Verdeling en onafhankelijkheid van kwadratensommen in de variantie-analyse

EINDHOVEN UNIVERSITY OF TECHNOLOGY
Department of Mathematics and Computing Science

Memorandum COSOR 90-29

Verdeling en onafhankelijkheid van kwadratensommen in de variantie-analyse

R. Doornbos

Eindhoven University of Technology
Department of Mathematics and Computing Science
P.O. Box 513
5600 MB Eindhoven
The Netherlands

Eindhoven, August 1990
The Netherlands

Verdeling en onafhankelijkheid van kwadratensommen in de variantie-analyse.

R. Doornbos

0. Inleiding

In de meeste boeken over variantie-analyse en/of lineaire modellen vinden we één of meer inleidende hoofdstukken over matrixrekening en over de verdeling van kwadratische vormen. Dit wordt dan later bij de variantie-analyse alleen toegepast bij modellen met factoren met vast nivo's. In deze notitie laten we zien hoe stochastische en gemengde modellen de voor de F-toets nodige eigenschappen eveneens rechtstreeks volgen uit de bekende stellingen over kwadratische vormen. Ook het geneste model kan op de zelfde wijze worden behandeld. Wij beperken ons, om de notatie eenvoudig te houden, tot het geval van 2 factoren met herhalingen. We zullen voor de gebruikte stellingen en notaties verwijzen naar de dictaten Lineaire Modellen (dictaatnr. 2.221, najaarssemester 1981) en Kansrekening en Statistiek 2, bestemd voor Bdk (2S750), herfsttrimester 1986), die we resp. met LM en K&S zullen aangeven.

1. Enkele stellingen

Stelling 1 $\underline{y} \sim N_p(\mu,\ V)$, V regulier

$$\underline{y}'A\underline{y} \sim \chi^2_r(\mu' A \mu) \ \Leftrightarrow\ AV \text{ idempotent en } r = r(A)\ .$$

(L.M. Stelling 4.11).

Speciaal geval:
Stelling 1a $\underline{y} \sim N_p(\mu,\ \sigma^2 I)$
P orthogonale projector, $r(P) = r$, dan $\dfrac{\underline{y}'P\underline{y}}{\sigma^2} \sim \chi^2_\mu(\delta),\ \delta = \dfrac{\mu' P \mu}{\sigma^2}$.

Stelling 2 $\underline{y} \sim N_p(\mu,\ V)$

$$\underline{y}'A\underline{y} \text{ en } \underline{y}'B\underline{y} \text{ onafhankelijk } \Leftrightarrow AVB = 0\ .$$

(L.M. Stelling 4.13*).

Speciaal geval:
Stelling 2a $\underline{y} \sim N_p(\mu,\ \sigma^2 I)$

$$\underline{y}'A\underline{y} \text{ en } \underline{y}'B\underline{y} \text{ onafhankelijk } \Leftrightarrow AB = 0\ .$$

In het bijzonder: $A = P_1$, $B = P_2$, orthogonale projectoren op onderling loodrechte deelruimten.

Stelling 3

$$\mathcal{E}\underline{y}A\underline{y} = sp(AV) + \mu' A\mu$$

(L.M. Eigenschap 2.36).
De toepassing van deze stellingen is zeer eenvoudig, zodra bij de verschillende modellen de variantie-covariantiematrix V is bepaald.

2. De tweevoudige variantie-analyse

We beschouwen 4 gevallen:

a. Vaste nivo's

b. Stochastische nivo's

c. Gemengd model

d. Geneste proefopzet.

