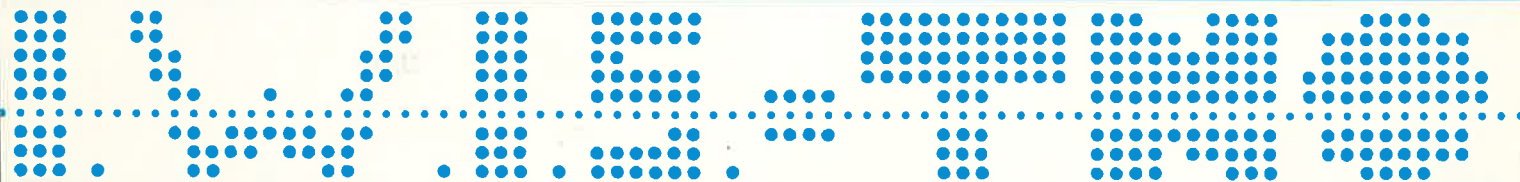


25 jaar
IWIS·TNO
afdeling
Wageningen

Enkele voorbeelden
van de toepassing
van statistiek
en informatica
bij landbouwkundig
onderzoek



25 JAAR IWIS-TNO, AFDELING WAGENINGEN

Enkele voorbeelden van de toepassing van statistiek en
informatica bij landbouwkundig onderzoek

Samengesteld onder redactie van Drs. J. de Bree

I N H O U D

Voorwoord	3
De methode van Beale en Little, toegepast op tijdreeksen waarin sommige waarnemingsuitkomsten ontbreken Drs. L.G. Barendregt	5
Principal components biplots and alpha and beta diversity Drs. C.J.F. ter Braak	17
'Most Probable Number' en Zaadkeuring Drs. J. de Bree	39
Informatica en het landbouwkundig onderzoek Drs. T. de Heer	55
Landbouw in de statistiek Ir. A.A.M. Jansen	67
Enkele aspecten van de analyse van groeicurven Ir. J. Jansen	87
Een toets op ruimtelijke autocorrelatie in rechthoekige roosters Dr. Michiel J.W. Jansen	99
De optimale opzet van een proef die wordt uitgevoerd om voor gedroogde produkten de effectiviteit van straling onder verschillende omstandigheden te schatten en te vergelijken Ir. A. Keen	115
Vogelbevolking en vegetatiestructuur Drs. J.H. Oude Voshaar	139
Onderzoek naar het effect van vitaminesuppletie op het prestatievermogen van CIOS-studenten Ir. J.Th.N.M. Thissen	163
Eischaalbeschadigingen op legbatterijen Ir. P.F.G. Vereijken	201

Voorwoord

De bedoeling van het boek dat voor u ligt is een indruk te geven van het dagelijks werk van IWIS-TNO, afdeling Wageningen.

Niet al het werk dat IWIS-Wageningen verricht is vertegenwoordigd; met name ontbreekt het werk van de programmeursgroep en van de computersectie. Het heeft weinig zin om aan dit boek een fraai computerprogramma toe te voegen, ook zonder dit zal iedereen begrijpen dat programmeurs en computermedewerkers substantieel bijdragen aan het werk van IWIS.

In de opgenomen artikelen komen drie belangrijke aspecten van het werk van de IWIS-statistici aan bod.

Het artikel van Ir. A.A.M. Jansen "Landbouw in de Statistiek" representeert de voorlichting, een activiteit zonder welke geen goede samenwerking met de onderzoeker tot stand kan worden gebracht en kan worden gehandhaafd. Ook het artikel van Drs. T. de Heer "Informatica en het Landbouwkundig Onderzoek" is hiervan een voorbeeld.

De artikelen van:

Drs. L.G. Barendregt: "De methode van Beale en Little, toegepast op tijdreeksen waarin sommige waarnemingsuitkomsten ontbreken",

Dr. M. Jansen: "Toepassing op ruimtelijke autocorrelatie in rechthoekige roosters",

Ir. A. Keen: "De optimale opzet van een proef die wordt uitgevoerd om voor gedroogde produkten de effectiviteit van straling onder verschillende omstandigheden te schatten en te vergelijken",

Drs. J.H. Oude Voshaar: "Vogelbevolking en vegetatiestructuur",

Ir. J.Th.N.M. Thissen: "Onderzoek naar het effect van vitaminesuppletie op het prestatievermogen van CIOS-studenten" en

Ir. P.F.G. Vereijken: "Eischaalbeschadiging op legbatterijen",

vertegenwoordigen het meest voorkomende statistische werk: het opstellen van een wiskundig-statistisch model, het ontwerpen van een proefopzet en de statistische analyse van waarnemingsresultaten.

Soms ligt hierbij de nadruk meer op de proefopzet, dan weer meer op de analyse.

In de consultatie blijkt vaak dat bestaande statistische methoden niet zonder meer voor een gegeven probleem kunnen worden gebruikt. Het is

daarom dikwijls nodig om reeds bestaande methoden en technieken aan te passen, te modificeren of uit te diepen. De artikelen van:

Drs. C.J.F. ter Braak: "Principal components biplots and alpha and beta diversity",

Drs. J. de Bree: "Most Probable Number en zaadkeuring" en

Ir. J. Jansen: "Enkele aspecten van de analyse van groeicurven"
geven hiervan een voorbeeld.

Het merendeel van de in dit boek opgenomen artikelen zijn normale IWIS-rapporten, zoals deze worden uitgebracht aan opdrachtgevers.

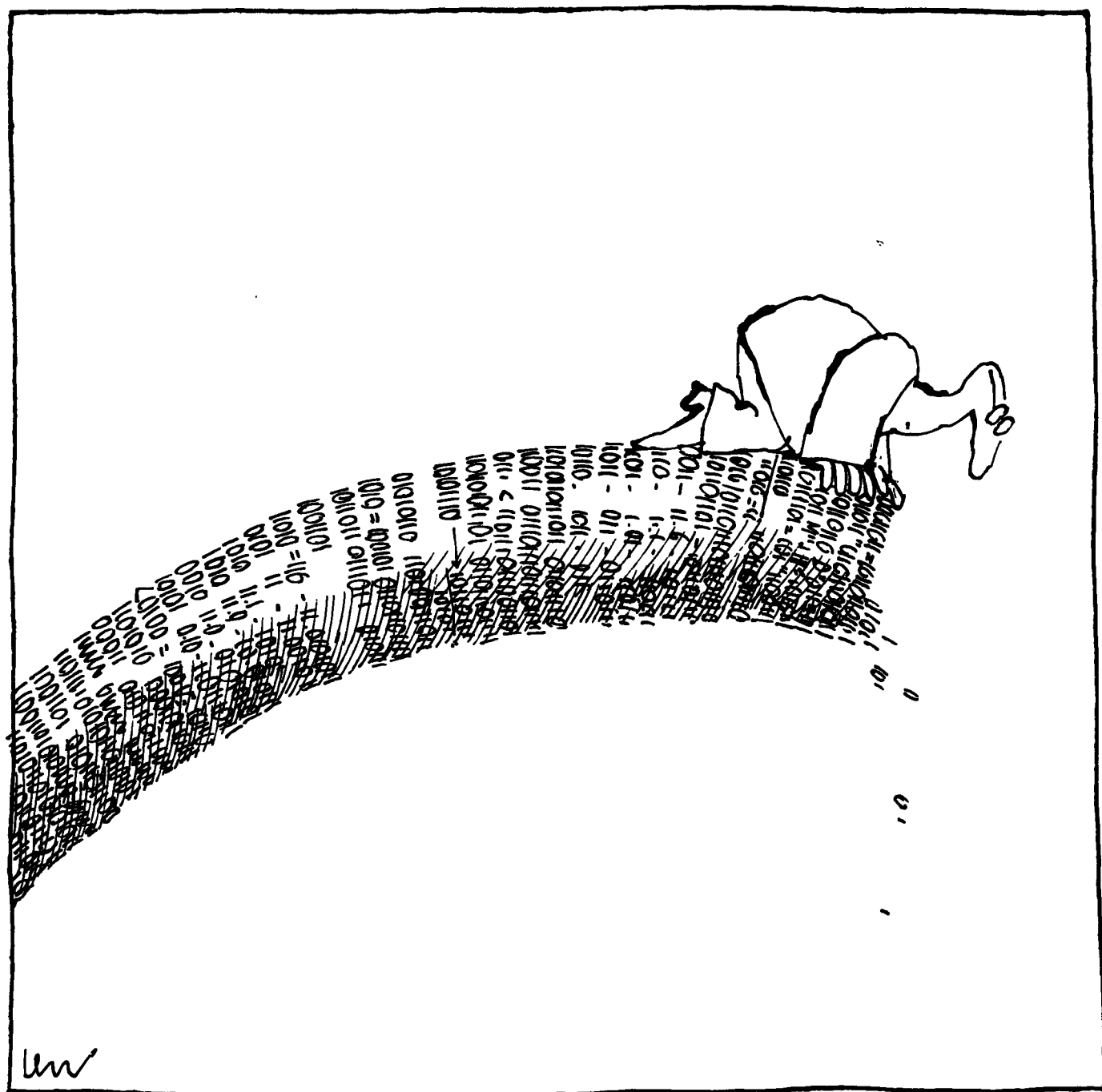
Voor het feit dat wij ze ook voor dit boek mochten gebruiken betuigen wij de opdrachtgevers gaarne onze dank.

Wij allen hopen dat deze artikelen iets zullen bijdragen aan het onderling begrip dat voor een goede samenwerking tussen onderzoeker en statisticus is vereist.

De directie

De methode van Beale en Little, toegepast op tijdreeksen waarin sommige waarnemingsuitkomsten ontbreken

Drs. L. G. Barendregt



lgr

Inhoud

1.	Inleiding	9
2.	Beschrijving van het waarnemingsmateriaal	9
3.	Multivariate variantie-analyse	10
4.	De methode van Beale en Little	11
5.	Toepassing van de methode Beale en Little op de Hb-gehaltenes	12
6.	Berekeningsresultaten	14
7.	Dankwoord	15
	<i>Referenties</i>	15
	Bijlage I	16

1. Inleiding

Waarnemingen van een meetgrootheid, die op elk van T tijdstippen aan elk van N individuen zijn gedaan, zijn in aanleg geschikt voor groeikromme-analyses. Een vervelend aspect van dergelijke waarnemingsuitkomsten is echter, dat er altijd wel enige van ontbreken. Die ontbrekende waarden komen her en der in het materiaal voor. De oorzaak van het ontbreken is vaak banaal; maar het uitvoeren van groeikrommeanalyse die de statisticus in gedachten had, wordt er wèl door bemoeilijkt.

Een soortgelijk probleem is er bij de multivariate analyses van waarnemingen van enquêtes, gedaan in het kader van een of ander gedragswetenschappelijk onderzoek. Daar gaat het om p variabelen die aan elk van n individuen zijn gemeten.

Voor het probleem van het schatten van de variantie-covariantie-matrix van de p variabelen, als er waarnemingsuitkomsten ontbreken, is een methode voorgesteld, die in ons land bekend staat als de methode van Beale en Little. Dit is, kort gezegd, een maximum-likelihood-schatting onder de veronderstelling van multivariate normaliteit; bij de berekening wordt een elegant en redelijk snel convergerend iteratieproces gebruikt.

In dit rapport zal ik beschrijven hoe de methode van Beale en Little kan worden toegepast bij het analyseren van tijdreeksen met ontbrekende waarnemingen.

2. Beschrijving van het waarnemingsmateriaal

De afdeling Parasitologie van het Centraal Diergeneeskundig Instituut te Lelystad heeft van 1977 tot en met 1980 een veldonderzoek gedaan naar de betekenis van een serologische titer (gemeten met de indirecte haemagglutinatietest = IHA) voor de vroeg-diagnose van leverbotbesmetting bij runderen. Op 12 rundveebedrijven zijn tien maal per jaar, gedurende drie jaar van dertig runderen bloedmonsters en mestmonsters genomen. Aan de bloedmonsters werden diverse waarden bepaald, zoals:

aantal leucocyten per mm³
leucocyten-differentiatie
hemoglobine-gehalte (Hb-gehalte)
indirecte haemagglutinatie-test (IHA)

Aan de mestmonsters werd het aantal leverbot-eieren per gram mest bepaald. Uit het aldus ontstane waarnemingsmateriaal heb ik de volgende getallenverzameling ontleend.

De dieren waaraan was gemeten, zijn kalveren op twee bedrijven, hierna te noemen bedrijf 1 en bedrijf 2. Op bedrijf 1 zijn aan tien kalveren metingen gedaan; op bedrijf 2 aan zeven kalveren. Van de in totaal zeventien kalveren heb ik de Hb-gehaltes beschouwd; aan elk kalf is het Hb-gehalte op tien tijdstippen gemeten. Deze tijdstippen zijn: de maanden mei t/m november 1977 en januari t/m maart 1978. De waarnemingsuitkomsten staan vermeld in de tabel van bijlage 1. Een aantal van deze uitkomsten is onbekend; zij zijn aangegeven door een liggend streepje.

3. Multivariate variantie-analyse

De Hb-gegevens, die in paragraaf 2 zijn beschreven, lenen zich voor een multivariate variantieanalyse.

Er zijn immers 17 individuen (namelijk de kalveren), aan elk waarvan tien variabelen zijn gemeten (de Hb-waarden op de tien tijdstippen).

Het is nu mogelijk om met een multivariate variantieanalyse te toetsen, of de kalveren van de twee bedrijven significant van elkaar verschillen ten aanzien van het gemiddeld niveau van één of meer van de tien Hb-waarden.

De toetsingsgrootte voor dit probleem is:

$$F \equiv \frac{n_1 n_2}{n_1 + n_2} \frac{n_1 + n_2 - p - 1}{n_1 + n_2 - 2} \frac{1}{p} \mathbf{d}' \mathbf{S}^{-1} \mathbf{d} \quad (3.1)$$

met

n_1 = aantal individuen in groep 1 (hier 10)

n_2 = aantal individuen in groep 2 (hier 7)

p = aantal variabelen (hier 10)

$\underline{d} = (d_1, d_2, \dots, d_p)'$ en waarin

d_j = het verschil tussen het gemiddelde van de j -de variabele bij groep 1 en het overeenkomstige gemiddelde bij groep 2. ($j = 1, \dots, p$)

$S = (S_{jk})$ $j, k = 1, \dots, p$ en waarin

S_{jk} = de gepoolde binnen-groepen-covariantie tussen de j -de en de k -de variabele.

4. De methode van Beale en Little

De methode van Beale en Little is bedoeld om uit een steekproef van n onafhankelijke trekkingen uit een p -variante normale verdeling de verwachtingswaarden en de variantie-covariantie-matrix van die verdeling te schatten, als sommige van de $n \times p$ waarnemingsuitkomsten ontbreken. Voorwaarde is wèl, dat het ontbreken van een bepaalde waarnemingsuitkomst niet samenhangt met de grootte van die of enige andere waarnemingsuitkomst. De log-likelihood-functie die hierbij wordt gemaximaliseerd, is afgeleid van de marginale kansdichtheidsfunctie van de bekende waarnemingsuitkomsten.

In formuletaal komt dit op het volgende neer:

Laat de gegevens-matrix zijn.

$x_{11} \text{ --- } x_{1p}$

| |

| |

$x_{n1} \text{ --- } x_{np}$

x_{ij} = waarnemingsuitkomst van de j -de variabele aan het i -de individu.

$i = 1, \dots, n; j = 1, \dots, p.$

Laat $x_i = (x_{i1}, \dots, x_{ip})'$ de waarnemingsvector van het i -de individu zijn.

Van de p getallen x_{i1}, \dots, x_{ip} kan slechts een gedeelte bekend zijn. Als r_i het aantal bekende gegevens van het i -de individu is, dan is er een $(r_i \times p)$ -matrix D_i zó, dat de r_i -dimensionale vector $D_i x_i$ juist de r_i bekende gegevens van het i -de individu als elementen heeft. D_i bestaat dan uit de getallen 0 en 1; in elke rij is er één 1 en in elke kolom hoogstens één 1. De bekende gegevens zijn dus:

$$D_1 x_1, D_2 x_2, \dots, D_n x_n.$$

De te maximaliseren log-likelihood-functie is nu:

$$\log L(\mu, \Sigma) = -\frac{1}{2} \sum_{i=1}^n \log \det (D_i \Sigma D_i')$$
$$-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)' D_i' (D_i \Sigma D_i')^{-1} D_i (x_i - \mu)$$

waarbij μ de verwachtingswaarde-vector en Σ de dispersie-matrix van de p variabelen is.

Het bijzondere van de methode van Beale en Little is, dat er gebruik wordt gemaakt van een elegante iteratieve procedure om $\log L(\mu, \Sigma)$ te maximaliseren. Zie hiervoor Beale and Little (1975).

5. Toepassing van de methode Beale en Little op de Hb-gehaltenes

De Hb-waarden van de tabel in bijlage 1 zijn niet alle bekend. Er mag echter vanuit worden gegaan dat het onbekend zijn van een waarneming niet is gecorreleerd met de grootte van die waarneming of met de grootte van enige andere waarneming. Aan een van de voorwaarden voor de toepasbaarheid van de methode van Beale en Little is dus voldaan.

In deze paragraaf zal worden aangegeven, hoe de multivariate variantie-analyse, zoals beschreven in paragraaf 3, ondanks het ontbreken van sommige waarnemingen toch kan worden uitgevoerd. Hierbij zal worden gebruik gemaakt van het feit dat de methode van Beale en Little ruimer toepasbaar is dan de beschrijving van paragraaf 4 doet vermoeden.

Little heeft namelijk opgemerkt, dat het voor de toepasbaarheid van die methode niet noodzakelijk is dat de p variabelen uit een p -variante normale verdeling afkomstig zijn (Little, 1979). Ze is ook nog toepasbaar als sommige van die variabelen niet stochastisch zijn; er moet dan wèl gelden:

- a. De waarden van die variabelen zijn voor alle individuen bekend.
- b. De voorwaardelijke verdeling van de overige variabelen is multivariaat normaal.

Deze situatie kunnen we in het bestand van de Hb-gehaltenes laten ontstaan door een elfde (dichotome) variabele in te voeren, die het bedrijf aangeeft, waartoe het kalf behoort. Aan elk van de zeventien kalveren wordt dus een getal $+1$ of -1 toegekend, namelijk -1 voor de kalveren uit bedrijf 1 en $+1$ voor de kalveren uit bedrijf 2.

Op de gegevens van 11 variabelen (tien Hb-gehaltenes en één bedrijfsvariabele bij zeventien kalveren kan de methode van Beale en Little worden toegepast. De 11×11 variantie-covariantie-matrix die daaruit ontstaat is op zich niet interessant. Wèl kan daaruit worden berekend zowel een schatting voor het verschil van de verwachtingswaarde-vectoren van beide bedrijven als ook een schatting voor de dispersie-matrix van de tien HB-gegevens bij één bedrijf. Dit gaat als volgt:

Laat M de (11×11) -matrix zijn, die het resultaat is van de Beale-en-Little-berekening op de Hb-gehaltenes en de bedrijfsvariabele. Noem de elementen van M M_{ij} . De nummers 1 t/m 10 hebben betrekking op de tien Hb-gehaltenes, het nummer 11 heeft betrekking op de bedrijfsvariabele.

Nu is

$$S_{jk} = M_{jk} - \frac{M_{j,11} M_{k,11}}{M_{11,11}} \quad j, k = 1, \dots, 10$$

een schatting voor de covariantie tussen het Hb-gehalte op tijdstip j en het Hb-gehalte op tijdstip k binnen een bedrijf.

Verder is

$$d_j = \frac{M_{j,11}}{M_{11,11}} \quad j = 1, \dots, 10$$

een schatting voor het verschil tussen de verwachtingswaarden van de HB-gehalten op tijdstip j bij beide bedrijven.

De tien-dimensionale vector $(d_1, \dots, d_{10})'$ en de (10×10) -matrix S kunnen worden gesubstitueerd in formule (3.1). De verdeling van de grootheid is niet exact bekend. Dit vloeit voort uit het feit dat de verdeling van bijna alle berekeningsresultaten van de methode van Beale en Little slechts voor heel speciale gevallen exact bekend is. Voor het algemene geval zijn slechts simulatieresultaten bekend.

6. Berekeningsresultaten

De toepassing van de methode van Beale en Little op de tien Hb-waarden en bedrijfsvariabele leverde geen bevredigend resultaat op: pas na 189 iteratiestappen werd een enigszins stabiele waarde voor M bereikt.

M bleek toen "bijna singulier" te zijn.

Omdat ik vermoedde dat dit werd veroorzaakt door de vele ontbrekende waarnemingen van tijdstip 8, heb ik de berekening ook toegepast op de Hb-waarden van de tijdstippen 1 t/m 7, 9 en 10 aangevuld met de bedrijfsvariabele. Nu convergeerde het iteratie-proces van Beale en Little vrij snel.

De waarden voor d_1 t/m d_9 waren alle positief, hetgeen duidde op een grotere Hb-waarde bij bedrijf 1 dan bij bedrijf 2, en dit op elk van de negen tijdstippen.

Voor F werd gevonden: 245.35

Hoewel de verdeling van F niet exact bekend is, mag toch wel worden vermoed dat deze niet al te zeer afwijkt van een F -verdeling met negen en zeven graden van vrijheid. Het 95 procent punt van deze F -verdeling is 3.29, zodat men veilig kan besluiten dat de bedrijven verschillen.

7. Dankwoord

Ik ben mevr. Drs. Y. Wetzlar en Dr. H.J. Over van de afdeling Parasitologie van het Centraal Diergeneeskundig Instituut in Lelystad dank verschuldigd voor hun toestemming om de Hb-gegevens van hun veldproef te gebruiken voor deze publikatie.

Referenties

- 1 Beale, E.M.L. and Little, R.J.A. (1975); Missing values in multivariate analysis; Journal of the Royal Statistical Society B, blz. 129-145

- 2 Little, R.J.A. (1979); Maximum Likelihood Inference for Multiple Regression with Missing Values: A Simulation Study; Journal of the Royal Statistical Society B; blz. 76-87.

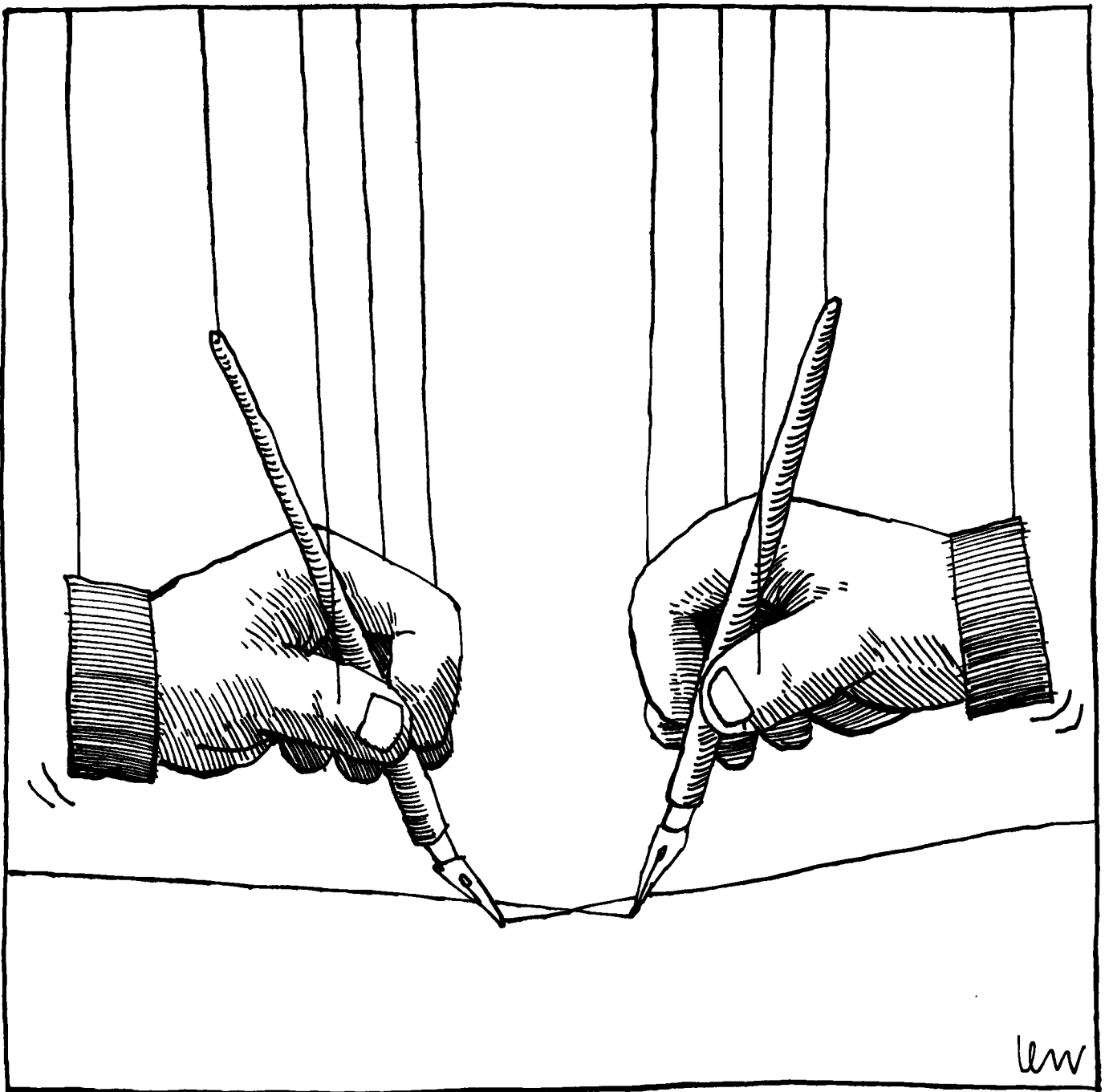
Bijlage I

Tabel van haemoglobin-gehaltenes

bedrijf	kalf nr.	tijdstoppen									
		1	2	3	4	5	6	7	8	9	10
1	1	10.2	10.8	10.1	-	11.2	7.5	9.4	9.1	9.6	8.5
	2	10.2	11.3	10.3	10.3	9.1	9.5	10.4	7.6	9.4	10.5
	3	9.3	11.4	9.9	9.8	9.1	9.6	10.7	7.6	10.2	11.4
	4	9.2	11.7	11.0	9.6	8.9	9.7	9.1	7.2	8.4	10.3
	5	10.7	11.0	9.7	9.2	9.3	10.2	10.5	10.5	-	-
	6	10.2	12.0	11.6	9.1	10.0	9.7	9.6	7.6	9.5	9.4
	7	10.0	10.2	10.5	10.7	9.5	9.1	9.8	7.2	8.8	8.6
	8	9.4	9.8	9.5	11.2	8.9	9.3	9.0	5.9	9.5	9.4
	9	9.9	9.7	9.0	10.5	9.0	9.1	8.7	6.4	8.7	8.6
	10	-	9.5	9.8	8.7	8.5	8.5	-	8.3	9.1	9.8
2	11	11.0	12.4	13.1	14.4	14.3	10.8	10.8	-	10.2	8.7
	12	10.2	13.2	11.3	12.0	10.3	11.2	10.6	9.7	8.9	8.9
	13	13.1	13.2	13.2	14.4	12.2	11.4	11.1	-	10.8	-
	14	14.7	12.8	11.0	13.6	11.0	12.8	12.0	-	11.0	10.4
	15	13.5	12.7	-	13.5	11.8	12.4	12.2	11.1	9.9	10.2
	16	13.3	-	12.4	11.4	10.5	-	-	-	-	-
	17	13.7	11.6	12.9	13.4	12.3	12.5	13.3	12.0	11.7	10.7

Principal components biplots and alpha and beta diversity

Drs. C. J. F. ter Braak



Contents

	Abstract	21
1.	Introduction	21
2.	Theory and methods	22
3.	Results	28
4.	Discussion	30
5.	Acknowledgements	32
6.	Literature cited	32

Abstract

Noncentred principal components analysis (PCA) on proportion-data gives site ordinations that display approximate alpha diversities of sites and beta diversities of groups of sites, as measured by Simpson index and mean squared Euclidean distance. Species-centring gives a better approximation to beta diversities; alpha diversities can still be visualized, if the true origin is projected into the plane of the ordination. The approximate species-composition of each site can be visualized if the site ordination is combined with a species ordination. The resulting plot of site scores and species loadings is called a biplot. In a biplot that displays both species-composition and diversity, diversity values can be associated with the names of the main contributing species. In such a biplot the sum of squares of the species loadings must be scaled to unity. This type of biplot is termed a 'distance biplot'. Noncentred and species-centred distance biplots were produced from diatom samples taken from Dutch moorland pools in the 1920's and 1978. These plots show the approximate change in the pools since the 1920's: impoverishment in species-composition and decrease in alpha diversity of pools as well as in beta diversity among pools. Species-centred PCA yielded more accurate approximations. Distortion was least in low-diversity samples.

