

MATHEMATISCH CENTRUM

2e BOERHAAVESTRAAT 49

AMSTERDAM

STATISTISCHE AFDELING

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

Rapport S 198

Proefopzet voor het opsporen van optimale condities bij  
serumzuivering (Eerste fase).

door

A.R. Bloemena

en

Ph. van Elteren

Mei 1956

## 1. Inleiding

Een zuiveringsprocédé voor een serum zal op grote schaal worden toegepast. Het is daarom van belang de voorwaarden, waaronder dit zuiveringsproces zal werken, zodanig te kiezen, dat het geheel optimaal verloopt. Wat in dit geval met "optimaal" wordt bedoeld, zal verderop worden aangegeven.

De factoren, welke invloed hebben op het verloop van de zuivering zijn:

### Eerste stadium:

zuurgraad ( $p_H$ )	$Z_1$
temperatuur	$T_1$ ( $^{\circ}C$ )
tijdsduur	$t_1$ (min.)
hoeveelheid toegevoegde pepsine	$p$

### Tweede stadium:

hoeveelheid toegevoegde $(NH_4)_2SO_4$	$a$
zuurgraad ( $p_H$ )	$Z_2$
temperatuur	$T_2$ ( $^{\circ}C$ )
tijdsduur	$t_2$ (min.)

### Derde stadium:

zuurgraad ( $p_H$ )	$Z_3$
---------------------	-------

Op grond van chemische overwegingen kan het onderzoek eerst apart uitgevoerd worden voor de eerste twee stadia, terwijl daarna de laatste factor kan worden onderzocht, uitgaande van de optimale condities voor de eerste acht factoren.

Zowel na het tweede als na het derde stadium worden gemeten:

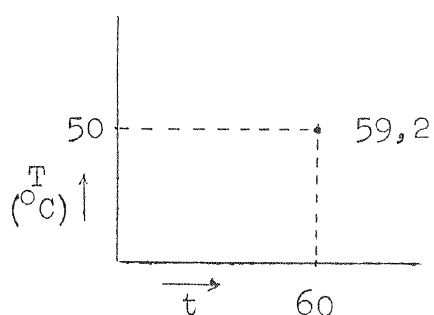
1. de opbrengst  $O$  aan gezuiverd serum, en
2. de reinigingsfactor  $R$  van het gezuiverde serum.

De reinigingsfactor moet minstens een bepaalde waarde  $R_1$  hebben, wil het product een acceptabele kwaliteit bezitten. Bovendien is het proces van dien aard, dat  $R$  daalt, indien de factoren zodanig gewijzigd worden, dat  $O$  stijgt. Het is om deze redenen, dat optimale condities in dit geval gedefinieerd moeten worden als die waarden van de negen factoren, waarbij  $O$  maximaal is, onder de voorwaarde dat  $R$  een bepaalde waarde  $R^*$  ( $\geq$  voorgeschreven minimum waarde) heeft.

## 2. Experimenteertechnieken

Het type van probleem, zoals dat in de inleiding gesteld is, komt in de chemische technologie zeer veel voor. Aan de hand van een eenvoudig voorbeeld zal worden aangegeven hoe men in het algemeen in deze gevallen optimale condities kan vinden.

Stel, men heeft een proces, dat slechts van twee factoren afhankelijk is, bijvoorbeeld tijdsduur  $T$ , en temperatuur  $t$ . Bij elk gekozen stel waarden van  $T$  en  $t$  kan een experiment worden verricht en die grootte gemeten worden, welke men wil optimaliseren. Deze grootte noemen wij een responsie; dit kan al naar gelang van het probleem een opbrengst of productiekosten, e.d. zijn. In een diagram als figuur 1 kan de gemeten responsie als



(minuten)

figuur 1

een staafje, loodrecht op het vlak van tekening worden uitgezet. Doet men dit voor zeer vele combinaties van  $t$  en  $T$ , en ziet men hierbij af van de fluctuaties in de responsie, die bij herhaling onder gelijkblijvende omstandigheden optreden (waarnemings- en proces-onnauwkeurigheden), dan vormen de toppen van deze staafjes een oppervlak: het responsie-oppervlak. In figuur 2 is een dergelijk oppervlak aangegeven door de lijnen van gelijke responsie ("hoogtelijnen"). Het doel van het onderzoek is nu de coördinaten van het punt A te bepalen.

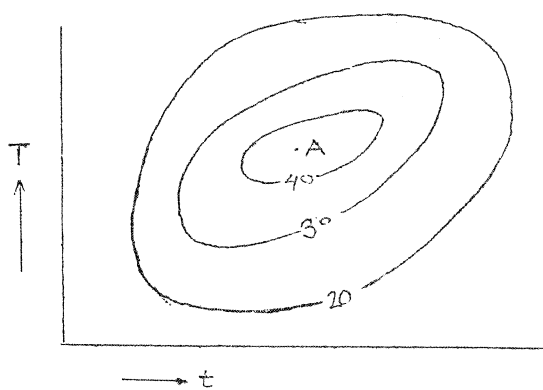


fig. 2. Responsie-oppervlak voor twee factoren.

