MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig
Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

1953 - 33(1)

Uittreksel van H.B.Mann,

Analysis and design of experiments

New York
Dover Publications, Inc.
1949

door

R.Doornbos

1953

Hoofdstuk I

De χ^2-verdeling en de F-verdeling.

1. De χ^2-verdeling. Als $\underline{x}_1, \ldots, \underline{x}_N$ onderling onafhankelijk en normaal verdeeld zijn met gemiddelde 0 en spreiding 1, dan heeft de uitdrukking

$$\underline{\chi}^2 = \underline{x}_1^2 + \underline{x}_2^2 + \cdots + \underline{x}_N^2$$

een χ^2-verdeling. De verdelingsdichtheid van $\underline{\chi}^2$ is

$$f(\chi^2) = \frac{1}{2^{N/2}\,\Gamma(N/2)} \left(\chi^2\right)^{\frac{N-2}{2}} e^{-\chi^2/2}$$

Het getal N wordt het aantal vrijheidsgraden genoemd. Voor grote waarden van N is $(2\underline{\chi}^2)^{1/2} - (2N-1)^{1/2}$ evenals $(\underline{\chi}^2 - N)/(2N)^{1/2}$ bij benadering normaal verdeeld met gemiddelde 0 en spreiding 1.

Als $\underline{\chi}_1^2$ n_1 vrijheidsgraden heeft en $\underline{\chi}_2^2$ heeft n_2 vrijheidsgraden, dan is $\underline{\chi}_1^2 + \underline{\chi}_2^2$ dus verdeeld als de som van $n_1 + n_2$ normaal verdeelde onderling onafhankelijke variabelen. Dus we hebben

Stelling 1.1: Als $\underline{\chi}_1^2, \underline{\chi}_2^2, \ldots, \underline{\chi}_s^2$ onafhankelijk verdeeld zijn, terwijl $\underline{\chi}_i^2$ ($i = 1, \ldots, s$) een χ^2-verdeling heeft met n_i vrijheidsgraden dan heeft

$$\underline{\chi}^2 = \underline{\chi}_1^2 + \underline{\chi}_2^2 + \cdots + \underline{\chi}_s^2$$

een χ^2-verdeling met $n = n_1 + n_2 + \cdots + n_s$ vrijheidsgraden.

2. De F-verdeling. Stel dat $\underline{\chi}_1^2$ een χ^2-verdeling bezit met n_1 vrijheidsgraden en dat $\underline{\chi}_2^2$ een χ^2-verdeling heeft met n_2 vrijheidsgraden. Onderstel verder dat $\underline{\chi}_1^2$ en $\underline{\chi}_2^2$ onderling onafhankelijk verdeeld zijn. Dan heeft de uitdrukking

$$F = \frac{n_2 \cdot \underline{\chi}_1^2}{n_1 \cdot \underline{\chi}_2^2}$$

een F-verdeling met n_1 en n_2 vrijheidsgraden. Deze verdeling speelt de belangrijkste rol in de variantie-analyse.

Hoofdstuk II

Matrices, kwadratische vormen en de meerdimensional normale verdeling.

1. Matrices en kwadratische vormen. Een matrix is een rechthoekig stelsel van coëfficiënten

$$\begin{pmatrix} \sigma_{11}, & \cdots, & \sigma_{1n} \\ & & \\ \sigma_{m1}, & \cdots, & \sigma_{mn} \end{pmatrix}$$

Als de getallen m en n bekend zijn kunnen we zo'n matrix aangeven met (σ_{ij})

Beschouwen we een stelsel lineaire vormen

(2.1) $\quad L_i = a_{i1}x_1 + \cdots a_{in}x_n \quad , i = 1, \ldots, m$

dan noemen we de matrix (a_{ij}) de matrix van de lineaire vormen L_i in $x_1, \ldots, x_n$. Als nu de x_i weer lineaire vormen in $y_1, \ldots, y_s$ zijn,

$$x_j = b_{j1}y_1 + \cdots + b_{js}y_s$$

dan is

$$L_i = \sum_{j=1}^{n} a_{ij}x_j = \sum_{j=1}^{n}\sum_{k=1}^{s} a_{ij}b_{jk}y_k = \sum_{k=1}^{s} c_{ik}y_k$$

waarbij

$$c_{ik} = \sum_{j=1}^{n} a_{ij}b_{jk} \qquad \begin{cases} i = 1, \ldots, m \\ k = 1, \ldots, s \end{cases}$$

Het ligt dus voor de hand om het product van twee matrices als volgt te definiëren:

$$(a_{ij})(b_{jk}) = \left(\sum_{j=1}^{n} a_{ij}b_{jk}\right)$$

Het product is dus alleen gedefinieerd als het aantal kolommen van de eerste matrix gelijk is aan het aantal rijen van de tweede. Schrijven we

$$(L) = \begin{pmatrix} L_1 \\ \vdots \\ L_m \end{pmatrix} \qquad (X) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

dan kunnen we (2.1) ook aldus schrijven:

$$(L) = (a_{ij})(X)$$

De vermenigvuldiging van matrices is associatief, d.w.z. als A, B en C matrices zijn zo dat (AB) en (BC) gedefinieerd zijn, dan geldt

$$(AB)C = A(BC)$$

De vermenigvuldiging is niet commutatief.

De determinant van een vierkante matrix $(a_{ij}) = A$ geven we aan met $|a_{ij}|$ of $|A|$. Er geldt:

$$|AB| = |A|.|B|$$

De matrix

$$\begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ & & & \\ 0 & 0 & \ldots & 1 \end{pmatrix} = I$$

noemt men de eenheidsmatrix. Voor iedere matrix A, waarvoor AI en IA gedefinieerd zijn geldt dan $AI = IA = A$. Een vierkante matrix A, waarvoor $|A| \neq 0$ bezit een inverse A^{-1}, waarvoor geldt

We gebruiken de notatie $(\sigma_{ij})^{-1} = (\sigma^{ij})$
Bij iedere matrix A bestaat een getransponeerde matrix A', die we verkrijgen door de rijen en kolommen te verwisselen. Het is eenvoudig te verifiëren dat

$$(AB)' = B'A', \quad (AB)^{-1} = B^{-1}A^{-1}, \quad (A^{-1})' = (A')^{-1}$$

We beschouwen kwadratische vormen

$$Q = x'Ax$$

waarbij

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \qquad A = (a_{ij}) \qquad \begin{cases} i = 1, \ldots, n \\ j = 1, \ldots, n \end{cases} \quad a_{ij} = a_{ji}$$

Stel

$$x = Py \quad , \quad P = \begin{pmatrix} p_{11} & \cdots & p_{1m} \\ p_{n1} & \cdots & p_{nm} \end{pmatrix} \quad , \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$$

Dan is $\quad Q = x'Ax = y'P'APy$

dus de matrix van Q als kwadratische vorm in de y's is $P'AP$

Een kwadratische vorm Q in $x_1, \ldots, x_n$ wordt positief (semi)-definiet genoemd als Q groter dan (of gelijk aan) nul is voor alle reële waarden van $x_1, \ldots, x_n$, uitgezonderd $0, \ldots, 0$. Iedere kwadratische vorm kan door een niet singuliere transformatie (d.w.z. dat de determinant waarde van de transformatiematrix $\neq 0$ is)

$$\overline{L}_i = \sum_{j=1}^{n} c_{ij} x_j \qquad , i = 1, \ldots, n$$

worden omgezet in

$$Q = \sum_{i=1}^{r} c_i \overline{L}_i^2 \qquad , r \leq n \; , c_i \neq 0$$

Het getal r heet de rang van Q en is onafhankelijk van de gebruikte transformatie.
Als Q positief semidefiniet is zijn alle $c_i > 0$. Dan kunnen we dus stellen $L_i = \sqrt{c_i}\, \overline{L}_i$ en

$$Q = \sum_{i=1}^{r} L_i^2$$

Lemma 2.1: De som van de rangen van een aantal kwadratische vormen is niet kleiner dan de rang van hun som.

Stel b.v. dat Q_1 de rang n_1 heeft en Q_2 de rang n_2.

$$Q_1 + Q_2 = \sum_{i=1}^{n_1} L_i^2 + \sum_{j=1}^{n_2} M_j^2$$

De L_i en de M_j kunnen alle onafhankelijk zijn, maar er kunnen ook lineaire relaties bestaan tussen sommige L_i en M_j. Met de hulp van deze relaties kunnen dan een aantal van de L_i en M_j worden geëlimineerd. De overblijvende noemen we $N_1, N_2, \ldots, N_{n'}$ ($n' \leq n_1 + n_2$). We vinden dus:

$$Q_1 + Q_2 = \sum_{i=1}^{n'} \sum_{j=1}^{n'} t_{ij} N_i N_j$$

waarbij de N_i onafhankelijke lineaire vormen zijn in de X_i. De rang van $Q_1 + Q_2$ is dus $n' \leq n_1 + n_2$, want volgens de hiervoor genoemde eigenschap is $Q_1 + Q_2$ weer te schrijven als de som van n' kwadraten.

2. De meerdimensionale normale verdeling. Stel $X_1, \ldots, X_n$ zijn simultaan normaal verdeeld met gemiddelde 0 en covariantiematrix (σ_{ij})
De verdelingsdichtheid is dan

$$f(X_1, \ldots, X_n) = \frac{1}{(2\pi)^{n/2} |\sigma_{ij}|^{1/2}} e^{-\frac{1}{2} x'(\sigma^{ij})x}$$

waarbij

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

We passen een niet singuliere lineaire transformatie toe

$$X_i = \sum_{j=1}^{n} p_{ij} Y_j \qquad , i = 1, \ldots, n$$

De determinant van Jacobi van deze transformatie is $|p_{ij}| = |P|$
De absolute waarde van $|P|$ geven we aan met $\|P\|$. De verdelingsdichtheid van de Y's is dus:

$$g(Y_1, \ldots, Y_n) = \frac{\|P\|}{(2\pi)^{n/2} |\sigma_{ij}|^{1/2}} e^{-\frac{1}{2} Y'(\sigma^{*ij})Y}$$

waarin $(\sigma^{*ij}) = P'(\sigma^{ij})P$
Daar $|P|^2 = |P| . |P'|$, kunnen we de constante factor schrijven als

$$\frac{\|P\|}{(2\pi)^{n/2} |\sigma_{ij}|^{1/2}} = \frac{1}{(2\pi)^{n/2} |P^{-1}(\sigma_{ij})P'^{-1}|^{1/2}} = \frac{1}{(2\pi)^{n/2} |\sigma^*_{ij}|^{1/2}}$$

We zien dus dat de Y's ook simultaan normaal verdeeld zijn met gemiddelde 0 en covariantiematrix $(\sigma^*_{ij}) = P^{-1}(\sigma_{ij})P'^{-1}$
De matrix P wordt orthogonaal genoemd als

$$PP' = P'P = I \quad \text{of} \quad P^{-1} = P'$$

Als $X_1, \dots, X_n$ onderling onafhankelijk en normaal verdeeld zijn met spreiding σ, dan is $\sigma_{ij} = \delta_{ij}\sigma^2$ ($\delta_{ij} = 0$ als $i \neq j$ en $= 1$ als $i = j$). Stel dat dit het geval is en dat P orthogonaal is.

Dus

$$(\sigma^*_{ij}) = P^{-1}(\sigma_{ij})P'^{-1} =$$

$$P'\begin{pmatrix}\sigma^2 & 0 & \dots & 0\\ & & & \\ 0 & 0 & \dots & \sigma^2\end{pmatrix}P = P'P\begin{pmatrix}\sigma^2 & 0 & \dots & 0\\ & & & \\ 0 & 0 & \dots & \sigma^2\end{pmatrix} = \begin{pmatrix}\sigma^2 & 0 & \dots & 0\\ & & & \\ 0 & 0 & \dots & \sigma^2\end{pmatrix}$$

want een scalaire matrix $\begin{pmatrix}\lambda & \dots & 0\\ 0 & \dots & \lambda\end{pmatrix}$ mogen we commuteren met iedere andere matrix.

Dus we hebben:

Lemma 2.2: Als $X_1, \dots, X_n$ o.o. normaal verdeeld zijn met gemiddelde 0 en spreiding σ en $x = Py$, waarbij P een orthogonale matrix is, dan zijn $Y_1, \dots, Y_n$ ook o.o. normaal verdeeld met gemiddelde 0 en spreiding σ.

Lemma 2.3: Stel

$$(2.23) \qquad \sum_{i=1}^{n} x_i^2 = Q_1(X) + \dots + Q_s(X)$$

waarbij $Q_i(X)$ een kwadratische vorm is in $X_1, \dots, X_n$ van de rang n_i. Dan bestaat er een orthogonale transformatie

$$(2.24) \qquad X'_i = \sum_{k=1}^{n} p_{ik} x_k \,, \qquad i = 1, \dots, n$$

zodat

$$(2.25) \qquad Q_i = \sum_{k=n_1+\dots+n_{i-1}+1}^{n_1+\dots+n_i} X'^2_k \,, \quad i = 1, \dots, s$$

dan en slechts dan als

$$n_1 + n_2 + \dots + n_s = n$$

Bewijs: Stel dat er een orthogonale transformatie bestaat die aan bovengenoemde eisen voldoet. Dan is de rang van het rechterlid van (2.23) dus $n_1 + \dots + n_s$. De rang van het linkerlid is n, dus $n = n_1 + n_2 + \dots + n_s$.

Stel nu dat $n_1 + \dots + n_s = n$. De rang van Q_1 is n_1. Er bestaat dus een transformatie

$$(2.28) \qquad L_j = \sum_{k=1}^{n} p_{jk} x_k \qquad j = 1, \dots, n_1$$

zodat

$$Q_1 = \sum_{j=1}^{n_1} L_j^2$$

We kunnen ook zeggen: de transformatie (p_{jk}) heeft n rijen, waarvan er n_1 gekwadrateerd en opgesteld Q_1 geven; we nummeren zo dat dit de eerste n_1 zijn; de overige doen er voor Q_1 niet toe.
Dit kunnen we doen voor alle Q_i. Er bestaat dus een transformatie(2.28)waarbij $j = 1, \ldots, n$, zodat

$$\sum_{i=1}^{n} x_i^2 = \sum Q_i = \sum_{j=1}^{n} L_j^2 \quad \text{waarbij}$$

$$Q_i = \sum_{j=n_1+\cdots+n_{i-1}+1}^{n_1+\cdots+n_i} L_j^2$$

De L_j 's zijn lineair onafhankelijk, want anders zou de rang van het rechterlid $< \sum n_i$ zijn en de rang is n en $n = \sum n_i$.
Dus $P = (p_{jk})$ is een niet singuliere matrix en heeft dus een inverse P^{-1}

$$L = Px \quad , \quad x = P^{-1}L$$

$$Q = x' I x = L' P^{-1\prime} I P^{-1} L = \sum L_j^2 = L' I L$$

Dus

$$P^{-1\prime} I P^{-1} = I \;, \text{ of } \; P^{-1\prime} P^{-1} = I$$

Dus P^{-1} is orthogonaal en dus P eveneens.

<u>Stelling 2.1</u>: Stel $\underline{x}_1, \ldots, \underline{x}_n$ zijn onderling onafhankelijk en normaal verdeeld met gemiddelde 0 en spreiding 1.
Is nu

$$\sum_{i=1}^{n} \underline{x}_i^2 = \underline{Q}_1 + \cdots + \underline{Q}_s \;,$$

waarbij $\underline{Q}_i$ ($i = 1, \ldots s$) een kwadratische vorm in $\underline{x}_1, \ldots, \underline{x}_n$ is met rang n_i, dan zijn de $\underline{Q}_i$ onderling onafhankelijk verdeeld en $\underline{Q}_i$ heeft de χ^2-verdeling met n_i vrijheidsgraden dan en slechts dan als $\sum_{i=1}^{s} n_i = n$

(Stelling van COCHRAN)

<u>Bewijs</u>: De noodzakelijkheid van de voorwaarde volgt uit stelling 1.1, het voldoende zijn uit lemma 2.3 en lemma 2.2.

<u>Gevolgtrekking uit stelling 2.1</u>:
Als $\underline{x}_1, \ldots, \underline{x}_n$ normaal en onderling onafhankelijk verdeeld zijn met gemiddelde 0 en spreiding σ en als $\underline{Q}_i$ ($i = 1, \ldots s$) kwadratische vormen zijn in $\underline{x}_1, \ldots, \underline{x}_n$ met rang n_i en

$$\underline{Q}_1 + \cdots + \underline{Q}_s = \sum_{i=1}^{n} \underline{x}_i^2$$

dan heeft $\frac{n_j \cdot Q_i}{n_i \cdot Q_j}$ de F-verdeling met n_i en n_j vrijheidsgraden dan en slechts dan als $\sum_{i=1}^{s} n_i = n$
(Volgt uit definitie van de F -verdeling.)

Hoofdstuk III

Variantieanalyse in het geval van een enkelvoudige classificatie.

Stel $\underline{X}_1, \dots, \underline{X}_s$ zijn onderling onafhankelijk en normaal verdeeld met gemeenschappelijke variantie σ^2 en gemiddelde $\mu_1, \dots, \mu_s$. We willen toetsen de hypothese dat $\mu_1 = \dots = \mu_s$. Stel dat een aselecte steekproef ter grootte n_i is getrokken uit $\underline{X}_i$. De waargenomen waarden noemen we $X_{i1}, \dots, X_{in_i}$

We stellen $X_i = \frac{X_{i1} + \dots + X_{in_i}}{n_i}$ en $X = \frac{\sum_{i,j} X_{ij}}{n}$ waarbij

$$n = \sum_{i=1}^{s} n_i$$

Als toetsingsgrootheid neemt men

$$\underline{F} = \frac{n-s}{s-1} \frac{\sum_i n_i (\underline{X}_i - \underline{X})^2}{\sum_{i,j} (\underline{X}_{ij} - \underline{X}_i)^2} \tag{3.10}$$

$$\mathcal{E}\left[\sum_{i,j} (\underline{X}_{ij} - \underline{X}_i)^2\right] = (n-s)\sigma^2 \text{ en } \mathcal{E}\left[\sum_i n_i (\underline{X}_i - \underline{X})^2\right] = (s-1)\sigma^2 + $$
$$+ \sum_i n_i (\mu_i - \mu)$$

waarbij $\mu = \frac{\sum_i n_i \mu_i}{n}$

Om dit te kunnen bewijzen moeten we de volgende eigenschap gebruiken: Als we t getallen $\alpha_1, \dots, \alpha_t$ hebben en $\alpha = \frac{\alpha_1 + \dots + \alpha_t}{t}$ dan is:

$$\sum_i \alpha_i^2 = \sum_i (\alpha_i - \alpha)^2 + t\alpha^2$$

Door deze eigenschap twee keer toe te passen kunnen we ook bewijzen dat

$$\sum_{i,j} (x_{ij} - \mu)^2 = \sum_{i,j} (x_{ij} - x_i)^2 + \sum_i n_i (x_i - x)^2 + n(x - \mu)^2 =$$
$$= Q_1 + Q_2 + Q_3$$

Daar er tussen de n lineaire vormen $L_{ij} = (x_{ij} - x_i)$ s lineair onafhankelijke relaties bestaan is de rang van $Q_1 \leq n-s$. De rang van Q is evenzo $\leq s-1$ en de rang van $Q_3 = 1$. Maar $(n-s)+(s-1)+1 = n$, dus volgens lemma 2.1 is de rang van Q_1 $(n-s)$ en van Q_2 $(s-1)$. Als $\mu_1 = \dots = \mu_s = \mu$, dan zijn $(X_{ij} - \mu)$ onderling onafhankelijk en normaal verdeeld met gemiddelde 0 en spreiding σ.