Key words: biplot; dissimilarity; diversity; graphical representation; ordination; principal components analysis; Simpson index; species-composition.

1. Introduction

Simultaneous site and species ordinations are not unique to reciprocal averaging (Hill 1973a). Gabriel (1971) has shown that in ordinary principal components analysis (PCA) both objects and variables, i.e. sites and species, can be combined into one plot: the biplot. The biplot is a powerful graphical tool.

Diversity has an elegant mathematical theory (Engen 1978, Patil and Taillie 1979, Kempton 1979) and an elegant ecological theory as well

(MacArthur and Wilson 1967, Pielou 1975, Whittaker 1972, May 1973, Connell and Orias 1964). But diversity measures suffer from a serious limitation: that their values are not associated with the names of the contributing species. Thus sites with totally different species-lists may have the same diversity. We would like to know which species are contributing most to the diversity of each site.

This paper presents a method, based on PCA biplots, to display sites and species in such a way that the diversity values of sites can be associated with names of species. For the sake of lucidity I shall concentrate on measures related to the Simpson index (Simpson 1949). The method is illustrated with data on diatom assemblages of six Dutch moorland pools in the 1920's and in 1978, collected in connexion with a project of the Research Institute for Nature Management (The Netherlands) to investigate the impact of acidification on these shallow water bodies (van Dam et al. in press). The samples include eighty-two species; data on the eight most abundant (mean proportion > 0.02) appear in Table 1.

2. Theory and methods

Geometrical representation of proportions and diversity

Suppose a survey of n sites lists the proportions of m species. Let p_{ki} be the proportion of individuals of species k at site i with respect to all individuals at site i . Each of the n sites may now be thought of as a vector, p_i , in a m -dimensional Euclidean space: $p_i = (p_{1i}, \dots, p_{mi})'$. (The axes or 'basis vectors' of this space are $e_1 = (1, 0, 0, \dots, 0)'$, $e_2 = (0, 1, 0, \dots, 0)'$, \dots , $e_m = (0, 0, 0, \dots, 1)'$, each basis vector (say e_k) being associated with a species (k). Formally, p_{ki} is the inner product of p_i and e_k , i.e. the product of their lengths and the cosine of the angle between them: this observation is relevant to the understanding of biplots.)

The squared length of each site vector p_i is given by

$$\|p_i\|^2 = \sum_{k=1}^m p_{ki}^2, \quad (1)$$

in other words, by the Simpson index (Simpson 1949), henceforth abbreviated to SI. SI can easily be transformed into a useful measure of alpha (within-site) diversity: for example $1-SI$ (Pielou 1969) or $1/SI$ (Hill 1973b).

Squared Euclidean distance between sites,

$$d_{ij}^2 = \sum_{k=1}^m (p_{ki} - p_{kj})^2, \quad (2)$$

may be regarded as a measure of beta (between-site) diversity (Whittaker 1960). Here its use is not restricted to coenoclines. 'Beta or between-habitat diversity refers to the . . . degree of contrast in species composition among samples of a set taken from a landscape.' (Whittaker 1973, p. 41). I define the beta diversity of two samples as the dissimilarity between them. If there are more than two samples, their mean dissimilarity is a measure of their beta diversity.

Noncentred PCA

The geometrical configuration of the sites has now been completely specified in terms of alpha and beta diversity measures. But this configuration has too many dimensions for convenience. We would like a two- or three-dimensional plot representing as accurately as possible the original configuration. We are anxious to preserve both the lengths of site vectors and the distances between them: we therefore need to know the position of the 'true' origin, i.e. a notional site without species, in the plot. Noncentred PCA guarantees this (Noy-Meir 1973). Noncentred PCA simply projects the sites on to the plane of best fit through the true origin.

Algebraically PCA proceeds as follows (see e.g. Jöreskog et al. 1976). Let the species \times site matrix Y contain the values y_{ki} , i.e. $Y = (y_{ki})$, or for our purposes the proportions, p_{ki} . The eigenvalues ($\lambda_1^2 > \lambda_2^2 > \dots$) and eigenvectors (u_1, u_2, \dots) of the site \times site matrix $Q = Y'Y$ contain the information needed for the plot. If the r^{th} eigenvector (r^{th} principal component) has unit length and coordinates $u_r = (u_{1r}, u_{2r}, \dots, u_{nr})'$, then the position of site i in a two-dimensional plot is given by the vector $g_i = (\lambda_1 u_{1i}, \lambda_2 u_{2i})'$. The coordinates of a site in this plot are usually termed scores on the first and second principal component.

An alternative way of computing the scores ('R-mode') is more economical if there are more sites than species, since the species-
 × species matrix $R = YY'$ has the same eigenvalues as Q , and its
 eigenvectors $(\mathbf{v}_1, \mathbf{v}_2, \dots)$ are related to the eigenvectors of
 Q : $\lambda_r \mathbf{u}_r = Y'\mathbf{v}_r$ and $\lambda_r \mathbf{v}_r = Y\mathbf{u}_r$ ($r = 1, 2, \dots$).

Again, \mathbf{v}_r has unit length and its coordinates are denoted by

$\mathbf{v}_r = (v_{1r}, v_{2r}, \dots, v_{mr})'$. The vector $\mathbf{h}_k = (v_{k1}, v_{k2})'$ is associated
 with species k and contains the loadings of species k on the first and
 second component. We do not need these for the present, but we shall
 need them in the construction of biplots.

The result of the above procedure is an ordination plot in which
 sites with high alpha diversity (low SI) are near the origin and sites
 with low alpha diversity (high SI) are far from the origin. Sites that
 are similar in species-composition, i.e. with low beta diversity as
 measured by mean squared Euclidean distance, will be close together,
 while sites that differ in species-composition will be more spread out -
 so long as the reduction in dimensionality has not resulted in too much
 distortion of the original configuration.

It happens that noncentred PCA on proportion-data allows a simple
 test for distortion. If two site vectors point in the same direction,
 they must also have the same length, since the coordinates of each site
 sum to unity. The more this relation is violated in the ordination plot,
 the more the original configuration has been distorted (type B distor-
 tion sensu Orłóci 1978, p. 182). If severe type B distortion is evident,
 then further dimensions should be inspected. Finally note that since PCA
 is a least-squares technique (Gabriel 1971, 1978), numerically high
 values will be proportionally best approximated. The difference between
 high and low diversity will therefore be accentuated.

The biplot

Distances between sites in an ordination plot are, of course, due
 to differences in species-composition (see (2)). These differences in
 species-composition can be displayed explicitly, as they were in the
 original m -dimensional space, by combining the site ordination with a
 species ordination. The site ordination resulted from projection of the
 sites on to an optimal plane. The appropriate species ordination now

follows from projection of the basis vectors of the m -dimensional space on to this plane, for these vectors are the ones that originally represented the species. The result is a principal components biplot (Gabriel 1971). The coordinates of species k in the biplot become $\mathbf{h}_k = (v_{k1}, v_{k2})'$, i.e. the loadings of species k as defined in the previous section. A biplot of three or more dimensions can be constructed in a similar way. Thus the biplot simply combines loadings and scores.

I shall now consider some mathematical properties of biplots. Remember that in the m -dimensional space the proportions were equal to the inner products between sites and species. Sites and species are now both projected on to an optimal plane; their inner products in that plane approximate the original proportions. In a biplot derived from a matrix $\mathbf{Y} = (y_{ki})$, let site i have vector $\mathbf{g}_i = (g_{i1}, g_{i2})'$ and let species k have vector $\mathbf{h}_k = (h_{k1}, h_{k2})'$.

Then their inner product approximates y_{ki} by, say, $y_{ki(2)}$. Now let θ be the angle between \mathbf{g}_i and \mathbf{h}_k and let $\sigma = 1$ if $\cos(\theta) > 0$ and -1 if $\cos(\theta) < 0$:

$$\begin{aligned}
 y_{ki(2)} &= h_{k1} g_{i1} + h_{k2} g_{i2} \\
 &= \|\mathbf{h}_k\| \|\mathbf{g}_i\| \cos(\theta) \\
 &= \sigma \|\text{projection of } \mathbf{g}_i \text{ on to } \mathbf{h}_k\| \|\mathbf{g}_i\| \\
 &= \sigma \|\mathbf{h}_k\| \|\text{projection of } \mathbf{h}_k \text{ on to } \mathbf{g}_i\|. \qquad (3)
 \end{aligned}$$

The first formula is the algebraic definition of an inner product; the others give more geometrical insight. In particular (given that we are working with proportions, $y_{ki} = p_{ki}$) note that the species-composition of a site can be approximated by projecting all species on to the appropriate site vector; and that the proportions of a species at all sites can be approximated by projecting the sites on to the appropriate species vector.

From (3) observe also that the scaling of a biplot is not necessarily unique, e.g. if g_{i1} is multiplied by a constant and h_{k1} is divided by the same constant for all i and k , then $y_{ki(2)}$ remains unchanged. However in the previous section it was required that lengths of and distances between sites in the biplot should approximate SI and squared Euclidean distance values. This constraint puts a fixed scaling on the biplot: the loadings on each principal component must have sum of squares equal to unity; the scores must have sum of squares equal to the corresponding eigenvalue. I propose calling this type of biplot a 'distance biplot', to distinguish it from the 'covariance biplot' considered in detail by Corsten and Gabriel (1976).

Diversity values can now be associated with names of species. By and large it is the species with large vectors pointing in roughly the same direction as a site vector that are present in high proportions at that site. A site with low diversity will have one or two long species vectors pointing in its direction. A site with high diversity will have several, shorter species vectors that point in its direction. The biplot thus displays which species make a site as diverse as it is.

Because of the least-squares properties of PCA high proportions will be approximated proportionally better than low proportions. The overall approximation error in a biplot of t dimensions can be defined as

$$\sum_{i=1}^n \sum_{k=1}^m (y_{ki} - y_{ki(t)})^2 = \lambda_{t+1}^2 + \dots + \lambda_n^2, \quad (4)$$

i.e. the sum of the higher-numbered eigenvalues (Gabriel 1971).

Species-centred PCA

Species-centred PCA is more commonly used than noncentred PCA. Here $y_{ki} = p_{ki} - p_{k.}$ (where $p_{k.} = \frac{1}{n} \sum_i p_{ki}$, the species mean) is used instead of $y_{ki} = p_{ki}$, but beyond this the derivation of vectors of sites and species for the biplot remains the same as in noncentred PCA.

In species-centred PCA the axes of the plane of the best fit pass through the centroid of the sites, instead of through the true origin as they do in noncentred PCA. Clearly distances between sites are not affected by this shift, but lengths are; hence Euclidean distances are

displayed, indeed more accurately than before, but not the SI values. For these we need to know the position of the true origin. This position is determined by projecting the origin on to the plane of the biplot and calculating the distance from the origin to its projection. The projection of the origin, z , has score

$$z_r = -(v_{1r}p_{1.} + v_{2r}p_{2.} + \dots + v_{mr}p_{m.}) \quad (5)$$

on the r^{th} principal component, i.e. z_r equals the inner product between the origin and the r^{th} eigenvector, both in coordinates with respect to the centroid. Now z can be plotted in the same way as could the sites.

The squared distance from the origin to z in a two-dimensional biplot is equal to

$$s = \sum_{k=1}^m p_k^2 - (z_1^2 + z_2^2). \quad (6)$$

The origin therefore lies exactly $s^{\frac{1}{2}}$ above its projection on to the biplot. The SI of a site with vector g_i in the biplot is approximated by

$$s + (g_{i1} - z_1)^2 + (g_{i2} - z_2)^2, \quad (7)$$

i.e. s plus the squared distance between the site and the projection of the origin. The order of SI's of the sites can therefore be seen at once by looking at distances from z . High-diversity sites will be near to z ; low-diversity sites will be further away. Note however that type B distortion is not detectable in the same way as for noncentred PCA.

Consider now an artificial example. Suppose there are just three species. Then sites can be plotted in the form of a triangular compositional diagram (e.g. Jöreskog et al. 1976, p. 93). The projection of the origin on to this diagram has coordinates $(1/3, 1/3, 1/3)$ in three-dimensional space and the squared distance from the origin to this point equals $1/3$. SI values can easily be derived from distances to the point $(1/3, 1/3, 1/3)$. A triangular compositional diagram is in fact a biplot in which the angle between the vectors representing variables equals 120° .

Finally: if z_1 is large in absolute value in relation to the other scores on the first principal component and $z_2 \approx 0$, then note that the first principal component will have high correlation with SI.

3. Results

Noncentred PCA

The last row of Table 1 shows SI values for each diatom sample; Table 2 gives squared Euclidean distances. These indices are approximated by noncentred PCA (Fig. 1). It is now of interest to compare SI values for samples with their squared lengths in Fig. 1, and squared Euclidean distances between samples with the corresponding squared distances in Fig. 1.

The highly diverse 1920-samples (D, E and F, with SI 0.1 to 0.2) are near the origin in Fig. 1. The other 1920-samples, which are somewhat less diverse (A, B and C, with SI 0.2 to 0.5), lie further away, as does 1978-sample e (SI ca. 0.3). The remaining 1978-samples have low diversity (SI 0.5 to 1.0). These are placed a long way from the origin.

The 1978-samples (except sample e) have low pairwise squared Euclidean distances (< 0.04), hence their proximity in Fig. 1. Their squared distances from sample e are higher (0.2 to 0.5), hence sample e is placed somewhat apart. Mean squared distances for samples from the 1920's and 1978 are 0.41 and 0.12 respectively; the 1920-samples are therefore more spread out in Fig. 1.

The vectors of the 1920-samples A, D, E and F point in the same direction, while differing greatly in length. This indicates type B distortion. The squared vector length of sample F is only half its SI value; the SI's of samples D and E are even worse approximated. On the other hand, the low-diversity 1978-samples have their SI's almost exactly approximated in Fig. 1.

A noncentred PCA biplot

The species vectors in Fig. 1 transform the site ordination into a biplot. Note that in this biplot Eunotia exigua has proportions near

unity in the 1978-samples, except in sample e. This is the only species of any importance pointing in the direction of the 1978-samples; in other words, the low diversity of these samples can be taken to result from the high proportion of Eunotia exigua in these samples. Obviously the 1920-samples B and C have a considerable proportion (ca. 0.4) of Eunotia exigua too, as does the 1978-sample e. The 1920-sample A is about as diverse as these samples, but this is due to another species, Tabellaria binalis, present with a proportion of 0.55 according to the biplot. Eunotia veneris has a high proportion in sample A too (0.20). Far more species point in the direction of the other highly diverse 1920-samples D, E and F. The main species in these samples are the remaining species of Table 1.

A comparison with Table 1 shows that proportions greater than 0.5 are more or less accurately approximated. These proportions concern Eunotia exigua and Tabellaria binalis. The proportions less than 0.5 (i.e. of all other species) are proportionally far less accurately approximated. The overall approximation error in Fig. 1 is 0.74; this is 13% of the sum of squares of all proportions (5.77). The first two eigenvalues are 4.51 and 0.52.

A species-centred PCA biplot

Fig. 2 shows a species-centred distance biplot. The proportions of each species are approximated as deviations from the species mean (last column of Table 1). These approximate deviations can be calculated from (3), if vectors in Fig. 2 are considered to originate from the point (0,0). This point is the centroid of the sites, not the true origin.

Eunotia exigua points west from the centroid. Projection of the samples on to this direction indicates that Eunotia exigua is present with high proportions in the 1978-samples except e; with a proportion about equal to its mean (0.48) in sample C; and with lower proportions in the remaining samples. These predictions agree with Table 1.

Tabellaria quadrisepata has mean 0.10 and points north from the centroid. Projection on to this direction shows that Tabellaria quadrisepata occurs with higher proportions in the samples B, E, e and D; with proportions of ca. 0.10 in samples C and F; and with lower proportions in the remaining samples. Table 1 does not confirm these

statements fully; sample D in particular contains less of the species than predicted.

The high proportions of Tabellaria binalis and Eunotia veneris in sample A are nicely represented in Fig. 2. The overall approximation error is lower than in Fig. 1. In Fig. 2 this error is 0.48, or 8% of the total sum of squares. The first two eigenvalues are 1.88 and 0.39.

In Fig. 1 the alpha and beta diversity values were approximated as well. What does Fig. 2 indicate about these? The squared Euclidean distances are, again, approximated by squared distances between samples, and note again that the 1920-samples are more spread out than the 1978-samples. But now we require the position of the true origin in order to infer approximate SI values. The origin lies 0.277 above its projection (indicated by a star). The SI for any site is therefore approximated by $0.277^2 = 0.077$ plus the squared distance from the site to the star. Again samples D, E and F are shown to have high alpha diversity (they lie near the star). Samples B, C and e are shown to be less diverse. Thereafter follow sample A and the remaining 1978-samples.

In conclusion, from Fig. 2 the impoverishment of the pools since the 1920's can be seen at once: many species have declined as Eunotia exigua has become the only dominant species. As a result the alpha diversity of the pools and the beta diversity among the pools have both decreased.

4. Discussion

Patterns of species-composition are commonly summarized by means of ordination-methods, by measures of diversity, or simply by reporting the main species with their abundances. The distance biplot unites these summaries: it is a form of PCA that allows approximate species-composition and diversity to be displayed simultaneously. This biplot allows alpha and beta diversity values to be associated with the main contributing species. Yet the biplot is simply a specially scaled combination of principal component scores and loadings. The display of alpha diversities in species-centred PCA requires just one extra point to be plotted, indicating the position of the true origin, i.e. the 'site without species'. Correlations between ordination results and values of

alpha diversity measures have been found empirically (Miracle 1974, Ritchie and Yarranton 1978, Goldsmith 1973). This paper contributes to an explanation of these findings.

The use of SI and squared Euclidean distance as alpha and beta diversity measures has certain advantages. Both measures have a convenient probabilistic interpretation in terms of random draws of two organisms from communities (Simpson 1949, Dyer 1978); their estimators have a small sampling variance in comparison with many other diversity measures; and sampling error in species-composition is conveniently related to SI. The main disadvantage of these indices is their relative insensitivity to rare species (Peet 1974). A rigorous solution to this problem was the creation of families of alpha diversity measures (Hurlbert 1971, Smith and Grassle 1977, Hill 1973b, Patil and Taillie 1979), indexed by a parameter that determines the weight given to common as against rare species. A similar approach to beta diversity measures is possible too (cf. Grassle and Smith 1976). The question arises whether these alternative diversity measures can also be linked to PCA through biplots. Elsewhere I shall show that diversity measures that fit into this framework include (with minor modifications) the families just mentioned - the only exception, strangely enough, being the Shannon index. These families induce new classes of useful data transformations in PCA:

$$y_{ki} = 1 - (1 - p_{ki})^m \quad m = 1, 2, 3, \dots \quad (8)$$

for diversity measures related to those of Hurlbert (1971) and Grassle and Smith (1976) and the power transformation

$$y_{ki} = p_{ki}^a \quad -\infty < a < \infty \quad (9)$$

for diversity measures related to those of Hill (1973b) and Patil and Taillie (1979). Note however that many ordination-methods that involve standardization of species or sites, e.g. reciprocal averaging (Hill 1973a), do not produce results that can be interpreted in terms of alpha diversities.

In conclusion, the biplot should prove to be a valuable graphical aid in summarizing species-composition and diversity. It may therefore

facilitate the exploratory stage of a study aimed at relating species-composition to environmental variables.

5. Acknowledgements

I would like to thank Mr H. van Dam, who suggested to me the relation between beta diversity and ordination-plots. He kindly permitted use of his data here. I am grateful to Drs I.C. Prentice, H.J.B. Birks, M.O. Hill and my colleagues for valuable comments.

6. Literature cited

- Connell, J.H., and E. Orias, 1964. The ecological regulation of species diversity. *American Naturalist* 98: 399-414.
- Corsten, L.C.A., and K.R. Gabriel, 1976. Graphical exploration in comparing variance matrices. *Biometrics* 32: 851-863.
- van Dam, H., G. Suurmond, and C.J.F. ter Braak. in press. Impact of acidification on diatoms and chemistry of Dutch moorland pools. *Hydrobiologia*.
- Dyer, D.P., 1978. An analysis of species dissimilarity using multiple environmental variables. *Ecology* 59: 117-125.
- Engen, S., 1978. *Stochastic abundance models*. Chapman and Hall, London, England.
- Gabriel, K.R., 1971. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58: 453-467.
- Gabriel, K.R., 1978. Least squares approximation of matrices by additive and multiplicative models. *Journal of the Royal Statistical Society Series B* 40: 186-196.
- Goldsmith, F.B., 1973. The vegetation of exposed sea cliffs at South Stack, Anglesey. I. The multivariate approach. *Journal of Ecology* 61: 787-818.
- Grassle, J.F., and W. Smith, 1976. A similarity measure sensitive to the contribution of rare species and its use in investigation of

- variation in marine benthic communities. *Oecologia* (Berl.) 25: 13-22.
- Hill, M.O., 1973a. Reciprocal averaging: an eigenvector method of ordination. *Journal of Ecology* 61: 237-249.
- Hill, M.O., 1973b. Diversity and evenness: a unifying notation and its consequences. *Ecology* 54: 427-432.
- Hurlbert, S.H., 1971. The non-concept of species diversity: a critique and alternative parameters. *Ecology* 52: 577-586.
- Jöreskog, K.G., J.E. Klován, and R.A. Reyment, 1976. Geological factor analysis. Elsevier Scientific Publishing Company, Amsterdam, The Netherlands.
- Kempton, R.A., 1979. The structure of species abundance and measurement of diversity. *Biometrics* 35: 307-321.
- MacArthur, R.H., and E.O. Wilson, 1967. The theory of island biogeography. Princeton University Press, Princeton, New Jersey, USA.
- May, R.M., 1973. Stability and complexity of model ecosystems. Princeton University Press, Princeton, New Jersey, USA.
- Miracle, M.R., 1974. Niche structure in freshwater zooplankton: a principal components approach. *Ecology* 55: 1306-1316.
- Noy-Meir, I., 1973. Data transformations in ecological ordination. I. Some advantages of non-centering. *Journal of Ecology* 61: 329-341.
- Orloči, L., 1978. Multivariate analysis in vegetation research. Second edition. Junk, The Hague, The Netherlands.
- Patil, G.P., and C. Taillie, 1979. An overview of diversity. Pages 3-27 in J.F. Grassle, G.P. Patil, W. Smith, and C. Taillie, editors. *Statistical ecology, volume S6. Ecological diversity in theory and practice*. International Cooperative Publishing House, Fairland, Maryland, USA.
- Peet, R.K., 1974. The measurement of species diversity. *Annual Review of Ecology and Systematics* 5: 285-307.
- Pielou, E.C., 1969. An introduction to mathematical ecology. Wiley-Interscience, New York, New York, USA.
- Pielou, E.C., 1975. Ecological diversity. John Wiley and Sons, New York, New York, USA.
- Ritchie, J.C., and G.A. Yarranton, 1978. The late-Quaternary history of the boreal forest of central Canada, based on standard pollen

- stratigraphy and principal components analysis. *Journal of Ecology* 66: 199-212.
- Simpson, E.H., 1949. Measurement of diversity. *Nature* 163: 688.
- Smith, W., and J.F. Grassle, 1977. Sampling properties of a family of diversity measures. *Biometrics* 33: 283-292.
- Whittaker, R.H., 1960. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs* 30: 279-338.
- Whittaker, R.H., 1972. Evolution and measurement of species diversity. *Taxon* 21: 213-251.
- Whittaker, R.H., 1973. Direct gradient analysis: results. Pages 33-51 in R.H. Whittaker, editor. *Handbook of vegetation science, volume 5. Ordination and classification of communities.* Junk, The Hague, The Netherlands.

Table 1. Proportions ($\times 100$) of eight diatom species in six Dutch moorland pools (columns indicated by letters) in the 1920's (upper case) and in 1978 (lower case), with species means and Simpson index values ($\times 100$).

	1920's						1978						mean
	A°	B	C	D	E	F	a	b	c	d	e	f	
<u>Eunotia lunaris</u> * (Ehrenberg) Grunow	0	0	1	1	23	1	0	0	0	1	0	0	2
<u>Navicula spec.</u>	0	0	2	30	0	0	0	0	0	1	0	0	3
<u>Eunotia veneris</u> (Kützing) De Toni	25	0	0	1	10	1	0	0	0	1	0	0	3
<u>Eunotia tenella</u> (Grunow) Hustedt	2	7	11	2	2	21	1	0	1	1	0	4	4
<u>Frustulia rhomboides</u> var. <u>saxonica</u> (Rabenhorst) De Toni	1	2	8	7	13	13	2	4	1	1	4	5	5
<u>Tabellaria binalis</u> (Ehrenberg) Grunow	63	0	0	0	0	19	0	0	0	0	0	0	7
<u>Tabellaria quadrisepta</u> <u>tata</u> Knudson	3	49	3	0	14	8	0	0	0	1	36	0	10
<u>Eunotia exigua</u> (De Brébisson) Rabenhorst	1	41	44	0	2	4	86	96	94	88	43	77	48
Simpson index	46	41	23	16	15	12	74	93	88	78	32	60	25

° Pools are arranged by Simpson index of the 1920's sample and denoted in van Dam et al. (in press) consecutively as B8, B3, B4, B5b, B2, B1.

* Species are arranged in order of mean abundance.

Table 2. Squared Euclidean distances ($\times 100$) between samples of Table 1, based on all 82 species.

A	0													
B	83	0												
C	68	23	0											
D	61	56	36	0										
E	55	40	32	27	0									
F	33	38	24	25	18	0								
a	119	45	21	89	85	78	0							
b	138	56	30	108	104	97	2	0						
c	133	52	27	103	99	92	1	0	0					
d	122	46	22	92	88	82	1	1	0	0				
e	75	3	14	47	34	34	31	41	38	32	0			
f	104	38	13	74	70	63	2	4	4	2	25	0		
A	B	C	D	E	F	a	b	c	d	e	f			

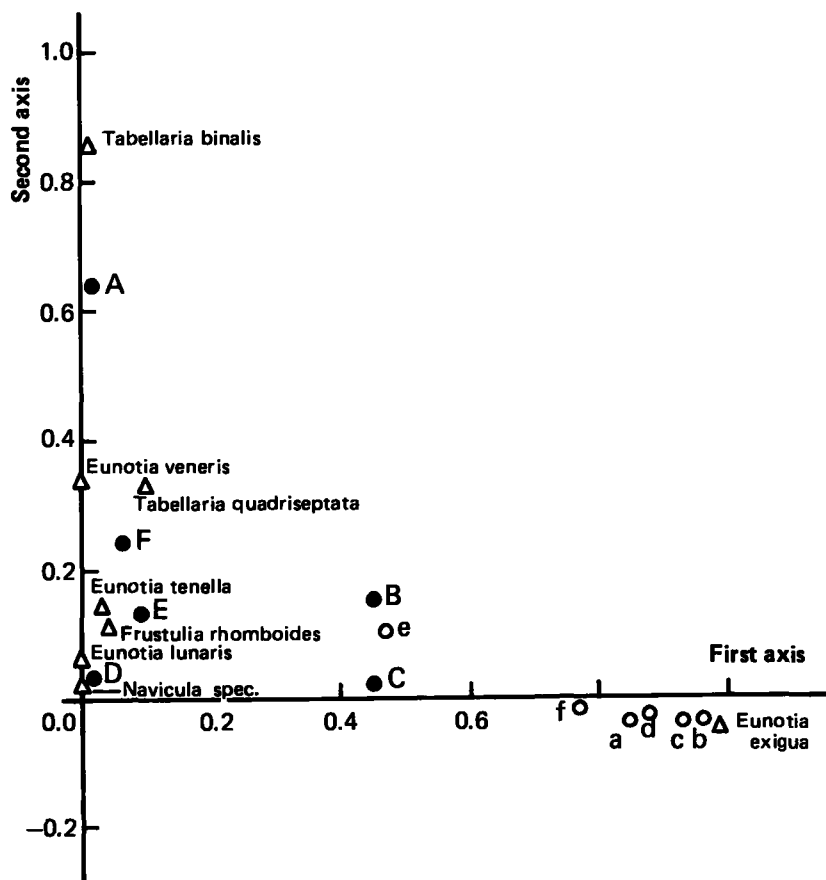


Fig. 1. Distance biplot by noncentred principal components analysis of the diatom samples, using all 82 species. The eight most abundant species are shown. Codes as in Table 1.

