = (A')^{-1} \text{ nl. } AA^{-1} = I \text{ geeft } (A^{-1})'A' = I' = I.$$

Gevolg: A regulier, symmetrisch  $\Leftrightarrow A^{-1}$  regulier, symmetrisch.

$$\text{Is } A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \text{ dan is } A' = \begin{pmatrix} A'_{11} & A'_{21} \\ A'_{12} & A'_{22} \end{pmatrix}.$$

1.52.  $\forall x, y (Ax, y) = (x, A'y)$ , nl.  $(Ax, y) = x'A'y = (x, A'y)$ .

Is A symmetrisch, dan geldt  $(Ax, y) = (x, Ay)$ .

Opmerking: Dit is eigenlijk de definitie van een symmetrische afbeelding. Op een orthonormale basis behoort bij een symmetrische afbeelding een symmetrische matrix en omgekeerd. Ditzelfde geldt voor een orthogonale afbeelding en een orthogonale matrix.

1.53. Is A symmetrisch en  $\lambda_1 \neq \lambda_2$ , dan zijn de bijbehorende eigenvectoren  $x_1$  en  $x_2$  orthogonaal, nl.  $\lambda_1(x_1, x_2) = (Ax_1, x_2) = (x_1, Ax_2) = \lambda_2(x_1, x_2)$  dus  $(x_1, x_2) = 0$ .

1.54. Uit 1.16 volgt voor een niet defecte matrix A : er is een reguliere matrix B zodat  $AB = BA$  oftewel  $A = BAB^{-1}$ ; B bestaat uit eigenvectoren en  $A = \text{diag}(\lambda_1, \dots, \lambda_p)$  uit de bijbehorende eigenwaarden van A. Nu geldt: een symmetrische matrix A is reëel diagonaliseerbaar (zie Wsk 30). Met 1.53 kan men er voor zorgen dat B orthogonaal is, zodat

$$A = HAH'$$

(H nu een orthogonale matrix van eigenvectoren en A de diagonaalmatrix van eigenwaarden; in het vervolg zullen H en A steeds deze betekenis hebben).

1.55. Zij  $A \in \mathcal{M}_{p,p}^{p-m}$ , symmetrisch en  $0 < m < p$ . Dan is er een  $B \in \mathcal{M}_{m,p}^m$  zodat

$\begin{pmatrix} A & B' \\ B & 0 \end{pmatrix}$  regulier is.

Bewijs:

$r(A) = p-m$ , dus  $\langle A^T \rangle = \langle A \rangle$  aan te vullen met  $m$  onafhankelijke rijen zodat  $r\begin{pmatrix} A \\ B \end{pmatrix} = p$ . Oftewel voor alle  $i$  geldt  $b_{i*} \notin \langle A \rangle$  dus ook

$\begin{pmatrix} b_{i*} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \notin \langle \begin{pmatrix} A \\ B \end{pmatrix} \rangle$  oftewel  $r\begin{pmatrix} A & B' \\ B & 0 \end{pmatrix} = p+m$  dus regulier.

1.56.  $f(x) = x'Bx = \frac{1}{2}(x'Bx + x'B'x) = x'\left(\frac{B+B'}{2}\right)x =: x'Ax = f(x)$  nu echter met  $A$  symmetrisch. Vandaar dat in een kwadratische vorm de matrix zonder beperking symmetrisch ondersteld kan worden, deze is bovendien uniek (met 1.57).

1.57.  $\forall x, y : x'Ay = 0 \iff A = 0$ , nl.  $\delta'_{*i} A \delta_{*j} = a_{ij} = 0 \forall ij$ .

$\forall x : x'Ax = 0 \iff A = -A'$  ( $A$  nu niet symmetrisch ondersteld!).

Bewijs:  $\iff x'Ax = \frac{1}{2}x'(A+A')x = 0$

$\implies \delta'_{*i} A \delta_{*i} = a_{ii} = 0$ . Nu  $(0, \dots, 1, \dots, 1, \dots, 0) A \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = a_{ij} + a_{ji} = 0$

geeft  $a_{ij} = -a_{ji}$ .

Opmerking: is  $A$  symmetrisch dan volgt dat  $A = 0$ .

1.58.  $A^2 = 0 \iff A = 0$ , nl.  $r(A^2) = r(A) = 0$  dus  $A = 0$ ,  
in het bijzonder  $\|A\| = 0 \iff A = 0$ .

1.59.  $A^2 = A^2 \iff A = A'$ , nl.

$\|A'-A\|^2 = \text{sp}(A-A')(A'-A) = \text{sp}[AA'-AA-A'A'+A'A] = 2\text{sp}(A'A) - 2\text{sp}(AA) = 0$   
daar  $\text{sp}(A^2) = \text{sp}(A'^2)$ , dus met (1.58)  $A = A'$ .

1.60.  $AX'X = BX'X \iff AX' = BX'$ , nl.

$(A-B)X'X = 0$ ;  $\langle (A-B)' \rangle \subset \langle X'X \rangle^\perp = \langle X' \rangle^\perp$  dus  $(A-B)X' = 0$ . Ander bewijs:  
 $0 = (A-B)X'X(A-B)' = [X(A-B)']^2$  dus  $X(A-B)' = (A-B)X' = 0$  (1.58).

Opmerking:  $A \in \mathcal{M}_{pq}^q$  (dus van volle kolomrang) dan  $AB = AC \Leftrightarrow B = C$ , nl.

$A'AB = A'AC$  en daar  $A'A$  nu regulier is (met 1.45) volgt  $B = C$ .

Analoog  $A \in \mathcal{M}_{pq}^p$  (dus van volle rijrang) dan  $BA = CA \Leftrightarrow B = C$ .

1.61.  $A, B \geq 0 \Rightarrow A+B \geq 0$ , nl.  $((A+B)x, x) = (Ax, x) + (Bx, x) \geq 0$ . Geldt bovendien:  
 $N(A) \cap N(B) = \{0\}$  dan is  $A+B > 0$ ; in het bijzonder is  $A+I > 0$ .

1.62.  $A'B = 0 \Rightarrow (A+B)^2 = A^2 + B^2$  (niet omgekeerd), nl.

$$(A+B)'(A+B) = (A'+B')(A+B) = A'A + B'B + A'B + B'A = A^2 + B^2.$$

1.63. Zij  $A$  symmetrisch, dan geldt  $A > 0 \Leftrightarrow \forall i \lambda_i > 0$ .

Bewijs: Volgt direct uit 1.66.

1.64.  $A \geq 0 \Leftrightarrow A = D^2$ .

Bewijs  $\Rightarrow$  : neem  $D = \Lambda^{\frac{1}{2}}H'$ , dan is  $D^2 = H\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}H' = H\Lambda H' = A$ .

Bewijs  $\Leftarrow$  :  $(Ax, x) = (D^2x, x) = (Dx, Dx) \geq 0$ .

Opmerking : Is  $A > 0$  dan is  $D$  van volle kolomrang en omgekeerd.

In de gebruikelijke notatie  $D^2 = DD$  is dit niet waar. Neem bijv.  $D = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$

$DD = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ , sp  $DD < 0$  oftewel met (1.63) is  $DD$  niet  $> 0$ .

Ook geldt  $A = BB$ , nl. met  $B := H\Lambda^{\frac{1}{2}}H'$ ;  $B := A^{\frac{1}{2}}$ .

Is  $A > 0$  dan is er zelfs een bovendriehoeksmatrix  $D = \begin{pmatrix} \times & & \\ & \times & \\ 0 & & \times \end{pmatrix}$  met  $A = D^2$ .

Deze is te vinden door een *Gram-Schmidt orthogonalisatieproces*.

$D$  is uniek als men eist  $d_{ii} \geq 0$  voor alle  $i$  (*Choleski-procedure*).

1.65.  $A > 0 \Leftrightarrow A^{-1} > 0$ .

Bewijs: volgt direct uit 1.66.



1.66.  $B \in \mathbb{M}_{q,p}^q : A_{p \times p} > 0 \Rightarrow BAB' > 0$ .

Bewijs:  $BAB'$  is symmetrisch en  $(BAB'x, x) = (AB'x, B'x) = (Ay, y) > 0$  daar  $y \neq 0$  als  $x \neq 0$ .

Opmerking:

- a) Is slechts gegeven dat A regulier is, dan kan  $BAB'$  wel singulier zijn.
- b) Is B regulier dan geldt:  $A > 0 \Leftrightarrow BAB' > 0$ .
- c) Daar  $A = H\Lambda H'$  geldt dus ook:  $A > 0 \Leftrightarrow \Lambda > 0$  (dus alle  $\lambda_i > 0$ ).
- d) Daar  $|A| = \prod \lambda_i$ , is A regulier.
- e)  $A > 0 \Leftrightarrow \Lambda > 0 \Leftrightarrow \Lambda^{-1} > 0 \Leftrightarrow A^{-1} > 0$  ( $A^{-1}$  is ook symmetrisch en bestaat).

1.67.  $A > 0 \Leftrightarrow$  elke hoofdondermatrix  $A_{ij} > 0$ . Stel  $A_{p \times p} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ .

Zij  $A_{11}$  van de orde  $p_1 \times p_1$ ,  $A_{21}$  van de orde  $p_2 \times p_1$ , dus  $p_1 + p_2 = p$ .

Neem  $x' = (x'_1, o')$  waarin  $x_1$  een vector met  $p_1$ ,  $o$  een vector met  $p_2$  componenten is. Voor alle  $x_1 \neq o$  is  $x'_1 A_{11} x_1 = x'Ax > 0$  dus met A ook  $A_{11} > 0$ .

Een willekeurige hoofdondermatrix B is altijd linksboven te plaatsen, nl  $\tilde{A} = \tilde{I}'A\tilde{I}$ . Daar I regulier is, is (met 1.66)  $\tilde{A}$  ook  $> 0$ .

1.68. A, B symmetrisch en  $A > 0$  dan is AB reëel diagonaliseerbaar (dus ook BA).

Bewijs:

$A = D^2$  (1.64) dus  $AB = D'DB = D'DBD'D^{-T}$  (A, dus D regulier).

Nu is  $DBD'$  symmetrisch dus  $= H\Lambda H^{-1}$  oftewel  $AB = D'H\Lambda H^{-1}D^{-T} =: CAC^{-1}$  met  $C := D'H$ .

Dit is niet juist als A s.p.d. is!

1.69. Is  $A \neq 0$  reëel diagonaliseerbaar dan geldt  $sp(AA) > 0$ .

Bewijs:

$A = BAB^{-1}$ ;  $sp(AA) = sp(BAB^{-1}BAB^{-1}) = sp(\Lambda^2) > 0$ .

Opmerking: is A defect dan is dit niet waar, voorbeeld:

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}; AA = 0 \text{ en } sp(AA) = 0.$$

1.70. Is A niet defect en  $\lambda_i = \lambda$  voor alle i, dan is  $A = \lambda I$ , nl.  $A = BAB^{-1} = \Lambda = \lambda I$ . Deze triviale uitspraak biedt de mogelijkheid defecte matrices te construeren, voorbeeld:

$$\begin{pmatrix} 3 & 0 & 1 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

1.71. Is A niet defect dan is  $r(A) = \sum m_i$ , waarin  $m_i$  de algebraïsche multipliciteit is van  $\lambda_i \neq 0$ ; nl.  $r(A) = r(BAB^{-1}) = r(A) = \sum m_i$ .

1.72. Gegeven  $A_{p \times q}$  met  $p \leq q$ . Dan zijn de eigenwaarden  $\neq 0$  van  $A'A$  dezelfde als die van  $AA'$ .

Bewijs:

Stel  $A'Ax = \lambda x$  en  $\lambda \neq 0$ , dus ook  $(AA')Ax = \lambda Ax$  oftewel  $\lambda$  eigenwaarde van  $AA'$  (nl.  $Ax \neq 0$ ). Analoog als  $\mu$  eigenwaarde is van  $AA'$ , dan ook van  $A'A$ .

1.73. Stel  $A, B > 0$  en van dezelfde orde. Dan is er een reguliere D zodat  $A = D^2$  en  $B = D'AD$  (A diagonaal met alle  $\lambda_i > 0$ ).

Bewijs:

$A = E^2$  (zie 1.64). Beschouw  $E^{-T}BE^{-1}$  (volgens 1.66 ook  $> 0$ ) dus  $= HAH'$  ( $\lambda_i > 0$ ). Dus  $B = E'HAH'E =: D'AD$  met  $D = H'E$ . Dit geeft  $A = E'E = E'HH'E = D'D = D^2$ .

Opmerking: is B alleen symmetrisch, dan geldt de stelling ook, alleen vervalt  $\lambda_i > 0$ .

1.74. Stel  $A > 0, B \geq 0$  en  $B \neq 0$ . Dan is  $|A+B| > |A| + |B|$ .

Bewijs:

Volgens 1.73 is

$$\begin{aligned} |A+B| &= |D'D+D'AD| = |D'| |I+\Lambda| |D| = |A| |I+\Lambda| > |A| (1+|\Lambda|) = \\ &= |A| + |A| |\Lambda| = |A| + |B| \quad \text{nl.} \\ |I+\Lambda| &= \prod(1+\lambda_i) > 1 + \prod \lambda_i = 1 + |\Lambda|. \end{aligned}$$

Gevolg: Is  $A > B \Rightarrow |A| > |B|$ .

1.75.  $sp(AA) \leq sp(A'A)$ .

Bewijs:

$sp(A' \pm A)'(A' \pm A) \geq 0$  (zie 1.64 en 1.63)

$= sp[(A \pm A')(A' \pm A)] = sp[AA' + A'A \pm AA' \pm A'A] = 2sp[A'A \pm AA]$ .

Dus  $sp(A'A) \geq \pm sp(AA)$  of anders geschreven:  $|sp AA| \leq \|A\|^2$ .

1.76.  $\|AB\|^2 \leq \|A\|^2 \|B\|^2$ . ( $A = a', B = b$  geeft Cauchy-Schwarz ongelijkheid).

Bewijs:

$$\begin{aligned} \|AB\|^2 &= \sum_{ij} [(AB)_{ij}]^2 = \sum_{ij} (a'_{i*} b_{*j})^2 \leq \sum_{ij} a_{i*}^2 b_{*j}^2 \quad (\text{Cauchy-Schwarz}) \\ &= \sum_i a_{i*}^2 \sum_j b_{*j}^2 = \|A\|^2 \|B\|^2. \end{aligned}$$

Speciaal geval:  $\|Ax\|^2 \leq \|A\|^2 \|x\|^2 = \|A\|^2 x^2$ .

Opmerking: Cauchy-Schwarz ongelijkheid luidt:  $(x,y)^2 \leq (x,x)(y,y)$  oftewel  $(x'y)^2 \leq \|x\|^2 \|y\|^2 = x^2 y^2$ ; we zien dat  $\|Ax\| \leq M\|x\|$  voor alle  $x$ .

1.77.  $\sum_0^\infty A^k = (I-A)^{-1}$  voor  $\|A\| < 1$ . ( $I - A$  is regulier, zie som 1.24).

Bewijs:

Definitie:  $\lim_{n \rightarrow \infty} S_n = S$  dus  $\|S_n - S\| < \varepsilon$  voor  $n > N$ .

Stel  $S_n := \sum_0^n A^k$  dus  $I - A^{n+1} = S_{n-1}(I-A)$  (som eindige meetkundige reeks met

reden  $A$  en eerste term  $I$ ). Daar  $\|A^n\| \leq \|A\|^n$  (1.76) en  $\|A\| < 1$  is  $\lim A^n = 0$  en dus  $\lim_{n \rightarrow \infty} S_n = (I-A)^{-1}$ .

1.78.  $A_{n \times n} > 0 \Rightarrow |A| \leq \prod_i a_{ii}$ .

Bewijs met inductie:

Waar voor  $n = 1$ ; stel waar voor  $n = n$ .

Partitioneer  $A_{(n+1)(n+1)} = \begin{pmatrix} A_{nn} & b \\ b' & a_{n+1,n+1} \end{pmatrix}$ .

Dus (zie 1.35)

$|A_{(n+1)(n+1)}| = |A_{nn}| |a_{n+1,n+1} - b'A_{nn}^{-1}b| \leq |A_{nn}| a_{n+1,n+1} \leq \prod_i^{n+1} a_{ii} \quad n!$ .

$b'A^{-1}b \geq 0$  en  $a_{n+1,n+1} > 0$  (zie 1.67). Hieruit volgt de

1.79. Hadamard ongelijkheid:  $|A|^2 \leq \prod_i \|a_{*i}\|^2$ .

Bewijs:

$B := A'A$  dus (1.78)  $|B| \leq \prod b_{ii}$ ; maar  $b_{ii} = a_{*i}^2 = \|a_{*i}\|^2$  en  $|B| = |A|^2$  oftewel  $|A|^2 \leq \prod_i \|a_{*i}\|^2$ .

In woorden meetkundig: het volume opgespannen door de kolommen van  $A$  (dus  $\det A$ ) is kleiner of gelijk aan het product van de lengten dezer kolommen.

1.80. Zij  $A \in \mathcal{M}_{p,q}^p \Rightarrow \exists B \in \mathcal{M}_{q,p}^p$  met  $AB = I_p$ ; B heet *rechterinverse* van A.

Neem  $B = A'(AA')^{-1}$ . Analoog:

Zij  $A \in \mathcal{M}_{p,q}^q \Rightarrow \exists B \in \mathcal{M}_{q,p}^q$  met  $BA = I_q$ ; B heet *linkerinverse* van A.

Neem  $B = (A'A)^{-1}A'$ .

1.81. *Hoofdwaaardendecompositie*: Zij  $r(A_{p \times q}) = r$ . Dan bestaan er matrices  $W_{pr}$ ,  $V_{qr}$  met  $W^2 = V^2 = I$  en een diagonaalmatrix  $\Lambda$  zodat  $A = W\Lambda^{\frac{1}{2}}V'$ .

Bewijs:

Neem voor de kolommen van V orthonormale eigenvectoren behorende bij de r eigenwaarden  $> 0$  van de matrix  $A^2$ . Dus  $A^2 = VAV'$ . Definieer  $W := AVA^{-\frac{1}{2}}$ . Dan is  $W^2 = \Lambda^{-\frac{1}{2}}V'A^2V\Lambda^{-\frac{1}{2}} = I$ . En  $(A - W\Lambda^{\frac{1}{2}}V')^2 = [A(I - VV')]^2 = (I - VV')A^2(I - VV') = 0$ , dus  $A = W\Lambda^{\frac{1}{2}}V'$ . q.e.d.

Opmerking:

1.  $\Lambda$  bevat de positieve eigenwaarden van  $A^2$ , deze worden de *hoofdwaaarden* van A genoemd (Eng. singular values).
2. De kolommen van W zijn de orthonormale eigenvectoren behorende bij de positieve eigenwaarden van  $AA'$  (zie ook 1.72) nl.:

$$AA'W = (W\Lambda^{\frac{1}{2}}V')(V\Lambda^{\frac{1}{2}}W')W = W\Lambda.$$

3. Men kan W en V aanvullen tot orthogonale matrices door toe te voegen orthonormale bases van  $N(A')$  resp.  $N(A)$ .  $\Lambda$  wordt dan met nullen aangevuld tot  $\Lambda_{pq}$ .
4. Andere versie: er bestaan reguliere matrices R en Q zodat

$$RAQ = \begin{pmatrix} I_{rr} & 0 \\ 0 & 0 \end{pmatrix}. \text{ Dit volgt direct uit opmerking 3.}$$

$$\text{Dan is } R = W' \text{ en } Q = V \begin{pmatrix} \Lambda^{-\frac{1}{2}} & 0 \\ 0 & I \end{pmatrix}.$$

1.82. A, B symmetrisch:  $|I-tA||I-tB| = |I-tA-tB| \Leftrightarrow AB = 0$ .

Bewijs  $\Leftarrow$ : triviaal.

Bewijs  $\Rightarrow$ :  $|I-tB| = |I-tA|^{-1}|I-tA-tB| = |I-tKB|$  met  $K := (I-tA)^{-1}$ .

Opmerking:  $I-tA$  is regulier als  $1/t$  geen eigenwaarde is van  $A$ .

$$\begin{aligned} \text{Dus ook is } |I-t^2 B^2| &= |I-t^2 (KB)^2| \text{ oftewel } \text{sp } B^2 = \text{sp}(KB)^2 = \\ &= \text{sp} [(I+tA+t^2 A^2 + \dots)B]^2 = \text{sp}[B^2+t(\dots)+t^2(ABAB+A^2 B^2+BA^2 B) + O(t^3)] \end{aligned}$$

$\therefore \text{sp}[BABA+2(BA)(BA)'] = 0$ . Stel  $BA = D$ . Nu is (1.75)

$$\text{sp}(DD+DD') \geq 0 \text{ dus } \text{sp}(DD') = 0 \text{ oftewel } D = BA = AB = 0 \text{ (1.58).}$$

1.83.  $A, B$  symmetrisch;  $AB = BA \iff \} \text{ orthog. } H \text{ zodat } H'AH = \Lambda_1 \text{ diag.}$

$$H'BH = \Lambda_2 \text{ diag.}$$

Bewijs  $\Leftarrow$  :  $AB = HA_1 H' HA_2 H' = HA_1 \Lambda_2 H' = HA_2 \Lambda_1 H' = HA_2 H' HA_1 H' = BA$ .

Bewijs  $\Rightarrow$  :

$$\text{Stel } R'AR = \Lambda_1 = \begin{pmatrix} \lambda_1 I & & 0 \\ & \ddots & \\ 0 & & \lambda_r I \end{pmatrix} \quad R \text{ orthogonaal, } \lambda_i \neq \lambda_j.$$

Stel  $R'BR =: C$  (dus symmetrisch)

$AB = BA$  geeft  $\Lambda_1 C = C \Lambda_1$ , of uitgeschreven,  $C$  gepartitioneerd:

$$\begin{pmatrix} \lambda_1 I & & 0 \\ & \ddots & \\ 0 & & \lambda_r I \end{pmatrix} \begin{pmatrix} C_{11} \dots C_{1r} \\ \vdots \\ C_{r1} \dots C_{rr} \end{pmatrix} = \begin{pmatrix} C_{11} \dots C_{1r} \\ \vdots \\ C_{r1} \dots C_{rr} \end{pmatrix} \begin{pmatrix} \lambda_1 I & & 0 \\ & \ddots & \\ 0 & & \lambda_r I \end{pmatrix}.$$

Dit geeft  $\lambda_i C_{ij} = \lambda_j C_{ij}$  en daar  $\lambda_i \neq \lambda_j$  is dus  $C_{ij} = 0$  voor  $i \neq j$ .

Stel  $Q_i' C_{ii} Q_i = D_i$  (diag) en  $Q_i$  orthog. (dit kan daar  $C$ , dus  $C_{ii}$  symmetrisch is)

$$Q := \begin{pmatrix} Q_1 & & 0 \\ & \ddots & \\ 0 & & Q_r \end{pmatrix} \quad \text{met } Q'Q = I.$$

$$\text{Nu is } Q' C Q = \begin{pmatrix} D_1 & & 0 \\ & \ddots & \\ 0 & & D_r \end{pmatrix} =: \Lambda_2 \text{ (diag.)}$$

$$\text{en } Q' \Lambda_1 Q = \begin{pmatrix} Q_1' & & 0 \\ & \ddots & \\ 0 & & Q_r' \end{pmatrix} \begin{pmatrix} \lambda_1 I & & 0 \\ & \ddots & \\ 0 & & \lambda_r I \end{pmatrix} \begin{pmatrix} Q_1 & & 0 \\ & \ddots & \\ 0 & & Q_r \end{pmatrix} = \Lambda_1.$$

Dus  $Q'R'ARQ = \Lambda_1$  en  $Q'R'BRQ = \Lambda_2$ ; noem  $RQ = H$ . q.e.d.

Gevolg:  $A, B > 0$  en  $AB$  symmetrisch  $\Rightarrow AB > 0$ .

Idempotente matrices en projecties

Nu volgen enkele eigenschappen betreffende idempotente matrices. Deze spelen een belangrijke rol o.a. bij het opsplitsen van kwadraat-sommen en het bepalen van kansdichtheden van kwadratische vormen.

1.90.  $P$  orthogonale projector  $\Leftrightarrow P = P^2$  (dus symmetrisch en idempotent en met (1.64) niet negatief).

Bewijs:

Zij  $R^n$  Euclidische vectorruimte met orthonormale basis.

$$[\forall x, y (y-Py)'Px = y'(I-P)'Px = 0] \Leftrightarrow (I-P)'P = 0 \Leftrightarrow P = P'P = P^2 .$$

Opmerkingen:

- i) Is  $P$  een orthogonale projector op een lineaire deelruimte  $D$  dan is  $P^\perp := I-P$  de orthogonale projector op het orthoplement  $D^\perp$  van  $D$ , nl.  $I-P$  is weer symmetrisch en idempotent en  $P^\perp P = 0$ .
- ii) Met  $P_A$  zal voortaan bedoeld zijn een orthogonale projector op de deelruimte  $A$  of  $\langle A \rangle$ .
- iii) Is  $A \perp B$  dan is  $P_A P_B = 0$  nl.  $(P_A x, P_B y) = x' P_A P_B y = 0$  voor alle  $x, y$  dus  $P_A P_B = 0$  (zie 1.57).

1.91. Gevolg:

- i)  $P_{n \times n} = \left\{ \frac{1}{n} \right\} = P_u = \frac{1}{n} U = \frac{1}{n} u u'$  is een orthogonale projector op  $\langle u \rangle$  (ga na).
- ii)  $P_X := X(X'X)^{-1}X'$  is de orthogonale projector op  $\langle X \rangle$  (ga na).  
 $P_X^\perp = I - P_X$  is de orthogonale projector op  $\langle X \rangle^\perp$ .
- iii) De orthogonale projectie van een vector  $y$  op een vector  $x$  is  $\frac{y'x}{x'x} x$ .  
Is  $x = u$ ,  $y' = (y_1, \dots, y_n)$  dan wordt dit juist  $\bar{y} := (\bar{y}_1, \dots, \bar{y}_n)'$  met  $\bar{y}_i := \Sigma y_i / n$ .

1.92.  $\langle P_X \rangle = \langle X \rangle$  nl.  $P_X X = X$  dus  $\langle X \rangle \subset \langle P_X \rangle$  (zie 1.48).  
Verder is  $r(P_X) = \dim \langle X \rangle = r(X)$  dus  $\langle X \rangle = \langle P_X \rangle$ .

1.93.  $A_{p \times p}$  niet defect en  $\forall i \lambda_i = 0$  of  $1 \iff A$  idempotent.

Bewijs  $\Rightarrow$  :  $A = BAB^{-1}$  geeft  $A^2 = BAB^{-1}BAB^{-1} = BA^2B^{-1} = BAB^{-1} = A$   
 daar  $A^2 = A$ .

$\Leftarrow$   $A^2 = A$ ;  $Ax = \lambda x$ ;  $A^2x = \lambda^2x = \lambda x$ ;  $\lambda(\lambda-1)x = 0$ ,  $\lambda = 0$  of  $1$ .

Dus  $\langle A \rangle = E_{\lambda=1}$ . Stel  $r(A) = r$ , dan  $\dim N(A) = p-r$ . Vul  $B$  aan met  $p-r$  onafhankelijke eigenvectoren uit  $N(A)$  bij  $\lambda = 0$ .

Dus  $A$  niet defect!

Opmerking: Is  $A$  symmetrisch, dan is  $A$  s.p.d. tenzij  $A = I_p$ .

1.94.  $A$  idempotent:  $\text{sp}(A) = r(A)$  nl. (met 1.93)

$\text{sp } A = \text{sp}(BAB^{-1}) = \text{sp } A = r(A)$ . De relatie is zeer belangrijk!

1.95.  $P_D$  orthogonale projector op  $r$ -dim.deelruimte  $D \iff P = G^2$  met  $GG' = I_r$ .

Bewijs  $\Rightarrow$  :  $P = HAH' = H \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} H' = H \begin{pmatrix} I \\ 0 \end{pmatrix} (I \ 0) H' =: G^2$  (zie ook 1.64).

$\Leftarrow$   $G^2$  is volgens 1.90 een orthogonale projector! En met 1.92 is  $\dim D = r(P_D) = r(G^2) = r(G) = r(GG') = r(I_r) = r$  (met 1.45).

Opmerking: Daar  $PG' = G'GG' = G'$  zijn de kolommen van  $G'$  orthonormale eigenvectoren bij  $\lambda = 1$ .

1.96.  $\Sigma A \Sigma'$  orthog projector  $\iff \Sigma^2 A \Sigma^2 A \Sigma^2 = \Sigma^2 A \Sigma^2$  en  $\Sigma^2 A \Sigma^2$  symmetrisch

$1^e$   $\Sigma A \Sigma' = \Sigma A' \Sigma' \iff \Sigma^2 A \Sigma^2 = \Sigma^2 A' \Sigma^2$  (met 1.60).

$2^e$   $\Sigma A \Sigma' \Sigma A \Sigma' = \Sigma A \Sigma' \iff \Sigma^2 A \Sigma^2 A \Sigma^2 = \Sigma^2 A \Sigma^2$  (met 1.60).

1.97. Zij  $B$  deelruimte  $\subset V$ ;  $a \in V$ ; dan geldt  $d[a, B] = \|a - P_B a\|$  nl.

$\forall b \in B$  is  $\|a-b\|^2 = \|a - P_B a\|^2 + \|P_B a - b\|^2$  dus  $\|a - P_B a\|^2 \leq \|a-b\|^2$  (zie 1.9).

1.98. Probleem. Bepaal bij gegeven matrix  $Y_{n \times p}$  en gegeven  $k < n$ , een matrix  $X_{n \times k}$  van volle rang en een matrix  $B$ , zodat  $\|Y - XB\|$  minimaal is.

Oplossing. De matrices van de orde  $n \times p$  vormen een lineaire vectorruimte en met het inproduct als gedefinieerd in 1.8 een Euclidische vectorruimte. Deze wordt metrisch met  $d[A, B] := \|A - B\|$  als afstandsfunctie. We proberen dus  $Y$  "zo goed mogelijk" door  $XB$  te benaderen.

Definieer  $P$  als de orthogonale projector op de, nog onbekende,  $k$ -dimensionale lineaire deelruimte  $\langle X \rangle$  opgespannen door de kolommen van  $X$ ;  $P^\perp$  is de orthogonale projectie op het orthoplement  $\langle X \rangle^\perp$  van  $\langle X \rangle$ . Uit 1.32 en de projectiestelling volgt nu

$$f(X, B) := \|Y - XB\|^2 = \sum_j \|y_{*j} - Xb_{*j}\|^2 = \sum_j \|P(y_{*j} - Xb_{*j})\|^2 + \sum_j \|P^\perp(y_{*j} - Xb_{*j})\|^2$$

oftewel

$$f(X, B) = \|P(Y - XB)\|^2 + \|P^\perp(Y - XB)\|^2$$

$$f(X, B(X)) := \min_B f(X, B) = \|P^\perp Y\|^2$$

voor  $(Y - XB)'X = 0$ ,  $B(X) = (X'X)^{-1}X'Y$ . Nu is  $YY' = H\Lambda H'$  ( $H$  orthonormale eigenvectoren;  $\Lambda$  eigenwaarden met  $\lambda_1 \geq \dots \geq \lambda_n$ ). Dus

$$\|P^\perp Y\|^2 = \text{sp}(P^\perp YY') = \text{sp}(P^\perp H\Lambda H') = \|P^\perp H\Lambda^{1/2}\|^2 = \sum_1^n \lambda_i \|P^\perp h_{*i}\|^2 =: \sum_1^n \lambda_i p_i$$

met  $0 \leq p_i \leq 1$  en  $\sum_1^n p_i = \text{sp}(P^\perp HH') = \text{sp } P^\perp = r(P^\perp) = n - k$  (zie 1.94).

Definieer  $s_i = \sum_1^i p_j$  voor  $i = 1, \dots, n$ ;  $s_0 = \lambda_{n+1} = 0$ . Dan is

$$\sum_1^n \lambda_i p_i = \sum_1^n \lambda_i (s_i - s_{i-1}) = \sum_1^n (\lambda_i - \lambda_{i+1}) s_i.$$

Nu is  $s_n = n - k$ , dus  $s_i \geq i - k$  daar  $p_i \leq 1$ . Verder is  $s_i \geq 0$ , dus

$$\sum_1^n (\lambda_i - \lambda_{i+1}) s_i \geq \sum_{k+1}^n (\lambda_i - \lambda_{i+1}) (i - k) = \sum_{k+1}^n \lambda_i,$$

oftewel

$$f(X_0, B_0) := \min_X f(X, B(X)) = \min_X \sum_1^n \lambda_i \|P^\perp h_{*i}\|^2 = \sum_{k+1}^n \lambda_i$$

voor  $\|P^\perp h_{*i}\| = 0$ ,  $i = 1, \dots, k$ , en  $\|P^\perp h_{*i}\| = 1$  voor  $i = k+1, \dots, n$ .

Conclusie: Kies voor de kolommen van  $X_0$  orthon. eigenvect. van  $YY'$  behorende bij de  $k$  grootste eigenwaarden. Inderdaad is dan  $\|P^\perp h_{*i}\| = \|h_{*i}\| = 1$  voor  $h_{*i} \in \langle X_0 \rangle^\perp$  en  $\|P^\perp h_{*i}\| = 0$  voor  $h_{*i} \in \langle X_0 \rangle$ .



Opmerking. We zien dat  $XB = X(X'X)^{-1}X'Y = PY$ . Ook zien we dat, wil deze benadering van  $Y$  door  $XB$  zin hebben,  $r(YY') = r(Y) = r(\Lambda) > k$  moet zijn (dus  $p > k$ ).

Uit dit probleem volgen direct enkele stellingen betreffende extrema:

Onderstel  $A_{n \times n} \geq 0$  met eigenwaarden  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ ;  $r(X_{n \times k}) = k$ .

Dan is

$$1.99. \min_X \text{sp}(P_X A) = \sum_{i=n-k+1}^n \lambda_i = \text{som van de } k \text{ kleinste eigenwaarden van } A.$$

De kolommen van  $X$  zijn de bijbehorende orthonormale eigenvectoren.

Analoog is natuurlijk:

$$\max_X \text{sp}(P_X A) = \sum_{i=1}^k \lambda_i = \text{som van de } k \text{ grootste eigenwaarden van } A, \text{ nl.}$$

$$= \max_X \text{sp}[(I - P_X^{\perp})A] = \sum_{i=1}^n \lambda_i - \min_X \text{sp}(P_X^{\perp} A) = \sum_{i=1}^n \lambda_i - \sum_{i=k+1}^n \lambda_i = \sum_{i=1}^k \lambda_i.$$

Bijzondere gevallen hiervan zijn:

$$1.100. \max_{X^2=I} \text{sp}(X'AX) = \sum_{i=1}^k \lambda_i,$$

nl.  $\text{sp}(X'AX) = \text{sp}(XX'A) = \text{sp}(P_X A)$ , nl.  $P_X = XX'$  is symmetrisch en idempotent.

Voor  $X = x$  wordt dit:

$$\max_{|x|=1} x'Ax = \lambda_1 = \text{grootste eigenwaarde van } A, \text{ en } x \text{ bijbehorende eigenvector.}$$

Dit is natuurlijk hetzelfde als  $\max_{x \neq 0} \frac{x'Ax}{x'x} = \lambda_1$  (geldt ook als  $A$  symmetrisch is).

Opmerking:

1) het quotiënt  $q(x) := \frac{x'Ax}{x'x}$  wordt het Rayleigh quotiënt genoemd.

2)  $\max_x \frac{x'Ax}{x'Bx} = \lambda_1 = \text{grootste eigenwaarde van } B^{-1}A$ .

Hierbij is  $A$  symmetrisch,  $B > 0$  (ga na).

§ 2. Enkele begrippen uit de mathematische statistiek

In deze paragraaf worden enkele begrippen uit de mathematische statistiek nog eens herhaald en enkele nieuwe begrippen ingevoerd, o.a. de maximum likelihood methode die veel wordt toegepast om schatters te vinden, en de variantie-covariantiematrix die een centrale rol speelt.

a. Definities, notaties

2.1. De  $\mathcal{E}$ -operator

$$\mathcal{E}_{\underline{x}} := \sum p_i x_i \quad , \quad \underline{x} \text{ diskreet, mits de som absoluut convergeert,}$$

en

$$\mathcal{E}_{\underline{x}} := \int_{-\infty}^{\infty} x f(x) dx \quad , \quad \underline{x} \text{ kontinu, mits de integraal absoluut convergeert.}$$

Stelling.  $\mathcal{E}_{\underline{x}} g(\underline{x}) := \int_{-\infty}^{\infty} g(x) f(x) dx = \mathcal{E} g(\underline{x}) .$

Ook meerdimensionaal:

$$\mathcal{E} g(\underline{x}_1, \dots, \underline{x}_n) = \int_G g(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \dots dx_n .$$

$\mathcal{E} g(\underline{x})$  heet de *verwachting* van  $g(\underline{x})$ .

Opmerking. Onderstreepte letters zijn stochastische grootheden.

2.2. Een *stochastische matrix*  $\underline{A} = \{a_{ij}\}$  is een matrix waarvan de elementen  $a_{ij}$  stochastisch zijn (niet te verwarren met een Markofmatrix).

2.3. De *verwachting* van  $\underline{A}$  is  $\mathcal{E}\underline{A} := \{\mathcal{E}a_{ij}\}$ .

2.4.  $\underline{x}$  is *isomoor* met  $\underline{y}$  (notatie  $\underline{x} \approx \underline{y}$ ) betekent  $\underline{x}$  en  $\underline{y}$  hebben dezelfde kansverdeling;  $\underline{x} \sim N(\mu, \sigma^2)$  betekent  $\underline{x}$  is normaal verdeeld met gemiddelde  $\mu$  en variantie  $\sigma^2$ .

2.5. *Karakteristieke functie*  $\varphi_{\underline{x}}(t) := \mathcal{E} e^{it\underline{x}}$  ( $t$  reëel). Analoog indien  $\underline{x}$  een  $p$ -dimensionale variabele (vektor) is:  $\varphi_{\underline{x}}(t) := \mathcal{E} e^{it'\underline{x}}$  met  $t' = (t_1, \dots, t_p)$  reëel.