Tot voor enkele jaren ging men in het algemeen op de volgende wijze te werk (zie [1] en [5]).

Van de factoren worden er  $k-1$  constant gehouden en bij verschillende waarden van de  $k^{\text{de}}$  factor experimenten uitgevoerd. Men houdt dan die waarde voor de  $k^{\text{de}}$  factor aan, die een zo groot mogelijke opbrengst geeft (dit behoeft niet precies één der waarden te zijn, waarbij een experiment verricht is). Men voert nu een volgende serie experimenten uit, waarbij één der andere factoren gevarieerd wordt. Indien men op deze wijze achtereenvolgens de  $k$  factoren alle eenmaal gevarieerd heeft, is de eerste fase van het onderzoek voltooid. Voert men enkele van dergelijke fasen uit, dan zal men uiteindelijk tot waarden van de factoren geraken, die niet veel verschillen van de optimale waarden. De eerste fase van een dergelijke experimenteertechniek is weergegeven in figuur 3 voor het geval  $k = 2$ .

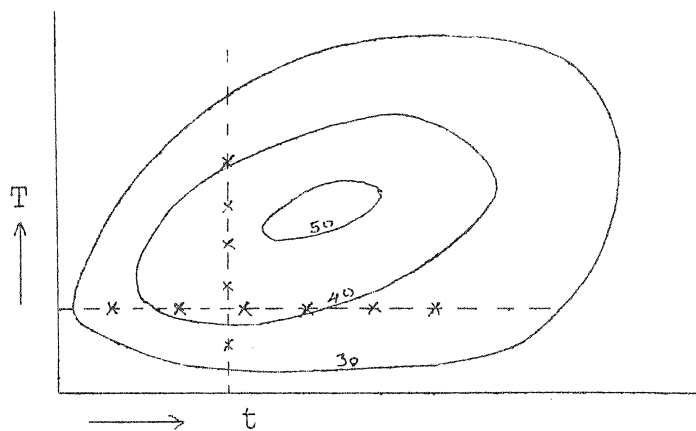
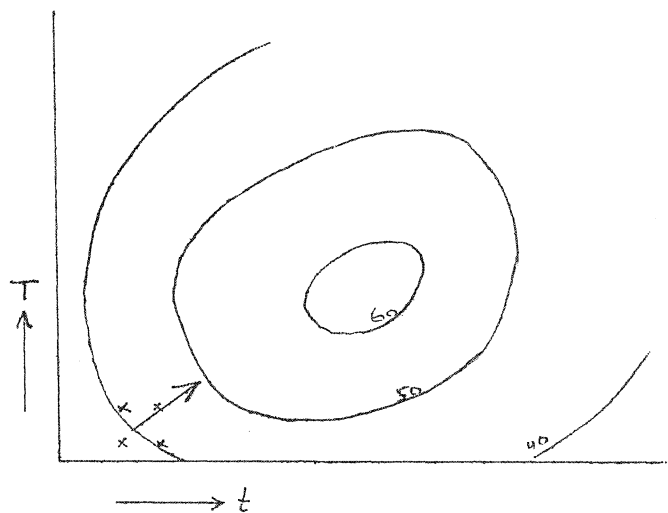


fig. 3: de "één factor per keer" methode.

De methode van de "één factor per keer" is evenwel zeer kostbaar en tijdrovend. Bovendien zijn gevallen denkbaar, waarbij men in het geheel niet de optimale waarden van de factoren bereikt, of zelfs benadert. Een belangrijke verbetering is mogelijk door toepassing van een methode, die enkele jaren geleden ontwikkeld en beschreven is door G.E.P. BOX (zie [1], [2] en [3]). Men kan een responsie-oppervlak, zoals bijvoorbeeld aangegeven in figuur 2, opvatten als een berg, waarvan men zo snel mogelijk de top wil bereiken. Het zal dan duidelijk zijn, dat men bij voorkeur zich loodrecht op de hoogtelijnen zal willen bewegen, om op deze wijze langs de kortste weg naar de top te komen.

Bevindt men zich (zie figuur 4) in een punt B, dan kan men, met B als centrum een serie experimenten uitvoeren. Aan de bij deze experimenten gevonden waarden van de responsie past men een oppervlak aan.

In het algemeen zal het responsie-oppervlak een zeer ingewikkelde vorm hebben, maar indien de experimenten slechts een klein deel betreffen van het gebied, waarover de factoren gevarieerd kunnen worden, zal in het algemeen het responsie-oppervlak plaatselijk goed door een eerste of tweede graads oppervlak benaderd kunnen worden. Uit de algebraïsche vergelijking van dit aangepaste oppervlak wordt dan de richting bepaald, waarin men kan verwachten, dat het optimum zich zal bevinden. Na verplaatsing over zekere afstand in deze richting zal men dan opnieuw experimenten uitvoeren. Indien men zover van het optimum verwijderd is, dat het responsie-oppervlak goed door een plat vlak te benaderen is, komt het bepalen van de richting van het optimum neer op het bepalen van de steilste weg (zie figuur 4). Het is deze methode, die zal worden toegepast bij het vinden van de optimale condities van het serumzuiveringsprocédé.