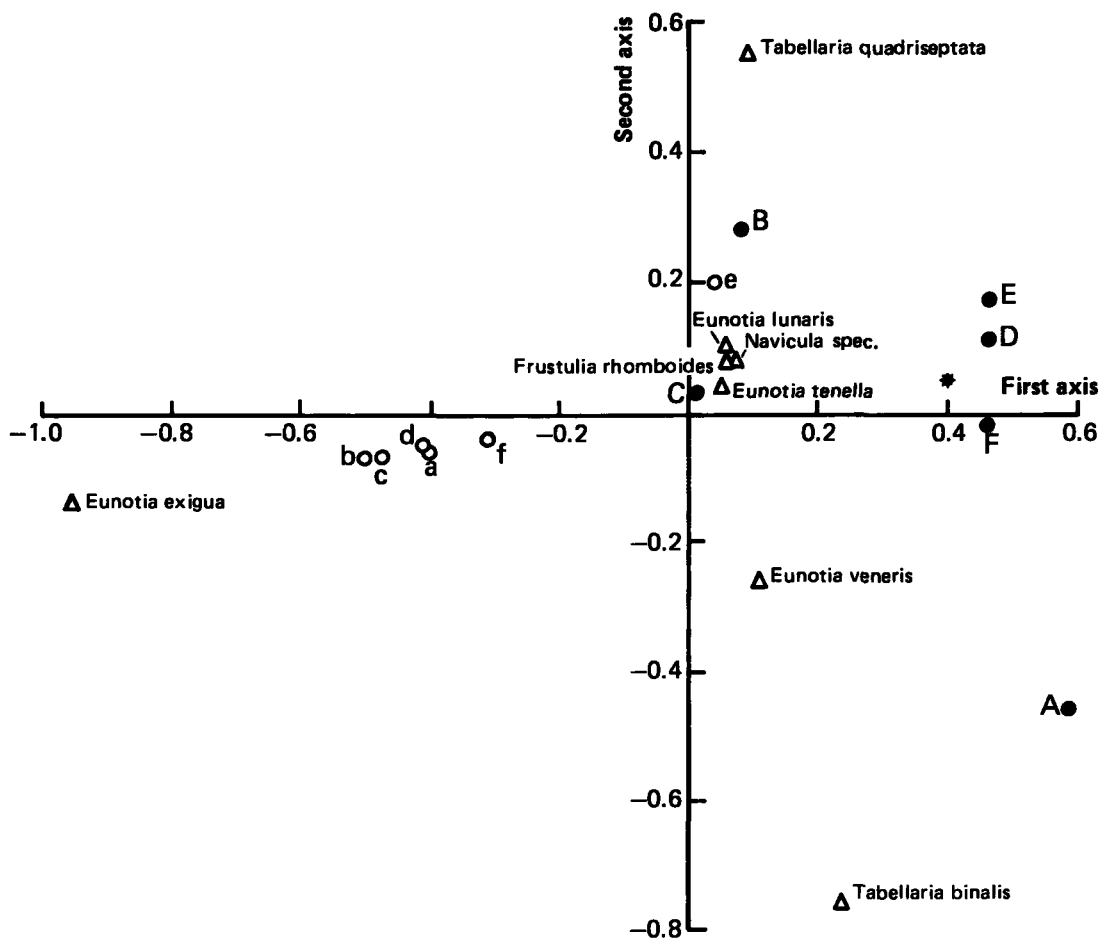
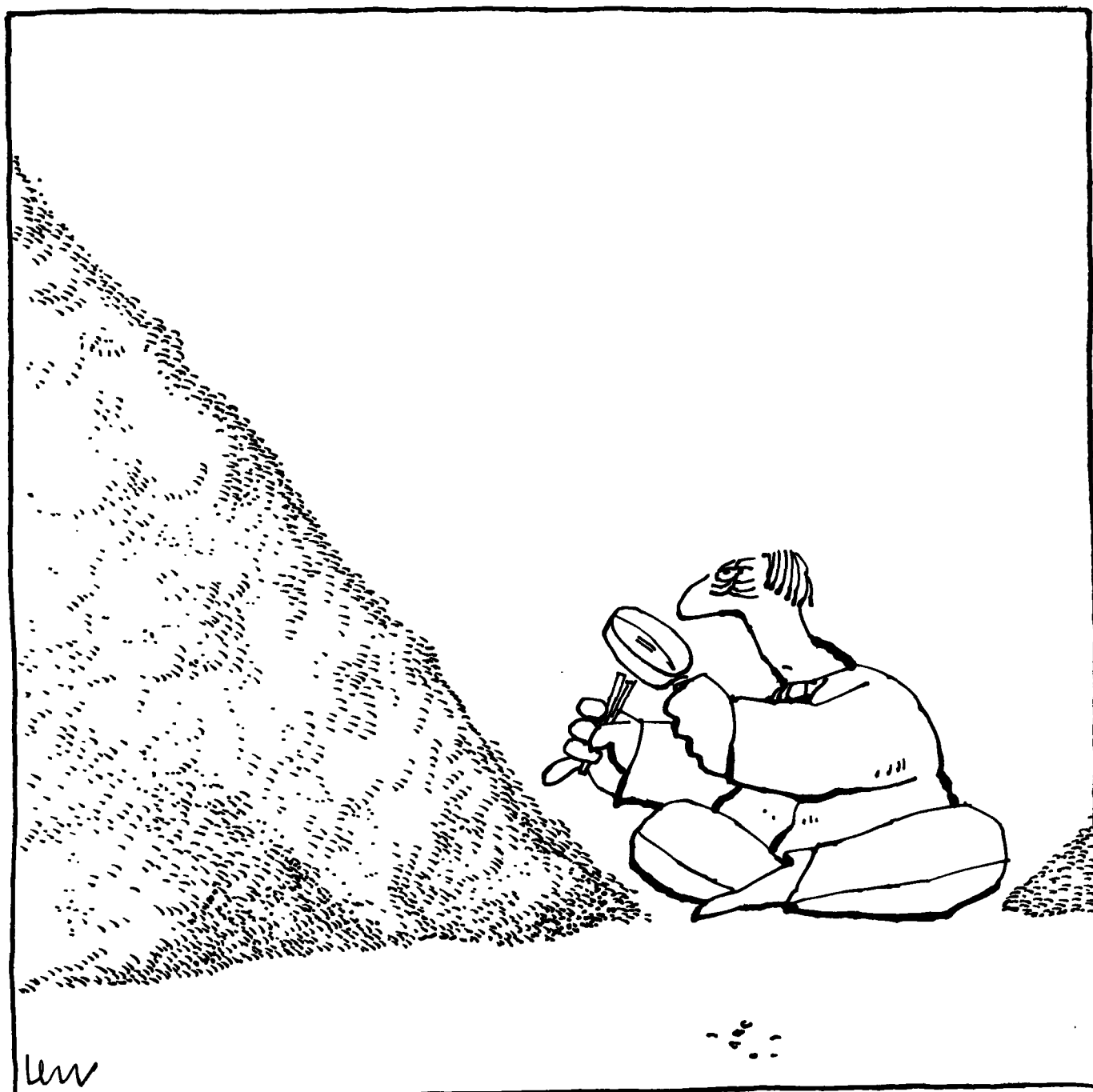


Fig. 2. Distance biplot by species-centred principal components analysis of the diatom samples, using all 82 species. Simpson index values are approximated by 0.077 plus squared distances from the projection of the origin (*). Codes as in Table 1.

'Most Probable Number' en Zaadkeuring

Drs. J. de Bree

Rapport ten behoeve van het Instituut voor Plantenziektenkundig
Onderzoek in het kader van de statistische begeleiding



Inhoud

1.	Inleiding	43
2.	Probleemstelling	44
3.	Puntschatting van f	45
4.	Intervalschatting van f	46
5.	Resultaten bij toepassing op twee voorbeelden	48
6.	Discussie	51
	Verwijzingen	51
	Appendix I	52
	Appendix II	53

1. Inleiding

Het is gebleken dat sommige plantenvirussen kunnen worden overgebracht door middel van het zaad van aangetaste planten. Derhalve is het van groot belang over een methode te beschikken voor keuring en certificering van partijen zaad. Een moeilijkheid is echter dat individuele zaden weliswaar op aanwezigheid van virus kunnen worden onderzocht, maar dat dit voor steekproeven van de hier vereiste omvang, zo al technisch uitvoerbaar, te tijdrovend en te kostbaar zou zijn. Het zou aantrekkelijk zijn wanneer het percentage geïnfecteerde zaden zou kunnen worden geschat op basis van het al dan niet aanwezig zijn van virus in een beperkt aantal steekproeven.

Deze situatie vertoont enige gelijkenis met de bepaling van de dichtheid van micro-organismen in vloeistoffen^{*)}). In dit geval kan men de aantallen door bacteriën gevormde kolonies op enige voedingsbodems waarop kleine hoeveelheden van de vloeistof zijn gedeponed, tellen en op voor de hand liggende wijze een schatting van de onbekende dichtheid berekenen. Het is echter aanzienlijk minder moeizaam om alleen maar vast te stellen of de opgebrachte hoeveelheden vloeistof al dan niet één of meer exemplaren van de bacterie in kwestie hebben bevat, terwijl uit deze informatie eveneens een schatting van de dichtheid kan worden gevonden, zij het één met niet al te grote precisie. Men gebruikt hiervoor de methode van de "grootste aannemelijkheid" (in het Engels: maximum likelihood). De aldus verkregen schatting van de dichtheid heet om historische redenen "Most Probable Number" (MPN).

Het is in sommige kringen thans gebruikelijk om de MPN-methode alsmede tabellen en dergelijke die daarvoor zijn verschenen, toe te passen op het hierboven beschreven zaadkeuringsprobleem. Dit blijkt in sommige gevallen merkwaardige of onlogische schattingen van het percentage geïnfecteerd zaad op te leveren. De oorzaak hiervan is het feit dat redelijkerwijze niet alle veronderstellingen die in het traditionele geval van MPN-bepalingen realistisch zijn, voor het zaadkeuringsprobleem

*) Het hier behandelde zaadkeuringsprobleem alsmede de genoemde analogie met de bepaling van de dichtheid van micro-organismen in vloeistoffen werden ons voorgelegd door de heer J.W.L. van Vuurde van het Rijksproefstation voor de Zaadcontrole te Wageningen.

kunnen worden gemaakt. Hierover is elders in het kort gerapporteerd (De Bree, 1979).

In paragraaf 2 wordt het te behandelen probleem geformuleerd in statistische termen. In paragraaf 3 wordt aangegeven hoe een puntschatting voor de onbekende fractie besmet zaad in de populatie kan worden verkregen, waarna in paragraaf 4 de constructie van betrouwbaarheidsintervallen voor die fractie wordt beschouwd. In paragraaf 5 worden de resultaten gepresenteerd van berekeningen, uitgevoerd voor twee vrij willekeurig gekozen steekproefschema's.

Tenslotte wordt in de appendices I en II een rechtvaardiging van de gebruikte statistische methoden geschetst.

2. Probleemstelling

Een partij zaden waarvan een onbekende fractie f met virus is besmet, moet worden onderzocht met het doel een schatting van f te vinden. In al wat volgt wordt de omvang van de partij als oneindig beschouwd, wat toegestaan is als de uit de partij te trekken steekproeven gezamenlijk slechts een zeer klein deel van de partij vormen.

We veronderstellen dat het niet uitvoerbaar is om grote aantallen zaden individueel op besmetting te onderzoeken, maar dat wel van alle te nemen steekproeven exact kan worden vastgesteld of zij al dan niet besmette zaden bevatten. In de praktijk is het inderdaad mogelijk om door middel van verfijnde technieken de aanwezigheid van één of meer besmette zaden in een steekproef van ettelijke duizenden zaden te detecteren.

Op grond van enige vooraf te verwerven informatie worden nu verscheidene aselechte steekproeven van verschillende aantallen zaden uit de partij getrokken. Het hierbij gevolgde steekproefschema dient zodanig te zijn, dat naar verwachting bij althans één van de gebruikte steekproefgroottes sommige steekproeven wel en andere niet besmet zullen zijn.

Het opzetten van een geschikt steekproefschema is een probleem op zichzelf, dat hier echter niet wordt behandeld.

Stel nu dat, voor $i = 1, 2, \dots, I$, n_i aselechte steekproeven ter grootte van m_i zaden worden getrokken en dat k_i van deze n_i steekproeven één of

meer besmette zaden bevatten. In totaal worden er dus $n_1 + n_2 + \dots + n_I$ steekproeven getrokken die alle onderling onafhankelijk worden verondersteld.

\underline{k}_i is een stochastische variabele, dat wil zeggen dat de waarde die \underline{k}_i aanneemt, wordt bepaald door een kansverdeling. In het bijzonder volgt \underline{k}_i een binomiale verdeling met parameters n_i en $p_i = 1 - (1-f)^{m_i}$ zoals gemakkelijk valt in te zien. De kans dat \underline{k}_i gelijk is aan k_i wordt dus gegeven door

$$\Pr(\underline{k}_i = k_i) = \binom{n_i}{k_i} \{1 - (1-f)^{m_i}\}^{k_i} \{(1-f)^{m_i}\}^{n_i - k_i} \quad [1]$$

De kansverdeling van het gezamenlijke steekproefresultaat $\underline{k}_1, \underline{k}_2, \dots, \underline{k}_I$ wordt gegeven door het produkt van de kansen volgens [1] en is gelijk aan

$$\Pr(\underline{k}_1 = k_1, \underline{k}_2 = k_2 \dots \underline{k}_I = k_I) = \prod_{i=1}^I \binom{n_i}{k_i} \{1 - (1-f)^{m_i}\}^{k_i} \{(1-f)^{m_i}\}^{n_i - k_i} \quad [2]$$

Het rechterlid van [2], voor gegeven waarden k_1, k_2, \dots, k_I en opgevat als functie van f , wordt de aannemelijkheid (in het Engels: likelihood) van het steekproefresultaat k_1, k_2, \dots, k_I genoemd. Formule [2] en de bijbehorende definities der gebruikte symbolen vormen het wiskundig-statistische model dat dient als uitgangspunt voor de schatting van f bij een gegeven realisatie k_1, k_2, \dots, k_I van de stochastische variabelen $\underline{k}_1, \underline{k}_2, \dots, \underline{k}_I$.

3. Punt-schatting van f

De punt-schatting van f , dat wil zeggen de bepaling van één getal als schatting van f , wordt uitgevoerd met de methode van de "grootste aannemelijkheid". De methode is algemeen bekend; hier wordt volstaan met op te merken dat bij een gegeven realisatie k_1, k_2, \dots, k_I een zodanige waarde van f wordt bepaald dat de aannemelijkheid, dat wil zeggen het rechterlid van [2], maximaal is als functie van f . Deze waarde van f , zeg \hat{f} , heet de meest aannemelijke schatting; zij wordt, met behulp van

een computer, langs iteratieve weg bepaald (Appendix I).

In paragraaf 5 worden bij wijze van voorbeeld voor alle mogelijke uitkomsten van het aantal "besmette" monsters de berekende meest aannemelijke schattingen weergegeven voor het steekproefschema met $n_1 = 1$, $n_2 = 5$, $m_1 = 500$, $m_2 = 100$ (tabel 1) en voor het schema met $n_1 = n_2 = n_3 = 2$, $m_1 = 5000$, $m_2 = 1000$, $m_3 = 100$ (tabel 2). Beide schema's worden in de praktijk gebruikt, maar zijn overigens willekeurig gekozen.

4. Intervalschatting van f

Puntschatting van f resulteert in slechts één getal als schatting van f , zonder enige aanduiding van de nauwkeurigheid van de gegeven waarde. Dit is uiteraard onbevredigend. Veel bruikbaar is een uitspraak dat f met een zekere betrouwbaarheid in een op te geven interval ligt. Deze laatste wijze van gevolgtrekking staat bekend als intervalschatting.

Een standaardmethode voor het berekenen van een betrouwbaarheidsinterval, in direct verband staande met de methode der grootste aannemelijkheid, berust op het benaderen van de variantie van de meest aannemelijke schatter. Dit geschiedt met behulp van de inverse van de tweede afgeleide naar f van de aannemelijkheid: de tegengestelde waarde van deze inverse, geëvalueerd ter plaatse van \hat{f} , is asymptotisch (dat wil zeggen voor grote n_1) een goede benadering van de variantie van de meest aannemelijke schatter. Deze methode is in ons geval niet bruikbaar gebleken. De aannemelijkheid ter plaatse van het maximum vertoont met name bij vrij kleine waarden van \hat{f} een zodanig scherpe piek dat de benadering van de variantie ongeloofwaardig klein is. Kennelijk gelden de plezierige asymptotische eigenschappen van \hat{f} hier bij lange na nog niet.

In het bovenstaande is getracht om het steekproefgedrag van de meest aannemelijke schatter te beschrijven met gebruikmaking van slechts zijn eerste twee (asymptotische) momenten. Men mag verwachten dat een betere benadering kan worden verkregen door ook hogere momenten van de schatter of een eenvoudige functie daarvan te gebruiken.

Dit is bijvoorbeeld mogelijk door een transformatie van een standaard-normaal verdeelde variabele te gebruiken, in het bijzonder een vierdegraads polynoom met coëfficiënten die functies zijn van de eerste vier momenten van de afgeleide naar f van de logaritmen van de aannemelijkheid in [2]. (Cornish - Fisher transformatie, zie bijvoorbeeld Kendall and Stuart, 1977, p. 175 e.v.)

Pogingen in deze richting hebben niet het gewenste resultaat opgeleverd: blijktbaar zijn de hier te gebruiken approximaties niet voldoende nauwkeurig.

Een uitweg uit de hierboven geschetste problemen wordt geboden door het besef, dat dank zij de formidabele rekensnelheid van de huidige computers het berekenen van de exacte kansverdeling van alle mogelijke steekproefresultaten, gegeven een veronderstelde waarde van f , een kleinigheid is. Men kan derhalve bij elk mogelijk steekproefresultaat een betrouwbaarheidsinterval voor f samenstellen uit die waarden f_0 die niet zouden worden verworpen bij toetsing van de

nulhypothese	$H_0 : f = f_0$ tegen de
alternatieve hypothese	$H_A : f < f_0$

(We beperken ons hier, zonder verlies van algemeenheid, tot het éézijdig toetsingsprobleem.)

Een toets voor dit probleem, met onbetrouwbaarheid α , wordt in principe gegeven door de beslissingsregel:

verwerp H_0 ten gunste van H_A als $\sum_{i=1}^I c_i k_i < c$.

- Hierbij geldt:
- c_1, c_2, \dots, c_I zijn constanten die van f_0 afhangen,
 - k_1, k_2, \dots, k_I is de steekproefuitkomst,

- c is een constante die wordt bepaald door de eis dat onder de nulhypothese de kans op de gebeurtenis

$$\sum_{i=1}^I c_i \frac{k_i}{n_i} < c \text{ gelijk is aan } \alpha .$$

Problemen die ontstaan doordat de kansverdeling van k_1, k_2, \dots, k_I discreet is, worden hier niet besproken; zij worden op de gebruikelijke manier opgelost (Appendix II).

De hier voorgestelde toets is afgeleid van de meest onderscheidende toets voor het toetsingsprobleem met

nulhypothese	$H_0: f = f_0$ en
alternatieve hypothese	$H'_A: f = f_1 < f_0$

Deze meest onderscheidende toets heeft van alle toetsen met onbetrouwbaarheid α het grootste onderscheidingsvermogen onder H'_A (Appendix II).

5. Resultaten bij toepassing op twee voorbeelden

In onderstaande tabellen worden voor alle mogelijke uitkomsten van de steekproefschema's genoemd in paragraaf 3 de meest aannemelijke schattingen en betrouwbaarheidsintervallen gegeven.

Voor de berekeningen is gebruik gemaakt van een door IWIS ontwikkeld computerprogramma.

aantal besmette steekproeven uit		meest aanneme- lijke schat- ting f	éénzijdig 0.95 betrouwbaar- heidsinterval bovengrens	tweezijdig 0.95 be- trouwbaarheidsinterval	
l van 500	5 van 100			ondergrens	bovengrens
0	0	0.0000	0.0030	0.0000	0.0037
0	1	0.0010	0.0043	0.0000	0.0052
0	2	0.0022	0.0052	0.0003	0.0062
0	3	0.0036	0.0071	0.0009	0.0084
0	4	0.0051	0.0090	0.0019	0.0090
0	5	0.0069	0.0109	0.0035	0.0127
1	0	0.0014	0.0068	0.0000	0.0080
1	1	0.0032	0.0108	0.0003	0.0127
1	2	0.0056	0.0165	0.0009	0.0190
1	3	0.0093	0.0254	0.0019	0.0290
1	4	0.0160	0.0448	0.0035	0.0515
1	5	1.0000	1.0000	0.0066	1.0000

Tabel 1: Meest aannemelijke schattingen, één- en tweezijdige betrouwbaarheidsintervallen voor de fractie besmet zaad in de populatie.

$n_1 = 1$, $m_1 = 500$; $n_2 = 5$, $m_2 = 100$.

aantal besmette steekproeven uit			meest aannemelijke schatting, f	éénzijdig 0.95 betrouwbaarheidsinterval, bovengrens	tweezijdig 0.95 betrouwbaarheidsinterval	
2 van 5000	2 van 1000	2 van 100			ondergrens	bovengrens
0	0	0	0.0000	0.0002	0.0000	0.0003
0	0	1	0.0001	0.0002	0.0000	0.0003
0	0	2	0.0002	0.0003	0.0000	0.0004
0	1	0	0.0001	0.0003	0.0000	0.0004
0	1	1	0.0002	0.0004	0.0000	0.0004
0	1	2	0.0003	0.0005	0.0001	0.0006
0	2	0	0.0002	0.0004	0.0000	0.0004
0	2	1	0.0003	0.0005	0.0001	0.0006
0	2	2	0.0004	0.0007	0.0002	0.0007
1	0	0	0.0001	0.0005	0.0000	0.0006
1	0	1	0.0002	0.0005	0.0000	0.0006
1	0	2	0.0003	0.0007	0.0001	0.0008
1	1	0	0.0002	0.0007	0.0000	0.0008
1	1	1	0.0004	0.0007	0.0001	0.0008
1	1	2	0.0005	0.0007	0.0002	0.0009
1	2	0	0.0004	0.0007	0.0001	0.0009
1	2	1	0.0005	0.0008	0.0002	0.0009
1	2	2	0.0007	0.0011	0.0003	0.0011
2	0	0	0.0003	0.0014	0.0000	0.0017
2	0	1	0.0006	0.0015	0.0001	0.0018
2	0	2	0.0010	0.0018	0.0002	0.0018
2	1	0	0.0007	0.0031	0.0001	0.0036
2	1	1	0.0013	0.0036	0.0002	0.0042
2	1	2	0.0021	0.0037	0.0006	0.0044
2	2	0	0.0024	0.0149	0.0002	0.0183
2	2	1	0.0070	0.0361	0.0006	0.0428
2	2	2	1.0000	1.0000	0.0020	1.0000

Tabel 2: Meest aannemelijke schattingen, één- en tweezijdige betrouwbaarheidsintervallen voor de fractie besmet zaad in de populatie.

$n_1 = 2, m_1 = 5000; n_2 = 2, m_2 = 1000; n_3 = 2, m_3 = 100$

6. Discussie

De hier beschreven constructie van de betrouwbaarheidsintervallen voor f heeft de plezierige eigenschap dat geen, al dan niet asymptotische, approximaties worden toegepast.

Het belangwekkende probleem van de keuze van een geschikt steekproef-schema is hier buiten beschouwing gebleven. Dit probleem hangt nauw samen met de vraag naar de kwaliteit van de statistische toets waarop de intervallen zijn gebaseerd. Wij hebben het voornemen om deze vraag in de nabije toekomst te bestuderen.

Verwijzingen

De Bree, J. (1979), "Most Probable Number" (MPN) en Zaadkeuring. IWIS-rapport D 79 ST 31 18.

Kendall, M and Stuart, A. (1977), The advanced theory of statistics, Vol. I. Griffin, London.

Lehmann, E.L. (1959), Testing statistical hypotheses. Wiley & Sons, New York,

Leppink, G.J. (1981), Mondelinge discussie.

Appendix I

De meest aannemelijke schatting \hat{f} is de waarde van f waarvoor de aannemelijkheid, zeg L , maximaal is.

Er geldt:

$$L = \prod_{i=1}^I \binom{n_i}{k_i} \{1 - (1-f)^{m_i}\}^{k_i} \{(1-f)^{m_i}\}^{n_i - k_i}$$

Na het nemen van de logaritme van L en differentiëren van $\log L$ naar f blijkt \hat{f} te worden gevonden als oplossing naar f van de vergelijking

$$\frac{1}{1-f} \left(\sum_{i=1}^I \frac{m_i k_i}{1 - (1-f)^{m_i}} - \sum_{i=1}^I m_i n_i \right) = 0$$

Voor de triviale steekproefuitkomsten $k_i = n_i$ ($i = 1, 2, \dots, I$) en $k_i = 0$ ($i = 1, 2, \dots, I$) is $\hat{f} = 1$ resp. $\hat{f} = 0$, zoals blijkt uit beschouwing van L zelf.

Het is eenvoudig te bewijzen dat L een concave functie van f is, zodat het maximum eenduidig bepaald is. Ter berekening van het maximum is een halveringsproces gebruikt.

Appendix II

Men kan bewijzen dat voor het toetsingsprobleem met

$$\left. \begin{array}{l} \text{nulhypothese} \\ \text{alternatieve hypothese} \end{array} \right\} \begin{array}{l} H_0: f = f_0 \text{ en} \\ H_A: f = f_1 < f_0 \end{array} \quad (I)$$

de meest onderscheidende toets met onbetrouwbaarheid α , gegeven wordt door Φ met

$$\Phi = \begin{cases} 1 & \text{als } \sum_{i=1}^I k_i \{g_i(f_0) - g_i(f_1)\} < c \\ \gamma & \text{als } \sum_{i=1}^I k_i \{g_i(f_0) - g_i(f_1)\} = c \quad (0 < \gamma < 1) \\ 0 & \text{als } \sum_{i=1}^I k_i \{g_i(f_0) - g_i(f_1)\} > c \end{cases}$$

waarin $g_i(f) = \log \left\{ \frac{1}{(1-f)^{m_i}} - 1 \right\}$. De constanten γ en c volgen uit de eis

dat de verwachtingswaarde van Φ onder H_0 gelijk is aan α , dat wil zeggen $E_{f_0} \Phi = \alpha$ (Lehmann, 1959, Theorem 1, p. 65).

We merken op dat $g_i(f_0) - g_i(f_1) > 0$ ($i = 1, 2, \dots, I$) zodat Φ een niet-stijgende functie van elk der k_i is. Derhalve is $E_f \Phi$ een niet-stijgende functie van f (Lehmann, 1959, Lemma 2, p. 74) en is Φ een zuivere toets van H_0 tegen het ruimere alternatief $H_A' : f \leq f_1$.

Laat men nu f_1 naderen tot f_0 en benadert men $g_i(f_0) - g_i(f_1)$ door

$$(f_0 - f_1) \left(\frac{dg_i(f)}{df} \right)_{f=f_0} = \frac{f_0 - f_1}{1 - f_0} \cdot \frac{m_i}{1 - (1 - f_0)^{m_i}},$$

dan lijkt het een goede keuze (Leppink, 1981) om als toets voor het probleem met

$$\begin{array}{ll}
 \text{nulhypothese} & H_0: f = f_0 \text{ en} \\
 & \hspace{20em} \} \quad \text{(II)} \\
 \text{alternatieve hypothese} & H_A: f < f_0
 \end{array}$$

te nemen de toets Ψ , gegeven door

$$\Psi = \begin{cases} 1 & \text{als } \sum k_i \frac{m_i}{1 - (1-f_0)^{m_i}} < d \\
 \delta & \text{als } \sum k_i \frac{m_i}{1 - (1-f_0)^{m_i}} = d \\
 0 & \text{als } \sum k_i \frac{m_i}{1 - (1-f_0)^{m_i}} > d \end{cases} \quad (0 < \delta < 1)$$

waarin δ en d volgen uit de eis $E_{f_0} \Psi = \alpha$

Aangezien $\frac{m_i}{1 - (1-f_0)^{m_i}} > 0$ ($i = 1, 2, \dots, I$), is Ψ , evenals ϕ , een

niet-stijgende functie van elk der k_i en derhalve is $E_f \Psi$ een niet-stijgende functie van f . Ψ is dus een zuivere toets voor het probleem (II).