- 2.6. Een *onverleete steekproef* ter grootte  $n$  van de  $p$ -dimensionale variabele  $\underline{x} = (\underline{x}_1, \dots, \underline{x}_p)'$  is een matrix  $\underline{X}_{n \times p}$  van  $n$  onafhankelijke, met  $\underline{x}$  isomere stochastische vektoren. De rijen geven dus de herhalingen aan.
- 2.7. Een *schatting* van een populatiegrootte (parameter) op basis van een steekproef  $\underline{X}_{n \times p}$  is een functie van deze  $\underline{X}_{n \times p}$ .

2.8. Een schatter  $\underline{t}$  voor de parameter  $\vartheta$  heet *zuiver* als geldt:

$$E \underline{t} = \vartheta \quad (\text{voor alle } \vartheta \in \text{parameterruimte}).$$

- 2.9. Een *schatting* is een realisatie van een schatter, dus  $t$  is een realisatie van  $\underline{t}$ .
- 2.10. De *aannemelijkheidsfunctie (likelihood function)* van een onafhankelijke steekproef  $\underline{X}_{n \times p}$  ter grootte  $n$  uit de verdeling  $f(x; \vartheta)$  waarin  $\underline{x} = (\underline{x}_1, \dots, \underline{x}_p)'$  en  $\vartheta = (\vartheta_1, \dots, \vartheta_k)'$ , is de simultane kansdichtheid van  $\underline{X}_{n \times p}$  opgevat als functie van  $\vartheta$ :

$$\underline{L}(\vartheta) = \prod_1^n f(x_{i*}; \vartheta) \quad \text{waarin } \underline{x}_{i*} = (\underline{x}_{i1}, \dots, \underline{x}_{ip})' .$$

- 2.11. De *meest aannemelijke schatter* voor  $\vartheta$  is die  $\hat{\vartheta}$  die  $\underline{L}(\vartheta)$  maximaliseert.
- 2.12. De *likelihood-ratio toets* voor de nulhypothese  $H_0 : \vartheta \in \omega$  tegen de alternatieve  $H_1 : \vartheta \notin \omega$ , heeft als kritieke zone die steekproefuitkomsten waarvoor

$$0 < \lambda = \frac{L_{\omega}}{L_{\Omega}} := \frac{\sup_{\vartheta \in \omega} L(\vartheta)}{\sup_{\vartheta \in \Omega} L(\vartheta)} < \lambda_{\alpha} \quad (\Omega = \text{parameterruimte}).$$

- |   |  |
|---|--|
| 2.13. <i>Populatie (ééndimensionaal)</i>                    | <i>Steekproef</i> $(\underline{x}_1, \dots, \underline{x}_n)$  |
| <i>populatiegrootte (parameter)</i>                         | <i>steekproefgrootte (schatting)</i>   |
| <i>pop.gemiddelde</i> $\mu := E \underline{x}$              | <i>steekpr.gemiddelde</i> $\bar{\underline{x}} := \sum_1^n \underline{x}_i / n$                      |
| <i>pop.variantie</i> $\sigma^2 := E(\underline{x} - \mu)^2$ | <i>steekpr.variantie</i> $\underline{s}^2 := \sum (\underline{x}_i - \bar{\underline{x}})^2 / (n-1)$ |
| <i>pop.spreiding</i> $\sigma := \sqrt{\sigma^2}$            | <i>steekpr.spreiding</i> $\underline{s} := \sqrt{\underline{s}^2}$                                   |

populatie (tweedimensionaal)

steekproef  $x_1, y_1; \dots; x_n, y_n$

$$\begin{aligned} \text{pop. covariantie } \text{cov}(\underline{x}, \underline{y}) &= \\ &= \sigma^2(\underline{x}, \underline{y}) := \mathcal{E}[(\underline{x} - \mu_x)(\underline{y} - \mu_y)] \end{aligned}$$

$$\begin{aligned} \text{steekpr. covariantie } s^2(\underline{x}, \underline{y}) &:= \\ &= \sum_1^n (\underline{x}_i - \bar{x})(\underline{y}_i - \bar{y}) / (n - 1) \end{aligned}$$

pop. correlatiecoëfficiënt

steekpr. correlatiecoëfficiënt

$$\rho(\underline{x}, \underline{y}) = \frac{\sigma^2(\underline{x}, \underline{y})}{\sigma(\underline{x})\sigma(\underline{y})}$$

$$r(\underline{x}, \underline{y}) = \frac{s^2(\underline{x}, \underline{y})}{s(\underline{x})s(\underline{y})}$$

2.14. De *variantie-covariantiematrix* van de stochastische vektor  $\underline{x} = (x_1, \dots, x_p)'$  is

$$\begin{aligned} \text{VAR}(\underline{x}) = \Sigma^2(\underline{x}) &:= \mathcal{E}(\underline{x} - \mu)(\underline{x} - \mu)' = \Sigma^2 \quad \text{met } \mu = \mathcal{E}\underline{x} = (\mu_1, \dots, \mu_p)' \\ &\text{en } \mu_i = \mathcal{E}x_i. \end{aligned}$$

Inderdaad is

$$(\Sigma^2)_{ij} = \mathcal{E}(x_i - \mu_i)(x_j - \mu_j) = \sigma_{ij}^2 = \text{cov}(x_i, x_j) = \sigma_{ji}^2.$$

Zijn de  $x_i$  alle gestandaardiseerd, dan wordt  $\Sigma^2 = \mathcal{R}$  de *correlatiematrix*.

Opmerking. In de literatuur is  $\Sigma$  gebruikelijk i.p.v.  $\Sigma^2$ . Toch biedt deze laatste notatie voordelen, o.a. door de analogie met  $\sigma^2$  voor  $p = 1$ .

$\Sigma^2 = \Sigma' \Sigma$ . Dat dit geoorloofd is, dus m.a.w. dat  $\text{VAR}(\underline{x}) \geq 0$  is, wordt bewezen in 2.39. Ook is  $\Sigma^2 = \Sigma \Sigma$ , nu met symmetrische  $\Sigma$  (zie 1.64).

2.15. De *covariantiematrix* der vektoren  $\underline{x} = (x_1, \dots, x_p)'$  en  $\underline{y} = (y_1, \dots, y_q)'$  is

$$\text{COV}(\underline{x}, \underline{y}) := \mathcal{E}(\underline{x} - \mu_x)(\underline{y} - \mu_y)'$$

In het bijzonder is  $\text{COV}(\underline{x}, \underline{x}) = \text{VAR } \underline{x}$  en  $\text{COV}(\underline{x}, \underline{y}) = \text{COV}'(\underline{y}, \underline{x})$ .

2.16. Een lineaire functie  $a'y$  is een BLUE (*best linear unbiased estimator*) voor  $c'\beta$  als  $\mathcal{E}(a'y) = c'\beta$  en  $\text{var}(a'y)$  minimaal is (onder alle zuivere lineaire schatters).

Opmerking. Ook gebruikelijke definitie voor BLUE is:  $\underline{t} = (t_1, \dots, t_p)'$  is een BLUE voor  $\vartheta$  als  $\mathcal{E}\underline{t} = \vartheta$ ,  $\underline{t}$  lineair en  $\underline{t}$  minimale gegeneraliseerde variantie heeft (binnen de klasse van alle lineaire zuivere schatters).

2.17.  $|\Sigma^2|$  wordt de *gegeneraliseerde variantie* van een verdeling genoemd.

b. Eigenschappen

2.30.  $\sigma^2(\underline{x}) = \xi_{\underline{x}^2} - (\xi_{\underline{x}})^2$ ;  $\sigma^2(\underline{x}, \underline{y}) = \xi(\underline{xy}) - \xi_{\underline{x}} \xi_{\underline{y}}$ .

2.31. Zijn  $\underline{x}$  en  $\underline{y}$  stochastisch onafhankelijk, dan geldt  $\xi(\underline{xy}) = \xi_{\underline{x}} \xi_{\underline{y}}$  en dus zijn  $\underline{x}$  en  $\underline{y}$  ongecorreleerd oftewel  $\rho(\underline{x}, \underline{y}) = 0$  (niet omgekeerd).

2.32.  $\text{var}(\alpha \underline{x} + \beta \underline{y}) = \alpha^2 \text{var } \underline{x} + \beta^2 \text{var } \underline{y} + 2\alpha\beta \text{cov}(\underline{x}, \underline{y})$ .

2.33. Is  $\underline{y} = (\underline{x} - \mu)/\sigma$ , dan is  $\mu(\underline{y}) = 0$  en  $\sigma(\underline{y}) = 1$ ;  $\underline{y}$  heet *gestandaardiseerd*.  
Analoog geldt: als  $y_i = (x_i - \bar{x})/s(x)$ , dan is  $\bar{y} = 0$  en  $s(y) = 1$ .

2.34. Is  $\underline{u} = \alpha \underline{x} + \beta$  en  $\underline{v} = \gamma \underline{y} + \delta$  met  $\alpha\gamma \neq 0$ , dan is  $\rho^2(\underline{u}, \underline{v}) = \rho^2(\underline{x}, \underline{y})$ , m.a.w.  $\rho^2$  is onder deze transformaties invariant.

2.35.  $\xi(A\underline{B} + \underline{C}) = A \xi_{\underline{B}} + \xi_{\underline{C}}$ .

Bewijs:

$$\begin{aligned} [\xi(A\underline{B} + \underline{C})]_{ij} &= \xi\left(\sum_k a_{ik} b_{kj} + c_{ij}\right) = \sum_k a_{ik} \xi_{b_{kj}} + \xi_{c_{ij}} = \\ &= \sum_k a_{ik} (\xi_{\underline{B}})_{kj} + (\xi_{\underline{C}})_{ij} = (A\xi_{\underline{B}})_{ij} + (\xi_{\underline{C}})_{ij} = (A\xi_{\underline{B}} + \xi_{\underline{C}})_{ij} . \end{aligned}$$

2.36. De verwachting van een kwadratische vorm:

$$\xi(\underline{x}'A\underline{x}) = \text{sp}(A\Sigma^2) + \mu'A\mu . \text{ Analoog } \xi(\underline{x}'A\underline{y}) = \text{sp}(A\Sigma_{21}) + \mu'_x A \mu_y .$$

Bewijs:  $\Sigma^2 = \xi(\underline{x} - \mu)(\underline{x} - \mu)' = \xi_{\underline{xx}'} - \mu\mu'$

dus  $A\Sigma^2 = \xi(A\underline{xx}') - A\mu\mu'$

en  $\text{sp}(A\Sigma^2) = \xi(\text{sp } A\underline{xx}') - \text{sp}(A\mu\mu') = \xi(\underline{x}'A\underline{x}) - \mu'A\mu$  (zie 1.32).

Is  $\Sigma^2 = \sigma^2 I$  en  $A = P$  een orthogonale projectie met  $r(P) = v$  dan ontstaat:

$$\xi\|\underline{P}\underline{x}\|^2 = \xi(\underline{x}'P\underline{x}) = \text{sp}(\sigma^2 P) + \mu'P\mu = v\sigma^2 + \mu'P\mu .$$

2.37.  $\text{COV}(A\underline{x} + B\underline{y}, C\underline{z}) = A \text{COV}(\underline{x}, \underline{z})C' + B \text{COV}(\underline{y}, \underline{z})C'$ .

Bewijs:

$$\begin{aligned} \text{COV}(A\underline{x} + B\underline{y}, C\underline{z}) &= \mathcal{E}[(A\underline{x} + B\underline{y} - A\underline{\mu}_x - B\underline{\mu}_y)(C\underline{z} - C\underline{\mu}_z)'] = \\ &= \mathcal{E}[\{A(\underline{x} - \underline{\mu}_x) + B(\underline{y} - \underline{\mu}_y)\}(\underline{z} - \underline{\mu}_z)'] = \mathcal{E}[A(\underline{x} - \underline{\mu}_x)(\underline{z} - \underline{\mu}_z)'] + \\ &+ \mathcal{E}[B(\underline{y} - \underline{\mu}_y)(\underline{z} - \underline{\mu}_z)'] = A \text{COV}(\underline{x}, \underline{z})C' + B \text{COV}(\underline{y}, \underline{z})C'. \end{aligned}$$

2.38. Gevolg:  $\text{VAR}(A\underline{x}) = A\Sigma^2A'$ , nl. met  $B = 0$ ,  $A = C$  en  $\underline{z} = \underline{x}$ .

In het bijzonder als  $\underline{y} = A\underline{x}$ ,  $\Sigma^2(\underline{x}) = \sigma^2I$  en  $A$  orthogonaal geldt:

$$\Sigma^2(\underline{y}) = A\Sigma^2(\underline{x})A' = A\sigma^2IA' = \sigma^2 \frac{AA'}{I} = \Sigma^2(\underline{x}).$$

2.39.  $0 < \text{var}(\alpha'\underline{x}) = \alpha'\Sigma^2\alpha$ , d.w.z.  $\Sigma^2 > 0$  (tenzij  $\alpha'\underline{x}$  constant is oftewel de  $\underline{x}_i$  lineair afhankelijk).

Hieruit blijkt dus dat de notatie  $\Sigma^2$  voor  $\text{VAR } \underline{x}$  geoorloofd is, daar  $\Sigma^2 := \Sigma'\Sigma$  dus (s.) p.d. is.

2.40. Is  $\hat{\vartheta}$  een maximum likelihoodschatter voor  $\vartheta$  en  $f$  een één-éénduidige functie, dan is  $f(\hat{\vartheta})$  een maximum likelihoodschatter voor  $f(\vartheta)$ . (Zie [1] pag.48.)

2.41. Als  $\varphi_{\underline{x}}(t) = \varphi_{\underline{y}}(t)$  voor alle  $t$ , dan is  $\underline{x} = \underline{y}$  (zie Cramèr, pag. 94).

2.42.  $\varphi_{A\underline{x}+\underline{\beta}}(t) = \mathcal{E}e^{it'(A\underline{x}+\underline{\beta})} = e^{it'\underline{\beta}} \varphi_{\underline{x}}(A't)$ , i.h.b. voor  $A = \alpha$ :  
 $\varphi_{\alpha\underline{x}+\underline{\beta}} = e^{i\underline{\beta}t} \varphi_{\underline{x}}(\alpha t)$ .

2.43. Is  $\underline{x} \sim N(0,1)$ , dan is  $\varphi_{\underline{x}}(t) = e^{-\frac{1}{2}t^2}$ .

Bewijs:

$$\varphi_{\underline{x}}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-\frac{1}{2}x^2} dx = \frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-it)^2} dx = e^{-\frac{1}{2}t^2}.$$

Is  $\underline{y} \sim N(\underline{\mu}, \sigma^2)$  oftewel  $\underline{y} = \sigma\underline{x} + \underline{\mu}$ , dan is volgens 2.42

$$\varphi_{\underline{y}}(t) = e^{i\underline{\mu}t} e^{-\frac{1}{2}(\sigma t)^2} = \exp(i\underline{\mu}t - \frac{1}{2}\sigma^2 t^2).$$

2.44.  $\varphi_{\underline{x}_i}(t) = \prod_i \varphi_{\underline{x}_i}(t)$  indien alle  $\underline{x}_i$  onderling stochastisch onafhankelijk zijn.

In het bijzonder is dus  $\varphi_{\underline{x+y}}(t) = \varphi_{\underline{x}}(t)\varphi_{\underline{y}}(t)$ , nl.

$$\int \int e^{it(\underline{x+y})} f(x)g(y) dx dy = \int e^{itx} f(x) dx \int e^{ity} g(y) dy = \varphi_{\underline{x}}(t)\varphi_{\underline{y}}(t).$$

Dit niet te verwarren met het volgende:

Is  $\underline{x} = (\underline{x}_1, \dots, \underline{x}_p)'$ , waarbij alle  $\underline{x}_j$  onderling onafhankelijk zijn, dan is

$$\varphi_{\underline{x}}(t) = \prod_{j=1}^p \varphi_{\underline{x}_j}(t_j) \quad \text{met } t = (t_1, \dots, t_p)' \quad (\text{Ook omgekeerd, zie Cramèr pag. 266, 296}).$$

2.45. Stel  $\varphi_{\underline{x}}(t) = \int e^{it'\underline{x}} f(\underline{x}) d\underline{x} = \int e^{i(t_1\underline{x}_1 + t_2\underline{x}_2)} f(\underline{x}_1, \underline{x}_2) d\underline{x}_1 d\underline{x}_2$ , dan is  $\frac{\partial^2 \varphi(0,0)}{\partial t_1 \partial t_2} = - \int \underline{x}_1 \underline{x}_2 f(\underline{x}_1, \underline{x}_2) d\underline{x}_1 d\underline{x}_2$ , nl.:

$$\frac{\partial \varphi}{\partial t_1} = \int i \underline{x}_1 e^{i(t_1\underline{x}_1 + t_2\underline{x}_2)} f(\underline{x}_1, \underline{x}_2) d\underline{x}_1 d\underline{x}_2; \quad \frac{\partial^2 \varphi}{\partial t_1 \partial t_2} = - \int \underline{x}_1 \underline{x}_2 e^{i(t_1\underline{x}_1 + t_2\underline{x}_2)} f(\underline{x}_1, \underline{x}_2) d\underline{x}_1 d\underline{x}_2.$$

2.46. Is  $\varphi_{\underline{x}}(t)$ ,  $\underline{x} = (\underline{x}_1, \dots, \underline{x}_p)'$ , bekend, dan vinden we  $\varphi_{\underline{x}_j}(t_j) = \int e^{it_j \underline{x}_j} f(\underline{x}_j) d\underline{x}_j$  uit  $\varphi_{\underline{x}} = \int e^{it'\underline{x}} f(\underline{x}) d\underline{x}$  door  $t' = (0, \dots, t_j, 0, \dots, 0)$  te nemen.

#### 2.47. Transformatie van stochastische variabelen

Zij  $\underline{x} = (\underline{x}_1, \dots, \underline{x}_p)'$  een stochastische vektor met simultane kansdichtheid  $f(\underline{x}) = f(x_1, \dots, x_p)$ .

Beschouw de één-éénduidige transformaties  $y_i = y_i(\underline{x}_1, \dots, \underline{x}_p)$  met resp. inversen  $\underline{x}_i = \underline{x}_i(y_1, \dots, y_p)$ ,  $i = 1, \dots, p$ .

Dan is de simultane kansdichtheid van  $\underline{y} = (y_1, \dots, y_p)'$ :

$$g(\underline{y}) = f[\underline{x}_1(y_1, \dots, y_p), \dots, \underline{x}_p(y_1, \dots, y_p)] \text{ abs } \left| \frac{\partial \underline{x}}{\partial \underline{y}} \right|.$$

Hierin is  $\left| \frac{\partial \underline{x}}{\partial \underline{y}} \right|$  de determinant van Jacobi  $\left| \left\{ \frac{\partial x_i}{\partial y_j} \right\} \right|$ .

Is in het bijzonder  $\underline{y} = A\underline{x}$  met  $|A| \neq 0$ , dan is  $g(\underline{y}) = f(A^{-1}\underline{y}) \text{ abs } |A^{-1}|$ .

Zijn  $\underline{x}$  en  $\underline{y}$  vectoren en is  $\underline{y} = \underline{y}(\underline{x})$ , dan is  $\frac{\partial \underline{y}}{\partial \underline{x}} := \left\{ \frac{\partial y_i}{\partial x_j} \right\}$ .

Is  $\underline{y} = A\underline{x}$ , dan is  $\frac{\partial \underline{y}}{\partial \underline{x}} = A$ .

### § 3. De multinormale verdeling

De ééndimensionale normale verdeling treedt vaak op omdat het te bestuderen effect veelal de som is van meerdere onafhankelijke variabelen.

Evenals de centrale limietstelling leidt tot de normale verdeling voor een enkelvoudige variabele, zo leidt ook de algemene centrale limietstelling tot de multinormale verdeling van meerdere variabelen.

Dit is een rechtvaardiging voor de centrale plaats die de multinormale verdeling in de multivariate analyse inneemt. Bovendien is het erg plezierig dat deze verdeling wiskundig goed hanteerbaar is.

Stel  $\underline{y} = (y_1, \dots, y_p)'$  met  $y_i \sim N(0,1)$  en alle  $y_i$  zijn stochastisch onafhankelijk. De vektor  $\underline{y}$  heet dan *p-dim. standaardnormaal verdeeld*:  $\underline{y} \sim N_p(0, I)$ . Dan heet  $\underline{x} = \mu + \Sigma' \underline{y}$   $|\Sigma| > 0$  *p-dim. normaal verdeeld*:  $\underline{x} \sim N_p(\mu, \Sigma^2)$ , nl.  $E\underline{x} = \mu$  en  $\text{VAR } \underline{x} = \Sigma' I \Sigma = \Sigma^2$ ; andere notatie  $\underline{y} \approx \underline{x}_p \sim N_p(0, I)$ .

Welke verdeling heeft  $\underline{x}$  ?

Nu is  $f_{\underline{y}}(\underline{y}) = \prod_1^p f_{y_i}(y_i) = (2\pi)^{-\frac{1}{2}p} e^{-\frac{1}{2}\underline{y}'\underline{y}}$ , dus volgens 2.47 is

$$f_{\underline{x}}(\underline{x}) = f_{\underline{y}}[\Sigma^{-T}(\underline{x} - \mu)] \text{ abs } |\Sigma^{-1}| = |\Sigma|^{-1} (2\pi)^{-\frac{1}{2}p} \exp[-\frac{1}{2}(\underline{x} - \mu)' \Sigma^{-2}(\underline{x} - \mu)].$$

Wat is de karakteristieke functie?

Volgens 2.43 en 2.44 is  $\varphi_{\underline{y}}(\underline{t}) = \prod_1^p \varphi_{y_j}(t_j) = \prod_j e^{-\frac{1}{2}t_j^2} = \exp(-\frac{1}{2}\underline{t}'\underline{t})$  en dus volgens 2.42 is

$$\varphi_{\underline{x}}(\underline{t}) = e^{it'\mu} \varphi_{\underline{y}}(\Sigma \underline{t}) = \exp(it'\mu - \frac{1}{2}\underline{t}'\Sigma^2 \underline{t}) \text{ (vergelijk met 2.43).}$$

Hierin mag  $\Sigma^2$  s.p.d. zijn. Dan heet  $\underline{x} \sim N_p^0(\mu, \Sigma^2)$  *sing. norm. verdeeld*.

Samengevat: is  $\underline{x}$  p-dim normaal verdeeld met parameters  $\mu, \Sigma^2$ , dan is

3.1.

$$f(\underline{x}) = |\Sigma|^{-1} (2\pi)^{-\frac{1}{2}p} \exp[-\frac{1}{2}(\underline{x} - \mu)' \Sigma^{-2}(\underline{x} - \mu)]$$

en

3.2.

$$\varphi_{\underline{x}}(\underline{t}) = \exp(it'\mu - \frac{1}{2}\underline{t}'\Sigma^2 \underline{t}).$$



Het geval  $p = 2$ , de tweedimensionale normale verdeling:

$$\Sigma^{-2} = \begin{pmatrix} 1/\sigma_1^2 & -\rho/\sigma_1\sigma_2 \\ -\rho/\sigma_1\sigma_2 & 1/\sigma_2^2 \end{pmatrix} \frac{1}{(1-\rho^2)}$$

en

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}} \exp\left\{-\frac{1}{2(1-\rho^2)} \left[ \frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \right]\right\}$$

Is  $\rho = 0$  dan is  $f(x_1, x_2) = f_1(x_1)f_2(x_2)$  dus  $\underline{x}_1$  en  $\underline{x}_2$  stochastisch onafhankelijk.

Daar  $|\Sigma|^2 = \sigma_1^2\sigma_2^2(1-\rho^2) > 0$  is  $|\rho| < 1$ .

Doorsnijding met vlak  $f(x) = \text{constant}$  geeft ellipsen, daar dan

$(x - \mu)' \Sigma^{-2} (x - \mu) = \text{constant}$  en de eigenwaarden van  $\Sigma^{-2}$  alle positief zijn (zie 1.63). Het middelpunt is  $\mu = (\mu_1, \mu_2)'$ .

### 3.3. Theorema:

Is  $\underline{x} \sim N_p(\mu, \Sigma^2)$  en  $A_{m \times p}$  heeft rang  $m$ , dan is

$$\underline{y} = \alpha + A\underline{x} \sim N_m(\alpha + A\mu, A\Sigma^2A')$$

Bewijs: Dit volgt direkt uit 2.42, 3.2 en 2.41, nl.:

$$\begin{aligned} \varphi_{\underline{y}}(t) &= e^{it'\alpha} \varphi_{\underline{x}}(A't) = e^{it'\alpha} \exp(it'A\mu - \frac{1}{2}t'A\Sigma^2A't) = \\ &= \exp[it'(\alpha + A\mu) - \frac{1}{2}t'(A\Sigma^2A')t] \end{aligned}$$

en daar volgens 2.41 de karakteristieke functie eënduidig de verdelingsdichtheid bepaalt, is het gestelde bewezen. Volgens 1.66 is  $A\Sigma^2A' > 0$ .

### Enkele belangrijke konsekwenties

3.4. Is  $\underline{x} \sim N_p(0, \sigma^2 I)$  en  $A_{r \times p}$  zodat  $AA' = I_r$ , dan is

$$\underline{y} = A\underline{x} \sim N_r(0, \sigma^2 I)$$

Is  $p = r$  (dus  $A$  orthogonaal), dan is dus  $\underline{y} = A\underline{x} \approx \underline{x}$ .

3.5. Neem  $A = (0, \dots, 1, 0, \dots, 0)$  dan volgt direkt  $A\underline{x} = \underline{x}_i \sim N(\mu_i, \sigma_{ii}^2)$ , oftewel elke komponent van  $\underline{x}$  is ook normaal verdeeld.

Pas op: het omgekeerde is niet waar: Als van een vektor  $\underline{x}$  elke komponent normaal verdeeld is, hoeft  $\underline{x}$  nog niet normaal verdeeld te zijn (zie appendix A).

3.6. Neem  $A = \alpha' = (\alpha_1, \dots, \alpha_p)$ , dan volgt uit dit theorema

$$\underline{y} = \alpha' \underline{x} = \alpha_1 \underline{x}_1 + \dots + \alpha_p \underline{x}_p \sim N(\alpha' \mu, \alpha' \Sigma^2 \alpha)$$

oftewel in woorden: Een lineaire combinatie van normaal verdeelde variabelen die een multinormale verdeling volgen, is weer normaal verdeeld.

Pas op: een lineaire combinatie van normaal verdeelde variabelen zonder meer hoeft niet normaal verdeeld te zijn (zie appendix A).

In het bijzonder geeft  $\alpha' = \frac{1}{n} (1, \dots, 1)$  en  $\underline{x} \sim N_n(\mu, \sigma^2 I)$  met  $\mu = (\mu, \dots, \mu)'$   $\alpha' \underline{x} = \bar{\underline{x}} \sim N(\mu, \sigma^2/n)$ .

Soms definieert men een multinormaal verdeelde variabele met behulp van

### 3.7. Theorema

Als elke lineaire combinatie van de componenten van een vektor  $\underline{x}$  normaal verdeeld is, dan is  $\underline{x}$  multinormaal verdeeld (dichtheidsvrije definitie).

Bewijs: Stel  $\underline{x} \sim F_p(\mu, \Sigma^2)$  en  $\alpha' \underline{x} \sim N(\alpha' \mu, \alpha' \Sigma^2 \alpha)$  voor alle  $\alpha \neq 0$ . Dus

$$\varphi_{\alpha' \underline{x}}(t) = \exp(it\alpha' \mu - \frac{1}{2} \alpha' \Sigma^2 \alpha t^2) ;$$

stel  $t = 1$  dan is

$$\varphi_{\alpha' \underline{x}}(t) = \varphi_{\underline{x}}(\alpha t) = \varphi_{\underline{x}}(\alpha) = \exp(i\alpha' \mu - \frac{1}{2} \alpha' \Sigma^2 \alpha)$$

dus  $\underline{x} \sim N_p(\mu, \Sigma^2)$ .

Uit theorema 3.3 volgt ook direkt:

### 3.8. Theorema

Is  $\underline{x} \sim N_p(\mu, \Sigma^2)$  dan is de marginale verdeling van elke deelverzameling van  $\underline{x}$  weer normaal met corresponderende gemiddelde en variantie-covariantiematrix.

Bewijs: Laat  $\underline{x}' = (\underline{x}^{(1)'}, \underline{x}^{(2)'}) = (\underline{x}_1, \dots, \underline{x}_{p_1}, \underline{x}_{p_1+1}, \dots, \underline{x}_p)'$ ;

$\mu' = (\mu^{(1)'}, \mu^{(2)'})$  en  $\Sigma^2 = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ . Hierin is dus  $\Sigma_{11} = \text{VAR } \underline{x}^{(1)}$ ,

$\Sigma_{22} = \text{VAR } \underline{x}^{(2)}$  (beide regulier ondersteld) en  $\Sigma_{12} = \text{COV}(\underline{x}^{(1)}, \underline{x}^{(2)}) = \Sigma_{21}'$ .

Neem nu  $A_{p_1 \times p} = \begin{pmatrix} I_{p_1 \times p_1} & 0 \end{pmatrix}$  dan volgt  $\underline{x}^{(1)} = A\underline{x} \sim N_{p_1}(\mu^{(1)}, \Sigma_{11})$ .

(Merk op dat in  $\Sigma_{11}$  het kwadraat ontbreekt; eventueel te schrijven als  $\Sigma_{11}^2$ .)

### 3.9. Theorema

Is  $\underline{x} \sim N_p(\mu, \Sigma^2)$ , dan geldt:  $\underline{x}^{(1)}, \underline{x}^{(2)}$  ongecorreleerd  $\Leftrightarrow \underline{x}^{(1)}, \underline{x}^{(2)}$  stochastisch onafhankelijk.

Bewijs: Dit volgt direkt uit de karakteristieke functie, nl.  $\Sigma_{12} = \Sigma_{21}' = 0$  geeft:

$$\begin{aligned} \varphi_{\underline{x}}(t) &= \exp[it^{(1)'}\mu^{(1)} + it^{(2)'}\mu^{(2)} - \frac{1}{2}t^{(1)'}\Sigma_{11}t^{(1)} - \frac{1}{2}t^{(2)'}\Sigma_{22}t^{(2)}] = \\ &= \varphi_{\underline{x}^{(1)}}(t^{(1)})\varphi_{\underline{x}^{(2)}}(t^{(2)}), \text{ dus } \underline{x}^{(1)} \text{ en } \underline{x}^{(2)} \text{ stochastisch onafhankelijk} \\ &\quad \text{(zie eind 2.44).} \end{aligned}$$

Hierin is  $t' = (t^{(1)'}, t^{(2)'}) = (t_1, \dots, t_{p_1}, t_{p_1+1}, \dots, t_p)'$ .

Pas op: ongecorreleerde, simultaannormaal verdeelde variabelen zijn onafhankelijk. Zijn ze echter niet simultaannormaal verdeeld dan is dat niet zonder meer juist (zie appendix A).

### 3.10. Theorema

Is  $\underline{x} \sim N_p(\mu, \Sigma^2)$  dan is de *voorwaardelijke verdeling* van  $\underline{x}^{(1)}$  bij gegeven  $\underline{x}^{(2)}$

$$f(\underline{x}^{(1)} | \underline{x}^{(2)}) = N_{p_1}[\mu^{(1)} + \beta(\underline{x}^{(2)} - \mu^{(2)}), \Sigma_{11 \cdot 2}]$$

met  $\beta := \Sigma_{12}\Sigma_{22}^{-1}$  en  $\Sigma_{11 \cdot 2} := \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$  (zie 1.35).

$\beta$  heet de *matrix der regressiecoëfficiënten* van  $\underline{x}^{(1)}$  op  $\underline{x}^{(2)}$  (beter  $\beta_{12}$ ).

Bewijs: Voer een reguliere transformatie uit  $\underline{y} = A\underline{x}$  zodat  $\underline{y}^{(1)}$  en  $\underline{y}^{(2)}$  ongecorrleerd zijn.

$$\begin{aligned} \underline{y}^{(1)} &= \underline{x}^{(1)} - \beta \underline{x}^{(2)} \\ \underline{y}^{(2)} &= \underline{x}^{(2)} \end{aligned} \quad \text{dus} \quad A = \begin{pmatrix} I_{p_1 \times p_1} & -\beta \\ 0 & I \end{pmatrix}; \quad |A| = 1.$$

Dus

$$\text{COV}(\underline{x}^{(1)} - \beta \underline{x}^{(2)}, \underline{x}^{(2)}) = \Sigma_{12} - \beta \Sigma_{22} = 0 \quad \text{oftewel} \quad \beta = \Sigma_{12} \Sigma_{22}^{-1}.$$

$$\text{VAR} \underline{y}^{(1)} = \Sigma_{11} + \beta \Sigma_{22} \beta' - \beta \Sigma_{21} - \Sigma_{12} \beta' = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} = \Sigma_{11 \cdot 2}$$

en

$$\xi_{\underline{y}}^{(1)} = \mu^{(1)} - \beta \mu^{(2)}.$$

Nu is  $\underline{x}$ , dus volgens 3.3 ook  $\underline{y}$ , eveneens  $\underline{y}^{(1)}$  volgens 3.8 normaal verdeeld.  $\underline{y}^{(1)}$  en  $\underline{y}^{(2)}$  zijn ongecorrleerd, dus volgens 3.9 ook onafhankelijk, dus  $f(\underline{y}^{(1)} | \underline{y}^{(2)}) = f(\underline{y}^{(1)})$ .

Daar  $f(\underline{x}^{(1)} | \underline{x}^{(2)}) = f(\underline{y}^{(1)} + \beta \underline{x}^{(2)} | \underline{x}^{(2)})$  is  $f$  een normale verdeling met  $\xi[\underline{x}^{(1)} | \underline{x}^{(2)}] = \xi_{\underline{y}}^{(1)} + \beta \underline{x}^{(2)} = \mu^{(1)} - \beta \mu^{(2)} + \beta \underline{x}^{(2)}$  en  $\text{VAR}[\underline{x}^{(1)} | \underline{x}^{(2)}] = \text{VAR} \underline{y}^{(1)} = \Sigma_{11 \cdot 2}$ . Dus

$$f(\underline{x}^{(1)} | \underline{x}^{(2)}) = N_{p_1}[\mu^{(1)} + \beta(\underline{x}^{(2)} - \mu^{(2)}), \Sigma_{11 \cdot 2}].$$

Enkele opmerkingen:

a) We zien inderdaad, indien  $\Sigma_{12} = 0$ , dus  $\beta = 0$  en  $\Sigma_{11 \cdot 2} = \Sigma_{11}$ , dat  $f(\underline{x}^{(1)} | \underline{x}^{(2)}) = N_{p_1}(\mu^{(1)}, \Sigma_{11}) = f(\underline{x}^{(1)})$  ofwel ongecorrleerd  $\Leftrightarrow$  onafhankelijk.

b) Het geval  $p = 2$  geeft:

3.11. 
$$f(x_1 | x_2) = N \left[ \mu_1 + \frac{\sigma_1}{\sigma_2} \rho (x_2 - \mu_2), \sigma_1^2 (1 - \rho^2) \right].$$

c)  $\text{VAR}[\underline{x}^{(1)} | \underline{x}^{(2)}] = \Sigma_{11 \cdot 2}$  is *onafhankelijk* van het niveau van  $\underline{x}^{(2)}$ . Dit is essentieel en geldt niet zonder meer voor andere verdelingen.

3.12. d) Is  $\mu = 0$ , dan is  $E[\underline{x}^{(1)} | x^{(2)}] = \beta x^{(2)}$  en  $E[\underline{x}_i | x^{(2)}] = \beta_{i*} x^{(2)}$ .

Stellen we nu  $\underline{x}_i = \underline{y}$ ,  $x^{(2)} = (x_0, \dots, x_p)'$  en  $\beta_{i*} = \beta' = (\beta_0, \dots, \beta_p)$ , dan wordt dit juist

$$E[\underline{y} | x] = \beta' x = \beta_0 x_0 + \dots + \beta_p x_p$$

of ook geschreven

$$\underline{y} = \beta' x + \underline{e} \quad \text{met } E\underline{e} = 0 \quad \text{en } \text{var } \underline{e} = \sigma^2 \text{ (onafhank. } x).$$

Dit is een rechtvaardiging voor het feit dat dit *lineaire model* in de multiple regressie analyse vaak wordt toegepast. In de regressie analyse schat men o.a. de *regressiecoëfficiënten*  $\beta_{i*}$ , dat is dus één rij uit de *matrix der regressiecoëfficiënten*  $\beta$  (zie § 5 multivariate analyse).

e)  $\Sigma_{11.2} > 0$  nl. het is een variantie-covariantie matrix, dus zeker  $\Sigma_{11.2} \geq 0$ .

Echter  $\Sigma_{11.2}$  is regulier als  $\Sigma^2$  regulier is, nl

$$0 < |\Sigma^2| = |\Sigma_{22}| |\Sigma_{11.2}| \quad \text{dus } |\Sigma_{11.2}| \neq 0$$

f) Er geldt  $\Sigma_{11.2} \leq \Sigma_{11}$ .

§ 4. Correlatie

Alvorens dit onderwerp te behandelen is het verhelderend om verband te leggen tussen de theorie der lineaire vektorruimten en de multivariate analyse.

a. Vektorruimten van stochastische variabelen

Beschouw de stochastische variabelen  $\underline{x}_1, \dots, \underline{x}_p$  en onderstel  $\text{var } \underline{x}_i < \infty$  en  $\mathcal{E}\underline{x}_i = 0, i = 1, \dots, p$ .

Uit de maattheoretische kansrekening weten we dat  $\underline{x}_1, \dots, \underline{x}_p$  een lineaire deelruimte  $D$  van de lineaire vektorruimte  $\mathcal{L}_2$  (dit is de verzameling van alle stochastische variabelen met eindig tweede moment) opspannen.

In deze  $D$  kunnen we als volgt een inproduct  $(\underline{x}, \underline{y})$  definiëren:

$$(\underline{x}, \underline{y}) := \mathcal{E}(\underline{xy}) = \text{cov}(\underline{x}, \underline{y}) .$$

Opmerking:  $\text{cov}(\underline{x}, \underline{x}) = \text{var } \underline{x}$ .

Dit bestaat voor elk paar  $\underline{x}, \underline{y} \in D$  en definieert inderdaad een inproduct (gana). Hierbij identificeren we  $\underline{x}$  met 0 als  $\underline{x} = 0$  "bijna overal".

$$\|\underline{x}\|^2 = \mathcal{E}\underline{x}^2 = \text{var } \underline{x} \quad \text{en} \quad \cos \varphi(\underline{x}, \underline{y}) = \frac{\mathcal{E}(\underline{xy})}{\sqrt{\mathcal{E}\underline{x}^2 \mathcal{E}\underline{y}^2}} = \rho(\underline{x}, \underline{y}) .$$

Eveneens kunnen we een metriek invoeren:

$$d(\underline{x}, \underline{y}) := \|\underline{x} - \underline{y}\| = \sqrt{\text{var}(\underline{x} - \underline{y})} .$$

Wat is het verdere verband?