Figuur 4: De steilste weg naar het optimum

### 3. Proefopzet voor de eerste fase

Als proefopzet voor de eerste fase van het onderzoek wordt voorgesteld een  $\frac{1}{16}$  deel van een  $2^8$  factorieel schema in duplo (zie [3] en [4] en het als bijlage aan dit rapport toegevoegde memorandum S 170 (M 74)). Het schema is ook op te vatten als een volledig  $2^4$  schema voor de factoren  $z_1$ ,  $T_1$ ,  $t$ , en  $\beta$ , waarbij vervolgens de overige 4 factoren verstrengeld zijn met de 4 derde orde interacties van de eerstgenoemde 4 factoren. Dit kan men ook aangeven met de groepsrelatie:

$$\text{eenheid} = 1235 = 1246 = 1347 = 2348.$$

Dit schema is in Tabel I gegeven in de volgorde waarin de experimenten zouden kunnen worden uitgevoerd. Alle factoren worden op twee niveaus in het onderzoek betrokken. De keuze van de waarden van de factoren, die geassocieerd zullen worden met resp. +1 en -1 zal nog moeten worden gedaan. Deze keuze is uiteraard gelijkwaardig met het vastleggen van de waarden van deze factoren in de oorsprong van het schema en de grootte van de eenheid. Kiest men bijvoorbeeld voor de eerste factor als oorsprong  $p_H = 3,2$  en als eenheid  $0,3$ , dan zullen de +1 en -1 voor deze factor in Tabel I overeenkomen met resp.  $3,2 + 0,3 = 3,5$  en  $3,2 - 0,3 = 2,9$ .

Wij zouden willen aanbevelen, de keuze van de oorsprong en de eenheid met zorg te doen geschieden, daar van een goede keuze veel kan afhangen. De eerste fase van het onderzoek is in hoofdzaak gericht op het verkrijgen van informatie over de grootte van de optredende effecten, en over de aard van het aan te passen oppervlak. De grootte van de effecten wordt zeer sterk beïnvloed door:

1. de eenheid, die voor de betreffende factoren gekozen wordt
- en 2. de afstand van de oorsprong tot het optimum.

Tabel I: Voorgestelde proefopzet  
 ( $\frac{1}{16} \times 2^8$  schema in duplo en in aselechte volgorde)

experiment nummer	factor:	$z_1$	$T_1$	$t_1$	$p$	$a$	$z_2$	$T_2$	$t_2$
1		+1	-1	+1	-1	-1	+1	-1	+1
2		-1	-1	-1	+1	-1	+1	+1	+1
3		-1	-1	+1	-1	+1	-1	+1	+1
4		+1	+1	+1	-1	+1	-1	-1	-1
5		-1	-1	-1	-1	-1	-1	-1	-1
6		-1	+1	+1	-1	-1	+1	+1	-1
7		+1	+1	-1	+1	-1	+1	-1	-1
8		+1	+1	+1	+1	+1	+1	+1	+1
9		-1	+1	-1	+1	+1	-1	+1	-1
10		+1	+1	-1	+1	-1	+1	-1	-1
11		+1	-1	+1	+1	-1	-1	+1	-1
12		+1	-1	-1	-1	+1	+1	+1	-1
13		+1	-1	-1	-1	+1	+1	+1	-1
14		-1	+1	-1	-1	+1	+1	-1	+1
15		+1	-1	-1	+1	+1	-1	-1	+1
16		-1	+1	-1	-1	+1	+1	-1	+1
17		+1	-1	-1	+1	+1	-1	-1	+1
18		-1	-1	+1	-1	+1	-1	+1	+1
19		-1	+1	+1	-1	-1	+1	+1	-1
20		-1	+1	+1	+1	-1	-1	-1	+1
21		+1	-1	+1	+1	-1	-1	+1	-1
22		-1	-1	-1	+1	-1	+1	+1	+1
23		+1	+1	+1	+1	+1	+1	+1	+1
24		+1	+1	-1	-1	-1	-1	+1	+1
25		+1	+1	+1	-1	+1	-1	-1	-1
26		+1	-1	+1	-1	-1	+1	-1	+1
27		-1	+1	-1	+1	+1	-1	+1	-1
28		-1	-1	+1	+1	+1	+1	-1	-1
29		+1	+1	-1	-1	-1	-1	+1	+1
30		-1	-1	-1	-1	-1	-1	-1	-1
31		-1	+1	+1	+1	-1	-1	-1	+1
32		-1	-1	+1	+1	+1	+1	-1	-1

Indien de eenheden te klein gekozen worden, bestaat de mogelijkheid, dat de optredende effecten door de waarnemings-  
onnauwkeurigheden niet ontdekt worden. Worden zij te groot  
gekozen, dan zal een te groot deel van het mogelijke variatie-  
gebied van de factoren in het onderzoek betrokken worden. In  
dit geval zal veelal een oppervlak van hogere graad aangepast  
dienen te worden, daar een plat vlak geen voldoende goede be-  
nadering het responsie-oppervlak ter plaatse is. Omdat voor  
het aanpassen van oppervlakken van een hogere graad relatief  
veel experimenten nodig zijn, is het van belang de eenheid  
vooral niet te groot te kiezen.