Uit de gevolgtrekking uit stelling 2.1 volgt dan, dat in
(3.10) F de F-verdeling heeft met $(s-1)$ en $(n-s)$ vrijheidsgraden.

Als we willen toetsen of $\mu_i = \mu_j$, dan is onze toetsingsgrootheid:

$$(3.19) \qquad F = \frac{n-s}{1} \; \frac{n_i n_j}{n_i + n_j} \; \frac{(\underline{x}_i - \underline{x}_j)^2}{Q_1}$$

Deze F heeft de F-verdeling met 1 en $(n-s)$ vrijheidsgraden.
Dit zien we als volgt:

$$\sum_{i,j} (x_{ij} - \mu)^2 = Q_1 + \sum_{\substack{l \neq i \\ l \neq j}} n_l (x_l - x)^2 + \frac{n_i n_j}{n_i + n_j} (x_i - x_j)^2 +$$

$$+ (n_i + n_j)(x' - x)^2 + n (x - \mu)^2,$$

waarin $\quad x' = \dfrac{n_i x_i + n_j x_j}{n_i + n_j}$

De rang van

$$\sum_{\substack{l \neq i \\ l = j}} n_l (x_l - x)^2 + (n_i + n_j)(x' - x)^2 \text{ is } \leq (s-2)$$

en de rang van $\dfrac{n_i n_j}{n_i + n_j} (x_i - x_j)^2$ is 1. Dus volgens de gevolgtrekking uit 2.1 heeft F de F-verdeling met 1 en $(n-s)$ vrijheidsgraden.

Voor het geval dat $s = 2$ komt dit overeen met de toets van STUDENT voor het verschil tussen de gemiddelden van twee normaal verdeeld grootheden met gelijke spreiding. De toetsingsgrootheid is dan

$$\underline{t} = \frac{\underline{x}_1 - \underline{x}_2}{\sqrt{\sum (\underline{x}_{1i} - \underline{x}_1)^2 + \sum (\underline{x}_{2i} - \underline{x}_2)^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

We zien dus dat als $s = 2$, $\underline{F} = \underline{t}^2$.
Als we in (3.10) $s=2$ nemen krijgen we dezelfde uitdrukking na enige herleidingen.

Hoofdstuk IV

λ-Toetsen en het toetsen van lineaire hypothesen.

We beschouwen de stochastische grootheid $\underline{x}$ met verdelingsdichtheid $f(x; \theta_1, \ldots, \theta_k)$; $\underline{x}$ mag ook een stochastische vector zijn, dit verandert niets aan het navolgende betoog. De parameterruimte met coördinaten $\theta_1, \ldots, \theta_k$ zullen we aangeven met Ω. Iedere f kunnen we dus voorstellen door een punt in Ω. Op grond van een steekproef $x_1, \ldots, x_n$ willen we toetsen de hypothese H_0, die inhoudt dat de verdelingsdichtheid behoort tot ω, een deelruimte van Ω.

De aannemelijkheidsfunctie van de steekproef is per definitie

$$L = \prod_{i=1}^{n} f(x_i; \theta_1, \ldots, \theta_k)$$

Het maximum van L over Ω zullen we aangeven met $L(\hat{\Omega})$; in ω heeft L het maximum $L(\hat{\omega})$. Als toetsingsgrootheid nemen we nu

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})}$$

Als critieke zone kiezen we $\lambda < \lambda_0$, waarbij λ_0 zo gekozen is dat $P[\lambda < \lambda_0 \mid H_0] = \alpha$, als α de onbetrouwbaarheidsdrempel is. Als k de dimensie van Ω en r die van ω is, dan is voor grote steekproeven $-2 \ln \lambda$ bij benadering χ^2-verdeeld met $k-r$ vrijheidsgraden, indien H_0 waar is.

In de variantieanalyse nemen we als toetsingsgrootheid een monotone functie van λ, waarvan de exacte verdeling bekend is.

Beschouw een aantal stochastische grootheden $\underline{y}_1, \ldots, \underline{y}_N$. We maken de volgende onderstellingen:

1) De $\underline{y}_\alpha$ zijn normaal en onderling onafhankelijk verdeeld met gelijke spreiding.

2) Stel $\mathcal{E}\underline{y}_\alpha = \mu_\alpha$, de μ_α zijn lineaire functies van p parameters $\beta_1, \ldots, \beta_p$; $p \leq N$

$$\mu_\alpha = \sum_i g_{i\alpha} \beta_i \,, \quad (\alpha = 1, \ldots N) \tag{4.6}$$

en de rang van $(g_{i\alpha}) = p$

Door de β_i uit (4.6) te elimineren zien we dat onderstelling 2) overeenkomt met:

$$\sum_\alpha l_{k\alpha} \mu_\alpha = 0 \quad , \quad (k = 1, \ldots, N-p) \tag{4.6'}$$

$$\text{rang } (l_{k\alpha}) = N-p$$

De hypothese die we willen toetsen is dat bovendien

$$\sum k_{ij} \beta_j = 0 \quad , \quad (i = 1, \ldots, s \,;\, s \leq p) \tag{4.7}$$

waarbij de vergelijking van (4.6) en (4.7) onafhankelijk zijn. Of, zoals we zien door de β_j uit (4.6) en (4.7) te elimineren

(4.7') $$\sum_{\alpha=1}^{N} r_{k\alpha}\mu_\alpha = 0\ ,\ (k = 1,\ldots,s),$$

terwijl de

(4.8) rang van $\begin{bmatrix} l_{11} & \cdots & l_{1N} \\ \vdots & & \\ l_{N-p,1} & \cdots & l_{N-p,N} \\ r_{11} & & r_{1N} \\ & & \\ r_{s1} & & r_{sN} \end{bmatrix}$ gelijk aan $N-p+s$ is.

De simultane verdelingsdichtheid van $y_1,\ldots,y_N$ is

$$L = \frac{1}{\sigma^N (2\pi)^{N/2}}\, e^{-\frac{1}{2}\sum_\alpha \frac{(y_\alpha-\mu_\alpha)^2}{\sigma^2}}\ ;$$

$$\ln L = -\tfrac{N}{2}\ln\sigma^2 - \tfrac{N}{2}\ln 2\pi - \tfrac{1}{2}\sum_\alpha \frac{(y_\alpha-\mu_\alpha)^2}{\sigma^2}$$

Als we voor de μ_α het rechterlid van (4.6) substitueren en $\ln L$ differentiëren naar σ^2 en naar $\beta_1,\ldots,\beta_p$ en de verkregen vergelijkingen nul stellen, dan kunnen we de maximum likelihood schattingen $b_1,\ldots,b_p$ voor $\beta_1,\ldots,\beta_p$ en $\hat{\sigma}^2$ voor σ^2 oplossen. In de p vergelijkingen waarin de $b_1,\ldots,b_p$ voorkomen, komt σ^2 niet voor. Stel

$$Q_\alpha = \sum_\alpha \Big(y_\alpha - \sum_i g_{i\alpha} b_i\Big)^2$$

Q_α is dus het minimum van $Q = \sum_\alpha (y_\alpha - \mu_\alpha)$ onder de voorwaarden (4.6).

$$\frac{\partial L}{\partial\sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{\sigma^4}\sum_\alpha (y_\alpha-\mu_\alpha)^2$$

Dus

$$-\frac{N}{2\hat{\sigma}^2} + \frac{1}{(\hat{\sigma}^2)} Q_\alpha = 0\ ,$$

$$\hat{\sigma}^2 = \frac{1}{N} Q_\alpha$$

Dus

$$L(\hat{\Omega}) = \left(\frac{N}{2\pi Q_\alpha}\right)^{N/2} e^{-N/2}$$

Analoog vinden we dat

$$L(\hat{\omega}) = \left(\frac{N}{2\pi Q_r}\right)^{N/2} e^{-N/2}$$

waarin Q_r het minimum is van Q onder de voorwaarde (4.6) en (4.7). Dus

$$\lambda = \left(\frac{Q_\alpha}{Q_r}\right)^{N/2} \tag{4.14}$$

Als toetsingsgrootheid gebruiken wij

$$\underline{F} = \frac{N-p}{s} \; \frac{Q_r - Q_\alpha}{Q_\alpha} \tag{4.15}$$

een monotone functie van λ. We verwerpen de hypothese (4.7) als $F \geq F_0$, waarbij

$$P[\underline{F} \geq F_0 \mid (4.7)] = \alpha$$

We zullen bewijzen dat het quotiënt (4.15) de F-verdeling bezit met s en $N-p$ vrijheidsgraden.

<u>Lemma 4.1.</u> <u>Stel</u>

$$\sum_{j=1}^{N} a_{ij}\mu_j = 0 \quad , (i=1,\dots,k) \tag{4.16}$$

<u>zijn k lineair onafhankelijke lineaire betrekkingen tussen $\mu_1,\dots\mu_N$. Dan bestaat er een systeem van betrekkingen</u>

$$\sum_{j=1}^{N} b_{ij}\mu_j = 0 \quad , (i=1,\dots,k) \tag{4.17}$$

<u>zodat de eerste l relaties van (4.16) aequivalent zijn met de eerste l van (4.17) ($l \leq k$) en zodanig dat de rijen van (b_{ij}) orthogonaal zijn t.o.v. elkaar, d.w.z.</u>

$$\sum_{j=1}^{N} a_{ij} b_{kj} = \delta_{ik}$$

Het bewijs is constructief en verloop zo dat eerst de eerste rij van (a_{ij}) genormeerd wordt en dan de tweede rij van (b_{ij}) wordt bepaald als lineaire combinatie van de 1e rij van (b_{ij}) en de 2e van (a_{ij}), waarna we weer normeren. Zo gaan we door, zodat b.v. de $(l+1)$e rij van (b_{ij}) wordt bepaald door een lineaire combinatie van de eerste l rijen van (b_{ij}) en de $(l+1)$e van (a_{ij}) te normeren.

Door dit lemma toe te passen, kunnen we aannemen dat de rijen van de matrix (4.8) orthogonaal en genormeerd zijn. Als $p > s$ vormen we een aanvullende rij $t_{11},\dots,t_{1N}$, zodat

$$\sum_{\alpha=1}^{N} l_{i\alpha}\, t_{1\alpha} = 0 \quad , (i = 1, \ldots, N-p),$$

$$\text{en} \quad \sum_{\alpha=1}^{N} r_{j\alpha}\, t_{1\alpha} = 0 \quad , (j = 1, \ldots, s)$$

Dit kan; we hebben nl. $N-p+s$ vergelijkingen in N onbekenden. Zo vullen we (4.8) aan tot een orthgonale matrix

$$(4.19) \qquad \begin{pmatrix} l_{11} & \cdots & l_{1N} \\ \vdots & & \vdots \\ l_{N-p,1} & \cdots & l_{N-p,N} \\ r_{11} & \cdots & r_{1N} \\ \vdots & & \vdots \\ r_{s1} & \cdots & r_{sN} \\ t_{11} & \cdots & t_{1N} \\ \vdots & & \vdots \\ t_{p-1,1} & \cdots , & t_{p-s,N} \end{pmatrix}$$

Wij beschouwen nu de transformatie:

$$y_\alpha^* = \sum_\alpha l_{i\alpha}\, y_\alpha \qquad (i = 1, \ldots, N-p)$$

$$y_{N-p+k}^* = \sum_\alpha r_{k\alpha}\, y_\alpha \qquad (k = 1, \ldots, s)$$

$$y_{N-p+s+u}^* = \sum_\alpha t_{u\alpha}\, y_\alpha \qquad (u = 1, \ldots, p-s)$$

De vergelijking $Q = \sum_\alpha (y_\alpha - \mu_\alpha)^2$ stelt in de N-dimensionale Euclidische ruimte met coördinaten $y_1, \ldots y_N$ een bol voor, die door bovengenoemde orthogonale transformatie in een bol met gelijke straal wordt omgezet. Daarbij wordt bovendien het middelpunt meegetransformeerd, d.w.z. het nieuwe middelpunt heeft als coördinaten

$$\mu_\alpha^* = \sum_\alpha l_{i\alpha}\, \mu_\alpha \qquad (i = 1, \ldots, N-p)$$

$$\mu_{N-p+k}^* = \sum_\alpha r_{k\alpha}\, \mu_\alpha \qquad (k = 1, \ldots, s)$$

$$\mu_{N-p+s+u}^* = \sum_\alpha t_{u\alpha}\, \mu_\alpha \qquad (u = 1, \ldots, p-s)$$

We hebben dus

$$Q = \sum_{\alpha} (y_\alpha - \mu_\alpha)^2 = \sum_{\alpha} (y_\alpha^* - \mu_\alpha^*)^2$$

Q_α is het minimum van Q onder de voorwaarden (4.6'), dus

$$Q_\alpha = \sum_{\alpha=1}^{N-p} y_\alpha^{*2},$$

de andere p termen kunnen we nl. nul maken door $\hat{\mu}_\alpha^* = y_\alpha$ te nemen. Evenzo geldt:

$$Q_r = \sum_{\alpha=1}^{N-p+s} y_\alpha^{*2}$$

Dus

$$Q_r - Q_\alpha = \sum_{\alpha=N-p+1}^{N-p+s} y_\alpha^{*2}$$

Volgens lemma 2.2 zijn de y_α^* normaal en onafhankelijk verdeeld en hun verwachting is, zoals gemakkelijk in te zien is, gelijk aan μ_α^*.

Volgens hoofdstuk I zijn Q_α en $Q_r - Q_\alpha$ dus, onder de nulhypothese, onafhankelijk verdeeld volgens χ^2-verdelingen met $N-p$ resp. s vrijheidsgraden.

Dus

$$F = \frac{N-p}{s} \; \frac{Q_r - Q_\alpha}{Q_\alpha}$$

heeft de F-verdeling met s en $N-p$ vrijheidsgraden.

We hebben dus nu bewezen

Stelling 4.1 . Stel dat $y_1, \ldots, y_N$ normaal en onafhankelijk verdeelde grootheden zijn met dezelfde variantie en gemiddelden $\mu_1, \ldots, \mu_N$. Stel dat de μ_α voldoen aan de onafhankelijke betrekkingen

(4.25) $$\sum_\alpha l_{i\alpha} \mu_\alpha = 0 \qquad i = 1, \ldots, N-p$$

Om de hypothese te toetsen dat de μ_α voldoen aan relaties

(4.26) $$\sum_\alpha r_{i\alpha} \mu_\alpha = 0 \qquad i=1,\ldots,s \qquad (s \leq r)$$

onafhankelijk van de betrekkingen (4.25) en van elkaar, vormen we het quotient

(4.27) $$\underline{F} = \frac{N-p}{s} \; \frac{\underline{Q}_r - \underline{Q}_\alpha}{\underline{Q}_\alpha},$$

waarbij $\underline{Q}_\alpha$ het minimum is van $\sum_\alpha (\underline{y}_\alpha - \mu_\alpha)^2$ met betrekking tot de μ_α onder de voorwaarden (4.25) en $\underline{Q}_r$ het minimum van dezelfde uitdrukking onder de voorwaarden (4.25) en (4.26). We verwerpen de hypothese (4.26) als $\underline{F} \geq F_0$, waarbij $P[\underline{F} \geq F_0 \mid (4.25) \cup (4.26)] = \alpha$ en α is een constante.

Dan geldt:

1) De beschreven toets is aequivalent met de λ-toets voor de hypothese (4.26)
2) Het quotient (4.27) heeft de F-verdeling met s en N-p vrijheidsgraden

Voorbeeld. We gaan nu stelling 4.1 toepassen op het in hoofdstuk III behandelde geval van de enkelvoudige klassificatie. We hadden een aantal stochastische grootheden $\underline{X}_1, \ldots, \underline{X}_s$, o.o. en $N(\mu_i, \sigma)$ verdeeld. We namen een steekproef ter grootte n_i uit $\underline{X}_i$, de waargenomen waarden waren $x_{i1}, \ldots, x_{in_i}$ $(i=1,\ldots,s)$.

We kunnen dit als volgt aanpassen aan de formulering van stelling 4.1: de stochastische grootheden $\underline{x}_{ij}$ $(i=1,\ldots,s; j=1,\ldots,n_i)$ zijn onderling onafhankelijk en normaal verdeeld, met gemeenschappelijke spreiding σ en gemiddelden μ_{ij}. De μ_{ij} voldoen aan $\sum_i n_i - s = \overline{\text{lineair}}\lfloor n-s$ onafhankelijke betrekkingen (analoog aan 4.6' en 4.25):

$$\mu_{11} - \mu_{12} = 0$$
$$\mu_{12} - \mu_{13} = 0$$
$$- - - - - - -$$
$$\mu_{1,n_1} - \mu_{1n_1} = 0$$
$$- - - - - - -$$
$$\mu_{s1} - \mu_{s2} = 0$$
$$- - - - - - -$$
$$\mu_{s,n_s-1} - \mu_{sn_s} = 0$$

In de vorm van (4.6) geschreven luidt de toegelaten hypothese:

$$\mu_{11} = \mu_1$$
$$\vdots$$
$$\mu_{1n_1} = \mu_1$$
$$\mu_{21} = \mu_2$$
$$\vdots$$
$$\mu_{2n_2} = \mu_2$$
$$------$$
$$\mu_{sn_1} = \mu_s$$
$$\mu_{sn_s} = \mu_s.$$

De hypothese die we willen toetsen, luidt:

$$\mu_{11} - \mu_{21} = 0$$
$$--------$$
$$\mu_{s-1,1} - \mu_{s,1} = 0.$$

Dit zijn dus s-1 lineair onafhankelijke betrekkingen.
In de tweede vorm:

$$\mu_1 - \mu_2 = 0$$
$$-------$$
$$\mu_{s-1} - \mu_s = 0.$$

Om Q_α te vinden moeten we naar de μ_i's minimaliseren

$$Q = \sum_{i,j} (x_{ij} - \mu_i)^2.$$

Stel dat m_i de hieruit verkregen schatting is van μ_i. Door differentiatie naar μ_i vinden we dan:

$$\sum_{j=1}^{n_i} (x_{ij} - m_i) = 0.$$

Dus:

$$m_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} = x_{i.}$$

Dus

$$Q_\alpha = \sum_{i,j} (x_{ij} - x_{i.})^2 = \sum_{i,j} x^2_{ij} - \sum_i n_i x^2_{i.}$$

Om Q_r te vinden minimaliseren we de uitdrukking

$$\sum_{i,j} (x_{ij} - \mu)^2 \text{ naar } \mu.$$

De schatting m van μ is dus $\frac{1}{n} \sum_{i,j} x_{ij} = x_{..}$.