Evenals het inproduct in een vektorruimte volledig bepaald is door de *inproductmatrix*  $\{(x_i, x_j)\}$ , dat zijn alle inproducten der basisvektoren, zo is ook  $\text{cov}(\underline{x}, \underline{y})$  voor alle  $\underline{x}, \underline{y}$  vastgelegd door  $\Sigma^2$ . Dus de variantie-covariantiematrix  $\Sigma^2$  is niets anders dan de inproductmatrix der basisvektoren  $\underline{x}_1, \dots, \underline{x}_p$ .

Analoog aan

$$\underline{x} = \sum_1^p \alpha_i \underline{x}_i, \quad \underline{y} = \sum_1^p \beta_j \underline{x}_j, \quad (\underline{x}, \underline{y}) = \sum_{ij} \alpha_i \beta_j (x_i, x_j)$$

geldt

$$\text{cov}(\underline{x}, \underline{y}) = \sum_{ij} \alpha_i \beta_j \text{cov}(\underline{x}_i, \underline{x}_j) .$$

De stelling uit de vektorrekening  $|\{(x_i, x_j)\}| = 0 \iff x_1, \dots, x_p$  lineair afhankelijk geeft als analogon  $|\Sigma^2| = 0 \iff \underline{x}_1, \dots, \underline{x}_p$  lineair afhankelijk

(almost sure). Is dus  $|\Sigma^2| \neq 0$ , dan spannen  $\underline{x}_1, \dots, \underline{x}_p$  een  $p$ -dimensionale lineaire deelruimte op.

*Ongecorrleerd* wordt dus *orthogonaal* en een stelsel *ongecorreleerde, gestandaardiseerde* stochastische variabelen vormen dus een *orthonormaal* stelsel vektoren.

Pas op: een *orthogonaal* stelsel vektoren is lineair onafhankelijk; een stelsel *ongecorreleerde* variabelen is dus ook *lineair onafhankelijk* (in vektortheoretische zin), doch hoeft niet *stochastisch onafhankelijk* te zijn (in statistische zin), m.a.w.:

Uit *stochastische* onafhankelijkheid volgt *lineaire* onafhankelijkheid, niet omgekeerd. Een stelsel stochastische variabelen is nl. *lineair* afhankelijk als er minstens één variabele onder is die met de overige een multiple correlatiecoëfficiënt 1 heeft (zie § 4.c) (kort gezegd  $\rho = 0 \Rightarrow \rho \neq 1$ , niet omgekeerd).

De *ongelijkheid van Cauchy-Schwarz*  $(\underline{x}, \underline{y})^2 \leq (\underline{x}, \underline{x})(\underline{y}, \underline{y})$  wordt nu:

$$\text{cov}^2(\underline{x}, \underline{y}) \leq \text{var } \underline{x} \text{ var } \underline{y} .$$

Hieruit volgt direkt dat  $|\rho| \leq 1$  en  $\rho = \pm 1$  als  $\underline{y} = \alpha \underline{x}$ .

Opmerking. Dit geldt ook als  $\mu_x, \mu_y \neq 0$  (volgt uit 2.34).

### Toepassing

Stelling.  $\underline{x} > 0$ ;  $\mathcal{E}\left(\frac{1}{\underline{x}}\right) \geq \frac{1}{\mathcal{E}\underline{x}}$ . (Jensens ongelijkheid:  $\mathcal{E}f(\underline{x}) \geq f(\mathcal{E}\underline{x})$  als  $f$  convex.)

Bewijs. Definieer  $\underline{u} := 1/\sqrt{\underline{x}}$  en  $\underline{v} := \sqrt{\underline{x}}$ . Dan is  $1 = [\mathcal{E}(\underline{u}\underline{v})]^2 \leq \mathcal{E}\frac{1}{\underline{x}} \mathcal{E}\underline{x}$ .

Tot zover betreffende een *populatie* van stochastische variabelen.

Stel we hebben een *steekproef*  $X_{n \times p}$ ,  $n$  realisaties van elk der  $p$  variabelen  $\underline{x}_j$ :  $\underline{x}_{*j} = (x_{1j}, \dots, x_{nj})'$ ,  $j = 1, \dots, p$ . Dit zijn dus  $p$  vektoren uit de  $R^n$ . Als inproduct definiëren we:  $(\underline{x}_{*i}, \underline{x}_{*j}) := \underline{x}_{*i}' \underline{x}_{*j}$ . Is er reeds gecorrigeerd voor het gemiddelde, d.w.z. is  $\bar{x}_{.j} = \sum_i x_{ij}/n = 0$ ,  $j = 1, \dots, p$ , dan is  $s^2(\underline{x}_j) = \|\underline{x}_{*j}\|^2/v$  ( $v =$  dimensie bijbehorende deelruimte).

b. Toepassing

In 3.12 zagen we  $\xi[\underline{x}^{(1)} | \underline{x}^{(2)}] = \underline{\beta}_x^{(2)}$  (weer  $\mu = 0$  ondersteld).

Noem  $\underline{\beta}_x^{(2)} = \underline{\hat{x}}_i^{(1)}$  dus  $\underline{\hat{x}}_i = \beta_{i \star x}^{(2)}$ ,  $i = 1, \dots, p_1$ . Dan is

$$\text{COV}(\underline{x}^{(1)} - \underline{\hat{x}}_i^{(1)}, \underline{x}^{(2)}) = \text{COV}(\underline{y}^{(1)}, \underline{y}^{(2)}) = 0,$$

dus

$$\underline{x}_i - \underline{\hat{x}}_i \perp \underline{x}_j; \quad i = 1, \dots, p_1, j = p_1 + 1, \dots, p.$$

Dus  $\underline{\hat{x}}_i$  is juist de orthogonale projectie van  $\underline{x}_i$  op de deelruimte opgespannen door de componenten van  $\underline{x}^{(2)}$ . De afstand van  $\underline{x}_i$  tot deze deelruimte is

$$\|\underline{x}_i - \underline{\hat{x}}_i\| = \sqrt{\text{var}(\underline{x}_i - \underline{\hat{x}}_i)}.$$

$\rho(\underline{x}_i, \underline{\hat{x}}_i) = \cos \varphi(\underline{x}_i, \underline{\hat{x}}_i)$  is maximaal, d.w.z. er is geen andere lineaire combinatie  $\alpha' \underline{x}^{(2)}$  die met  $\underline{x}_i$  hoger gecorreleerd is.

Daarom wordt  $\underline{\hat{x}}_i$  de *beste lineaire benadering* van  $\underline{x}_i$  genoemd.  $\underline{x}_i - \underline{\hat{x}}_i$  wordt het *residu* van  $\underline{x}_i$  genoemd.

Er geldt:

$$4.1. \quad \text{VAR} \underline{\hat{x}}_i^{(1)} = \text{VAR} \underline{\beta}_x^{(2)} = \underline{\beta} \underline{\Sigma}_{22}^{-1} \underline{\beta}' = \underline{\Sigma}_{12} \underline{\Sigma}_{22}^{-1} \underline{\Sigma}_{21} =: \underline{\Sigma}_{1 \cdot 2}$$

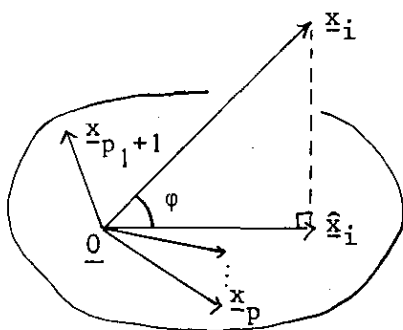
$$\text{VAR}(\underline{x}^{(1)} - \underline{\hat{x}}_i^{(1)}) = \text{VAR} \underline{y}^{(1)} = \underline{\Sigma}_{11 \cdot 2} \quad (\text{zie 3.10})$$

$$\text{COV}(\underline{x}^{(1)}, \underline{\hat{x}}_i^{(1)}) = \text{COV}(\underline{x}^{(1)}, \underline{\beta}_x^{(2)}) = \underline{\Sigma}_{12} \underline{\beta}' = \underline{\Sigma}_{12} \underline{\Sigma}_{22}^{-1} \underline{\Sigma}_{21} = \underline{\Sigma}_{1 \cdot 2}.$$

$\underline{\Sigma}_{1 \cdot 2}$  = variantie-covariantiematrix der beste lineaire benaderingen.

$\underline{\Sigma}_{11 \cdot 2}$  = variantie-covariantiematrix der residuen.

Nog eens meetkundig bekeken:



$$\|\underline{x}_i\|^2 = \|\underline{x}_i - \underline{\hat{x}}_i\|^2 + \|\underline{\hat{x}}_i\|^2 \quad \text{oftewel}$$

$$(\underline{\Sigma}_{11})_{ii} = (\underline{\Sigma}_{11 \cdot 2})_{ii} + (\underline{\Sigma}_{1 \cdot 2})_{ii} \quad (\text{ook voor } ij \text{ geldig}).$$

Laten we na deze inleiding de verschillende correlatiecoëfficiënten definiëren.



c. De enkelvoudige, partiële en multiple correlatiecoëfficiënt

Het begrip "x en y zijn stochastisch onafhankelijk" is gedefinieerd en hiermee ook het begrip afhankelijk (= niet onafhankelijk). Als maat voor afhankelijkheid hadden we de correlatiecoëfficiënt  $\rho(\underline{x}, \underline{y})$ .

Hiermee moet men echter zeer kritisch te werk gaan.

"Thurstone, one of the chief architects of modern factor analysis, was obliged to call the correlation coefficient a symbol of complete ignorance."

Wat is nl. het geval?

De correlatiecoëfficiënt  $\rho(\underline{x}, \underline{y})$  is een maat voor *lineaire* afhankelijkheid.  $\rho = 0$  betekent dus "niet lineair afhankelijk" dus onafhankelijk in vektor-theoretische zin, doch x en y kunnen zeker wel stochastisch afhankelijk zijn. Dus de eerste fout die men vaak maakt is, dat ongecorreleerd zou betekenen onafhankelijk, geen causaal verband aanwezig. Dit is niet juist.

Een voorbeeld:

Zij x het aantal ogen geworpen met een zuivere dobbelsteen en y het aantal ogen dat naar de speler toegekeerd ligt. Dan is  $\rho(\underline{x}, \underline{y}) = 0$  terwijl x en y toch afhankelijk zijn (ga na).

De tweede fout die men vaak maakt is, dat een hoge correlatie een sterk causaal verband zou aangeven tussen beide variabelen. Zo ontstaat de bekende "nonsens correlatie".

Een voorbeeld:

De prijs van de rumbonen en het salaris van de dominëes in de loop der jaren is hoog gecorreleerd. Daarentegen: het aantal gevallen longkanker is hoog gecorreleerd met het aantal sigaretten dat men rookt. Men voelt aan (of neemt aan) dat het eerste voorbeeld nonsens correlatie geeft, het tweede echter niet.

Waarin zit het verschil? Hierop komen we later terug.

Zij nu  $\underline{x}' = (\underline{x}^{(1)'}, \underline{x}^{(2)'})$ ,  $\Sigma_{11}$ , etc. als in 3.8. We definiëren nu:

4.2. De (enkelvoudige) correlatiecoëfficiënt tussen  $\underline{x}_i$  en  $\underline{x}_j \in \underline{x}^{(1)}$  is

$$\rho_{ij} = \frac{(\Sigma_{11})_{ij}}{\sqrt{(\Sigma_{11})_{ii}(\Sigma_{11})_{jj}}} \quad (= \rho(\underline{x}_i, \underline{x}_j)) .$$

4.3. De *partiële* correlatiecoëfficiënt tussen  $\underline{x}_i$  en  $\underline{x}_j$  bij gegeven  $\underline{x}^{(2)}$  is de enkelvoudige correlatiecoëfficiënt tussen  $\underline{x}_i - \hat{\underline{x}}_i$  en  $\underline{x}_j - \hat{\underline{x}}_j$

$$\rho_{ij \cdot 2} = \frac{(\Sigma_{11 \cdot 2})_{ij}}{\sqrt{(\Sigma_{11 \cdot 2})_{ii}(\Sigma_{11 \cdot 2})_{jj}}} \quad (= \rho(\underline{x}_i - \hat{\underline{x}}_i, \underline{x}_j - \hat{\underline{x}}_j)) .$$

4.4. De *multiple* correlatiecoëfficiënt tussen  $\underline{x}_i \in \underline{x}^{(1)}$  en  $\underline{x}^{(2)}$  is de enkelvoudige correlatiecoëfficiënt tussen  $\underline{x}_i$  en  $\hat{\underline{x}}_i$

$$\rho_{i \cdot 2} = \frac{(\Sigma_{1 \cdot 2})_{ii}}{\sqrt{(\Sigma_{11})_{ii}(\Sigma_{1 \cdot 2})_{ii}}} \quad (= \rho(\underline{x}_i, \hat{\underline{x}}_i)) .$$

Laten we deze eens nader bekijken:

(1) Zijn  $\underline{x}_i$  en  $\underline{x}_j$  ongecorreleerd met  $\underline{x}^{(2)}$ , d.w.z.  $(\Sigma_{12})_{i*} = (\Sigma_{12})_{j*} = (0, \dots, 0)'$ , dan is dus  $\rho_{ij \cdot 2} = \rho_{ij}$  en  $\rho_{i \cdot 2} = \rho_{j \cdot 2} = 0$ .

(2) Zij  $p = 3$ ,  $\underline{x}^{(1)} = (\underline{x}_1, \underline{x}_2)'$  en  $\underline{x}^{(2)} = \underline{x}_3$ , dan is

$$\Sigma_{11 \cdot 2} = \begin{pmatrix} \sigma_{11}^2 - \sigma_{13}^2 \sigma_{23}^2 / \sigma_{33}^2 & \sigma_{12}^2 - \sigma_{13}^2 \sigma_{23}^2 / \sigma_{33}^2 \\ \sigma_{12}^2 - \sigma_{13}^2 \sigma_{23}^2 / \sigma_{33}^2 & \sigma_{22}^2 - \sigma_{23}^2 \sigma_{33}^2 / \sigma_{33}^2 \end{pmatrix}$$

en

4.5. 
$$\rho_{12 \cdot 3} = \frac{\rho_{12} - \rho_{13} \rho_{23}}{\sqrt{(1 - \rho_{13}^2)(1 - \rho_{23}^2)}} .$$

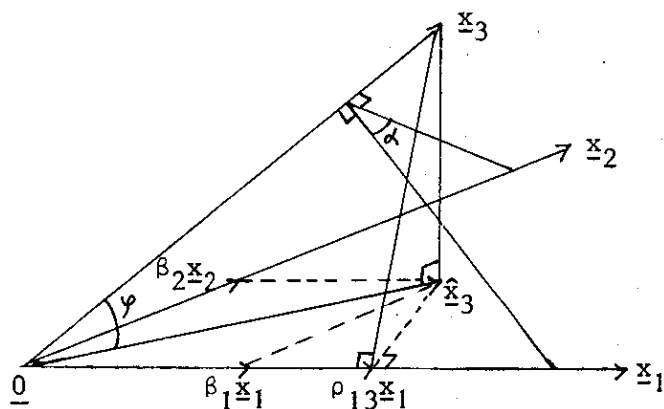
(3)  $\rho_{i \cdot 2} \geq \rho_{i\ell}$  voor  $\ell = p_1 + 1, \dots, p$  en tevens  $\rho_{i \cdot 2} \geq 0$ , nl.  $\rho_{i \cdot 2} = \cos(\underline{x}_i, \hat{\underline{x}}_i)$  en deze is maximaal en altijd  $\geq 0$  (ga na).

$\rho_{i \cdot 2} = 1$  indien  $\underline{x}_i$  dus in het projektievlak ligt oftewel als  $\underline{x}_i = \alpha' \underline{x}^{(2)}$  (zie opmerking § 4.a).

(4) Er geldt  $\text{var}(\underline{x}_i | \underline{x}^{(2)}) = (1 - \rho_{i \cdot 2}^2) \sigma_{ii}^2 \leq \sigma_{ii}^2$  (vergelijk met 3.11) nl. meetkundig in te zien:

$$\frac{\text{var}(\underline{x}_i | \underline{x}^{(2)})}{\text{var} \underline{x}_i} = \frac{\|\underline{x}_i - \hat{\underline{x}}_i\|^2}{\|\underline{x}_i\|^2} = \sin^2 \varphi = 1 - \cos^2 \varphi = 1 - \rho_{i \cdot 2}^2 .$$

In de volgende figuur zien we voor  $p = 3$  een meetkundige interpretatie van de resp. correlatiecoëfficiënten en hun onderlinge samenhang. Ondersteld is dat  $\bar{x}_i = 0$  en  $\text{var } x_i = 1$ ,  $i = 1, 2, 3$ .



$$\rho_{12} = \cos(x_1, x_2)$$

$$\rho_{13} = \cos(x_1, x_3)$$

$$\rho_{23} = \cos(x_2, x_3)$$

$$\rho_{12 \cdot 3} = \cos \alpha$$

$$\rho_{3 \cdot 12} = \cos \varphi = \cos(x_3, \hat{x}_3)$$

$$\hat{x}_3 = \beta_1 x_1 + \beta_2 x_2 .$$

(5) We zien in de figuur dat, indien  $x_3 \perp$  vlak, d.w.z.  $\rho_{13} = \rho_{23} = 0$ ,  $\rho_{3 \cdot 12} = 0$  en  $\rho_{12 \cdot 3} = \rho_{12}$  (zie ook (1)).

Is  $\rho_{12} = 0$  (dus  $x_1 \perp x_2$ ) dan is  $\beta_1 = \rho_{13}$  en  $\beta_2 = \rho_{23}$ . Dit klopt met

$$\beta = \Sigma_{12} \Sigma_{22}^{-1} = \Sigma_{12} .$$

(6) Meetkundig is te bewijzen (ga na):

$$\cos \alpha = \frac{\cos(x_1, x_2) - \cos(x_1, x_3)\cos(x_2, x_3)}{\sin(x_1, x_3)\sin(x_2, x_3)} , \text{ juist formule 4.5.}$$

Eveneens is meetkundig te bewijzen (ga na):

$$4.6. \quad \rho_{3 \cdot 12}^2 = \frac{\rho_{13}^2 + \rho_{23}^2 - 2\rho_{13}\rho_{23}\rho_{12}}{1 - \rho_{12}^2} .$$

(7) Uit de figuur zien we duidelijk dat  $\rho_{12}$  bijna +1 kan zijn, terwijl  $\rho_{12 \cdot 3}$  nul of zelfs negatief is (dit doet zich voor indien  $x_3$  dicht bij het vlak ligt oftewel  $\rho_{3 \cdot 12}$  groot is).

Een aardig voorbeeld geeft Anderson [1], pag. 82.

Gedurende 20 jaar heeft men ergens in Engeland de hooiopbrengst ( $x_1$ ), de regenval ( $x_2$ ) en de temperatuur ( $x_3$ ) gemeten voor een hooiland.

De berekende correlatiematrix was:

$$\begin{pmatrix} 1 & 0.80 & -0.40 \\ 0.80 & 1 & -0.56 \\ -0.40 & -0.56 & 1 \end{pmatrix} .$$

Opbrengst en temperatuur zijn negatief gecorreleerd ( $\hat{\rho}_{13} = -0.40$ ).  
 Betekent dit dat hogere temperaturen lagere opbrengsten geven? Er blijkt echter dat  $\hat{\rho}_{13.2} = 0.097$ , d.w.z. na eliminatie van het "regeneffekt" zijn opbrengst en temperatuur wel positief gecorreleerd. De conclusie is dat hogere regenval en hogere temperatuur de opbrengst vergroten, doch dat meestal bij hogere temperatuur een lagere regenval behoort. Hiermee kan men vaak de eerder genoemde "nonsens correlatie" verklaren. In het geval van de rumbonen zal de partiële correlatie, na eliminatie van de tijd, wel ongeveer 0 zijn.

4.7. Opm.: Ga na dat geldt:  $\prod_2^p (1 - \rho^2_{i.1, \dots, i-1}) = |R|$

oftewel:  $\prod_1^p \sigma^2_{i.1, \dots, i-1} = |\Sigma^2|$  met  $\sigma^2_{1.0} = \sigma^2_1$

Gevolg 1:  $|\Sigma^2| \leq \prod_1^p \sigma^2_i$

2:  $0 \leq |R| \leq 1$

3:  $\sigma^2_{p.1, \dots, p-1} = \frac{|\Sigma^2|}{|\Sigma^2|_{pp}} = 1/(\Sigma^2)^{pp}$ .

4.8. Er geldt:  $\Sigma_{11.2} = [(\Sigma^{-2})_{11}]^{-1}$ .

Bewijs:  $I = \Sigma^2 \Sigma^{-2} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} A & -DB \\ -CA & B \end{pmatrix}$ , dus  $\begin{matrix} (\Sigma_{11} - \Sigma_{12}C)A = I \\ (\Sigma_{21} - \Sigma_{22}C)A = 0 \end{matrix}$ .

Dus  $C = \Sigma_{22}^{-1} \Sigma_{21} = B'_{12}$ ;  $A = (\Sigma_{11.2})^{-1} = (\Sigma^{-2})_{11}$ . Analoo

$D = \Sigma_{11}^{-1} \Sigma_{12} = B'_{21}$ ;  $B = (\Sigma_{22.1})^{-1}$ .

Gevolg:  $\Sigma^{-2} = \begin{pmatrix} (\Sigma_{11.2})^{-1} & -B'_{21}(\Sigma_{22.1})^{-1} \\ -B'_{12}(\Sigma_{11.2})^{-1} & (\Sigma_{22.1})^{-1} \end{pmatrix}$

§ 5. Meervoudige lineaire regressie

- a. Het onderwerp regressie-analyse is zo uitgebreid dat we er hier slechts even op in kunnen gaan. Meestal wordt dit onderwerp, evenals de variantie-analyse, apart behandeld en niet opgenomen in de multivariate analyse. Voor verdere studie wordt verwezen naar het duidelijke boek van Draper en Smith [3].

Zij gegeven het model  $\underline{y} = \eta + \underline{e} = \beta'x + \underline{e}$ ;  $E\underline{e} = 0$ ,  $\text{var } \underline{e} = \sigma^2$ . Ook vaak geschreven als  $E(\underline{y} | x) = \beta'x$  (zie 3.12). Hierin is

$x = (x_0, x_1, \dots, x_p)'$  de vektor der *verklarenden*;  $x_0 = 1$ ;

$y$  de te *verklaren*, afhankelijke variabele;

$\underline{e}$  het *residu*;

$\beta' = (\beta_0, \dots, \beta_p)$  de *parametervektor*, vektor der *regressiecoëfficiënten*.

In een chemisch proces is bijvoorbeeld de opbrengst  $y$  afhankelijk van de druk  $x_1$ , de concentratie  $x_2$  en de temperatuur  $x_3$ .

*Essentieel* voor het vervolg is, dat  $x$  niet stochastisch ondersteld is. In de praktijk betekent dit dat de fout in  $x$  te verwaarlozen is t.o.v. de fout in  $y$ . Het geval dat ook  $x$  stochastisch is, is veel ingewikkelder en beschouwen we hier niet.

De *lineaire* regressie-analyse behandelt *lineaire* modellen, dat wil zeggen lineair in de *parameters*. Zo zijn dus ook lineair:

$$\underline{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \underline{e} \quad \text{en} \quad \underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_2^2 + \underline{e}.$$

Sommige zijn direkt door een transformatie lineair te maken, bv.

$\underline{y} = \alpha x_1^{\beta_1} x_2^{\beta_2} \underline{e}$  geeft  $\ln \underline{y} = \ln \alpha + \beta_1 \ln x_1 + \beta_2 \ln x_2 + \ln \underline{e}$ . Deze noemt men "intrinsically linear". Een voorbeeld van "intrinsically nonlinear",

d.w.z. niet door transformatie lineair te maken is  $\underline{y} = \beta_0 + \beta_1 e^{-\beta_2 x} + \underline{e}$ . Deze beschouwen we echter niet.

Het doel van de regressie-analyse kan o.a. zijn dat men  $y$  als functie van de  $x_i$  wil kennen om te gebruiken bij interpoleren of calibreren. Een rechtvaardiging voor het feit dat men vaak lineaire regressie toepast zagen we reeds in 3.12. Daar was  $E(\underline{y} | x) = \beta'x$ . Bovendien geldt zelfs dat  $\text{var}(\underline{y} | x)$  onafhankelijk van de waarde van  $x$  is, een aanname die we eveneens zullen maken.

Het probleem is nu uit waarnemingen, schattingen te vinden voor de modelparameters  $\beta$  (en  $\eta$ ) en  $\sigma^2$ , om hiermee het regressie-oppervlak (de regressielijn)  $\hat{\eta} = \underline{b}'x$  te bepalen inclusief de betrouwbaarheden van deze schatters. Hiertoe doet men aan elke variabele  $n$  metingen (of instellingen) en zo ontstaat het *steekproefmodel*:  $\underline{y} = \eta + \underline{e} = X\beta + \underline{e}$ ,  $\underline{E}\underline{e} = 0$ . Hierin is nu

$\underline{y} = (y_1, \dots, y_n)'$  de *response vektor*;

$\underline{e} = (e_1, \dots, e_n)'$  de *storingsvektor (random disturbance)*

en

$$X_{n \times (p+1)} = \begin{pmatrix} x_{10} & x_{11} & \dots & x_{1p} \\ \vdots & \vdots & & \vdots \\ x_{n0} & x_{n1} & & x_{np} \end{pmatrix} \quad \text{de matrix der waarnemingen (instellingen)} \\ \text{(design matrix); } x_{*0} = (1, \dots, 1)'$$

Tevens nemen we voorlopig aan dat  $\text{VAR } \underline{e} = \sigma^2 I$ , d.w.z. de fouten  $e_i$  zijn onderling ongecorreleerd en hebben dezelfde variantie. Onder  $c$  laten we deze aanname vallen.

Stel deze schatters zijn resp.  $\underline{b}$  (en  $\hat{\eta} = \underline{b}'x$  voor  $\eta$ ) en  $\underline{s}^2$ . Natuurlijk kunnen we op verschillende manieren schatters vinden. Een gebruikelijke methode is die der *kleinste kwadraten (least squares method)*. Hierbij wordt  $\underline{b}$  zó bepaald dat

$$5.1. \quad \|\underline{y} - \hat{\underline{y}}\|^2 = \|\underline{y} - X\underline{b}\|^2 = \min_{\underline{b}} \|\underline{y} - X\underline{b}\|^2 = \min(\underline{e}'\underline{e}) .$$

5.2. Voor  $\underline{s}^2$  wordt dan genomen:  $\underline{s}^2 = \|\underline{y} - \hat{\underline{y}}\|^2 / (n - p - 1)$ , de *residuele variantie*.  $\underline{b}$  heet in dit geval de kleinste kwadratenschatter voor  $\beta$ .

Neemt men echter aan dat  $\underline{e} \sim N_n(0, \sigma^2 I)$  dan is  $\underline{b}$  tevens de meest aannemelijke schatter voor  $\beta$ , nl.

$$\begin{aligned} \max L(\beta) &= \max \prod_i f(y_i; \beta) = \max \prod_i \frac{1}{\sigma\sqrt{2\pi}} \exp(-e_i^2/2\sigma^2) = \\ &= \max \sigma^{-n} (2\pi)^{-\frac{1}{2}n} \exp(-\underline{e}'\underline{e}/2\sigma^2) \end{aligned}$$

oftewel maximaal voor  $\underline{e}'\underline{e}$  minimaal. Dit is dus een rechtvaardiging voor het gebruik van de methode der kleinste kwadraten en berust op de (vaak reële) aanname dat  $\underline{e} \sim N_n(0, \sigma^2 I)$ .

De oplossing volgt direkt uit 1.72:  $\hat{\underline{y}} = P_{\underline{X}} \underline{y}$  waarin  $P_{\underline{X}} := X(X'X)^{-1}X'$  de orthogonale projectie is op de kolomruimte van  $X$ , en dus

$$5.3. \quad \underline{b} = (X'X)^{-1}X'\underline{y} .$$

Wil  $X'X$  regulier zijn, dan moet  $r(X'X) = r(X) = p + 1$ , dus zeker  $n > p$ . In de statistiek is dat meestal te verwezenlijken, zelfs neemt men vaak de  $x$ -waarnemingen zó dat ze orthogonaal zijn. In de econometrie is dat niet altijd mogelijk. Daar is  $X'X$  vaak singulier. In dat geval werkt men met een *gegeneraliseerde* (pseudo-) *inverse* (zie Rao [7]), hetgeen in dit kader te ver voert.

We zien dat, als  $\underline{e}$  normaalverdeeld is, dan ook  $\underline{y}$  en volgens 3.3 ook  $\underline{b}$  en  $\hat{\underline{y}}$ .

5.4.  $X'Xb = X'y$  oftewel  $X'(\hat{\underline{y}} - \underline{y}) = 0$  worden de *normaalvergelijkingen* genoemd.

Hieruit volgt dat  $\hat{\underline{y}} - \underline{y}$  loodrecht elke kolom van  $X$  is, dus ook  $\perp \underline{x}_0 = (1, \dots, 1)'$  hetgeen betekent dat  $\sum (\hat{y}_i - y_i) = 0$  oftewel de *som der residuen is nul*. Het stelsel normaalvergelijkingen is *consistent*, ook als  $X'X$  singulier is (zie 1.50).

$\underline{b}$  is een BLUE voor  $\beta$  (zie 2.16). We zullen de zuiverheid bewijzen:

$$\mathcal{E}\underline{b} = (X'X)^{-1}X'\mathcal{E}\underline{y} = (X'X)^{-1}X'X\beta = \beta \text{ dus } \underline{b} \text{ is zuiver en hiermee tevens}$$

$$\mathcal{E}\hat{\underline{y}} = \mathcal{E}\underline{b}'\underline{x} = \beta'\underline{x} = \eta .$$

Opmerking. Ook als  $\underline{X}$  stochastisch is en ongecorrleerd met  $\underline{e}$ , is  $\underline{b}$  zuiver, dan

$$\text{geldt nl: } \mathcal{E}\underline{b} = \mathcal{E}_{\underline{x}}(\mathcal{E}_{\underline{y}|\underline{x}} \underline{b}) = \mathcal{E}_{\underline{x}}\beta = \beta .$$

De varianties van deze schatters worden:

$$5.5. \quad \text{VAR } \underline{b} = (X'X)^{-1}X'\Sigma^2(\underline{y})X(X'X)^{-1} = (X'X)^{-1}\sigma^2 \text{ daar } \Sigma^2(\underline{y}) = \Sigma^2(\underline{e}) = \sigma^2 I .$$

$$5.6. \quad \text{var } \hat{\underline{y}} = \text{var}(\underline{x}'\underline{b}) = \underline{x}' \text{VAR } \underline{b}\underline{x} = \underline{x}'(X'X)^{-1}\underline{x}\sigma^2 .$$

$$\text{VAR } \hat{\underline{y}} = \text{VAR } \underline{X}\underline{b} = X(X'X)^{-1}X'\sigma^2 .$$

5.7. Is  $\underline{e} \sim N_n(0, \sigma^2 I)$  dan is dus  $\underline{b} \sim N_{p+1}(\beta, (X'X)^{-1}\sigma^2)$ .

5.8. Als schatter voor  $\sigma^2$  namen we  $\underline{s}^2 = \|\underline{y} - \hat{\underline{y}}\|^2 / (n - p - 1)$ . Eigenlijk is dit een schatter voor de residuele populatievariantie  $\|\underline{x}_i - \hat{\underline{x}}_i\|^2$ , nu met  $\underline{x}_i = \underline{y}$

en  $\hat{x}_i = \beta'x$ , zie § 4b;  $n - p - 1 = n - r(X) = \dim$  orthoplement van kolomruimte van  $X$ . Deze schatter volgt ook logisch uit de vorm van  $\sigma^2$ , nl.:  $n\sigma^2 = \text{sp}(\xi e e') = \xi e' e$ , dus  $\sigma^2 = \xi e' e/n = \xi \|y - \eta\|^2/n$ , d.w.z. we schatten  $\eta$  door  $\hat{y}$  en corrigeren hiervoor de noemer. Met deze noemer is  $s^2$  nl. een zuivere schatter voor  $\sigma^2$ :

volgens 2.36 is:

$$\begin{aligned} \xi [(n - p - 1)s^2] &= \xi \|y - \hat{y}\|^2 = \xi \|P_x^\perp y\|^2 = r(P_x^\perp)\sigma^2 = \\ &= (n - p - 1)\sigma^2 \text{ daar } r(P_x^\perp) = r(I - P_x) = n - (p+1); \\ P_x^\perp \eta &= 0 \text{ daar } \eta \in \langle X \rangle. \end{aligned}$$

5.9. Er geldt:  $\|y - \bar{y}\|^2 = \|\hat{y} - \bar{y}\|^2 + \|y - \hat{y}\|^2$  oftewel: de *totale* (gecorrigeerde) KS = KS t.g.v. *regressie* + KS t.g.v. *fout* (KS = kwadratensom).

De analyse wordt

bron	KS	v	EMS
regressie	$\ \hat{y} - \bar{y}\ ^2 = b'X'y - \bar{y}'\bar{y}$	p	MS = KS/v .
fout	$\ y - \hat{y}\ ^2 = y'y - b'X'y$	n - p - 1	
totaal	$\ y - \bar{y}\ ^2 = y'y - \bar{y}'\bar{y}$	n - 1	

Stel dat de  $x_i$  in het model stochastisch waren (dit kan men stellen door aan te nemen dat  $X$  de gehele populatie omvat), dan was een schatting voor  $\rho_{y \cdot x}$  juist

$$r_{y \cdot x} = r(\underline{y}, \underline{\hat{y}}) = \cos(\underline{y} - \bar{y}, \underline{\hat{y}} - \bar{y}) = \frac{\|\hat{y} - \bar{y}\|}{\|y - \bar{y}\|} =: R .$$

Oftewel

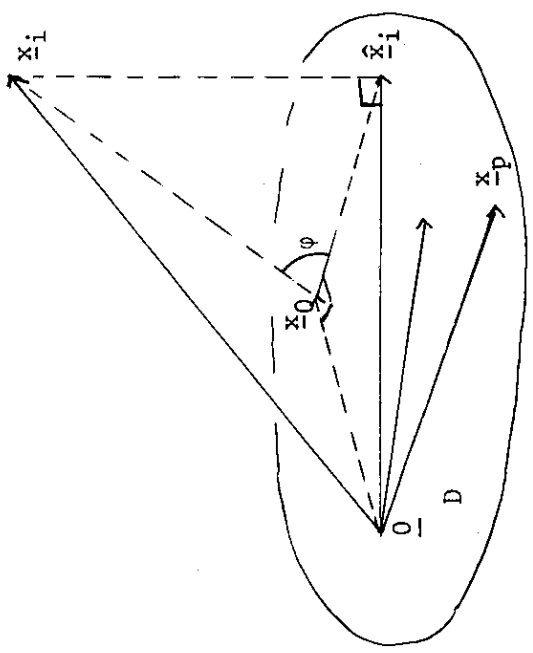
$$R^2 = \frac{\|\hat{y} - \bar{y}\|^2}{\|y - \bar{y}\|^2} = \frac{\text{KS regressie}}{\text{KS totaal}} .$$

Dan is  $\frac{R^2}{1 - R^2} = \frac{\text{KS regressie}}{\text{KS fout}}$  een toetsingsgrootheid voor de hypothese

$\rho_{y \cdot x} = 0$  (of  $(\beta_1, \dots, \beta_p) = 0$ ), zie § 7.

Een samenvatting en tevens de samenhang tussen populatie en steekproef vindt U in de volgende twee figuren:





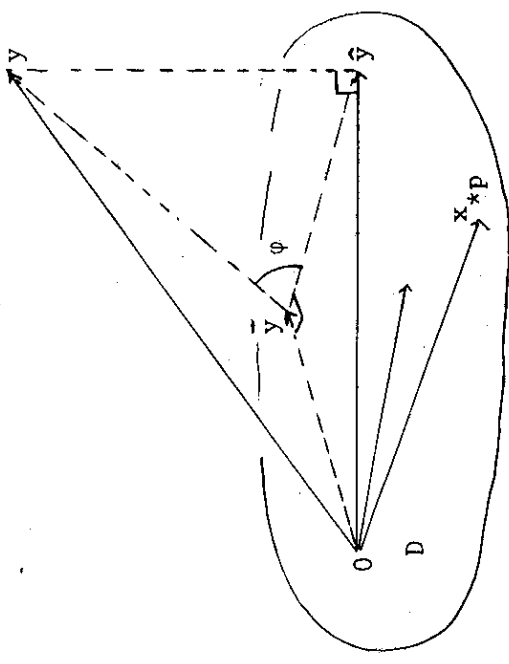
populatie  $L_2$ .

D: (p+1) dim. deelruimte opgespannen door  $x_0, \dots, x_p$ .

$$\begin{aligned}
 x_i &= \hat{x}_i + (x_i - \hat{x}_i) \\
 x_i &= \beta_0 x_0 + \dots + \beta_p x_p + e_i \\
 x_i \mid x &= y = \beta_0 x_0 + \dots + \beta_p x_p + e \\
 \sum e &= 0
 \end{aligned}$$

$\sigma^2 = \text{var } e = \|x_i - \hat{x}_i\|^2$  (residuele variantie)

$$\rho_{i \cdot 2} = \rho_{x_i, \hat{x}_i} = \cos \varphi = \frac{\|\hat{x}_i - \xi \hat{x}_i\|}{\|x_i - \xi x_i\|}$$



Steekproefrealisaties  $R^n$ .