In alle gevallen geldt de splitsing van kwadratensommen (vgl. K&S)

$$SS_T = SS_A + SS_B + SS_{AB} + SS_R \quad ,$$

waarbij in het geneste geval, met $B : A$ geldt:

$$SS_{B:A} = SS_B + SS_{AB} \quad .$$

De corresponderende splitsing van de vrijheidsgraden is

$$abn - 1 = (a-1) + (b-1) + (a-1)(b-1) + ab(n-1)$$

en dit komt meer overeen met de notatie in bruto kwadratensommen:

$$\begin{aligned} S_{ABC} - S_0 \quad &= (S_A - S_0) + (S_B - S_0) + (S_{AB} - S_A - S_B + S_0) \\ &+ (S_{ABC} - S_{AB}) \quad . \end{aligned}$$

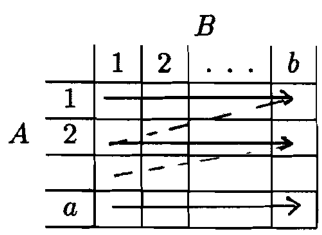
We beschouwen nu de projectoren

$$P,\ P_A,\ P_B \text{ en } P_{AB} \quad .$$

Zoals in L.M. stelt $P := \frac{1}{n}uu'$ de orthogonale projector voor op $u := \{1, \ldots, 1\}'$, zodat $P_y = \{\bar{y}, \ldots, \bar{y}\}'$.
P_A is de projector die middelt per nivo van de factor A, P_B middelt per nivo van B en P_{AB} per nivocombinatie van A en B. Als we voor de waarnemingvector y als volgorde afspreken

$$y_{111}, \ldots, y_{11n}, y_{121}, \ldots, y_{12n}, \ldots, y_{1b1}, \ldots, y_{1bn} \quad ,$$

$$y_{211}, \ldots, y_{21n}, \ldots, y_{ab1}, \ldots, y_{abn} \quad ,$$



dan is

$$P_A = \frac{1}{nb}\begin{pmatrix} \begin{bmatrix}1 & \cdots & 1\\ \vdots & & \vdots\\ 1 & \cdots & 1\end{bmatrix} & & & \emptyset \\ & \begin{bmatrix}1 & \cdots & 1\\ \vdots & & \vdots\\ 1 & \cdots & 1\end{bmatrix} & & \\ & & \ddots & \\ \emptyset & & & \begin{bmatrix}1 & \cdots & 1\\ \vdots & & \vdots\\ 1 & \cdots & 1\end{bmatrix} \end{pmatrix}$$

a matrices $(nb \times nb)$

P_B bestaat uit $a \times a$ blokken van de gedaante

$$\frac{1}{na}\begin{pmatrix} \boxed{\begin{matrix}1 & \cdots & 1\\ \cdots & \cdots & \cdots \\ 1 & \cdots & 1\end{matrix}} \to b \text{ matrices } (n \times n) & & & \\ & \boxed{\begin{matrix}1 & \cdots & 1\\ \cdots & \cdots & \cdots \\ 1 & \cdots & 1\end{matrix}} & & \phi \\ & & \ddots & \\ \phi & & & \boxed{\begin{matrix}1 & \cdots & 1\\ 1 & \cdots & 1\end{matrix}} \end{pmatrix}$$

$$P_{AB} = \frac{1}{n}\begin{pmatrix} \boxed{\begin{matrix}1 & \cdots & 1\\ \cdots & \cdots & \cdots \\ 1 & \cdots & 1\end{matrix}} \to ab \text{ matrices } (n \times n) & & & \\ & \boxed{\begin{matrix}1 & \cdots & 1\\ \cdots & \cdots & \cdots \\ 1 & \cdots & 1\end{matrix}} & & \\ & & \ddots & \\ & & & \boxed{\begin{matrix}1 & \cdots & 1\\ \cdots & \cdots & \cdots \\ 1 & \cdots & 1\end{matrix}} \end{pmatrix}$$

In andere woorden:
P_A is de projectie op de deelruimte opgespannen door a vectoren, die elke constant zijn ($bv = 1$) op één nivo van A en 0 op de $(a-1)$ andere.
P_B is de projectie op de deelruimte opgespannen door b vectoren, elk constant ($\neq 0$) op één van de b nivo's van B en $= 0$ op de andere. P_{AB} is de projectie op de deelruimte opgespannen door ab vectoren, constant ($\neq 0$) op één van de ab nivocombinaties van A en B en $= 0$ op de andere combinaties.
De producten van P, P_A, P_B en P_{AB} volgen het schema:

	P	P_A	P_B	P_{AB}
P	P	P	P	P
P_A	P	P_A	P	P_A
P_B	P	P	P_B	P_B
P_{AB}	P	P_A	P_B	P_{AB}

Met de splitsing in kwadratensommen en met die in vrijheidsgraden correspondeert nu:

$$I - P = (P_A - P) + (P_B - P) \quad +(P_{AB} - P_A - P_B + P)$$
$$+(I - P_{AB})$$

Het is gemakkelijk te verifieren dat de onderlinge producten van de 4 leden van het rechterlid gelijk zijn aan 0. Het zijn 4 projecties op onderling loodrechte deelruimten. Verder is

$$SS_A \quad = y'(P_A - P)y$$

$$SS_B \quad = y'(P_B - P)y$$

$$SS_{AB} \quad = y'(P_{AB} - P_A - P_B + P)y$$

$$SS_R \quad = y'(I - P_{AB})y$$

De modellen met de gebruikelijke veronderstellingen voor de 4 gevallen zijn te vinden in K&S.
Voor de verwachtingen van de gemiddelde kwadratensommen is het voor ons doel beter onderscheid te blijven maken tussen vaste en stochastische nivo's en bv $\frac{1}{a-1}\Sigma\alpha_i^2$ niet te vervangen door σ_A^2.
Deze verwachtingen zijn dan:

	vast	stochastisch	gemengd*)
GK_A	$\frac{nb}{a-1}\Sigma\alpha_i^2 + \sigma^2$	$nb\sigma_A^2 + n\sigma_{AB}^2 + \sigma^2$	$\frac{nb}{a-1}\Sigma\alpha_i^2 + n\sigma_{AB}^2 + \sigma^2$
GK_B	$\frac{na}{b-1}\Sigma\beta_j^2 + \sigma^2$	$na\sigma_B^2 + \sigma_{AB}^2 + \sigma^2$	$na\sigma_B^2 + \sigma^2$
GK_{AB}	$\frac{n}{(a-1)(b-1)}\Sigma(\alpha\beta)_{ij}^2 + \sigma^2$	$n\sigma_{AB}^2 + \sigma^2$	$n\sigma_{AB}^2 + \sigma^2$
GK_R	σ^2	σ^2	σ^2
			*) A vast, B stochastisch

	genest
GK_A	$\frac{nb}{a-1}\Sigma\alpha_i^2 + n\sigma_{B:A}^2 + \sigma^2$
$GK_{B:A}$	$n\sigma_{B:A}^2 + \sigma^2$
GK_R	σ^2

Al deze verwachtingen zijn zonder moeite uit stelling 3 af te leiden. Hierbij levert $\mu' A\mu$ steeds de uidtrukkingen in $\Sigma\alpha_i^2$, $\Sigma\beta_j^2$ en $\Sigma(\alpha\beta)_{ij}^2$ op, terwijl *sp AV* de variantie-componenten geeft.

a. Bij het model met vaste nivo's volgt uit stelling 1a dat de vier kwadratensommen, gedeeld door σ^2, een χ^2-verdeling hebben.
Immers hier is $V = \sigma^2 I$.