Dus

$$Q_r = \sum_{i,j} (x_{ij} - x_{..})^2 = \sum_{i,j} x^2_{ij} - n\, x^2_{..}$$

En

$$Q_r - Q_\alpha = \sum_i n_i x_{i.}^2 - n\, x_{..}^2 = \sum_i n_i (x_{i.} - x_{..})^2$$

We vinden dus de toetsingsgrootheid (3.10) terug:

$$\underline{F} = \frac{n-s}{s-1} \; \frac{\sum_i n_i (\underline{x}_{i.} - x_{..})^2}{\sum_{i,j} (\underline{x}_{ij} - \underline{x}_{i.})^2}$$

Om te toetsen $\mu_i = \mu_j = \mu'$, vinden we Q_r door naar de μ_k's en μ' te minimaliseren:

$$\sum_{\substack{k,l \\ k \neq i,j}} (x_{kl} - \mu_k)^2 + \sum_{l=1}^{n_i} (x_{il} - \mu')^2 + \sum_{l=1}^{n_j} (x_{jl} - \mu')^2$$

De schattingen m_k van μ_k worden dus weer $x_{k.}$ Door naar μ' te differentieren, vinden we voor m' de schatting van μ';

$$(n_i x_{i.} - n_i m') + (n_j x_{j.} - n_j m') = 0.$$

$$m' = \frac{n_i x_{i.} + n_j x_{j.}}{n_i + n_j} .$$

Dus

$$Q_r = \sum_{\substack{k,l \\ k \neq i,j}} (x_{kl} - x_{k.})^2 + \sum_{l=1}^{n_i} (x_{il} - m')^2 + \sum_{l=1}^{n_j} (x_{jl} - m')^2$$

$$= \sum_{k,l} x_{kl}^2 - \sum_{\substack{k=1 \\ k \neq i,j}}^{s} n_k x_{k.}^2 - \frac{(n_i x_{i.} + n_j x_{j.})^2}{n_i + n_j} .$$

Dus

$$Q_r - Q_\alpha = n_i x_{i.}^2 + n_j x_{j.}^2 - \frac{(n_i x_{i.} + n_j x_{j.})^2}{n_i + n_j} =$$

$$= \frac{n_i n_j}{n_i + n_j} (x_{i.} - x_{j.})^2 .$$

We vinden dus opnieuw de in hoofdstuk III verkregen toetsingsgrootheid:

$$\underline{F} = \frac{n-s}{1} \; \frac{n_i n_j}{n_i + n_j} \; \frac{(\underline{x}_{i.} - \underline{x}_{j.})^2}{\underline{Q}_\alpha} .$$

We zullen nu enige stellingen afleiden, die ons in staat zullen stellen de toetsingsgrootheid (4.27) in vele gevallen eenvoudig te berekenen. Stel

$$Y_\alpha = b_1 g_{1\alpha} + \ldots + b_p g_{p\alpha} .$$

Y_α is dus de schatting van μ_α onder de voorwaarde (4.6)

$b_1, \ldots, b_p$ minimaliseren de uitdrukking

$$\sum_\alpha (y_\alpha - \beta_1 g_{1\alpha} - \ldots - \beta_p g_{p\alpha})^2 .$$

Dus moet gelden

(4.36) $\sum_\alpha (y_\alpha - b_1 g_{1\alpha} \ldots b_p g_{p\alpha}) g_{i\alpha} = 0$ $\qquad (i = 1, \ldots, p)$

Door de 1e vergelijking van (4.30) te vermenigvuldigen met b_i en daarna over i te sommeren, vinden we

(4.32) $\sum_\alpha (y_\alpha - Y_\alpha) Y_\alpha = 0$, of

$$\sum_\alpha y_\alpha Y_\alpha = \sum_\alpha Y_\alpha^2 .$$

Per definitie geldt:

$$(4.33) \quad Q_\alpha = \sum_\alpha (y_\alpha - Y_\alpha)^2 = \sum_\alpha y_\alpha^2 - \sum_\alpha Y_\alpha^2$$

Stel Y_α^* is de schatting van μ_α onder de voorwaarden (4.6) en (4.7). Dan vinden we dus analoog aan het voorgaande:

$$(4.34) \quad Q_r = \sum_\alpha (y_\alpha - Y_\alpha^*)^2 = \sum_\alpha y_\alpha^2 - \sum_\alpha Y_\alpha^{*2}$$

Dus

$$(4.35) \quad Q_r - Q_\alpha = \sum_\alpha Y_\alpha^2 - \sum_\alpha Y_\alpha^{*2}$$

De voorwaarden (4.7) zijn aequivalent met

$$(4.36) \quad \beta_i = \sum_{j=1}^{p-s} k_{ij}\gamma_j \qquad (i=1,\dots,p)$$

Als $c_1,\dots,c_{p-s}$ de maximum likelihood schattingen zijn van $\gamma_1,\dots,\gamma_{p-s}$, dan geldt

$$Y_\alpha^* = \sum_l c_l \sum_t g_{t\alpha} k_{tl}$$

Door nu de t^e vergelijking van (4.30) te vermenigvuldigen met $k_{tl}\, c_l$ en te sommeren over t en l krijgen we

$$(4.38) \quad \sum_\alpha (y_\alpha - Y_\alpha) Y_\alpha^* = 0$$

Bovendien geldt, analoog aan (4.32) $\sum_\alpha (y_\alpha - Y_\alpha^*) Y_\alpha^* = 0$

Dus

$$\sum_\alpha (Y_\alpha - Y_\alpha^*) Y_\alpha^* = \sum_\alpha (y_\alpha - Y_\alpha^*) Y_\alpha^* - \sum_\alpha (y_\alpha - Y_\alpha) Y_\alpha^* = 0$$

Dus

$$(4.39) \quad \sum_\alpha (Y_\alpha - Y_\alpha^*)^2 = \sum_\alpha Y_\alpha^2 - \sum_\alpha Y_\alpha^{*2}$$

Dus

$$Q_r - Q_\alpha = \sum_\alpha (Y_\alpha - Y_\alpha^*)^2$$

We hebben dus gevonden dat

$$(4.41) \quad F = \frac{N-p}{s} \frac{\sum (Y_\alpha - Y_\alpha^*)^2}{\sum_\alpha (y_\alpha - Y_\alpha)^2} = \frac{N-p}{s} \frac{\sum_\alpha Y_\alpha^2 - \sum_\alpha Y_\alpha^{*2}}{\sum_\alpha y_\alpha^2 - \sum_\alpha Y_\alpha^2}$$

Uit (4.33) en (4.39) volgt verder dat

$$\sum_\alpha (y_\alpha - Y_\alpha)^2 = \sum_\alpha y_\alpha^2 - \sum_\alpha Y_\alpha^{*2} - \sum_\alpha (Y_\alpha - Y_\alpha^*)^2$$

Dit resultaat is te generaliseren als volgt:

<u>Stelling 4.2.</u> <u>Stel $H_1,\dots,H_s$ zijn een stelsel hypothesen betreffende de gemiddelden van de stochastische variabelen y_α,</u> <u>met $\mathcal{E}\, y_\alpha = \mu_\alpha$</u> <u>van de vorm</u>

$$H_1: \quad \mu_\alpha = \sum_{i=1}^{k} g_{i\alpha} \beta_i$$

$$H_t: \quad H_1 \cup H_2 \cup \cdots \cup H_{t-1} \cup \sum_{j=1}^{k} a_{kj} \beta_j = 0$$

$$(k = s_{t-1}+1, \ldots, s_t), \; s_t \leq p$$

zodanig dat de lineaire vormen in de β's, die gelijk zijn aan nul volgens H_s, lineair onafhankelijk zijn. Stel $Y_\alpha^{(t)}$ is de schatting van μ_α onder de hypothese H_t, dan is

$$\sum_\alpha y_\alpha^2 = \sum_\alpha (y_\alpha - Y_\alpha^{(1)})^2 + \sum_\alpha (Y_\alpha^{(1)} - Y_\alpha^{(2)})^2 + \cdots$$

$$\cdots + \sum_\alpha (Y_\alpha^{(s-1)} - Y_\alpha^{(s)})^2 + \sum_\alpha Y_\alpha^{(s)2}$$

We willen nu de vergelijkingen (4.50) gaan oplossen. Stel

$$\sum_\alpha g_{i\alpha} g_{j\alpha} = a_{ij} \quad , \quad \sum_\alpha y_\alpha g_{i\alpha} = a_i$$

$$(i, j = 1, \ldots, p)$$

Lemma 4.2. Stel $(g_{i\alpha})$ $(i = 1, \ldots, p \; ; \alpha = 1, \ldots, N)$ is een matrix van de rang $p \leq N$. Definieer $gg' = (a_{ij})$. Dan is de kwadratische vorm $\sum_{ij} a_{ij}$ positief definiet.

Bewijs:

$$Q = \sum_{ij} a_{ij} x_i x_j = \sum_{\alpha, ij} g_{i\alpha} g_{j\alpha} x_i x_j = \sum_\alpha (\sum_i g_{i\alpha} x_i)^2$$

Dus Q is zeker positief semidefiniet. Als $Q = 0$, dan is dus

(4.45) $\qquad \sum_i g_{i\alpha} x_i = 0 \qquad (\alpha = 1, \ldots, N)$

De rang van $(g_{i\alpha})$ is p. Dus (4.45) heeft alleen de triviale oplossing $x_i = 0$ ($i = 1, \ldots, p$). Dus is Q positief definiet.

Als gevolg hiervan is de determinant $|a_{ij}| \neq 0$. We kunnen (4.30) schrijven als

$$\sum_{j=1}^{p} a_{ij} b_j = a_i \qquad (i = 1, \ldots, p)$$

Of in matrix-notatie

$$(a_{ij})(b) = (a) \quad , \text{ waarin } (b) = \begin{bmatrix} b_1 \\ \vdots \\ b_p \end{bmatrix} \quad (a) = \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix}$$

Daar (a_{ij}) volgens lemma 4.2 niet singulier is, hebben we

$$(b) = (a^{ij})(a)$$

of

$$(4.48) \qquad b_i = \sum_{j=1}^{p} a^{ij} a_j \qquad (i = 1, \ldots, p)$$

We zien dat b_i een lineaire functie is van de y_α. Als dus de y_α normaal verdeeld zijn, dan zijn de b_i ook simultaan normaal verdeeld. Uit (4.48) is eenvoudig af te leiden dat

$$\mathcal{E}\, b_i = \beta_i$$

$\sigma_{b_i b_j} = a^{ij}\sigma^2$, als σ de spreiding is van y_α.

We geven nu een stelling zonder bewijs:

<u>Stelling 4.3.</u> <u>Stel Q_α is het minimum van de kwadratische vorm $\sum (y_\alpha - \mu_\alpha)^2$, onder de voorwaarde</u>

$$(4.51) \qquad \mu_\alpha = \sum_{i=1}^{s} \beta_i g_{i\alpha} + \sum_{d=s+1}^{p} \beta_d g_{d\alpha} \quad (\alpha = 1, \ldots, N)$$

<u>en Q_r het minimum onder de hypothese $\beta_i = 0$ $(i = 1, \ldots, s)$.</u>
<u>Stel b_i $(i = 1, \ldots, s)$, b_d $(d = s+1, \ldots, p)$ zijn de maximum likelihood schattingen van $\beta_1, \ldots, \beta_p$ onder de voorwaarden (4.51) en stel</u>

$$\left(\frac{\sigma_{b_i b_j}}{\sigma^2}\right)^{-1} = c_{ij} \qquad (i, j = 1, \ldots, s)$$

<u>Dan geldt</u>

$$Q_r - Q_\alpha = \sum_{i,j} c_{ij}\, b_i\, b_j$$

Uit deze stelling volgt: Stel dat de hypothese in stelling 4.3 is:

$$\beta_i^* = \sum_{t=1}^{p} l_{it}\, \beta_t = 0 \qquad (i = 1, \ldots, s \leq p)$$

waarbij de rang van $(l_{it}) = s$

Stel verder

$$\sum_t l_{it}\, b_t = b_i^* \quad \text{en} \quad \left(\frac{\sigma_{b_i^* b_j^*}}{\sigma^2}\right)^{-1} = c_{ij}$$

Dan geldt

$$Q_r - Q_\alpha = \sum_{i,j} c_{ij}\, b_i^*\, b_j^*$$

<u>Bewijs</u>: Daar de rang van (l_{it}) gelijk is aan s, kunnen we $(p-s)$ rijen toevoegen zodat we een niet singuliere matrix (l_{vt}) krijgen met p rijen en p kolommen. De β_t zijn dan ook lineaire functies van $\beta_v^* = \sum_t l_{vt}\, \beta_t$ $(v = 1, \ldots, p)$. Door nu stelling 4.3. toe te passen op de β_v^* zien we dat het gestelde waar is.

Vaak komt het geval voor dat $s = 1$, dan geldt dus

$$Q_r - Q_\alpha = \frac{b_1^2 \sigma^2}{\sigma_{b_1}^2}$$

Met behulp van de gevolgtrekking uit stelling 4.3 kunnen we op zeer eenvoudige wijze het tweede gedeelte van het op p. 16 behandelde voorbeeld oplossen.

Stel
$$\beta^* = \mu_{i.} - \mu_{j.}$$

Dus
$$b^* = x_{i.} - x_{j.} ,$$
$$\sigma_{b^*}^2 = \sigma_{x_{i.}}^2 + \sigma_{x_{j.}}^2 = \left(\frac{1}{n_i} + \frac{1}{n_j}\right)\sigma^2 = \frac{n_i + n_j}{n_i n_j}\,\sigma^2 .$$

Dus
$$c = \frac{n_i n_j}{n_i + n_j}$$

en
$$Q_r - Q_\alpha = \frac{n_i n_j}{n_i + n_j} (x_{i.} - x_{j.})^2 .$$

Als tweede voorbeeld beschouwen we de regressie-vergelijking
$$\mathcal{E}\, y = \beta_1 + \beta_2 x .$$
waarbij we N waarnemingen $y_1, \ldots, y_N$ doen voor de waarden $x_1, \ldots, x_N$ van x. Minstens 2 x-waarden moeten verschillend zijn; stel dat dit x_1 en x_2 zijn. De toegelaten hypothese heeft dus de vorm
$$\mu_\alpha = \beta_1 + \beta_2 x .$$
We kunnen dit ook schrijven als $N-2$ lineair onafhankelijke betrekkingen tussen de μ_α:
$$\mu_1 (x_2 - x_i) + \mu_2 (x_i - x_1) + \mu_i (x_1 - x_2) = 0 \quad (i = 3, \ldots, N)$$
We toetsen de hypothese $\beta_2 = 0$, of in de tweede vorm: $\mu_1 = \mu_2$.
$Q_\alpha = \sum_\alpha (y_\alpha - b_1 - b_2 x_\alpha)^2$ waarin b_1 en b_2 de kleinste kwadratenschattingen van β_1 en β_2 zijn. We bepalen b_1 en b_2 met behulp van formule (4.48). In ons geval is $g_{1\alpha} = 1$ en $g_{2\alpha} = x_\alpha$ $(\alpha = 1, \ldots, N)$

Dus
$$a_{11} = N$$
$$a_{21} = a_{12} = \sum_\alpha x_\alpha = N\bar{x}.$$
$$a_{22} = \sum_\alpha x_\alpha^2$$
$$a_1 = \sum_\alpha y_\alpha = N\bar{y}.$$
$$a_2 = \sum_\alpha x_\alpha y_\alpha$$

$$|a| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = N \sum_{\alpha} x_{\alpha}^2 - N^2 x_{.}^2 = N \sum_{\alpha} (x_{\alpha} - x_{.})^2$$

$$a^{11} = \frac{\sum_{\alpha} x_{\alpha}^2}{N \sum_{\alpha} (x_{\alpha} - x_{.})^2}$$

$$a^{12} = \frac{-x_{.}}{\sum_{\alpha} (x_{\alpha} - x_{.})^2}$$

$$a^{22} = \frac{1}{\sum_{\alpha} (x_{\alpha} - x_{.})^2}$$

We vinden dus voor onze parameters:

$$b_1 = a^{11} a_1 + a^{22} a_2 = \frac{y_{.} \sum_{\alpha} x_{\alpha}^2 - x_{.} \sum_{\alpha} x_{\alpha} y_{\alpha}}{\sum_{\alpha} (x_{\alpha} - x_{.})^2}\,.$$

$$b_2 = a^{21} a_1 + a^{22} a_2 = \frac{-N x_{.} y_{.} + \sum x_{\alpha} y_{\alpha}}{\sum_{\alpha} (x_{\alpha} - x_{.})^2} = \frac{\sum_{\alpha} (x_{\alpha} - x_{.})(y_{\alpha} - y_{.})}{\sum_{\alpha} (x_{\alpha} - x_{.})^2}$$

Het is gemakkelijk te verifiëren dat

$$b_1 = y_{.} - b_2 x_{.}$$

Dus

$$Y_{\alpha} = b_1 + b_2 x_{\alpha} = y_{.} + b_2 (x_{\alpha} - x_{.})\,.$$

$$Q_{\alpha} = \sum (y_{\alpha} - b_1 - b_2 x_{\alpha})^2 = \sum_{\alpha} \left\{ (y_{\alpha} - y_{.}) - b_2 (x_{\alpha} - x_{.}) \right\}^2 =$$

$$\sum_{\alpha} (y_{\alpha} - y_{.})^2 - b_2^2 \sum_{\alpha} (x_{\alpha} - x_{.})^2$$

Om $Q_r - Q_{\alpha}$ met behulp van stelling (4 .) te bepalen, moeten we $\sigma_{b_2}^2$ berekenen. Zoals we op p. 19 hebben afgeleid is

$$\sigma_{b_2}^2 = a^{22} \sigma^2 = \frac{\sigma^2}{\sum (x_{\alpha} - x_{.})^2}\,.$$

Dus

$$Q_r - Q_{\alpha} = b_2^2 \sum_{\alpha} (x_{\alpha} - x_{.})^2.$$

Dit is ook als volgt in te zien. Om Q_r te vinden, moeten we minimaliseren naar β_1 : $\sum_{\alpha} (y_{\alpha} - \beta_1)^2$. Dit geeft voor de schatting b_1' van β_1 : $b_1' = y_{.}$.

Dus

$$Q_r = \sum_{\alpha} (y_{\alpha} - y_{.})^2 \quad \text{en}$$

$$Q_r - Q_{\alpha} = b_2^2 \sum_{\alpha} (x_{\alpha} - x_{.})^2.$$

De gezochte toetsingsgrootheid is dus:

$$F = \frac{N-2}{1} \; \frac{b_2^2 \sum_\alpha (x_\alpha - x_.)^2}{\sum_\alpha (y_\alpha - y_.)^2 - b_2^2 (x_\alpha - x_.)^2} ,$$

waarin

$$b_2 = \frac{\sum_\alpha (x_\alpha - x_.)(y_\alpha - y_.)}{\sum_\alpha (x_\alpha - x_.)^2}$$

Soms moet Q_α worden bepaald als het minimum van $Q = \sum_\alpha \left(y_\alpha - \sum_j g_{j\alpha} \beta_j\right)^2$ onder zekere bijvoorwaarden opgelegd aan de β's.

(4.65) $$\sum_{j=1}^{p} l_{ij} \beta_j = L_i = 0 \quad (i = 1, \dots, s)$$

$$\text{rang } (l_{ij}) = s < p$$

Door middel van (4.56) kunnen al de β's worden voorgesteld als lineaire functies van $\beta_1^*, \dots, \beta_t^*, \beta_{t+1}^*, \dots \beta_{p-s}^*$, waarin de β_i^* lineaire functies van $\beta_1, \dots, \beta_p$ zijn. Stel dat we willen toetsen de hypothese $\beta_1^* = \dots = \beta_t^* = 0$. We kunnen Q schrijven als functie van $\beta_1^*, \dots, \beta_{p-s}^*$, dus we kunnen stelling (4.3) toepassen op de maximum likelihoodschattingen $b_1, \dots, b_t$.