D: (p+1) dim. deelruimte opgespannen door  $x_{*0}, \dots, x_{*p}$ .

$$\begin{aligned}
 \underline{y} &= X\underline{b} + \underline{e} = \hat{y} + (\underline{y} - \hat{y}) \\
 y &= x_{*0}b_0 + \dots + x_{*p}b_p + e \\
 \sum_i e_i &= 0 = \Sigma(y_i - \hat{y}_i)
 \end{aligned}$$

$s^2 = \text{var } e = \|\underline{y} - \hat{y}\|^2 / (n - p - 1)$  (residuele variantie)

$$r_{y \cdot 2} = r_{y, \hat{y}} = \cos \varphi = \frac{\|\hat{y} - \bar{y}\|}{\|y - \bar{y}\|}$$

b, Aanpassing van een rechte lijn

Dit is het speciale geval  $p = 1$ . Dus

$$\underline{y} = \underline{n} + \underline{e} = \beta_0 + \beta_1 \underline{x} + \underline{e}$$

en het steekproefmodel:

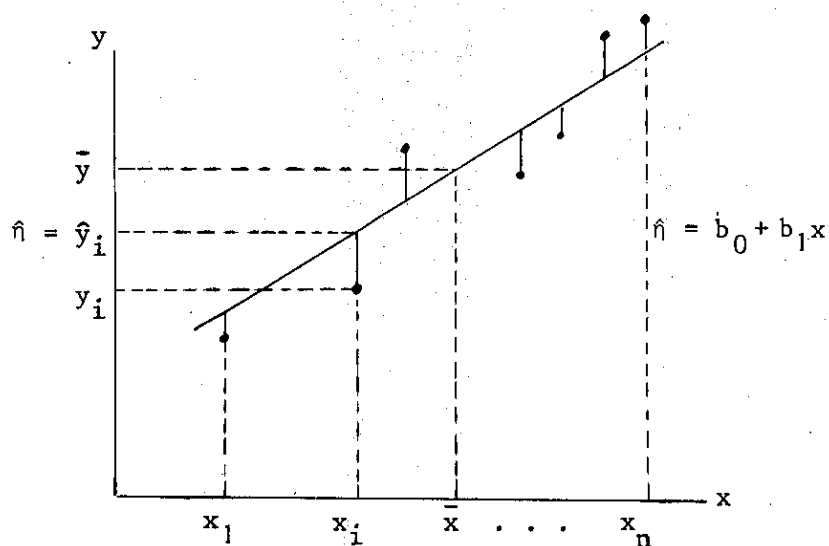
$$\underline{y} = X\beta + \underline{e}$$

met

$$\beta = (\beta_0, \beta_1)'$$

$$\underline{y} = (y_1, \dots, y_n)'; \quad \underline{e} = (e_1, \dots, e_n)'$$

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}.$$



Nu is

$$X'X = \begin{pmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{pmatrix}; \quad X'y = \begin{pmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{pmatrix}; \quad (X'X)^{-1} = \begin{pmatrix} \Sigma x_i^2 & -\Sigma x_i \\ -\Sigma x_i & n \end{pmatrix} / [n \Sigma x_i^2 - (\Sigma x_i)^2].$$

Dit geeft

$$5.10. \quad b_0 = \bar{y} - b_1 \bar{x}; \quad b_1 = \frac{n \Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n \Sigma x_i^2 - (\Sigma x_i)^2} = \frac{s^2(x,y)}{s^2(x)} = r \frac{s(y)}{s(x)};$$

$$\text{var } b_0 = \frac{\Sigma x_i^2}{n \Sigma x_i^2 - (\Sigma x_i)^2} \sigma^2; \quad \text{var } b_1 = \frac{\sigma^2}{\Sigma x_i^2 - (\Sigma x_i)^2 / n}.$$

$$5.11. \text{ var } \hat{\eta} = \sigma^2 \left[ \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum x_i^2 - (\sum x_i)^2/n} \right]; \quad s^2 = \|y - \hat{y}\|^2 / (n - 2) = \sum (y_i - \hat{y}_i)^2 / (n - 2).$$

Enkele opmerkingen

- (1) De berekende regressielijn gaat door het "zwaartepunt  $(\bar{x}, \bar{y})$ " van de puntenwolk, nl.  $\hat{\eta} = b_0 + b_1 x = \bar{y} + b_1 (x - \bar{x})$ .
- (2) We zien uit 5.10 dat, hoe groter  $\sum (x_i - \bar{x})^2$ , des te kleiner var  $b_1$ , d.w.z. doe de waarnemingen in beide uiteinden van het meetgebied (aangenomen dat het model korrek is!).
- (3) Uit 5.11 volgt dat var  $\hat{\eta}$  kleiner wordt naarmate  $x$  dichter bij  $\bar{x}$  ligt. In de uiteinden is  $\hat{\eta}$  het minst nauwkeurig.
- (4) Er geldt  $\hat{\eta} - \bar{y} = b_1 (x - \bar{x})$  oftewel  $\hat{\eta}^* = r_{xy} x^*$  waarbij  $x^*$  en  $\hat{\eta}^*$  gestandaardiseerde variabelen zijn.
- (5) Voor een betrouwbaarheidsgebied voor  $\beta$  zie § 7.  
 Voor  $\eta$ :  $\hat{\eta} - t_{n-2}(\frac{1}{2}\alpha)s(\hat{\eta}) < \eta < \hat{\eta} + t_{n-2}(\frac{1}{2}\alpha)s(\hat{\eta})$ .

c. Gewogen kleinste kwadraten

In § 6a luidde een nevenvoorwaarde: VAR  $\underline{e} = \sigma^2 I$ . Soms is hieraan in de praktijk niet voldaan. Daarom beschouwen we nu het geval VAR  $\underline{e} = \Sigma^2$  (willekeurig, doch regulier).

We minimaliseren nu  $f = (y - \eta)' \Sigma^{-2} (y - \eta)$

$$f = y' \Sigma^{-2} y - 2\eta' \Sigma^{-2} y + \eta' \Sigma^{-2} \eta = y' \Sigma^{-2} y - 2\beta' X' \Sigma^{-2} y + \beta' X' \Sigma^{-2} X \beta.$$

$$\frac{\partial f}{\partial \beta} = -2X' \Sigma^{-2} y + 2X' \Sigma^{-2} X \beta = 0; \quad \frac{\partial^2 f}{\partial \beta^2} = 2X' \Sigma^{-2} X \text{ is p.d. als } r(X) = p+1$$

geeft een minimum voor

$$\underline{b} = (X' \Sigma^{-2} X)^{-1} X' \Sigma^{-2} y, \quad (\text{zie Ba, B8})$$

en

$$\text{VAR } \underline{b} = (X' \Sigma^{-2} X)^{-1} X' \Sigma^{-2} \Sigma^2 \Sigma^{-2} X (X' \Sigma^{-2} X)^{-1} = (X' \Sigma^{-2} X)^{-1}.$$

Een bijzonder geval is weer dat

$$\Sigma^{-2} = \begin{pmatrix} w_1 & \theta \\ \theta & w_n \end{pmatrix} =: W \quad (\text{matrix der gewichten}).$$

Dit geeft een eenvoudige doch interessante toepassing:

Stel  $\eta = \beta x$  een lijn door de oorsprong en we willen de helling  $\beta$  schatten uit een steekproef  $y$  en  $X' = (x_1, \dots, x_n)$ ;  $\underline{y} = X\beta + \underline{e}$ . Neem aan dat  $\text{VAR } \underline{e} = \Sigma^2 = W^{-1}$ . Dus

$$\underline{b} = (X'WX)^{-1}X'W\underline{y} = \frac{\sum w_i x_i y_i}{\sum w_i x_i^2}.$$

(1) Als  $\sigma_i^2 = kx_i$ , dan wordt  $\underline{b} = \bar{y}/\bar{x}$ .

(2) Als  $\sigma_i^2 = kx_i^2$  (d.w.z.  $V_i^2 = \frac{\sigma_i^2}{x_i^2} = \text{constant}$ ), dan is  $\underline{b} = \Sigma(y_i/x_i)/n$ , dus

juist het gemiddelde van de  $n$  schattingen voor de helling.

De juiste procedure hangt dus sterk af van de verdeling van  $\underline{e}$ .

§ 6. De Wishart- en Hotelling's  $T^2$ -verdeling

Beide verdelingen treden op in de multivariate analyse bij het bepalen van een betrouwbaarheidsgebied of het toetsen van hypothesen betreffende parameters (dat zijn vektoren of matrices) van meerdimensionale verdelingen. Ze zijn dan ook generalisaties van de verdelingen die optreden bij het toetsen van parameters van eëndimensionale verdelingen.

a. Definities, notaties

- 6.1. standaard normaal :  $\underline{x} \sim N(0, I)$   
 normaal :  $\mu + \sigma \underline{x} \sim N(\mu, \sigma^2)$   
 p-dim. stand. normaal :  $\underline{x}_p \sim N_p(0, I)$   
 p-dim. normaal :  $\mu + \Sigma' \underline{x}_p \sim N_p(\mu, \Sigma^2) \quad |\Sigma| > 0$

$\underline{x}'_v := ((\underline{x})_1, \dots, (\underline{x})_v)$  alle stochastisch onafhankelijk

$\underline{x}'_{v,p} := ((\underline{x}_p)_1, \dots, (\underline{x}_p)_v)$  alle stochastisch onafhankelijk

$\underline{x}_{v,p} := \mathfrak{M} + \underline{x}_{v,p} \Sigma$  met  $\mathfrak{M} = (\mu, \dots, \mu)'$ .

- 6.2. centrale chi-kwadraat :  $\chi^2_v = \underline{x}'_v \underline{x}_v$   
 niet centrale chi-kwadraat:  $\chi^2_v(\mu^2) = (\mu + \underline{x}_v)^2$

- 6.3. standaard Wishart :  $\chi^2_{v,p} = \underline{x}'_{v,p} \underline{x}_{v,p} \sim W_p(v, I) \quad v \geq p$   
 centrale Wishart :  $(\underline{x}_{v,p} \Sigma)^2 = \Sigma' \chi^2_{v,p} \Sigma \sim W_p(v, \Sigma^2)$   
 niet centrale Wishart :  $\chi^2_{v,p} \sim W_p(v, \Sigma^2, D^2)$

- 6.4. centrale Hotelling :  $\underline{t}^2_{v,p} = \frac{v \chi^2_p}{\chi^2_{v-p+1}}$  beide var. onafhank.  
 niet centrale Hotelling :  $\underline{t}^2_{v,p}(\mu^2) = \frac{v \chi^2_p(\mu^2)}{\chi^2_{v-p+1}}$  beide var. onafhank.  
 centrale Student :  $\underline{t}^2_v = \frac{v \chi^2_p}{\chi^2_v}$  beide var. onafhank.

$v$  heet steeds het aantal *vrijheidsgraden*,  $p$  de *dimensie* van de verdeling. We zullen steeds weer onderstellen dat  $\Sigma$  regulier is.

Enkele opmerkingen

- (1) In Kendall en ook in Anderson wordt de verdeling van de steekproefvariance-covariantiematrix  $\underline{S}^2$  de Wishart verdeling genoemd.
- (2) We zien dat inderdaad de Wishart verdeling een generalisatie is van de  $\chi^2$ -verdeling en Hotelling's verdeling ( $T^2$  in de literatuur) een generalisatie van de Student verdeling, nl.:

$$\underline{X}_1 = \underline{X}; \quad \underline{X}_{v,1}^2 = \underline{X}_v^2 \quad \text{en} \quad \underline{t}_{v,1}^2 = \underline{t}_v^2 \quad (\text{dus } p = 1).$$

- (3) Evenals  $\underline{t}_v^2 \approx \underline{F}_v^1$ , zo is  $\underline{t}_{v,p}^2 \approx \frac{pv}{v-p+1} \underline{F}_{v-p+1}^p$ . Via de F-verdeling is de  $\underline{t}_{v,p}^2$ -verdeling dus te tabelleren.
- (4)  $\mu^2$  resp.  $D^2$  zijn de *niet centraliteitsparameters* van de betreffende verdelingen.  $D^2 = (M\Sigma^{-1})^2$ .

b. Eigenschappen, stellingen

In analogie met  $\underline{X}_v^2 = \underline{X}_v' \underline{X}_v = \sum_1^v (\underline{X}^2)_i$  is volgens 1.30  $\underline{X}_{v,p}^2 = \sum_1^v (\underline{X}_p \underline{X}_p')_i$ . Hiermee is te bewijzen:

$$E \underline{X}_{v,p}^2 = \sum_1^v E (\underline{X}_p \underline{X}_p')_i = \sum_1^v \underline{I} = v \underline{I}$$

en analoog

6.5.  $E(\underline{X}_{v,p} \Sigma)^2 = \Sigma' E \underline{X}_{v,p}^2 \Sigma = \Sigma' v \underline{I} \Sigma = v \Sigma^2$  (= produkt parameters uit  $W_p(v, \Sigma^2)$ ).

6.6. Zij  $G_{r \times v}$  zodat  $GG' = I$ , dan is  $G \underline{X}_{v,p} \approx \underline{X}_{r,p}$ .

Bewijs:

$$\begin{aligned} G \underline{X}_{v,p} &= G((\underline{X}_v)_1, \dots, (\underline{X}_v)_p) = (G(\underline{X}_v)_1, \dots, G(\underline{X}_v)_p), \text{ volgens 3.4} \\ &= ((\underline{X}_r)_1, \dots, (\underline{X}_r)_p) = \underline{X}_{r,p} \end{aligned}$$

daar  $\text{COV}(G(\underline{X}_v)_i, G(\underline{X}_v)_j) = GOG' = 0$  voor  $i \neq j$ , d.w.z. ongecorrleerd dus onafhankelijk. (Ga zelf na dat  $G(\underline{X}_v)_i$  en  $G(\underline{X}_v)_j$  simultaan normaal verdeeld zijn.)

6.7. Hieruit volgt direkt: Is  $P$  een orthogonale projektie met  $r(P) = r$ , dan is

$$(P\underline{\chi}_{v,p}, \Sigma)^2 \approx (\underline{\chi}_{r,p}, \Sigma)^2 .$$

Bewijs. Volgens 1.95 is  $P = G^2$  met  $GG' = I$ , dus

$$(P\underline{\chi}_{v,p}, \Sigma)^2 = (G'G\underline{\chi}_{v,p}, \Sigma)^2 = (G\underline{\chi}_{v,p}, \Sigma)^2 \approx (\underline{\chi}_{r,p}, \Sigma)^2 .$$

In het bijzonder voor  $p = 1$  geldt dus  $(P\underline{\chi}_v)^2 \approx \underline{\chi}_r^2$  (Pas op:  $P\underline{\chi}_v \neq \underline{\chi}_r$ ).

6.8. Stel  $P_A$  en  $P_B$  zijn orthogonale projekties resp. op de orthogonale lineaire deelruimten  $A$  en  $B$ ;  $\underline{x}$  en  $\underline{y}$  zijn stochastische vektoren zodat  $\text{COV}(\underline{x}, \underline{y}) = \alpha I$ . Dan zijn  $P_A \underline{x}$  en  $P_B \underline{y}$  ongecorreleerd.

Bewijs.

$$\text{COV}(P_A \underline{x}, P_B \underline{y}) = P_A \text{COV}(\underline{x}, \underline{y}) P_B' = P_A \alpha I P_B = 0$$

daar  $P_A P_B = 0$ .

Een belangrijk, bijzonder geval hiervan is: (speciaal geval Fisher-Cochran theoreem

6.9. Stel  $\underline{x} = \underline{y} \approx \underline{\chi}_v$ ;  $P_A = P$ ,  $P_B = P^\perp$  (zie 1.90, de orthogonale projektie op het orthoplement  $A^\perp$  van  $A$ );  $r(P) = \dim A = r$ ;  $\dim A^\perp = v - r$ . Dan geldt

$$\underline{\chi}_v^2 = \left[ \underbrace{(P + P^\perp)}_I \underline{\chi}_v \right]^2 = \underline{\chi}_v' (P + P^\perp)^2 \underline{\chi}_v = \underline{\chi}_v' (P^2 + P^{\perp 2}) \underline{\chi}_v$$

$$\underline{\chi}_v^2 = (P\underline{\chi}_v)^2 + (P^\perp \underline{\chi}_v)^2 = \underline{\chi}_r^2 + \underline{\chi}_{v-r}^2$$

en beide onafhankelijk volgens 6.8.

Dit resultaat in algemenere vorm luidt:

6.10. Is  $\underline{W}_1 \sim W_p(v_1, \Sigma^2)$  en  $\underline{W}_2 \sim W_p(v_2, \Sigma^2)$  en zijn beide stochastisch onafhankelijk dan geldt

$$\underline{W}_1 + \underline{W}_2 \sim W_p(v_1 + v_2, \Sigma^2) .$$

Bewijs:

$$\begin{aligned} \underline{W}_1 + \underline{W}_2 &\approx (\underline{X}_{v_1, p} \Sigma)^2 + (\underline{X}_{v_2, p} \Sigma)^2 = \Sigma' (\underline{X}_{v_1, p}^2 + \underline{X}_{v_2, p}^2) \Sigma = \\ &= \Sigma' \left[ \sum_{i=1}^{v_1} (\underline{X}_{-p} \underline{X}'_{-p})_i + \sum_{i=v_1+1}^{v_1+v_2} (\underline{X}_{-p} \underline{X}'_{-p})_i \right] \Sigma = \Sigma' \underline{X}_{v_1+v_2, p}^2 \Sigma = (\underline{X}_{v_1+v_2, p} \Sigma)^2. \end{aligned}$$

Voor  $p = 1$  en  $\Sigma^2 = 1$  ontstaat inderdaad  $\underline{X}_{v_1}^2 + \underline{X}_{v_2}^2 \approx \underline{X}_{v_1+v_2}^2$  indien beide onafhankelijk zijn.

6.11. Als  $\underline{A} \sim W_p(v, \Sigma^2)$  en  $B = B_{k \times p}$  met  $r(B) = k$ , dan is  $\underline{BAB}' \sim W_k(v, B\Sigma^2 B')$ .

Bewijs:

$$\underline{A} \approx (\underline{X}_{v, p} \Sigma)^2 = \sum_{i=1}^v (\Sigma' \underline{X}_{-p})_i (\Sigma' \underline{X}_{-p})'_i$$

dus

$$\underline{BAB}' \approx \sum_{i=1}^v (B\Sigma' \underline{X}_{-p})_i (B\Sigma' \underline{X}_{-p})'_i \sim W_k(v, B\Sigma^2 B')$$

nl. volgens 3.3 is  $B\Sigma' \underline{X}_{-p} \sim N_k(0, B\Sigma^2 B')$  en tevens, daar  $(\underline{X}_{-p})_i$  en  $(\underline{X}_{-p})_j$  onafhankelijk zijn voor  $i \neq j$ , zijn ook  $(B\Sigma' \underline{X}_{-p})_i$  en  $(B\Sigma' \underline{X}_{-p})_j$  onafhankelijk.

6.12. Zij  $\underline{X}_{v, p} := (\underline{X}_1, \underline{X}_2)$ ;  $\underline{X}_1$  bevat  $p_1$ ,  $\underline{X}_2$  bevat  $p_2$  kolommen;  $p_1 + p_2 = p$ .  
Stel  $\underline{X}_{v, p} \approx \underline{X}_{v, p} \Sigma$ , dus  $\mathfrak{M} = 0$ .

$$\underline{A} := \underline{X}' \underline{X} = \begin{pmatrix} \underline{X}'_1 \underline{X}_1 & \underline{X}'_1 \underline{X}_2 \\ \underline{X}'_2 \underline{X}_1 & \underline{X}'_2 \underline{X}_2 \end{pmatrix} =: \begin{pmatrix} \underline{A}_{11} & \underline{A}_{12} \\ \underline{A}_{21} & \underline{A}_{22} \end{pmatrix}.$$

Dan geldt:

$$\underline{A} \sim W_p(v, \Sigma^2) \text{ en } \underline{A}_{22} \sim W_{p_2}(v, \Sigma_{22}^2).$$

Bewijs: Dit volgt direkt uit 6.11, nl. met  $B = (0 \ I_{p_2 \times p_2})$ . Dan is

$$\underline{A}_{22} = \underline{BAB}' \sim W_{p_2}(v, B\Sigma B') = W_{p_2}(v, \Sigma_{22}^2).$$



Verder geldt

6.13.  $\underline{A}_{11.2} := \underline{A}_{11} - \underline{A}_{12} \underline{A}_{22}^{-1} \underline{A}_{21} \sim W_{p_1}(\nu - p_2, \Sigma_{11.2})$  en onafhankelijk  $\underline{X}_2$ .

Bewijs: Noem de i-de rij van  $\underline{X}_1$  en  $\underline{X}_2$  resp.  $\underline{x}^{(1)}$  en  $\underline{x}^{(2)}$ . Daar de rijen onafhankelijk zijn, is de verdeling van  $\underline{x}^{(1)}$  bij gegeven  $\underline{X}_2$  hetzelfde als de verdeling van  $\underline{x}^{(1)}$  bij gegeven  $\underline{x}^{(2)}$ . Deze is volgens 3.10:

$$f(\underline{x}^{(1)} | \underline{x}^{(2)}) = N_{p_1}(\underline{\beta}_{\underline{x}^{(2)}}, \Sigma_{11.2}) .$$

Bij gegeven  $\underline{X}_2$  is dus

$$f(\underline{X}'_1 \underline{X}_1 | \underline{X}_2) = W_{p_1}(\nu, \Sigma_{11.2}, \cdot) .$$

Nu is

$$f(\underline{A}_{11.2} | \underline{X}_2) = f[\underline{X}'_1 (\underline{I} - \underline{X}_2 \underline{A}_{22}^{-1} \underline{X}'_2) \underline{X}_1 | \underline{X}_2] = f[(\underline{P} \underline{X}_1)^2 | \underline{X}_2] = W_{p_1}(\nu - p_2, \Sigma_{11.2})$$

nl.  $\underline{P} := \underline{I} - \underline{X}_2 \underline{A}_{22}^{-1} \underline{X}'_2$  is een orthogonale projectie op  $\underline{X}_2^{\perp}$  (idempotent en symmetrisch en  $r(\underline{P}) = sp(\underline{P}) = \nu - p_2$ ).

Dus bij vaste  $\underline{X}_2$  is  $f(\underline{A}_{11.2} | \underline{X}_2) = W_{p_1}(\nu - p_2, \Sigma_{11.2})$  en daar deze de voorwaarde niet bevat, geldt de verdeling ook onvoorwaardelijk en onafhankelijk  $\underline{X}_2$ .

Bij vaste  $\underline{X}_2$  is de niet-centraliteitsparameter inderdaad 0, nl.:

$$\mathcal{E}[\underline{P} \underline{X}_1 | \underline{X}_2] = \mathcal{E}[(\underline{I} - \underline{X}_2 \underline{A}_{22}^{-1} \underline{X}'_2) \underline{X}_1 | \underline{X}_2] = (\underline{I} - \underline{X}_2 \underline{A}_{22}^{-1} \underline{X}'_2) \underline{X}_2 \Sigma_{22}^{-1} \Sigma_{21} = 0 .$$

De voorlaatste stap volgt uit:  $\mathcal{E}[\underline{x}^{(1)} | \underline{x}^{(2)}] = \underline{\beta}_{\underline{x}^{(2)}} = \Sigma_{12} \Sigma_{22}^{-1} \underline{x}^{(2)}$ , dus

$$\mathcal{E}[\underline{x}^{(1)'} | \underline{x}^{(2)}] = \underline{x}^{(2)'} \Sigma_{22}^{-1} \Sigma_{21} \text{ en nu overgaan op } \underline{X}_1 \text{ en } \underline{X}_2 .$$

6.14. Stelling.  $1/(\chi_{\nu, p}^2)^{11} \sim \chi_{\nu - p + 1}^2$  en is onafhankelijk van de laatste  $p - 1$  kolommen van  $\underline{X}_{\nu, p}$ .

Bewijs: Neem  $\underline{A}$  als in 6.12 met  $\Sigma = \underline{I}$ , dus  $\underline{A} \approx \underline{X}_{\nu, p}^2$ . Neem verder  $p_1 = 1$  dus  $p_2 = p - 1$ ;  $\underline{X}_1$  is dus één kolom. Dan is dus volgens 1.35 en 1.33 (scalar)  $\underline{A}_{11.2} = |\underline{A}| / |\underline{A}_{22}| = |\underline{A}| / |\underline{A}|_{11} = 1/\underline{a}^{11}$  en is volgens 6.13 verdeeld als  $W_1(\nu - p + 1, \underline{I}) = \chi_{\nu - p + 1}^2$ , nl.  $\Sigma_{11.2} = \underline{I}$  daar  $\Sigma^2 = \underline{I}$  dus  $\Sigma_{11} = \underline{I}$  en  $\Sigma_{12} = 0$ .

§ 7. Enkele toetsen

Zij  $\underline{x} = (x_1, \dots, x_p)'$  een p-dimensionale stochastische vektor en  $\underline{X}_{n \times p}$  een steekproef van  $\underline{x}$  ter grootte n, d.w.z.  $\underline{X} = (\underline{x}_{*1}, \dots, \underline{x}_{*p})$  waarbij  $\underline{x}_{*j}$  n waarnemingen zijn van  $x_j$ ,  $j = 1, \dots, p$ .

Schatters voor  $\mu$  en  $\Sigma^2$  zijn:

7.1. Het steekproefgemiddelde  $\underline{\bar{x}} := \sum_1^n \underline{x}_{i*} / n$ ;  $\bar{x} = (\bar{x}_1, \dots, \bar{x}_p)'$ .

7.2. De steekproefvariantie-covariantiematrix

$$\underline{s}^2 := (\underline{X} - \underline{\bar{X}})^2 / (n - 1) = \sum_1^n (\underline{x}_{i*} - \underline{\bar{x}})(\underline{x}_{i*} - \underline{\bar{x}})' / (n - 1).$$

Soms wordt  $\underline{s}^2$  gedefinieerd met n i.p.v. n - 1 in de noemer. Dan is  $\underline{s}^2$  juist de maximumlikelihoodschatter voor  $\Sigma^2$  in het geval dat  $\underline{x}$  multinormaal verdeeld is (zie appendix B).

$\bar{X} = PX$  met  $P = \{1/n\}$ ; elke kolom van  $X$  wordt loodrecht geprojecteerd op de vektor  $(1, \dots, 1)$ , zie ook 1.91. Tevens is  $\bar{X} = (\bar{x}, \dots, \bar{x})'$ , vergelijk met 6.1  $\mathcal{M} = (\mu, \dots, \mu)'$ . Dus  $\bar{X}$  is een schatter voor  $\mathcal{M}$ .

7.3. Er geldt:

$$(\underline{X} - \mathcal{M})^2 = (\underline{X} - \bar{X} + \bar{X} - \mathcal{M})^2 = (\underline{X} - \bar{X})^2 + (\bar{X} - \mathcal{M})^2 \quad (\text{ga na!})$$

7.4. Analoog aan  $s^2 = (\Sigma x_i^2 - n\bar{x}^2) / (n - 1)$  is  $S^2 = (X^2 - \bar{X}^2) / (n - 1)$ , nl.:

$$\begin{aligned} (n - 1)S^2 &= (X - \bar{X})^2 = [(I - P)X]^2 = X'(I - P)X = \\ &= X^2 - X'PX = X^2 - X'P'PX = X^2 - \bar{X}^2. \end{aligned}$$

$\bar{X}$  en  $\underline{s}^2$  (met n - 1 in de noemer) zijn zuivere schatters voor  $\mu$  resp.  $\Sigma^2$  nl.:

$$\mathcal{E}\bar{x} = \sum_1^n \mathcal{E}x_{i*} / n = \sum_1^n \mu / n = \mu$$

en

$$\mathcal{E}\underline{s}^2 = \{\mathcal{E}s_{ij}^2\} = \{\sigma_{ij}^2\} = \Sigma^2.$$

7.5. Stelling. Is  $\underline{x} \sim N_p(\mu, \Sigma^2)$ , dan zijn  $\bar{\underline{x}}$  en  $\underline{S}^2$  stochastisch onafhankelijk.

Bewijs:  $\bar{\underline{x}}'(X - \bar{\underline{x}}) = (PX)'(P^\perp X) = 0$  (zie ook 1.90), d.w.z.  $Px_{*i} \perp P^\perp x_{*j}$ .

Nu is  $\text{COV}(x_{*i}, x_{*j}) = \sigma_{ij}^2 I$ , dus volgens 6.8 zijn  $Px_{*i}$  en  $P^\perp x_{*j}$  ongecorrleerd dus onafhankelijk, oftewel  $\bar{\underline{x}}$  en  $\underline{X} - \bar{\underline{x}}$  zijn onafhankelijk en dus ook  $\bar{\underline{x}}$  en  $\underline{S}^2$ .

7.6. Stelling. Is  $\underline{x} \sim N_p(\mu, \Sigma^2)$  en  $\bar{\underline{x}}$  en  $\underline{S}^2$  gedefinieerd als in 7.1 en 7.2, dan is  $\bar{\underline{x}} \sim N_p(\mu, \Sigma^2/n)$  en  $\underline{S}^2 \sim W_p(n-1, \Sigma^2/(n-1))$ .

Bewijs:  $\underline{X}_{n,p} \approx \mathfrak{M} + \underline{X}_{n,p} \Sigma$ , dus  $\bar{\underline{x}}' = \frac{u'}{n} \underline{X}_{n,p} \approx \mu' + \frac{\chi_p' \Sigma / \sqrt{n}}{n} \sim N_p(\mu', \Sigma^2/n)$ ,  
 $u' = (1, \dots, 1)$ .

$(\underline{X} - \bar{\underline{x}})^2 \approx [(I - P)\underline{X}_{n,p} \Sigma]^2$ , nl.  $(I - P)\mathfrak{M} = 0$  en nu met 6.7  $\approx (\underline{X}_{n-1,p} \Sigma)^2$  dus

$$\underline{S}^2 \approx (\underline{X}_{n-1,p} \Sigma / \sqrt{n-1})^2 \sim W_p(n-1, \Sigma^2/(n-1)).$$

Opmerking. We zien dat  $\mathcal{E}\bar{\underline{x}} = \mu$  en  $\mathcal{E}\underline{S}^2 = \Sigma^2$  (met 6.5).  $v = n - 1$  heet het aantal vrijheidsgraden van  $\underline{S}^2$ ;  $p = 1$  geeft  $v \underline{s}^2 / \sigma^2 \approx \chi_v^2$ , nl.  $\underline{s}^2 \approx \chi_v^2 \sigma^2 / v$ .

7.7. Stelling. Zijn  $\underline{S}_1^2$  en  $\underline{S}_2^2$  onafhankelijke schatters voor  $\Sigma^2$  uit  $N_p(\mu, \Sigma^2)$  met resp.  $v_1$  en  $v_2$  vrijheidsgraden, dan geldt:

$$\underline{S}^2 := \frac{v_1 \underline{S}_1^2 + v_2 \underline{S}_2^2}{v_1 + v_2} \sim W_p(v_1 + v_2, \Sigma^2 / (v_1 + v_2)) \quad \text{en} \quad \mathcal{E}\underline{S}^2 = \Sigma^2.$$

Bewijs:  $v_1 \underline{S}_1^2 \sim W_p(v_1, \Sigma^2)$ ;  $v_2 \underline{S}_2^2 \sim W_p(v_2, \Sigma^2)$ , dus volgens 6.10

$v_1 \underline{S}_1^2 + v_2 \underline{S}_2^2 \sim W_p(v_1 + v_2, \Sigma^2)$  en dus  $\underline{S}^2 \sim W_p(v_1 + v_2, \Sigma^2 / (v_1 + v_2))$ .

Dat  $\mathcal{E}\underline{S}^2 = \Sigma^2$  volgt uit 6.5.

Nu volgen enkele toetsen.

Zij  $\underline{X}_{n \times p}$  een onafhankelijke, aselechte steekproef ter grootte  $n$  uit  $N_p(\mu, \Sigma^2)$ .

Bij gebruik van een likelihood-ratio toets wordt de nulhypothese

$$H_0 : \theta \in \omega$$

getoetst tegen de alternatieve hypothese

$$H_a : \theta \in \Omega \setminus \omega$$

waarbij  $\theta$  de vector van onbekende parameters voorstelt,  $\Omega$  de gehele parameter-ruimte en  $\omega \subset \Omega$  de parameterruimte onder  $H_0$ . De likelihood-functie  $L$  wordt

zowel voor  $\theta \in \omega$  als voor  $\theta \in \Omega$  gemaximaliseerd. De maxima geven we aan met  $L_\omega$  resp.  $L_\Omega$ . Het toetsingscriterium is dan

$$\lambda = L_\omega / L_\Omega.$$

In een aantal gevallen kennen we de verdeling van  $\lambda$ , of die van een monotone functie van  $\lambda$ . In de andere gevallen moeten we de verdeling van  $\lambda$  benaderen. Volgens Wilks (zie An.Math.Stat. 9.p. 60) is

$$-2 \log \lambda = -2(\log L_\omega - \log L_\Omega) \sim \chi_v^2$$

met

$$v = v_\Omega - v_\omega$$

dat wil zeggen het verschil in dimensie tussen  $\Omega$  en  $\omega$ .

Vooraf voor grote  $n$  is deze benadering goed. Voor verschillende in dit dictaat beschreven toetsen bestaan vermenigvuldigingsfactoren voor de toetsingsgrootheden. Deze dienen om voor kleine  $n$  een betere aanpassing aan de  $\chi^2$ -verdeling te krijgen. (zie Bartlett. J. Royal. Stat. Society, series B, 16 pp.296-298).

### 7.8. Toets voor de hypothese $\mu = \mu_0$ .

a. Stel  $\Sigma^2$  bekend. Bepaal  $\bar{x}$  uit steekproef. Voor  $p = 1$  gold als toetsingsgroot-

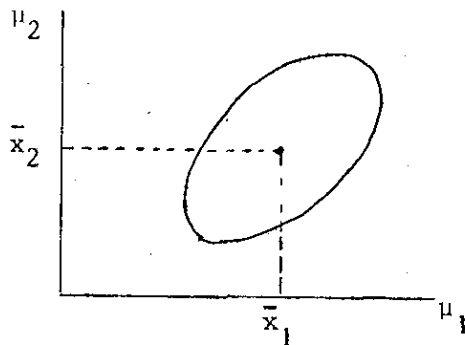
heid  $\left(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right)^2 = n(\bar{x} - \mu_0)\sigma^{-2}(\bar{x} - \mu_0) \approx \chi_1^2$ . Analoog nemen we nu als toetsingsgrootheid:  $n(\bar{x} - \mu_0)' \Sigma^{-2} (\bar{x} - \mu_0) \approx \chi_p^2$ . Dat deze toetsingsgrootheid

onder  $H_0$  inderdaad  $\chi_p^2$ -verdeeld is volgt uit:  $\bar{x} \approx \mu_0 + \Sigma' \chi_p / \sqrt{n}$  (zie 7.6) dus  $\sqrt{n} \Sigma^{-T} (\bar{x} - \mu_0) \approx \chi_p$  en dus

$$n(\bar{x} - \mu_0)' \Sigma^{-2} (\bar{x} - \mu_0) = [\sqrt{n} \Sigma^{-T} (\bar{x} - \mu_0)]^2 \approx \chi_p^2.$$

De kritieke zône wordt  $n(\bar{x} - \mu_0)' \Sigma^{-2} (\bar{x} - \mu_0) \geq \chi_p^2(\alpha)$ .

$n(\bar{x} - \mu)' \Sigma^{-2} (\bar{x} - \mu) \leq \chi_p^2(\alpha)$  geeft een *betrouwbaarheidsgebied* voor  $\mu$  met betrouwbaarheid  $P = 1 - \alpha$ . Dit is het oppervlak + inwendige van een  $p$ -dimensionale ellipsoïde met middelpunt  $\bar{x}$ . Voor  $p = 2$  wordt dit een ellips:



Betrouwbaarheidsgebied voor  $\mu = (\mu_1, \mu_2)'$  bij gegeven  $\bar{x} = (\bar{x}_1, \bar{x}_2)'$  en  $\alpha$ .

b. Stel  $\Sigma^2$  onbekend. Bepaal  $\bar{x}$  en  $S^2$  uit steekproef. Voor  $p = 1$  gold als toetsingsgrootheid  $\left(\frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right)^2 = n(\bar{x} - \mu_0)' \underline{s}^{-2} (\bar{x} - \mu_0) = \underline{t}_{n-1}^2$ . Analoog nemen we nu als toetsingsgrootheid:  $n(\bar{x} - \mu_0)' \underline{S}^{-2} (\bar{x} - \mu_0) =: \underline{T}^2$  (van Hotelling). Deze volgt uit het likelihoodratio-criterium, nl.:

$$\begin{aligned} \lambda &= \frac{\max_{\omega} L(\mu_0, \Sigma^2)}{\max_{\Omega} L(\mu, \Sigma^2)} = (\text{zie B12}) \frac{|\underline{\Sigma}_{\Omega}^2|^{\frac{1}{2}n} |(X - \bar{X})^2|^{\frac{1}{2}n}}{|\underline{\Sigma}_{\omega}^2|^{\frac{1}{2}n} |(X - \mu_0)^2|^{\frac{1}{2}n}} \\ &= (\text{zie 7.3}) \frac{|(X - \bar{X})^2|^{\frac{1}{2}n}}{|(X - \bar{X})^2 + (\bar{X} - \mu_0)^2|^{\frac{1}{2}n}}; \\ \lambda^{\frac{2}{n}} &= \frac{|(X - \bar{X})^2|}{|(X - \bar{X})^2 + n(\bar{x} - \mu_0)(\bar{x} - \mu_0)'|} = \frac{|(X - \bar{X})^2|}{\begin{vmatrix} 1 & (\bar{x} - \mu_0)' \\ -n(\bar{x} - \mu_0) & (X - \bar{X})^2 \end{vmatrix}} \\ &= (\text{zie 1.35}) \frac{1}{1 + n(\bar{x} - \mu_0)'(X - \bar{X})^{-2}(\bar{x} - \mu_0)} = \frac{1}{1 + \underline{T}^2/(n-1)}. \end{aligned}$$

Het kritieke gebied  $\lambda \leq \lambda_{\alpha}$  geeft als kritiek gebied  $\underline{T}^2 \geq \underline{T}_{\alpha}^2$ .

Onder  $H_0$  geldt eveneens  $\underline{T}^2 \approx \underline{t}_{n-1, p}^2$ .