Daar in het betreffende geval weinig bekend is over de  
optredende effecten is het aantal experimenten voor de eerste  
fase van het onderzoek zeer klein gehouden, zodat zonder hoge  
kosten een gedeelte hiervan eventueel met een andere keuze  
van eenheden kan worden herhaald.

Literatuur:

- 1 BOX, G.E.P.: The exploration and exploitation of response  
surfaces, Biometrics 1954, 10 16-61.
- 2 BOX, G.E.P. and K.B. WILSON: On the experimental attainment  
of optimum conditions. Journ. Roy. Stat. Soc.  
1951, B13, 1-45.
- 3 DAVIES, O.L. et al.: Design and Analysis of experiments,  
Oliver and Boyd, Edinburgh, 1953.
- 4 COCHRAN, W.G. and G.M. COX: Experimental Designs, Wiley,  
New York, 1950.
- 5 FRIEDMAN, M. and L.J. SAVAGE: Planning Experiments seeking  
maxima. Hoofdstuk 13 van EISENHART, HASTAY and  
WALLIS: Techniques of Statistical Analysis,  
Mc Graw Hill, New York, 1947.

-----

MATHEMATISCH CENTRUM,  
2e Boerhaavestraat 49,  
A m s t e r d a m - O .

Statistische Afdeling  
S 190 (M 74)

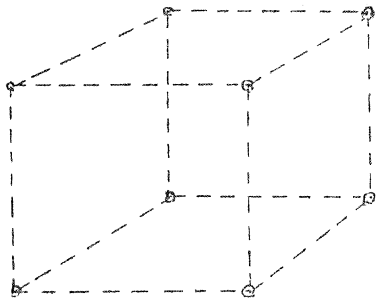
$2^k$ -de factoriële schema's <sup>1)</sup>

0. Inleiding

Door het uitvoeren van experimenten volgens een factoriël schema kunnen de uitwerkingen van een aantal factoren gelijktijdig bestudeerd worden. Wij zullen dit laten zien aan de hand van een eenvoudig voorbeeld.

Stel, de proefopzet, die uitgevoerd is (of zal worden), is een  $2^3$  factoriël schema, d.w.z. er zijn drie factoren, die ieder op twee niveaus bij de experimenten betrokken worden. In totaal worden er dus  $2^3 = 8$  experimenten verricht, waarbij voor elk van deze 8 experimenten een andere combinatie van de twee niveaus van de drie factoren gebruikt wordt.

De eenheden, waarin de niveaus van de factoren uitgedrukt kunnen worden, zijn vrij te kiezen. Door invoering van schaal-factoren kan dit zodanig worden gedaan, dat het hoge niveau van een factor door +1 wordt aangegeven, en het lage niveau door -1. Met deze keuze van schaal-factoren is een  $2^3$  factoriël schema voor te stellen door een kubus, waarvan ieder hoekpunt correspondeert met een der experimenten (zie figuur 1).



Figuur 1: een  $2^3$  factoriël schema.

Ook kunnen wij een programmamatrix aangeven, d.w.z. een matrix, waarvan de  $u$ -de rij de coördinaten aangeeft van het punt corresponderende met het  $u$ -de experiment. In dit geval is de programmamatrix

-----  
1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.



$$P = \begin{pmatrix} +1 & +1 & +1 \\ +1 & +1 & -1 \\ +1 & -1 & +1 \\ +1 & -1 & -1 \\ -1 & +1 & +1 \\ -1 & +1 & -1 \\ -1 & -1 & +1 \\ -1 & -1 & -1 \end{pmatrix} \dots (0,1)$$

1. Het mathematisch model

Geven wij een waarneming aan met  $y_u$  <sup>2)</sup>, dan wordt voor een  $2^3$  factorieel schema uitgegaan van het model, dat de  $y_u$  ( $u = 1, \dots, 8$ ) onderling onafhankelijke stochastische variabelen zijn met mathematische verwachtingen  $\eta_u$ , dus:

waarbij dus  $E y_u = \eta_u$  en waarbij bovendien een onderstelling gemaakt wordt:

$$y_u = \eta_u + \varepsilon_u \quad (1.1)$$

$$E \varepsilon_u^2 = \sigma^2 \quad (1.2)$$

Voorts wordt de  $\eta_u$  als volgt opgebouwd gedacht:

$$\eta_u = \mu_0 x_{0u} + \mu_1 x_{1u} + \mu_2 x_{2u} + \mu_3 x_{3u} + \mu_{12} x_{1u} x_{2u} + \mu_{13} x_{1u} x_{3u} + \mu_{23} x_{2u} x_{3u} + \mu_{123} x_{1u} x_{2u} x_{3u} \quad (1.3)$$

hetgeen geen enkele beperking oplegt aan de waarden, die de  $\eta_u$  kunnen bezitten.