Vaak is het eenvoudiger om de methode van de multiplicatoren van Lagrange toe te passen. Hierbij wordt de uitdrukking

(4.66) $$Q' = Q + \lambda_1 L_1 + \dots + \lambda_s L_s$$

gedifferentieerd naar $\beta_1, \dots, \beta_p$ en dan worden uit

(4.67) $$\frac{\partial Q'}{\partial \beta_i} = 0 \quad (i = 1, \dots, p)$$

en (4.65) de $p+s$ onbekenden $\beta_1, \dots, \beta_p$ en $\lambda_1, \dots, \lambda_s$ opgelost. De aldus verkregen waarden voor $\beta_1, \dots, \beta_p$ zijn juist de maximum likelihoodschattingen $b_1, \dots, b_p$ onder de bijvoorwaarden (4.65).

Stelling 4.4: Stel

$$\mathcal{E}\, y_\alpha = \mu_\alpha = \sum_{i=1}^{p} g_{i\alpha} \beta_i$$

Neem aan dat

1) $g_{p\alpha} = 1 \quad (\alpha = 1, \dots, N)$

Als dit het geval is noemen we β_p het algemeen gemiddelde.

2) $g_{i\alpha} = 0$ of $1 \quad (i = 1, \dots, s)$

3) $\sum_{i=1}^{s} g_{i\alpha} = 1 \quad (\alpha = 1, \dots, N)$.

Als

$$Q = \sum_{\alpha} (y_{\alpha} - \sum_{i=1}^{p} g_{i\alpha} \beta_i)^2$$

wordt geminimaliseerd met betrekking tot

$$\beta_1, \ldots, \beta_p \text{ en } \sum_{i=1}^{s} t_i \beta_i = 0 \quad , \sum t_i \neq 0$$

als de enige beperking opgelegd aan $\beta_1, \ldots, \beta_s$, β_p en als λ_1 de multiplicator van Lagrange is die behoort bij

$$\sum_{i=1}^{s} t_i \beta_i \quad , \quad \text{dan geldt} \quad \lambda_1 = 0$$

Er mag op worden gewezen dat nog een willekeurig aantal beperkingen mag zijn opgelegd aan $\beta_{s+1}, \ldots, \beta_{p-1}$ Uit 3) volgt.

$$3') \sum_{\alpha} g_{i\alpha} g_{j\alpha} = 0 \qquad \text{voor } i \neq j \text{ en } i,j \leq s$$

want van de getallen $g_{t\alpha}$ $(t = 1, \ldots, s)$ is er een gelijk aan 1 en de andere zijn 0. Als we de schattingen voor $\beta_1, \ldots, \beta_p$ aangeven met $b_1, \ldots, b_p$ dan vinden we door Q' te differentiëren naar $\beta_1, \ldots, \beta_s$ en β_p de volgende vergelijkingen (rekening houdende met 3'))

$$\sum_{\alpha} y_{\alpha} g_{i\alpha} - b_i \sum_{\alpha} g_{i\alpha}^2 - \sum_{d=s+1}^{p} b_d \sum_{\alpha} g_{d\alpha} g_{i\alpha} - \frac{t_i \lambda_1}{2} = 0$$

(4.68)

$$\sum_{\alpha} y_{\alpha} - \sum_{i=1}^{s} b_i \sum_{\alpha} g_{i\alpha} - \sum_{d=s+1}^{p} b_d \sum_{\alpha} g_{d\alpha} = 0$$

Volgens voorwaarde 3) is $\sum_{\alpha} y_{\alpha} = \sum_{\alpha} \sum_{i=1}^{s} y_{\alpha} g_{i\alpha}$.

Uit 2) volgt: $\sum_{\alpha} g_{i\alpha} = \sum_{\alpha} g_{i\alpha}^2$.

Volgens 3) geldt verder

$$\sum_{\alpha} \sum_{i=1}^{s} g_{i\alpha} g_{d\alpha} = \sum_{\alpha} g_{d\alpha}.$$

Als we nu de eerste s relaties van (4.68) sommeren over i en daarna de laatste relatie van (4.68) er van af trekken zien we dat

$$- \frac{\sum_i t_i \lambda_1}{2} = 0$$

Volgens het gegeven is $\sum_i t_i \neq 0$, dus $\lambda_1 = 0$

Stelling 4.5: Stel

$$Q = \sum_{\alpha} (y_{\alpha} - \sum_i g_{i\alpha} \beta_i)^2$$

Stel dat Q_{α} het minimum is van Q onder de bijvoorwaarde

$$\sum_{i=1}^{s} \beta_i = 0$$

en andere bijvoorwaarden die alleen betrekking hebben op $\beta_{s+1}, \ldots, \beta_p$. Noem de schattingen van $\beta_1, \ldots, \beta_s$ onder deze voorwaarden $b_1, \ldots, b_s$. Laat Q_r het minimum zijn onder de voorwaarde $\beta_1 = \ldots = \beta_s = 0$ en alle andere voorwaarden.

Als $Q_r - Q_\alpha$ een symmetrische functie is van $b_1, \ldots, b_s$ en als de voorwaarden van stelling 4.4 vervuld zijn dan geldt

(4.70) $$Q_r - Q_\alpha = \frac{s-1}{sc} \sum_{i=1}^{s} b_i^2 \quad \text{, waarin } \sigma_{b_i}^2 = c\sigma^2$$

Bewijs: Volgens stelling 4.3 is $Q_r - Q_\alpha$ een kwadratische vorm in $b_1, \ldots, b_s$. Volgens het gegeven is het een symmetrische vorm in $b_1, \ldots, b_s$ en daar bovendien $\sum b_i = 0$, geldt

(4.71) $$Q_r - Q_\alpha = h \sum_{i=1}^{s} b_i^2 + k \sum_{i,j=1}^{s} b_i b_j = (h-k) \sum_{i=1}^{s} b_i^2 .$$

Vanwege de symmetrie moeten de varianties van de b_i alle gelijk zijn. Volgens stelling 4.1 heeft $Q_r - Q_\alpha$ de χ^2-verdeling met $(s-1)$ vrijheidsgraden. Dus

$$\mathcal{E}(Q_r - Q_\alpha) = (s-1)\sigma^2 = (h-k)\, s\, c\, \sigma^2$$

Dus $$(h-k) = \frac{s-1}{sc}$$

Uit (4.71) volgt nu meteen het gestelde.

en andere bijvoorwaarden, die alleen betrekking hebben op $\beta_{s+1},\dots,\beta_p$. Noem de schattingen van $\beta_1,\dots,\beta_s$ onder deze voorwaarden $b_1,\dots,b_s$. Laat Q_r het minimum zijn onder de voorwaarde $\beta_1=\dots=\beta_s=0$ en alle andere voorwaarden.

Als $Q_r - Q_\alpha$ een symmetrische functie is van $b_1,\dots,b_s$ dan geldt

(4.70) $Q_r - Q_\alpha = \frac{s-1}{sc}\sum_{i=1}^{s} b_i^2$, waarin $\sigma^2_{b_i} = c\,\sigma^2$.

Bewijs: We kunnen b_s vervangen door $-(b_1+\dots+b_{s-1})$. Volgens stelling 4.3 is $Q_r - Q_\alpha$ een kwadratische vorm in $b_1,\dots,b_{s-1}$ en dus ook in $b_1,\dots,b_s$. Volgens het gegeven is het een symmetrische vorm en daar bovendien $\sum_{i=1}^{s} b_i = 0$, geldt

(4.71) $$Q_r - Q_\alpha = h\sum_{i=1}^{s} b_i^2 + k\sum_{\substack{i,j=1\\ i\neq j}}^{s} b_i b_j = (h-k)\sum_{i=1}^{s} b_i^2 =$$
$$2(h-k)\sum_{i=1}^{s-1} b_i^2 + (h-k)\sum_{\substack{i,j=1\\ i\neq j}}^{s-1} b_i b_j$$

Volgens stelling (4.3) geldt dus nu

$$\left(\frac{\sigma_{b_i b_j}}{\sigma^2}\right)^{-1} = \left([1+\delta_{ij}][h-k]\right)$$

Dus voor $i=1,\dots,s-1$ is

$$c = \frac{\sigma^2_{b_i}}{\sigma^2} = \frac{1}{h-k}\,\frac{\Delta_{s-2}}{\Delta_{s-1}},\quad \text{waarin } \Delta_n = \begin{vmatrix} 2 & 1 & \dots & 1 \\ 1 & 2 & \dots & 1 \\ \vdots & & & \vdots \\ 1 & 1 & \dots & 2 \end{vmatrix}$$ (n rijen en kolommen)

ofwel

$$c = \frac{1}{h-k}\,\frac{s-1}{s},$$
$$h-k = \frac{s-1}{sc}$$

Maar ook $\frac{\sigma^2_{b_s}}{\sigma^2} = c$, want het voorgaande blijft gelden als we niet b_s maar een willekeurige b_i ($i \leq s$) uitdrukken in de andere b's en stelling 4.3 toepassen.

Dus $$Q_r - Q_\alpha = (h-k)\sum_{i=1}^{s} b_i^2 = \frac{s-1}{sc}\sum_{i=1}^{s} b_i^2.$$

We behandelen nu het voorbeeld van de tweevoudige classificatie. Stel b.v. dat $r.s$ varkens van r verschillende rassen s verschillende diëten krijgen toegediend, zodanig dat juist één varken van het i^e ras ($i=1,\dots r$) het j^e dieet krijgt ($j=1,\dots s$). We willen zowel de verschillen tussen de rassen als tussen de diëten onderzoeken. $\underline{x}_{ij}$ is de gewichtstoename van het varken van het i^e ras dat het j^e dieet krijgt.

We onderstellen nu dat de gewichtstoename wordt veroorzaakt door twee factoren, ras en voedsel, die onafhankelijk van elkaar werken. Verder onderstellen we dat de $\underline{x}_{ij}$ normaal en onafhankelijk verdeeld zijn, alle met dezelfde spreiding. De toegelaten hypothese is de volgende:

(4.82) $$\mathcal{E}\underline{x}_{ij} = \mu_{i.} + \mu_{.j} + \mu \;, \sum_i \mu_{i.} = \sum_j \mu_{.j} = 0 .$$

Indien we de aanname, dat de beide factoren onafhankelijk van elkaar werken laten varen, gebruiken we het model:

(4.82') $$\mathcal{E}\underline{x}_{ij} = \mu_{ij} + \mu_{i.} + \mu_{.j} + \mu . \left(\sum_i \mu_{ij} = \sum_j \mu_{ij} = \sum_i \mu_{i.} = \sum_j \mu_{.j} = 0\right.$$

We hebben nu evenveel parameters (nl. $(r-1)(s-1)+(r-1)+(s-1)+1=rs$) als stochastische grootheden. Welke hypothese we dan ook willen toetsen, de toetsingsgrootheid (4.27) wordt onbepaald. In $F=\frac{N-p}{s}\frac{Q_r-Q_\alpha}{Q_\alpha}$, is nl. $N-p=0$ en $Q_\alpha=0$. We kunnen Q_α nl. nul maken door voor de schattingen van de parameters te nemen:

$$m = x_{..} \;,$$

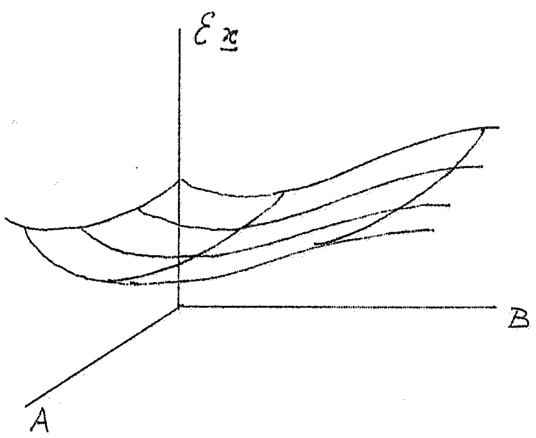
$$m_{.j} = x_{.j} - x_{..} \quad (j=1,\dots,s) \;,$$

$$m_{i.} = x_{i.} - x_{..} \quad (i=1,\dots,r) \;,$$

$$m_{ij} = x_{ij} - x_{i.} - x_{.j} + x_{..} \; (i=1,\dots,r\,;\, j=1,\dots s) .$$

In het algemeen als we N stochastische ~~grootheden~~ x_α hebben en de μ_α hangen van N parameters af, dan kunnen we de parameters zo schatten uit N vergelijkingen met N onbekenden, dat $Q_\alpha = \min \sum_\alpha (y_\alpha - \mu_\alpha)^2$ gelijk wordt aan nul.

In het model (4.82') levert de interpretatie van de parameters $\mu_{i.}$ en $\mu_{.j}$ moeilijkheden op. Stel we hebben een rechthoekig veld, waarop een graansoort wordt verbouwd, die wordt bemest met de meststoffen A, in hoeveelheden $0,1,2,\dots,r-1$ en B in de hoeveelheden $0,1,2,\dots,s-1$. Het is nu niet zo dat $\mu_{i.}$ de invloed is van een bemesting i met A, zonder dat B wordt toegediend, zelfs niet dat $\mathcal{E}(\underline{x}_{i0} - \underline{x}_{k0}) = \mu_{i.} - \mu_{k.}$. Dit laatste geldt alleen, indien $\mu_{i0} = \mu_{k0}$ ofwel: $\sum_{j\neq 0} \mu_{ij} = \sum_{j\neq 0} \mu_{kj}$. Als $\mu_{ij}=0$ voor iedere i en j, dan wordt het oppervlak $\mathcal{E}\underline{x} = \mu_{i.} + \mu_{.j} + \mu$ boven het A-B-vlak vastgelegd door de randwaarden langs de lijnen $A=0$ en $B=0$. Het ontstaat door de lijn $A=0$ evenwijdig aan zichzelf te verplaatsen



langs de lijn $B = 0$ of omgekeerd. Als de $\mu_{ij} \neq 0$ zijn is het een willekeurig vlak, waarvan alle punten bekend moeten zijn om het te kunnen karakteriseren.

Om Q_α te vinden moeten we $Q = \sum_{i,j} (x_{ij} - \mu_{i.} - \mu_{.j} - \mu)^2$ minimaliseren met inachtneming van de relatie (4.82).

Nu gelden de voorwaarden van stelling 4.4 zowel voor de $\mu_{i.}$ als voor de $\mu_{.j}$. We hoeven dus geen rekening te houden met de beperkingen $\sum \mu_{i.} = 0$, $\sum \mu_{.j} = 0$. Dus de meest aannemelijke schattingen $m_{i.}$, $m_{.j}$ en m van $\mu_{i.}$, $\mu_{.j}$ en μ zijn:

$$m = \frac{1}{rs} \sum_{i,j} x_{ij} = x_{..}$$

$$m_{i.} = \frac{1}{s} \sum_j x_{ij} - m = x_{i.} - x_{..} \,,$$

$$m_{.j} = \frac{1}{r} \sum_i x_{ij} - m = x_{.j} - x_{..} \,.$$

Dus de regressiewaarde Y_{ij} is

(4.83) $$Y_{ij} = x_{i.} + x_{.j} - x_{..} \,.$$

We gaan nu stelling 4.2 toepassen en beschouwen het stelsel hypothesen

(4.84)
$$\begin{aligned} H_1 &: \mu_{ij} = \mu_{i.} + \mu_{.j} + \mu \,, \\ H_2 &: H_1 \;\&\; \mu_{i.} = 0 \quad (i = 1, \ldots, r) \,, \\ H_3 &: H_2 \;\&\; \mu_{.j} = 0 \quad (j = 1, \ldots, s) \,. \end{aligned}$$

De bijbehorende regressiewaarden zijn:

$$Y_{ij}^{(1)} = x_{i.} + x_{.j} - x_{..} \,,$$

$$Y_{ij}^{(2)} = x_{.j} \,,$$

$$Y_{ij}^{(3)} = x_{..}$$

Volgens stelling 4.2 geldt nu dus

$$\sum_{i,j} x_{ij}^2 = \sum_{i,j} (x_{ij} - x_{i.} - x_{.j} + x_{..})^2 + s \sum_i (x_{i.} - x_{..})^2 + r \sum_j (x_{.j} - x_{..})^2 + rs\, x_{..}^2 \,.$$

Dus $$Q_\alpha = \sum_{i,j} (x_{ij} - x_{i.} - x_{.j} + x_{..})^2 =$$
$$= \sum_{i,j} x_{ij}^2 - s\sum_i (x_{i.} - x_{..})^2 - r\sum_j (x_{.j} - x_{..})^2 - rs\, x_{..}^2 .$$

Om te toetsen $\mu_{i.} = 0$ hebben we
$$Q_r - Q_\alpha = \sum_{i,j} (Y_{ij}^{(1)} - Y_{ij}^{(2)})^2 = s\sum_i (x_{i.} - x_{..})^2 .$$

Onze toetsingsgrootheid is dus
$$F_2 = \frac{(r-1)(s-1)}{(r-1)} \; \frac{s\sum_i x_{i.}^2 - rs\, x_{..}^2}{\sum_{i,j} x_{ij}^2 - s\sum_i x_{i.}^2 - r\sum_j x_{.j}^2 + rs\, x_{..}^2}$$

Als we $\mu_{i.} = 0$ & $\mu_{.j} = 0$ willen toetsen is
$$Q_r - Q_\alpha = \sum_{i,j} (Y_{ij}^{(1)} - Y_{ij}^{(3)})^2 = \sum_{i,j} (x_{i.} + x_{.j} - 2x_{..})^2$$

Dus $$F_3 = \frac{(r-1)(s-1)}{r+s-2} \; \frac{\sum_{i,j} (x_{i.} + x_{.j} - 2x_{..})^2}{Q_\alpha}$$

Willen we bv. toetsen $H_{12} : \mu_{1.} = \mu_{2.}$, dan passen we de gevolgtrekking uit stelling 4.3 toe:
$$\beta^* = \mu_{1.} - \mu_{2.} .$$

Dus $$b^* = x_{1.} - x_{2.} .$$
$$\sigma_{b^*}^2 = \sigma_{x_{1.}}^2 + \sigma_{x_{2.}}^2 = \frac{2}{s}\sigma^2$$

Dus $$Q_r - Q_\alpha = \frac{s}{2}(x_{1.} - x_{2.})^2 .$$

Hoofdstuk V

Variantie-analyse bij een r-voudig classificatieschema.

In het laatste voorbeeld van het vorige hoofdstuk hadden we rs grootheden x_{ij}. De waarnemingen konden op twee manieren in klassen worden gerangschikt en x_{ij} was de waarde die werd waargenomen in de i^e klasse van de eerste en de j^e klasse van de tweede classificatie.

In dit hoofdstuk zal hiervan een generalisatie worden behandeld, nl. de r-voudige classificatie. Om practische redenen zal r beperkt blijven tot hoogstens 4 of 5.

Als voorbeeld van een drievoudige classificatie beschouwen we het volgende geval. Stel we hebben t_1 weerstations. Gedurende t_3 opeenvolgende jaren is door deze stations voor t_2 maanden de gemiddelde regenval gemeten. Iedere waarneming is dan gekarakteriseerd door drie getallen, het nummer van het station, de maand en het jaar waarin de waarneming werd verricht. De waarnemingen kunnen we dus aangeven met $x_{a_1 a_2 a_3}$ $(a_1 = 1, \dots, t_1\ ;\ a_2 = 1, \dots, t_2\ ;\ a_3 = 1, \dots t_3)$.