Bewijs: Volgens 7.6 is  $v\underline{S}^2 \approx (\chi_{v, p}^2)_{\Sigma}$  en onafhankelijk  $\bar{x}$ ;  $v = n - 1$ . Dus de toetsingsgrootheid is isomoor met

$$(\Sigma' \underline{\chi}_p)' (\Sigma' \underline{\chi}_{v, p}^2 \Sigma / v)^{-1} (\Sigma' \underline{\chi}_p)$$

en  $\underline{\chi}_p$  stochastisch onafhankelijk van  $\underline{\chi}_{v, p}$

$$= v \underline{\chi}_p' (\underline{\chi}_{v, p}^2)^{-1} \underline{\chi}_p. \text{ Opm. } r(\Sigma' \underline{\chi}_{v, p}^2 \Sigma) = r(\underline{\chi}_{v, p}^2) = r(\underline{\chi}_{v, p}) = p \text{ (bijna ov}$$

Houd nu  $\underline{\chi}_p$  vast en voer een orthogonale basistransformatie  $H$  uit zodat de eerste rij van  $H$  valt langs  $\underline{\chi}_p$ . Dan krijgen we

$$\begin{aligned}
 &= v \chi_p' (H'H) (\chi_{v,p}^2)^{-1} H'H \chi_p \approx v (H \chi_p)' (\chi_{v,p}^2)^{-1} (H \chi_p) \\
 &= v \chi_p^2 (1, 0, \dots, 0) (\chi_{v,p}^2)^{-1} (1, 0, \dots, 0)' = v \chi_p^2 (\chi_{v,p}^2)^{-1}
 \end{aligned}$$

en nu onvoorwaardelijk daar  $\chi_{v,p}$  onafhankelijk  $\chi_p$  is:

$$= v \chi_p^2 / \chi_{v-p+1}^2 \quad (\text{volgens 6.14})$$

$$= \underline{t}_{v,p}^2 \quad (\text{zie 6.4}).$$

Opmerking. Voor  $v \rightarrow \infty$ , d.w.z.  $S^2 \rightarrow \Sigma^2$  gaat  $\underline{t}_{v,p}^2 \rightarrow \underline{t}_{\infty,p}^2 \approx \chi_p^2$  (geval a).

De niet-centrale Hotelling verdeling treedt op bij het onderscheidingsvermogen van deze toets.

### 7.9. Toets voor de hypothese $\mu_1 = \mu_2$ .

Stel  $\underline{x}_1 \sim N_p(\mu_1, \Sigma_1^2)$  en  $\underline{x}_2 \sim N_p(\mu_2, \Sigma_2^2)$ ;  $\underline{X}_{n_1,p}$  en  $\underline{X}_{n_2,p}$  onafhankelijke steekproeven.

a. Stel  $\Sigma_1^2$  en  $\Sigma_2^2$  beide bekend. Toetsingsgrootheid

$$(\bar{\underline{x}}_1 - \bar{\underline{x}}_2)' \left( \frac{\Sigma_1^2}{n_1} + \frac{\Sigma_2^2}{n_2} \right)^{-1} (\bar{\underline{x}}_1 - \bar{\underline{x}}_2) \approx \chi_p^2.$$

b. Stel  $\Sigma_1^2 = \Sigma_2^2$  doch onbekend. Toetsingsgrootheid:

$$\frac{n_1 n_2}{n_1 + n_2} (\bar{\underline{x}}_1 - \bar{\underline{x}}_2)' S^{-2} (\bar{\underline{x}}_1 - \bar{\underline{x}}_2) \approx \underline{t}_{v_1+v_2,p}^2$$

met  $S^2$ ,  $v_1$  en  $v_2$  als in 7.7.

Opmerking. Het geval  $\Sigma_1^2 \neq \Sigma_2^2$  en beide onbekend behandelen we niet (zie [1]).

### 7.10. Toets voor de hypothese $\beta = \beta^*$ ( $\beta = (\beta_0, \dots, \beta_p)'$ is de vektor der regressie-coëfficiënten).

Ondersteld is dat  $\underline{e} \sim N_n(0, \sigma^2 I)$ ;  $\underline{X}_{n,p+1}$  de steekproefrealisatie. Dan geldt onder  $H_0$  (zie 5.7)

$$\underline{b} \sim N_{p+1}[\beta^*, (X'X)^{-1} \sigma^2],$$

oftewel

$$(\underline{b} - \beta^*)' \frac{X'X}{\sigma^2} (\underline{b} - \beta^*) \approx \chi_{p+1}^2.$$

Nu is

$$\begin{aligned} (\underline{b} - \beta^*)' \frac{X'X}{s^2} (\underline{b} - \beta^*) &= (\underline{b} - \beta^*)' \frac{X'X}{\sigma^2} (\underline{b} - \beta^*) \frac{1}{s^2/\sigma^2} \approx \\ &\approx \frac{\chi_{p+1}^2}{\chi_v^2/v} \text{ (beide onafhankelijk)} \approx (p+1) F_v^{p+1} \quad \text{met } v = n - p - 1. \end{aligned}$$

De kritieke zône wordt:  $(\underline{b} - \beta^*)' \frac{X'X}{(p+1)s^2} (\underline{b} - \beta^*) \geq F_v^{p+1}(\alpha)$ ,

$(\underline{b} - \beta)' \frac{X'X}{(p+1)s^2} (\underline{b} - \beta) \leq F_v^{p+1}(\alpha)$  geeft een *betrouwbaarheidsgebied* voor  $\beta$  met betrouwbaarheid  $P = 1 - \alpha$ . Voor  $p = 1$  is dit het inwendige van een elips.

7.11. Toets voor de hypothese  $\rho_{1 \cdot 2} = 0$ .

Stel  $\underline{x} = (\underline{x}_1, \underline{x}^{(2)})' \sim N_{p+1}(\mu, \Sigma^2)$  en  $X_{n,p+1}$  een onafhankelijke steekproef ter grootte  $n$  van  $\underline{x}$ .

Een schatting voor  $\rho_{1 \cdot 2}$  is

$$R := r_{1 \cdot 2} = \frac{\sqrt{S_{12} S_{22}^{-1} S_{21}}}{\sqrt{S_{11}}}$$

met  $S_{12} = (X_1 - \bar{X}_1)'(X_2 - \bar{X}_2)/n$  etc.;  $X_1$  etc. als in 6.12 met  $v = n$ ,  $p_1 = 1$ ,  $p_2 =$

Uit het likelihoodratio-criterium volgt als toetsingsgroottheid (zie [1], pag. 90)

$$\frac{R^2}{1 - R^2} = \frac{S_{12} S_{22}^{-1} S_{21}}{S_{11 \cdot 2}}.$$

Beschouw  $X_2$  vast en noem  $X_2 - \bar{X}_2 = Y_2$  en  $X_1 - \bar{X}_1 = Y_1$ . Dan wordt

$$f\left(\frac{R^2}{1-R^2} \mid X_2\right) = f\left(\frac{Y_1' [Y_2 (Y_2' Y_2)^{-1} Y_2'] Y_1}{Y_1' [I - Y_2 (Y_2' Y_2)^{-1} Y_2'] Y_1}\right) = f\left(\frac{(PY_1)^2}{(P^\perp Y_1)^2} \mid X_2\right)$$

waarin  $P := Y_2 (Y_2' Y_2)^{-1} Y_2'$  de orthogonale projektie is op kolomruimte van  $Y_2$  met rang  $p$  en  $P^\perp$  de orthogonale projektie op het orthoplement van  $Y_2$  met rang  $(n - p - 1)$  (zie ook 1.90).

$$f(Y_1 \mid X_2) = N_n(\gamma, \Sigma_{11 \cdot 2}^{-1} I) \text{ met } \gamma = (X_2 - \bar{X}_2) \Sigma_{22}^{-1} \Sigma_{21}.$$

7.12. Dus ook onvoorwaardelijk:

$$\frac{\underline{R}^2}{1 - \underline{R}^2} \approx \frac{\chi_p^2(\delta^2) \Sigma_{11.2}}{\chi_{n-p-1}^2 \Sigma_{11.2}} \approx \frac{t_{v,p}^2(\delta^2)}{v} \quad \text{met } v = n - 2.$$

Opm.:  $\Sigma_{11.2}$  is een scalar;  $\delta^2 = (PY/\sqrt{\Sigma_{11.2}})^2 = Y^2/\Sigma_{11.2}$ .

$\delta^2 = 0$  als  $\rho_{1.2} = 0$ , nl. dan is  $\Sigma_{12} = 0$ , dus onder  $H_0$  een centrale F stochastiek. De noemer van 7.12 is inderdaad een *centrale*  $\chi$ -kwadraat, daar

$$E(PY_{-1} | X_2) = [I - Y_2(Y_2'Y_2)^{-1}Y_2'](Y_2'\Sigma_{22}^{-1}\Sigma_{21}) = 0.$$

Teller en noemer in 7.12 zijn onafhankelijk volgens 6.8.

7.13. Het bijzondere geval  $p = 1$  geeft een toets voor de hypothese  $\rho = 0$ . Dan is

$$\frac{(n-2)\underline{r}^2}{1 - \underline{r}^2} \approx t_{n-2}^2 \quad (\underline{r} \text{ is de steekproefcorrelatiecoëfficiënt}).$$

Opmerking. Vergelijk 7.12 met 5.9. De toets voor  $\rho_{1.2} = 0$  is dus te vergelijken met de toets voor  $(\beta_1, \dots, \beta_p) = 0$ .

7.14. Toets voor de hypothese  $\Sigma_1^2 = \dots = \Sigma_k^2$ .

Stel  $\underline{x}^{(i)} \sim N_p(\mu^{(i)}, \Sigma_i^2)$ ,  $i = 1, \dots, k$ . Aan elke  $\underline{x}^{(i)}$  doen we  $n_i$  waarnemingen:

$$X^{(i)} = \begin{pmatrix} x_{11}^{(i)} & \dots & x_{1p}^{(i)} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_{n_i 1}^{(i)} & \dots & x_{n_i p}^{(i)} \end{pmatrix}; \quad \sum_{i=1}^k n_i = n.$$

$$S_i^2 := (X^{(i)} - \bar{X}^{(i)})^2 / n_i.$$

De likelihoodfunctie is dan (zie appendix Bc):

$$L = \prod_{i=1}^k \{ |\Sigma_i^2|^{-\frac{1}{2}n_i} (2\pi)^{-\frac{1}{2}n_i p} \exp[-\frac{1}{2} \sum_{j=1}^{n_i} (x_{j*}^{(i)} - \mu^{(i)})' \Sigma_i^{-2} (x_{j*}^{(i)} - \mu^{(i)})] \}.$$

Nu geldt in  $\Omega$ :  $\hat{\mu}^{(i)} = \bar{x}^{(i)}$ ,  $\hat{\Sigma}_i^2 = S_i^2$

$i = 1, \dots, k$ .

en in  $\omega$ :  $\hat{\mu}^{(i)} = \bar{x}^{(i)}$ ,  $\hat{\Sigma}^2 = \sum_{i=1}^k n_i S_i^2 / n$



Dit laatste volgt aldus:

$$\text{in } \omega \text{ is } \log L = c - \frac{1}{2} \sum_1^k n_i [\log |\Sigma^2| + \text{sp}(S_i^2 \Sigma^{-2})], \text{ zie B9.}$$

Met behulp van B5, B6 en B7 vinden we

$$\begin{aligned} d \log L &= - \frac{1}{2} \sum_1^k n_i [\text{sp}(\Sigma^{-2} d\Sigma^2) - \text{sp}(S_i^2 \Sigma^{-2} d\Sigma^2 \Sigma^{-2})] = \\ &= - \frac{1}{2} \text{sp}[(n\Sigma^{-2} - \sum_1^k n_i \Sigma^{-2} S_i^2 \Sigma^{-2}) d\Sigma^2] = 0 \end{aligned}$$

geeft  $\hat{\Sigma}^2 = \Sigma n_i S_i^2 / n$ . Met B12 wordt

$$\lambda = \frac{\max_{\omega} \underline{L}}{\max_{\Omega} \underline{L}} = \frac{\prod |\hat{\Sigma}^2|^{-\frac{1}{2}n_i} (2\pi)^{-\frac{1}{2}n_i} \exp(-\frac{1}{2}n_i p)}{\prod |\underline{\Sigma}_i^2|^{-\frac{1}{2}n_i} (2\pi)^{-\frac{1}{2}n_i} \exp(-\frac{1}{2}n_i p)} = \frac{\prod_{i=1}^k |\hat{\Sigma}_i^2|^{\frac{1}{2}n_i}}{|\hat{\Sigma}^2|^{\frac{1}{2}n}}$$

Nu is  $-2 \log \lambda \sim \chi_{\nu}^2$

met  $\nu = \nu_{\Omega} - \nu_{\omega} = k \cdot \frac{1}{2}p(p+1) - \frac{1}{2}p(p+1) = \frac{1}{2}p(k-1)(p+1)$

(aantal onbekende parameters in de variantie-covariantiematrices).

## 8. Componentenanalyse

Men beschouwt de stochastische variabelen  $\underline{x}_1, \dots, \underline{x}_p$ , samengevat als

$$\underline{x}' = (\underline{x}_1, \dots, \underline{x}_p) ,$$

met

$$E\underline{x} = \mu,$$

$$E(\underline{x}-\mu)(\underline{x}-\mu)' = \Sigma^2.$$

Er kan hierbij bijv. gedacht worden aan  $p$  eigenschappen die aan individuen kunnen worden gemeten. Men stelt zich nu de vraag of het mogelijk is de individuen te karakteriseren door een kleiner aantal variabelen, lineaire combinaties van de oorspronkelijke, die onderling niet gecorreleerd zijn. Deze nieuwe eigenschappen heten hoofdcomponenten ("principal components").

Als  $\Sigma^2$  de rang  $m < p$  heeft dan bestaan er  $p-m$  lineaire relaties tussen de  $\underline{x}$ -en, dus er zijn  $p-m$  lineaire combinaties met variantie 0. Er zijn dan slechts  $m$  lineair onafhankelijke variabelen en alle eigenschappen van een individu kunnen door  $m$  lineaire combinaties worden vastgelegd. In het algemeen zal  $\Sigma^2$  echter de rang  $p$  hebben. Als men zich ook in dat geval wil beperken tot een kleiner aantal variabelen dan worden daarvoor die lineaire combinaties gekozen die een zo groot mogelijke variantie hebben. (Hotelling 1933) .

Willen we ons beperken tot  $k$  ( $1 \leq k < p$ ) hoofdcomponenten  $\underline{y}_1, \dots, \underline{y}_k$ , dan moeten we zoeken naar  $\underline{y} = (\underline{y}_1, \dots, \underline{y}_k)'$  waarvoor geldt

$$\underline{y} = G' \underline{x} \text{ met } G = G_{p \times k}, G'G = I_{k \times k}$$

zodanig dat

$$\sum_{i=1}^k \text{var } \underline{y}_i$$

maximaal is. Omdat

$$\text{var } \underline{y}_i = E[G'(\underline{x}-\mu)(\underline{x}-\mu)'G]_{ii}$$

komt het probleem neer op het zoeken naar  $G_{p \times k}$ ,  $G'G = I_{k \times k}$  zodanig dat

$$\text{sp}(G'\Sigma^2G) = \text{sp}(GG'\Sigma^2)$$

maximaal is. Nu geldt voor elke  $G_{p \times k}$  met  $G'G = I$ , dat

$$P := GG'$$

de orthogonale projector is op  $\langle G \rangle$  (ga na). Uit 1.99 volgt nu dat de kolommen van  $G$  orthogonale eigenvectoren zijn van  $\Sigma^2$  behorend bij de  $k$  grootste eigenwaarden  $\lambda_1, \dots, \lambda_k$ . (Anderson, 1958).

Het is het meest gebruikelijk om de hoofdcomponenten één voor één te berekenen.

Bovenstaande wordt dan toegepast voor  $k = 1, \dots, k_0$  voor zekere  $k_0$ ,  $1 \leq k_0 \leq p$ .

In dat geval vinden we de eigenvectoren bij de eigenwaarden in afnemende grootte, zodat dan

$$\text{var } y = \text{var}(G'\underline{x}) = G'\Sigma^2G = G'GA = \Lambda$$

met  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_k)$  en  $\lambda_1 \geq \dots \geq \lambda_k$ .

Opmerking. We zien ook, dat aan de eis, dat de hoofdcomponenten onderling ongecorreleerd moeten zijn, voldaan is.

De componentenanalyse is gevoelig voor een schaaltransformatie. Stel

$$\underline{x}^* = D\underline{x},$$

met

$$D := \begin{bmatrix} d_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & d_p \end{bmatrix}$$

Dan is de nieuwe variantie-covariantiematrix:

$$\Sigma^{2*} = D \Sigma^2 D.$$

Het nieuwe eigenwaardeprobleem wordt nu

$$D \Sigma^2 D H^* = H^* \Lambda^*, \quad \text{of} \quad \Sigma^{2*} D H^* = D^{-1} H^* \Lambda^*$$

in plaats van

$$\Sigma^2 H = H \Lambda.$$

Dit is alleen dan essentieel hetzelfde probleem als

$$D = cI.$$

Vaak wordt in de toepassingen met gestandaardiseerde variabelen gewerkt en gaat men uit van de correlatiematrix. Dit is natuurlijk geen oplossing van de principiële moeilijkheid. (Zie Anderson, 1963, p. 139)

Tot dusver hebben we populatie-eigenschappen besproken. Nu gaan we uit van een steekproef van  $x$ . dus van een matrix

$$X_{n \times p} = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \dots & x_{np} \end{pmatrix}$$

We zoeken nu  $k$  lineaire combinaties  $y_{*1}, \dots, y_{*k}$  van de kolommen van  $x$  met maximale steekproefvariantie. Voor  $\hat{G}_{p \times k}$  moet gelden

$$Y = (y_{*1}, \dots, y_{*k}) = X\hat{G} \text{ met } \hat{G}'\hat{G} = I_k$$

$$SP = (Y - \bar{Y})'(Y - \bar{Y}) \text{ maximaal.}$$

Hierin is  $\bar{Y}$  gedefinieerd als in 7.2. Nu is

$$\begin{aligned} sp(Y - \bar{Y})'(Y - \bar{Y}) &= sp(\hat{G}'(X - \bar{X})'(X - \bar{X})\hat{G}) \\ &= sp(\hat{G}\hat{G}'(X - \bar{X})(X - \bar{X})) \\ &= n \, sp(\hat{G}\hat{G}' S^2) \end{aligned}$$

Op dezelfde manier als bij de populatie volgt nu dat de  $k$  kolommen van  $\hat{G}$  worden gevormd door orthogonale eigenvectoren behorend bij de  $k$  grootste eigenwaarden van  $S^2$ .

Opmerking. Na schatting van  $\hat{G}$  volgen uit  $Y = X\hat{G}$  direct de individuele scores op de  $k$  componenten.

Als we aannemen dat  $\underline{x}$ , en dus ook  $\underline{y}$ , multinormaal verdeeld is, kunnen bepaalde hypothesen worden getoetst.

In de eerste plaats kunnen we ons afvragen, of het wel zin heeft om te gaan transformeren, omdat wellicht de  $x_i$  al onafhankelijk zijn. We maximaliseren de likelihood-functie  $L$  zowel met als zonder bijvoorwaarden (in  $\omega$  resp.  $\Omega$  zie § 7).

De logaritme van de simultane kansdichtheid van de waarnemingen is, na substitutie van  $\hat{\mu} = \bar{x}$  (vgl. B9)

$$\log L = c - \frac{1}{2}n \log |\Sigma^2| - \frac{1}{2}n \, sp(S^2 \Sigma^{-2}).$$

Onder de bijvoorwaarde dat alle covarianties nul zijn, levert dit

$$\log L = c - \frac{1}{2}n \log \prod_{i=1}^p \sigma_i^2 - \frac{1}{2}n \sum_{i=1}^p S_i^2 \sigma_i^{-2}.$$

Het maximum wordt aangenomen voor  $\hat{\sigma}_i^2 = S_i^2$  ( $i = 1, \dots, p$ ), en is

$$\log L_{\omega} = c - \frac{1}{2}n \log \prod_{i=1}^p S_i^2 - \frac{1}{2}np$$

Maximaliseren zonder bijvoorwaarden levert (vgl. B 12)

$$\log L_{\Omega} = c - \frac{1}{2}n \log |S^2| - \frac{1}{2}np$$

Het toetsingscriterium is (zie §7)

$$- 2 \log \lambda = 2 (\log L_{\Omega} - \log L_{\omega}) = n \log \frac{\prod_{i=1}^p s_i^2}{|S^2|}$$

De steekproefcorrelatiematrix R is gedefinieerd als

$$R := V^{-1} S^2 V^{-1}, \text{ met } V := \text{diag}(s_1, \dots, s_p).$$

We kunnen de toetsingsgrootte dus ook schrijven als

$$- 2 \log \lambda = n \log |R|^{-1}.$$

Deze toetsingsgrootte heeft voor grote n bij benadering een  $\chi^2$ -verdeling met  $\frac{1}{2}p(p-1)$  vrijheidsgraden. Dit is het aantal covarianties, die in  $\Omega$  wel, in  $\omega$  niet geschat worden, ofwel het verschil in aantal parameters (dimensies) tussen  $\Omega$  en  $\omega$ . (zie ook §7).

Daarna rijst de vraag: als er een aantal relatief grote eigenwaarden zijn, heeft het dan nog zin tussen de overige te onderscheiden? Stel k eigenwaarden van  $S^2$  zijn zo groot (t.o.v.  $\text{sp}(S^2)/p$ ) dat ze met  $\lambda_1, \dots, \lambda_k$  uit de populatie zijn te identificeren. We willen toetsen dat de overige p-k alle gelijk zijn.

Stel  $S^2 = KHK'$  met K orthogonaal,  $H = \text{diag}(\eta_1, \dots, \eta_p)$ . Voor het gemak nemen we  $\eta_1 > \eta_2 > \dots > \eta_k > \eta_{k+1} \geq \dots \geq \eta_p$ , en  $\Sigma^2 = GAG'$  met G orthogonaal,

$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_k, v, \dots, v)$ . Dan is

$$\begin{aligned} \log L &= c - \frac{1}{2}n \log |\Sigma^2| - \frac{1}{2}n \text{sp}(S^2 \Sigma^{-2}) = \\ &= c - \frac{1}{2}n \sum_{i=1}^k \log \lambda_i - \frac{1}{2}n(p-k) \log v - \frac{1}{2}n \text{sp}(G'KHK'GA^{-1}). \end{aligned}$$

Uit differentiëren van  $\log L$  naar  $\lambda_i$  ( $i=1, \dots, k$ ) resp. v volgt dat voor vaste G  $\log L$  maximaal is voor

$$\lambda_i = (G'KHK'G)_{ii} \quad (i=1, \dots, k), \quad v = \frac{1}{p-k} \sum_{j=k+1}^p (G'KHK'G)_{jj}.$$

We zien nu dat

$$\text{sp}(G'KHK'GA^{-1}) = p$$

en ons rest nu nog het minimaliseren van

$$\sum_{i=1}^k \log \lambda_i + (p-k) \log v = \sum_{i=1}^k \log(G'KHK'G)_{ii} + (p-k) \log \left( \frac{1}{p-k} \sum_{j=k+1}^p (G'KHK'G)_{jj} \right)$$

Dit is minimaal als de eerste k kolommen van G gelijk zijn aan (een permutatie van) de eerste k kolommen van K. De laatste p-k kolommen van G spannen dan dezelfde ruimte op als de eerste p-k kolommen van K. Het bewijs hiervan dat hier achterwege blijft, maakt gebruik van het konkaaf zijn van de logaritme-functie.

Hiermee volgt dat  $\hat{\lambda}_1, \dots, \hat{\lambda}_k$  gelijk zijn aan (een permutatie van)  $\eta_1 \dots \eta_k$ , en  $\hat{v} = \frac{1}{p-k} \sum_{j=k+1}^p \eta_j$ , en nu is

$$\sum_{i=1}^k \log \hat{\lambda}_i + (p-k) \log \hat{v} = \sum_{i=1}^k \log \eta_i + (p-k) \left( \frac{\eta_{k+1} + \dots + \eta_p}{p-k} \right)$$

Concluderend is

$$\log L_{\omega} = c - \frac{1}{2}n \log \left\{ \eta_1 \dots \eta_k \left( \frac{\eta_{k+1} + \dots + \eta_p}{p-k} \right)^{p-k} \right\} - \frac{1}{2}np.$$

Zonder bijvoorwaarden

$$\log L_{\Omega} = c - \frac{1}{2}n \log \prod_{i=1}^p \eta_i - \frac{1}{2}np.$$

Derhalve is

$$\begin{aligned} -2 \log \lambda &= 2(\log L_{\Omega} - \log L_{\omega}) \\ &= n \log \left\{ (\eta_{k+1} \dots \eta_p)^{-1} \left( \frac{\eta_{k+1} + \dots + \eta_p}{p-k} \right)^{p-k} \right\}. \end{aligned}$$

Het gedeelte tussen accoladen is de (p-k) de macht van het rekenkundig gemiddelde van de p-k kleinste eigenwaarden van  $S^2$ , gedeeld door het meetkundig gemiddelde. Zoals we al eerder gezien hebben, is het aantal geschatte parameters in  $\Omega$  gelijk aan  $\frac{1}{2}p(p+1)$ , nl. het aantal varianties en covarianties. We kunnen de dimensie van  $\Omega$  ook a.v. vinden: we schatten p eigenwaarden en p eigenvektoren, die genormeerd en onderling loodrecht zijn, d.w.z.  $\sum_{j=1}^p (p-j)$  parameters bepalen de eigenvektoren (nl. p componenten, 1 eraf voor normering, j-1 eraf voor het loodrecht zijn).

De dimensie van  $\Omega$  is  $p + \sum_{j=1}^p (p-j) = p + p^2 - \frac{1}{2}p(p+1) = \frac{1}{2}p(p+1)$ .

In  $\omega$  schatten we k+1 eigenwaarden en k eigenvektoren. De dimensie van  $\omega$  is dus  $k+1 + \sum_{j=1}^k (p-j) = k+1 + pk - \frac{1}{2}k(k+1)$ .

Het aantal vrijheidsgraden is dus

$$\frac{1}{2}p(p+1) - k-1 - pk + \frac{1}{2}k(k+1) = \frac{1}{2}(p-k-1)(p-k+2).$$

Rotatie. Het zoeken van de hoofdcomponenten kunnen we op de manier van §4 ook meetkundig bekijken. Het zoeken van k lineaire combinaties  $y_1, \dots, y_k$  van  $x_1, \dots, x_p$  zo dat  $\sum_{i=1}^k \text{var } y_i$  maximaal is, komt overeen met het bepalen van een

$k$ -dimensionale deeltuimte van  $\mathbb{R}_p$ , met  $\mathbb{R}_p$  opgespannen door  $\underline{x}_1, \dots, \underline{x}_p$ , z6 dat de som van de kwadraten van de lengtes van de projecties van  $\underline{x}_1, \dots, \underline{x}_p$  op die deelruimte maximaal is. De vectoren  $\underline{y}_1, \dots, \underline{y}_k$  vormen een orthonormale basis van die deelruimte. Het kan echter zijn dat een andere basis meer inzicht geeft in de samenhang van de variabelen. Zo'n andere basis kunnen we vinden door  $\underline{y}_1, \dots, \underline{y}_k$  te roteren. Hoe we roteren, en volgens welke criteria we verschillende bases vergelijken wordt in §9 beschreven.

#### Literatuur

Anderson, T.W.: "Introduction to multivariate statistical analysis". John Wiley & Sons Inc., New York, 1958.

Anderson, T.W.: "Asymptotic theory for principal component analysis". Annals of Mathematical statistics 34 (1963) pp. 122-148.

Hottelling, H.: "Analysis of a complex of statistical variables into principal components". Journal of educational psychology 24 (1933) pp. 417-441, 498-520.

Meer over componentenanalyse vindt men in

Harman, H.H.: "Modern factor analysis", the university of Chicago press. Chicago & London, 1960.

Morrison, D.F.: "Multivariate statistical methods". Mc. Graw-Hill Book Company, 1967.

Tevens geven deze werken uitgebreide literatuur-verwijzingen.

## 9. Factoranalyse

### Inleiding

De factoranalyse is ontstaan in de psychologie. Het doel is om uit eigenschappen die aan individuen zijn waargenomen of uit testcores, informatie te krijgen over niet direct meetbare eigenschappen, de zogenaamde factoren.

Stel er zijn  $p$  eigenschappen  $\underline{x}_i$  ( $i=1, \dots, p$ ) gegeven. Het model dat in de factoranalyse gebruikt wordt, is nu dat elke  $\underline{x}_i$  geschreven kan worden als de som van zijn verwachting, een lineaire combinatie van  $k$  ( $k < p$ ) factoren  $\underline{y}_j$  en een storings-term  $\underline{e}_i$ .

$$9.1. \quad \underline{x}_i = \mu_i + \sum_{j=1}^k f_{ij} \underline{y}_j + \underline{e}_i \quad (i=1, \dots, p; k < p)$$

We noemen  $f_{ij}$  de  $i$ -de factorlading op de  $j$ -de factor. In matrixnotatie

$$9.2. \quad \underline{x} = \mu + F\underline{y} + \underline{e} \quad (F = F_{p \times k}, k < p)$$

We veronderstellen hierbij

$$9.3. \quad \underline{E}\underline{y} = 0; \underline{E}\underline{e} = 0; \text{VAR } \underline{y} = I_k; \text{VAR } \underline{e} = V^2 = \text{diag}(v_1^2, \dots, v_p^2); \text{COV}(\underline{y}, \underline{e}) = 0.$$

Uit 9.2 en 9.3 volgt

$$9.4. \quad \Sigma^2 := \text{VAR } \underline{x} = FF' + V^2$$

Dit is de fundamentele formule van de factor-analyse. Uitgeschreven luidt ze

$$9.5. \quad \sigma_{ii}^2 = \sum_{j=1}^k f_{ij}^2 + v_i^2 \quad (i=1, \dots, p)$$

$$9.6. \quad \sigma_{ij}^2 = \sum_{l=1}^k f_{il} f_{jl} \quad (i, j=1, \dots, p; i \neq j)$$

We noemen

$$9.7. \quad h_i^2 := \sum_{j=1}^k f_{ij}^2 \quad (i=1, \dots, p)$$

de communaliteit en  $v_i^2$  de uniciteit of specifieke variantie van  $\underline{x}_i$ .

Het factoranalyse-model onderscheidt zich van het in § 5 besproken regressie-model doordat de variabelen  $\underline{y}$  in 9.2 niet direct kunnen worden waargenomen.

Het verschil tussen componenten-analyse en factor-analyse bestaat daarin, dat in de componenten-analyse gepoogd wordt met  $k$  lineaire combinaties zoveel



mogelijk variantie te verklaren, terwijl in de factor-analyse  $k$  onbekende variabelen de covarianties bepalen. In meetkundige termen zij  $D$  de  $p$ -dimensionale ruimte, opgespannen door  $\underline{x}_1 - \mu_1, \dots, \underline{x}_p - \mu_p$ ; de orthogonale projector op een deelruimte  $K \subset D$  geven we aan met  $P_k$ . In de componenten-analyse zoeken we  $K$  met dimensie  $k$  zo, dat  $\|P_k(\underline{x}_1 - \mu_1)\|^2 + \dots + \|P_k(\underline{x}_p - \mu_p)\|^2$  maximaal is, in de factor-analyse gaan we ervan uit, dat  $\underline{x}_1 - \mu_1 - \underline{e}_1, \dots, \underline{x}_p - \mu_p - \underline{e}_p$ , met  $\underline{e}_1, \dots, \underline{e}_p$  onderling onafhankelijke meetfouten, in een  $k$ -dimensionale ruimte liggen (die geen deelruimte van  $D$  is).

Geschiedenis. Het eerste factoranalyse-model was het "twee factor-model" van Charles Spearman (1904). In dit model nam Spearman aan dat intellectuele vaardigheden verklaard worden door een algemene factor en voor iedere vaardigheid een speciale factor.

In de hierboven ingevoerde notatie:  $k=1$ . Uit 9.4 volgt dat dit model alleen voldoet als

$$\sigma_{ij}^2 \sigma_{lm}^2 - \sigma_{lj}^2 \sigma_{im}^2 = 0 \quad (i, j, l, m = 1, \dots, p; i \neq j \neq l \neq m \neq i).$$

Karl J. Holzinger kwam hierna (1925-1930, zie Holzinger & Harman, 1941) met het "bi-factor-model", waarin iedere variabele geschreven werd in termen van een algemene factor, een groepsfactor en een specifieke factor. Bij  $m$  groepsfactoren betekent dit, dat de factormatrix  $m+1$  kolommen heeft, en in elke rij het eerste en één van de overige elementen niet nul zijn.

Als toegestaan wordt, dat de groepen elkaar overlappen, komen we bij het "multiple-group-model" van Horst (1937) en Guttman (1944). Dit model komt al dicht in de buurt van de verder in deze paragraaf te behandelen maximum-likelihood-factoranalyse. De ontwikkeling van dit model is voornamelijk het werk van Lawley (voor een overzicht, zie Lawley & Maxwell, 1963).

Identificatie. Voor de rest van de paragraaf nemen we aan dat de kansverdelingen van  $\underline{y}$  en  $\underline{e}$ , en dus ook die van  $\underline{x}$ , multinormaal zijn. Voorlopig nemen we ook aan dat  $k$  bekend is. Verderop in deze paragraaf zullen we zien hoe de veronderstelling dat  $k$  een zekere waarde heeft, getoetst kan worden.

We vragen ons nu af of, uit waarnemingen aan de variabelen  $\underline{x}$ , de elementen van  $F$  en  $V^2$  geschat kunnen worden. Dit is niet het geval. Wat we kunnen schatten uit waarnemingen zijn  $\mu$  en  $\Sigma^2$ , onder de voorwaarde dat  $\Sigma^2$  aan de fundamentele formule voldoet.

Echter, als

$$\Sigma^2 = FF' + V^2$$

dan geldt ook

$$9.8. \quad \Sigma^2 = GG' + V^2$$

Voor elke  $G = FA$  met  $A_{k \times k}$  orthogonaal, immers  $GG' = FAA'F' = FF'$ .

Het bij 9.8 behorende factor-model luidt

$$\begin{aligned} \underline{x} &= \mu + F'\underline{y} + \underline{e} \\ &= \mu + GA'\underline{y} + \underline{e} \\ &= \mu + G\underline{z} + \underline{e} \quad \text{met } \underline{z} := A'\underline{y} \end{aligned}$$

en dit model voldoet ook aan de eisen, zoals genoemd in 9.3:

$$\sum \underline{z} = 0; \text{VAR } \underline{z} = I; \text{COV } (\underline{z}, \underline{e}) = 0$$

Uit meetkundig oogpunt: bij elke orthogonale  $A_{k \times k}$  hoort een andere orthonormale basis in  $\mathbb{R}_k$ .

Opmerking. We kunnen ook de eis dat de factoren onafhankelijk moeten zijn, laten vervallen. Bij elke reguliere  $A_{k \times k}$  behoort een (eventueel "scheve") basis in  $\mathbb{R}_k$ . Voorlopig beperken we ons tot orthogonale matrices.

Omdat de ruimte der orthogonale matrices van  $k \times k$  dimensie  $\frac{1}{2}k(k-1)$  heeft (d.i.  $\sum_{j=1}^k (j-1)$ , vgl. §8), heffen we deze onbepaaldheid op door  $\frac{1}{2}k(k-1)$  beperkingen op te leggen aan  $F$ . In navolging van Lawley eisen we (tijdelijk) dat

$$9.9. \quad F'V^{-2}F = \Lambda$$

waarin  $\Lambda$  een onbekende diagonale  $k \times k$ -matrix is, ofwel:

$$9.10. \quad \sum_{t=1}^p f_{ti}f_{tj}/v_t^2 = 0 \quad (i, j = 1, \dots, k; i \neq j)$$

Dit zijn  $\frac{1}{2}k(k-1)$  beperkingen. Dat aan 9.9 voldaan kan worden, volgt uit 1.5.4.

Stel nl.  $F$  is een matrix die aan 9.4 voldoet. Dan is  $F'V^{-2}F$  symmetrisch, dus er bestaat een orthogonale matrix  $A$  zo, dat  $F'V^{-2}F = A \Lambda A'$  met  $\Lambda$  diagonaal.

Dan voldoet  $G := FA$  aan de fundamentele formule, en  $G'V^{-2}G$  is diagonaal.

De eis 9.9 is slechts tijdelijk, want hebben we uit waarnemingen  $F$  en  $V$  geschat, dan zoeken we een  $G$  met  $GG' = FF'$ , die "mooier" is dan  $F$ . Wat "mooier" hier betekent zien we verderop in deze paragraaf.

Eerst gaan we na of invoering van de eis 9.9 het ons mogelijk maakt om op grond van een steekproefvariantie-covariantiematrix  $S^2$  de matrices  $F$  en  $V^2$  te schatten. Uit  $S^2$  kunnen maximaal  $\frac{1}{2}p(p+1)$  parameters geschat worden. We moeten schatten: het aantal parameters van  $F$ , dit is het aantal elementen van  $F$ , verminderd met het aantal opgelegde beperkingen, plus het aantal parameters van  $V^2$ , dus totaal

$$pk - \frac{1}{2}k(k-1) + p$$

Een nodige voorwaarde om F en  $V^2$  uit  $S^2$  te kunnen schatten, is dus

$$\frac{1}{2}p(p+1) \geq kp + p - \frac{1}{2}k(k-1)$$

of, na enige herleiding

$$k \leq \frac{1}{2}\{(2p+1) - \sqrt{8p+1}\}. \quad (\text{Ledermann, 1939})$$

B.V.

p =	3	6	36
k ≤	1	3	28

Opmerking. De eis 9.9 neemt niet alle onbepaaldheid uit het factormodel weg. Alleen als  $FF'$  regulier is, ligt F, op een orthogonale transformatie na, vast. Maar we kunnen opmerken dat als  $FF'$  singulier is, k te groot verondersteld is. Een meer principiële moeilijkheid is, dat kan gelden

$$9.12. \quad FF' + V^2 = GG' + W^2 \text{ met } V^2 \neq W^2$$

Het is niet altijd zo, dat de niet-diagonaalelementen van  $FF'$  de diagonaal-elementen bepalen. Er zijn voorwaarden geformuleerd, waaronder 9.12 niet kan gelden. Het zou ons te ver voeren, om hier op in te gaan (geïnteresseerden worden verwezen naar Anderson and Rubin, 1956).

Een voorbeeld. Van 20 leerlingen zijn de cijfers gegeven van wiskunde, natuurkunde, scheikunde, frans, Duits en Engels.