Ψ is een "gerandomiseerde" toets": voor bepaalde waarden van f_0 en bepaalde realisaties k_1, k_2, \dots, k_I dient H_0 met kans δ te worden verworpen (door middel van een lotingsprocedure). Wij hebben echter dergelijke f_0 zonder meer in het betrouwbaarheidsinterval opgenomen. Ten gevolge hiervan is de betrouwbaarheid van sommige intervallen in werkelijkheid groter dan $1-\alpha$.

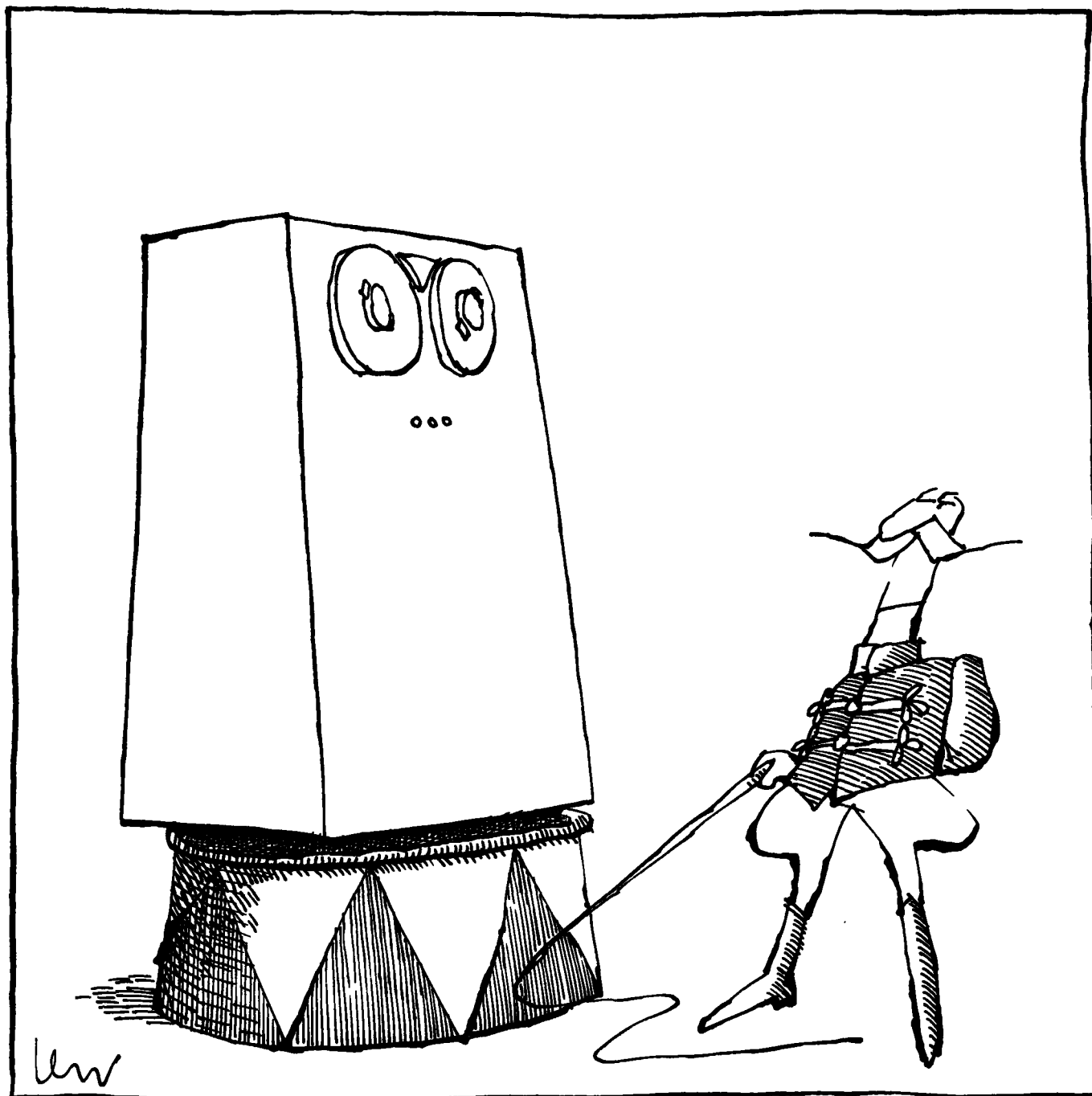
Opmerking:

De uit Lehmann (1959) geciteerde stellingen worden bewezen voor stochastische variabelen die onafhankelijk en identiek verdeeld zijn; zij kunnen echter ook worden bewezen voor het geval dat de variabelen verschillen in verdeling.

Informatica en het landbouwkundig onderzoek

Drs. T. de Heer

Voordracht gehouden op 19 mei 1981 ter gelegenheid van het
25-jarig bestaan van IWIS-TNO afdeling Wageningen



Als mensen horen dat je in je werk met computers te maken hebt, worden ze vaak een tikje emotioneel geïnteresseerd. Een computer is voor hen een spookachtig apparaat dat dingen vermag te doen waartoe wij tot voor kort alleen het menselijk brein in staat achtten.

Meestal vragen ze dan ook gretig "wat doe je nou eigenlijk precies". Dan antwoord ik naïef en plichtsgetrouw "ik beoefen het vak informatica. Maar o wee, dan heb je de poppen aan het dansen. Je krijgt onmiddellijk de wedervraag "wat is dat". Zie dan maar dat je dat weer voor elkaar krijgt.

Kijk, en dat vind ik nou eigenlijk zo onbarmhartig. Want een groenteboer of een professor in de economie overkomt dat nooit. Stel dat u zegt "ik verkoop groente" en iemand vraagt dan "wat is dat".

Ik ben benieuwd wat u dan ten antwoord geeft.

Ludwig Wittgenstein schijnt de eer te beurt te vallen de eerste filosoof te zijn geweest die erop gewezen heeft dat de betekenis van een woord helemaal niet gedefinieerd kán worden. Wat wel kan volgens hem, is de betekenis te tonen. Gewoon door te laten zien hoe een woord gebruikt kan worden. In het geval van 'groente' kan men een groentewinkel binnengaan, een bloemkool in de hand nemen en zeggen "kijk, dit is nou groente". Bij het woord 'stoel' kan men op een stoel gaan zitten en zeggen "datgene waarop ik nu zit is een stoel". Enzovoort.

Wanneer iemand naar de betekenis van het woord 'informatica' informeert, zeg ik meestal "het heeft iets met taal te maken". De mensen reageren daarop in de trant van "nou breekt m'n klomp, ik dacht dat computers rekenmachines waren en taal is toch heel wat anders dan rekenen?".

Deze wedervraag brengt mij precies waar ik wil wezen. Want nu kan ik ongestoord een lange monoloog opzetten. Immers, om de computer iets te laten uitrekenen moet je om te beginnen een taal hebben waarin je de computer kunt vertellen wat ie moet doen.

Zo'n taal heet een programmeertaal. Het is een kunstmatige taal net als Esperanto, alleen een stuk eenvoudiger. Het maken van programmeertalen behoort tot het vakgebied van de informatica. Programmeertalen vormen de bakermat van de informatica.

En waarvoor dient een taal nou anders dan voor het overbrengen van informatie. Zolang die overbrenging plaatsgrijpt van mens tot mens via de natuurlijke taal, heeft het allemaal nog niets met informatica te maken. Dan heet het gewoon spreken, praten, telefoneren, opbellen, schrijven, vergaderen, bevel, les, college, advies, verzoek, preek, rede, boek, tijdschrift, noem maar op.

Zodra er overbrenging van informatie, niet via, maar tussen mens en machine plaatsgrijpt, beweegt men zich echter onmiddellijk wél op het terrein van de informatica. Er is dan namelijk duidelijk sprake van een machine die (re)ageert met behulp van het medium taal. Die machine doet dan dus iets dat mensen beslist niet kunnen zonder hun hersens te gebruiken. En inderdaad, als je er probeert achter te komen hoe je een machine kunt laten schaken, rekenen, vertalen, boekhouden, administreren, of welke intelligentie vereisende handeling dan ook, dan ben je bezig om informatica te bedrijven.

De gewraakte machine is dan al gauw een digitale elektronische rekenmachine met aangekoppelde of perifere apparatuur zoals beeldschermen, toetsenborden, schrijfmachines, regeldrukkers en wat al niet meer.

Ik neem maar aan dat u nu niet meer vraagt wat informatica is.

Het vak, of zo u wilt, de nog jonge wetenschap informatica, is in zijn korte bestaan enorm snel uitgegroeid en bestrijkt een behoorlijk groot aantal onderwerpen. Zoals bij de meeste wetenschappen is de grens met enkele andere wetenschappen, namelijk taalkunde, wiskunde en logica, niet duidelijk te trekken. Onderzoekers en toepassers van de informatica bepalen zich veelal wijselijk tot een of ander deelgebied. Zo zijn er programmeertaalspecialisten, compilerbouwers, data base management-specialisten, systeem software-specialisten.

De Afdeling Informatica van het IWIS bevindt zich tot nu toe voornamelijk op het terrein van de verzorging en de verspreiding van informatie met behulp van computers. Bij de verzorging komt men vaak te staan voor het inschakelen van computers bij het inrichten van archieven. Daarbij wordt gebruik gemaakt van data base management-systemen en retrieval-systemen. Voor sommige toepassingen moeten speciale nieuwe computerprogramma's worden ontwikkeld. Voor andere kunnen bestaande programma's worden ingeschakeld. De programmeurs werken meestal op de eigen computer van het IWIS. Hoewel dat niet strikt noodzakelijk is, heeft dit vaak toch wel de nodige voordelen omdat dan de produktiebegeleiding en het onderhoud in ieder geval in vakbekwame handen is.

Toch is de Afdeling Informatica heus niet de enige afdeling van het IWIS die de informatica beoefent. Echter wel de enige afdeling die zich ertoe bepaalt. Andere afdelingen beoefenen informatica alleen zodra het nodig is om een gesteld doel te bereiken. Bijvoorbeeld de Afdeling Operations Research, waar simulatietalen worden gebezigd om wiskundige modellen te beschrijven en te realiseren op een computer. Of de Afdeling Administratieve Automatisering. Het gaat om de administratie van TNO, maar altijd met behulp van de computer. Dat zegt het woord automatisering al.

De Afdeling Statistiek ook. Zodra er rekenwerk van betekenis moet worden verricht, wordt de informatica, althans de computer, te hulp geroepen. Standaardpakketten zoals GENSTAT en SPSS, die ook met speciale talen via de terminal aan het werk worden gezet, stellen de statisticus veelal in staat om zonder raadpleging van informatici zijn rekenwerk aan de computer op te dragen.

De geschiedenis van de informatica binnen het IWIS, voorheen ABW-TNO, loopt in grote lijn parallel met de ontwikkeling van dat vakgebied in de wereld er omheen. De allereerste vorm van automatisering bestond uit het toepassen van tabelleermachines voor het maken van statistieken.

Eerst werd er alleen gerekend. Ieder probleem werd apart geprogrammeerd in het begin in assembleertaal, later in ALGOL60. De eerste generatie computers was nog niet zo snel, had kleine geheugens en weinig in- en uitvoermogelijkheden. Men communiceerde toen altijd met papieren ponsband. Later kwam de programmeertaal COBOL erbij en de magneetband als invoer-, uitvoer- en opslagmedium. Nu werkt men met grote magnetische schijfengeheugens en de data base management-systemen. Men heeft de beschikking over de programmeertalen FORTRAN, COBOL, ALGOL, SIMULA, PASCAL, SIMPL (SPL), APL en SNOBOL. Er zijn standaardpakketten voor statistische en numerieke bewerkingen. De computer verwerkt meerdere programma's gelijktijdig met behulp van een supervisieprogramma (operating system).

Interactieve communicatie met de computer behoort tot de gewone gang van zaken. Programmeurs beschikken over een eigen terminal waaraan zij opdrachten doorspelen aan de computer direct via een toetsenbord. Externe klanten van het Rekencentrum worden in staat gesteld hetzelfde te doen.

Ten aanzien van de 'wetenschap' informatica en haar toepassingen heeft het IWIS maar een uiterst bescheiden bijdrage kunnen leveren. Alleen binnen de Afdeling Informatica hebben eigenlijk zelfstandige ontwikkelingen in dat opzicht plaats gehad die zich beperken tot het gebied van de informatiesystemen, dus data base-systemen en retrieval-systemen, en dan nog gespecialiseerd op het omgaan met natuurlijke taal in een computer. Die verbinding met natuurlijke taal is historisch zo gegroeid omdat de Afdeling Informatica gestart is met het ontwikkelen van een retrieval-programma voor een geautomatiseerde literatuurattenderingsdienst op het gebied van de chemie. Dat programma wordt tegenwoordig ook gebruikt voor literatuurattendering op het terrein van de landbouw in samenwerking met PUDOC.

Sindsdien is er ook een programma ontwikkeld waarmee de computer te hulp kan worden geroepen bij het aanleggen van archieven. Dit programma heet DARIUS.

Het is bijzonder geschikt voor het maken van catalogusachtige overzichten. Het programma kan tevens informatie gereed maken voor een computer-gestuurde zetmachine. Met dit programma worden uitgaven verzorgd van inventarisaties van beschrijvingen van wetenschappelijk en technisch onderzoek in Nederland in samenwerking met ZWO, het Ministerie van Volksgezondheid en Milieuhygiëne, het Waterloopkundig Laboratorium, het Sociaal Wetenschappelijk Informatie- en Documentatiecentrum en de Staatsdrukkerij.

Hoewel ik geen futuroloog ben, verwacht ik toch dat juist dit omgaan met natuurlijke taal in de nabije toekomst een geweldig belangrijke rol in de informatica zal gaan spelen. Waarom?

Omdat de elektronische computer in wezen een machine voor informatieverwerking is. En dat is lang niet altijd alleen maar het uitvoeren van rekenwerk. Op dit moment kan al worden vastgesteld dat het echte rekenwerk nog maar een klein bestanddeel vormt van al het werk dat door computers wordt verricht. Dat is niet zo vreemd als het lijkt.

Bij rekenen is er namelijk wel degelijk sprake van informatieverwerking, maar de verwerkte informatie bestaat uit getallen. Getallen op zichzelf bevatten echter nauwelijks of geen informatie als er niet een stuk tekst, een verhaal bij gegeven wordt, waarin staat wat die getallen betekenen. Ik durf te beweren, dat de eigenlijke informatie, waarover wij mensen beschikken, niet kwantitatief maar overwegend kwalitatief van aard is. Of anders gezegd, dat de informatie waarover wij mensen beschikken niet zonder de natuurlijke taal kan worden vastgelegd en overgebracht.

Ik bedoel daarmee gewoon dit. Kunt u iemand uitleggen wat een brug is, een spiraalnevel, een chemische verbinding of wat dan ook, enkel en alleen door gebruik te maken van getallen?

Het klinkt u misschien allemaal wat triviaal in de oren.

Maar de gevolgen voor de informatica zijn niet gering en allerminst triviaal te noemen. Want wat voor ons mensen de gewoonste zaak van de wereld is, namelijk het omgaan met taal, is voor de informatica een gigantisch probleem.

We zien en horen namelijk wel wat er gebeurt als we een taal gebruiken, maar we weten niet hoe het gebeurt. Wat is er bijvoorbeeld het afgelopen half uur hier in de zaal gebeurd? Ik heb u toegesproken en ik hoop dat u

mij hebt verstaan en begrepen. Maar hoe gaat dat in zijn werk? Wat is er in al die hoofden omgegaan?

Ik denk, dat dit taalprobleem ook speelt in de sfeer van het landbouwkundig onderzoek. Ik kan me tenminste nauwelijks voorstellen van niet, omdat ik weet dat bij veel landbouwkundig onderzoek het inventariseren een belangrijke rol speelt. Dat wat geïnventariseerd wordt kan vaak alleen verbaal worden weergegeven. De steeds maar groeiende hoeveelheid gegevens rijst vaak de pan uit. In het beste geval richt men hiervoor kaartsystemen in. Kaartenbak na kaartenbak wordt gevuld.

Aan het eind van het onderzoek is de verzamelde informatie meestal onoverzichtelijk en slecht toegankelijk. Om conclusies te kunnen trekken is het nodig om dwarsdoorsneden er uit te kunnen lichten die combinaties van bepaalde aspecten vertegenwoordigen. Vaak is dit bijna ondoenlijk. Niettemin is deze informatie en de mogelijkheid om er conclusies uit te trekken nou juist het enige wat een dergelijk onderzoek kan opleveren. Daarom verdient het overweging om de hanteerbaarheid van de informatie bij dit soort onderzoeken te verbeteren door het inschakelen van computers.

Ik heb een vermoeden dat het bijvoorbeeld zelden of nooit voorkomt dat er een post 'informatieverzorging' op de begroting van een dergelijk onderzoek voorkomt. En ik ben er zeker van dat dat precies de reden is waarom er nog steeds zo weinig plaats is voor geautomatiseerde informatieverzorging, alias informatica, binnen de lopende onderzoeken, terwijl er eigenlijk wel behoefte aan bestaat. Hier doet zich het verschijnsel voor dat men zich aan de verzorging van het uiteindelijke resultaat van het onderzoek, namelijk de voortgebrachte kennis, te weinig gelegen laat liggen.

In de bouwwereld zou dat vergeleken kunnen worden met het verwaarlozen van juist dié eigenschappen van een gebouw die te maken hebben met het eigenlijke doel ervan, namelijk de bescherming die het gebouw tegen het buitenklimaat moet bieden. Bijvoorbeeld een gebouw met een kartonnen dak om het goedkoop te houden, of ramen van cellofaan, of een gebouw zonder schoorsteen zodat verwarming uitgesloten is.

Hoeveel goedkoper zou je een gebouw niet kunnen maken als je met al die dingen geen rekening hoefde te houden? Of, om op ons onderwerp terug te komen, welk gedeelte van het onderzoekbudget mag men op de post 'informatieverzorging' opvoeren?

In mijn verbeelding hoor ik u nu denken "dat is nou allemaal goed en wel, maar wat kan de informatica voor mij doen in het onderzoek?"

En u hebt gelijk, ik ben het antwoord op deze vraag nog schuldig. Aan de andere kant moet u begrijpen dat ik in het bestek van deze voordracht ook weer niet te diep op de zaak kan ingaan. Ik wil slechts volstaan met het noemen van drie punten waarbij het inschakelen van de informatica een overweging alleszins waard kan zijn.

- Het literatuuronderzoek voorafgaande aan en tijdens het onderzoek.

Er zijn in Nederland instanties die tegen betaling via terminals literatuurrecherches op grote internationale literatuurdatabanken voor u uitvoeren. Voor de landbouw vindt u deze mogelijkheid bij PUDOC.

- Het uitvoeren van berekeningen.

In veel gevallen is het niet nodig om zelf nog programma's te maken die de gewenste berekeningen voor u verrichten. Er zijn rekencentra die de mogelijkheid bieden om via terminals statistische of numerieke berekeningen door standaardprogramma's te laten verrichten. Meestal zijn daar ook mensen te vinden die u daarbij gaarne willen adviseren.

- Inventarisatie van grote hoeveelheden gegevens.

Er bestaan systemen waarmee gegevens die gedeeltelijk of geheel verbaal van aard zijn, via terminals kunnen worden geplaatst in een magnetisch geheugen. Het maken van alfabetische ingangen (indexen) op bepaalde kenmerken geschiedt automatisch. Het selecteren van gegevens op combinaties van kenmerken (retrieval) behoort ook tot de mogelijkheden.

Welk systeem het beste bij uw onderzoek past, kunt u het beste laten adviseren door specialisten op het gebied van de geautomatiseerde informatieverzorging.

Tenslotte wil ik nog even terugkomen op de natuurlijke taal.

Op dit moment schiet de informatica zeker nog te kort waar het gaat om wat ik zou willen noemen 'omgang met de natuurlijke taal door computers'. De meeste information retrieval-systemen bieden de gebruiker een speciale vraagtaal. De gebruiker wordt gedwongen om datgene wat hij wil weten in deze kunstmatige taal te formuleren. Dat kost extra moeite en het pakt lang niet altijd goed uit.

In feite komt het erop neer dat de vragensteller zelf een aantal opdrachten aan de computer geeft. Hij moet zich daarbij goed kunnen realiseren wat de gevolgen zijn, want de computer voert die opdrachten stompzinig en slaafs uit. Er is geen sprake van enig taalbegrip van de kant van de computer. Dit is precies de reden van teleurstellingen die optreden wanneer computers voor het omgaan met gegevens in natuurlijke taal worden gebruikt.

Mijns inziens is de toestand op dit punt inderdaad erg onbevredigend. Natuurlijk is het de informatica als wetenschap die hier soelaas zal moeten bieden. Maar hoe? Allereerst moet opgehelderd worden waar de moeilijkheid precies ligt. Ik bedoel het probleem van de kunstmatige intelligentie. Wat is intelligentie? Het valt te betwijfelen of we moeten streven naar iets als een zelfstandige kunstmatige intelligentie.

Waar het om gaat is in te zien dat de computer in principe een werktuig is. Wat wil dat zeggen? De eerste menselijke werktuigen, maar ook veel moderne werktuigen, zijn in wezen niets anders dan onzelfstandige 'verlengstukken' van de menselijke hand. Pas samen met die hand, en er zeker niet zonder, vormen zij als het ware een compleet werktuig bestuurd door de menselijke intelligentie.

Vergelijkenderwijs zou men de computer dan moeten zien als een even onzelfstandig verlengstuk van het menselijk brein.

De 'greep' van de hersens op dit verlengstuk, het 'handvat', wordt gevormd door de taal. Zoals bij elk werktuig zullen wij ons moeten afvragen welke voordelen wij nastreven door het inschakelen van dit werktuig. Bij computers begint dat zich op sommige punten al duidelijk af te tekenen.

- de bewerkingen worden met zeer grote snelheid feilloos uitgevoerd (alleen de programmeur maakt fouten);
- er treedt geen vermoeidheid op, hoe groot het aantal bewerkingen ook is;
- computergeheugens leiden niet aan geheugenverlies of geheugenstoornis.

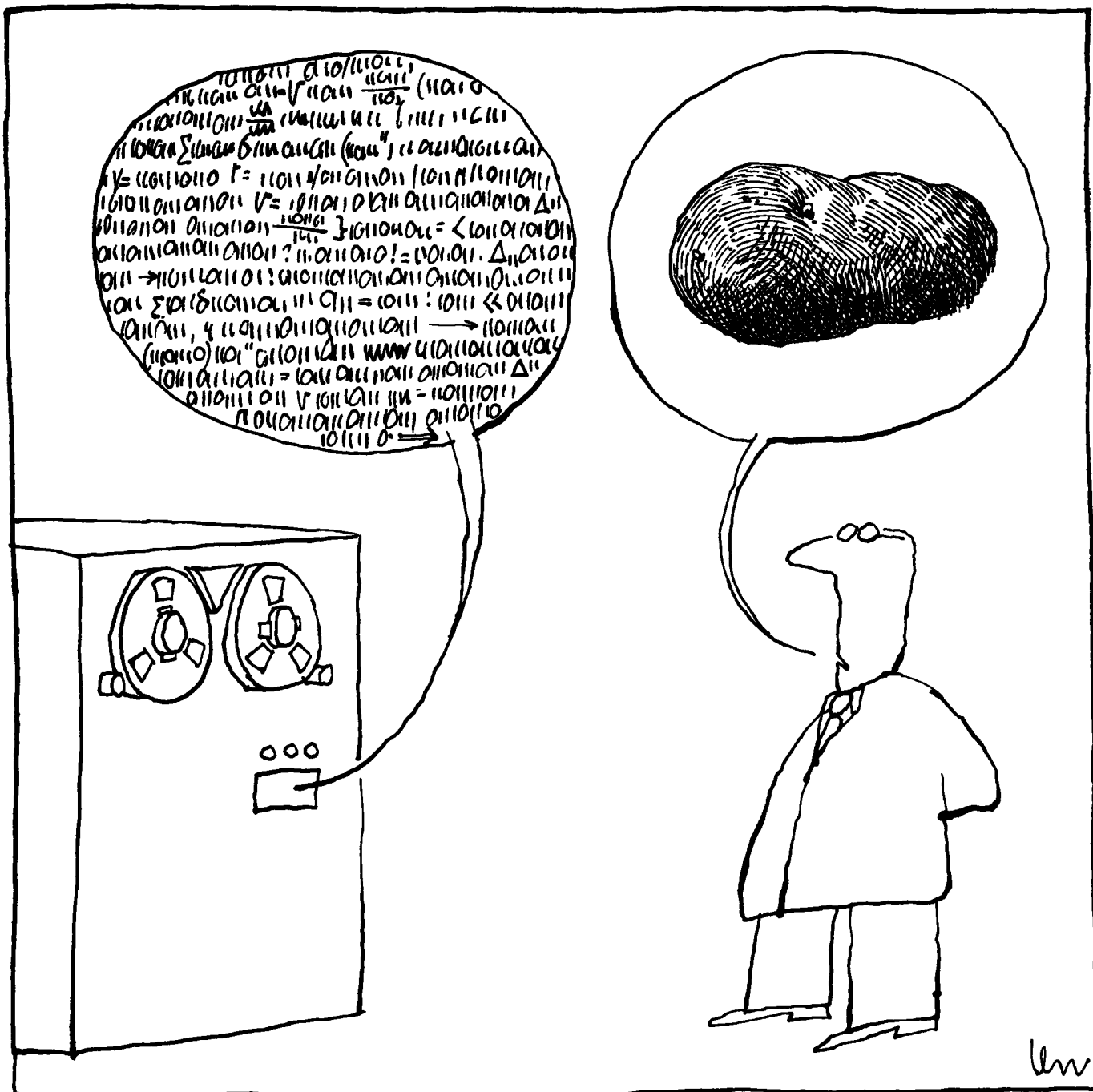
Wanneer moet u er dus over gaan denken om bij een onderzoekproject de informatica te hulp te roepen?

- 1° Als u in korte tijd een vreselijk groot aantal bewerkingen moet verrichten of ingewikkelde berekeningen moet maken, bijvoorbeeld een statistisch onderzoek dat op een groot aantal waarnemingen berust.
- 2° Als u met dezelfde informatie of kennis voortdurend opnieuw iets wilt doen, vooral als u er steeds weer vragen aan of over wilt stellen, bepaalde selecties uit wilt maken of statistische bewerkingen op los wilt laten. Bijvoorbeeld alle grondwaterstanden in Nederland, die op vele plaatsen (roosterpunten) periodiek (eenmaal per maand) worden gemeten die voor velerlei doeleinden steeds opnieuw worden geraadpleegd en onderling vergeleken.
- 3° Als u catalogusachtige boekwerken periodiek wilt uitgeven die voortdurend up to date moeten worden gehouden, bijvoorbeeld een catalogus van het lopend landbouwkundig onderzoek of van landbouwwerktuigen.

Landbouw in de statistiek

Ir. A. A. M. Jansen

Voordracht gehouden op 19 mei 1981 ter gelegenheid van het 25-jarig bestaan van IWIS-TNO afdeling Wageningen



Inhoud

1.	Inleiding	71
2.	De omvang van landbouw en de omvang van statistiek	73
3.	Waarmee een statisticus zich zoal bemoeit	76
3.1.	<i>Met de doelstelling van het onderzoek</i>	76
3.2.	<i>Met de wijze van uitvoering</i>	79
3.3.	<i>Met de omvang van het onderzoek</i>	81
3.4.	<i>Met de analyse van de resultaten</i>	83
4.	Besluit	85

1. Inleiding

Ongeveer 25 jaar geleden werd te Wageningen de dependance gevestigd van de toenmalige Afdeling Bewerking Waarnemingsuitkomsten TNO. In dezelfde periode werd ik aan de Landbouwhogeschool ingewijd in de beginselen van de wiskundige statistiek.

Ik herinner mij dat in de eerste les het begrip 'stochastische variabele' werd geïntroduceerd aan de hand van een landbouwkundig voorbeeld: de diameter van een aselect gekozen aardappel uit een hoop aardappelen. Die hoop aardappelen bevond zich rechts van de hoogleraar, vóór in de zaal, en was alleen via zijn suggestieve gebaren waar te nemen. Wat ik destijds niet zozeer beseftte is dat, terwijl wiskundige statistiek gezien kan worden als een tak van wiskunde, dit niet geldt voor het toepassen daarvan. En dat daarom het suggereren van een echte hoop aardappelen meer dan als didactische steun in de rug zou kunnen worden opgevat. Maar ik ben er niet zeker van of het geven van landbouwkundige voorbeelden niet meer de bedoeling had te laten zien hoe je daarvan kunt abstraheren dan wel aandacht te vragen voor de praktische relevantie van de voorgestelde analysemethode.