We willen bv. weten of de regenval verschil voor de verschillende plaatsen of in verschillende jaren. Dat er tussen de verschillende maanden verschillen zijn is wel zeker. Daarnaast is het ook van belang te weten of de combinatie van een bepaalde plaats met een bepaalde maand invloed heeft op de regenval, of dat de regenval ongewoon groot was in Juli van een bepaald jaar. Zo beschouwen we de gemiddelde regenval in een bepaald station in een bepaalde maand en in een bepaald jaar als te zijn samengesteld uit de effecten van station, maand en jaar even goed als uit het effect van de "interacties" (interaction) van maand en jaar, maand en station, jaar en station en ten slotte één effect tengevolge van de interactie van maand, jaar en station. Dus is

$$
\begin{aligned}
\mathcal{E}\, \underline{x}_{a_1 a_2 a_3} &= \mu(1,2,3\,;a_1,a_2,a_3) + \mu(2,3\,;a_2,a_3) + \\
(5.1) \qquad &+ \mu(1,3\,;a_1,a_3) + \mu(1,2\,;a_1,a_2) + \\
&+ \mu(1;a_1) + \mu(2\,;a_2) + \mu(3\,;a_3) + \mu,
\end{aligned}
$$

waarbij

$$
\begin{aligned}
&\sum_{a_1} \mu(1,2,3\,;a_1,a_2,a_3) = \sum_{a_2} \mu(1,2,3\,;a_1,a_2,a_3) = \\
&= \sum_{a_3} \mu(1,2,3\,;a_1,a_2,a_3) = \sum_{a_{i_1}} \mu(i_1,i_2\,;a_{i_1},a_{i_2}) = \\
&= \sum_{a_{i_2}} \mu(i_1,i_2\,;a_{i_1},a_{i_2}) = \sum_{a_j} \mu(j\,;a_j) = 0 \\
&(i_1,i_2 = 2,3\,;\ 1,3 \text{ of } 1,2 \text{ en } j = 1,2,3).
\end{aligned}
$$

Ook hier zullen we weer in het model moeten opnemen:
$\mu(1,2,3\,;a_1,a_2,a_3) = 0,\ (a_1 = 1, \dots, t_1\ ;\ a_2 = 1, \dots, t_2\ ;\ a_3 = 1, \dots t_3)$,
daar we anders geen hypothese kunnen toetsen.

In het algemeen stellen we bij een z -voudige variantie-analyse:

$$
\begin{aligned}
\mathcal{E}\, \underline{x}_{a_1,a_2,\dots,a_z} &= \sum_{\alpha=0}^{z} \sum^{*}_{1,\dots,z} \mu(i_1,\dots,i_\alpha\,;a_{i_1},\dots,a_{i_\alpha}) \\
(5.3) \qquad & \qquad\qquad\qquad\qquad (a_\ell = 1,\dots,t_\ell) \\
&\sum_{a_{i_j}=1}^{t_{i_j}} \mu(i_1,\dots,i_k\,;a_{i_1},\dots,a_{i_k}) = 0 \ (j = 1,\dots,k).
\end{aligned}
$$

Met $\sum\limits_{1,\ldots,r}^{*} \mu(i_1,\ldots,i_\alpha; a_{i_1},\ldots,a_{i_\alpha})$ geven we aan dat gesommeerd wordt over alle combinaties $i_1,\ldots,i_\alpha$ gekozen uit $1,\ldots,r$ met $i_1 < i_2 < \cdots < i_\alpha$.

We hebben nu juist $t_1 \cdots t_r$ onafhankelijke parameters. Dat is als volgt in te zien: In de $\mu(1,\ldots,r; a_1,\ldots,a_r)$ hebben we $(t_1-1)(t_2-1)\cdots(t_r-1)$ onafhankelijke parameters, in de $\mu(i_1,\ldots,i_\alpha; a_{i_1},\ldots,a_{i_\alpha})$ $(t_{i_1}-1)\cdots(t_{i_\alpha}-1)$, enz. Er zijn in totaal dus

$$\sum_{\alpha=0}^{r} \sum_{1,\ldots,r}^{*} (t_{i_1}-1)\cdots(t_{i_\alpha}-1) = \{(t_1-1)+1\}\cdots\{(t_r-1)+1\} =$$

$t_1 \ldots t_r$ onafhankelijke parameters. $\mu(i_1,\ldots,i_\alpha; a_{i_1},\ldots,a_{i_\alpha})$ noemen we een interactie van de orde $\alpha-1$.

We berekenen nu eerst de schattingen $m(i_1,\ldots,i_\alpha; a_{i_1},\ldots,a_{i_\alpha})$ van de $\mu(i_1,\ldots,i_\alpha; a_{i_1},\ldots,a_{i_\alpha})$ onder de voorwaarden (5.3). Deze berkening zullen we uitvoeren voor het geval $r = 3$.

We definiëren $x(i_1,\ldots,i_\alpha; a_{i_1},\ldots,a_{i_\alpha})$ als

$$x(i_1,\ldots,i_\alpha; a_{i_1},\ldots,a_{i_\alpha}) = \frac{1}{\prod t_j} \sum_{j} \sum_{a_j} x_{a_1,\ldots,a_r} \quad .$$
$$(j \neq i_1,\ldots,i_\alpha)$$

Bij het berekenen van de m's zullen we eerst dat deel berekenen dat we krijgen door de multiplicatoren van Lagrange nul te stellen. Daarna verifiëren we dat de schattingen voor de μ's die we aldus verkrijgen voldoen aan de restricties van (5.3) en dat ze bovendien de eenduidig bepaalde oplossing van het algemene minimumprobleem vormen. Het minimaliseren van

$$Q = \sum_{a_1,a_2,a_3} \left[x_{a_1,a_2,a_3} - \sum_{\alpha=0}^{3} \sum_{1,2,3}^{*} \mu(i_1,\ldots,i_\alpha; a_{i_1},\ldots,a_{i_\alpha}) \right]^2 =$$

$$= \sum_{a_1,a_2,a_3} \Big[x_{a_1,a_2,a_3} - \mu(1,2,3; a_1,a_2,a_3) - \mu(2,3; a_2,a_3) - \mu(1,3; a_1,a_3)$$
$$- \mu(1,2; a_1,a_2) - \mu(1; a_1) - \mu(2; a_2) - \mu(3; a_3) - \mu \Big]^2$$

leidt tot de vergelijkingen:

$$x = m$$

$$x(3; a_3) = m(3; a_3) + m$$

$$x(2;a_2) = m(2;a_2) + m$$

$$x(1;a_1) = m(1;a_1) + m$$

$$x(1,2;a_1,a_2) = m(1,2;a_1,a_2) + m(1;a_1) + m(2;a_2) + m$$

$$x(1,3;a_1,a_3) = m(1,3;a_1,a_3) + m(1;a_1) + m(3;a_3) + m$$

(5.4')
$$x(2,3;a_2,a_3) = m(2,3;a_2,a_3) + m(2;a_2) + m(3;a_3) + m$$

$$x(1,2,3;a_1,a_2,a_3) = m(1,2,3;a_1,a_2,a_3) + m(2,3;a_2,a_3) +$$
$$+ m(1,3;a_1,a_3) + m(1,2;a_1,a_2) + m(1;a_1) + m(2;a_2) +$$
$$+ m(3;a_3) + m.$$

Hierin stelt $m(i_1,\ldots,i_\alpha;a_{i_1},\ldots,a_{i_\alpha})$ de schatting voor van $\mu(i_1,\ldots,i_\alpha;a_{i_1},\ldots,a_{i_\alpha})$.

Lossen we hieruit m's op, dan vinden we:

$$m = x$$

$$m(1;a_1) = x(1;a_1) - x$$

— — — — — — —

(5.5')
$$m(2,3;a_2,a_3) = x(2,3;a_2,a_3) - x(2;a_2) - x(3;a_3) + x$$

— — — — — — — — — — —

$$m(1,2,3;a_1,a_2,a_3) = x(1,2,3;a_1,a_2,a_3) - x(2,3;a_2,a_3) - x(1,3;a_1,a_3)$$
$$- x(1,2;a_1,a_2) + x(1;a_1) + x(2;a_2) + x(3;a_3) - x.$$

In het algemene geval geldt:

(5.4)
$$x(i_1,\ldots,i_\alpha;a_{i_1},\ldots,a_{i_\alpha}) = \sum_{\beta=0}^{\alpha} \sum_{i_1,\ldots,i_\alpha}^{*} m(k_1,\ldots,k_\beta;a_{k_1},\ldots,a_{k_\beta})$$

en

(5.5)
$$m(i_1,\ldots,i_\alpha;a_{i_1},\ldots,a_{i_\alpha}) = \sum_{\beta=0}^{\alpha} (-1)^{\alpha-\beta} \sum_{i_1,\ldots,i_\alpha}^{*} x(k_1,\ldots,k_\beta;a_{k_1},\ldots,a_{k_\beta})$$

Voor $z = 3$ is het eenvoudig na te gaan dat

$$\sum_{\substack{a_i \\ (i=1,2,3)}} m(1,2,3;a_1,a_2,a_3) = \sum_{\substack{a_{i_j} \\ (j=1,2)}} m(i_1,i_2;a_{i_1},a_{i_2}) = \sum_{\substack{a_i \\ (i=1,2,3)}} m(i;a_i) = 0$$

Vullen we deze m's in in Q, dan zien we dat $Q_\alpha = 0$, daar de m's voldoen aan de laatste betrekking van (5.4'), of voor het r-voudige geval aan die vergelijking van (5.4) waarin $\alpha = r$. Onder de restricties van (5.3) vormen de m's uit (5.5) dus de oplossing van het algemene minimumprobleem.

Uit (5.4) volgt

$$(5.10)\qquad \sum_{a_{i_1}}\cdots\sum_{a_{i_\alpha}}\left[x(i_1,\ldots,i_\alpha;a_{i_1},\ldots,a_{i_\alpha})\right]^2 = \\ = \sum_{\beta=0}^{\alpha}\sum^{*}_{i_1,\ldots,i_\alpha}\frac{t_{i_1}\cdots t_{i_\alpha}}{t_{k_1}\cdots t_{k_\beta}}\sum_{a_{k_1}}\cdots\sum_{a_{k_\beta}}\left[m(k_1,\ldots,k_\beta;a_{k_1},\ldots,a_{k_\beta})\right]^2$$

Gaan we het rechterlid van (5.4) nl. kwadrateren en dan sommeren, dan vallen alle dubbele producten weg en blijven alleen de kwadraten, met als coëfficiënten die uit het rechterlid van (5.10) over.

Voor $\alpha = r$ krijgen we het speciale geval

$$(5.7)\qquad \sum_{a_1}\cdots\sum_{a_r}\left[x_{a_1,\ldots,a_r}\right]^2 = \\ = \sum_{\alpha=0}^{r}\sum^{*}_{1,\ldots,r}\frac{t_1\cdots t_r}{t_{i_1}\cdots t_{i_\alpha}}\sum_{a_{i_1}}\cdots\sum_{a_{i_\alpha}}\left[m(i_1,\ldots,i_\alpha;a_{i_1},\ldots,a_{i_\alpha})\right]^2$$

We vervangen nu in (5.7) $x_{a_1,\ldots,a_r}$ door $x'_{a_1,\ldots,a_r} =$

$$= x_{a_1,\ldots,a_r} - \sum_{\alpha=0}^{r}\sum^{*}_{1,\ldots,r}\mu(i_1,\ldots,i_\alpha;a_{i_1},\ldots,a_{i_\alpha})$$

Dus:

$$x'(j_1,\ldots,j_\beta;a_{j_1},\ldots,a_{j_\beta}) = \\ = x(j_1,\ldots,j_\beta;a_{j_1},\ldots,a_{j_\beta}) - \sum_{\gamma=0}^{\beta}\sum^{*}_{j_1,\ldots,j_\beta}\mu(k_1,\ldots,k_\gamma;a_{k_1},\ldots,a_{k_\gamma})$$

Stel nu $m'(k_1,\ldots,k_\gamma;a_{k_1},\ldots,a_{k_\gamma}) =$

$$= m(k_1,\ldots,k_\gamma;a_{k_1},\ldots,a_{k_\gamma}) - \mu(k_1,\ldots,k_\gamma;a_{k_1},\ldots,a_{k_\gamma}).$$

Dan zien we dat (5.4) blijft gelden als we x door x' en m door m' vervangen.

Maar de m werd juist ondubbelzinnig bepaald door (5.4). Dus als we in (5.7) x vervangen door x', moet m door m' worden vervangen. Uit (5.7) volgt dus

$$(5.9)\quad Q = \sum_{a_1} \cdots \sum_{a_r} \left[x_{a_1,\ldots,a_r} - \sum_{\alpha=0}^{r} \sum^{*}_{1,\ldots,r} \mu(i_1,\ldots,i_\alpha; a_{i_1},\ldots,a_{i_\alpha}) \right]^2 =$$

$$= \sum_{\alpha=0}^{r} \sum^{*}_{1,\ldots,r} \frac{t_1\ldots t_r}{t_{i_1}\ldots t_{i_\alpha}} \sum_{a_{i_1}} \cdots \sum_{a_{i_\alpha}} \left[m(i_1,\ldots,i_\alpha; a_{i_1},\ldots,a_{i_\alpha}) - \mu(i_1,\ldots,i_\alpha; a_{i_1},\ldots,a_{i_\alpha}) \right]^2$$

Willen we nu Q minimaliseren onder de restricties van (5.3) en bv. onder de hypothese $H(i_1,\ldots,i_k)$: $\mu(i_1,\ldots,i_k; a_{i_1},\ldots,a_{i_k}) = 0$ voor alle $a_{i_1},\ldots,a_{i_k}$, dan wordt

$$Q_r = \frac{t_1\ldots t_r}{t_{i_1}\ldots t_{i_k}} \sum_{a_{i_1}} \cdots \sum_{a_{i_k}} \left[m(i_1,\ldots,i_k; a_{i_1}\ldots a_{i_k}) \right]^2$$

Het minimum wordt nl. bereikt door in (5.9) als schatting voor de $\mu(i_1,\ldots,i_\alpha; a_{i_1},\ldots,a_{i_\alpha})$ te nemen $m(i_1,\ldots,i_\alpha; a_{i_1},\ldots,a_{i_\alpha})$ voor alle μ's die niet nul zijn volgens de hypothese, want deze m's voldoen aan de restricties uit (5.3) zoals we lieten zien.

We zien dus dat als zekere μ's volgens de hypothese gelijk zijn aan nul, de schattingen voor de andere μ's dezelfde blijven als in het geval dat alleen de restricties (5.3) gelden.

Indien een stelsel interacties $\mu(i_1,\ldots,i_\alpha; a_{i_1},\ldots,a_{i_\alpha})$ ondersteld wordt gelijk te zijn aan nul voor iedere $a_{i_1},\ldots,a_{i_\alpha}$, noemen we dit interacties van het type I. Andere interacties $\mu(i_1,\ldots,i_\alpha; a_{i_1},\ldots,a_{i_\alpha})$ zijn onbekend voor alle $a_{i_1},\ldots,a_{i_\alpha}$; dit noemen we interacties van het type II.

Indien we hypothese willen toetsen betreffende enkele individuele waarden van zo'n stelsel interacties, dan zijn dit interacties van het type III.

Uit vergelijking (5.9) zien we, dat we om Q_r en Q_α te vinden $\mu(i_1,\ldots,i_\alpha; a_{i_1},\ldots,a_{i_\alpha}) = 0$ moeten stellen voor interacties van het type I en $\hat{\mu}(i_1,\ldots,i_\alpha; a_{i_1},\ldots,a_{i_\alpha}) = m(i_1,\ldots,i_\alpha; a_{i_1},\ldots,a_{i_\alpha})$ voor interacties van het type II.

Voor interacties van het type III moeten we b.v.

$$\sum_{a_1} \cdots \sum_{a_k} \left[m(j_1, \ldots, j_k; a_{j_1}, \ldots, a_{j_k}) - \mu(j_1, \ldots, j_k; a_{j_1}, \ldots, a_{j_k}) \right]^2$$

minimaliseren onder de voorwaarden

$$\sum_{a_{j_i}} \mu(j_1, \ldots, j_k; a_{j_1}, \ldots, a_{j_k}) = 0 \quad (i = 1, \ldots, k).$$

en zekere andere voorwaarden opgelegd door onderstelling en hypothese.

Stel b.v. we hebben een drievoudige classificatie en we onderstellen $\mu(1,2,3; a_1, a_2, a_3) = 0$ voor alle a_1, a_2 en a_3. We willen de hypothese toetsen dat $\mu(1,2; 1,1) = \mu(1,2; 1,2)$. Om Q_α te vinden stellen we alle

$$\mu(i_1, \ldots, i_\alpha; a_{i_1}, \ldots, a_{i_\alpha}) = m(i_1, \ldots, i_\alpha; a_{i_1}, \ldots, a_{i_\alpha})$$

voor $\alpha \leq 2$ (interacties van het type II). De interacties van de tweede orde zijn van het type I, dus

$$Q_\alpha = \sum_{a_1} \sum_{a_2} \sum_{a_3} \left[m(1,2,3; a_1, a_2, a_3) \right]^2$$

Om Q_r te vinden moeten we minimaliseren

$$Q_\alpha + t_3 \sum_{a_1} \sum_{a_2} \left[m(1,2; a_1, a_2) - \mu(1,2; a_1, a_2) \right]^2$$

onder de voorwaarden $\sum_{a_1} \mu(1,2; a_1, a_2) = \sum_{a_2} \mu(1,2; a_1, a_2) = 0$
en $\mu(1,2; 1,1) = \mu(1,2; 1,2)$.

Het is nu evenwel eenvoudiger om het gevolg uit stelling 4.3. toe te passen op de lineaire vorm

$$\mu(1,2; 1,1) - \mu(1,2; 1,2).$$

We moeten dan de variantie vinden van

$$m(1,2; 1,1) - m(1,2; 1,2) =$$

$$\left[x(1,2; 1,1) - x(1; 1) - x(2; 1) + x \right] -$$

$$\left[x(1,2; 1,2) - x(1; 1) - x(2; 2) + x \right] =$$

$$x(1,2; 1,1) - x(1,2; 1,2) + x(2; 2) - x(2; 1).$$

Deze is gelijk aan:

$$\sigma^2_{x(1,2;1,1)} + \sigma^2_{x(1,2;1,2)} + \sigma^2_{x(2;2)} + \sigma^2_{x(2;1)} - 2\sigma_{x(1,2;1,1)\,x(2;1)} - 2\sigma_{x(1,2;1,2)\,x(2;2)} =$$

$$= \frac{1}{t_3}\sigma^2 + \frac{1}{t_3}\sigma^2 + \frac{1}{t_1 t_3}\sigma^2 + \frac{1}{t_1 t_3}\sigma^2 - \frac{2}{t_1 t_3}\sigma^2 - \frac{2}{t_1 t_3}\sigma^2 =$$

$$= 2\sigma^2\left(\frac{1}{t_3} - \frac{1}{t_1 t_3}\right) = 2\sigma^2\,\frac{t_1 - 1}{t_1 t_3}\,.$$

Dus $$Q_2 - Q_\alpha = \frac{t_1 t_3}{2(t_1-1)}\left[m(1,2;1,1) - m(1,2;1,2)\right]^2$$

De toetsingsgrootheid is dus:

$$F = \frac{(t_1-1)(t_2-1)(t_3-1)}{1}\;\frac{t_1 t_3}{2(t_1-1)}\;\frac{\left[m(1,2;1,1) - m(1,2;1,2)\right]^2}{\sum\limits_{a_1}\sum\limits_{a_2}\sum\limits_{a_3}\left[m(1,2,3;a_1,a_2,a_3)\right]^2}$$

Om het berekenen van de voorkomende sommen van kwadraten te vergemakkelijken kunnen we gebruik maken van de formule

$$\sum_{a_{i_1}}\cdots\sum_{a_{i_k}}\left[m(i_1,\ldots,i_k;a_{i_1},\ldots,a_{i_k})\right]^2 =$$

(5.11) $$= \sum_{\alpha=0}^{k}(-)^{k-\alpha}\sum^{*}_{i_1,\ldots,i_k}\;\sum_{a_{j_1}}\cdots\sum_{a_{j_\alpha}}\frac{t_{i_1}\cdots t_{i_k}}{t_{j_1}\cdots t_{j_\alpha}}\left[x(j_1,\ldots,j_\alpha;a_{j_1},\ldots,a_{j_\alpha})\right]^2$$

B.v.

$$\sum_{a_1}\sum_{a_2}\sum_{a_3}\left[m(1,2,3;a_1,a_2,a_3)\right]^2 = \sum_{a_1}\sum_{a_2}\sum_{a_3}\left[x(1,2,3;a_1,a_2,a_3)\right]^2 - t_1\sum_{a_2}\sum_{a_3}\left[x(2,3;a_2,a_3)\right]^2 - \cdots + t_1 t_2\sum_{a_3}\left[x(3;a_3)\right]^2 - t_1 t_2 t_3\, x^2$$

Dit laatste is gemakkelijk te verifiëren. Formule (5.11) bewijst men met behulp van het principe van volledige inductie.