WIS	NAT	SCH	FRA	DUI	ENG
9	8	10	6	7	8
9	8	6	6	5	7
8	8	7	5	6	7
7	8	7	5	5	7
7	9	7	5	5	5
8	9	7	4	6	6
9	7	7	8	7	8
7	6	6	9	7	9
5	5	6	7	6	7
7	7	7	5	6	8
6	6	7	9	8	7
6	5	6	5	5	6
5	6	6	6	6	6
6	4	6	7	8	7
6	6	4	5	5	5
4	4	5	10	9	9
6	5	6	9	8	8
9	9	7	4	4	7
8	8	9	6	7	7
8	8	8	6	7	8

We willen nagaan of we deze 6 variabelen terug kunnen brengen tot twee factoren. De steekproefvariantie-covariantiematrix  $S^2$  is

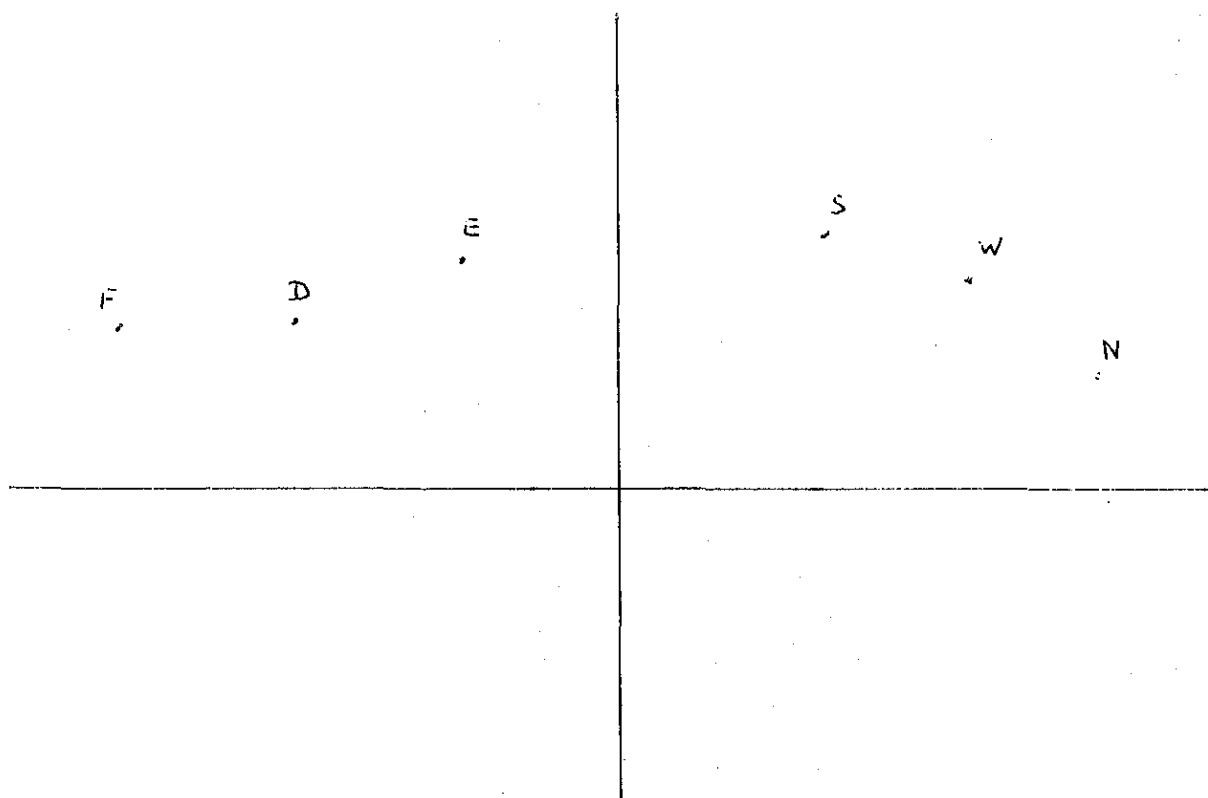
	WIS	NAT	SCH	FRA	DUI	ENG
WIS	2.21	1.95	1.21	-1.16	-.68	.11
NAT		2.69	1.25	-1.93	-1.19	-.45
SCH			1.69	-.52	.11	.30
FRA				3.19	1.92	1.28
DUI					1.71	.91
ENG						1.25

Via een rekenproces, dat hierna beschreven zal worden, vinden we

$$\hat{F}' = \begin{pmatrix} 1.08 & 1.46 & .57 & -1.53 & -1.01 & -.51 \\ .77 & .47 & .84 & .60 & .61 & .71 \end{pmatrix}$$

$\hat{V}^2$  heeft als diagonaalelementen:

.45    .33    .66    .48    .32    .44



De zes variabelen verklaard door twee factoren

De zes variabelen verklaard door twee geroteerde factoren.

$$\hat{\Sigma}^2 = \begin{bmatrix} 2.21 & 1.94 & 1.26 & -1.19 & -.62 & .02 \\ & 2.68 & 1.23 & -1.95 & -1.19 & -.40 \\ & & 1.69 & -.37 & -.06 & .33 \\ & & & 3.18 & 1.91 & 1.22 \\ & & & & 1.71 & .97 \\ & & & & & 1.25 \end{bmatrix}$$

De eerste factor verklaart het verschil tussen exacte vakken en talen, terwijl de tweede factor zoiets als "algemene aanleg" vertegenwoordigt.

De ML-schatters. We willen nu nagaan, hoe we uit een steekproef  $X_{n \times p}$  schatters voor  $F$ ,  $V^2$  en  $\Lambda$  vinden.

Voor een behandeling van de verschillende type schatters, zie het college-dictaat "Kansrekening en Statistiek", §6. De afleiding van de maximum-likelihoodschatters  $\hat{F}$ ,  $\hat{V}^2$  en  $\hat{\Lambda}$ , die nu volgt, is origineel van Lawley (1940). Na substitutie van  $\hat{\mu} = \hat{x}$  geldt (zie B9)

$$9.13. \quad \text{Log } L = C - \frac{1}{2}n \log |\Sigma^2| - \frac{1}{2}n \text{sp}(S^2 \Sigma^{-2}).$$

We willen log L maximaliseren onder de bijvoorwaarden

$$9.14. \quad \Sigma^2 = FF' + V^2$$

$$9.15. \quad \Lambda = F'V^{-2}F$$

$$9.16. \quad F \text{ afmetingen } p \times k, V^2 \text{ diagonaal } p \times p, \Lambda \text{ diagonaal } k \times k$$

Met B14 en B15 vinden we dat in stationaire punten geldt

$$9.17. \quad F' \Sigma^{-2} S^2 = F'$$

$$9.18. \quad (\Sigma^{-2} S^2 \Sigma^{-2} - \Sigma^{-2})_{ii} = 0 \quad (i = 1, \dots, p)$$

We kunnen het stelsel 9.14 t/m 9.18 vereenvoudigen. Eerst wordt 9.18 voorvermenigvuldigd met de diagonaalmatrix  $\Sigma^2 - FF'$ . Dit levert

$$9.19. \quad \text{diag}(S^2 \Sigma^{-2} - FF' \Sigma^{-2} S^2 \Sigma^{-2} - I + FF' \Sigma^{-2}) = 0$$

Nu is volgens 9.17  $F' \Sigma^{-2} S^2$  gelijk aan  $F'$ , dus 9.19 wordt

$$9.20. \quad \text{diag}(S^2 \Sigma^{-2} - I) = 0$$

Dit wordt navermenigvuldigd met  $\Sigma^2 - FF'$  en weer wordt 9.17 toegepast. We vinden dan

$$9.21. \quad \text{diag}(S^2 - S^2 \Sigma^{-2} FF' - \Sigma^2 + FF') = 0$$

$$9.22. \quad \text{diag}(S^2 - \Sigma^2) = 0, \text{ ofwel } V^2 = \text{diag}(S^2 - FF')$$

We definiëren nu

$$9.23. \quad U := V^{-1}F$$

Hiermee herschrijven we 9.14, 9.15 en 9.17.

$$9.24. \quad \Sigma^2 = VUU'V + V^2$$

$$9.25. \quad A = U'U$$

$$9.26. \quad S^2\Sigma^{-2}VU = VU$$

Nu wordt 9.24 voorvermenigvuldigd met  $\Sigma^{-2}$  en navermenigvuldigd met  $V^{-1}U$ :

$$9.27. \quad V^{-1}U = \Sigma^{-2}VUU'U + \Sigma^{-2}VU$$

Volgens 9.26 is  $\Sigma^{-2}VU = S^{-2}VU$ . Pas dit toe op 9.27, pas hierop ook 9.25 toe:

$$9.28. \quad V^{-1}U = S^{-2}VUA + S^{-2}VU$$

Dit wordt voorvermenigvuldigd met  $V^{-1}S^2$

$$9.29. \quad V^{-1}S^2V^{-1}U = UA + U$$

ofwel

$$9.30. \quad (V^{-1}S^2V^{-1} - I)U = UA$$

De oplossingen van het stelsel 9.22, 9.24, 9.25, 9.30 zijn de stationaire punten van  $\log L$ . We willen het maximum vinden. We nemen daartoe eerst  $V$  vast. Omdat  $U'U$  en  $UU'$  dezelfde eigenwaarden ongelijk nul hebben (1.72) kunnen we met 9.24 afleiden dat

$$9.31. \quad \log |\Sigma^2| = \log |V^2| + \sum_{i=1}^k \log(1 + \lambda_i).$$

Gebruikmakend van 9.24, 9.30 en

$$(I + UU')^{-1} = I - U(\Lambda + I)^{-1}U'$$

valt af te leiden (ga na) dat

$$9.32. \quad \text{sp}(S^2\Sigma^{-2}) = \text{sp}(V^{-1}S^2V^{-1}) - \sum_{i=1}^k \lambda_i$$

Substitutie van 9.31 en 9.32 in 9.13 levert

$$9.33. \quad \log L = c - \frac{1}{2}n \log |V^2| - \frac{1}{2}n \text{sp}(V^{-1}S^2V^{-1}) - \frac{1}{2}n \sum_{i=1}^k (\log(1 + \lambda_i) - \lambda_i).$$

Het maximum van  $\log L$  wordt voor vaste  $V$  gevonden als  $\sum_{i=1}^k (\log(1+\lambda_i) - \lambda_i)$  minimaal is, wat betekent dat we in 9.30 de  $k$  grootste eigenwaarden van  $V^{-1}S^2V^{-1} - I$  moeten nemen.

Nu is de vraag hoe  $\hat{V}$  gevonden wordt. Dit gebeurt iteratief. Gestart wordt met een eerste schatting van  $V$ : gewoonlijk wordt gebruikt  $V_i^2 = S_{ii}^2 - \max_{j \neq i} S_{ij}^2$ .

Met een schatting voor  $V^2$  wordt het eigenwaardenprobleem opgelost, met een schatting van  $F$  wordt 9.22 toegepast.

Deze iteratie-methode is oorspronkelijk van Lawley (1942) en werd nader uitgewerkt door Tao (1955) en Maxwell (1964). De convergentie is erg slecht. Men kan zelfs aantonen dat het proces niet altijd convergeert. Een alternatieve manier is die van Corsten. Voor elke  $V^2$  definieert hij  $f_v$  als het maximum van 9.33 en hij zoekt een richting in  $\mathbb{R}_p$  waarin  $f_v$  toeneemt. In die richting doet het iteratieproces een stap.

Opmerking. In het factor-analysemodel van Jöreskog (1963) komt in plaats van de eis 9.9 de eis

$$V^2 = \alpha(\text{diag } \Sigma^{-2})^{-1} \text{ met } \alpha \in \mathbb{R}^+$$

Hierbij hoort ook een ander iteratieproces.

Opmerking. Door 9.23 en 9.25 worden de eigenschappen in een fundamentele schaal uitgedrukt. Aan 9.30 is n.l. te zien dat de methode invariant is voor een schaalverandering in de  $x_i$ . Een schaalverandering in de  $x_i$  kan n.l. worden opgevangen door een zelfde schaalverandering in de  $f_{ij}$  en in  $V_i$  (Vergelijk ook 9.22). Op grond van deze overwegingen kan ook worden uitgegaan van de correlatiematrix in plaats van de variantiematrix. Deze schaalinvariantie geeft de factoranalyse volgens deze methode een groot voordeel t.o.v. de in de vorige paragraaf besproken componentenanalyse.

Rotatie. Als we de maximum-likelihoodschatters  $\hat{F}$  en  $\hat{V}^2$  gevonden hebben, kunnen we nagaan wat er gebeurt als we in plaats van

$$\underline{x} = \mu + F\underline{y} + \underline{e}$$

het equivalente model

$$\underline{x} = \mu + G\underline{z} + \underline{e}$$

beschouwen, waarin  $G$  en  $\underline{z}$  afhangen van een reguliere matrix  $A_{k \times k}$  en wel zo dat

$$G = FA, \underline{z} = A^{-1}\underline{y}$$

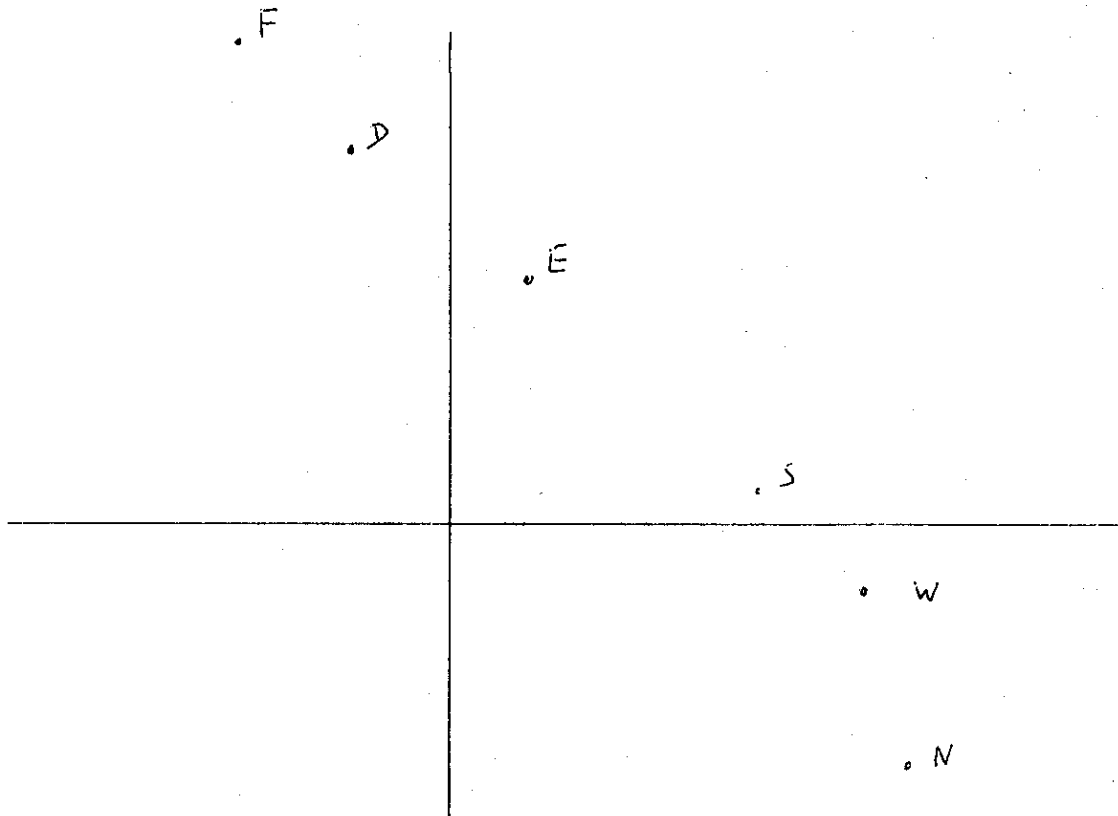
We bekijken eerst het geval dat A orthogonaal is.

In meetkundige termen:  $\underline{y}_1, \dots, \underline{y}_k$  vormen een orthonormaal assenstelsel en  $A^{-1}\underline{y}_1, \dots, A^{-1}\underline{y}_k$  ook en deze spannen beide  $\mathbb{R}_k$  op; de rijen van F drukken de vectoren  $\underline{x}_1 - \mu_1 \underline{e}_1, \dots, \underline{x}_p - \mu_p \underline{e}_p$  uit in het eerste stelsel, die van G in het tweede stelsel.

We gaan even terug naar ons voorbeeld op blz. 9.6 t/m 9.8. De matrix

$$\begin{aligned} \hat{G}' &= \begin{pmatrix} 1.30 & 1.35 & 1.00 & -.62 & -.26 & .18 \\ -.25 & -.73 & .17 & 1.52 & 1.115 & .88 \end{pmatrix} \\ &= \begin{pmatrix} .69 & .72 & 1.08 & 1.46 & .57 & -.153 & -1.01 & -.51 \\ -.72 & .69 & .77 & .47 & .84 & .60 & .61 & .74 \end{pmatrix} \end{aligned}$$

geeft een duidelijker beeld van de factoren.





We zien nu een "exacte factor" en een "talenfactor".

Algemeen willen we een nieuw assenstelsel zo, dat de p punten zo dicht mogelijk bij de k assen liggen. Er zijn vele methodes bekend om zo'n assenstelsel te vinden.

Bij de quartimax-methode (zie Sauders, 1949, Carroll.1953, Neuhaus en Wrigley, 1954) proberen we voor ieder punt de kwadraten van de coördinaten zo extreem mogelijk te maken. Als  $g_{j1}, \dots, g_{jk}$  de coördinaten van het  $j^e$  punt zijn, willen we de variantie van  $g_{j1}, \dots, g_{jk}$  maximaliseren. We definiëren

$$\sigma_{g_j}^2 := \frac{1}{k} \sum_{i=1}^k (g_{ji}^2 - \frac{1}{k} \sum_{\ell=1}^k g_{j\ell}^2)^2$$

en willen nu

$$\sum_{j=1}^p \sigma_{g_j}^2$$

maximaliseren. Echter, omdat

$$\begin{aligned} \sigma_{g_j}^2 &= \frac{1}{k} \sum_{i=1}^k g_{ji}^4 - \frac{1}{k^2} (\sum_{\ell=1}^k g_{j\ell}^2)^2 \\ &= \frac{1}{k} \sum_{i=1}^k g_{ji}^4 - \frac{1}{k^2} h_j^4 \end{aligned}$$

en  $h_j$  onafhankelijk is van de keuze van het assenstelsel, valt het maximum van  $\sum_{j=1}^p \sigma_{g_j}^2$  samen met het maximum van

$$9.34. \quad Q := \sum_{j=1}^p \sum_{i=1}^k g_{ji}^4$$

Dit verklaart de naam quartimax. Het quartimax-criterium luidt nu dat we Q moeten maximaliseren over alle

$$G = FA, \quad A \text{ direct orthogonaal}$$

(De beperking tot direct orthogonale matrices is niet essentieel). Hoe vinden we A? Voor  $k = 2$  is A van de vorm

$$A = \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix}$$

en vinden we G uit een rotatie van de assen. Voor  $k > 2$  moeten we een iteratief proces toepassen waarin bij elke iteratie voor elk paar assen een rotatie

uitgevoerd wordt (d.w.z.  $\frac{1}{2}k(k-1)$  rotaties per iteratie). De toename van Q en de grootste hoek waarover geroteerd is, zijn stopcriteria.

Ook bij de varimax-methode (ontwikkeld door Henry F. Kaiser, 1958) proberen we de kwadraten van de elementen van G zo extreem mogelijk te maken, maar we werken hierbij kolomsgewijs.

$$\sigma_j^2 := \frac{1}{p} \sum_{i=1}^p (g_{ij}^2)^2 - \frac{1}{p^2} \left( \sum_{i=1}^p g_{ij}^2 \right)^2 \quad (j=1, \dots, k)$$

is de variantie van de gekwadrateerde getallen uit de  $j^e$  kolom van G. Omdat de getallen  $g_{ij}$  ( $j=1, \dots, k$ ) afhankelijk zijn van de schaling van  $\underline{x}_i$ , en  $g_{ij}/h_i$  niet, prefereerde Kaiser maximalisatie van

$$9.35 \quad V := p \sum_{j=1}^k \sum_{i=1}^p (g_{ij}/h_i)^4 - \sum_{j=1}^k \left( \sum_{i=1}^p g_{ij}^2/h_i^2 \right)^2$$

boven maximalisatie van  $\sum_{j=1}^p \sigma_j^2$  als varimax-criterium. Maximalisatie van V moet gebeuren over alle

$$G = FA, \quad A \text{ direct orthogonaal.}$$

De werkwijze is analoog aan die bij de quartimax-methode.

Opmerking. De vele methoden die gebruikt worden bij niet orthogonale (scheve of "oblique") transformaties, worden hier niet behandeld. Een overzicht vindt men in Catell en Khanna, 1977.

Opmerking. In de componenten-analyse spannen de k hoofdcomponenten een k-dimensionale deelruimte van  $\mathbb{R}_p$  op. Ook hier kan een andere basis van  $\mathbb{R}_k$  meer inzicht geven in de structuur van het probleem. Op de componentenmatrix G kunnen dezelfde rotatietechnieken toegepast worden als op de factormatrix F.

Factorscores. We zijn uitgegaan van de scores van n individuen op p tests, en hebben de p tests teruggebracht tot k factoren. We vragen ons nu af, wat de scores zijn op deze factoren. Thompson (1951) geeft de volgende afleiding. Uit (9.2) en (9.3) volgt dat de variantie-covariantiematrix van  $\begin{pmatrix} \underline{x} \\ \underline{y} \end{pmatrix}$  is

$$\begin{pmatrix} FF' + V^2 & F \\ F' & I \end{pmatrix}$$

De matrix der regressiecoëfficiënten van  $\underline{x}$  op  $\underline{y}$  (zie collegedictaat lineaire

Modellen) is

$$F'(FF' + V^2)^{-1}$$

Als  $x_i$  de  $i^e$  waarneming is  $\bar{x}$  het gemiddelde over de  $n$  waarnemingen, dan is de schatting van de  $i^e$  factorscore

$$y_i = \hat{F}' (\hat{F}\hat{F}' + \hat{V}^2)^{-1} (x_i - \bar{x})$$

en hiermee wordt de  $n \times k$  matrix  $Y$  van factorladingen

$$9.36. \quad Y = (X - \bar{X})(\hat{F}\hat{F}' + \hat{V}^2)^{-1} \hat{F}$$

Uit (9.14) en (9.17) volgt dat

$$S^2(\hat{F}\hat{F}' + \hat{V}^2)^{-1} \hat{F} = \hat{F}$$

dus we kunnen (9.36) herleiden tot

$$9.37. \quad Y = (X - \bar{X})S^{-2} \hat{F}$$

Hierin moeten we een  $p \times p$  matrix inverteren. Dit kunnen we als volgt vermijden:

$$\begin{aligned} \hat{F}(I + \hat{F}'\hat{V}^{-2}\hat{F}) &= \hat{F} + \hat{F}\hat{F}'\hat{V}^{-2}\hat{F} \\ &= (\hat{V}^2 + \hat{F}\hat{F}')\hat{V}^{-2}\hat{F} \end{aligned}$$

ofwel

$$(\hat{F}\hat{F}' + \hat{V}^2)^{-1} \hat{F} = \hat{V}^{-2} \hat{F} (I + \hat{F}'\hat{V}^{-2}\hat{F})^{-1}$$

Uit (9.14) en (9.17) volgt  $S^{-2} \hat{F} = \hat{\Sigma}^{-2} \hat{F} = (\hat{F}\hat{F}' + \hat{V}^2)^{-1} \hat{F}$ , dus we kunnen in plaats van 9.37 ook schrijven

$$9.38. \quad Y = (X - \bar{X})\hat{V}^{-2} \hat{F} (I + \hat{F}'\hat{V}^{-2}\hat{F})^{-1}$$

Bij deze berekening van de thomsonfactorscores hoeven we slechts een  $k \times k$  matrix te inverteren.

Volgens de methode van Bartlett (1938) vinden we de factorscores als oplossing van een kleinste kwadraten-probleem. Voor  $t=1, \dots, n$  minimaliseren we

$$9.39. \quad \sum_{j=1}^p \hat{V}_j^{-2} e_{ij}^2 = (x_i - \bar{x} - \hat{F}' y_i)' \hat{V}^{-2} (x_i - \bar{x} - \hat{F}' y_i)$$

Uit differentiëren naar  $y_i$  volgt

$$y_i = (\hat{F}'\hat{V}^{-2}\hat{F})^{-1} \hat{F}'\hat{V}^{-2} (x_i - \bar{x})$$

ofwel

$$9.40. \quad Y = (X - \bar{X}) \hat{V}^{-2} \hat{F} (\hat{F}' \hat{V}^{-2} \hat{F})^{-1}$$

De factorscores van de 20 leerlingen uit ons voorbeeld zijn ter illustratie berekend, waarbij de geroteerde factoren gebruikt zijn. In de eerste kolom staan de scores op de "exacte factor", in de tweede die op de "talenfactor".

thomsonfactorscores

1.57	.70
.65	-.46
.62	-.38
.28	-.76
.33	-1.19
.78	-.86
.81	.94
-.03	1.18
-1.21	-.13
.81	-.22
-.31	1.09
-1.18	-1.04
-1.04	-.60
-1.10	.57
-1.32	-1.45
-1.41	1.76
-.65	1.23
1.02	-1.18
1.02	.34
.99	.45

bartlett factorscores

1.78	.86
.70	-.47
.66	-.39
.27	-.82
.31	-1.30
.82	-.90
.95	1.08
.02	1.30
-1.35	-.21
.19	-.24
-.29	1.18
-1.36	-1.21
-1.19	-.72
-1.19	.58
-1.54	-1.67
-1.48	1.87
-.65	1.32
1.07	-1.25
1.15	.43
1.12	.55

Opmerking. Volgens 9.22 geldt

$$s_i^2 = s^2(x_i) = \sum_{j=1}^k f_{ij}^2 + v_i^2$$

Dus

$$9.41. \quad sp S^2 = \sum_{i=1}^p s^2(x_i) = \sum_{i=1}^p \sum_{j=1}^k f_{ij}^2 + \sum_{i=1}^p v_i^2 = sp (FF') + sp V^2.$$

of na schaaltransformatie:

$$9.42. \quad sp(V^{-1} S^2 V^{-1}) = sp(V^{-1} FF' V^{-1}) + sp I_{p \times p} = sp \Lambda + p$$

Als men in sommige beschrijvingen van toepassingen van factoranalyse leest dat "zoveel procent van de variantie door de k factoren is verklaard", wordt meestal bedoeld het percentage van  $sp S^2$  dat in 9.41 wordt uitgemaakt door

$$\sum_{i=1}^p \sum_{j=1}^k f_{ij}^2, \text{ dus } 100 \frac{sp FF'}{sp S^2}$$

Het aantal factoren. Nu gaan we na hoe getoetst kan worden of het aantal factoren  $k$  juist is gekozen (zie ook Lawley, 1940). De nulhypothese  $H_0$  is dan dat 9.2 het juiste model is en de alternatieve hypothese is dat de  $\underline{x}_i$  multinormaal verdeeld zijn zonder verdere restricties. In beide gevallen is  $\hat{\mu} = \bar{x}$  en de maximalisering kan gericht worden op

$$\log L = c - \frac{1}{2}n \log |\Sigma^2| - \frac{1}{2}n \text{sp}(S^2 \Sigma^{-2})$$

Onder de veronderstelling  $\Omega$  (zie paragraaf 7) wordt weer gevonden (vgl. B12)

$$\log L_{\Omega} = c - \frac{1}{2}n \log |S^2| - \frac{1}{2}np$$

Verder is

$$\log L_{\omega} = c - \frac{1}{2}n \log |\hat{\Sigma}^2| - \frac{1}{2} \text{sp}(S^2 \hat{\Sigma}^{-2}),$$

waarin  $\hat{\Sigma}^2 = \hat{F}\hat{F}' + \hat{V}^2$  voldoet aan 9.22, 9.23 en 9.30. Als deze vergelijkingen exact zijn opgelost, geldt volgens 9.20

$$\text{diag}(S^2 \hat{\Sigma}^{-2} - I) = 0$$

Dus ook hier is  $\text{sp}(S^2 \hat{\Sigma}^{-2}) = p$

Dus

$$-2 \log \lambda = n \{ \log |\hat{\Sigma}^2| - \log |S^2| \} = -n \log |S^2 \hat{\Sigma}^{-2}|$$

Voor grote  $n$  is  $-2 \log \lambda$  by benadering  $\chi^2$  - verdeeld met zoveel vrijheidsgraden als in  $\omega$  minder parameters zijn geschat dan in  $\Omega$ . Afgezien van de gemiddelden zijn dat

in  $\Omega$  :  $\frac{1}{2}p(p+1)$  (elementen van  $\Sigma^2$ )

in  $\omega$  :  $\{p(k+1) - \frac{1}{2}k(k-1)\}$ , zoals eerder afgeleid

Het aantal vrijheidsgraden is dus:

$$\frac{1}{2}p(p+1) - \{p(k+1) - \frac{1}{2}k(k-1)\} = \frac{1}{2}\{(p-k)^2 - (p+k)\}.$$

Een significante uitkomst duidt er op dat  $k$  te klein is.

#### Literatuur.

Anderson, T.W. and H. Rubin: Statistical inference in factor analysis, "Proceedings of the third Berkeley Symposium" (Vol.V), ed: J. Neyman, University of California Press. Berkeley and Los Angeles 1956, pp. 111-150.

Bartlett, M.S.: "Methods of estimating mental factors", Nature, vol. 141 (1938), pp. 609-610.

- Caroll, J.B.: "An analytical solution for approximating simple structure in factor analysis", *Psychometrika*, vol. 18 (1953), pp. 23-38.
- Catell, B.C. and D.V. Khanna: Principles and procedures for unique rotation in factor analysis, "Mathematical Methods for digital computers," vol. 3: Statistical methods for digital computers, Ed. A. Ralston, H.S. Wilf ; (0-ed.: K. Enslein, Wiley, 1977, pp. 166-202.
- Corsten, L.C.A.: "Factoranalyse", colloquium-voordrachten voor ISA-Research-Eindhoven, Sektie Mathematische Statistiek, ongepubliceerd.
- Guttman, L.: "General theory and methods for matrix factoring", *psychometrika* 9 (1944) pp. 1-16.
- Holzinger, K.J. and H.H. Harman: "Factor Analysis", University of Chicago Press, Chicago, I 11, 1941.
- Horst, P.: "A method of factor analysis by means of which all coordinates of the factor matrix are given simultaneously", *psychometrika* 2 (1937), pp. 225-236.
- Jöreskog, K.G.: "Statistical estimation in factor analysis", Almqvist & Wiksell, Uppsala, 1963.
- Kaiser, H.F.: "The varimax criterion for analytic rotation in factor analysis", *psychometrika* 23 (1958), pp. 187-200.
- Lawley, D.N.: "The estimation of factor loadings by the method of maximum likelihood" proceedings of the royal society of Edinburgh, series A, 60 (1940), pp. 64-82.
- Lawley, D.N.: "Further investigations in factor estimation", proceedings of the royal society of Edinburgh, series A, 61 (1942), pp. 176-185.
- Lawley, D.N. and A.E. Maxwell: "Factor Analysis as a statistical Method", Butterworth & Co. ltd., London, 1963.
- Ledermann, W.: "On a problem concerning matrices with variable diagonal elements", Proceedings on the Royal society of Edinburgh, series A 60 (1939) pp. 1-17.
- Maxwell, A.E.: "Calculating maximum likelihood factor loadings". *Journal of the royal statistical society, series A* 127 (1964), pp. 238-241.
- Neuhaus, J. and C. Wrigley : " The quartimax method: An analytical approach to orthogonal simple structure", *British journal of statistical psychology* 7 (1954) pp. 81-91.
- Rao, C.R.: "Estimation and tests of significance in factor analysis", *psychometrika* 20 (1955), pp. 93-111.
- Saunders, D.R.: "Factor analysis II: A note concerning rotation of axes to simple

structure", Educational and Psychological measurement 9 (1949), pp. 753-756.

Spearman, C.: "General intelligence objectively determined and measured",  
American Journal of psychology 15 (1904), pp. 201-293.

Thomson, G.H. "The factorial analysis of human ability", University of London  
Press. LTD, London, 1951.

Meer over factoranalyse vindt men in

Harman, H.H.: "Modern factor analysis", the university of Chicago press, Chicago &  
London, 1960.

Morrison, D.F.: "Multivariate statistical methods", McGraw-Hill Book Company, 1967.  
Tevens geven deze werken uitgebreide literatuur-verwijzingen.

§10. Canonieke correlatie

De multiële correlatie, zoals gedefinieerd in 4.4 geeft een maat voor het verband tussen één variabele  $\underline{x}_i$  en een verzameling van andere variabelen  $\underline{x}^{(2)}$ .

De multiële correlatiecoëfficiënt bleek het maximum te zijn van de correlatie tussen  $\underline{x}_i$  en een lineaire combinatie van  $\underline{x}^{(2)}$ , terwijl de lineaire combinatie waarvoor dit maximum bereikt wordt de beste lineaire benadering  $\hat{\underline{x}}_i$  van  $\underline{x}_i$  is (vergelijk pag. 4.3, punt b, en 4.5).

Dit idee werd door Hotelling in 1936 ggeneraliseerd om de samenhang tussen twee groepen van stochastische variabelen te bestuderen.

Stel

$$10.1. \quad \underline{x}' = (\underline{x}^{(1)'}, \underline{x}^{(2)'}) = (\underline{x}_1, \dots, \underline{x}_{p_1}, \underline{x}_{p_1+1}, \dots, \underline{x}_p)' , \quad p = p_1 + p_2 .$$

We nemen aan dat  $p_1 \leq p_2$ , eventueel is dit te bereiken door de groepen om te wisselen.

De variantie-covariantie matrix is:

$$10.2. \quad \text{VAR}(\underline{x}) = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} .$$

We zoeken  $p_1$  lineaire combinaties  $\underline{y}_i^{(1)}$  van de  $\underline{x}^{(1)}$  en eveneens  $p_1$  lineaire combinaties  $\underline{y}_i^{(2)}$  van de  $\underline{x}^{(2)}$  zodanig dat deze lineaire combinaties paarsgewijze een maximale correlatie hebben, dus  $\rho(\underline{y}_i^{(1)}, \underline{y}_i^{(2)})$  is maximaal, terwijl alle andere correlatiecoëfficiënten nul zijn en de varianties gelijk aan 1.

In formulevorm:

$$\begin{aligned} \underline{y}^{(1)} &= G_1' \underline{x}^{(1)} , & \text{met } G_1' \text{ een } p_1 \times p_1 \text{ matrix} \\ \underline{y}^{(2)} &= G_2' \underline{x}^{(2)} , & \text{met } G_2' \text{ een } p_1 \times p_2 \text{ matrix} \end{aligned}$$

zodanig dat

$$\text{VAR}(\underline{y}^{(1)}) = G_1' \Sigma_{11} G_1 = I$$

$$\text{VAR}(\underline{y}^{(2)}) = G_2' \Sigma_{22} G_2 = I$$

$$\text{COV}(\underline{y}^{(1)}, \underline{y}^{(2)}) = G_1' \Sigma_{12} G_2$$

diagonaal met maximale diagonaalelementen. Dit laatste wil zeggen dat deze



diagonaalelementen achtereenvolgens worden gemaximaliseerd. Dus eerst worden  $y_1^{(1)}$  en  $y_1^{(2)}$  bepaald met  $\text{var}(y_1^{(1)}) = \text{var}(y_1^{(2)}) = 1$  en  $\rho(y_1^{(1)}, y_1^{(2)})$  maximaal. Daarna  $y_2^{(1)}$  en  $y_2^{(2)}$  met varianties 1,  $\rho(y_1^{(1)}, y_2^{(1)}) = \rho(y_1^{(2)}, y_2^{(2)}) = 0$  en  $\rho(y_2^{(1)}, y_2^{(2)})$  maximaal, etc.

We zullen zien dat, als wij maximaliseren bovenstaande bijvoorwaarden, bovendien voldaan is aan  $\rho(y_1^{(1)}, y_2^{(2)}) = \rho(y_2^{(1)}, y_1^{(2)}) = 0$ .

Definieer:

$$\Sigma_{11} := \Sigma_1^2 \quad \text{en} \quad \Sigma_{22} := \Sigma_2^2 \quad (1.64)$$

en

$$L_1 := \Sigma_1 G_1 ; \quad L_2 := \Sigma_2 G_2 .$$

Wij eisen dus:

$$G_1' \Sigma_{11} G_1 = G_1' \Sigma_1' \Sigma_1 G_1 = L_1' L_1 = I ,$$

10.3.

$$G_2' \Sigma_{22} G_2 = G_2' \Sigma_2' \Sigma_2 G_2 = L_2' L_2 = I ,$$

terwijl onder de voorwaarde 10.3 de diagonaalelementen van

10.4.

$$G_1' \Sigma_{12} G_2 = L_1' \Sigma_1^{-T} \Sigma_{12} \Sigma_2^{-1} L_2 =: L_1' A L_2 ,$$

waarin  $A := \Sigma_1^{-T} \Sigma_{12} \Sigma_2^{-1}$ , maximaal moeten zijn.

Dat wil zeggen  $\ell_1' A \ell_2$  maximaal, als  $\ell_1'$  de eerste rij van  $L_1'$  en  $\ell_2$  de eerste kolom van  $L_2$  voorstelt. Bij vaste  $\ell_1$  is  $\ell_1' A \ell_2 = (A' \ell_1, \ell_2)$  maximaal als  $\alpha \ell_2 = A' \ell_1$ , dus

10.5.

$$\frac{1}{\alpha} \ell_1' A A' \ell_1 \text{ maximaal.}$$

$$\text{Maar } \ell_2' \ell_2 = \frac{1}{\alpha^2} \ell_1' A A' \ell_1 = 1 \Rightarrow \alpha = \sqrt{\ell_1' A A' \ell_1} .$$

Dus 10.5 gaat over in

$$\sqrt{\ell_1' A A' \ell_1} \text{ maximaal}$$

en dus

$$\ell_1' A A' \ell_1 \text{ maximaal,}$$

terwijl  $\ell_1' \ell_1 = 1$ .

Volgens 1.100 wordt dit bereikt door voor  $\ell_1$  de eigenvector te nemen die hoort bij de grootste eigenwaarde  $\lambda_1$  van  $AA'$ . Dus  $\rho = \alpha = \sqrt{\lambda_1}$ .