Hierin noemen wij:

$\mu_0$  = het gemiddelde

$\mu_i$  = het i-de hoofdeffect

$\mu_{ij}$  = de i,j interactie (een interactie van de eerste orde)

$\mu_{ijk}$  = de i,j,k interactie (" " " " tweede " )

Verder is:

$$x_{0u} \equiv 1$$

en voor  $i = 1, 2, 3$

$$x_{iu} = \begin{cases} +1 & \text{als de i-de factor bij het u-de experiment op het +1} \\ & \text{niveau gehouden werd} \\ -1 & \text{als de i-de factor bij het u-de experiment op het -1} \\ & \text{niveau gehouden werd.} \end{cases}$$

-----  
 2) Een stochastische grootheid is een grootheid, die een waarschijnlijkheidsverdeling bezit. Stochastische grootheden worden door onderstreepte letters aangegeven; dezelfde letters niet onderstreept worden vaak gebruikt voor waarden, die zij aan kunnen nemen of aangenomen hebben.

De interpretatie van (1.3) is dat men de verwachting van een waarneming opgebouwd denkt uit een gemiddelde waarde en een aantal bijdragen van de factoren: de hoofdeffecten. Daar evenwel de effecten van twee factoren veelal niet onafhankelijk zijn, moet men teneinde de hoofdeffecten te kunnen optellen, hierbij een correctiefactor invoeren, de interacties. Een interactie tussen factor 1 en 2 betekent, dus dat factor 2 bij een experiment, waarbij factor 1 op het +1 niveau betrokken is, een andere invloed heeft, dan bij een experiment, waarbij factor 1 op het -1 niveau gehouden werd.

In 1.3 is  $\eta_u$  uitgedrukt in  $\mu_0, \mu_1, \mu_2, \mu_3, \mu_{12}, \mu_{13}, \mu_{23}$  en  $\mu_{123}$ . Gebruik makende van enkele bijzondere eigenschappen van de factoriele proefopzetten kan men ook ieder der  $\mu$ 's uitdrukken in de  $\eta_u$ . Zoals nl. uit de definitie van de  $x_{iu}$  volgt, kan de u-de rij van de programmamatrix (0,1) geschreven worden als

$$(x_{1u} \quad x_{2u} \quad x_{3u})$$

Hieruit blijkt dan, dat voor  $i = 1, 2, 3$  en  $j = 0, 1, 2$  en  $3$  geldt:

$$\left. \begin{aligned} \sum_{u=1}^8 x_{iu} &= 0 \\ \frac{1}{8} \sum_{u=1}^8 x_{ju}^2 &= 1 \end{aligned} \right\} \quad (1.4)$$

en als  $i \neq j$

$$\sum_{u=1}^8 x_{iu} x_{ju} = 0$$

Vermenigvuldigt men 1.3 rechts en links met en sommert men over u, dan is in verband met (1.4):

$$\begin{aligned} \sum_{u=1}^8 x_{iu} \eta_u &= \sum_{u=1}^8 (x_{0u} x_{iu} \mu_0 + x_{1u}^2 \mu_1 + x_{1u} x_{2u} \mu_2 + x_{1u} x_{3u} \mu_3 + \\ &+ x_{1u}^2 x_{2u} \mu_{12} + x_{1u}^2 x_{3u} \mu_{13} + x_{1u} x_{2u} x_{3u} \mu_{23} + \\ &+ x_{1u}^2 x_{2u} x_{3u} \mu_{123}) = \\ &= 8 \mu_1 \end{aligned}$$

Op analoge wijze verkrijgt men:

$$\mu_0 = \frac{1}{8} \sum_{u=1}^8 \eta_u$$

$$\begin{aligned}
 \mu_1 &= \frac{1}{8} \sum_{u=1}^8 X_{1u} \eta_u \\
 \mu_2 &= \frac{1}{8} \sum_{u=1}^8 X_{2u} \eta_u \\
 \mu_3 &= \frac{1}{8} \sum_{u=1}^8 X_{3u} \eta_u \\
 \mu_{12} &= \frac{1}{8} \sum_{u=1}^8 X_{1u} X_{2u} \eta_u \\
 \mu_{13} &= \frac{1}{8} \sum_{u=1}^8 X_{1u} X_{3u} \eta_u \\
 \mu_{23} &= \frac{1}{8} \sum_{u=1}^8 X_{2u} X_{3u} \eta_u \\
 \mu_{123} &= \frac{1}{8} \sum_{u=1}^8 X_{1u} X_{2u} X_{3u} \eta_u
 \end{aligned}
 \tag{1.5}$$

Men kan ook direct na (1.1) deze  $\mu$ 's volgens (1.5) definiëren en vervolgens aantonen dat (1.3) geldt.