Een belangrijk speciaal geval doet zich voor wanneer we bij een r-voudige classificatie niet één, maar meerdere waarnemingen van iedere combinatie doen. Dit kunnen we opvatten als een $(r+1)$-voudig schema, waarbij we aannemen dat de $(r+1)^e$ classificatie het gemiddelde niet beïnvloedt.

Dus $\mu(i_1,\dots,i_\alpha,r+1;a_{i_1},\dots,a_{i_\alpha},a_{r+1})=0$ voor alle $i_1<\dots<i_\alpha<r+1$ en alle $a_{i_1},\dots,a_{i_\alpha},a_{r+1}$. Volgens (5.9) wordt Q_α dan

$$Q_\alpha = \sum_{\alpha=0}^{r}\sum^{*}_{1,\dots,r}\frac{t_1\cdots t_r\cdot t_{r+1}}{t_{j_1}\cdots t_{j_\alpha}\,t_{r+1}}\sum_{a_{j_1}}\cdots\sum_{a_{j_\alpha}}\sum_{a_{r+1}}\left[m(j_1,\dots,j_\alpha,r+1;a_{j_1},\dots,a_{j_\alpha},a_{r+1})\right]^2 =$$

(volgens 5.10)

$$= \sum_{a_1}\cdots\sum_{a_{r+1}}\left[x(1,\dots,r+1;a_1,\dots,a_{r+1})\right]^2 - t_{r+1}\sum_{\alpha=0}^{r}\sum^{*}_{1,\dots,r}\frac{t_1\cdots t_r}{t_{j_1}\cdots t_{j_\alpha}}\sum_{a_{j_1}}\cdots\sum_{a_{j_\alpha}}\left[m(j_1,\dots,j_\alpha;a_{j_1},\dots,a_{j_\alpha})\right]^2 =$$

(volgens 5.10)

$$= \sum_{a_1}\cdots\sum_{a_{r+1}}\left[x(1,\dots,r+1;a_1,\dots,a_{r+1})\right]^2$$

$$\text{(5.14)}\qquad - t_{r+1}\sum_{a_1}\cdots\sum_{a_r}\left[x(1,\dots,r;a_1,\dots,a_r)\right]^2 =$$

$$\sum_{a_1}\cdots\sum_{a_{r+1}}\left[x(1,\dots,r+1;a_1,\dots,a_{r+1})-x(1,\dots,r;a_1,\dots,a_r)\right]^2$$

Deze formule is gemakkelijk te interpreteren. Daar de verdeling van $x_{a_1,\dots,a_{r+1}}$ onafhankelijk is van a_{r+1}, staat hier een schatting van de spreiding met $t_1\cdots t_r(t_{r+1}-1)$ vrijheidsgraden.

Dit is ook het aantal dat we volgens onze theorie vinden. Het aantal stochastische grootheden N is nl. $t_1\cdots t_{r+1}$ en het aantal onafhankelijke parameters p dat overblijft na onze onderstelling is juist dat van een r-voudige classificatie en dus $t_1\cdots t_r$. Dus $N-p=t_1\cdots t_r(t_{r+1}-1)$.

Een eenvoudig voorbeeld is het volgende: We hebben 2 geneesmiddelen A en B, die 's morgens of 's avonds kunnen worden toegediend. De uitwerking van een geneesmiddel kunnen we meten. 4n patiënten worden in 4 groepen van n verdeeld. We hebben dan het volgende schema:

$$\text{morgen}\begin{cases} x_1, \ldots, x_n & \text{(A)} \\ y_1, \ldots, y_n & \text{(B)} \end{cases}$$

$$\text{avond}\begin{cases} x'_1, \ldots, x'_n & \text{(A)} \\ y'_1, \ldots, y'_n & \text{(B)} \end{cases}$$

Hierin stellen $x_1, \ldots, x_n$ de uitwerkingen voor die het geneesmiddel A heeft 's morgens toegediend aan één van de groepen van n patiënten enz. We kunnen dit opvatten als een tweevoudige classificatie met n herhalingen. In de notatie van het voorafgaande is dus $a_1 = 1,2$ (A of B), $a_2 = 1,2$ (morgen of avond) en $a_3 = 1, \ldots, n$. Het model is

$$H_1 : \ \mathcal{E}\, \underline{x}_{a_1,a_2,a_3} = \mu(1,2;a_1,a_2) + \mu(1;a_1) + \mu(2;a_2) + \mu ,$$

met als bijvoorwaarden

$$\sum_{a_1=1}^{2} \mu(1,2;a_1,a_2) = \sum_{a_2=1}^{2} \mu(1,2;a_1,a_2) =$$

$$= \sum_{a_1=1}^{2} \mu(1;a_1) = \sum_{a_2=1}^{2} \mu(2;a_2) = 0$$

Volgens vergelijking (5.14) is

$$Q_\alpha = \sum_{a_1,a_2,a_3} \left[x_{a_1,a_2,a_3} - x(1,2;a_1,a_2) \right]^2 ,$$

$$x(1,2;1,1) = 1/n \sum_i x_i = \bar{x} ,$$

$$x(1,2;2,1) = 1/n \sum_i y_i = \bar{y} ,$$

$$x(1,2;1,2) = 1/n \sum_i x'_i = \bar{x}' ,$$

$$x(1,2;2,2) = 1/n \sum_i y'_i = \bar{y}' .$$

We vinden dus

$$Q_\alpha = \sum_i (x_i - \bar{x})^2 + \sum_i (y_i - \bar{y})^2 + \sum_i (x'_i - \bar{x}')^2 + \sum_i (y'_i - \bar{y}')^2$$

$$\overset{\text{def}}{=} S_1^2 + S_2^2 + S_3^2 + S_4^2 .$$

Q_α heeft $2.2.(n-1) = 4(n-1)$ vrijheidsgraden.

Stel $\mu(1,2;1,1) = -\mu(1,2;1,2) = \mu(1,2;2,2) = -\mu(1,2;2,1) = a,$

$\mu(1;1) = -\mu(1;2) = b$ en
$\mu(2;1) = -\mu(2;2) = c$

In het volgende schema vinden we de verwachtingen van de x_{a_1,a_2,a_3}. De derde classificatie heeft geen invloed en hoeven we daarom dus niet in aanmerking te nemen.

	A	B
morgen	$a+b+c+\mu$	$-a-b+c+\mu$
avond	$-a+b-c+\mu$	$a-b-c+\mu$

We willen toetsen of de geneesmiddelen A en B gelijkwaardig zijn. Dan moeten zowel 's morgens als 's middags de verwachtingen gelijk zijn. Dus moet gelden

$$a+b+c+\mu = -a-b+c+\mu$$
$$-a+b-c+\mu = a-b-c+\mu$$

of

$$a = 0 \text{ en } b = 0$$

De hypothese is dus

$H_2 : H_1 \,\&\, \mu(1,2;a_1,a_2) = \mu(1;a_1) = 0$ voor iedere a_1 en a_2

$$Q_r - Q_\alpha = \sum_{a_1,a_2,a_3} \left(Y^{(1)}_{a_1,a_2,a_3} - Y^{(2)}_{a_1,a_2,a_3}\right)^2 =$$
$$= t_3 \sum_{a_1,a_2} \left[m(1,2;a_1,a_2) + m(1;a_1)\right]^2 = t_3 \sum_{a_1,a_2} \left[x(1,2;a_1,a_2) - x(2;a_2)\right]^2.$$

Na enige herleiding vinden we

$$Q_r - Q_\alpha = \tfrac{1}{2} n \left[(\bar{x}-\bar{y})^2 + (\bar{x}'-\bar{y}')^2\right].$$

De toetsingsgrootheid is dus

$$F_2 = \frac{4(n-1)}{2} \; \frac{\tfrac{1}{2} n \left[(\bar{x}-\bar{y})^2 + (\bar{x}'-\bar{y}')^2\right]}{S_1^2 + S_2^2 + S_3^2 + S_4^2}.$$

Willen we toetsen $H_3 : H_1 \,\&\, \mu(1,2;a_1,a_2) = 0$ dan vinden we met behulp van (5.9)

$$Q_r - Q_\alpha = \tfrac{1}{4} n \left[(\bar{x}-\bar{y}) - (\bar{x}'-\bar{y}')\right]^2$$

F_3 wordt dus

$$F_3 = \frac{4(n-1)}{1} \; \frac{\tfrac{1}{4} n \left[(\bar{x}-\bar{y}) - (\bar{x}'-\bar{y}')\right]^2}{S_1^2 + S_2^2 + S_3^2 + S_4^2}.$$

Nemen we in het model op: $\mu(1;2;a_1,a_2) = 0$, dan wordt de toegelaten hypothese de nulhypothese van de vorige toets. De nieuwe Q_α wordt dus gelijk aan de Q_r in het vorige geval. Dus

$$Q_\alpha = S_1^2 + S_2^2 + S_3^2 + S_4^2 + \tfrac{1}{4} n \left[(\bar{x}-\bar{y}) - (\bar{x}'-\bar{y}')\right]^2$$

We kunnen nu toetsen

$$H_4 : H_3 \,\&\, \mu(1;a_1) = 0.$$

Uit (5.9) volgt:

$$Q_r - Q_\alpha = t_2 t_3 \left[m(1;a_1)\right]^2 = \tfrac{1}{4} n \left[(\bar{x}-\bar{y}) + (\bar{x}'-\bar{y}')\right]^2$$

dus

$$F_4 = \frac{4(n-1)+1}{1} \quad \frac{\tfrac{1}{4} n \left[(\bar{x}-\bar{y}) + (\bar{x}'-\bar{y}')\right]^2}{S_1^2 + S_2^2 + S_3^2 + S_4^2 + \tfrac{1}{4} n \left[(\bar{x}-\bar{y}) - (\bar{x}'-\bar{y}')\right]^2}$$

Gaan we ten slotte weer uit van het oorspronkelijke model en willen we toetsen $H_5 : H_1 \,\&\, \mu(1;a_1) = 0$, dan vinden we voor onze toetsingsgrootheid

$$F_5 = \frac{4(n-1)}{1} \quad \frac{\tfrac{1}{4} n \left[(\bar{x}-\bar{y}) + (\bar{x}'-\bar{y}')\right]^2}{S_1^2 + S_2^2 + S_3^2 + S_4^2}.$$

De laatste drie toetsen kunnen we ook vervangen door Student-toetsen door de wortel uit de desbetreffende toetsingsgrootheid te nemen.

Hoofdstuk VII

Latijnse vierkanten en evenwichtige schema's met onvolledige blokken.

Stel dat we m tarwevariëteiten willen vergelijken met betrekking tot hun gemiddelde opbrengst op een bepaald type grond. We hebben daartoe tot onze beschikking een rechthoekig veld, verdeeld in m^2 vakken. We spreken nu van een Latijns vierkant, indien iedere variëteit één keer in iedere rij en één keer in iedere kolom voorkomt. De opbrengst van het vak in de i^e rij en de j^e kolom, dat de k^e variëteit bevat geven we aan met y_{ijk} . We onderstellen dat

$$\text{(7.1)} \quad \begin{aligned} &\mathcal{E} y_{ijk} = \mu_i + \nu_j + \rho_k + \rho\,, \\ &\sum_i \mu_i = \sum_j \nu_j = \sum_k \rho_k = 0. \end{aligned}$$

Deze onderstelling is b.v. steeds vervuld als de vruchtbaarheid een lineaire functie is van de coördinaten. Verder wordt ondersteld dat de y_{ijk} onderling onafhankelijk en normaal verdeeld zijn met dezelfde spreiding σ .

Om Q_α te vinden moeten we minimaliseren

$$Q = \sum_i \sum_j (y_{ijk} - \mu_i - \nu_j - \rho_k - \rho)^2$$

onder de bijvoorwaarden van (7.1). Volgens stelling 4.4 hoeven we met deze bijvoorwaarden evenwel geen rekening te houden. Dit levert als schattingen

$\hat{\mu}_i$, $\hat{\nu}_j$, $\hat{\rho}_k$ en $\hat{\rho}$ van μ_i, ν_j, ρ_k en ρ op

$$\hat{\rho} = \frac{1}{m^2}\sum_i \sum_j y_{ijk} = y_{...},$$

(7.2)
$$\hat{\mu}_i = y_{i..} - y_{...},$$
$$\hat{\nu}_j = y_{.j.} - y_{...},$$
$$\hat{\rho}_k = y_{..k} - y_{...}.$$

Daar we te maken hebben met een Latijns vierkant, komen in de termen van $\frac{\partial Q}{\partial \mu_i}$ die μ_i bevatten alle ν_j en ρ_k juist één keer voor, zodat in de som de kolom- en variëteitseffecten wegvallen. Dus

(7.3)
$$Q_\alpha = \sum_i \sum_j (y_{ijk} - y_{i..} - y_{.j.} - y_{..k} + 2y_{...})^2.$$

Q_α heeft $m^2 - [3(m-1)+1] = m^2-3m+2 = (m-1)(m-2)$ vrijheidsgraden. Door stelling 4.2 toe te passen op het stelsel hypothesen

H_1 : de onderstelling van (7.1),
H_2 : H_1 & $\mu_i = 0$ $(i=1,\ldots,m)$,
H_3 : H_2 & $\nu_j = 0$ $(j=1,\ldots,m)$,
H_4 : H_3 & $\rho_k = 0$ $(k=1,\ldots,m)$,

vinden we

$$\sum_{i,j} y_{ijk}^2 = \sum_{i,j}(y_{ijk} - y_{i..} - y_{.j.} - y_{..k} + 2y_{...})^2 + m\sum_i (y_{i..} - y_{...})^2 + m\sum_j (y_{.j.} - y_{...})^2 + m\sum_k (y_{..k} - y_{...})^2 + m^2 y_{...}^2 .$$

We vinden dezelfde ontbinding als we de hypothesen H_2, H_3 en H_4 anders nummeren. Dus de hypothese H_2' : H_1 & $\rho_k = 0$ wordt getoetst door

$$F = \frac{m^2-3m+2}{m-1} \; \frac{m\sum_k y_{..k}^2 - m^2 y_{...}^2}{\sum_i \sum_j y_{ijk}^2 - m\left[\sum_i y_{i..}^2 + \sum_j y_{.j.}^2 + \sum_k y_{..k}^2\right] + 2m^2 y_{...}^2}$$

Voor de berekening van de verschillende toetsingsgrootheden gebruikt men vaak het volgende schema:

	vrijheidsgraden	sommen van kwadraten	gem. s. v. kw	F
rijen	$m-1$	$Q_1 = m\sum_i y_{i..}^2 - m^2 y_{...}^2$	$Q_1/m-1$	$\frac{(m-1)(m-2)}{m-1} \frac{Q_1}{Q_\alpha}$
kolommen	$m-1$	$Q_2 = m\sum_j y_{.j.}^2 - m^2 y_{...}^2$	$Q_2/m-1$	$\frac{(m-1)(m-2)}{(m-1)} \frac{Q_2}{Q_\alpha}$
variëteiten	$m-1$	$Q_3 = m\sum_k y_{..k}^2 - m^2 y_{...}^2$	$Q_3/m-1$	$\frac{(m-1)(m-2)}{m-1} \frac{Q_3}{Q_\alpha}$
Q_α	$(m-1)(m-2)$	$Q - Q_1 - Q_2 - Q_3$	$Q_\alpha/(m-1)(m-2)$	
totaal	m^2-1	$Q = \sum_{i,j} y_{ijk}^2 - m^2 y_{...}^2$		

Als we b.v. willen toetsen $\mu_i = 0$ en $\rho_k = 0$ voor iedere i en k, dan tellen we de grootheden F van de 1^e^ en de 2^e^ rij van het schema op en delen de som door 2. De grootheid die we dan krijgen heeft onder de te toetsen hypothese de F -verdeling met $2(m-1)$ en $(m-1)(m-2)$ vrijheidsgraden.

De teller van de toetsingsgrootheid is nl. $Q_1 + Q_2$ zoals eenvoudig is na te gaan en deze heeft $2(m-1)$ vrijheidsgraden.

Worden m variëteiten uitgezaaid op r verschillende velden, die alle onderverdeeld zijn als een Latijns vierkant, dan hebben we het volgende model:

$$\mathcal{E}\, y_{ijk}^{\ell} = \mu_i^{\ell} + \nu_j^{\ell} + \rho_k + \alpha_\ell + \mu \quad (i,j,k=1,\ldots,m;\ \ell=1,\ldots,r)$$

(7.7)

$$\sum_i \mu_i^{\ell} = \sum_j \nu_j^{\ell} = \sum_k \rho_k = \sum_\ell \alpha_\ell = 0$$

De index ℓ geeft het Latijnse vierkant, i de rij, j de kolom en k de variëteit aan. De schattingen voor de parameters onder de onderstellingen (7.7) worden:

$$\hat{\mu} = y^{\cdot}_{\cdots}\ ,$$
$$\hat{\alpha}_\ell = y^{\ell}_{\cdots} - y^{\cdot}_{\cdots}\ ,$$
$$\hat{\rho}_k = y^{\cdot}_{\cdot\cdot k} - y^{\cdot}_{\cdots}\ ,$$
$$\hat{\nu}_j^{\ell} = y^{\ell}_{\cdot j \cdot} - y^{\ell}_{\cdots}\ ,$$
$$\hat{\mu}_i^{\ell} = y^{\ell}_{i\cdot\cdot} - y^{\ell}_{\cdots}\ .$$

Door stelling (4.2) toe te passen op het stelsel hypothesen

H_1 : onderstelling (7.7).

$H_2 : H_1\ \&\ \mu_i^{\ell} = 0 \quad (i=1,\ldots,m;\ \ell=1,\ldots,r)$,

$H_3 : H_2\ \&\ \nu_j^{\ell} = 0 \quad (j=1,\ldots,m;\ \ell=1,\ldots,r)$,

$H_4 : H_3\ \&\ \rho_k = 0 \quad (k=1,\ldots,m)$,

$H_5 : H_4\ \&\ \alpha_\ell = 0 \quad (\ell=1,\ldots,r)$,

vinden we

$$\sum_{i,j,\ell} (y_{ijk}^{\ell})^2 = Q_\alpha + m\sum_{i,\ell}(y^{\ell}_{i\cdot\cdot} - y^{\ell}_{\cdots})^2 + m\sum_{j,\ell}(y^{\ell}_{\cdot j\cdot} - y^{\ell}_{\cdots})^2 +$$
$$+ mr\sum_k (y^{\cdot}_{\cdot\cdot k} - y^{\cdot}_{\cdots})^2 + m^2\sum_\ell (y^{\ell}_{\cdots} - y^{\cdot}_{\cdots})^2 + m^2 r (y^{\cdot}_{\cdots})^2 .$$

We verkrijgen weer dezelfde splitsing als we in plaats van $H_1,\ldots,H_5$ een ander stelsel nemen met b.v. $H_2' : H_1\ \&\ \rho_k = 0$. De toetsingsgrootheid voor deze hypothese is dus

$$F = \frac{(m-1)(rm-r-1)}{m-1}\ \frac{mr\sum_k (y^{\cdot}_{\cdot\cdot k} - y^{\cdot}_{\cdots})^2}{Q_\alpha}\ ,$$

waarin we Q_α vinden uit de bovenstaande splitsing van $\sum (y_{ijk}^{\ell})^2$.