Of:

$$10.6. \quad \Sigma_1^{-T} \Sigma_{12} \Sigma_2^{-1} \Sigma_2^{-T} \Sigma_{21} \Sigma_1^{-1} \ell_1 = \ell_1 \lambda_1.$$

Nu is  $\ell_1 = \Sigma_1 g_1$  ( $g_1$  is de 1<sup>e</sup> kolom van  $G_1$ ), dus:

$$\Sigma_1^{-T} \Sigma_{12} \Sigma_2^{-1} \Sigma_{21} g_1 = \Sigma_1 g_1 \lambda_1,$$

of:

$$10.7. \quad \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} g_1 = g_1 \lambda_1.$$

Door uit te gaan van het maximaliseren van  $\ell_2' A' \ell_1$  in plaats van  $\ell_1' A \ell_2$ , hetgeen neerkomt op het verwisselen van de indices 1 en 2 wordt in plaats van 10.6 en 10.7 gevonden:

$$10.8. \quad \Sigma_2^{-T} \Sigma_{21} \Sigma_1^{-1} \Sigma_1^{-T} \Sigma_{12} \Sigma_2^{-1} \ell_2 = \ell_2 \lambda_1$$

en

$$10.9. \quad \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} g_2 = g_2 \lambda_1.$$

Opmerking: Als  $\ell_1$  en  $\lambda_1$  berekend zijn kan  $\ell_2$  ook worden gevonden uit  $\sqrt{\lambda_1} \ell_2 = A' \ell_1$ .

De grootste eigenwaarden moeten bij 10.7 en bij 10.9 uiteraard gelijk zijn, omdat  $\lambda_1 = \rho^2$ .

De volgende kolommen van  $L_1$  en  $L_2$ , en dus van  $G_1$  en  $G_2$  worden gevonden als de eigenvectoren behorende bij de volgende eigenwaarden in volgorde van afnemende grootte van 10.6 en 10.8, respectievelijk 10.7 en 10.9. Dit volgt uit de afleiding zoals gegeven voor de grootste canonieke correlatie en uit 1.100.

Dus, in formulevorm:

$$AA' L_1 = L_1 \Lambda,$$

waarin de kolommen van  $L_1$  orthonormale eigenvectoren zijn en

$$L_2 \Lambda^{\frac{1}{2}} = A' L_1, \quad \text{of} \quad A' A L_2 = L_2 \Lambda.$$

Aan de verdere eisen is dan voldaan, immers

$$L_1' L_1 = I,$$

$$\Lambda^{\frac{1}{2}} L_2' L_2 \Lambda^{\frac{1}{2}} = L_1' A A' L_1 = \Lambda, \text{ dus } L_2' L_2 = I \text{ (mits alle } \lambda_i \neq 0)$$

$$L_1' A L_2 = \Lambda^{\frac{1}{2}} L_2' L_2 = \Lambda^{\frac{1}{2}} \text{ (diagonaal).}$$

Opmerkingen:

1)  $r(AA') = r(A'A) = r(A)$  (1.45).

$\Sigma_1$  en  $\Sigma_2$  zijn regulier, omdat  $\Sigma_{11}$  en  $\Sigma_{22}$  positief definit zijn. Dus  $r(AA') = r(A'A) = r(\Sigma_{12})$  (1.44).

Als we veronderstellen  $r(\Sigma_{12}) = p_1$ , dan hebben zowel  $AA'$  als  $A'A$   $p_1$  eigenwaarden  $> 0$ ,  $A'A$  heeft dan verder nog  $p_2 - p_1$  eigenwaarden  $= 0$ .

2) Men kan eventueel  $\underline{y}^{(2)}$  nog aanvullen met  $\underline{z}' = (z_1, \dots, z_{p_2 - p_1})$ , of

$\underline{z} = H' \underline{x}^{(2)}$  (waarin  $H'$  een  $(p_2 - p_1) \times p_2$  matrix is), zodanig dat  $H' \Sigma_{22} G_2 = 0$  en  $H' \Sigma_{22} H = I$ . Hierdoor is  $H$  niet bepaald. Door extra voorwaarden (zie Anderson p. 293) kan deze onbepaaldheid worden opgeheven.

Als  $H' \Sigma_{22} G_2 = 0$ , dan geldt ook dat  $H' \Sigma_{21} G_1 = 0$ .

Definieer nl.

$$K := \Sigma_2 H.$$

Dan geldt:

$$H' \Sigma_{21} G_1 = K' \Sigma_2^{-1} \Sigma_{21} \Sigma_1^{-1} L_1 = K' A' L_1 = K' L_2 \Lambda^{\frac{1}{2}} = H' \Sigma_2' \Sigma_2 G_2 \Lambda^{\frac{1}{2}} = H' \Sigma_{22} G_2 \Lambda^{\frac{1}{2}} = 0.$$

3) De variantie-covariantiematrix van  $\underline{y}$ , met  $\underline{y}' = (\underline{y}^{(1)'}, \underline{y}^{(2)'}, \underline{z}')$  is dan

$$\text{VAR } \underline{y} = \begin{pmatrix} I & \Lambda^{\frac{1}{2}} & 0 \\ \Lambda^{\frac{1}{2}} & I & 0 \\ 0 & 0 & I \end{pmatrix} \begin{matrix} p_1 \text{ rijen} \\ p_1 \text{ rijen} \\ (p_2 - p_1) \text{ rijen} \end{matrix}.$$

Zoals aan het begin van deze paragraaf reeds werd opgemerkt bestaat er een nauw verband tussen de in §4 behandelde multipele correlatie en de canonieke correlatie. We zagen op pag. 4.3 dat de correlatie tussen  $\underline{x}_i$  en een lineaire combinatie van  $\underline{x}^{(2)}$  maximaal werd indien als lineaire combinatie werd genomen  $\hat{\underline{x}}_i = \beta_{i*}' \underline{x}^{(2)}$ , waarin  $\beta_{i*}'$  de  $i^e$  rij is van  $B := \Sigma_{12} \Sigma_{22}^{-1}$ . Met een lineaire combinatie  $g_j' \underline{x}^{(1)}$  geeft dus  $g_j' B \underline{x}^{(2)}$  de hoogste correlatie. Dat betekent dat het maximum gevonden moet worden van

$$\text{cov}(g_1' \underline{x}^{(1)}, g_1' B \underline{x}^{(2)}) = g_1' \Sigma_{12} B' g_1 = g_1' \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} g_1$$

onder de bijvoorwaarde  $g_1' \Sigma_{11} g_1 = 1$ .

Of:

$$\lambda_1' \Sigma_{11}^{-T} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \lambda_1 \text{ maximaal,}$$

onder de voorwaarde  $\lambda_1' \lambda_1 = 1$ , hetgeen ons weer terugvoert tot 10.6.

- 4) In plaats van met de variantie-covariantiematrix van  $\underline{x}$  kan ook met de correlatiematrix worden gewerkt. Het is gemakkelijk na te gaan dat dit essentieel hetzelfde resultaat geeft.
- 5) Ook in het geval van de canonieke correlatie zijn toetsen ontwikkeld door Bartlett, Lawley, e.a., voor hypothesen zoals:  $k$  canonieke correlaties zijn  $> 0$ , de overige zijn  $= 0$  ( $k < p_1$ ). Daarbij moet dan weer verondersteld worden dat  $\underline{x}^{(1)}$  en  $\underline{x}^{(2)}$  multinormaal verdeeld zijn.

### §11. Discriminant analyse

Het probleem is het volgende. Gegeven zijn 2 of meer populaties en van elke populatie een steekproef. Gevraagd wordt een beslissingsregel, gebaseerd op deze waarnemingen, die ons in staat stelt een nieuw element met een zo groot mogelijke kans toe te wijzen aan de juiste populatie. Wij nemen aan dat geen twijfelgevallen worden toegelaten.

Voorlopig gaan we uit van 2 populaties. De variabelen hebben dimensie  $p$ . Een waarneming is dan een punt  $\underline{x}$  in  $R^p$ . De opgave is nu een gebied  $R \subset R^p$  te bepalen. Als  $\underline{x} \in R$  dan is onze conclusie dat we te maken hebben met de populatie met kansdichtheid  $f_1$  en als  $\underline{x} \in R^*$  (als  $R^*$  het complement van  $R$  voorstelt) dan luidt de conclusie dat de verdeling  $f_2$  is.

We stellen nu eerst als eis dat de kansen op een verkeerde toewijzing naar beide kanten even groot moet zijn:

$$11.1. \quad \int_R f_2 \, dx = \int_{R^*} f_1 \, dx = 1 - \int_R f_1 \, dx$$

of:

$$11.2. \quad \int_R (f_1 + f_2) \, dx = 1 .$$

Er geldt dan vanzelfsprekend ook:

$$11.3. \quad \int_{R^*} (f_1 + f_2) \, dx = 1 .$$

Willen we nu onder deze voorwaarde de kans op een verkeerde indeling minimaliseren, dan zoeken we dus het minimum van

$$\int_R f_2 \, dx$$

onder de voorwaarde

$$\int_R (f_1 + f_2) \, dx = 1 .$$

Of:

$$11.4. \quad \int_R \{f_2 - \lambda(f_1 + f_2)\} dx + \lambda$$

moet minimaal zijn. Of:

$$11.5. \quad \int_R \{-\lambda f_1 + (1 - \lambda)f_2\} dx + \lambda$$

minimaal. Hierin zijn  $R$  en  $\lambda$  de variabelen.

Het is duidelijk dat bij vaste  $\lambda$  het minimum wordt bereikt door voor  $R$  te nemen de verzameling van alle punten  $x$  waarvoor geldt

$$-\lambda f_1 + (1 - \lambda)f_2 < 0 .$$

De grens tussen  $R$  en  $R^*$  wordt dus gevormd door

$$11.6. \quad f_1/f_2 = (1 - \lambda)/\lambda =: \beta .$$

$\lambda$  kan dan worden berekend uit

$$11.7. \quad \int_R (f_1 + f_2) dx = 1 .$$

Het criterium 11.6 is blijkbaar een likelihood-ratio criterium. Het wordt nu verder uitgewerkt onder de aanname dat de beide verdelingen  $f_1$  en  $f_2$  multivariëer normaal zijn met dezelfde variantie-covariantiematrix  $\Sigma^2$ . De logaritme van de likelihood-ratio is dan:

$$\begin{aligned} & -\frac{1}{2}\{(x - \mu_1)' \Sigma^{-2}(x - \mu_1) - (x - \mu_2)' \Sigma^{-2}(x - \mu_2)\} = \\ & = -\frac{1}{2}\{-(\mu_1 - \mu_2)' \Sigma^{-2}x - x' \Sigma^{-2}(\mu_1 - \mu_2) + \mu_1' \Sigma^{-2}\mu_1 - \mu_2' \Sigma^{-2}\mu_2\} . \end{aligned}$$

Het van  $x$  afhankelijke gedeelte is:

$$11.8. \quad \frac{1}{2}\{(\mu_1 - \mu_2)' \Sigma^{-2}x + x' \Sigma^{-2}(\mu_1 - \mu_2)\} = (\mu_1 - \mu_2)' \Sigma^{-2}x .$$

Op de grens tussen  $R$  en  $R^*$  moet 11.8 constant zijn. Dit is de vergelijking van een hypervlak. Op symmetriegronden mag worden verwacht dat dit hypervlak het lijnstuk  $[\mu_1, \mu_2]$  doormidden deelt. Dat dit waar is is in te zien door de transformatie  $x^* = \Sigma^{-T}x$  toe te passen, waardoor de variabelen onafhankelijk worden en variantie 1 hebben, immers  $\text{var } \underline{x}^* = \Sigma^{-T}\Sigma'\Sigma^{-1} = I$ . De uitdrukking 11.8 gaat dan over in

11.9.  $(\mu_1^* - \mu_2^*)' x^* = \text{constant} .$

Dit is een hypervlak loodrecht op de lijn door de gemiddelden. De verdelingen zijn door de transformatie bolsymmetrisch geworden en de kansen op verkeerde klassificatie kunnen één-dimensionaal worden bekeken door de  $\mu_1^*, \mu_2^*$ -lijn als coördinaatas te nemen. De kansen worden gelijk als het scheidingsvlak het lijnstuk  $[\mu_1^*, \mu_2^*]$  doormidden deelt en dus wordt voor de transformatie ook  $[\mu_1, \mu_2]$  doormidden gedeeld. Dit bepaalt meteen de constante in 11.8, want we weten nu dat  $\frac{1}{2}(\mu_1 + \mu_2)$  in het scheidingshypervlak ligt. Dus de constante is:

11.10.  $C = \frac{1}{2}(\mu_1 - \mu_2)' \Sigma^{-2}(\mu_1 + \mu_2)$

en de discriminantfunctie is:

11.11.  $(\mu_1 - \mu_2)' \Sigma^{-2} x = \frac{1}{2}(\mu_1 - \mu_2)' \Sigma^{-2}(\mu_1 + \mu_2) .$

De geschatte discriminantfunctie, op grond van twee steekproeven, is:

11.12.  $(\bar{x}_1 - \bar{x}_2)' S^{-2} x = \frac{1}{2}(\bar{x}_1 - \bar{x}_2)' S^{-2}(\bar{x}_1 + \bar{x}_2) ,$

waarin  $S^2$  de "gepoolde" steekproefvariantie-covariantiematrix is (als in 7.7 met  $n_i$  i.p.v.  $v_i$ ). De lineariteit van 11.8 en 11.11 volgt uit de aanname van gelijke variantie-covariantiematrices. Laten we deze aanname los, dan wordt de logaritmische van de likelihood-ratio (afgezien van constanten)

11.13.  $(x - \mu_1)' \Sigma_1^{-2}(x - \mu_1) - (x - \mu_2)' \Sigma_2^{-2}(x - \mu_2) .$

De grens tussen beide gebieden wordt nu een kwadratisch oppervlak. Er is blijkbaar, ook voor het geval van meer dan twee populaties dat hierna wordt behandeld, behoefte aan een toets voor de gelijkheid van een aantal variantie-covariantiematrices. Deze toets is besproken in 7.14.

Er kunnen twee redenen zijn waarom de voorgaande behandeling van twee populaties gewijzigd moet worden.

- a) De kans dat we te maken hebben met een element van de eerste populatie kan sterk verschillen van de kans op een element uit de tweede populatie.
- b) De consequenties van een verkeerde klassificatie kunnen verschillen.

We stellen nu dat we de a priori kansen  $\pi_1$  en  $\pi_2$  kennen op het aantreffen van een element uit de betreffende populatie en verder dat de kosten van een verkeerde klassificatie resp.  $c_1$  en  $c_2$  bedragen. In plaats van de kans op een fout minimaliseren we nu de kosten. Dus we zoeken het minimum van

$$11.14. \quad c_2 \int_R \pi_2 f_2 dx + c_1 \int_{R^*} \pi_1 f_1 dx = c_1 \pi_1 + \int_R (c_2 \pi_2 f_2 - c_1 \pi_1 f_1) dx .$$

Het minimum wordt bereikt als we voor  $R$  het gebied nemen waarin geldt:

$$11.15. \quad f_1/f_2 > \frac{c_2 \pi_2}{c_1 \pi_1} .$$

Opnieuw dus een likelihood-ratio criterium, alleen met een andere grenswaarde dan in 11.6.

Tenslotte het geval van  $k > 2$  populaties, waarbij we aannemen dat de kosten  $c_i$  alle gelijk zijn. We hebben dus de verdelingsdichtheden

$$f_1, f_2, \dots, f_k$$

en de a priori kansen

$$\pi_1, \pi_2, \dots, \pi_k .$$

Als  $\underline{x} \in R_i$  zullen wij het betreffende element aan  $f_i$  toewijzen. De kans op een verkeerde klassificatie is dan:

$$11.16. \quad \int_{R_1^*} \pi_1 f_1 dx + \int_{R_2^*} \pi_2 f_2 dx + \dots + \int_{R_k^*} \pi_k f_k dx = \\ = (\pi_1 + \dots + \pi_k) - \int_{R_1} \pi_1 f_1 dx - \int_{R_2} \pi_2 f_2 dx - \dots - \int_{R_k} \pi_k f_k dx .$$

Dit is minimaal als

$$\sum_i \int_{R_i} \pi_i f_i dx$$

maximaal is en dat is het geval als  $R_i$  zodanig wordt gekozen dat



$$11.17. \quad \pi_i f_i = \max_{j=1, \dots, k} \pi_j f_j,$$

dus

$$R_i := \{x \mid \pi_i f_i(x) = \max_j \pi_j f_j(x)\}.$$

Bij multinormale verdelingen met dezelfde  $\Sigma^2$  en met gelijke a priori kansen betekent dit dat het maximum gezocht moet worden van

$$11.18. \quad -\frac{1}{2}(x - \mu_i)' \Sigma^{-2} (x - \mu_i) = -\frac{1}{2}\{x' \Sigma^{-2} x - x' \Sigma^{-2} \mu_i - \mu_i' \Sigma^{-2} x + \mu_i' \Sigma^{-2} \mu_i\}.$$

Er moet dus het maximum worden gekozen van

$$11.19. \quad \mu_i' \Sigma^{-2} x - \frac{1}{2} \mu_i' \Sigma^{-2} \mu_i.$$

Als  $k = 2$  is 11.19 weer eenvoudig te herleiden tot 11.11.

Op grond van een steekproef wordt het criterium: ken het element  $x$  toe aan de verdeling  $f_j$  als voor  $j$  het maximum wordt bereikt van

$$11.20. \quad \bar{x}_i' S^{-2} x - \frac{1}{2} \bar{x}_i' S^{-2} \bar{x}_i.$$

Er bestaan verschillende andere methoden van discriminant analyse. Wij zullen iets nader ingaan op de multipele discriminant analyse. Deze methode komt voort uit de benadering van Fisher (1936) voor twee populaties.

Stel weer dat wij te maken hebben met  $k$  populaties waarvan steekproeven bekend zijn. Wij hebben dus steekproeven  $X^{(i)}$  ( $i = 1, \dots, k$ ), waarbij  $X^{(i)}$  een

$n_i \times p$  matrix is, met  $r(X^{(i)}) = p$  voor minstens één  $i$ . Stel  $\sum_{i=1}^k n_i = n$ .

We definiëren:

$$(X')_{p \times n} := (X^{(1)'}, X^{(2)'}, \dots, X^{(k)'})$$

$$T' := (X - \bar{X})'$$

11.21.

$$W' := \{(X^{(1)} - \bar{X}^{(1)})', (X^{(2)} - \bar{X}^{(2)})', \dots, (X^{(k)} - \bar{X}^{(k)})'\}$$

$$A' := \{(\bar{X}^{(1)} - \bar{X})', (\bar{X}^{(2)} - \bar{X})', \dots, (\bar{X}^{(k)} - \bar{X})'\}.$$

Opmerking: Om het aantal boven- en onderindices te beperken komt  $\bar{X}$  in bovenstaande definities in verschillende betekenissen voor. Het gemiddelde van de vector  $x$  over alle  $n$  waarnemingen noemen we  $\bar{x}$  (vergelijk 7.1). De matrix  $\bar{X}$  bestaat nu telkens uit een aantal rijen die alle gelijk zijn aan  $\bar{x}$ . Het aantal rijen is telkens verschillend en blijkt uit de context.

Wij passen nu 7.3 toe op de  $i^e$  steekproef waarbij op de plaats van  $X$ ,  $\bar{X}$  en  $\mathcal{M}$  respectievelijk wordt gesubstitueerd:  $X^{(i)}$ ,  $\bar{X}^{(i)}$  en  $\bar{X}$  ( $n_i$  rijen). Er geldt dus blijkbaar voor elke  $i$ :

$$1.22. \quad (X^{(i)} - \bar{X})^2 = (X^{(i)} - \bar{X}^{(i)})^2 + (\bar{X}^{(i)} - \bar{X})^2 .$$

Sommatie over  $i$  levert ons

$$1.23. \quad T^2 = A^2 + W^2 .$$

Er wordt nu gezocht naar een lineaire combinatie  $\ell_1'x$  zodanig dat het quotiënt

$$\frac{\ell_1' A^2 \ell_1}{\ell_1' W^2 \ell_1} ,$$

maximaal is. Definieer  $D^2 := W^2$  met  $D_{p \times p}$  regulier. Stel  $D \ell_1 =: d_1$ , dan zoeken wij het maximum van

$$1.24. \quad \frac{d_1' D^{-T} A^2 D^{-1} d_1}{d_1' d_1} .$$

$D^{-T} A^2 D^{-1}$  is symmetrisch, dus volgens 1.100 is het maximum gelijk aan  $\lambda_1$ , de grootste eigenwaarde van  $D^{-T} A^2 D^{-1}$  en is  $d_1$  de bijbehorende eigenvector. Dus

$$1.25. \quad D^{-T} A^2 D^{-1} d_1 = d_1 \lambda_1 ,$$

of

$$D^{-1} D^{-T} A^2 D^{-1} d_1 = D^{-1} d_1 \lambda_1$$

of

$$1.26. \quad W^{-2} A^2 \ell_1 = \lambda_1 \ell_1 .$$

Dus  $\ell_1$  is een eigenvector behorende bij de grootste eigenwaarde van  $W^{-2} A^2$ . Een punt  $x$  wordt bij de  $i^e$  populatie ingedeeld als  $\ell_1' x$  het dichtst ligt bij  $\ell_1' \bar{x}_i$  ( $i = 1, \dots, k$ ).

Als  $k = 2$  wordt

$$A_2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{x}_1 - \bar{x}_2)(\bar{x}_1 - \bar{x}_2)' \quad (r(A^2) = 1) .$$

Verder is

$$W^2 = (n_1 + n_2) S^2 .$$

Dus

$$W^{-2} A^2 = \frac{n_1 n_2}{(n_1 + n_2)^2} S^{-2} (\bar{x}_1 - \bar{x}_2)(\bar{x}_1 - \bar{x}_2)' .$$

In dit geval is er één eigenwaarde  $> 0$ , want  $r(W^{-2} A^2) = r(A^2) = \min(k-1, p)$ .

De bijbehorende eigenvector is

$$11.27. \quad \ell = S^{-2} (\bar{x} - \bar{x}_2) .$$

Immers

$$\frac{n_1 n_2}{(n_1 + n_2)^2} S^{-2} (\bar{x} - \bar{x}_2)(\bar{x}_1 - \bar{x}_2)' S^{-2} (\bar{x}_1 - \bar{x}_2) = S^{-2} (\bar{x}_1 - \bar{x}_2) \lambda ,$$

als

$$\lambda = \frac{n_1 n_2}{(n_1 + n_2)^2} (\bar{x}_1 - \bar{x}_2)' S^{-2} (\bar{x}_1 - \bar{x}_2) .$$

Op de grens tussen beide gebieden is

$$\ell'x = \frac{1}{2}(\ell'\bar{x}_1 + \ell'\bar{x}_2) ,$$

hetgeen, na substitutie van 11.27 precies het criterium 11.12 oplevert.

In het algemeen is de rang  $r$  van  $W^{-2} A^2$  groter dan 1. Er zijn dan  $r$  eigenwaarden  $> 0$ . De eigenvectoren  $\ell_1, \dots, \ell_r$  die horen bij deze eigenwaarden, in volgorde van afnemende grootte gerangschikt, vormen de kolommen van een  $p \times r$  matrix  $L$ . De eigenvectoren worden zo genormeerd dat

$$11.28. \quad L'W^2L = I_{r \times r} .$$

Dit is mogelijk omdat, als  $\lambda_i \neq \lambda_j$ , geldt:

$$\lambda_i' A^2 \lambda_j = \lambda_i' W^2 \lambda_j = 0 .$$

Dit volgt geheel analoog aan het bewijs van 1.53 uit

$$A^2 \lambda_i = W^2 \lambda_i \lambda_i$$

en uit de symmetrie van  $A^2$  en  $W^2$ .

De matrix  $L'$  bepaalt een  $R^P \rightarrow R^r$  transformatie

$$y = L'x .$$

Men kan nu in deze  $R^r$  nagaan tot welk (getransformeerd) groepsgemiddelde  $L'x$  de kleinste Euclidische afstand heeft. Dan wordt  $x$  toegewezen aan die populatie waarvoor deze afstand

$$11.29. \quad (L'x - L'\bar{x}_i)'(L'x - L'\bar{x}_i) = (x - \bar{x}_i)'LL'(x - \bar{x}_i)$$

minimaal is.

We tonen nu aan dat de criteria 11.29 en 11.20 hetzelfde resultaat opleveren. Uit de afleiding van 11.20 zien we dat dit gelijkwaardig is met het zoeken van het minimum van

$$(x - \bar{x}_i)'S^{-2}(x - \bar{x}_i) .$$

Maar  $S^2$  en  $W^2$  zijn op een constante na aan elkaar gelijk, dus 11.20 komt neer op het vinden van de  $i$ -waarde die

$$11.30. \quad (x - \bar{x}_i)'W^{-2}(x - \bar{x}_i)$$

minimaal maakt.

We beschouwen eerst het geval  $r = p$ . Dan is volgens 11.28

$$W^{-2} = LL'$$

en dus zijn 11.30 en 11.29 aan elkaar gelijk.

Als  $r < p$  is dan vullen we de matrix  $L$  aan met  $p-r$  eigenvectoren die horen bij de  $p-r$  eigenwaarden gelijk aan 0. Deze vormen de matrix  $L_0$ . Ze kunnen zo worden gekozen dat, als

$$L_p := (L, L_0) ,$$

geldt

$$L_p' W_p^2 L_p = I_{p \times p}$$

en dus

$$11.31. \quad W^{-2} = L_p L_p' = LL' + L_0 L_0' .$$

Voor een kolom  $\ell_0$  van  $L_0$  geldt dat  $A^2 \ell_0 = 0$ .

Maar  $A^2$  kunnen we schrijven als (zie 11.21)

$$A^2 = \sum_{i=1}^k (\bar{X}^{(i)} - \bar{X})^2 = \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})' .$$

Dus

$$\ell_0' A^2 \ell_0 = \sum_{i=1}^k n_i \ell_0' (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})' \ell_0 = \sum_{i=1}^k n_i \{(\bar{x}_i - \bar{x})' \ell_0\}^2 = 0 .$$

Dus voor elke  $i$  geldt  $(\bar{x}_i - \bar{x})' \ell_0 = 0$  en dus

$$(\bar{x}_i - \bar{x})' L_0 = 0 \quad (i = 1, \dots, k)$$

en dus ook

$$(\bar{x}_i - \bar{x}_j)' L_0 = 0 \quad (i, j = 1, \dots, k) .$$

Hieruit volgt

$$(x - \bar{x}_i)' L_0 = c'(x)$$

en

$$11.32. \quad (x - \bar{x}_i)' L_0 L_0' (x - \bar{x}_i) = c^2(x) .$$

Als we nu 11.30, 11.31 en 11.32 combineren dan zien we dat 11.30 voor dezelfde waarde van  $i$  minimaal wordt als 11.29 waarmee de equivalentie van de beide criteria is aangetoond.

In de praktijk wordt vaak aanbevolen om in  $L$  slechts de eigenvectoren op te nemen die horen bij de  $q$  ( $< r$ ) grootste eigenwaarden en dan het criterium 11.29 toe te passen, hetgeen nu uiteraard niet meer hetzelfde resultaat oplevert als toepassing van 11.20.

## §12. Clusteranalyse

Met clusteranalyse worden technieken aangeduid die een groot aantal objecten (personen, planten, etc.) op grond van een groot aantal gegevens (d.m.v. enquêtes, tests, metingen, etc.) indelen in groepen die van te voren niet bekend zijn. Men gaat uit van de veronderstelling dat de verzameling objecten een steekproef is uit een onbekend aantal onbekende populaties, waartussen verschillen bestaan wat betreft de verzamelde gegevens.

Zijn van  $n$  objecten  $E_1, E_2, \dots, E_n$  een  $p$ -tal gegevens bekend, dan is het uitgangspunt voor de clusteranalyse een matrix  $X_{n \times p}$ . De meeste methoden voeren de analyse nu in twee fasen uit:

- 1) Voor ieder objectenpaar wordt de mate van overeenkomst tussen de gegevens berekend. Dit wordt uitgedrukt in een afstand of (dis)similarity-coëfficiënt. Het resultaat is een symmetrische matrix  $D_{n \times n}$ , waarin  $d_{ij}$  de afstand tussen de objecten  $E_i$  en  $E_j$  aangeeft.
- 2) Uitgaande van de matrix  $D$  worden de clusters  $C_1, \dots, C_k$  bepaald.

De meest gebruikte afstand is de Euclidische afstand zodat

$$12.1. \quad d_{ij}^2 = (x_{i*} - x_{j*})'(x_{i*} - x_{j*}) .$$

Bezwaren tegen dit afstandsbegrip zijn dat  $d$  afhankelijk is van de eenheden waarin de variabelen zijn gemeten en dat sterk afhankelijke variabelen als het ware dubbel tellen omdat verschillen bij die variabelen gemeenschappelijk optreden.

Beide punten vragen naar een tegen elkaar afwegen van de verschillen op de verschillende variabelen; het kiezen van een kleinere meeteenheid is hetzelfde als het toekennen van een groter gewicht aan de verschillen op die variabele. Er moeten dus eigenlijk gewichtsfactoren  $\lambda_1, \dots, \lambda_p$  worden gekozen waarna een gewogen Euclidische afstand wordt berekend:

$$12.2. \quad d_{ij}^2 = (x_{i*} - x_{j*})' \Lambda^2 (x_{i*} - x_{j*}) ,$$

waarin  $\Lambda^2$  een diagonale matrix is met  $\lambda_1^2, \dots, \lambda_p^2$  op de diagonaal. Voor de gewichtsfactoren worden wel genomen

$$\lambda_i^2 = \frac{1}{s_i} ,$$

waarin  $s_i^2$  de geschatte variantie is van de  $i^e$  variabele. Deze moet echter worden berekend uit de hele steekproef en bevat daarom een component tengevolge van de verschillen tussen de clusters. Om  $\sigma_i^2$  goed te kunnen schatten zouden we groepen objecten moeten kennen die uit één verdeling afkomstig zijn, maar deze indeling in groepen is juist het doel van de berekening!

Eenzelfde bezwaar kleeft aan de zogenaamde Mahalanobisafstand

$$12.3. \quad d_{ij}^2 = (x_{i*} - x_{j*})' \Sigma^{-2} (x_{i*} - x_{j*}) ,$$

bedoeld om voor de afhankelijkheid van de variabelen te corrigeren.

Opmerking. Deze zelfde maat kwamen we in het vorige hoofdstuk tegen, vgl. 11.18.

In de praktijk moet  $\Sigma^2$  weer uit de totale steekproef worden geschat en dan hebben we

$$12.4. \quad d_{ij}^2 = (x_{i*} - x_{j*})' S^{-2} (x_{i*} - x_{j*}) .$$

Een combinatie van 12.2 en 12.4 is voorgesteld door Morrison (1967). Hij beveelt aan

$$12.5. \quad d_{ij}^2 = (x_{i*} - x_{j*})' \Lambda S^{-2} (x_{i*} - x_{j*}) .$$

De gewichtsfactoren  $\lambda_i$  kunnen dan worden gekozen in verband met het belang dat de onderzoeker hecht aan de betreffende variabele. Samenvattend kunnen we zeggen dat meestal 12.1 zal worden toegepast of eventueel 12.2 waarbij dan bij de interpretatie met de weging rekening moet worden gehouden. Varianties  $s_i^2$  en variantie-covariantie matrices  $S^2$  zijn eigenlijk slechts dan bruikbaar als ieder object een representant is van een populatie, waarbij gegevens beschikbaar zijn over de variantie-covariantiestructuur binnen die populaties. Een voorbeeld hiervan is te vinden in hoofdstuk 9 van het boek van Rao (1952).

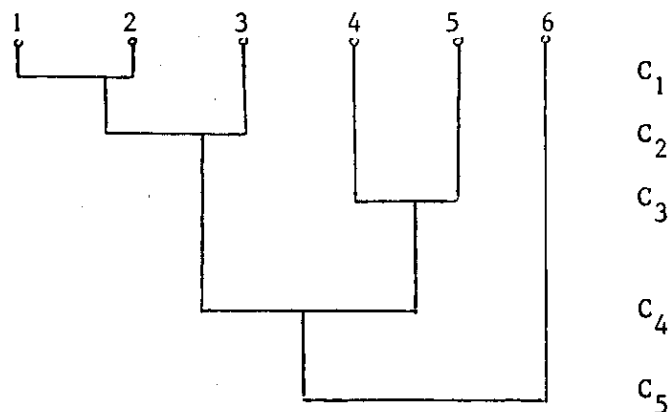
Er zijn vervolgens vele manieren om uit een afstandenmatrix D te komen tot een indeling in clusters.

Thorndike (1953) geeft een methode waarbij de  $n$  objecten over een van tevoren vastgesteld aantal ( $k$ ) clusters moeten worden verdeeld. Hij gaat er van uit dat de  $k$  punten met onderling de grootste afstanden tot verschillende clusters behoren, zodat deze als startpunten kunnen worden genomen. De overige punten worden vervolgens verdeeld door beurtelings aan ieder cluster dat, nog niet opgenomen, object toe te wijzen dat de kleinste gemiddelde afstand tot de al opgenomen objecten heeft. Het bezwaar van de arbitraire keuze van  $k$  wordt ondervangen door de clustering uit te voeren voor meerdere waarden van  $k$  en dan te kiezen voor die waarde van  $k$  waarbij de gemiddelde afstand binnen de clusters weinig groter is dan in het geval met  $k+1$  clusters.

Overblijvende bezwaren zijn:

- De methode resulteert in clusters van gelijke grootte, wat in het algemeen onjuist zal zijn.
- De meest uiteenliggende punten zullen randelementen zijn van de clusters waartoe ze horen. Het formeren van clusters rond deze punten lijkt dan ook niet juist.
- De methode is tijdrovend door de omslachtige manier waarop  $k$  wordt bepaald.

Veel methoden zijn hiërarchisch, d.w.z. dat er een criterium bestaat volgens welk in elke situatie waarin een clusterindeling gegeven is twee clusters tot één worden samengevoegd. Aan het begin worden alle objecten als één-puntsclusters opgevat, terwijl het proces eindigt met de situatie waarin alle objecten tot één cluster behoren. Het verloop van een dergelijk proces is in figuur 12.1 schematisch aangegeven.



Figuur 12.1



De knopen van de getekende boom geven het niveau aan waarop twee clusters worden samengevoegd. Dit niveau hangt direct samen met het gehanteerde criterium. Meestal is dit een functie die in iedere fase moet worden geoptimaliseerd. Achteraf wordt nu het aantal clusters bepaald op grond van de criterium-functie waarden. De boom wordt nl. doorgesneden op een punt waar deze functie een "sprong" maakt.

In het voorbeeld van fig. 12.1 is dit bijv. de sprong  $C_3 - C_4$ , zodat tot de clustering  $\{1,2,3\}$ ,  $\{4,5\}$  en  $\{6\}$  wordt besloten.

Een hierarchische methode is bijv. die van Ward en Hook (1963) waarbij als criterium wordt genomen de som van de gekwadraterde binnen-groep afwijkingen om het groepsgemiddelde, gesommeerd over alle clusters. Dus als

$$C = \sum_{\ell=1}^k \sum_{E_i \in C_\ell} (x_{i^*}^{(\ell)} - \bar{x}_i^{(\ell)})^2 (x_{i^*}^{(\ell)} - \bar{x}_i^{(\ell)})$$

dan wordt iedere samenvoeging zo gekozen dat  $\Delta C$  minimaal is. Het is eenvoudig aan te tonen dat  $\Delta C$  voor de achtereenvolgende samenvoegingen monotoon stijgend is wat de keuze van een afbreekpunt vergemakkelijkt.

Methoden als deze, die werken met kwadratische afwijkingen rond het groepsgemiddelde en dus goed aansluiten bij het gebruik van de Euclidische afstand, worden vaak aangeduid als minimum variantie methoden. Er moet op worden gewezen dat deze methoden sterk de nadruk leggen op interne samenhang en meestal resulteren in (hyper) bolvormige clusters. Wat meer langgerekte wolken worden in twee of meer delen gesplitst.

Een methode die dit bezwaar niet heeft is de nearest-neighbour of single linkage methode. Daarbij worden in iedere fase die clusters samengevoegd die de kleinste afstand hebben, waarbij de afstand tussen twee clusters  $C_\ell$  en  $C_m$  wordt gedefinieerd als

$$d(C_\ell, C_m) := \min_{E_i \in C_\ell, E_j \in C_m} d(E_i, E_j) .$$

Een variant op deze methode ontstaat door van te voren een getal  $S$  te kiezen en dan twee objecten in één cluster te verenigen als hun afstand  $\leq S$  is. De keuze van  $S$  komt overeen met een keuze van het afbreekpunt in de hierarchische methode. Voor ieder tweetal objecten  $E_i$  en  $E_j$  uit verschillende clusters geldt dus dat  $d(E_i, E_j) > S$  is, maar er geldt niet voor ieder tweetal uit eenzelfde cluster dat  $d(E_i, E_j) \leq S$  is. In tegenstelling tot de minimum variantie method

ligt hier de nadruk op externe scheiding. Het belangrijkste nadeel van de single-linkage methode is het "kettingeffect": door een paar "storingspunten" aan de rand worden twee verschillende clusters aan elkaar verbonden. De methode is verbeterd door Wishart (1969) en daarna weer door Praagman (1972).

Aanvullende literatuur

- J. Praagman (1972), Clusteranalyse. Een verbeterde versie van de single linkage methode. Afstudeerverslag Onderafdeling der Wiskunde, T.H. Eindhoven, juni 1972.
- C.R. Rao (1952), Advanced Statistical Methods in Biometric Research, Wiley 1952.
- R.L. Thorndike (1953), Who belongs in the family? Psychometrika 18, p.267-276.
- J.H. Ward and M.E. Hook (1963), Application of an hierarchical grouping procedure to a problem of grouping profiles. Educ. Psychol. Measur. 23, p.69-82.
- D. Wishart (1969), Mode analysis. In: Numerical Taxonomy (ed. A.J. Cole), p.282-311.

Appendix A: Normaal verdeelde, doch niet simultaan normaal verdeelde variabelen.

Stel  $x$  en  $y$  hebben als simultane kansverdeling (voorbeeld J. van IJzeren)

$$f(x,y) = \frac{1}{2\pi\sqrt{2}} (e^{-x^2-xy\sqrt{2}-y^2} + e^{-x^2+xy\sqrt{2}-y^2}), \quad -\infty < x,y < \infty.$$

- 1) Er geldt: zijn  $x$  en  $y$  *simultaan* normaal verdeeld, dan zijn  $x$  en  $y$  ook elk normaal verdeeld. Echter het omgekeerde geldt niet. Stel nl.  $x$  en  $y$  als boven verdeeld. Dan is

$$f_x(x) = \frac{1}{2\pi\sqrt{2}} e^{-\frac{1}{2}x^2} \left[ \int_{-\infty}^{\infty} e^{-(y+\frac{1}{2}x\sqrt{2})^2} dy + \int_{-\infty}^{\infty} e^{-(y-\frac{1}{2}x\sqrt{2})^2} dy \right] = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

Analoog  $f_y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$ . Dus  $x$ ,  $y$  elk normaal verdeeld, doch niet simultaan normaal.