## 2. Schattingen van de effecten

Uitgaande van de waargenomen waarden  $y_u$  zullen nu schattingen van de effecten worden berekend. Deze schattingen zullen wij met een  $m$  met overeenkomstige indices aangeven.

Allereerst zullen schattingen voor de  $\eta_u$  berekend worden, daar indien deze bekend zijn uit (1.5) de schattingen voor de effecten volgen. Volgens het principe der kleinste quadraten kiezen wij de schattingen  $y_u^*$  van  $y_u$  zodanig, dat

$$Q = \sum_{u=1}^8 (y_u - y_u^*)^2$$

minimaal wordt. Zoals zonder meer in te zien is, wordt  $Q$  minimaal (en wel gelijk aan 0), indien voor  $u = 1, \dots, 8$

$$y_u = y_u^* \tag{2.1}$$

Volgens (2.1) zijn de waarnemingen  $y_u$  dus de beste schattingen van  $\eta_u$ . Invoering in (1.5) geeft dan voor de schattingen van de  $\mu$ :

$$\begin{aligned}
 m_0 &= \frac{1}{8} \sum_{u=1}^8 y_u \\
 m_1 &= \frac{1}{8} \sum_{u=1}^8 X_{1u} y_u \\
 m_2 &= \frac{1}{8} \sum_{u=1}^8 X_{2u} y_u \\
 m_3 &= \frac{1}{8} \sum_{u=1}^8 X_{3u} y_u \\
 m_{12} &= \frac{1}{8} \sum_{u=1}^8 X_{1u} X_{2u} y_u \\
 m_{13} &= \frac{1}{8} \sum_{u=1}^8 X_{1u} X_{3u} y_u \\
 m_{23} &= \frac{1}{8} \sum_{u=1}^8 X_{2u} X_{3u} y_u \\
 m_{123} &= \frac{1}{8} \sum_{u=1}^8 X_{1u} X_{2u} X_{3u} y_u
 \end{aligned}
 \tag{2.2}$$

Uit deze uitdrukkingen blijkt dat de schattingen op zeer eenvoudige wijze te verkrijgen zijn. De  $m_1$  wordt bijvoorbeeld verkregen door de  $y_u$  een + of een - teken toe te kennen naar gelang de 1e factor op het hoge of het lage niveau bij het u-de experiment betrokken is, daarna over deze waarden te sommeren en door 8 te delen. De  $m_{12}$  wordt op analoge wijze verkregen; in dit geval krijgt de  $y_u$  een + of een - teken, al naar gelang  $X_{1u} X_{2u} + 1$  of  $-1$  is. De berekening van de schattingen verloopt dan ook zeer eenvoudig, indien men gebruik maakt van de programmatrix.

Daar alle waarnemingen een inherente waarnemingsonnauwkeurigheid bezitten, kan uit het simpele feit van het verkrijgen van een getalwaarde voor  $m_1, \dots, m_{123}$  nog niet besloten worden tot het bestaan van systematische hoofdeffecten en interacties. De variantieanalyse biedt evenwel een mogelijkheid dit nader te onderzoeken. Hiervoor wordt verwezen naar de handboeken op dit gebied o.a. [1], [2] en [5], waarin bovendien verdere literatuurverwijzingen zijn opgenomen. Indien bij toetsing bijvoorbeeld blijkt, dat behoudens een zekere (kleine) onbetrouwbaarheid aangenomen moet worden, dat er een interactie  $m_{12}$  aanwezig is als systematisch effect, dan wil dit zeggen, dat de eerste factor en de tweede factor niet onafhankelijk werken.

De schattingen, die op hierboven beschreven wijze berekend zijn, hebben enkele eigenschappen, die hier slechts kort zullen worden aangegeven. Zoals alle kleinste quadraten-schattingen zijn zij zuiver. Bovendien zijn zij onderling ongecorrleerd. Is bijvoorbeeld  $m_1$  zeer groot en  $m_2$  en  $m_3$  zeer klein, dan toch worden de schattingen van  $m_{12}$  etc. hier niet door beïnvloed. Deze eigenschap hangt zeer nauw samen met de orthogonaliteitsrelaties (1.4) en men zegt daarom wel dat alle schattingen "orthogonaal" zijn. Door deze eigenschappen en door hun eenvoud zijn factoriële schema's proefopzetten, die zeer bruikbaar zijn o.a. voor industriële experimenten.