De restkwadratensom heeft altijd een centrale χ^2-verdeling, de andere drie een niet- centrale, behalve onder de drie te toetsen nulhypothesen.
De onderlinge onafhankelijkheid volgt direct uit stelling 2a.

b. Bij het model met stochastische nivo's is de variantie-covariantiematrix

$$V = nbP_A\sigma_A^2 + naP_B\sigma_B^2 + nP_{AB}\sigma_{AB}^2 + I\sigma^2 \quad .$$

De verdelingen van de kwadratensommen volgen nu uit stelling 1. Neem bv.

$$\underline{SS}_A = \underline{y}'(P_A - P)\underline{y} \quad .$$

$$(P_A - P)V = (P_A - P)(nb\sigma_A^2 + n\sigma_{AB}^2 + \sigma^2) \quad .$$

Dus $\underline{SS}_A/(nb\sigma_A^2 + n\sigma_{AB}^2 + \sigma^2) \sim \chi_{a-1}^2$.
Hier zijn dus alle vier de kwadratensommen, gedeeld door de verwachting van de gemiddelde kwadratensom, centraal χ^2 verdeeld.
De onafhankelijkheid van bnP_A en P_{AB} volgt uit stelling 2 omdat

$$(P_A - P)\ V(P_{AB} - P_A - P_B + P) =$$

$$= (P_A - P)(P_{AB} - P_A - P_B + P)\ V = 0 \quad ,$$

want alle vermenigvuldigingen zijn commutatief.
Dus $\underline{F} = \dfrac{\underline{SS}_A}{\underline{SS}_{AB}} \cdot \dfrac{(a-1)(b-1)}{(a-1)} \cdot \dfrac{n\sigma_{AB}^2 + \sigma^2}{nb\sigma_A^2 + \sigma_{AB}^2 + \sigma^2}$ heeft een centrale F-verdeling en derhalve is $\underline{GK}_A$ / $\underline{GK}_{AB}$ onder de hypothese $H_0\ :\ \sigma_A^2 = 0$ eveneens centraal F-verdeeld.

c. Het gemengde model

$$\underline{y}_{ijk} = \mu + \alpha_i + \underline{\beta}_j + (\alpha\underline{\beta})_{ij} + \underline{e}_{ijk}$$

met als voorwaarde $\Sigma_i(\alpha\underline{\beta})_{ij} = 0$ en $\text{var}(\alpha\underline{\beta})_{ij} = \dfrac{a-1}{a}\sigma_{AB}^2$ kan (zie K&S) ook geschreven worden als

$$\underline{y}_{ijk} = \mu + \alpha_i + \underline{\gamma}_j + (\alpha\gamma)_{ij} + \underline{e}_{ijk} \quad ,$$

met var $\underline{\gamma}_j = \sigma_C^2$ en $\text{var}(\alpha\underline{\gamma})_{ij} = \sigma_{AC}^2$ en alle stochastische variabelen onafhankelijk. Het verband tussen beide modellen is

$$\begin{aligned} \underline{\beta}_j &= \underline{\gamma}_j + (\alpha\underline{\gamma})_{.j} \\ (\alpha\underline{\beta})_{ij} &= (\alpha\underline{\gamma})_{ij} - (\alpha\underline{\gamma})_{.j} \end{aligned}$$

Er geldt dan

$$
\begin{aligned}
\sigma_B^2 &= \sigma_C^2 + \tfrac{1}{a}\sigma_{AC}^2 \\
\sigma_{AB}^2 &= \sigma_{AC}^2 \\
V &= naP_C\sigma_C^2 + nP_{AC}\sigma_{AC}^2 + I\sigma^2 = \\
&= naP_B(\sigma_B^2 - \tfrac{1}{a}\sigma_{AB}^2) + nP_{AB}\sigma_{AB}^2 + IC^2 \quad .
\end{aligned}
$$

Met stelling 1 vinden we weer de verdelingen van de kwadratensommen, uit stelling 2 volgt hun onderline onafhankelijkheid.

d. In het genest model

$$\underline{y}_{ijk} = \mu + \alpha_i + \underline{\beta}_{j:i} + \underline{e}_{ijk}$$

is de variantie-covariantiematrix

$$V = n\sigma_{B:A}^2 P_{AB} + \sigma^2 I \quad .$$

en

$$SS_{B:A} = \underline{y}'(P_{AB} - P_A)\,\underline{x} \quad .$$

Verdelingen en onafhankelijkheid volgen opnieuw onmiddellijk uit de stellingen 1 en 2.