- 2) Er geldt: zijn  $x$  en  $y$  *simultaan* normaal verdeeld met  $\rho(x,y) = 0$ , dan zijn  $x$  en  $y$  stochastisch onafhankelijk. Zijn  $x$  en  $y$  elk normaal verdeeld, doch niet simultaan normaal verdeeld, dan is dit niet juist; nl. neem  $x$  en  $y$  weer als boven verdeeld:

$$\rho(x,y) = \iint xy f(x,y) dx dy.$$

Nu is

$$\begin{aligned} \int y f(x,y) dy &= \frac{1}{2\pi\sqrt{2}} e^{-\frac{1}{2}x^2} \left[ \int_{-\infty}^{\infty} y e^{-(y+\frac{1}{2}x\sqrt{2})^2} dy + \int_{-\infty}^{\infty} y e^{-(y-\frac{1}{2}x\sqrt{2})^2} dy \right] = \\ &= \frac{1}{2\pi\sqrt{2}} e^{-\frac{1}{2}x^2} \left[ \int_{-\infty}^{\infty} (u - \frac{1}{2}x\sqrt{2}) e^{-u^2} du + \int_{-\infty}^{\infty} (v + \frac{1}{2}x\sqrt{2}) e^{-v^2} dv \right] = 0, \end{aligned}$$

dus  $\rho(x,y) = 0$ .

Toch zijn  $x$  en  $y$  stochastisch afhankelijk daar  $f(x,y) \neq f_x(x)f_y(y)$ .

3) Er geldt: de som van *simultaan* normaal verdeelde variabelen is normaal verdeeld. Dit geldt niet als we het woord "simultaan" weglaten. Stel weer  $\underline{x}$  en  $\underline{y}$  als boven verdeeld. Voer de volgende transformatie uit:  $\underline{x} = (\underline{u} + \underline{v})/\sqrt{2}$  en  $\underline{y} = (\underline{u} - \underline{v})/\sqrt{2}$

We zoeken de kansdichtheid van  $\underline{x} + \underline{y} = \underline{u}\sqrt{2}$ . Dan is (det. v. Jacobi = 1)

$$g(u,v) = \frac{1}{2\pi\sqrt{2}} \{ \exp[-u^2(1 + \frac{1}{2}\sqrt{2}) - v^2(1 - \frac{1}{2}\sqrt{2})] + \exp[-u^2(1 - \frac{1}{2}\sqrt{2}) - v^2(1 + \frac{1}{2}\sqrt{2})] \}$$

en dus

$$g_u(u) = \int g(u,v)dv = \frac{1}{2\sqrt{\pi}} \left[ \frac{1}{\sqrt{2 - \sqrt{2}}} e^{-u^2(1 + \frac{1}{2}\sqrt{2})} + \frac{1}{\sqrt{2 + \sqrt{2}}} e^{-u^2(1 - \frac{1}{2}\sqrt{2})} \right]$$

d.w.z.  $\underline{x} + \underline{y}$  niet normaal verdeeld!

Appendix B. Differentiatie van scalaire functies van de elementen van een matrix

a. Is  $y = f(x_1, \dots, x_n)$  een minstens tweemaal differentieerbare scalaire functie van  $x_1, \dots, x_n$ , dan is bekend uit de theorie van functies van meerdere variabelen:

$$dy = \sum_i \frac{\partial y}{\partial x_i} dx_i \quad \text{en} \quad d^2y = \sum_{i,j} \frac{\partial^2 y}{\partial x_i \partial x_j} dx_i dx_j .$$

Voldoende voor het optreden van een maximum in een punt  $P_0$  is dat  $(dy)_{P_0} = 0$  en  $(d^2y)_{P_0} < 0$ .

Zij nu  $y = y(X)$  een minstens tweemaal differentieerbare scalaire functie van de elementen van een matrix  $X_{n \times p}$ .

Definitie 1:  $\frac{dy}{dX} := \left\{ \frac{\partial y}{\partial x_{ji}} \right\}$  (dus een matrix van de orde  $p \times n$ ).  $\frac{\partial y}{\partial X'} = \left( \frac{\partial y}{\partial X} \right)'$ ,  
 $i = 1, \dots, n; j = 1, \dots, p$ .

Definitie 2:  $dX := \{dx_{ij}\}$ .

b. Eigenschappen:

B1.  $dy = \text{sp} \left( \frac{\partial y}{\partial X} dX \right)$ . Eveneens als  $y = y(X(t))$  :  $\frac{dy}{dt} = \text{sp} \left( \frac{\partial y}{\partial X} \frac{dX}{dt} \right)$ .

Bewijs:  $y = f(x_{11}, \dots, x_{np})$  dus

$$dy = \sum_{i,k} \frac{\partial y}{\partial x_{ik}} dx_{ik} = \sum_k \sum_i \left( \frac{\partial y}{\partial x} \right)_{ki} (dX)_{ik} = \sum_k \left( \frac{\partial y}{\partial X} dX \right)_{kk} = \text{sp} \left( \frac{\partial y}{\partial X} dX \right) .$$

Opmerking:  $dX' = (dX)'$ . Is  $X = x$  dan is  $dy = \frac{\partial y}{\partial x} dx$ ;  $y = a'x$  geeft  $\frac{\partial y}{\partial x} = a'$ . Ook geldt omgekeerd:  $dy = \text{sp}(AdX) \Rightarrow A = \frac{\partial y}{\partial X}$ . Analoog geldt:

B2.  $d^2y = d(dy) = d(\text{sp} \frac{\partial y}{\partial X} dX) = \text{sp}(d \frac{\partial y}{\partial X} dX) .$

Is  $X = x$  dan is  $d^2y = dx'Hdx$  met  $H := \left\{ \frac{\partial^2 y}{\partial x_i \partial x_j} \right\}$  de matrix van Hesse ( $H = H'$  als de afgeleiden continu zijn). Voldoende voor een maximum is  $\frac{\partial y}{\partial x} = 0$  en  $H$  negatief definitief.

B3.  $\frac{\partial |X|}{\partial X} = |X|X^{-1}$  ( $X$  regulier ondersteld) .

Bewijs:  $|X| = \sum_i x_{ij} |X|_{ij}$ , dus  $\frac{\partial |X|}{\partial x_{ij}} = |X|_{ij}$ . Nu is  $(X^{-1})_{ij} = |X|_{ji} / |X|$  dus  $|X|_{ji} = |X| (X^{-1})_{ij} = \left( \frac{\partial |X|}{\partial X} \right)_{ij}$ .

B4.  $z = z(y(X))$  geeft  $\frac{\partial z}{\partial X} = \frac{dz}{dy} \cdot \frac{\partial y}{\partial X}$ .

Een voorbeeld:

B5.  $\frac{\partial \log |X|}{\partial X} = \frac{1}{|X|} \overset{\text{zie B3}}{|X|X^{-1}} = X^{-1}$ . ( $|X| > 0$  ondersteld).

B6. Door uitschrijven ziet men eenvoudig in dat

$$d(X + Z) = dX + dZ ; \quad d(XZ) = (dX)Z + XdZ ;$$

$$dX' = (dX)' = dX \text{ indien } X \text{ symmetrisch; } \quad d \operatorname{sp} X = \operatorname{sp} dX .$$

B7.  $dX^{-1} = -X^{-1}(dX)X^{-1}$ .

Bewijs:

$$0 = dI = d(XX^{-1}) = (dX)X^{-1} + Xd(X^{-1}) .$$

B8. Stel  $y = x'Ax$  met  $A$  symmetrisch, dan is  $\frac{\partial y}{\partial x} = 2Ax$ , nl.

$$dy = dx'Ax + x'Adx = dx'Ax + dx'A'x = 2dx'Ax = \operatorname{sp}(2Axdx') .$$

c. Toepassingen in de multivariate analyse.

Zij  $\underline{x} \sim N_p(\mu, \Sigma^2)$  en  $X_{n \times p}$  een steekproef van  $\underline{x}$  ter grootte  $n$ .

Gevraagd maximum likelihoodschatters voor  $\mu$  en  $\Sigma^2$  ( $> 0$  ondersteld).

Oplossing: De likelihoodfunctie (simultane kansdichtheid van  $\underline{X}$ ) is volgens 3.1

$$L = |\Sigma^2|^{-\frac{1}{2}n} (2\pi)^{-\frac{1}{2}np} \exp\left[-\frac{1}{2} \sum_1^n (x_{i*} - \mu)' \Sigma^{-2} (x_{i*} - \mu)\right]$$

ofwel met  $A := \Sigma^{-2}$

$$y = \log L = \text{const} + \frac{1}{2}n \log |A| - \frac{1}{2} \sum_1^n (x_{i*} - \mu)' A (x_{i*} - \mu) .$$

(1) schatteer voor  $\mu$ :

$$\frac{\partial \log L}{\partial \mu'} = + \frac{1}{2} \sum_1^n 2A(x_{i*} - \mu) = A \sum_1^n (x_{i*} - \mu) = 0 \quad (\text{zie B8})$$

en daar  $A > 0$ , is

$$\hat{\underline{\mu}} = \sum_1^n \underline{x}_{i*} / n =: \bar{\underline{x}} .$$

Dit geeft inderdaad een maximum, nl.

$$dy = \text{sp}\{(A \Sigma \underline{x}_{i*} - nA\mu) d\mu'\}$$

$$d^2y = \text{sp}(-nA d\mu d\mu') = -n \text{sp}(d\mu' A d\mu) < 0 .$$

(2) schatter voor  $\Sigma^2$ :

Nu wordt

$$\begin{aligned} y &= c + \frac{1}{2} n \log |A| - \frac{1}{2} \sum_1^n (\underline{x}_{i*} - \bar{\underline{x}})' A (\underline{x}_{i*} - \bar{\underline{x}}) = \\ &= c + \frac{1}{2} n \log |A| - \frac{1}{2} \text{sp} \sum (\underline{x}_{i*} - \bar{\underline{x}}) (\underline{x}_{i*} - \bar{\underline{x}})' A \end{aligned}$$

oftewel met

$$S^2 := \sum_1^n (\underline{x}_{i*} - \bar{\underline{x}}) (\underline{x}_{i*} - \bar{\underline{x}})' / n$$

wordt dit

B9.  $y = \log L = c + \frac{1}{2} n \log |A| - \frac{1}{2} n \text{sp}(S^2 A) .$

Dit geeft:  $dy = \text{sp}(\frac{\partial y}{\partial A} dA)$  oftewel

B10.  $dy = \frac{1}{2} n \text{sp}(A^{-1} dA - S^2 dA) = \frac{1}{2} n \text{sp}[(A^{-1} - S^2) dA]$

(A symmetrisch; B1, B5, B6).

B11.  $d^2y = -\frac{1}{2} n \text{sp}(A^{-1} dA A^{-1} dA) .$

Het maximum van L (van logL) wordt bereikt voor  $\hat{A}^{-1} = \underline{S}^2 = \hat{\Sigma}^2$ . *p x p-matrix*

Dit geeft inderdaad een maximum, daar  $d^2y < 0$  volgens 1.68 en 1.69.

Het maximum wordt dan:

B12.  $L_{\max} = |\hat{\Sigma}^2|^{-\frac{1}{2}n} (2\pi)^{-\frac{1}{2}np} \exp(-\frac{1}{2}np) .$

Samengevat: de *maximum likelihoodschatters* voor  $\mu$  en  $\Sigma^2$  zijn resp.:

$$B13. \quad \hat{\underline{\mu}} = \bar{\underline{x}} := \sum_1^n \underline{x}_{i*} / n \quad \text{en} \quad \hat{\underline{\Sigma}}^2 = \underline{S}^2 := \sum_1^n (\underline{x}_{i*} - \bar{\underline{x}})(\underline{x}_{i*} - \bar{\underline{x}})' / n.$$

In de faktoranalyse moet dezelfde functie  $y$  uit B9 gemaximaliseerd worden, doch nu naar  $F$  en  $V^2$ . Daar is nl.  $A^{-1} := \Sigma^2(\underline{x}) = FF' + V^2$ ;  $V^2$  is een  $p \times p$  diagonaalmatrix,  $F$  is een  $p \times k$  matrix.

Nu is  $dA = d\Sigma^{-2} = -\Sigma^{-2}(d\Sigma^2)\Sigma^{-2}$  (zie B7).

En  $d\Sigma^2 = d(FF') + dV^2 = (dF)F' + F dF' + dV^2$ .

B10 wordt nu:

$$\begin{aligned} dy &= \frac{1}{2}n \operatorname{sp} [(\Sigma^2 - S^2)(-\Sigma^{-2}(d\Sigma^2)\Sigma^{-2})] = \\ &= \frac{1}{2}n \operatorname{sp} [(\Sigma^{-2}S^2\Sigma^{-2} - \Sigma^{-2})d\Sigma^2] = \\ &= \frac{1}{2}n \operatorname{sp} [(\Sigma^{-2}S^2\Sigma^{-2} - \Sigma^{-2})d(FF') + (\Sigma^{-2}S^2\Sigma^{-2} - \Sigma^{-2})dV^2] \end{aligned}$$

Nu is

$$\begin{aligned} &\operatorname{sp} [(\Sigma^{-2}S^2\Sigma^{-2} - \Sigma^{-2})((dF)F' + F dF')] = \\ &= \operatorname{sp} [(F'\Sigma^{-2}S^2\Sigma^{-2} - F'\Sigma^{-2})dF + dF'(\Sigma^{-2}S^2\Sigma^{-2}F - \Sigma^{-2}F)] = \\ &= \operatorname{sp} [2(F'\Sigma^{-2}S^2\Sigma^{-2} - F'\Sigma^{-2})dF] \end{aligned}$$

ofwel

$$dy = \frac{1}{2}n \operatorname{sp} [2(F'\Sigma^{-2}S^2\Sigma^{-2} - F'\Sigma^{-2})dF + (\Sigma^{-2}S^2\Sigma^{-2} - \Sigma^{-2})dV^2]$$

Dus maximaal voor  $\hat{F}'\hat{\Sigma}^{-2}\hat{S}^2\hat{\Sigma}^{-2} - \hat{F}'\hat{\Sigma}^{-2} = 0$ , ofwel

$$B14. \quad \hat{F}'\hat{\Sigma}^{-2}\hat{S}^2 = \hat{F}$$

en

$$B15. \quad (\hat{\Sigma}^{-2}\hat{S}^2\hat{\Sigma}^{-2} - \hat{\Sigma}^{-2})_{ii} = 0 \quad (i=1, \dots, p)$$

alleen voor diagonaalelementen, omdat  $V^2$  een diagonaalmatrix is.



Appendix C: Overzicht van het verband tussen één- en meerdimensionale analyse

a. $\underline{x}, \mu, \mu_x, \bar{x}, x_i, \chi_1$	$x', \mu', \mu'_x, \bar{x}', x'_{i*}, \chi'_p$
b. $\chi_v, \underline{t}_v, N_1, W_1, \alpha_1, \beta_1, v_1, \chi_{v_1}$	$\chi_{v,p}, \underline{t}_{v,p}, N_p, W_p, \alpha_p, \beta_p, v_p, \chi_{v_p}$
c. $x_*, \mu_*, s, \rho, \sigma_w, \text{var}, \text{cov}, a, \delta$	$X, M, S, R, \Sigma( \Sigma ), W( W ), \text{VAR}, \text{COV}, A, D$
d. $\exp, 2$	$\text{exptr}, 2$ (if necessary)
e. $2, t, n, v, y, y_i, \mu_y, f(x), \varphi_x(t), P, P^\perp$	invariant

A One-dimensional  $\underline{x}$

1  $\mu := \mathcal{E} \underline{x}$

2  $\sigma^2 := \mathcal{E} (\underline{x} - \mu)^2 = \mathcal{E} \underline{x}^2 - \mu^2$   
 $\sigma^2 = \sigma^2(\underline{x}) = \text{var } \underline{x} \geq 0$   
 $\mathcal{E} (a\underline{x}^2) = a(\sigma^2 + \mu^2)$

3  $\text{cov}(\underline{y}, \underline{x}) := \mathcal{E} [(\underline{y} - \mu_y)(\underline{x} - \mu_x)]$

4  $\underline{x} = \mu + \chi_1 \sigma \sim N_1(\mu, \sigma^2) \quad \sigma > 0$   
 $\chi_1 \sim N_1(0, 1)$

5  $f(x) = \frac{1}{\sigma(2\pi)^{\frac{1}{2}\alpha_1}} \exp[-\frac{1}{2}(x-\mu)^2 \sigma^{-2}]$

6  $\varphi_{\underline{x}}(t) = \exp[i\mu t - \frac{1}{2}(\sigma t)^2]$

7  $\chi_v = \begin{pmatrix} (\chi_1)_1 \\ \vdots \\ (\chi_1)_v \end{pmatrix}$  mutually independent

8  $(\chi_v \sigma)^2 \sim : W_1(v, \sigma^2) \quad v \geq 1$   
 $\chi_v^2 \sim : W_1(v, 1)$   
 $\mathcal{E} (\chi_v \sigma)^2 = v\sigma^2 ; \mathcal{E} \chi_v^2 = v$

B p-dimensional  $\underline{x}' := (x_1, \dots, x_p)$

$\mu' := \mathcal{E} \underline{x}' = (\mu_1, \dots, \mu_p)$

$\Sigma^2 := \mathcal{E} (\underline{x}' - \mu')^2 = \mathcal{E} (\underline{x}')^2 - (\mu')^2$   
 $\Sigma^2 = \Sigma^2(\underline{x}) = \text{VAR } \underline{x} \geq 0$   
 $\mathcal{E} [A(\underline{x}')^2] = A [\Sigma^2 + (\mu')^2]$

$\text{COV}(\underline{y}, \underline{x}) := \mathcal{E} [(\underline{y} - \mu_y)(\underline{x}' - \mu'_x)]$

$\underline{x}' = \mu' + \chi'_p \Sigma \sim N_p(\mu', \Sigma^2) \quad |\Sigma| > 0$   
 $\chi'_p \sim N_p(0', I)$

$f(x) = \frac{1}{|\Sigma| (2\pi)^{\frac{1}{2}\alpha_p}} \text{exptr}[-\frac{1}{2}(\underline{x}' - \mu')^2 \Sigma^{-2}]$

$\varphi_{\underline{x}'}(t) = \exp[i\mu' t - \frac{1}{2}(\Sigma t)^2]$

$\chi_{v,p} := \begin{pmatrix} (\chi'_p)_1 \\ \vdots \\ (\chi'_p)_v \end{pmatrix}$  mutually independent

$(\chi_{v,p} \Sigma)^2 \sim : W_p(v, \Sigma^2) \quad v \geq p$   
 $\chi_{v,p}^2 \sim : W_p(v, I)$   
 $\mathcal{E} (\chi_{v,p} \Sigma)^2 = v\Sigma^2 ; \mathcal{E} \chi_{v,p}^2 = v I$

$$9 \quad \underline{\chi}_{-v}^2 = \sum_1^v (\underline{\chi}_{-1})_i^2$$

$$10 \quad (P\underline{\chi}_{-v, \sigma})^2 \approx (\underline{\chi}_{-q, \sigma})^2 \quad r(P) = q$$

$$11 \quad \underline{t}_{-v}^2(\mu^2) := v \|\mu + \underline{\chi}_{-1}\|^2 / \|\underline{\chi}_{-1}\|^2$$

(indep.)

$$\underline{t}_{-v}^2 := v \|\underline{\chi}_{-1}\|^2 / \|\underline{\chi}_{-1}\|^2 \quad (\approx F_{-v, 1}^1)$$

$$12 \quad \underline{\chi}_{-v}^2 = \underline{W} \sim \beta_1 \underline{W}^{\frac{1}{2}v-1} \exp[-\frac{1}{2}\underline{W}]$$

$$(\underline{\chi}_{-v, \sigma})^2 = \underline{W} \sim \beta_1 \frac{\underline{W}^{\frac{1}{2}v-1}}{\sigma^v} \exp[-\frac{1}{2}\underline{W}\sigma^{-2}]$$

$$\underline{\chi}_{-v, p}^2 = \sum_1^v (\underline{\chi}'_{-p})_i^2$$

$$(P\underline{\chi}_{-v, p, \Sigma})^2 \approx (\underline{\chi}_{-q, p, \Sigma})^2$$

$$\underline{t}_{-v, p}^2(\mu^2) := v \|\mu + \underline{\chi}_{-p}\|^2 / \|\underline{\chi}_{-p}\|^2$$

(indep.)

$$\underline{t}_{-v, p}^2 := v \|\underline{\chi}_{-p}\|^2 / \|\underline{\chi}_{-p}\|^2 \quad (\approx \frac{vP}{v_p} F_{-v_p}^P)$$

$$\underline{\chi}_{-v, p}^2 = \underline{W} \sim \beta_p |\underline{W}|^{\frac{1}{2}v_p-1} \exp\text{tr}[-\frac{1}{2}\underline{W}]$$

$$(\underline{\chi}_{-v, p, \Sigma})^2 = \underline{W} \sim \beta_p \frac{|\underline{W}|^{\frac{1}{2}v_p-1}}{|\Sigma|^{v_p}} \exp\text{tr}[-\frac{1}{2}\underline{W}\Sigma^{-2}]$$

Sample of order n

$$13 \quad \underline{x}_* := \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}; \quad \underline{\mu}_* := \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix}; \quad \bar{\underline{x}}_* := \begin{pmatrix} \bar{x} \\ \vdots \\ \bar{x} \end{pmatrix}$$

$$\text{with } \bar{\underline{x}} := \sum_1^n x_i / n$$

$$14 \quad \underline{x}_* = \underline{\mu}_* + \underline{\chi}_{-n, \sigma}$$

$$15 \quad \bar{\underline{x}} \sim N_1(\mu, \sigma^2/n)$$

$$16 \quad \bar{\underline{x}}_* = P\underline{x}_* \quad P = (1/n)$$

$$(\underline{x}_* - \underline{\mu}_*)^2 = (\underline{x}_* - \bar{\underline{x}}_*)^2 + (\bar{\underline{x}}_* - \underline{\mu}_*)^2$$

$$17 \quad (\underline{\mu}_* + \underline{\chi}_{-n, \sigma})^2 \sim W_1(n, \sigma^2, \delta^2)$$

$$18 \quad v_s^2 := \sum_1^n (x_i - \bar{x})^2 = (\underline{x}_* - \bar{\underline{x}}_*)^2$$

$$= \sum_1^n x_i^2 - n\bar{x}^2 = \underline{x}_*^2 - \bar{\underline{x}}_*^2$$

$$\underline{X} := \begin{pmatrix} x'_{1*} \\ \vdots \\ x'_{n*} \end{pmatrix}; \quad M := \begin{pmatrix} \mu' \\ \vdots \\ \mu' \end{pmatrix}; \quad \bar{\underline{X}} := \begin{pmatrix} \bar{x}' \\ \vdots \\ \bar{x}' \end{pmatrix}$$

$$\text{with } \bar{\underline{x}}' := \sum_1^n x'_{i*} / n = (\bar{x}'_1, \dots, \bar{x}'_p)$$

$$\underline{X} = M + \underline{\chi}_{-n, p, \Sigma}$$

$$\bar{\underline{X}}' \sim N_p(\mu', \Sigma^2/n)$$

$$\bar{\underline{X}} = P\underline{X}$$

$$(\underline{X} - M)^2 = (\underline{X} - \bar{\underline{X}})^2 + (\bar{\underline{X}} - M)^2$$

$$(M + \underline{\chi}_{-n, p, \Sigma})^2 \sim W_p(n, \Sigma^2, D^2)$$

$$v_S^2 := \sum_1^n (x'_{i*} - \bar{x}')^2 = (\underline{X} - \bar{\underline{X}})^2$$

$$= \sum_1^n (x'_{i*})^2 - n(\bar{x}')^2 = \underline{X}^2 - \bar{\underline{X}}^2$$

$$19 \quad \begin{aligned} v \underline{s}^2 &= (P^\perp \underline{x}_*)^2 \\ &= (\underline{X}_v \sigma)^2 \sim W_1(v, \sigma^2) \\ \xi_{\underline{s}^2} &= \sigma^2 \end{aligned}$$

$$\left| \begin{aligned} v \underline{S}^2 &= (P^\perp \underline{X})^2 \\ &= (\underline{X}_{v,p} \Sigma)^2 \sim W_p(v, \Sigma^2) \\ \xi_{\underline{S}^2} &= \Sigma^2 \end{aligned} \right.$$

Test-statistics

$$20 \quad \begin{aligned} \|(\bar{\underline{x}} - \mu_0)(\sigma/\sqrt{n})^{-1}\|^2 &\approx \|\underline{X}_1\|^2 \\ \|(\bar{\underline{x}} - \mu_0)(\underline{s}/\sqrt{n})^{-1}\|^2 &\approx \underline{t}_v^2 \end{aligned}$$

$$\left| \begin{aligned} \|(\bar{\underline{x}}' - \mu_0')(\Sigma/\sqrt{n})^{-1}\|^2 &\approx \|\underline{X}_p\|^2 \\ \|(\bar{\underline{x}}' - \mu_0')(\underline{S}/\sqrt{n})^{-1}\|^2 &\approx \underline{t}_{v,p}^2 \end{aligned} \right.$$

$$21 \quad \frac{v \hat{\rho}^2}{1 - \hat{\rho}^2} \approx \underline{t}_v^2$$

$$\left| \frac{v \hat{R}^2}{1 - \hat{R}^2} \approx \underline{t}_{v,p}^2 \right.$$

$$22 \quad \text{var}(\underline{y}_i | \underline{x}) = (1 - \rho^2) \text{var} \underline{y}_i$$

$$\left| \text{var}(\underline{y}_i | \underline{x}') = (1 - R^2) \text{var} \underline{y}_i \right.$$

$$23 \quad \text{Ef}(\underline{x}) = \log \sigma + \frac{1}{2} \alpha_1 \log(2\pi e)$$

$$\text{Ef}(\underline{x}) = \log |\Sigma| + \frac{1}{2} \alpha_p \log(2\pi e) \quad (\text{Entropy}).$$

Opmerking:  $\alpha_p = p$  dus  $\alpha_1 = 1$  ;  $v_p = v - p + 1$  dus  $v_1 = v$

$$P^\perp = I - P.$$

Appendix E: Het Kronecker- of directproduct van matrices

a. Definitie:  $A \otimes B := \{a_{ij} \cdot B\}$ . Is  $A_{p \times q}$ ,  $B_{r \times s}$  dan is  $A \otimes B$  een matrix van de orde  $pr \times qs$ .

Elk element  $a_{ij}$  van  $A$  wordt met de gehele  $B$  vermenigvuldigd.

Opmerking: In het algemeen is  $A \otimes B \neq B \otimes A$ .

$A \otimes I = \{a_{ij} \cdot I\}$  maar

$$I \otimes A = \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix}.$$

$$I_p \otimes I_q = I_q \otimes I_p = I_{pq}.$$

Voor vectoren geldt:  $a' \otimes b = b \otimes a' = ba'$ .

b. Eigenschappen:

E1. Partitioneren:  $(A_1 A_2 A_3) \otimes B = (A_1 \otimes B, A_2 \otimes B, A_3 \otimes B)$ .

E2.  $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$  indien de producten  $AC$  en  $BD$  bestaan, nl.:  
 $(A \otimes B)(C \otimes D) = \{a_{ik} \cdot B\} \{c_{kj} \cdot D\} = \{a_{i*} \cdot c_{*j} \cdot BD\} = (AC) \otimes (BD)$ .

Gevolg 2a:  $A, B$  idempotent  $\Rightarrow A \otimes B$  idempotent.

E3.  $A \otimes (B + C) = A \otimes B + A \otimes C$ . Deze volgen direct uit de definitie.

$$(B + C) \otimes A = B \otimes A + C \otimes A.$$

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

E4.  $(A \otimes B)' = A' \otimes B'$ , nl.  $\{a_{ij} \cdot B\}' = \{a_{ji} \cdot B'\} = A' \otimes B'$ .

Gevolg 4a:  $A, B$  symmetrisch  $\Leftrightarrow A \otimes B$  symmetrisch.

4b:  $A, B$  orthogonale projecties  $\Rightarrow A \otimes B$  orthogonale projectie (met 2a en 4a).

4c:  $A, B$  orthogonaal  $\Rightarrow A \otimes B$  orthogonaal.

- E5. Zij  $A_{p \times p}$  met eigenwaarden  $\lambda_i$  en eigenvectoren  $x_i$ ,  $i = 1, \dots, p$ .  
 En  $B_{q \times q}$  met eigenwaarden  $\mu_j$  en eigenvectoren  $y_j$ ,  $j = 1, \dots, q$ .

Dan heeft  $A \otimes B$  als eigenwaarden  $\lambda_i \mu_j$  met eigenvectoren  $x_i \otimes y_j$ , nl.:

$$(A \otimes B)(x_i \otimes y_j) = Ax_i \otimes By_j = \lambda_i \mu_j (x_i \otimes y_j) .$$

E6.  $|A_{p \times p} \otimes B_{q \times q}| = \prod_{ij} \lambda_i \mu_j = \prod_i \lambda_i^q \prod_j \mu_j^p = |A|^q |B|^p$  (pas op, p en q wisselen!).

E7.  $\text{sp}(A_{p \times p} \otimes B_{q \times q}) = \sum_{ij} \lambda_i \mu_j = \sum_i \lambda_i \sum_j \mu_j = \text{sp } A \text{ sp } B$ .

E8.  $r(A \otimes B) = r(A)r(B)$ .

Bewijs: De grootste reguliere submatrix van A, stel  $\bar{A}$ , heeft de afmeting  $r(A)$ .

Analoog van B, stel  $\bar{B}$ , de afmeting  $r(B)$ . Nu is  $\bar{A} \otimes \bar{B}$  submatrix van  $A \otimes B$  en deze is regulier daar  $|\bar{A} \otimes \bar{B}| = |\bar{A}|^{r(B)} |\bar{B}|^{r(A)} \neq 0$ .

Elke grotere submatrix van  $A \otimes B$  is singulier, nl. stel deze  $A_{kk}^*$  en  $B_{nn}^*$  dus  $|A^* \otimes B^*| = |A^*|^n |B^*|^k = 0$  voor  $k > r(B)$  of  $n > r(A)$ . De afmeting van  $\bar{A} \otimes \bar{B}$  is juist  $r(A)r(B)$ .

Gevolg 8a:  $A \otimes B$  regulier  $\iff A, B$  regulier (en dus vierkant!).

Bewijs  $\Leftarrow$ : Triviaal.

Bewijs  $\Rightarrow$ : Stel  $A_{pq} \otimes B_{rs}$  regulier. Dan is dus  $pr = qs = r(A)r(B)$ .

Stel  $r(A) = p$  ( $< p$  uitgesloten anders  $r(B) > r$ ). Dus

$r(B) = r$ . Als  $q > p$  of  $s > r$  is  $pr \neq qs$ , dus  $p = q$  en

$r = s$ .

8b:  $A, B > 0 \implies A \otimes B > 0$  (met E4a, E5 en 1.63).

- E9.  $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ . Uit 8a volgt dat A en B regulier moeten zijn. Eveneens geldt voor een gegeneraliseerde inverse  $(A \otimes B)^{-} = A^{-} \otimes B^{-}$ .

c. De vec-notatie

Definitie:  $\vec{A} := (a'_{*1} \dots a'_{*q})'$ , d.w.z. de kolommen van A achtereenvolgens onder elkaar genoteerd. Is A van de orde  $p \times q$ , dan is  $\vec{A}$  (ook wel vec A genoteerd) een kolomvector met pq componenten. In het bijzonder is  $\vec{a} = \vec{a}' = a$ .

E10. Stelling:  $\overrightarrow{A+B} = \vec{A} + \vec{B}$ . Triviaal (ga na).

E11. Stelling:  $\overrightarrow{ABC} = (C' \otimes A)\vec{B}$ .

Bewijs:  $(ABC)_{*j} = (Ab_{*1} \dots Ab_{*s})C_{*j} = \sum_k c_{kj} Ab_{*k}$ . En

$$(C' \otimes A)\vec{B} = \{c_{kj}A\} \begin{pmatrix} b_{*1} \\ \vdots \\ b_{*s} \end{pmatrix} = \begin{pmatrix} \vdots \\ \sum_k c_{kj} Ab_{*k} \\ \vdots \end{pmatrix} \quad \text{q.e.d.}$$

Gevolg 11a:  $Y = AX$  geeft  $\vec{Y} = (I \otimes A)\vec{X}$  en  
 $Y = XA$  geeft  $\vec{Y} = (A' \otimes I)\vec{X}$ .

E12. Stelling:  $\text{sp}(A'B) = \sum_i a'_{*i} b_{*i} = (\vec{A}, \vec{B})$  zodat  $\|A\| = \|\vec{A}\|$  en  $(A, B) = (\vec{A}, \vec{B})$  (zie 1.8 en 1.9).

d. Enkele toepassingen

Definitie:  $\text{VAR } \underline{X} := \text{VAR } \vec{X}$  (pas op:  $\text{VAR } \underline{X} \neq \text{VAR } \underline{X}'$ ).

E13. Stelling:  $\underline{Y} = \underline{AX}$  geeft:  $\text{VAR } \underline{Y} = (I \otimes A)\text{VAR } \underline{X}(I \otimes A')$   
 $\underline{Y} = \underline{XA}$  geeft:  $\text{VAR } \underline{Y} = (A' \otimes I)\text{VAR } \underline{X}(A \otimes I)$ , (met 2.37 en E11a).

E14. Voorbeeld 1: Zij  $\underline{y} = \underline{XB} + \underline{e}$  met  $\underline{e} = \chi_n \sigma$  dus  $\underline{E} = 0$  en  $\text{var } \underline{e} = \sigma^2 I_n$ . We weten nu:

$$\underline{b} = (X'X)^{-1} X' \underline{y} \text{ en } \text{var } \underline{b} = \sigma^2 (X'X)^{-1}.$$

Beschouwen we nu simultaan q vergelijkingen, dan ontstaat:

$$\underline{Y} = \underline{XB} + \underline{E} \text{ met } \underline{E} \simeq \chi_{n,q} \Sigma_{qq} \text{ dus } \underline{E} = 0.$$

Nu blijkt:

$$\hat{\underline{B}} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y} \text{ met } \text{VAR } \hat{\underline{B}} = \Sigma^2 \otimes (\underline{X}'\underline{X})^{-1} \text{ en } \text{VAR } \underline{E} = \Sigma^2 \otimes \underline{I}_n .$$

Bewijs:

i)  $\text{VAR } \underline{E} = (\Sigma' \otimes \underline{I}_n)(\underline{I}_q \otimes \underline{I}_n)(\Sigma \otimes \underline{I}_n) = \Sigma^2 \otimes \underline{I}_n$  (met E2, E13).

ii)  $\underline{\hat{Y}} = (\underline{I} \otimes \underline{X})\underline{\hat{B}} + \underline{\hat{E}}$  dus  $\underline{\hat{B}} = [(\underline{I} \otimes \underline{X}')(\underline{I} \otimes \underline{X})]^{-1}(\underline{I} \otimes \underline{X}')\underline{\hat{Y}} =$   
 $= [\underline{I} \otimes (\underline{X}'\underline{X})^{-1}][\underline{I} \otimes \underline{X}']\underline{\hat{Y}} = [\underline{I} \otimes (\underline{X}'\underline{X})^{-1}\underline{X}']\underline{\hat{Y}}$  oftewel met E11  
 $\underline{\hat{B}} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}.$

iii)  $\text{VAR } \underline{\hat{B}} = [\underline{I} \otimes (\underline{X}'\underline{X})^{-1}\underline{X}'][\Sigma^2 \otimes \underline{I}_n][\underline{I} \otimes \underline{X}(\underline{X}'\underline{X})^{-1}] = \Sigma^2 \otimes (\underline{X}'\underline{X})^{-1}.$

Volgens E8b is  $\Sigma^2 \otimes (\underline{X}'\underline{X})^{-1} > 0.$

E15. Voorbeeld 2: Beschouw het volgende variantie-analyse model:

$$y_{ij} = \mu + \alpha_i + \gamma_j + e_{ij}, \quad i = 1, \dots, p, \quad j = 1, \dots, q .$$

Dit is te schrijven als:

$$\underline{Y} = \mu \underline{U} + \alpha \underline{u}'_q + \underline{u}_p \gamma' + \underline{E} .$$

Dus

$$\underline{\hat{Y}} = \mu \underline{u}_{pq} + (\underline{u}_q \otimes \underline{I}_p)\alpha + (\underline{I}_q \otimes \underline{u}_p)\gamma + \underline{\hat{E}}$$

oftewel  $\underline{\hat{Y}} = \underline{X}\beta + \underline{\hat{E}}$  met  $\beta := (\mu, \alpha', \gamma')'$  en  $\underline{X} := (\underline{u}_{pq} \quad \underline{u}_q \otimes \underline{I}_p \quad \underline{I}_q \otimes \underline{u}_p).$

E16. Voorbeeld 3: Zij  $\underline{A}\underline{x} = \underline{y}$  consistent, dan worden alle oplossingen

$$\underline{x}^0 = \underline{A}^- \underline{y} + (\underline{I} - \underline{A}^- \underline{A})\underline{z} .$$
 Beschouw nu simultaan het consistente stelsel  $\underline{A}\underline{X} = \underline{Y}.$

Wat zijn nu alle oplossingen?

$$(\underline{I} \otimes \underline{A})\underline{\hat{X}} = \underline{\hat{Y}} \text{ geeft } \underline{\hat{X}}^0 = (\underline{I} \otimes \underline{A}^-)\underline{\hat{Y}} + [\underline{I} - (\underline{I} \otimes \underline{A}^-)(\underline{I} \otimes \underline{A})]\underline{\hat{Z}} =$$

$$= (\underline{I} \otimes \underline{A}^-)\underline{\hat{Y}} + [\underline{I} - (\underline{I} \otimes \underline{A}^- \underline{A})]\underline{\hat{Z}} ,$$

oftewel

$$\underline{\hat{X}}^0 = \underline{A}^- \underline{Y} + (\underline{I} - \underline{A}^- \underline{A})\underline{Z} .$$

Opmerking: Dit kan natuurlijk ook direct, daar  $\underline{A}\underline{X} = \underline{Y}$  betekent  $\underline{A}_{*j} x_{*j} = y_{*j}$  voor alle  $j!$