Vaak kunnen we nog tegelijkertijd andere factoren toetsen, die invloed hebben op de opbrengst, bijv. kunnen we m verschillende meststoffen toedienen en wel zo dat het proefveld een Latijns vierkant vormt voor variëteiten en meststoffen en dat iedere combinatie van variëteit met meststof juist ééns voorkomt. We zeggen dan dat de beide Latijnse vierkanten orthogonaal zijn. Dit kunnen we generaliseren tot het geval van r factoren, waarbij ieder tweetal van de r betreffende Latijnse vierkanten orthogonaal is.

De toetsingsgrootheden vinden we door een directe toepassing van stelling 4.1, volgens een methodie die geheel analoog is aan die bij één Latijns vierkant. De ondersteling die we hierbij moeten maken is, dat er geen interactie bestaat tussen de factoren.

Indien het aantal variëteiten groot is wordt het Latijns vierkant zo groot dat de onderstelling (7.1) niet meer zal zijn vervuld. Hieraan wordt tegemoetgekomen door het gebruik van de schema's met onvolledige blokken. Door kleine blokken te nemen bereiken we, dat we kunnen aannemen, dat de vruchtbaarheid constant is in zo'n blok.

We noemen een schema van v variëteiten, verdeeld over b blokken van k velden (plots), een evenwichtig schema met onvolledige blokken (balanced incomplete block design), indien het aan de volgende voorwaarden voldoet:

1) Geen enkel blok bevat dezelfde variëteit meer dan één keer;
2) Iedere variëteit komt r keer voor;
3) Iedere variëteit v_i komt met iedere andere variëteit v_j λ keer samen in één blok voor.

Het totale aantal velden is bk, maar ook rv, dus

(7.12) $$b.k = r.v.$$

Iedere variëteit v_j komt voor in r verschillende blokken. Ieder van deze blokken bevat nog $(k-1)$ andere variëteiten. Dus het totale aantal keren, dat v_j met een andere variëteit in een blok voorkomt is $r(k-1)$. Eveneens moet v_j met iedere der $(v-1)$ andere variëteiten λ in een blok voorkomen, dus

(7.13) $$r(k-1) = \lambda(v-1).$$

Verder is nog bewezen, dat, als $k<v$ is, moet gelden:

(7.14) $$b \geq v.$$

Deze 3 voorwaarden zijn evenwel nog niet voldoende. Van het schema $v=b=43$, $r=k=7$, $\lambda=1$ is b.v. bekend dat het onbestaanbaar is.

We onderstellen dat de verwachting van de opbrengst een lineaire functie is van variëteit- en blokinvloed:

(7.8) $$\mathcal{E}\, y_{ij} = v_i + b_j + \mu, \quad \sum_{i=1}^{v} v_i = \sum_{j=1}^{b} b_j = 0.$$

Het minimaliseren van

$$Q = \sum_{i,j} (y_{ij} - v_i - b_j - \mu)^2,$$

leidt tot de vergelijkingen

$$(7.15) \qquad \begin{cases} \sum_{i,j} y_{ij} = r v \hat{\mu}, \\ V_i = r \hat{v}_i + \sum_j^{(i)} \hat{b}_j + r \hat{\mu}, \\ B_j = \sum_i^{(j)} \hat{v}_i + k \hat{b}_j + k \hat{\mu}. \end{cases}$$

Hierin stellen $\hat{\mu}$, $\hat{b}_j$ en $\hat{v}_i$ de meest aannemelijke schattingen van μ, b_j en v_i voor. V_i is de som van de opbrengsten van de i^e variëteit, B_j is de som van de opbrengsten in het j^e blok, $\sum_j^{(i)}$ betekent dat gesommeerd wordt over die blokken, die de i^e variëteit bevatten. $\sum_i^{(j)}$ betekent dat we sommeren over de variëteiten, die voorkomen in het j^e blok.
We stellen $T_i = \sum_j^{(i)} B_j$ en sommeren de derde vergelijking in (7.15) over alle blokken die de i^e variëteit bevatten; dit levert

$$(7.16) \qquad T_i = (r-\lambda)\hat{v}_i + k \sum_j^{(i)} \hat{b}_j + r k \hat{\mu}.$$

Iedere variëteit $v_j \neq v_i$ komt in de som nl. λ keer voor en v_i zelf r keer, terwijl $\sum_{j \neq i} \hat{v}_j = -\hat{v}_i$. Door de tweede vergelijking van (7.15) met k te vermenigvuldigen vinden we:

$$(7.17) \qquad k V_i = r k \hat{v}_i + k \sum_j^{(i)} \hat{b}_j + r k \hat{\mu}.$$

We trekken nu (7.16) af van (7.17); dit geeft

$$(7.18) \qquad k V_i - T_i = (rk - r + \lambda)\hat{v}_i.$$

Volgens (7.12) en (7.13) is

$$(rk - r + \lambda) = r(k-1) + \lambda = \lambda(v-1) + \lambda = \lambda v.$$

Dus.

$$(7.19) \qquad \begin{cases} \hat{\mu} = y_{..} \\ \hat{v}_i = \frac{1}{\lambda v}(k V_i - T_i) \\ \hat{b}_j = \frac{1}{k} B_j - \frac{1}{\lambda k v} \sum_i^{(j)} (k V_i - T_i) - y_{..} \end{cases}$$

We passen nu stelling (4.2) toe op het stelsel hypothesen

$$H_1 : \mathcal{E} y_{ij} = v_i + b_j + \mu,$$
$$H_2 : H_1 \ \& \ v_i = 0 \quad (i = 1, \ldots, v).$$

Dus
$$Y_{ij}^{(1)} = \frac{1}{k} B_j + (\hat{v}_i - 1/k \sum_i^{(j)} \hat{v}_i),$$
$$Y_{ij}^{(2)} = \frac{1}{k} B_j .$$

Dus

$$(7.21)\quad \begin{cases} \sum\limits_{i,j} y_{ij}^2 = Q_\alpha + \sum\limits_{i,j} \left(\hat{v}_i - \frac{1}{k}\sum\limits_i^{(j)} \hat{v}_i\right)^2 + \frac{1}{k}\sum\limits_j B_j^2 , \\ Q_r - Q_\alpha = \sum\limits_{i,j} \left(Y_{ij}^{(1)} - Y_{ij}^{(2)}\right)^2 = \sum\limits_{i,j} \left(\hat{v}_i - \frac{1}{k}\sum\limits_i^{(j)} \hat{v}_i\right)^2 . \end{cases}$$

$Q_r - Q_\alpha$ is dus een symmetrische vorm in de $\hat{v}_i$. Volgens stelling 4.5 is dus

$$Q_r - Q_\alpha = \frac{v-1}{vc}\sum_{i=1}^{v} \hat{v}_i^2 , \quad \text{waarin} \quad c = \frac{\sigma^2_{\hat{v}_i}}{\sigma^2}$$

Nu is $\hat{v}_i = \frac{1}{\lambda v}\left(kV_i - T_i\right) = \frac{1}{\lambda v}\left[(k-1)V_i - (T_i - V_i)\right]$

$(k-1)V_i$ en $T_i - V_i$ zijn onafhankelijk, dus

$$\frac{\sigma^2_{\hat{v}_i}}{\sigma^2} = \frac{1}{\lambda^2 v^2}\left[(k-1)^2 r + r(k-1)\right] = \frac{rk(k-1)}{\lambda^2 v^2} = \frac{k(v-1)}{\lambda v^2} .$$

Dus $Q_r - Q_\alpha = \frac{v-1}{v}\,\frac{\lambda v^2}{k(v-1)} \sum\limits_i \hat{v}_i^2 = \frac{\lambda v}{k}\sum\limits_i \hat{v}_i^2$ en volgens (7.21)

$$(7.22)\quad Q_\alpha = \sum_{i,j} y_{ij}^2 - \frac{1}{k}\sum_j B_j^2 - \frac{\lambda v}{k}\sum_i \hat{v}_i^2 .$$

De toetsingsgrootheid voor de hypothese H_2 is dus

$$F = \frac{rv - b - v + 1}{v-1} \; \frac{\lambda v / k \sum\limits_i \hat{v}_i^2}{\sum\limits_{i,j} y_{ij}^2 - \frac{1}{k}\sum\limits_j B_j^2 - \frac{\lambda v}{k}\sum\limits_i \hat{v}_i^2} ,$$

waarin $\hat{v}_i = \frac{1}{\lambda v}\left(kV_i - T_i\right)$

We hadden onze afleiding ook kunnen geven zonder stelling 4.5 te gebruiken. We hadden nl.

$$Q_r - Q_\alpha = \sum_{i,j}\left(\hat{v}_i - \frac{1}{k}\sum_i^{(j)} \hat{v}_i\right)^2 =$$

$$\sum_i \left[r\hat{v}_i^2 - \frac{2r}{k}\hat{v}_i^2 - \frac{2\lambda}{k}\hat{v}_i \sum_{j \neq i} \hat{v}_j + \frac{1}{k^2}\sum_j^{(i)}\left(\sum_i^{(j)} \hat{v}_i\right)^2 \right] =$$

$$\sum_i \left[\left(r - \frac{2r}{k} - \frac{2\lambda}{k}\right)\hat{v}_i^2 + \frac{1}{k^2}\sum_j^{(i)}\left(\sum_i^{(j)} \hat{v}_i\right)^2 \right] =$$

$$\sum_i \hat{v}_i^2 \left[r - \frac{2(r-\lambda)}{k} + \frac{k}{k^2}(r-\lambda) \right] = \frac{rk - r + \lambda}{k}\sum_i \hat{v}_i^2 =$$

$\frac{\lambda v}{k}\sum\limits_i \hat{v}_i^2$ Er geldt nl.

$$\sum_i \left[\sum_j^{(i)}\left(\sum_i^{(j)} \hat{v}_i\right)^2 \right] = k\sum_j \left(\sum_i^{(j)} \hat{v}_i\right)^2 = k(r-\lambda)\sum_i \hat{v}_i^2 ,$$

want alle $\hat{v}_i^2$ komen r keer voor in $\sum\limits_j\left(\sum\limits_i^{(j)} \hat{v}_i\right)^2$ en alle dubbele producten $\hat{v}_i\hat{v}_j$ komen λ keer voor, dus

$$\sum_j \left(\sum_i^{(j)} \hat{v}_i\right)^2 = \lambda\left(\sum_i \hat{v}_i\right)^2 + (r-\lambda)\sum_i \hat{v}_i^2 = (r-\lambda)\sum_i \hat{v}_i^2 .$$

Willen we toetsen H_2' : H_1 & $b_j = 0$ $(j=1,\dots,b)$, dan moeten we de schattingen van v_i en μ onder H_2' bepalen.
Deze zijn:

$$\hat{\mu} = y_{..}$$
$$\hat{v}_i = \frac{1}{r} V_i - y_{..}$$

Dus $Q_r = \sum_{i,j} (y_{ij} - \frac{1}{r} V_i)^2 = \sum_{i,j} y_{ij}^2 - \frac{1}{r} \sum_i V_i^2$ en

$$Q_r - Q_\alpha = \frac{1}{k} \sum_j B_j^2 - \frac{1}{r} \sum_i V_i^2 + \frac{\lambda v}{k} \sum_i \hat{v}_i^2,$$

waarin $\hat{v}_i = \frac{1}{\lambda v} (k V_i - T_i)$.

Om de hypothese $v_i = v_j$ te kunnen toetsen moeten we de variantie kennen van $\hat{v}_i - \hat{v}_j$. Er geldt

(7.23) $$\sigma^2_{\hat{v}_i - \hat{v}_j} = \sigma^2_{\hat{v}_i} + \sigma^2_{\hat{v}_j} - 2\sigma_{\hat{v}_i \hat{v}_j}$$

Nu worden $\hat{v}_i$ en $\hat{v}_j$ gegeven door (7.19). V_i en V_j hebben geen termen gemeen. V_i en T_j, zowel als V_j en T_i hebben λ termen gemeen en T_i en T_j $k\lambda$ termen. Dus:

$$\sigma_{\hat{v}_i \hat{v}_j} = \frac{1}{\lambda^2 v^2}\left(-k\lambda - k\lambda + k\lambda\right)\sigma^2 = \frac{-k}{\lambda v^2}\sigma^2$$

Dus, daar $\sigma^2_{\hat{v}_i} = \frac{k(v-1)}{\lambda v^2}\sigma^2$ (zie pag. 44), is

$$\sigma^2_{\hat{v}_i - \hat{v}_j} = \left(\frac{2k(v-1)}{\lambda v^2} + \frac{2k}{\lambda v^2}\right)\sigma^2 = \frac{2k}{\lambda v}\sigma^2.$$

Met het gevolg uit stelling 4.3 vinden we dan voor de toetsingsgrootheid:

$$F = \frac{rv - b - v - 1}{1} \; \frac{\lambda v}{2k} \; \frac{(\hat{v}_i - \hat{v}_j)^2}{Q_\alpha}.$$

De doeltreffendheidsfactor (efficiency factor) van het schema met betrekking tot de schatting $\hat{v}_i$ definiëren we als volgt. Stel dat we het experiment uit konden voeren met volledige blokken. We hadden dan r blokken en ieder blok had v velden. We vonden dan een andere schatting $\hat{v}_i'$ met variantie $\sigma^2_{\hat{v}_i'}$. De doeltreffendheidsfactor e met betrekking tot $\hat{v}_i$ is nu

$$e \overset{def}{=} \frac{\sigma^2_{\hat{v}_i'}}{\sigma^2_{\hat{v}_i}}$$

Nu geldt

$$\hat{v}_i' = y_{i.} - y_{..} = \frac{v-1}{v} y_{i.} - \frac{1}{v}\sum_{j \neq i} y_j$$

$$\sigma^2_{\hat{v}_i'} = \left\{\left(\frac{v-1}{v}\right)^2 + \frac{v-1}{v^2}\right\}\frac{1}{r}\sigma^2 = \frac{v-1}{rv}\sigma^2$$

Dus $$e = \frac{\frac{v-1}{rv}}{\frac{k(v-1)}{\lambda v^2}} = \frac{\lambda v}{rk}$$

We kunnen ook de doeltreffendheidsfactor bepalen met betrekking tot de schatting van $v_i - v_j$.

In het schema met de volledige blokken hebben we:

$$\hat{v}_i' - \hat{v}_j' = y_{i\cdot} - y_{j\cdot} \qquad \text{en}$$

$$\sigma^2_{\hat{v}_i' - \hat{v}_j'} = \sigma^2_{y_{i\cdot}} + \sigma^2_{y_{j\cdot}} = \frac{2}{r}\sigma^2 .$$

Dus $$e = \frac{2/r}{2k/\lambda v} = \frac{\lambda v}{rk} .$$

We vinden dus in beide gevallen dezelfde factor.

Voor het berekenen van de toetsingsgrootheden gebruikt men het volgende schema:

1e methode	vrijh. graden	Sommen van kwadraten (1)	Sommen van kwadraten (2)	2e methode
	$v-1$	$\frac{[T_i]}{k(r-\lambda)}$	$\frac{[W_i]}{rv(v-k)(k-1)}$	
	$b-v$	$\dagger$	$\dagger$	
Blokken (zonder rekening te houden met variëteiten)	$b-1$	$\frac{[B_j]}{k}$	$\dagger$	Blokken (variëteitsinvloed geëlimineerd)
Variëteiten (blokinvloed geëlimineerd)	$v-1$	$\frac{[Q_i]}{\lambda k v}$	$\frac{[V_i]}{r}$	Variëteiten (zonder rekening te houden met blokken)
Q_α	$rv-v-b+1$	$\dagger$	$\dagger$	Q_α
Totaal	$rv-1$	$[y]$	$[y]$	Totaal

De vierkante haken betekenen: de som nemen van de kwadraten van de desbetreffende grootheid verminderd met zijn gemiddelde, b.v.

$$[B_j] = \sum_{j=1}^{b} \left[B_j - \frac{\sum B_j}{b}\right]^2 .$$

$$W_i \overset{def}{=} (v-k)V_i - (v-1)T_i + (k-1)G ,$$

waarin G de som van alle waarnemingen voorstelt;

$$Q_i \overset{def}{=} kV_i - T_i .$$

De termen aangegeven met het teken $\dagger$ vindt men door optellen of aftrekken.

De tellers van de toetsingsgrootheden voor de toetsen $H_2 : v_i = 0 \ (i=1,\dots,v)$ en $H_2' : b_j = 0 \ (j=1,\dots,b)$, vinden we in het schema met de aanduiding variëteiten (blokinvloed geëlimineerd) en blokken (variëteitsinvloed geëlimineerd). Door een eenvoudige berekening is na te gaan dat dit overeenkomt met wat op p. 44 en p. 45 is afgeleid.

Bepalen we de verwachtingen van de verschillende sommen van kwadraten uit dit schema, dan vinden we (zelfde volgorde als in het vorige schema):

$$
\begin{array}{lcl}
(v-1)\sigma^2 + \frac{r-\lambda}{k}\sum_i v_i^2 + 2\sum_{i,j}{}' v_i b_j + \frac{k}{r-\lambda}\sum_i\Big(\sum_j^{(i)} b_j\Big)^2 & & (v-1)\sigma^2 + \frac{\lambda v}{r(r-\lambda)}\sum_i\Big(\sum_j^{(i)} b_j\Big)^2 \\
(b-v)\sigma^2 + k\sum_j b_j^2 - \frac{k}{r-\lambda}\sum_i\Big(\sum_j^{(i)} b_j\Big)^2 & \longleftrightarrow & - - - - - - - \\
\hline
(b-1)\sigma^2 + k\sum_j b_j^2 + \frac{r-\lambda}{k}\sum_i v_i^2 + 2\sum_{i,j}{}' v_i b_j \quad + & & (b-1)\sigma^2 + k\sum b_j^2 + \frac{1}{r}\sum_i\Big(\sum_j^{(i)} b_j\Big)^2 \quad + \\
(r-1)\sigma^2 + \frac{\lambda v}{k}\sum_i v_i^2 & & (r-1)\sigma^2 + r\sum v_i^2 + \frac{1}{r}\sum_i\Big(\sum_j^{(i)} b_j\Big)^2 + 2\sum_{i,j}{}' v_i b_j \\
(rv-b-v+1)\sigma^2 & & (rv-b-v+1)\sigma^2 \\
\hline
(rv-1)\sigma^2 + r\sum_i v_i^2 + k\sum_j b_j^2 + 2\sum_{i,j}{}' v_i b_j \quad + & \longleftrightarrow & - - - - - - - \quad +
\end{array}
$$

$\sum_{i,j}{}' v_i b_j$ wil zeggen dat we alleen sommeren over de producten $v_i b_j$ waarvan de variëteit v_i in het blok b_j voorkomt.