Ook in de statistische vakliteratuur worden nieuwe methoden meestal geïllustreerd met behulp van voorbeelden waarvan de gegevens afkomstig zijn uit de praktijk van het onderzoek. Aan dergelijke voorbeelden wordt blijkbaar een grotere overtuigingskracht toegedacht dan aan geconstrueerde voorbeelden. Op het eerste gezicht lijkt dit een wat merkwaardige gewoonte. Immers voor het illustreren van een statistische methode kan het alleen maar een voordeel zijn als men de methode toepast op gegevens die precies de veronderstelde eigenschappen hebben. Wat ligt dan meer voor de hand, zeker in dit computertijdperk, dan deze gegevens zelf door middel van simulatie te construeren? Dit heeft bovendien allerlei bijkomende voordelen, zoals de mogelijkheid het aantal stellen gegevens naar behoefte te kiezen. Men ziet in de literatuur de laatste tijd dan ook geregeld resultaten van zogenaamde simulatiestudies gepresenteerd die tot doel hebben de merites van methoden te beschrijven of te vergelijken. Meestal worden in zulke studies stellen gegevens in grote aantallen gegenereerd; maar er lijkt weinig tegen in te brengen als men ook voor een enkelvoudig voorbeeld de benodigde gegevens zou construeren.

Is dan het kiezen van praktijkvoorbeelden als illustratie van een methode alleen maar een historisch gegroeide gewoonte, waarvoor eigenlijk geen goede motivatie gegeven kan worden? Een gewoonte die misschien zijn oorsprong vindt in het natuurwetenschappelijk onderzoek, waar het gemeengoed is geworden dat wetenschappelijke theorieën aan waarnemingen moeten worden getoetst? Of geldt ook hier dat men de relevantie van de methode als het ware wil bewijzen door te laten zien dat men er in een gegeven praktische situatie metterdaad iets mee kan doen?

Ik ben geneigd te veronderstellen dat het laatste het geval moet zijn. Men kan vaststellen dat de in de literatuur gegeven voorbeelden lang niet altijd bijdragen tot een beter begrip van de voorgestelde methoden of tot een beter inzicht in de toepassingsmogelijkheden. Soms zelfs blijkt het zeer moeilijk om overtuigende voorbeelden te vinden, zoals bijvoorbeeld de literatuur over variantieanalyse voor meer eigenschappen tegelijk, of over methoden voor simultane uitspraken over de verschillen tussen een aantal gemiddelden, bewijst. Het voert te ver nu dieper hierop in te gaan. Ik kan mij echter niet onttrekken aan de indruk dat het zoeken en geven van relevante toepassingen niet de sterkste zijde van sommige capita uit de statistische literatuur is.

In de statistische literatuur, en ook in het onderwijs, is nauwelijks te vermijden dat men vanuit de methodologie zoekt naar toepassingen. In de praktijk van het onderzoek echter ligt dit precies omgekeerd. Het gegeven is het onderzoeksgebied, met de daarop betrekking hebbende onderzoeksvragen, waarvoor een oplossing wordt gezocht. Wiskundige statistiek is daarbij een wellicht onmisbaar hulpmiddel, maar blijft desondanks ondergeschikt aan het gestelde doel. Dit houdt in dat de rechtvaardiging voor de formulering van hypothesen en voor de keuze van een model voor de waarnemingen en daarop gebaseerde methoden vanuit dat onderzoeksgebied moet worden gegeven. Daarvoor is nodig dat de onderzoeker naast een gedegen kennis van zijn eigen vakgebied voldoende inzicht heeft in de toe te passen statistische methoden. Het behoeft geen betoog dat het in de praktijk wenselijk kan zijn de verantwoordelijkheid voor de statistische aspecten van een onderzoek over te dragen aan een specialist, zoals dat wellicht ook het geval zal zijn voor bijvoorbeeld chemische of

physische aspecten daarvan. Immers, men kan nauwelijks van een onderzoeker verwachten dat hij zijn kennis op het toch wel specialistisch vakgebied, dat de statistiek inmiddels is geworden, afdoende op peil kan brengen en houden zonder zich een onevenredige inspanning te moeten getroosten. Maar een samenwerking van een onderzoeker met een statisticus heeft alleen kans van slagen als zij belangstelling voor elkaars vakgebied kunnen opbrengen. Voor de statisticus betekent dit dat hij moet trachten tot een zelfstandige beoordeling te komen van de bruikbaarheid van het beschikbare statistische arsenaal voor de specifieke situatie in het onderzoek waarbij hij betrokken is. Hij dient te beseffen dat de waarde van de informatie die uit gegevens wordt verkregen niet alleen afhangt van de kwaliteit van die gegevens, maar ook in belangrijke mate van het gestelde model, van de formulering van de hypothesen en van de daarop gebaseerde analyses. Toepassen van statistiek is dan niet zozeer een wiskunde-vak, maar meer een verlengstuk van het onderzoeksgebied waarop het wordt toegepast. Ik denk dat de aanduiding 'biometrie', of zo men wil 'agrometrie', dit accent-verschil met wiskundige statistiek tot uitdrukking wil brengen.

In deze voordracht wil ik proberen het werkterrein van een consulterend statisticus te illustreren juist vanuit zijn betrokkenheid bij het onderzoek. Ik zal dit doen aan de hand van een aantal praktische voorbeelden, die elk een bepaald facet van de bemoeienis van een statisticus belichten. Het verhaal gaat dus in eerste instantie over landbouw en landbouwkundig onderzoek, en slechts indirect over statistiek.

2. De omvang van landbouw en de omvang van statistiek

Zoals gezegd is het wenselijk dat een statisticus, die zich met consultatie bezighoudt, zich enige kennis eigen maakt van het onderzoeksterrein waarin hij actief is. Daarom is het nuttig om te beginnen met een vluchtige verkenning van het toepassingsgebied: landbouw.

Bladeren in het boekwerk 'Wageningen, Centre of Agricultural Research' maakt spoedig duidelijk dat de term 'landbouw' zeer veel disciplines omvat.

Ik noem er enkele, in een willekeurige volgorde:

tuinbouw, bosbouw, veeteelt, veevoeding, diergeneeskunde, plantenziektekunde, atoomenergietoepassingen, bewaring en verwerking van produkten, natuurbeheer, cultuurtechniek, bodemkartering, mechanisatie en gebouwen, visserij.

Op elk van deze terreinen zijn een of meer gespecialiseerde onderzoeksinstituten werkzaam en vanuit elk van deze terreinen bereiken ons verzoeken om statistisch advies. Deze kunnen betrekking hebben op traditionele landbouwkundige onderwerpen, maar even goed op bijvoorbeeld de kwaliteitseigenschappen van dubbel glas voor toepassingen in kassen, of op de aantrekkelijkheid van dagrecreatie in het kader van een landinrichtingsprobleem. Maar zelfs binnen één instituut treft men vaak een enorme variatie in onderzoeksprojecten aan van waaruit een beroep op statistische consultatie wordt gedaan. Bijvoorbeeld bij het Instituut voor Veeteeltkundig Onderzoek in Zeist, waar ik zelf betrokken ben bij de statistische consultatie, treft men projecten aan op het gebied van fokkerij, voortplanting, huisvestingsonderzoek, bedwelmingsonderzoek, classificatie van geslachte dieren, vleeskwaliteitsonderzoek, melkerijonderzoek. Daarbij komt men onder meer in aanraking met onderwerpen zoals: kwantitatieve genetica, gedragsleer, druksterktemeting van schedels, beoordeling door keurmeesters (van vorm, kleur, geur, smaak, malsheid), calibratie van meetmethoden, anatomie van de huisdieren, capita uit de diergeneeskunde, meting van hormonale activiteit, enzovoort. Zelfs binnen een instituut is het nauwelijks doenlijk voor een statisticus het door dat instituut bestreken werkterrein te overzien. Om een zinvolle bijdrage te kunnen leveren aan de onderzoeksprojecten waarvoor statistische hulp wordt ingeroepen, blijkt een voortdurende wetenschappelijke nieuwsgierigheid op allerlei gebied noodzakelijk.

Het spreekt vanzelf dat een statisticus, die zich met consultatie bezighoudt, zo vertrouwd moet zijn met de gangbare methoden zoals regressie- en variantieanalyse, en technieken zoals het statistisch verantwoord opzetten van experimenten, dat hij hun voordelen en nadelen in een praktische situatie gemakkelijk overziet. Maar naast de op de praktijk van het onderzoek gerichte nieuwsgierigheid moet hij veel belangstelling kunnen opbrengen voor de ontwikkeling van de toepasbare methodologie op het gebied van de statistiek zelf. Dit vraagt een geregelde oriëntatie

over wat er zoal aan bruikbaar is gepresenteerd is en wordt in de wetenschappelijke literatuur, ook die op het terrein van het onderzoek. Zoals gesteld verschaft de literatuur lang niet altijd een helder inzicht in de praktische relevantie van voorgestelde methoden, zodat onderzoek daarnaar door de statisticus geboden kan zijn. Ook het signaleren van het ontbreken van een geëigende methodologie voor het oplossen van een probleem dient de nodige aandacht te krijgen, en waar mogelijk moet deze worden ontwikkeld of aangepast. Tenslotte wordt verwacht dat een statisticus deskundige hulp kan bieden bij het verwerken van gegevens via een computer, al dan niet gebruik makend van beschikbare programmapakketten, en rekening houdend met de eigenschappen van de daarbij gebruikte numerieke methoden.

Een statisticus moet dus, om goed te kunnen functioneren, van vele markten thuis zijn en kwaliteiten bezitten die nauwelijks alle in één persoon zijn te verenigen. Het behoeft daarom geen betoog dat ook hij behoefte zal hebben aan een zekere specialisatie. In de huidige praktijk blijkt zo'n specialisatie tot op zekere hoogte aanwezig te zijn. Zo besteedt iedere IWIS-statisticus zijn tijd grotendeels aan problemen van één, of hoogstens twee instituten voor landbouwkundig onderzoek. Dit brengt een zekere beperking met zich mee, zowel ten aanzien van de breedte van het toepassingsgebied als ten aanzien van het spectrum van frequent gehanteerde statistische methoden. Uiteraard geldt dit ook voor statistici die direct bij de instituten in dienst zijn. Voor een deel wordt het beheer van gegevensbestanden en het rekenwerk dat uit de consultatie voortvloeit, overgelaten aan daarin geschoolde krachten, hoewel de tendens tot meer zelf doen sterker lijkt te worden naarmate de programmatuur daarvoor hanteerbaarder wordt. Men kan zich afvragen of men in een centrale opstelling van statistici verdergaande specialisatie zou moeten nastreven, of dat men juist de statistici zou moeten laten rouleren over verschillende toepassingsgebieden om hen een zo breed mogelijk gezichtsveld te bieden. Dit valt echter buiten het kader van deze voordracht zodat ik daarop niet nader zal ingaan.

3. Waarmee een statisticus zich zoal bemoeit

Het is niet doenlijk om in dit bestek een volledig beeld te schetsen van de wijze waarop een statisticus bij het toegepaste onderzoek functioneert. Zijn bemoeienis strekt zich vaak uit over het gehele gebied vanaf het formuleren van de onderzoeksdoelen tot en met de publikatie van de resultaten. Ter wille van de duidelijkheid is het goed hier vast te stellen dat dit niet impliceert dat hij ook verantwoordelijkheid zou dragen voor alle keuzen die mede op basis van zijn bemoeienis worden gemaakt: deze ligt uiteraard bij de onderzoeker zelf. Wel is een statisticus verantwoordelijk voor zijn adviezen over de inrichting van het onderzoek, over de wenselijke omvang daarvan, en over de te gebruiken statistische procedures. Maar om de deugdelijkheid van zijn eigen adviezen te kunnen beoordelen is het nodig dat hij diepgaand kennis neemt van het doel van het onderzoek en van de beschikbare mogelijkheden; vaak zal het daarbij nodig zijn te sleutelen aan formuleringen en randvoorwaarden. Ik zal aan de hand van een aantal voorbeelden trachten te schetsen waarmee een statisticus zich zoal bemoeit, en waarom.

3.1. *Met de doelstelling van het onderzoek*

Een recent concreet voorbeeld betreft het onderzoek naar de fractie schaalbreuk van eieren, gelegd door kippen in legbatterijen. Deze economisch belangrijke grootheid kan onder meer afhangen van een aantal factoren die de kooiconstructie betreffen, zoals de helling van de kooibodems, de wijze van ophanging daarvan en dergelijke. Wil men het effect van deze factoren onderzoeken dan is een standaard mogelijkheid die men aantreft in alle boeken over proefopzetten een zogenaamde 2^n proef, waarin men alle combinaties van factoren, elk ingesteld op twee niveaus, onderzoekt. Als 2^n te groot wordt met betrekking tot het vereiste of beschikbare aantal experimentele eenheden, zijn er mogelijkheden om de proef slechts gedeeltelijk uit te voeren zonder op wezenlijke punten informatie te verliezen. Bij alle opzetten is het nodig een aantal varianten van de kooiconstructie te onderzoeken. De onderzoekers blijken een sterke voorkeur te hebben voor het bijeenplaatsen van identieke eenheden omdat men beducht is voor allerlei vormen van onderlinge beïnvloeding; de stabiliteit van een batterij van kooien blijkt namelijk te worden bepaald door de wijze waarop de afzonderlijke eenheden met elkaar worden

verbonden en dus wellicht ook door de schikking van de varianten die elkaars burens vormen. Het ligt voor de hand te eisen dat de in het experiment vastgestelde verschillen in schaalbreuk overdraagbaar moeten zijn naar situaties waarin de batterij geheel bestaat uit eenheden van gelijke constructie. In zo'n geval dienen diverse alternatieve mogelijkheden voor de proefopzet op hun merites te worden onderzocht. Daarbij kan men denken aan vergaande beperkingen ten aanzien van de plaatsing van de varianten, met als bezwaar de bijbehorende beperkingen voor de statistische analyse. Of aan het opnemen van veel buffers tussen verschillende varianten, wat een relatief inefficiënte benutting van de beschikbare proefaccommodatie met zich meebrengt. Ook kan men denken aan het onderling onafhankelijk bevestigen van de eenheden, wat gevolgen heeft voor de definitie van de behandelingen. Welke mogelijkheid moet worden gekozen hangt af van velerlei omstandigheden, waarop ik hier niet zal ingaan. De hier gemaakte keuze is echter bepalend voor de extrapoleerbaarheid van de resultaten.

Zulke voorbeelden kan men in allerlei vormen aantreffen. Zonder op details in te gaan is het wellicht instructief nog enkele andere voorbeelden kort aan te duiden.

- Bij het onderzoek naar de effectiviteit van verschillende bestrijdingsmiddelen voor mobiele insecten, bijvoorbeeld de peenvlieg, moet men zijn experimentele eenheden zodanig kiezen dat de gemeten schade overeenkomt met de schade die zou worden gevonden als het gehele veld op gelijke wijze zou worden behandeld. Dit betekent onder meer dat men middelen met een verschillende effectiviteit niet zonder meer direct naast elkaar op relatief kleine veldjes kan beproeven. Doet men dit wel, dan worden waarschijnlijk de resultaten, zowel die van de effectieve als die van de niet effectieve middelen, vertekend. Hoe groot die veldjes dan wel moeten zijn, of hoe groot de rand tussen de veldjes moet zijn, kan een statisticus niet vanuit zijn specialisme beantwoorden. Hij kan wel behulpzaam zijn bij het formuleren van de daarop betrekking hebbende eisen of bij het verzamelen van relevante informatie daarover.

- Bij het vergelijken van drie rassen van melkkoeien ten aanzien van hun graastijd in de weide is het om allerlei redenen aantrekkelijk alle dieren van de drie rassen samen in te scharen in dezelfde weide. Het onderzoek veronderstelt uiteraard dat er verschillen zullen zijn. Zullen dan die verschillen niet worden vertekend doordat het graasgedrag van elk ras wordt beïnvloed door de aanwezigheid van beide andere rassen in dezelfde weide? Omdat deze onderlinge beïnvloeding gestrengeld is met de verschillen tussen de rassen, kan men deze niet afzonderlijk in handen krijgen, en daarom is het verstandig in zo'n geval een proefopzet te kiezen waarin onderlinge beïnvloeding is uitgesloten. Immers, niemand weet hoe groot deze is, en in hoeverre deze de gezochte relatief kleine verschillen tussen de rassen vertekent. Dat onderlinge beïnvloeding van dieren kwantitatief belangrijk kan zijn, kan nauwelijks in twijfel worden getrokken gezien de plaats die voor gedragsonderzoek is ingeruimd in het veeteeltkundig onderzoek.

- Bij onderzoek naar milieu-genotype interactie is het in principe de bedoeling na te gaan in hoeverre de genotypen in het ene milieu op dezelfde manier kunnen worden gerangschikt als in het andere milieu. Het hierbij gehanteerde interactiebegrip is beperkter dan het klassieke interactiebegrip, dat naast volgordeverschillen ook de absolute grootte van de verschillen tussen de genotypen in de beschouwde milieus omvat. Afgezien hiervan kan men in principe zowel de milieus als de genotypen aselekt kiezen uit een gegeven verzameling milieus of uit een gegeven verzameling genotypen. Het hangt enerzijds van de formulering van de doelstellingen af en anderzijds van de randvoorwaarden van het onderzoek of het wenselijk is de in het experiment op te nemen milieus, of genotypen, of beide, niet aselekt maar zeer gericht te kiezen. Zo zal men bijvoorbeeld voor de vraag in hoeverre het mogelijk is onder beperkte voedingsomstandigheden te selecteren naar de groeipotentie van slachtdieren onder rijkere omstandigheden, of omgekeerd, de te onderzoeken genotypen aselekt kiezen uit de relevante populatie, maar de milieukeuze beperken tot extremen.

3.2. Met de wijze van uitvoering

De zojuist genoemde voorbeelden hebben vooral tot doel duidelijk te maken dat het belangrijk is zorgvuldig na te gaan of de bij het onderzoek verkregen resultaten met vertrouwen zullen kunnen worden geëxtrapoleerd naar de omstandigheden die in de praktijk gelden. Zijdelings kwam daarbij een volgend aspect aan de orde, dat bij veel onderzoek van groot belang is, namelijk de keuze van de experimentele eenheden. Dat zijn de eenheden die men onafhankelijk van elkaar kan toewijzen aan de te beproeven behandelingen. Bij veldproeven zijn dat meestal afzonderlijke veldjes, maar het kunnen ook veel grotere eenheden zijn, waarvan de afmetingen afhangen van de eisen van de mechanische bewerkingen die in het onderzoek worden betrokken. Bij proeven met dieren kunnen afzonderlijke dieren experimentele eenheden zijn, maar ook groepen van dieren (tomen, hokken, stallen) of achtereenvolgende perioden gedurende het leven van een dier, zoals bij wisselproeven het geval is. De keuze van die experimentele eenheden, en vooral de manier van toewijzen daarvan aan de behandelingen, bepaalt de vorm van de statistische analyse voorzover deze is gebaseerd op door aselekt toewijzen geïnduceerde onderlinge onafhankelijkheid van de gegevens.

Als niet voldaan is aan de eis dat de gegevens onderling onafhankelijk zijn moet men bij de analyse voor de afhankelijkheid een model stellen, de relevantie daarvan aantonen en de parameters daarvan schatten. Dit bemoeilijkt de analyse en maakt de resultaten, zeker bij kleine aantallen gegevens, meer afhankelijk van subjectieve modelveronderstellingen. Voorbeelden waarbij men zou kunnen denken aan speciale modellen voor de afhankelijkheid van experimentele eenheden is het onderzoek in kassen (waarbinnen men de optredende variatie met de traditionele proefveldtechnieken niet of nauwelijks adequaat kan modelleren), of onderzoek naar de afhankelijkheid van bronstwaarnemingen in een stal met individueel gehuisveste zeugen.

Een ander belangrijk aspect betreft de vraag of men een experiment zal uitvoeren waarin men de instellingen van een aantal factoren in de hand heeft, dan wel of men gegevens zal verzamelen uit de praktijk of uit het vrije veld. Men kan niet altijd vrij kiezen tussen deze mogelijkheden. Soms wijst de doelstelling van het onderzoek eerder in de richting van

een inventarisatie, soms ook is het in het geheel niet mogelijk een experiment uit te voeren. Maar in de meeste gevallen waarin de keus wel open is, is er alle aanleiding aan het uitvoeren van een experiment de voorkeur te geven. Ik zal trachten dit met enkele voorbeelden duidelijk te maken.

- Bij een praktijkonderzoek naar het verschil tussen ligboxenstallen en grupstallen zal men zelfs bij een zo vergelijkbaar mogelijke keuze van bedrijven in beide groepen steeds beducht moeten zijn voor niet geregistreerde, maar wel met de door de bedrijven gerealiseerde keuze van het staltype samenhangende factoren, zoals de vooruitstrevenheid en het economisch inzicht van de bedrijfsleider, of wellicht ook minder voor de hand liggende factoren. Men zal een in zo'n onderzoek vastgesteld verschil altijd met grote voorzichtigheid moeten hanteren omdat geenszins vaststaat dat dit verschil ook metterdaad zal optreden als willekeurige boeren van het ene staltype op het andere overgaan bij overigens gelijkblijvende omstandigheden. Het is bij dit type onderzoek moeilijk te bewijzen dat de gevonden verschillen voor het doel van het onderzoek relevant zijn. Hoewel dit de waarde van zulk onderzoek beperkt is er hiervoor geen werkelijk alternatief: men kan niet vrijelijk het staltype van bestaande bedrijven veranderen. Evenmin kan men zijn uitspraken baseren op toevallig in de praktijk optredende overgangen van het ene naar het andere staltype, omdat deze meestal zullen samenhangen met toevallig aanwezige omstandigheden.

- Praktijkonderzoek zoals in het vorige voorbeeld blijft dus altijd een min of meer hachelijke onderneming, omdat men door goed nadenken allerlei mogelijk storende invloeden moet proberen op te sporen om deze vervolgens bij de analyse te kunnen uitsluiten of kwantificeren.

Daarom is onderzoek waarin men de instelling van de invloedsfactoren in de hand heeft veel aantrekkelijker, en dit heeft als het uitvoerbaar is altijd de voorkeur. Maar ook bij op gerichte wijze opgezette experimenten kunnen zich problemen voordoen die verwant zijn met die in het voorgaande praktijkonderzoek. Stelt u zich bijvoorbeeld de volgende situatie voor. In een onderzoek naar de invloed van de zaai- diepte op de opkomst van suikerbieten worden grote verschillen in de standdichtheid van de bieten verkregen. Uiteraard werken standdicht-

heidsverschillen door op de opbrengstverschillen. Kan men nu met de bij zo'n proef verkregen gegevens de relatie tussen de opbrengst en de standdichtheid schatten? Het antwoord moet zijn dat dit niet kan: de bij dit onderzoek verkregen verschillen in standdichtheid zijn direct gestrengeld met de verschillen in de eigenschappen van de opgekomen zaden. Immers bij de grote zaaidiepte komen alleen de zeer vitale zaden tot ontwikkeling en bij de geringe zaaidiepte ook zaden met een geringe vitaliteit. En wie kan zeggen dat vitaliteit van een zaad en opbrengst van de plant van elkaar onafhankelijk zijn?

Dit voorbeeld herinnert mij aan een voorval dat kan dienen om te illustreren wat het effect kan zijn van het niet onderkennen van de invloed van een factor. Bij een van de proeven werden de zaaidiepten gerealiseerd door de zaai-elementen aan één machine links en rechts verschillend in te stellen en deze instelling van zaaigang tot zaaigang te variëren. Bij de analyse van de resultaten bleek dat er een allesoverheersende interactie was opgetreden tussen de zaaidiepten en de kant van de machine waar deze waren gerealiseerd. Dit gaf zowel bij de onderzoeker als bij de statisticus aanleiding tot allerlei gissingen over hoe dit tot stand kon zijn gekomen. Pas veel later bleek dat bij het zaad van de ene helft van de machine een onschuldige geordeelde kleuring was toegepast om de zaden voor een directe controle van de zaaidiepte in de grond te kunnen terugvinden. Kiemproeven bevestigden dat deze kleuring de oorzaak was van de teleurstellende resultaten.

3.3. *Met de omvang van het onderzoek*

Een stereotype vraag aan het adres van een statisticus betreft altijd het gewenste aantal waarnemingen. Bij veel onderzoekers leeft de hoopvolle verwachting dat de statisticus kans zal zien hierover een absolute uitspraak te doen die voortkomt uit de eisen die de statistische methodologie stelt. Bovendien hoopt hij dat het georakelde getal precies zal overeenkomen met de beschikbare faciliteiten. Een statisticus zal die vraag altijd terugspelen naar de onderzoeker in die zin dat deze zelf moet aangeven hoe groot de precisie van de resultaten of hoe groot het onderscheidingsvermogen van de proef moet zijn, en moet opsommen wat hij weet over de te verwachten variatiebronnen. Met informatie hierover,

eventueel in de vorm van verstandige gissingen, is het mogelijk vooraf alternatieve proefopzetten onderling te vergelijken en de gewenste omvang van een proef vast te stellen. Mede op grond daarvan kan een weloverwogen keuze uit alternatieve proefopzetten worden gemaakt, of kan zelfs van een onderzoek worden afgezien omdat met de beschikbare faciliteiten geen acceptabel resultaat kan worden verwacht.

Ter illustratie wil ik even bij één voorbeeld stilstaan. Het betreft de vraag naar de waarde van het vaststellen van de groeiprestatie van fokstieren tijdens hun eigen opfok, waarbij een beperkt voederregiem wordt aangehouden, voor het voorspellen van de groeiprestaties van hun nakomelingen onder mesterijcondities, die een veel ruimer voederregiem inhouden. Relevant voor het beantwoorden van die vraag is de regressie van de prestatie van de nakomelingen op de prestatie van de vaders. Beschouwingen over de gewenste omvang van de proef kan men dus ophangen aan de precisie waarmee die regressiecoëfficiënt kan worden geschat. Deze is echter niet een eenvoudige functie van het aantal waarnemingen. Men moet immers rekening houden met verschillen in verwantschap tussen de nakomelingen van dezelfde stier, en tussen nakomelingen van verschillende stieren. Variantie-verschillen die hiermee samenhangen kan men goed beschrijven met behulp van de erfelijkheidsgraad van en correlaties tussen de beschouwde kenmerken, en daarvoor kunnen meestal wel enigszins bruikbare gissingen worden gemaakt. Als proefopzetten komen in aanmerking alternatieve mogelijkheden met betrekking tot de keuze van de vaders waarvan men nakomelingen onderzoekt. Uiterste mogelijkheden zijn de volgende twee. Van alle vaders waarvan nakomelingen beschikbaar zijn ten tijde van het onderzoek worden een bepaald aantal nakomelingen onderzocht, of er worden alleen nakomelingen onderzocht van de twee vaders die de hoogste en de laagste groeiprestatie lieten zien; dan uiteraard met evenredig meer nakomelingen per vader. Deze uitersten zijn in het algemeen beide niet optimaal. De beste strategie blijkt te zijn het kiezen van meer dan één nakomelingengroep aan beide zijden van het traject van eigen groeiprestaties van de vaders. Hoeveel groepen precies kan men berekenen met behulp van acceptabele veronderstellingen over de grootte van de relevante parameters. Daarmee kan ook de wenselijke omvang van de proef worden vastgesteld. Voor dit voorbeeld is gebleken dat een aantal jaren onderzoek nodig is om een bruikbare precisie te

bereiken; het is uiteraard nuttig om vooraf hierin een behoorlijk inzicht te hebben.

Men hoort wel eens beweren, dat om te kunnen zeggen hoe groot een experiment moet zijn men van te voren moet weten wat er uit komt. Dit is in principe juist, maar het is geen rechtvaardiging om dan maar af te zien van een zo goed mogelijke gissing, die gebaseerd is op aanwezige kennis en op verwachtingen over de resultaten voordat het experiment wordt uitgevoerd. Zulke gissingen zijn beter naarmate deze op betere informatie gebaseerd zijn. Helaas moet worden geconstateerd dat onderzoekers vaak weinig oog hebben voor de waarde van informatie over de aard en grootte van optredende toevalsvariatie. Dit blijkt uit de weinige documentatie die daarover uit vroeger onderzoek of uit de literatuur wordt bijgehouden en ook uit de geringe animo om zulke informatie in publikaties op te nemen. Ik reken het tot de taak van de statistici om bij de onderzoekers aandacht hiervoor te vragen, en er bijvoorbeeld op aan te dringen dat in publikaties volledigheid wordt nagestreefd met betrekking tot het vermelden van zulke informatie.