De schattingen (2.2) krijgen een zeer eenvoudige vorm indien men van matrixnotatie gebruik maakt. Hiertoe definieert men de matrix X als volgt: de u-de rij van X bestaat uit de elementen ( $u = 1, \dots, 8$ )

$$\left( \begin{array}{cccccccc} X_{0u} & X_{1u} & X_{2u} & X_{3u} & X_{1u}X_{2u} & X_{1u}X_{3u} & X_{2u}X_{3u} & X_{1u}X_{2u}X_{3u} \end{array} \right)$$

De kolommen zullen wij aangeven met:

$$(E) \quad (1) \quad (2) \quad (3) \quad (12) \quad (13) \quad (23) \quad (123)$$

Uitgaande van de matrix P vinden wij dan:

$$X = \begin{matrix} & \begin{matrix} (E) & (1) & (2) & (3) & (12) & (13) & (23) & (123) \end{matrix} \\ \begin{matrix} +1 \\ +1 \\ +1 \\ +1 \\ +1 \\ +1 \\ +1 \\ +1 \end{matrix} & \begin{pmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & +1 & +1 & -1 & +1 & -1 & -1 & -1 \\ +1 & +1 & -1 & +1 & -1 & +1 & -1 & -1 \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ +1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ +1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 \end{pmatrix} \end{matrix} \quad (2.3)$$

Wij verstaan onder een vector een kolomvector en geven een rijvector aan als de getransponeerde kolomvector (dus als  $\hat{a}$ ).  
Definieren wij

$$\hat{\mu} = (\mu_0, \mu_1, \mu_2, \mu_3, \mu_{12}, \mu_{13}, \mu_{23}, \mu_{123})$$

terwijl de overeenkomstige schattingen aangegeven worden door

$$\hat{m} = (m_0, \dots, m_{123})$$

Is voorts de vector der waarnemingsresultaten:

$$\hat{y} = (y_1, \dots, y_8)$$

dan gaat (2.5) over in:

$$\hat{m} = \frac{1}{8} X' \hat{y} \quad 2.4$$

Bij een  $2^k$  experiment is het aantal uit te voeren experimenten zeer groot, als k niet al te klein is. Om het aantal experimenten binnen redelijke grenzen te houden kan men op de volgende wijze te werk gaan:

1. een  $2^{k-1}$  schema uitvoeren en de k-de factor door verstremgeling bij het experiment betrekken (Engels: confounding)
2. van een  $2^k$  schemaslechts een gedeelte, bijvoorbeeld de helft uit te voeren: dit leidt tot partiële schema's (Engels: fractional designs)

### 3. Verstremgeling

Stel men wenst de invloed van vier factoren te onderzoeken. Men kan dan een  $2^3$  factorieel schema uitvoeren, waar de 4e factor verstremgeld is met een van de interacties van de eerste drie. Men kan dit bijvoorbeeld doen door in de matrix P, een nieuwe kolom in te voegen waarvan ieder element hetzelfde teken heeft als het product van de elementen op diezelfde rij in de eerste drie kolommen. Men verkrijgt op deze wijze een nieuwe programmamatrix P'.

$$P' = \begin{pmatrix} +1 & +1 & +1 & +1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & +1 \\ +1 & -1 & -1 & -1 \\ -1 & +1 & +1 & +1 \\ -1 & +1 & -1 & -1 \\ -1 & -1 & +1 & +1 \\ -1 & -1 & -1 & -1 \end{pmatrix} \quad (3.1)$$

Bij het u-de experiment houdt men de vierde factor op het +1 (resp. -1) niveau als in de programmamatrix op de u-de rij en vierde kolom een +1, resp. -1 voorkomt.

Uit de wijze, waarop de vierde factor werd ingevoerd blijkt, dat voor elke u geldt:

$$x_{1u}^2 = x_{2u}^2 = x_{3u}^2 = x_{4u}^2 = x_{1u}x_{2u}x_{3u}x_{4u} = 1. \quad (3.2)$$

Men geeft dit aan door de z.g. groepsrelatie:

$$11 = 22 = 33 = 44 = 1\ 2\ 3\ 4 = \text{eenheid}. \quad (3.3)$$

Indien men de laatste gelijkheid in (3.2) links en rechts met  $x_{1u}$  vermenigvuldigt, is:

$$x_{1u}^2 x_{2u} x_{3u} x_{4u} = x_{1u}$$

of daar (weer volgens (3.2))  $x_{1u}^2 = \text{eenheid}$ , geldt:

$$\left. \begin{aligned} x_{2u} x_{3u} x_{4u} &= x_{1u} \\ x_{1u} x_{2u} x_{3u} &= x_{4u} \\ x_{1u} x_{2u} x_{4u} &= x_{3u} \\ x_{1u} x_{3u} x_{4u} &= x_{2u} \end{aligned} \right\} \quad (3.4)$$

en

Het model voor een volledig factorieel schema met 4 factor is: (naar analogie van 1.1 en 1.3)

$$\begin{aligned} \text{en} \quad y_u &= \eta_u + \underline{v}_u \\ \eta_u &= \mu_0 x_{0u} + \mu_1 x_{1u} + \mu_2 x_{2u} + \mu_3 x_{3u} + \mu_4 x_{4u} + \mu_{12} x_{1u} x_{2u} + \\ &+ \dots + \mu_{34} x_{3u} x_{4u} + \mu_{123} x_{1u} x_{2u} x_{3u} + \dots + \\ &+ \mu_{234} x_{2u} x_{3u} x_{4u} + \mu_{1234} x_{1u} x_{2u} x_{3u} x_{4u}. \end{aligned} \quad (3.5)$$