De berekeningen worden eenvoudiger als $b = v$, de term op de tweede regel van het schema wordt in dat geval namelijk gelijk aan nul. Om dit te kunnen bewijzen leiden we eerst het volgende lemma af:

<u>Lemma 7.1. In een evenwichtig schema met onvolledige blokken, waarin het aantal variëteiten gelijk is aan het aantal blokken komt in ieder tweetal blokken eenzelfde aantal gemeenschappelijke variëteiten voor. Dit aantal is gelijk aan λ, het aantal malen dat twee variëteiten samen in één blok voorkomen.</u>

Beschouw een willekeurig blok. Hierin komen k variëteiten voor en dus $\frac{1}{2}k(k-1)$ paren variëteiten. Deze paren komen in de andere blokken ieder nog $\lambda-1$ keren voor. De $v-1$ andere blokken nummeren we van $1, \ldots, v-1$. Stel dat van de k variëteiten, die in ons blok voorkomen er λ_i in het i^e blok voorkomen, dus $\frac{1}{2}\lambda_i(\lambda_i-1)$ paren. Er geldt dus

$$(7.25) \qquad \tfrac{1}{2}\sum_{i=1}^{v-1}\lambda_i(\lambda_i-1) = \tfrac{1}{2}k(k-1)(\lambda-1) = \tfrac{1}{2}\lambda(\lambda-1)(v-1),$$

immers $b = v$ en daar $bk = rv$ is ook $r = k$ en bovendien geldt $r(k-1) = \lambda(v-1)$.

Iedere variëteit komt in de andere blokken nog $(r-1)$ keren voor, dit geeft ons dus

$$(7.26) \qquad \sum_{i=1}^{v-1}\lambda_i = k(r-1) = \lambda(v-1)$$

Uit (7.25) en (7.26) volgt

$$\sum_{i=1}^{v-1} (\lambda_i^2 - \lambda^2) = 0.$$

Maar $\sum_{i=1}^{v-1} (\lambda_i^2 - \lambda^2) = \sum_{i=1}^{v-1} (\lambda_i - \lambda)^2$, want volgens (7.26) is λ het gemiddelde van de λ_i.

Dus $\sum_{i=1}^{v-1} (\lambda_i - \lambda)^2 = 0 \longrightarrow \lambda_i = \lambda \; (i = 1, \ldots, v-1)$.

M.a.w. het door ons beschouwde blok heeft met elk der andere blokken λ variëteiten gemeen

Uit een evenwichtig schema met b onvolledige blokken en v variëteiten kunnen we een ander schema afleiden met b variëteiten en v blokken, door nl. de j^e variëteit in het i^e blok te plaatsen als in het oorspronkelijke schema de i^e variëteit in het j^e blok voorkwam. Het nieuwe schema wordt alleen dan weer een evenwichtig schema indien in het eerste schema ieder tweetal blokken evenveel variëteiten gemeen heeft. Een voldoende voorwaarde hiervoor is volgens lemma 7.1 dus dat $b = v$. Deze voorwaarde is ook nodig; men zie hiervoor b.v. K.M.MAJUMDAR, On some theorems in combinatories relating to incomplete block designs, A.M.S. 23 (1953), p. 377-389.

In het geval dat $b = v$, geldt

$$\sum_i T_i^2 = (k-\lambda) \sum_j B_j^2 + \lambda G^2 \tag{7.28}$$

Immers in het linkerlid komen alle kwadraten van de bloktotalen B_j k keren voor en alle dubbele producten λ keer, op grond van lemma 7.1. Dus

$$\sum_i T_i^2 = \lambda \Big(\sum_j B_j\Big)^2 + (k-\lambda) \sum_j B_j^2 = (k-\lambda) \sum_j B_j^2 + \lambda G^2$$

We wilden bewijzen dat de term op de tweede regel van ons rekenschema gelijk werd aan nul indien $b = v$. Dit is dus het geval indien geldt

$$\frac{[T_i]}{k(r-\lambda)} = \frac{[B_j]}{k} \tag{7.29}$$

Het gemiddelde van T_i is gelijk aan $\frac{\sum T_i}{v} = \frac{k}{v} G$

Dus

$$\frac{[T_i]}{k(r-\lambda)} = \frac{1}{k(r-\lambda)} \sum_i T_i^2 - \frac{k}{v(r-\lambda)} G^2 = \qquad \text{(volgens 7.28)}$$

$$\frac{k-\lambda}{k(r-\lambda)}\sum_j B_j^2 + \left\{\frac{\lambda}{k(r-\lambda)} - \frac{k}{v(r-\lambda)}\right\} G^2 = \quad \text{(daar } k=r \text{ als } b=v\text{)}$$

$$= \frac{1}{k}\sum_j B_j^2 + \frac{\lambda v - k^2}{vk(r-\lambda)} G^2 = \frac{1}{k}\sum_j B_j^2 - \frac{1}{kb} G^2 = \frac{[B_j]}{k}$$

Immers $\lambda(v-1) = r(k-1) = k(k-1)$, dus $\lambda v - \lambda = k^2 - k$,

$\lambda v - k^2 = -(k-\lambda) = -(r-\lambda)$, dus $\frac{\lambda v - k^2}{k(r-\lambda)} = \frac{1}{rk} = \frac{1}{bk}$.

Hiermee is het gestelde dus bewezen.

Een voorbeeld van toepassing van Latijnse vierkanten.

Dit voorbeeld werd ontleend aan R.A.FISHER, The design of experiments, pg. 196 en verder.

We denken ons het hele proefveld met dezelfde variëteit bezaaid, terwijl de verschillende vakken verschillend bemest worden. Als model kan nu weer (7.1) gebruikt worden, waarin nu y_{ijk} de opbrengst is van het vak in de i^e rij en de j^e kolom dat de k^e soort bemesting heeft ontvangen. Ter bepaling van de gedachten onderstellen we $m=6$; er kunnen dus 6 verschillende bemestingen vergeleken worden. Van deze soorten bemesting is verder gegeven dat ze als in het onderstaande schema uit verschillende doses phosphaat- en nitraat-verbindingen zijn samengesteld:

dosis nitraat	dosis phosphaat 0	dosis phosphaat 1
0	ρ_1	ρ_4
1	ρ_2	ρ_5
2	ρ_3	ρ_6

Op grond hiervan beschrijft Fisher de toetsing van de volgende 5 hypothese, welke samen met de hypothese $\rho_k = 0$ voorstellen:

1) $\rho_1+\rho_2+\rho_3-\rho_4-\rho_5-\rho_6 = 0$
2) $\rho_1-\rho_3+\rho_4-\rho_6 = 0$
3) $\rho_1+\rho_3-2\rho_2+\rho_4+\rho_6-2\rho_5 = 0$
4) $\rho_1-\rho_3-(\rho_4-\rho_6) = 0$
5) $\rho_1+\rho_3-2\rho_2-(\rho_4+\rho_6-2\rho_5) = 0$

Ten einde de betekenis van deze hypothesen nader te analyseren, bekijken we in plaats van model (7.1) het volgende

(7.1') $\quad \mathcal{E}\, y_{ijst} = \mu_i + \nu_j + n_s + p_t + \rho_{st} + \rho,$

met

$$\sum_i \mu_i = \sum_j \nu_j = \sum_s n_s = \sum_t p_t = \sum_s \rho_{st} = \sum_t \rho_{st} = 0,$$

waarin s en t resp. op de s^e nitraat dosis en de t^e phosphaat dosis slaan. Daar μ_i, ν_j en ρ dezelfde betekenis bezitten als bij (7.1) en n_s, p_s en ρ_{st} samen weer 5 onafhankelijke parameters opleveren evenals de ρ_k, zullen de modellen (7.1) en (7.1') volkomen aequivalent zijn en dezelfde uitdrukking (7.3) voor Q_α opleveren.

De parameters ρ_k uit (7.1) zijn bij (7.1') als volgt vervangen:

$$\begin{aligned} &\rho_1 \to n_0 + p_0 + \rho_{00} \quad &&; \rho_4 \to n_0 + p_1 + \rho_{01} \\ &\rho_2 \to n_1 + p_0 + \rho_{10} \quad &&; \rho_5 \to n_1 + p_1 + \rho_{11} \\ &\rho_3 \to n_2 + p_0 + \rho_{20} \quad &&; \rho_6 \to n_2 + p_1 + \rho_{21} \quad , \end{aligned}$$

zodat de hypothesen 1),...,5) overgaan in (na uitwerking met behulp van de bijvoorwaarden):

1) $p_1 = p_0 = 0$ ("geen phosphaateffect") ,

2) $n_0 = n_2$,

3) $n_1 = 0$,

4) $\rho_{01} = \rho_{21}$ $(\rho_{00} = \rho_{20})$,

5) $\rho_{10} = \rho_{11} = 0$

Hypothesen 2) en 3) samen zeggen: $n_s = 0$ (geen "nitraateffect"), terwijl de hypothesen 4) en 5) samen: $\rho_{st} = 0$ opleveren (geen interactie).

Analoog met het op pg. 25 e.v. besprokene bij model (4.82') zijn ook hier de parameters n_s en p_t moeilijk te interpreteren, indien de interactie parameters $\rho_{st} \neq 0$ zijn, zodat toetsing van alleen hypothese 1) of 2) of 3) minder zinvol is. De hypothese dat de phosphaat doses bij geen der nitraatdoses een verschillend effect hebben luidt:

$$\rho_1 = \rho_4 \;\&\; \rho_2 = \rho_5 \;\&\; \rho_3 = \rho_6$$

en blijkt bij uitwerking identiek te zijn met de hypothesen 1), 4) en 5) samen, dus: $p_t = 0 \;\&\; \rho_{st} = 0$.

Evenzo is de hypothese dat de nitraat-doses bij geen van beide phosphaatdoses in uitwerking verschillen, dus $\rho_1 = \rho_2 = \rho_3$ $\&$ $\rho_4 = \rho_5 = \rho_6$, is identiek met 2), 3), 4) en 5) samen, dus met: $n_s = 0 \;\&\; \rho_{st} = 0$.

Hoofdstuk X

Ontbrekende waarnemingen

Het komt vaak voor dat bv. in Latijns vierkant één of meer waarnemingen ontbreken tengevolge van een mislukking van het experiment, of dat in een r -voudige klassificatie de aantallen in de klassen niet gelijk zijn. Dit laatste kan o.a. een gevolg zijn van het feit dat we de factoren die de klassegrootte bepalen niet in de hand hebben.

In al deze gevallen moeten we Q_α en $Q_r - Q_\alpha$ zien te vinden. Dat wil zeggen we moeten een kwadratische vorm

(10.1) $$Q = \sum_1^N {}^{\alpha} \left(y_\alpha - \sum_1^k {}^{i} g_{i\alpha}\beta_i \right)^2$$

minimaliseren onder de bijvoorwaarden

(10.2) $$\sum_1^k {}^{j} c_{uj}\beta_j = 0 \quad , \quad u = 1, \dots r < k \; , \quad \text{rang } (c_{uj}) = r$$

Dit minimum is gelijk aan Q_α, indien (10.2) de voorwaarden zijn onder de toegelaten hypothesen opgelegd aan de β_i, terwijl het Q_r is als (10.2) de voorwaarden zijn onder de nulhypothese.

We zullen nu bewijzen:

Stelling 10.1: Stel S is het minimum van Q met betrekking tot de β_i onder de voorwaarden (10.2).

Stel

$$a_{pq} = \sum_1^N {}^{\alpha} g_{p\alpha} g_{q\alpha} \quad \text{voor } k \geq p > 0,\; k \geq q > 0.$$

(10.3) $$a_{q0} = a_{0q} = \sum_1^N {}^{\alpha} y_\alpha g_{q\alpha} \, ,$$

$$a_{00} = \sum_1^N {}^{\alpha} y_\alpha^2 .$$

Dan is

(10.4) $$S = \frac{\Delta}{\Delta_{00}} \, ,$$

indien

$$\Delta = \begin{vmatrix} a_{00} & a_{01} & \dots & a_{0k} & 0 & \dots & 0 \\ a_{10} & a_{11} & \dots & a_{1k} & c_{11} & \dots & c_{r1} \\ \vdots & & & & & & \vdots \\ a_{k0} & a_{k1} & \dots & a_{kk} & c_{1k} & \dots & c_{rk} \\ 0 & c_{11} & \dots & c_{1k} & 0 & \dots & 0 \\ \vdots & & & & & & \vdots \\ 0 & c_{r1} & \dots & c_{rk} & 0 & \dots & 0 \end{vmatrix}$$

en Δ_{00} de minor is van a_{00} in (Δ).

Om S te vinden moeten we

$$Q + \sum^{u}\sum^{j} \lambda_u c_{uj}\beta_j$$

differentiëren naar de β_i en de λ_u en in de gevonden uitdrukkingen $\hat{\beta}_i$ substitueren voor β_i, waarna we de $\hat{\beta}_i$ en de λ_u kunnen oplossen uit de vergelijkingen gelijk gesteld aan nul. We vinden dan voor de eerste k vergelijkingen

$$\left(\frac{\partial Q}{\partial \beta_p}\right)_{\beta_q=\hat{\beta}_q} + \sum_{1}^{r}{}^{u} \lambda_u c_{up} =$$

$$= 2\left(-a_{0p} + \sum_{1}^{k}{}^{q} a_{pq}\hat{\beta}_q\right) + \sum_{1}^{r}{}^{u} \lambda_u c_{up} = 0,$$

$$(p=1,\dots,k).$$

We vermenigvuldigen de p^e vergelijking met $\hat{\beta}_p$ en sommeren daarna over p. Daar $\sum^{p} c_{up}\hat{\beta}_p = 0$ levert dit

(10.7) $$-\sum^{p} a_{0p}\hat{\beta}_p + \sum^{p}\sum^{q} a_{pq}\hat{\beta}_p\hat{\beta}_q = 0$$

We ontwikkelen S als volgt (vgl. (10.1))

(10.8) $$S = a_{00} - 2\sum_{1}^{k}{}^{p} a_{0p}\hat{\beta}_p + \sum^{p}\sum^{q} a_{pq}\hat{\beta}_p\hat{\beta}_q =$$

(volgens 10.7) $= a_{00} - \sum^{p} a_{0p}\hat{\beta}_p$

We krijgen nu de volgende $k+r+1$ vergelijkingen voor de onbekenden $\hat{\beta}_0=1, \hat{\beta}_1,\dots,\hat{\beta}_k, \frac{\lambda_1}{2},\dots,\frac{\lambda_r}{2}$:

$$(S-a_{00})\hat{\beta}_0 + \sum^{p} a_{0p}\hat{\beta}_p = 0 \;;$$

$$-a_{0p}\hat{\beta}_0 + \sum^{q} a_{pq}\hat{\beta}_q + \sum^{u} \frac{\lambda_u}{2} c_{up} = 0, \quad (p=1,\dots,k),$$

$$\sum^{q} c_{uq}\hat{\beta}_q = 0, \quad (u=1,\dots,r).$$

Dit systeem heeft een niet triviale oplossing, immers $\hat{\beta}_0 = 1$, dus de determinant van het stelsel is gelijk aan nul.

$$\begin{vmatrix} S-a_{00} & a_{01} & \cdots & a_{0k} & 0 & \cdots & 0 \\ -a_{10} & a_{11} & \cdots & a_{1k} & c_{11} & \cdots & c_{r1} \\ \vdots & & & & & & \\ -a_{k0} & a_{k1} & \cdots & a_{kk} & c_{1k} & \cdots & c_{rk} \\ 0 & c_{11} & \cdots & c_{1k} & 0 & \cdots & 0 \\ \vdots & & & & & & \vdots \\ 0 & c_{r1} & \cdots & c_{rk} & 0 & \cdots & 0 \end{vmatrix} = 0.$$

ofwel $$S\Delta_{00} - \Delta = 0, \quad S = \frac{\Delta}{\Delta_{00}}, \quad q.e.d.$$

In sommige gevallen is het evenwel eenvoudiger om direct de kleinste kwadratenvergelijkingen op te lossen. Nemen we b.v. het geval van een tweevoudige klassificatie, waarin beide klassificaties uit 2 klassen bestaan, terwijl van iedere combinatie 4 waarnemingen zijn gedaan. In één van de groepen ontbreekt een waarneming. We hebben dus het model:

$$\begin{aligned} \mathcal{E}\, \underline{x}_{11i} &= \mu+\nu+\rho+\tau, \quad i=1,2,3,4, \\ \mathcal{E}\, \underline{x}_{12i} &= \mu-\nu-\rho+\tau, \quad i=1,2,3,4, \\ \mathcal{E}\, \underline{x}_{21i} &= -\mu+\nu-\rho+\tau, \quad i=1,2,3, \\ \mathcal{E}\, \underline{x}_{22i} &= -\mu-\nu+\rho+\tau, \quad i=1,2,3,4. \end{aligned} \tag{10.5}$$

Differentiëren we Q naar τ, ρ, ν en μ, dan vinden we

$$\begin{aligned} 4x_{11.}+4x_{12.}+3x_{21.}+4x_{22.}-\hat{\mu}+\hat{\nu}-\hat{\rho}-15\hat{\tau} &= 0, \\ -4x_{11.}+4x_{12.}+3x_{21.}-4x_{22.}-\hat{\mu}+\hat{\nu}+15\hat{\rho}+\hat{\tau} &= 0, \\ -4x_{11.}+4x_{12.}-3x_{21.}+4x_{22.}+\hat{\mu}+15\hat{\nu}+\hat{\rho}-\hat{\tau} &= 0, \\ -4x_{11.}-4x_{12.}+3x_{21.}+4x_{22.}+15\hat{\mu}+\hat{\nu}-\hat{\rho}+\hat{\tau} &= 0. \end{aligned} \tag{10.6}$$

Hieruit lossen we $\hat{\mu}$, $\hat{\nu}$, $\hat{\rho}$ en $\hat{\tau}$ op:

$$\begin{aligned} \hat{\mu} &= \tfrac{1}{4}(x_{11.}+x_{12.}-x_{21.}-x_{22.}), \\ \hat{\nu} &= \tfrac{1}{4}(x_{11.}-x_{12.}+x_{21.}-x_{22.}), \\ \hat{\rho} &= \tfrac{1}{4}(x_{11.}-x_{12.}-x_{21.}+x_{22.}), \\ \hat{\tau} &= \tfrac{1}{4}(x_{11.}+x_{12.}+x_{21.}+x_{22.}). \end{aligned} \tag{10.7}$$

We hadden dus ditzelfde resultaat gevonden, wanneer we in plaats van de ontbrekende waarneming in de 3e groep het gemiddelde van de drie andere waarnemingen hadden gesubstitueerd.

We willen nu toetsen $\rho = 0$, dan berekenen we

$$\sigma^2_{\hat{\rho}} = \frac{1}{16}\left(\sigma^2_{\underline{x}_{11.}}+\sigma^2_{\underline{x}_{12.}}+\sigma^2_{\underline{x}_{21.}}+\sigma^2_{\underline{x}_{22.}}\right) =$$

$$= \frac{1}{16}\left(\frac{1}{4}+\frac{1}{4}+\frac{1}{3}+\frac{1}{4}\right)\sigma^2 = \frac{1}{16}\cdot\frac{13}{12}\,\sigma^2.$$

Dus volgens stelling 4.3 geldt

$$Q_2 - Q_\alpha = 16 \cdot \frac{12}{13}\,\hat{\rho}^2 = \frac{12}{13}\left(x_{11.}-x_{12.}-x_{21.}+x_{22.}\right)^2.$$