3.4. Met de analyse van de resultaten

De analyse van de gegevens die beschikbaar komen bij een goed opgezet experiment hoeft in principe weinig problemen op te leveren. Immers, van de aanvang af is gedacht aan een bepaald model voor de waarnemingen, en is getracht de onderzoeksvragen duidelijk te formuleren. De statistische analyse is daarmee in feite vastgelegd, en het berekenen van de gewenste toetsen en schattingen van parameters met hun precisie een routinematige zaak. Dit is de theorie. In de praktijk heeft men te maken met allerlei oorzaken die kunnen noodzaken tot afwijken van het uitgestippelde pad. Ik noem er enkele: het optreden van uitbijters, de verdeling van de waarnemingen is duidelijk anders dan verondersteld, het ontbreken van waarnemingen, al of niet opzettelijke afwijkingen van het proefplan, waarnemingen over onverwacht storende invloeden, waarnemingen over meer en over andere variabelen dan oorspronkelijk de bedoeling was. Al zulke zaken vragen aandacht, en de te nemen actie hangt af van de omstandigheden. Ik heb u aan de hand van het voorbeeld van de bietenopkomstproef reeds laten zien dat het gewenst is voorzichtig te zijn met het schatten van relaties waarvoor de proef niet is bedoeld. Ook, dat het vaak loont

na te gaan of onverwachte merkwaardige uitkomsten misschien berusten op aanwijsbare defecten in de proefprocedure of in de modelveronderstellingen. De aard van de waarnemingen kan het nodig maken robuuste analysetechnieken te hanteren die niet of minder gevoelig zijn voor afwijkingen van de veronderstellingen ten aanzien van hun verdeling.

Statistici kunnen ook belangrijke bijdragen leveren aan onderzoek dat niet of veel minder experimenteel, maar meer beschrijvend of inventariserend van aard is. Men treft zulk onderzoek veel aan op instituten zoals het Rijksinstituut voor Natuurbeheer, de Stichting voor Bodemkartering, en het Instituut voor Cultuurtechniek en Waterhuishouding, maar ook op de meeste andere instituten wordt dergelijk onderzoek wel uitgevoerd. Daarbij kan met vrucht gebruik worden gemaakt van methodologie die, mede onder invloed van de snelle evolutie van de rekenmogelijkheden, recentelijk is ontwikkeld of uitvoerbaar geworden.

Men denke bijvoorbeeld aan allerlei vormen van multivariate analyse zoals clusteranalyse en matrix-approximatiemethoden, aan methoden voor het schatten van de parameters van ruimtelijke correlatiemodellen, of aan efficiënte steekproefmethoden.

Een voorbeeld van een toepassing van matrixapproximatie kan worden ontleend aan het onderzoek naar de diversiteit van soorten in de ecologie. De diversiteit van de soorten in een ecosysteem wordt beschreven door een lijst van de soorten, elk met hun mate van voorkomen (hun abundantie). Omdat dit een tamelijk onoverzichtelijke lijst is, tracht men deze te comprimeren tot een frequentieverdeling van de abundanties en zelfs tot één parameter ter beschrijving daarvan, die dan een diversiteitsmaat wordt genoemd. Verschillende ecosystemen worden vervolgens vergeleken met behulp van zo'n diversiteitsmaat. Het behoeft nauwelijks betoog dat dit een gebrekkige procedure is omdat elke informatie over de soorten zelf daarbij verloren is gegaan: twee ecosystemen met dezelfde diversiteitsmaat kunnen er wat betreft het soortenassortiment heel verschillend uitzien. Veel vruchtbaarder blijkt een aanpak te zijn waarbij men de tabel van ecosystemen \times soorten, met in de cellen de betreffende abundanties, tracht te beschrijven met een beperkt aantal parameters ter karakterisering van de verschillen tussen de ecosystemen en een beperkt aantal parameters ter karakterisering van de soorten. Met behulp van

deze parameters kan men zogenaamde biplots maken; dit zijn grafische voorstellingen die de gegevensmatrix benaderen en waarin men de samenhangen tussen ecosystemen en soorten gemakkelijk kan overzien.

4. Besluit

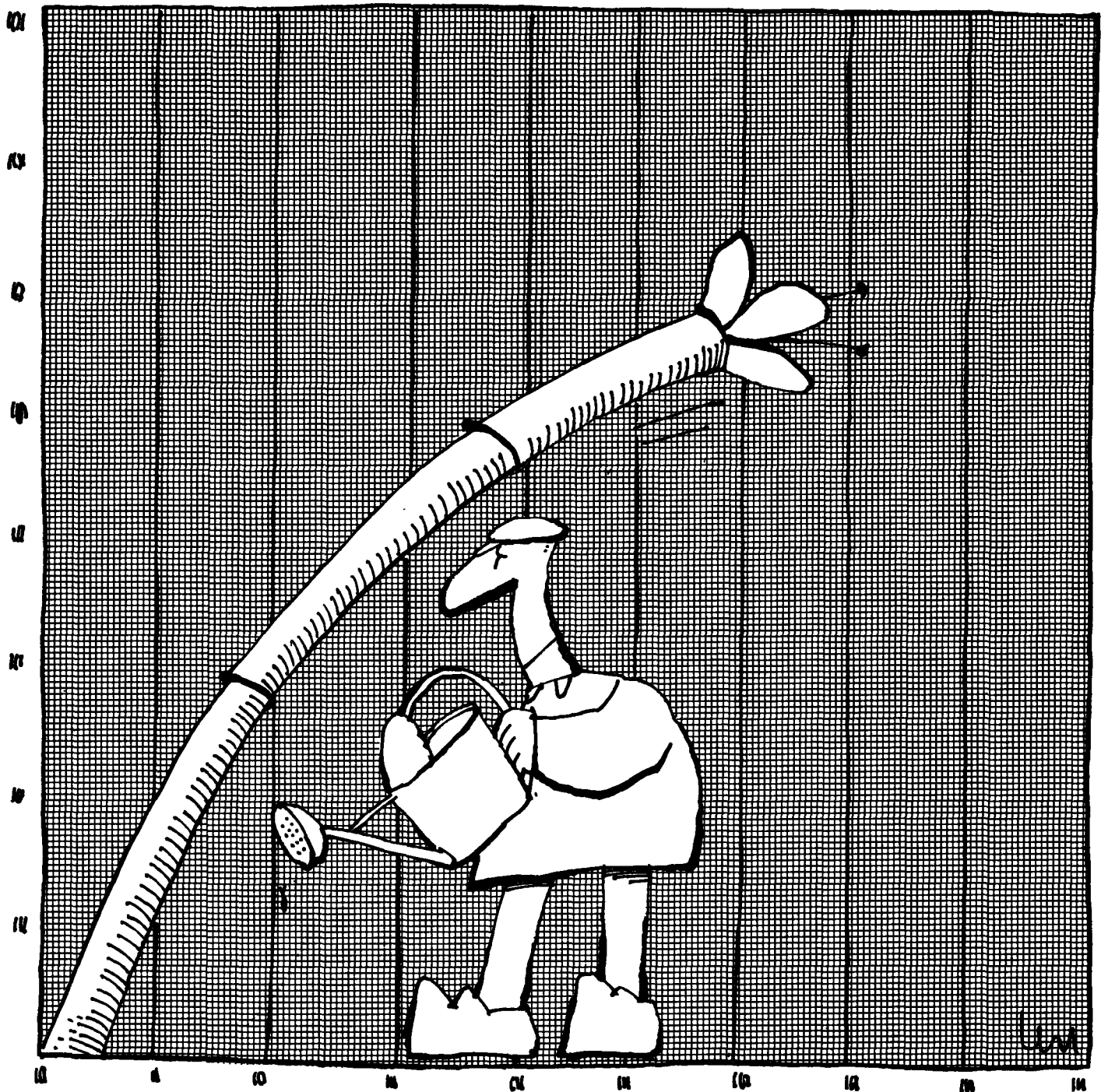
De voorbeelden die ik heb aangedragen om het werk van een statisticus bij het landbouwkundig onderzoek enigszins te illustreren geven uiteraard slechts een zeer onvolledig beeld van wat er zoal ter tafel kan komen. Ik heb getracht te laten zien dat een statisticus een wezenlijke bijdrage kan leveren aan het onderzoek, vanaf het moment dat het wordt opgezet tot en met de publikatie van de resultaten. Maar is het ook mogelijk de waarde van die bijdrage te evalueren, dat wil zeggen in geld uit te drukken? Een redelijke vraag, vooral in deze tijd waarin elke gulden nog een keer schijnt te worden omgedraaid alvorens hij wordt uitgegeven. Helaas zie ik geen kans precies aan te geven wat de waarde van de statistische hulp is. Op de schaal van een enkel experiment kan men wellicht nagaan hoeveel efficiënter de gewenste informatie kan worden verkregen met een statistisch verantwoorde opzet van het onderzoek, in vergelijking met een opzet die die kwalificatie mist. Maar hoe moet men een waarde toekennen aan het om statistische redenen niet doorgaan van onderzoeksplannen, of het geheel anders formuleren van doelstellingen, zodanig dat deze duidelijker en meer bereikbaar worden? Hoe moet men het meer verantwoord trekken van conclusies waarderen? Aan een publikatie, versierd met P-waarden en standaardafwijkingen, is lang niet altijd de kwaliteit van het onderzoek te herkennen.

Er lijkt eigenlijk maar één mogelijkheid voor een evaluatie en dat is het meten van het verschil tussen het rendement van elke gulden die aan het landbouwkundig onderzoek wordt uitgegeven met en zonder statistische begeleiding. Dit is vanuit statistisch oogpunt gezien een onuitvoerbaar onderzoek, en moet daarom worden ontraden. U zult het daarom ook in de toekomst moeten blijven doen met gedeeltelijk subjectieve evaluaties per situatie. En verder met de stellige verwachting dat een zorgvuldige statistische begeleiding van het landbouwkundig onderzoek, waarbij een nauwe samenwerking met onderzoekers en instituten wordt nagestreefd, in

belangrijke mate de kwaliteit van het onderzoek ten goede komt. Ik hoop tenminste dat ik erin geslaagd ben met mijn betoog en de daarin verwerkte voorbeelden aan te tonen of te bevestigen dat die verwachting niet zonder grond hoeft te worden uitgesproken.

Enkele aspecten van de analyse van groeicurven

Ir. J. Jansen



Inhoud

	Samenvatting	91
1.	Inleiding	91
2.	Het schatten van de individuele parameters	93
3.	De variatie tussen individuen	94
4.	Slotopmerkingen	96
5.	Literatuur	97

Samenvatting

In dit artikel wordt ingegaan op de analyse van groeicurven, waarbij de analyse wordt uitgevoerd in twee stappen. Bij de eerste stap wordt een niet-lineaire functie aangepast aan de data van de individuele planten of dieren. Daarna wordt beslist of de parameterwaarden kunnen worden gebruikt voor het vergelijken van behandelingen. Een en ander wordt toegepast met gebruikmaking van de Richards functie.

1. Inleiding

Een deel van de experimenten bij het landbouwkundig onderzoek heeft betrekking op de groei van planten of dieren. Een groot aantal van deze experimenten wordt opgezet om een aantal behandelingen te vergelijken, bijvoorbeeld voedersystemen of teeltmethoden. In dit artikel zal de aandacht worden gericht op experimenten waarbij op gezette tijdstippen waarnemingen worden verricht bij alle individuen in het experiment. De reeks van waarnemingen aan een zeker individu kan worden weergegeven door middel van een vector $Y = (y_1, y_2, \dots, y_m)'$, waarbij m het aantal metingen is.

Er zal worden verondersteld dat de individuen die dezelfde behandeling hebben ondergaan kunnen worden beschouwd als een aselechte steekproef uit een relevante populatie. In dat geval kan de vector Y worden beschouwd als een realisatie van een vector van stochastische variabelen

$\underline{Y} = (y_1, y_2, \dots, y_m)'$ met verwachting $\mu = (\mu_1, \mu_2, \dots, \mu_m)'$ en covariantiematrix

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \dots & \sigma_{mm} \end{bmatrix}$$

De vector van waarnemingen aan individu i in de steekproef wordt genoteerd als $Y_i = (y_{i1}, y_{i2}, \dots, y_{im})'$. Op grond van een steekproef van n

individuen kan μ worden geschat door

$$\hat{\mu} = \bar{Y} = n^{-1} \sum_{i=1}^n Y_i \text{ en } \hat{\Sigma} \text{ door } \hat{\Sigma} = [\hat{\sigma}_{st}], \text{ waarbij}$$

$$\hat{\sigma}_{st} = (n-1)^{-1} \sum_{i=1}^n (y_{is} - \bar{y}_s)(y_{it} - \bar{y}_t) \quad (s, t = 1, 2, \dots, m).$$

De diversiteit in het groeiverloop in de populatie wordt op deze wijze weergegeven door middel van de matrix $\hat{\Sigma}$. In het algemeen zal het groeiverloop van de afzonderlijke individuen een bepaalde vorm vertonen, bijvoorbeeld een sigmoïde vorm. In dat geval kan men proberen om via het aanpassen van een functie aan de individuele waarnemingen tot een compactere beschrijving van de populatie te komen. De functie zal worden aangegeven met $f_{\Delta}(t)$, waarbij $\Delta = (\delta_1, \delta_2, \dots, \delta_p)'$ een vector van parameters is.

Er zal worden verondersteld dat de waarnemingsvector \underline{Y} , behorende bij een aselekt getrokken individu kan worden geschreven als som van twee componenten, te weten een vector van functiewaarden $f(\underline{\Delta})$ en een vector van afwijkingen \underline{e} , beide behorende bij dit individu. Er geldt dan:

$$\underline{Y} = f(\underline{\Delta}) + \underline{e} \quad [1]$$

met $f(\underline{\Delta}) = (f_{\Delta}(t_1), f_{\Delta}(t_2), \dots, f_{\Delta}(t_m))'$. De vector $\underline{\Delta}$ heeft verwachting Δ_0 en covariantiematrix $\Lambda = [\lambda_{kl}]$, ($k, l = 1, 2, \dots, p$) in de relevante populatie van individuen. De matrix Λ representeert de variatie tussen individuen (in het Engels: between-individual variation). De vector $\underline{e} = (e_1, e_2, \dots, e_m)$ bevat de afwijkingen van het model. De vector \underline{e} heeft verwachting 0 en covariantiematrix Ψ "binnen" elk individu: de matrix Ψ representeert de variatie van de waarnemingen van elk individu om diens eigen verwachte curve (in het Engels: within-individual variation). In het navolgende zal worden verondersteld dat $\Psi = \sigma^2 I_m$, waarbij σ^2 niet varieert over de individuen in de populatie. Verder zal worden verondersteld dat $\underline{\Delta}$ en \underline{e} onderling onafhankelijk zijn.

De analyse van [1] in het geval dat $f_{\Delta}(t)$ een lineaire functie is, wordt beschreven door Rao (1965). In het navolgende zal worden ingegaan op de situatie waarin $f_{\Delta}(t)$ niet lineair is in minstens een van de parameters $\delta_1, \delta_2, \dots, \delta_p$. In dat geval bestaat er geen eenvoudige relatie tussen μ en \int enerzijds en $f(\Delta_0)$, Λ en $\sigma^2 I$ anderzijds. Er wordt aandacht besteed aan het schatten van de individuele parameters, met name aan het schatten van de parameters van de Richardsfunctie (hoofdstuk 2). In hoofdstuk 3 zal aandacht worden besteed aan het schatten van Δ_0 , Λ en σ^2 met een praktische illustratie. In beide hoofdstukken zal worden ingegaan op moeilijkheden die kunnen optreden eventuele oplossingen zullen worden aangedragen.

2. Het schatten van de individuele parameters

Het schatten van de individuele parameters kan worden uitgevoerd met de kleinste kwadratenmethode. De kleinste kwadratenschatter van Δ is die waarde $\hat{\Delta}$ waarvoor

$$\{Y - f(\Delta)\}' \{Y - f(\Delta)\} \quad [2]$$

minimaal is. Een probleem dat hierbij optreedt is dat er geen expliciete formule voor $\hat{\Delta}$ bestaat, maar dat $\hat{\Delta}$ moet worden berekend via iteratieve weg. De klassieke methode voor het minimaliseren van [2] is de methode van Gauss-Newton. Deze methode is gebaseerd op de lineaire benadering van $f(\Delta)$, die wordt gegeven door

$$f(\Delta) \doteq f(\Delta^{(0)}) + F(\Delta^{(0)}) \{\Delta - \Delta^{(0)}\} \quad [3]$$

waarbij $F(\Delta)$ de $m \times p$ matrix van partiële afgeleiden van $f(\Delta)$ is en $\Delta^{(0)}$ een startwaarde voor Δ . De algemene vergelijking voor het iteratieproces is

$$\Delta^{(i+1)} = \Delta^{(i)} + [F(\Delta^{(i)})', F(\Delta^{(i)})]^{-1} F(\Delta^{(i)})' \{Y - f(\Delta^{(i)})\}$$

en dit proces wordt gestopt bij stap i als de coördinaten van $\Delta^{(i+1)} - \Delta^{(i)}$ in absolute waarde voldoende klein zijn.

Indien lineaire benadering [3] van toepassing is in het gebied rond de werkelijke waarde van Δ , dan geldt dat

$$\hat{\Delta} \doteq \Delta + [F(\Delta)' F(\Delta)]^{-1} F(\Delta)' \{Y - f(\Delta)\} \quad [4]$$

waaruit volgt dat $\hat{\Delta}$ (bij benadering) een zuivere schatter is voor Δ met covariantiematrix $\sigma^2 [F(\Delta)' F(\Delta)]^{-1}$. Een schatter voor de covariantiematrix is $\hat{\sigma}^2 [F(\hat{\Delta})' F(\hat{\Delta})]^{-1}$, waarbij $\hat{\sigma}^2 = (m-p)^{-1} \{Y - f(\hat{\Delta})\}' \{Y - f(\hat{\Delta})\}$. Benadering [4] is in sommige gevallen niet geldig. Dit kan als gevolg hebben dat $\hat{\Delta}$ een niet-zuivere schatter van Δ is (zie Box, (1971)).

Een veel gebruikte groeifunctie is de Richards functie (Richards, (1959)). In het navolgende zal gebruik worden gemaakt van de parametrisering die wordt gegeven door Nelder (1961):

$$f_{\Delta}(t) = \frac{\delta_1}{\{1 + e^{(-\delta_2 t + \delta_3)/\delta_4}\} \delta_4} \quad [5]$$

Het schatten van de parameters kan gebeuren op de hiervoor beschreven wijze. Hierbij treden soms moeilijkheden op, die opgelost kunnen worden door een modificatie van het iteratieve proces (zie Ross, (1975)).

De schatters van de parameters δ_2 en δ_4 vertonen een structureel hoge correlatie, die niet door een eenvoudige herparametrisering verwijderd kan worden (zie Jansen, (1981)). De waarde van de correlatiecoëfficiënt wordt onder meer bepaald door de waarde van δ_4 , de parameter die de vorm van de curve bepaalt. Er is gebleken dat, indien $0 < \delta_4 < 1$, [5] redelijk bruikbaar is. In dat geval is [5] bij benadering exponentieel in de linkerstaart.

3. De variatie tussen individuen

In het navolgende zal aan de hand van een voorbeeld worden getracht inzicht te verkrijgen in de variaties die een rol spelen bij de groei van de scheuten van jonge appelbomen. Aan de data van negen individuele

bomen is de Richards functie aangepast. Als schatting voor Δ_0 wordt gevonden het gemiddelde van de individuele schattingen gekozen, met als resultaat

$$\hat{\Delta}_0 = (123.24, .02010, .9105, .3348)'$$

Tevens is het van belang om naar de covariantiematrix van de schattingen te kijken. Deze matrix is van belang in het geval men behandelingen wil vergelijken. In dit geval wordt deze matrix gegeven door

$$L = \begin{bmatrix} 63.791 & & & & \\ & .0030456 & .000003130 & & \\ & -.29512 & -.000015255 & .0025307 & \\ & .29711 & .000078940 & -.0018111 & .0041725 \end{bmatrix}$$

De covariantiematrices van de schatters kunnen worden gemiddeld met als resultaat

$$C = \begin{bmatrix} 1.9291 & & & & \\ & .0024456 & .0000075061 & & \\ & -.014868 & -.000042560 & .0010121 & \\ & .10961 & .00027708 & -.0018909 & .019795 \end{bmatrix}$$

In het geval er geen variatie is tussen individuele curves ($\Lambda \equiv 0$), moeten L en C bij benadering gelijk zijn. In het geval dat er variatie is tussen individuele curves moeten de diagonale elementen van L groter zijn dan de corresponderende waarden van C. Noteer de diagonale elementen van L en C door middel van l_{ii} en c_{ii} ($i = 1, 2, \dots, 4$). In dit voorbeeld is $l_{22}/c_{22} = .42$ en $l_{44}/c_{44} = .39$. Van belang is het op te merken dat $c_{24}/c_{22} c_{44} = .97$, hetgeen erop duidt dat het opnemen van zowel δ_2 en δ_4 als parameters die bijdragen aan de beschrijving van de variatie tussen individuen niet mogelijk is. Day (1966) suggereert om de waarde van δ_4 voor alle individuen gelijk te kiezen. In dit geval is het mogelijk om δ_4 de waarde .3348 toe te kennen, omdat deze parameter vrij

stabiele geschat kan worden. Dan worden de uitkomsten:

$$\hat{\Delta}_0 = (123.27, .02011, .9156)$$

$$L = \begin{bmatrix} 57.319 & & \\ -.0039094 & .0000018686 & \\ -.27905 & .000023103 & .0022713 \end{bmatrix}$$

$$C = \begin{bmatrix} .72030 & & \\ -.00031661 & .00000033358 & \\ .00087595 & .0000057777 & .00027734 \end{bmatrix}$$

Door het fixeren van de waarde van δ_4 worden de varianties van de schatters van de overige parameters sterk gereduceerd. Een schatting voor de matrix Λ is in dit geval gelijk aan (met correlatiecoëfficiënten boven de diagonaal)

$$\hat{\Lambda} = L - C = \begin{bmatrix} 56.599 & | & -.3855 & | & -.8333 \\ -.0035928 & | & .000001535 & | & .3132 \\ -.27993 & | & .000017325 & | & .0019940 \end{bmatrix}$$

De hoge (geschatte) correlatie (in absolute waarde) tussen

δ_1 en δ_3 duidt erop dat grote eindwaarden samengaan met grote beginwaarden. Deze hoge correlatie maakt het mogelijk om bij een vergelijking van behandelingen alleen gebruik te maken van δ_1 en δ_2 .

4. Slotopmerkingen

Een belangrijk probleem is hoe de hier behandelde zaken kunnen worden toegepast in de praktijk. Hiervoor is het belangrijk dat individuele parameters goed geschat kunnen worden. De Richards-functie biedt deze mogelijkheid niet, althans niet zonder een restrictie op de parameterwaarden.

Het blijkt dat in vele gevallen de vormparameter δ_4 niet gebruikt kan worden voor het vergelijken van behandelingen.

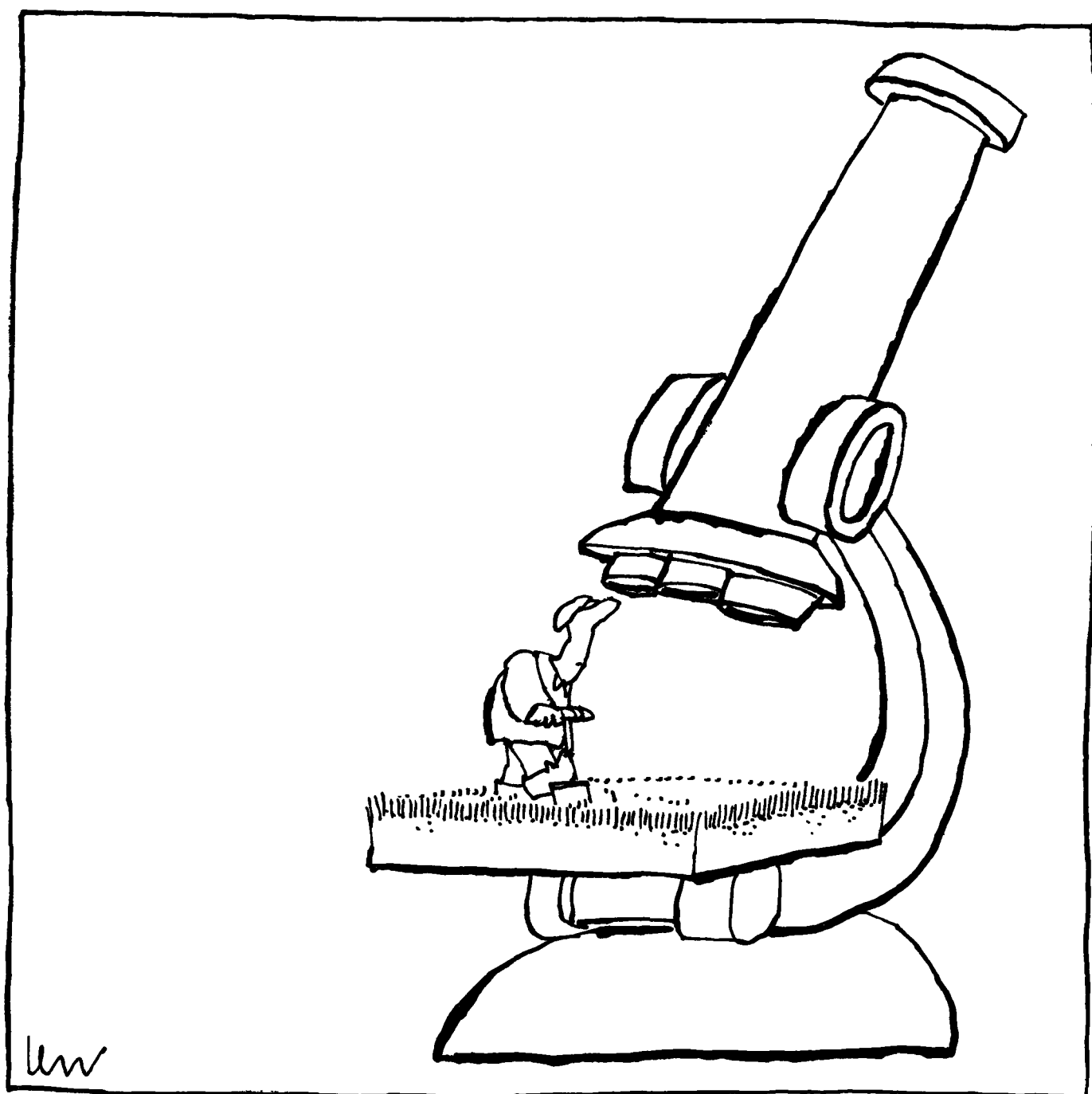
Verder is het van belang te weten of de variatie tussen individuen bij alle behandelingen gelijk is. Is dit niet het geval dan zijn vergelijkingen van de bijbehorende behandelingen niet exact uitvoerbaar. In het geval de covariantiestructuren gelijk zijn, kan $2n^{-1}(L_1+L_2)$ worden gebruikt om na te gaan of Δ_{01} significant verschilt van Δ_{02} (via de T^2 -toets van Hotelling, het multivariate equivalent van de t-toets). Hierbij zijn Δ_{01} en Δ_{02} de parametervectoren behorende bij twee behandelingen en L_1 en L_2 de bijbehorende covariantiematrices van de schattingen. Bij meer ingewikkelde proeven kan men denken aan de toepassing van multivariate variantieanalyse. Veel onderzoek op dit gebied is echter nog vereist.

5. Literatuur

- Box, M.J. (1971), Bias in nonlinear estimation (with discussion) *J.R.S.S., B* 31, 171-201.
- Day, N.E. (1966), Fitting curves to longitudinal data. *Biometrics*, 22, 276-291.
- Jansen, J. (1981), Some aspects of growth curve analysis. Thesis to be submitted for the degree of Master of Science. University of Kent, Canterbury.
- Nelder, J.A. (1961), The fitting of a generalization of the logistic curve. *Biometrics*, 17, 89-110.
- Rao, C.R. (1965), The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika*, 52, 447-458.
- Richards, F.J. (1959), A flexible growth function for empirical use. *J. Exper. Bot.*, 10, 290-300.
- Ross, G.J.S. (1975), Simple non-linear modelling for the general user. *Proc. 40th Session ISI, Warsaw*, 2, 585-593.

Een toets op ruimtelijke autocorrelatie in rechthoekige roosters

Dr. Michiel J. W. Jansen



Inhoud

1.	Inleiding	103
2.	De ruimtelijke correlatiecoëfficiënt bij normaal verdeelde toevalstermen	107
3.	Een proef in een kas	109
4.	Het verspreidingspatroon van een bomenziekte	111
5.	Conclusies	112
	Literatuur	113

1. Inleiding

Dit rapport handelt over de analyse van waarnemingen op een rechthoekig rooster van, zeg, $r \times k$ roosterpunten of cellen. We nemen voor de eenvoud aan dat in iedere cel één waarneming wordt gedaan.