Eindhoven, augustus 1990

EINDHOVEN UNIVERSITY OF TECHNOLOGY
Department of Mathematics and Computing Science
PROBABILITY THEORY, STATISTICS, OPERATIONS RESEARCH AND SYSTEMS THEORY
P.O. Box 513
5600 MB Eindhoven - The Netherlands
Secretariate: Dommelbuilding 0.03
Telephone: 040 - 47 3130

List of COSOR-memoranda - 1990

Number	Month	Author	Title
M 90-01	January	I.J.B.F. Adan J. Wessels W.H.M. Zijm	Analysis of the asymmetric shortest queue problem Part 1: Theoretical analysis
M 90-02	January	D.A. Overdijk	Meetkundige aspecten van de productie van kroonwielen
M 90-03	February	I.J.B.F. Adan J. Wessels W.H.M. Zijm	Analysis of the assymmetric shortest queue problem Part II: Numerical analysis
M 90-04	March	P. van der Laan L.R. Verdooren	Statistical selection procedures for selecting the best variety
M 90-05	March	W.H.M. Zijm E.H.L.B. Nelissen	Scheduling a flexible machining centre
M 90-06	March	G. Schuller W.H.M. Zijm	The design of mechanizations: reliability, efficiency and flexibility
M 90-07	March	W.H.M. Zijm	Capacity analysis of automatic transport systems in an assembly factory
M 90-08	March	G.J. v. Houtum W.H.M. Zijm	Computational procedures for stochastic multi-echelon production systems

Number	Month	Author	Title
M 90-09	March	P.J.M. van Laarhoven W.H.M. Zijm	Production preparation and numerical control in PCB assembly
M 90-10	March	F.A.W. Wester J. Wijngaard W.H.M. Zijm	A hierarchical planning system versus a schedule oriented planning system
M 90-11	April	A. Dekkers	Local Area Networks
M 90-12	April	P. v.d. Laan	On subset selection from Logistic populations
M 90-13	April	P. v.d. Laan	De Van Dantzig Prijs
M 90-14	June	P. v.d. Laan	Beslissen met statistische selectiemethoden
M 90-15	June	F.W. Steutel	Some recent characterizations of the exponential and geometric distributions
M 90-16	June	J. van Geldrop C. Withagen	Existence of general equilibria in infinite horizon economies with exhaustible resources. (the continuous time case)
M 90-17	June	P.C. Schuur	Simulated annealing as a tool to obtain new results in plane geometry
M 90-18	July	F.W. Steutel	Applications of probability in analysis
M 90-19	July	I.J.B.F. Adan J. Wessels W.H.M. Zijm	Analysis of the symmetric shortest queue problem
M 90-20	July	I.J.B.F. Adan J. Wessels W.H.M. Zijm	Analysis of the asymmetric shortest queue problem with threshold jockeying
M 90-21	July	K. van Harn F.W. Steutel	On a characterization of the exponential distribution
M 90-22	July	A. Dekkers J. van der Wal	Performance analysis of a volume shadowing model

Number	Month	Author	Title
M 90-23	July	A. Dekkers J. van der Wal	Mean value analysis of priority stations without preemption
M 90-24	July	D.A. Overdijk	Benadering van de kroonwielflank met behulp van regeloppervlakken in kroonwieloverbrengingen met grote overbrengverhouding
M 90-25	July	J. van Oorschot A. Dekkers	Cake, a concurrent Make CASE tool
M 90-26	July	J. van Oorschot A. Dekkers	Measuring and Simulating an 802.3 CSMA/CD LAN
M 90-27	August	D.A. Overdijk	Skew-symmetric matrices and the Euler equations of rotational motion for rigid systems
M 90-28	August	A.W.J. Kolen J.K. Lenstra	Combinatorics in Operations Research
M 90-29	August	R. Doornbos	Verdeling en onafhankelijkheid van kwadratensommen in de variantie-analyse