Uit de relaties (3.4) volgt, dat (3.5) voor het geval dat de vierde factor op de bovenbeschreven wijze geïntroduceerd is, ook kan worden geschreven als:

$$\begin{aligned} \eta_u &= (\mu_0 + \mu_{1234}) x_{0u} + (\mu_1 + \mu_{234}) x_{1u} + (\mu_2 + \mu_{134}) x_{2u} + \\ &+ (\mu_3 + \mu_{124}) x_{3u} + (\mu_{12} + \mu_{34}) x_{1u} x_{2u} + (\mu_{13} + \mu_{24}) x_{1u} x_{3u} + \\ &+ (\mu_{23} + \mu_{14}) x_{2u} x_{3u} + (\mu_{123} + \mu_4) x_{1u} x_{2u} x_{3u}. \end{aligned} \quad (3.6)$$

Op dezelfde wijze als in 2 kunnen de  $\mu$ 's uitgedrukt worden in de  $y$ 's:

$$\begin{aligned}
 \mu_0 + \mu_{1234} &= \frac{1}{8} \sum_{u=1}^8 \eta_u \\
 \mu_1 + \mu_{234} &= \frac{1}{8} \sum_{u=1}^8 X_{1u} \eta_u \\
 \mu_2 + \mu_{134} &= \frac{1}{8} \sum_{u=1}^8 X_{2u} \eta_u \\
 \mu_3 + \mu_{124} &= \frac{1}{8} \sum_{u=1}^8 X_{3u} \eta_u \\
 \mu_{12} + \mu_{34} &= \frac{1}{8} \sum_{u=1}^8 X_{1u} X_{2u} \eta_u \\
 \mu_{13} + \mu_{24} &= \frac{1}{8} \sum_{u=1}^8 X_{1u} X_{3u} \eta_u \\
 \mu_{23} + \mu_{14} &= \frac{1}{8} \sum_{u=1}^8 X_{2u} X_{3u} \eta_u \\
 \mu_{123} + \mu_4 &= \frac{1}{8} \sum_{u=1}^8 X_{1u} X_{2u} X_{3u} \eta_u
 \end{aligned}
 \tag{3.7}$$

Dus in dit geval zal men, indien men voor  $\eta_u$  de schattingen  $y_u$  invoert, volgens 3.7 niet alle  $\mu$ 's apart schatten, maar twee aan twee bij elkaar.

Zo bijvoorbeeld schat men de som van  $\mu_1 + \mu_{234}$ , en men kan uit deze acht experimenten niets meer te weten komen over  $\mu_1$  en  $\mu_{234}$  apart. Men zegt dan dat  $\mu_1$  en  $\mu_{234}$  verstrengeld zijn. Men ziet dus uit 3.7 dat elk effect verstrengeld is met een ander effect. Door vermenigvuldiging van de indices van een effect met de groepsrelatie (3.3), verkrijgt men de indices van het effect, dat ermee verstrengeld is. Zo is bijvoorbeeld  $\mu_{12}$  verstrengeld met  $\mu_{34}$ , daar  $12 \cdot 1234 = 34$ .

Het gebruik van verstrengeling is meestal dan slechts zinvol, als van de met de hoofdeffecten verstrengelde interacties op redelijke gronden aangenomen mag worden, dat zij zeer gering of afwezig zijn.

Het zou in dit verband te ver voeren, de verschillende methoden van verstrengeling te bespreken. Men wordt hiervoor verwezen naar de bestaande handboeken.

#### 4. Partiële schema's

Men komt tot dezelfde matrix  $P'$  (2.1), indien van een  $2^4$  factoriële proefopzet slechts die experimenten uitgevoerd worden waarvoor:

$$X_{1u} X_{2u} X_{3u} X_{4u} = +1, \tag{4.1}$$

en die experimenten achterwege laat waarvoor:

$$X_{1u} X_{2u} X_{3u} X_{4u} = -1$$

In dit geval noemt men dit een half factorieel schema. Indien men op de door (4.1) gedefinieerde wijze te werk gaat, blijven de geschatte effecten orthogonaal.

De consequenties van het uitvoeren van een partieel schema t.a.v. de schattingen zijn geheel gelijk aan die, besproken in 3. Zo volgt uit 4.1 weer de groepsrelatie:

$$1234 = \text{eenheid.}$$

## 5. Literatuur

- 1 ANDERSON, R.L. and T.A. BANCROFT: Statistical theory in research. McGrawHill, New York 1952.
- 2 COCHRAN, W.G. and G.M. COX: Experimental designs, Wiley, New York, 1950.
- 3 DAVIES, O.L. and W.A. HAY: The construction and use of fractional designs in industrial research. Biometrics 1950, 6, 233-249.
- 4 DAVIES, O.L. Design and analysis of industrial experiments (editor) Oliver and Boyd, Edinburgh 1953.
- 5 MANN, H.B. Analysis and design of experiments. Dover publications, New York, 1949.