In het meest gebruikelijke model voor de analyse van dit type proef wordt de waarneming in een roosterpunt beschouwd als de som van twee termen: een systematische of deterministische term die afhangt van de bekende omstandigheden in dat roosterpunt (de behandeling in ruime zin), en een term die aan het toeval wordt toegeschreven. Meestal wordt verondersteld dat de toevalstermen in de verschillende cellen geen enkele samenhang vertonen en dat ze beschouwd kunnen worden als onafhankelijke trekkingen uit dezelfde verdeling.

In dit rapport wordt besproken hoe men kan toetsen of de toevalstermen in naburige roosterpunten inderdaad onafhankelijk zijn, oftewel hoe men kan toetsen op ruimtelijke autocorrelatie. Er wordt niet systematisch ingegaan op de vraag wat er geconcludeerd of gedaan moet worden wanneer is gebleken dat de toevalstermen van buurcellen afhankelijk zijn.

Een opmerking van Scheffé (1959, p. 364) volstaat om aan te geven dat autocorrelatie, dus afhankelijkheid, de analyse sterk kan beïnvloeden: "Van de besproken afwijkingen van de gebruikelijke veronderstellingen levert afhankelijkheid in het algemeen de meest formidabele problemen."

In §2 bespreken we kort de theorie van de toets op ruimtelijke autocorrelatie voor het geval van een willekeurig aantal behandelingen, of verklarende variabelen, bij normaal verdeelde toevalstermen. (Er bestaat ook een toets voor willekeurig verdeelde toevalstermen bij één behandeling; hiervoor wordt de geïnteresseerde lezer verwezen naar Cliff & Ord, 1973, sectie 2.3.2.) De inhoud van §2 is misschien wat technisch van aard; we hopen echter dat de verdere paragrafen goed leesbaar zijn voor niet-specialisten op statistisch gebied.

Dat de toevalstermen niet altijd onafhankelijk zijn blijkt uit de volgende twee voorbeelden, die in §3 en §4 worden uitgewerkt.

Bij landbouwkundige proeven zijn belangrijke factoren zoals klimaat en bodemvruchtbaarheid meestal zeer onvolledig geregistreerd. De niet geregistreerde onderdelen van de behandeling kunnen niet op simpele wijze in de deterministische modelterm worden opgenomen en hun effect wordt soms beschouwd als *toeval*. Het is echter redelijk te veronderstellen dat klimaat en bodemvruchtbaarheid in het proefveld geleidelijk kunnen veranderen zodat er bijvoorbeeld "goede" gebieden kunnen ontstaan met hoge toevalstermen en "slechte" met lage. Zulke plaatsgebonden effecten worden vaak standplaatseffecten genoemd. In kassen worden door glazen wanden, verwarmingsbuizen, dakgoten etc. aanzienlijke standplaatsverschillen veroorzaakt. Dit blijkt uit Figuur 1.

Soms wordt voor onbekende klimaat- of bodemeffecten gecorrigeerd door op de één of andere manier de standplaats op te nemen in de *deterministische* modelterm (zie bijvoorbeeld Scheffé, 1959, p. 170). De in §2 te bespreken toets zou een criterium kunnen leveren om na te gaan in hoeverre zo'n correctie geslaagd is. In §3 wordt de toets gebruikt om twee correctiemethoden voor standplaats met elkaar te vergelijken.

Het tweede voorbeeld van ruimtelijke afhankelijkheid wordt geleverd door het in §4 onderzochte verspreidingspatroon van een bomenziekte. Als de ziekte besmettelijk is, mag men verwachten dat de toevalstermen in naburige cellen met elkaar samenhangen. De toets op ruimtelijke autocorrelatie levert een objectief criterium om na te gaan of er zo'n samenhang bestaat. In Figuur 2 is het onderzochte verspreidingspatroon van de ziekte weergegeven.

Het is niet moeilijk de lijst van voorbeelden verder uit te breiden: denk bijvoorbeeld aan onderlinge concurrentie van planten of aan de verspreiding van ongedierte. We hopen echter dat uit de twee gegeven voorbeelden blijkt dat de toets op ruimtelijke autocorrelatie een bruikbaar instrument is bij de analyse van waarnemingen in het vlak.

Tot slot van deze inleiding wil ik graag mijn dank betuigen aan het Proefstation voor de Tuinbouw onder Glas te Naaldwijk en aan het Rijksinstituut voor het onderzoek in Bos- en Landschapsbouw "De Dorschkamp" te Wageningen voor de toestemming om gebruik te maken van hun waarnemingsmateriaal.

3	10	5	9	1	11	6	7
2	12	4	8	3	10	5	9
1	11	6	7	2	12	4	8
2	12	4	9	1	11	6	7
3	10	5	7	2	12	4	8
1	11	6	8	3	10	5	9
(a) behandelingen							
60	115	109	59	126	76	72	45
87	9	109	59	77	123	125	29
116	61	65	100	116	52	153	42
77	8	98	22	116	67	86	59
54	105	108	88	79	51	122	52
100	53	64	34	69	83	79	34
(b) waarnemingen							
-5	8	4	23	11	12	0	-28
-3	-21	-11	12	12	16	20	-7
2	-3	-7	27	26	22	32	-5
-13	-22	-22	-14	1	3	14	-14
-11	-1	3	15	-11	21	1	5
-14	-11	-8	-13	4	-23	-26	-2
(c) residuen							

Figuur 1. Een proef in een kas, bestaande uit twaalf verschillende behandelingen die ieder vier maal herhaald zijn. In (a) is de plattegrond getekend van de behandelingen, in (b) die van de waarnemingen en in (c) die van de afgeronde residuen. De residuen zijn verkregen door iedere waarneming te verminderen met het gemiddelde van de bijbehorende behandeling. Er is met opzet geen rekening gehouden met plaatseffecten, opdat deze effecten goed zichtbaar zouden worden. Men ziet dat residuen met hetzelfde teken opmerkelijk vaak naast elkaar liggen. De ruimtelijke aspecten van deze proef worden besproken in §3.

4	4	3	6	6	6	5	2	1	6	1	3	3	3	4	5	5	4	4	3
4	6	4	5	5	5	5	2	3	4	4	5	1	1	1	3	3	5	4	2
6	4	6	4	5	6	4	4	6	6	4	2	3	2	1	2	2	5	5	2
4	3	4	6	6	5	5	6	3	5	4	3	3	2	2	2	1	1	3	5
4	3	6	7	4	5	5	3	7	5	5	2	6	3	5	6	1	1	2	4
7	3	3	6	3	6	2	4	5	4	5	3	7	7	4	5	7	1	1	3
6	7	4	3	1	2	3	6	4	4	5	3	3	3	5	5	3	3	3	3
5	4	4	4	3	4	7	4	5	4	7	7	7	5	7	3	2	2	6	6
3	7	5	7	6	6	6	7	5	3	3	7	2	3	5	3	2	5	3	4
4	7	6	4	4	4	4	4	4	4	3	3	2	3	7	7	7	4	7	2
4	5	2	2	4	3	2	5	3	2	1	3	1	3	6	2	5	2	6	5
7	3	2	3	4	4	6	6	5	6	2	6	2	6	3	3	6	7	5	5
3	5	2	6	6	4	6	2	7	7	6	6	3	4	6	6	2	5	7	5
6	3	5	7	6	4	3	3	4	7	4	6	4	5	7	7	6	7	6	3
2	6	2	3	5	4	3	4	5	7	6	7	7	7	4	4	5	6	3	6
1	3	3	3	5	4	6	3	7	6	7	7	7	7	7	7	7	7	2	7
2	7	2	6	2	2	1	4	7	4	4	6	6	2	3	4	3	3	3	4
3	2	2	3	1	1	2	3	6	2	4	6	4	7	5	6	3	3	2	3
6	3	5	4	2	2	1	3	4	4	4	3	5	4	5	3	2	2	2	4
6	5	3	6	2	1	1	2	6	6	6	5	3	3	7	4	4	4	2	3

Figuur 2. Verspreidingspatroon van een ziekte (naalduitval) in een beplanting van grove dennen. Ieder cijfer stelt een boom voor. Deze cijfers, lopend van 1 tot 7, zijn een maat voor de aantasting van de bomen. Steeds zijn bomen van dezelfde afkomst (familie) in een 4x4-opstelling bij elkaar geplant, zoals aangegeven door de lijnen. Iedere familie komt één keer voor; er zijn dus vijftientwintig families, ieder van zestien bomen. De door ons gestelde vraag is of het waargenomen verspreidingspatroon voldoende verklaard kan worden door aanleg (familie). Of moet men concluderen dat er ook nog sprake is van andere factoren zoals onderlinge besmetting? Deze proef wordt besproken in §4.

2. De ruimtelijke correlatiecoëfficiënt bij normaal verdeelde toevalstermen

Het gebruikelijke lineaire model heeft de vorm

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

hierin zijn \underline{y} en $\underline{\varepsilon}$ toevalsvectoren van lengte n ;
 $n = rk$ met r het aantal rijen en k het aantal kolommen van het rooster;
 de vectoren worden gemaakt door het rooster in leesvolgorde te doorlopen;
 β is een vector van lengte m ; X is een $n \times m$ matrix; de m kolommen van X
 noemen we de regressors; de toevalsvector $\underline{\varepsilon}$ is normaal verdeeld.

De nulhypothese luidt

$$H_0 : \underline{\varepsilon} \stackrel{d}{=} N(0, \lambda^2 I),$$

waarin I de identieke afbeelding op \mathbb{R}^n is en waarin " $\stackrel{d}{=}$ " betekent: "is verdeeld als". Volgens H_0 zijn de componenten van $\underline{\varepsilon}$ dus onafhankelijk en identiek normaal verdeeld.

Het alternatief, H_1 , is dat de verdeling van bijvoorbeeld de i -de component van $\underline{\varepsilon}$, ε_i , afhangt van de waarden van $\underline{\varepsilon}$ in de buurcellen van cel i . Om dit alternatief te kunnen formuleren, voeren we een $n \times n$ -matrix W in door zijn werking te geven op een willekeurige n -vector x :

$$(Wx)_i \stackrel{\text{def}}{=} \frac{1}{4} (\text{som van de } x\text{-waarden in de buurcellen van cel } i),$$

waarin de burens gedefinieerd zijn als de noorder-, ooster-, zuider- en westerburens (cellen in het binnengebied hebben dus vier burens, cellen op de rand hebben drie of - in de hoeken - twee burens). Uiteraard hadden we W ook anders kunnen definiëren; we gaan daar niet verder op in.

Het alternatief kan nu geformuleerd worden als:

$$H_1 : \underline{\varepsilon} = \zeta W \underline{\varepsilon} + \underline{u},$$

met

$$\underline{u} \stackrel{d}{=} N(0, \lambda^2 I),$$

waarin $\zeta \neq 0$ een getal is dat om hier niet te noemen redenen tussen -1 en 1 moet liggen. In dit alternatieve model is de verwachting van $\underline{\varepsilon}_i$ gegeven de ε -waarden van de burens van i , hoger of lager naarmate het ε -buurgemiddelde van cel i hoger of lager is.

Omdat de toevalsvector ε niet meetbaar is, moet een toets van H_0 tegen H_1 gebaseerd zijn op de gevonden residuvector, e , van de waarnemingen bij regressie naar X . De toetsingsgrootte moet een hoge waarde hebben als naburige residuen vaak tegelijk hoog of laag zijn. Hieraan voldoet het inwendig product (e, We) . Verder mag de grootte niet veranderen wanneer alle residuen met dezelfde factor worden vermenigvuldigd. Een intuïtief aantrekkelijke toetsingsgrootte is dus

$$c = (e, We)/(e, e).$$

Cliff en Ord (1973) laten zien dat c inderdaad zeer geschikt is om H_0 van H_1 te onderscheiden (in het bijzonder bij kleine waarden van ζ).

De eerste twee momenten van \underline{c} onder H_0 , zeg μ en σ^2 , die uitsluitend van X en W afhangen, kunnen met enige moeite berekend worden. Hoe, dat doet er hier niet toe (zie Durbin en Watson, 1950). We merken slechts op dat de residuen van de regressie naar X uiteraard gecorreleerd zijn, zodat de momenten in ieder geval van X moeten afhangen.

Men kan bewijzen dat de grootte \underline{c} onder H_0 asymptotisch normaal verdeeld is voor grote aantallen waarnemingen:

$$\underline{c}^* \rightarrow N(0, 1) \quad (n-m \rightarrow \infty) \quad (\text{in distributie}),$$

waarin

$$\underline{c}^* \text{ def } (\underline{c} - \mu) / \sigma$$

de gestandaardiseerde waarde van \underline{c} is (zie Durbin en Watson, 1950).

3. Een proef in een kas

We zullen de ruimtelijke aspecten (en geen andere) bespreken van de in Figuur 1 geïntroduceerde proef.

Bij de proefopzet was besloten tot een blokstructuur zoals aangegeven in Figuur 3.a: ieder van de zestien blokken wordt geacht een onbekend, deterministisch, effect te hebben op de waarnemingen in dat blok. Op deze wijze wordt in belangrijke mate gecorrigeerd voor standplaats-effecten. De residuen die overblijven na aftrek van behandelings- en blokeffecten zijn weergegeven Figuur 3.a.

Een alternatieve methode om voor standplaatseffecten te corrigeren is de volgende: neem als benadering voor het werkelijke standplaatseffect bij de verklarende variabelen (de deterministische term) een willekeurig polynoom op van het rijnummer, zeg i , en het kolomnummer, zeg j . De residuen van de regressie naar behandeling en naar een derdegraads polynoom in i en j zijn weergegeven in Figuur 3.b.

Deterministische modeltermen	c*	vrijheids- graden van residu	gemiddeld kwadraat van residu
behandeling	3.0	36	301
behandeling en blokken	1.6	24	208
behandeling en derdegraads polynoom	0.8	27	188

Tabel 1.

-3	14	9	2	-5	-5	-17	-15
-1	-16	-7	-8	-5	0	2	6
4	2	-2	6	10	5	15	8
0	-10	-13	-10	3	3	18	-10
2	10	12	19	-9	21	5	9
-2	0	1	-9	6	-24	-23	1
(a)							
-4	9	-3	5	0	0	0	-5
4	-14	-10	-2	-5	-3	8	1
14	7	-6	12	11	5	21	-5
-3	-7	-12	-21	-4	-11	0	-19
-4	12	13	13	-12	16	0	10
-10	4	7	-3	13	-18	-19	15
(b)							

Figuur 3. Afgeronde residuen van de proef uit Figuur 1 bij twee verschillende methodes om voor plaatseffecten te corrigeren. In (a) is "geblokt": de zestien blokken zijn door de lijnen aangegeven. In (b) is gecorrigeerd met een derdegraads plaatspolynoom. Het is niet op het oog duidelijk welke methode de minste autocorrelatie achterlaat.

Welke van de twee correctiemethoden is in dit geval de beste?

De blokmethode is aantrekkelijk vanwege haar eenvoud en elegantie.

Doch de polynoommethode is geloofwaardiger omdat de in rekening gebrachte plaatseffecten wat geleidelijker verlopen. Aangezien een visuele vergelijking van de residuen in Figuur 3a en 3b weinig oplevert, zullen we proberen met de c^* -toets tot een beslissing te komen. De resultaten van de toets zijn weergegeven in Tabel 1. Inspectie van de tabel leert dat in dit geval de polynoomcorrectie doeltreffender is dan de blokcorrectie: zij laat de minste autocorrelatie achter. Deze conclusie wordt bevestigd door het feit dat het gemiddelde restkwadraat kleiner is (en dat bij een groter aantal vrijheidsgraden van de rest).

4. Het verspreidingspatroon van een bomenziekte

We bespreken nu de in Figuur 2 geïntroduceerde proef. De proefopzet en de vraagstelling zijn in de tekst bij die figuur kort uiteengezet.

De vraag die we ons voorlopig stellen is of het gewone lineaire model, H_0 , met uitsluitend familie als verklarende factor, in vergelijking met het autocorrelatiemodel H_1 , voldoende goed past bij de waarnemingen. De c^* -toets levert een gestandaardiseerde autocorrelatiecoëfficiënt $c^* = 3.1$; dat is hoogst significant: de eenzijdige overschrijdingskans onder H_0 is 0.001.

Er is dus alle aanleiding om wat te filosoferen over de interpretatie van het feit dat uit de toets blijkt dat H_0 niet houdbaar is bij deze waarnemingen. We zullen een aantal denkbare verklaringen opsommen voor het resultaat van de toets.

- (A) De toevalscomponenten zijn niet normaal verdeeld. Dat is natuurlijk juist, maar er bestaat in dit geval weinig aanleiding om te twijfelen aan de robuustheid van de gebruikte techniek. Bovendien blijkt uit een grafiek op normaalpapier dat de residuen, waar de toets uiteindelijk op werkt, in verbazend goede benadering normaal verdeeld zijn (al zijn de staarten van de verdeling wat dun). Al met al lijkt niet-normaliteit geen aantrekkelijke uitleg van de toetsuitslag.
- (B) Een andere verklaring zou kunnen zijn dat plaatseffecten niet voldoende verwerkt zijn in de deterministische modelterm. Dit is geen erg aannemelijke verklaring: bomen van dezelfde familie staan steeds keurig bij elkaar, en familie-effecten zijn wel in rekening gebracht. Verder is de afstand tussen burens klein (1,5 m), zodat deze verklaring alleen geloofwaardig is als de bodem- en klimaatcondities plaatselijk uiterst variabel zijn.

(C) Misschien is het alternatief, H_1 , wel juist.

Hiermee zou de autocorrelatie van de residuen verklaard zijn, doch H_1 is biologisch moeilijk interpreteerbaar.

(D) We zullen nu een biologisch plausibele verklaring geven die met H_1 verwant is. Dit tweede alternatief van de nulhypothese zullen we aanduiden als H_2 . In H_2 nemen we aan dat de verspreiding van de naalduitval beschreven kan worden door een zogenaamd gemengd regressief-autoregressief model;

$$H_2 : \underline{y} = \zeta W \underline{y} + X \beta + \underline{u}, \quad \underline{u} \underline{d} N(0, \sigma^2 I).$$

In dit model is de verwachte besmettingsgraad van bijvoorbeeld boom i , gegeven de besmettingsgraad van alle andere bomen, hoger naarmate de buurbomen van boom i zwaarder besmet zijn (en omgekeerd). Dit is een plausibel model wanneer de bomen meer ziekteverwekkers produceren naarmate zij zwaarder besmet zijn. De toevalsterm in dit model gedraagt zich net zoals de toevalsterm in het model H_1 ; het verschil tussen H_1 en H_2 ligt in de deterministische term. Ook onder H_2 mag men een residu met autocorrelatie verwachten: daarmee zou de toetsuitslag verklaard zijn op een wijze die goed interpreteerbaar is.

Op dit moment wordt er programmatuur ontwikkeld om de parameters van de diverse modellen te schatten (maximum likelihood). Zulke programmatuur levert tevens de middelen om H_1 van H_2 te onderscheiden. Nu het rekenwerk nog niet kan worden uitgevoerd, is onze voorkeur voor H_2 uitsluitend gebaseerd op het feit dat H_2 goed interpreteerbaar is.

5. Conclusies

De besproken toets op ruimtelijke autocorrelatie is een soort "breedspectrum"-toets. Significantie hoeft lang niet altijd op zuivere autocorrelatie te duiden. Een positief toetsresultaat wijst er uitsluitend op dat er méér samenhang bestaat tussen naburige elementen dan in de nulhypothese tot uitdrukking komt. Doordat de toets zo weinig specifiek is kan men hem in zeer uiteenlopende situaties toepassen.

Literatuur

Cliff, A.D. & Ord, J.K., 1973,
Spatial autocorrelation,
London, Pion.

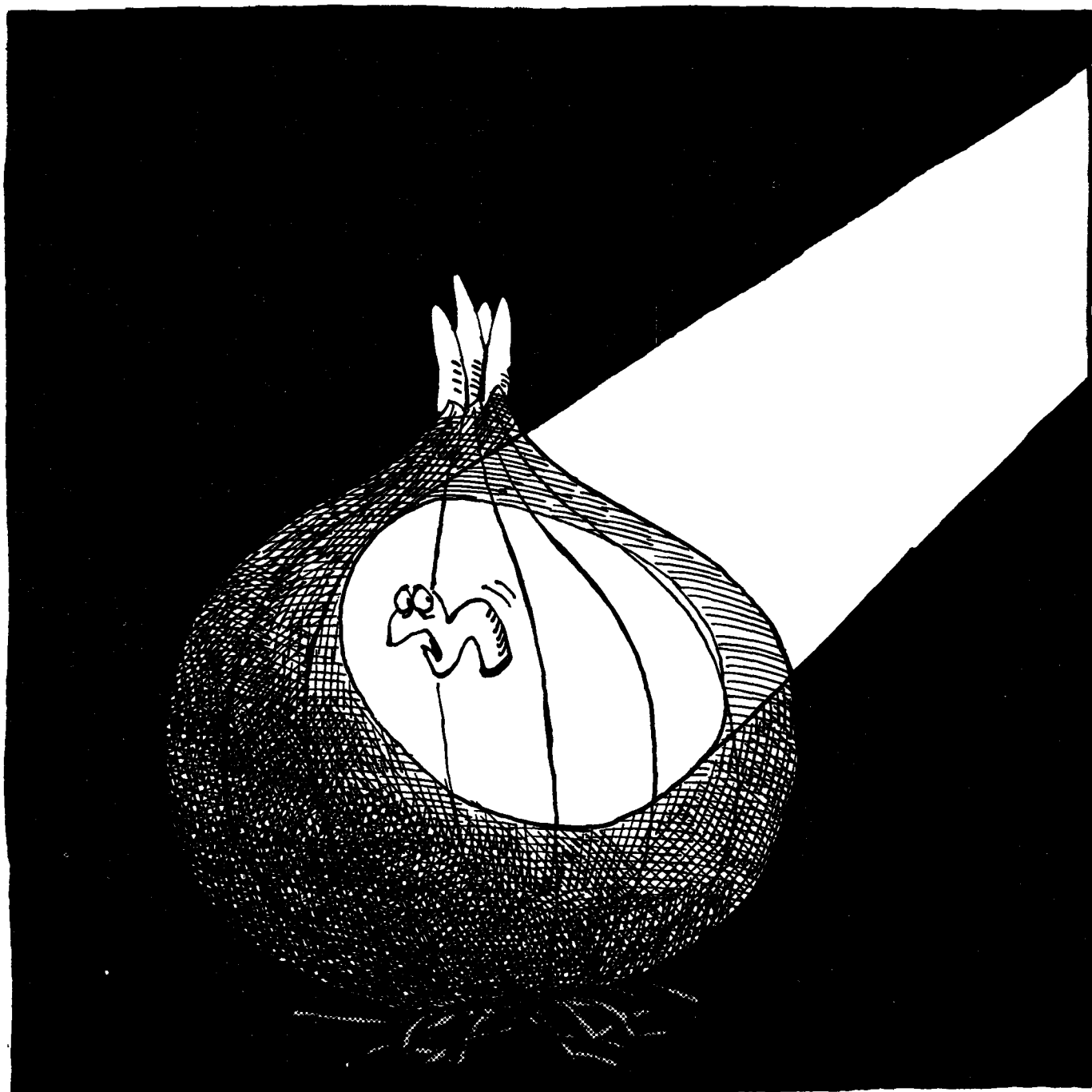
Durbin, J. & Watson, G.S., 1950,
Testing for serial correlation in least squares regression. I.
Biometrika, 1950, p 409-428.

Scheffé, H, 1959,
The analysis of variance,
New York, Wiley.

**De optimale opzet van een proef die wordt uitgevoerd
om voor gedroogde produkten de effectiviteit van
straling onder verschillende omstandigheden te
schatten en te vergelijken**

Ir. A. Keen

**Rapport ten behoeve van de Stichting ITAL, in het kader van de
statistische begeleiding van deze Stichting**



Inhoud

	Samenvatting	119
1.	Inleiding	119
2.	Uitvoering van de proeven	121
3.	Definities/notatie	121
4.	Uitgangspunten/veronderstellingen en hun relevantie; model voor de waarnemingen	123
5.	Het schatten en vergelijken van D_{10}-waarden	129
6.	De optimale proefopzet	131

Samenvatting

Voor een proef om de effectiviteit van straling onder verschillende omstandigheden te schatten en te vergelijken, dient de optimale opzet te worden gemaakt. Hierbij wordt gebruik gemaakt van gegevens uit een reeds uitgevoerde proef. Uiteengezet wordt hoe men tot een keuze komt en welke onzekerheden een rol spelen. Eerst wordt een (simpel regressie-)model voor de waarnemingen ontwikkeld met daarin de effectiviteit van de straling gerepresenteerd door de parameter D_{10} , zijnde de stralingsdosis die het aantal micro-organismen met een factor 10 vermindert. Eisen ten aanzien van de conclusies worden geformuleerd als eisen ten aanzien van de verwachte lengte van een betrouwbaarheidsinterval en/of ten aanzien van de kans om een van te voren aangegeven verschil significant te vinden. Deze lengtes en kansen zijn voor een groot aantal proefsituaties berekend. De onderzoeker kan hieruit op grond van praktische overwegingen (kosten) de "optimale" opzet kiezen.

1. Inleiding

Ook in gedroogde produkten blijken vaak (te) veel micro-organismen voor te komen, wat een gevaar is voor de volksgezondheid. Het aantal micro-organismen dient daarom sterk gereduceerd te worden. Via op De Stichting ITAL uit te voeren onderzoek wil men voor gedroogde peterselie, uien en knoflook bepalen hoe zo'n reductie het best tot stand kan worden gebracht. Het is bekend, dat straling hiertoe een bruikbaar middel is, waarvan de effectiviteit onder andere afhangt van het interne vochtgehalte van de produkten (uitgedrukt in de A_w -waarde). Vaststellen in welke mate deze afhankelijkheid aanwezig is, is de directe doelstelling van het uit te voeren onderzoek.

Het effect van de bestraling wordt uitgedrukt in de D_{10} -waarde, de dosis die nodig is om het aantal micro-organismen met een factor 10 te verminderen. De relatie tussen de logaritme van het aantal micro-organismen en de stralingsdosis wordt verondersteld constant te zijn. Het gaat er dus om de D_{10} -waarde vast te stellen bij alle mogelijke A_w -waarden. In een proef houdt dit in dat de D_{10} -waarden moeten worden geschat en vergeleken.

Uit een vooronderzoek zijn gegevens beschikbaar waarin bij een aantal 13 combinaties van stralingsdosis en A_w -waarde het aantal micro-organismen is bepaald. Naast het vaststellen van de orde van grootte van het aantal micro-organismen en het uitproberen van de onderzoeksmethodiek, is dit vooronderzoek vooral bedoeld om een inzicht te krijgen in de nauwkeurigheid van de procedure. De nauwkeurigheid wordt gekarakteriseerd door de lengte van betrouwbaarheidsintervallen (intervallen die met gespecificeerde kans de werkelijke waarde bevatten) en het onderscheidingsvermogen (de kans, dat bij een werkelijk aanwezig verschil in D_{10} -waarden, ook inderdaad wordt geconcludeerd dat een verschil bestaat) bij verschillende proefopzetten. Dit is de basis om de optimale proefopzet te kiezen.

De volgende hoofdstukken bevatten een uitvoerige uiteenzetting over hoe dit probleem wordt aangepakt. Dit is bedoeld als een illustratie van hoe met wat rekening wordt gehouden om tot een antwoord te komen op een veel voorkomende vraagstelling in een relatief simpele situatie. Alles is uiteindelijk uitsluitend gebaseerd op hoe de waarnemingen worden verkregen en hoe de micro-organismen reageren op de A_w -waarden onder de omstandigheden gedurende de proef. Het antwoord kan alleen worden verkregen voor een model van de werkelijkheid. In hoeverre dit antwoord ook geldt voor de werkelijkheid zelf, wordt voornamelijk bepaald door de mate waarin variaties goed worden beschreven door het wiskundig/statistische model en hangt onder andere af van:

- * of de monsters aselekt kunnen worden genomen,
- * of op de goede manier rekening kan worden gehouden met de factoren die invloed hebben op de waarneming.

De redenering is als volgt opgebouwd: eerst wordt een nadere beschrijving gegeven van de manier waarop de proeven worden uitgevoerd (hoofdstuk 2), waarbij in woorden wordt omschreven welke variaties mogelijk een rol spelen. In hoofdstuk 3 volgt dan een overzicht van definities en notaties van grootheden die verband houden met de waarnemingen en met het uitgangsmateriaal. Dit is uitsluitend een omschrijving zonder ook maar iets te veronderstellen. De veronderstellingen (de niet volledig bewijsbare uitgangspunten) die in eerste instantie worden gedaan en waarmee een fraaie recht toe recht aan oplossing van het

probleem mogelijk wordt, worden gegeven in hoofdstuk 4, resulterend in het te hanteren model voor de waarnemingen. Het gestelde probleem, uitgedrukt in termen van het model en de formulematige oplossing ervan, wordt gegeven in hoofdstuk 5. In hoofdstuk 6 wordt ingegaan op hoe de optimale proefopzet dan wordt gevonden. Hierin worden de voor peterselie uit de beschikbare gegevens verkregen schattingen verwerkt.

2. Uitvoering van de proeven

Van elk van de produkten peterselie, uien en knoflook is één grote gemengde partij van 5 kg beschikbaar. Er wordt gewerkt met monsters van 20 g, voor iedere behandeling een apart monster. Elk monster wordt op de (voor de proef) vereiste A_w -waarde gebracht. Hiermee is een hoeveelheid tijd gemoeid die afhangt van de desbetreffende A_w -waarde. De monsters worden hierna verpakt en bestraald met de (voor de proef) vereiste doses. Daarna wordt het aantal micro-organismen (verschillende groepen bacteriën) bepaald via het maken van een verdunningsreeks, uitplaten en tellen van het aantal gevormde kolonies.

De variaties die een rol (kunnen) spelen zijn:

- * variaties in werkelijke aantallen micro-organismen tussen de monsters: de heterogeniteit van de partij,
- * variaties in effect van de behandeling, gevolg van het niet helemaal constant zijn van de omstandigheden gedurende de proef,
- * bepalingfouten: pipetteerfouten, telfouten, enzovoort.

3. Definities/notatie

Alle aantallen worden uitgedrukt per gewichtseenheid. Een streepje onder een grootheid geeft aan, dat deze grootheid stochastisch is, dat wil zeggen een waarde aanneemt die in zekere mate onvoorspelbaar is.

grootheid	betekenis
x	de stralingsdosis
N/x	het aantal kweekbare micro-organismen in de partij, na bestraling met dosis x
\underline{Y}_i/x	de logaritme van het aantal micro-organismen in monster i bij dosis x . \underline{Y}_i/x is, gegeven het monster en de omstandigheden tijdens de proef weliswaar een vast getal, maar omdat het monster een toevallige greep is uit de partij en ook de omstandigheden in zekere mate toevallig zijn, is \underline{Y}_i/x toch stochastisch
\underline{y}_i/x	de waarneming; de logaritme van het geschatte (bepaalde, getelde) aantal micro-organismen bij dosis x in monster i
$\frac{\varepsilon}{b} \underline{y}_i/x$	$\underline{y}_i/x - \underline{Y}_i/x$: de bepalingsfout
μ/x	$E \underline{Y}_i/x$: de verwachtingswaarde over de monsters van \underline{Y}_i/x . Dit is het gemiddelde van alle hypothetisch mogelijke \underline{Y}_i/x
$\frac{\varepsilon}{m} \underline{y}_i/x$	$\underline{Y}_i/x - \mu/x$: de bemonsteringsfout en de afwijking als gevolg van variaties in de omstandigheden samen
$\underline{\varepsilon}_i/x$	$\frac{\varepsilon}{b} \underline{y}_i/x + \frac{\varepsilon}{m} \underline{y}_i/x = \underline{y}_i/x - \mu/x$: de totale afwijking van \underline{y}_i/x t.o.v. μ/x

σ_y^2/x	var $\underline{\epsilon}/x$: de variantie*) van de totale afwijking, in de populatie van herhaalde bepalingen aan steeds nieuwe monsters, bij dosis x
σ^2	var $(\frac{\hat{1}}{D_{10}})$: de variantie*) van de schatter van $\frac{1}{D_{10}}$
TC	log (totale aantal bacteriën)
TSC	log (aantal sporenvormende bacteriën)
TANC	log (aantal anaerobe bacteriën)
TANSC	log (aantal anaerobe sporenvormende bacteriën)
f	het aantal vrijheidsgraden
α	de onbetrouwbaarheid
R	de range van stralingsdoses waarbinnen het verband lineair is
Δ	het verschil tussen twee D_{10} -waarden
Δ_{inv}	het verschil tussen twee $\frac{1}{D_{10}}$ -waarden

4. Uitgangspunten/veronderstellingen en hun relevantie; model voor de waarnemingen

In de eerste plaats moeten de stochastische componenten voldoen aan enkele eigenschappen. In feite hoeven deze eigenschappen alleen te gelden voor $\underline{\epsilon}/x$, maar omdat $\underline{\epsilon}/x$ de som is van twee begrijpelijke termen, komt het neer op eisen die aan deze termen afzonderlijk worden gesteld. $\underline{\epsilon}_m/x$ heeft een verdeling in de populatie van werkelijke

*) variantie = gemiddelde gekwadrateerde afwijking t.o.v. de gemiddelde waarde in de populatie

aantallen bacteriën in steeds nieuwe monsters onder de proefomstandigheden, $\frac{\epsilon}{x}$ heeft een verdeling in de populatie van steeds herhaalde waarnemingen aan hetzelfde monster onder de proefomstandigheden. Voor deze verdelingen moet gelden, dat ze:

- 1^e) niet afhangen van x of van de A_w -waarde,
- 2^e) een verwachtingswaarde gelijk aan 0 hebben,
- 3^e) zogenaamde "normale verdelingen" zijn,
- 4^e) bovendien moet gelden, dat de verschillende trekkingen uit deze verdelingen onafhankelijk zijn.

Als de verdeling van $\frac{\epsilon}{x}$ afhangt van x , dan ontstaat een systematische fout bij het schatten van D_{10} (zie blz.126). De 2^e en 3^e eigenschap maken het mogelijk, dat de systematische fout bij het schatten van N/x , die het gevolg is van het overgaan op logaritmen, wordt geëlimineerd (zie blz.125 en 126).

De 3^e en de 4^e eigenschap moeten gelden opdat de nauwkeurigheid kan worden geschat en opdat toetsen kunnen worden uitgevoerd. Hierbij wordt er gewoonlijk ook van uitgegaan, dat de verdeling van $\frac{\epsilon}{x}$ ook niet afhangt van de A_w -waarden. Dit geheel van eigenschappen maakt het mogelijk om de statistische analyse op eenvoudige wijze uit te voeren.

Voor de 2^e en 4^e eigenschap kan in belangrijke mate worden gezorgd via de proefopzet door:

- * aselechte keuze van de monsters,
- * aselechte volgorde van werken (niet de behandelingen één voor één na elkaar afwerken) en
- * te zorgen voor afwezigheid van systematische fouten in de bepaling-methodiek.

Echter, de gebruikelijke praktijk is, dat de bepalingen in een van te voren vastgelegde volgorde worden afgewerkt, veelal behandeling na behandeling, omdat dit zo'n overzichtelijke manier van werken is. Vaak ook is het vrijwel ondoenlijk aselechte monsters te verkrijgen. De gevolgen hiervan hangen af van welke variaties overheersend zijn. De praktijk-werkwijze zal zijn ingegeven door het uitgangspunt, dat variaties in de tijd die alle waarnemingen op een tijdstip beïnvloeden nauwelijks een rol spelen. Het zou echter overweging verdienen om een tussenweg te

bewandelen tussen een volledig aselecte volgorde en de bovengenoemde volgorde, zodat een overzichtelijke volgorde gepaard gaat met goed analyseerbare waarnemingen (zonder veronderstellingen nodig te hebben die vermijdbaar zijn).

Of de 1^e eigenschap al of niet geldt, valt nauwelijks via de proefopzet te beïnvloeden. Alleen is achteraf enigszins na te gaan of de variantie afhangt van x . De 3^e eigenschap, normaliteit, is verder ook nodig zodra wordt getoetst.

Voorts is nodig:

5^e) de eventuele fout bij het toedienen van de stralingsdosis is klein: het effect ervan is verwaarloosbaar ten opzichte van $\underline{\epsilon}/x$.

Zou dit niet gelden, dan zijn de waarnemingen bij eenzelfde x (die allemaal tegelijk worden bestraald) afhankelijk en wel positief gecorreleerd. Dit heeft consequenties voor de nauwkeurigheid van de schatting van D_{10} en van de vergelijking van D_{10} -waarden bij verschillende A_w -waarden. Er bestaat overigens weinig twijfel over dat dit effect inderdaad verwaarloosbaar is.

6^e) het verband tussen $\log(N/x)$ en x is bij elke A_w -waarde exact lineair:

$$\log(N_j/x) = A_j - B_j x,$$

waarin A_j het intercept en B_j de richtingscoëfficiënt bij A_w -waarde j voorstelt. Dit is nodig om überhaupt een zinnige D_{10} -waarde te kunnen definiëren: D_{10_j} is dan gelijk aan $\frac{1}{B_j}$.

Het model voor de i -de waarneming bij A_w -waarde j en dosis x , gebaseerd op de proefprocedure en de bovengenoemde uitgangspunten is:

$$y_{ij}/x = \mu_j/x + \underline{\epsilon}_{ij}/x$$

met alle $\underline{\epsilon}_{ij}$ onderling onafhankelijk normaal verdeeld met verwachtingswaarde 0 en variantie σ_y^2/x .

Aangezien het gaat om het schatten van D_{10} moet in ieder geval D_{10} als parameter in het model voorkomen. Hiertoe is het nodig, dat de relatie tussen μ_j/x en $\log(N_j/x)$ wordt gegeven. Deze twee grootheden zijn niet aan elkaar gelijk, omdat het gemiddelde van de logaritmen van een aantal getallen (μ_j/x) niet gelijk is aan de logaritme van het gemiddelde van deze getallen ($\log(N_j/x)$). Hoe groot het verschil is tussen μ_j/x en $\log(N_j/x)$ is alleen te berekenen als de wijze waarop de waarnemingen bij dezelfde x variëren bekend is, dus als de verdeling van $\underline{\varepsilon}_{1j}/x$ bekend is. Volgens de veronderstellingen is dit de normale verdeling met verwachtingswaarde 0 en variantie σ_y^2 , op grond waarvan kan worden afgeleid:

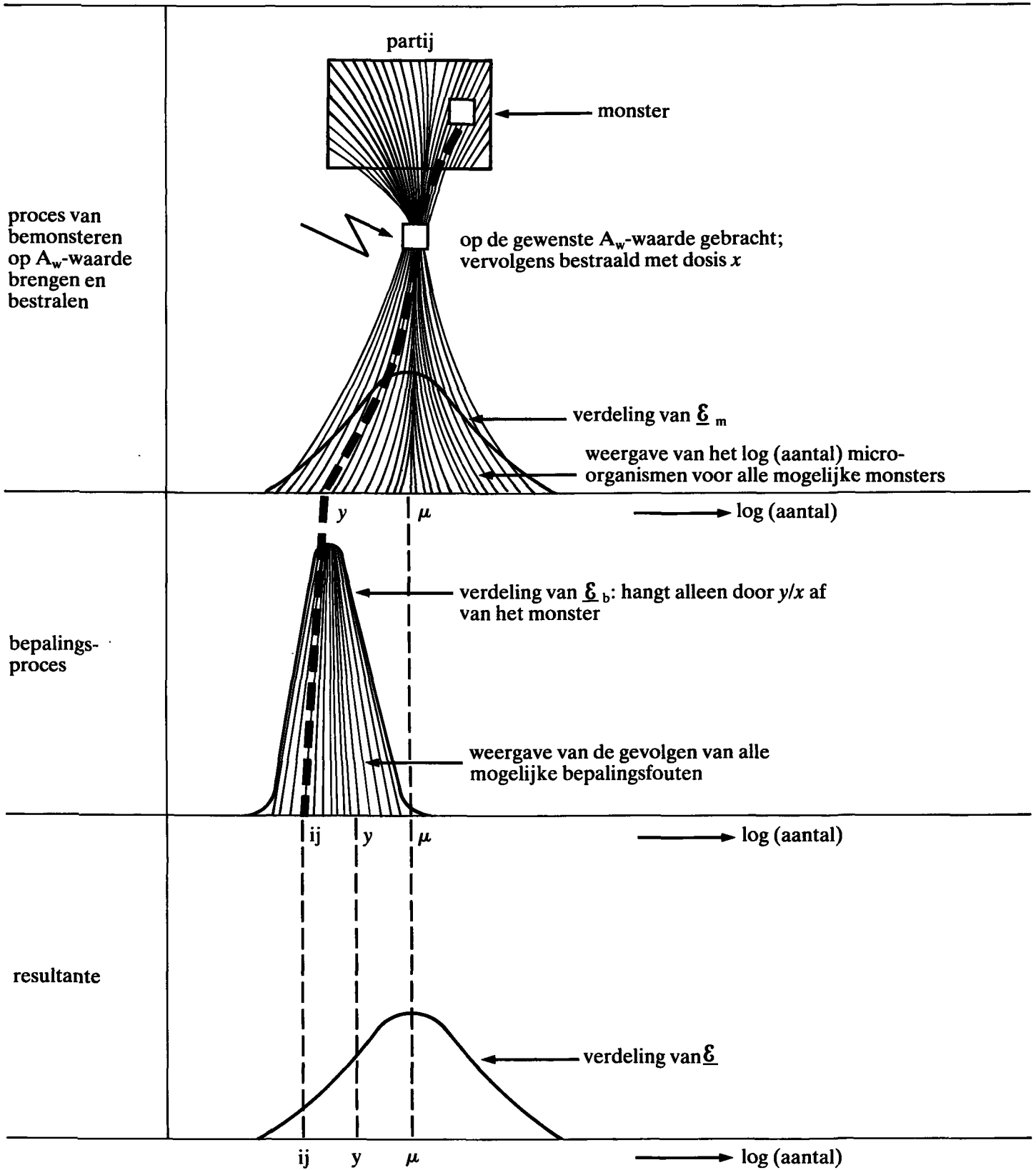
$$\mu_j/x = \log(N_j/x) - \frac{1}{2} \sigma_y^2,$$

zodat het model voor de waarnemingen wordt:

$$\begin{aligned} \underline{y}_{1j}/x &= \log(N_j/x) - \frac{1}{2} \sigma_y^2 + \underline{\varepsilon}_{1j}/x = \\ &= A_j - \frac{1}{2} \sigma_y^2 - B_j x + \underline{\varepsilon}_{1j}/x = \\ &= A^* - \frac{1}{D_{10}} x + \underline{\varepsilon}_{1j}/x \end{aligned}$$

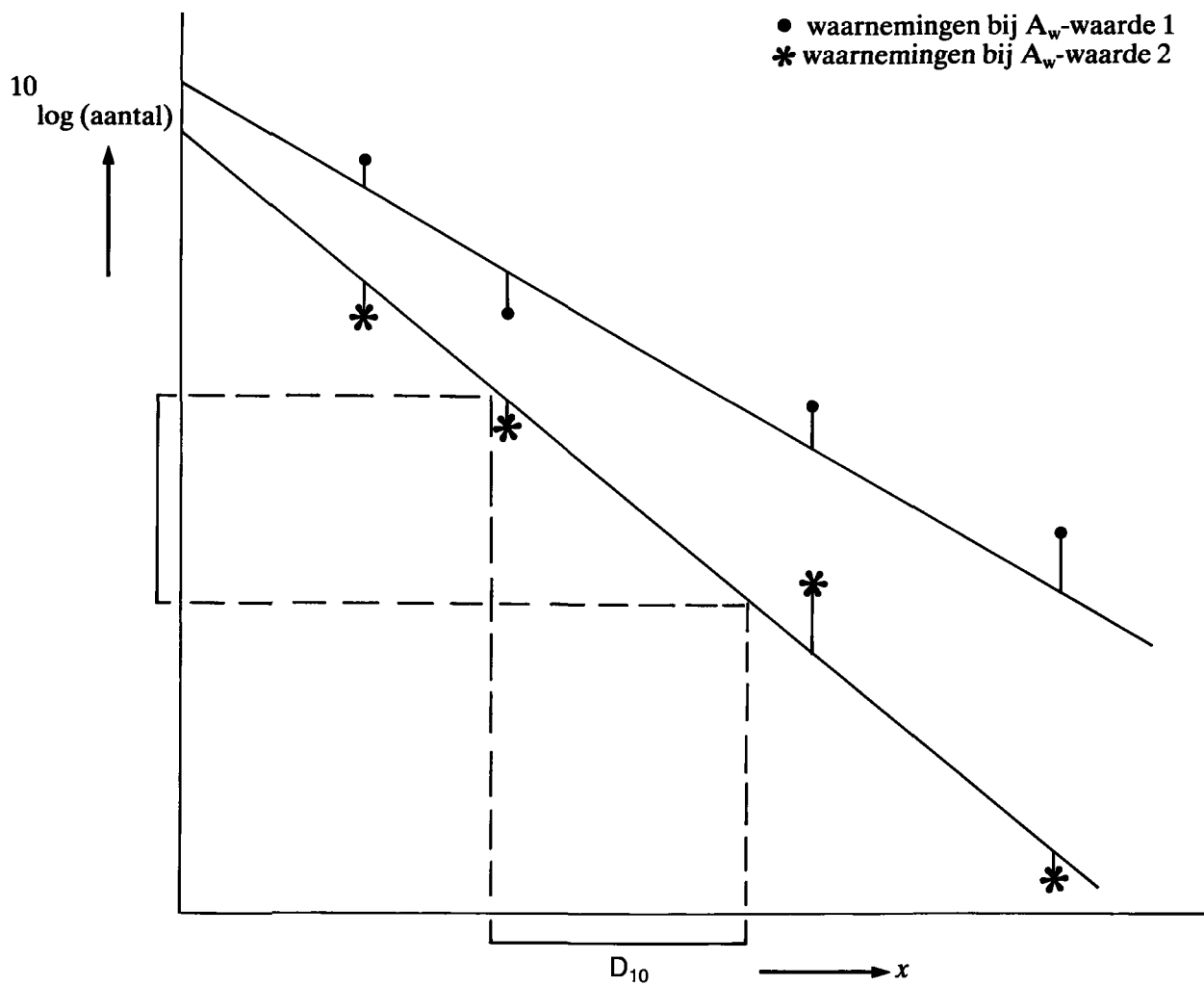
Hieruit valt ook af te lezen waarom σ_y^2/x niet af mag hangen van x . Was deze afhankelijkheid er wel, dan zou uit de lineaire relatie met x niet zonder meer D_{10} geschat kunnen worden.

De figuren 1 en 2 geven een samenvatting van wat de modelcomponenten voorstellen en van wat de veronderstellingen zijn.



Alleen de plaats van μ op de horizontale as hangt af van de A_w -waarde en de stralingsdosis, niet ξ_b en ξ_m (zie ook figuur 2)

Figuur 1: Visuele weergave van de variaties die een rol spelen. De veronderstellingen zijn ter plaatse vermeld. De indices i en j , evenals de aanduiding van de stralingsdosis, zijn weggelaten. Een mogelijke realisatie wordt aangegeven door: — — —



Figuur 2: Een voorbeeld van hoe μ_j/x afhangt van A_w -waarde j en van stralingsdosis x , met enkele mogelijke waarnemingen.

5. Het schatten en vergelijken van D_{10} -waarden

De situatie is nu die van een enkelvoudige lineaire regressie van y op x , waarbij het gaat om het schatten en vergelijken van hellingen bij verschillende A_w -waarden. Vragen zijn hoe de x 'en te kiezen zo, dat de hellingen zo nauwkeurig mogelijk worden geschat, en hoeveel herhalingen nodig zijn om een gegeven verschil tussen hellingen nog met een behoorlijke kans te kunnen aantonen. Standaard enkelvoudige lineaire regressietechniek geeft zowel een schatting van de helling van de lijn

$(\frac{1}{D_{10}})$ als van de variantie rond de lijn (σ_y^2). Als de proef zo wordt uitgevoerd, dat r waarnemingen worden verkregen bij iedere combinatie van A_w -waarde en stralingsdosis x , dan is de variantie (σ^{*2}) van de schatter van $(\frac{1}{D_{10}})$:

$$\text{var} \left(\frac{\hat{1}}{D_{10}} \right) = \sigma^{*2} = \frac{\sigma_y^2}{r S_1^2} \quad [5.1.]$$

met S_1^2 de som van de gekwadrateerde afwijkingen van de stralingsdoses ten opzichte van de gemiddelde dosis in de situatie $r = 1$. Op grond van de veronderstelde normaliteit van y is ook de schatter $\frac{1}{D_{10}}$ normaal verdeeld.

Deze schatter is de basis van de verdere beschouwing.

Een $100(1-\alpha)\%$ betrouwbaarheidsinterval voor $\frac{1}{D_{10}}$ is

$$\frac{\hat{1}}{D_{10}} - t_{f,\alpha} \hat{\sigma}^* < \frac{1}{D_{10}} < \frac{1}{D_{10}} + t_{f,\alpha} \hat{\sigma}^* \quad [5.2.]$$

met $t_{f,\alpha}$ het $100(1-\frac{\alpha}{2})$ procentpunt van de Student-verdeling met f vrijheidsgraden en $\hat{\sigma}^*$ de schatting van σ^* , verkregen via regressieanalyse.

Uit [5.2.] volgt direct, dat

$$\frac{1}{\frac{\hat{1}}{D_{10}} + t_{f,\alpha} \hat{\sigma}^*} < D_{10} < \frac{1}{\frac{\hat{1}}{D_{10}} - t_{f,\alpha} \hat{\sigma}^*} \quad [5.3.]$$

een $100(1-\alpha)\%$ betrouwbaarheidsinterval is voor D_{10} . De lengte van dit betrouwbaarheidsinterval is

$$L = \frac{2t_{f,\alpha} \hat{\sigma}^*}{\left(\frac{\hat{1}}{D_{10}}\right)^2 - (t_{f,\alpha} \hat{\sigma}^*)^2} \quad [5.4.]$$

\underline{L} is een stochastische grootte omdat $\frac{\hat{1}}{D_{10}}$ en $\hat{\sigma}^*$ stochastisch zijn.

De gemiddelde waarde van L over alle mogelijke schattingen $\frac{\hat{1}}{D_{10}}$ en $\hat{\sigma}^*$ (de verwachtingswaarde van \underline{L} : $E\underline{L}$) lijkt een redelijke maat om de nauwkeurigheid van de schatter aan te geven, maar is niet eenvoudig exact te berekenen. Als benadering hiervoor wordt genomen:

$$E\underline{L} \approx \frac{2t_{f,\alpha} \hat{\sigma}^*}{\left(\frac{\hat{1}}{D_{10}}\right)^2 - (t_{f,\alpha} \hat{\sigma}^*)^2} \quad [5.5.]$$

Het verschil tussen twee D_{10} -waarden wordt aangeduid met Δ , het verschil tussen twee $\frac{1}{D_{10}}$ -waarden met Δ_{inv} .

Het vergelijken van twee D_{10} -waarden kan worden gedaan via de toets met hypothesen:

$$H_0: \Delta = 0 \text{ (of } \Delta_{inv} = 0)$$

$$H_a: \Delta = \Delta_a \neq 0 \text{ (of } \Delta_{inv} = \Delta_{inv_a} \neq 0)$$

De toetsingsgrootte kan worden gebaseerd op $\hat{\Delta}_{inv}$ (= het geschatte verschil tussen twee hellingen op grond van regressie), normaal verdeeld met verwachtingswaarde

$$\Delta_{inv} = \frac{\Delta}{\left(\bar{D}_{10}\right)^2 - \frac{1}{t} \Delta^2} \quad [5.6.]$$

en variantie $2 \hat{\sigma}^{*2}$. Het onderscheidingsvermogen is hierbij, gegeven de onbetrouwbaarheid α , alleen afhankelijk van

$$\delta = \frac{\Delta \ln v}{\sigma_y \sqrt{2}} \quad [5.7.]$$

δ is de niet-centraliteitsparameter van de niet-centrale Student-verdeling van de toetsingsgrootte onder de alternatieve verdeling H_a . Naarmate δ groter wordt neemt het onderscheidingsvermogen toe. Uit [5.1.], [5.5.], [5.6.] en [5.7.] volgt dat, naarmate $r S_1^2$ groter is en σ_y^2 en de betrokken D_{10} -waarden kleiner zijn, het betrouwbaarheidsinterval voor D_{10} korter is en het onderscheidingsvermogen van de toets op gelijkheid van twee D_{10} -waarden groter is. Het onderscheidingsvermogen van de toets hangt verder alleen nog van $\Delta \ln v$ af. Uit [5.6.] kan worden afgelezen hoe $\Delta \ln v$ samenhangt met Δ en \bar{D}_{10} .

6. De optimale proefopzet

Van de grootheden die de nauwkeurigheid beïnvloeden zijn er slechts enkele via de proefopzet te beïnvloeden. α (het risico dat men wil lopen om ten onrechte de nulhypothese te verwerpen) heeft niets te maken met de proef, σ_y^2 wordt bepaald door de proefprocedure (bijvoorbeeld de bepalingsmethode) en kan worden geschat (met een zekere en onvermijdelijke onnauwkeurigheid) uit de beschikbare waarnemingen. $\Delta \ln v$, \bar{D}_{10} en Δ zijn uiteraard van te voren niet bekend, maar volledig bepaald door de A_w -waarden die in de proef worden onderzocht. Het gaat er juist om hierover iets te weten te komen. Te beïnvloeden via de proefopzet zijn alleen S_1^2 door de keuze van de stralingsdoses en het aantal herhalingen r . De volgende beschouwing is erop gericht deze keuzen zo verantwoord mogelijk te doen.

De keuze van de stralingsdoses

Uitgangspunt bij de keuze is, dat het geen wezenlijk verschil maakt in kosten of moeite hoeveel stralingsdoses worden gekozen, maar dat de kosten uitsluitend bepaald worden door het aantal waarnemingen. Bij een lineair verloop met de stralingsdosis zoals hier verondersteld, is de optimale keuze: twee doses die zo ver mogelijk uit elkaar liggen. Onderstaande tabel, die overgenomen is uit Cox (1958, blz. 140)*) geeft aan

*) Cox, D.R. (1958), Planning of Experiments. Wiley & Sons, New York

het aantal waarnemingen dat naar verhouding meer zou moeten worden gedaan in geval meer dan twee equidistante stralingsdoses zouden worden gebruikt, om $\frac{1}{D_{10}}$ met dezelfde precisie te schatten als bij twee stralingsdoses.

aantal stralingsdoses	relatief aantal waarnemingen
2	1
3	1.225
4	1.342
5	1.414

Tabel 1: De efficiëntie van verschillende keuzen van aantallen stralingsdoses (zie tekst).

Voor verdere berekeningen wordt uitgegaan van twee stralingsdoses. Als R de range van deze doses voorstelt, geldt:

$$S_1^2 = \frac{1}{2} R^2 \quad [6.1.]$$

De keuze van R hangt samen met de bij de ingestelde A_w -waarden behorende D_{10} -waarden en met het aantal bacteriën in de produkten. Een eerste vereiste is, dat ook bij de hoogste stralingsdosis het aantal bacteriën goed telbaar is. Bij een lineair verband in het traject 0 tot R KGY moet dus gelden:

$$y/0 - R \cdot \frac{1}{D_{10}} > c, \text{ ofwel } R < D_{10}(y/0 - c) \quad [6.2.]$$

met $y/0$ de logaritme van het aantal bacteriën bij dosis 0 en c de logaritme van het minimale aantal goed telbare bacteriën; c is afhankelijk van de bepalingsmethodiek. Bij de toegepaste methodiek lijkt $c = 1$ een redelijke waarde. Een tweede eis is uiteraard, dat de relatie tussen de logaritme van het aantal bacteriën en de stralingsdosis lineair moet zijn. In dit verband verdient het aanbeveling R niet te groot te kiezen. Wanneer er ook dan nog onzekerheid is omtrent lineariteit is het

verstandig meerdere stralingsdoses te kiezen opdat er in ieder geval controle bestaat op afwijking van een lineair verband.

De gegevens van de peterselie betreffen vier variabelen die zijn gemeten bij iedere combinatie van niveaus van A_w -waarden (0.1, 0.33 en 0.75) en van stralingsdoses (0, 1, 2, 4 en 8 KGY); $r = 1$, $S_1^2 = 40$. Enkele waarnemingen ontbreken. Toepassing van regressieanalyse per variabele levert de schattingen van de parameters D_{10} en σ_y^2 , met f het aantal vrijheidsgraden op basis waarvan σ_y^2 is geschat. Deze zijn vermeld in tabel 2.

Op grond van de modelveronderstellingen is $\frac{f\hat{\sigma}_y^2}{\sigma_y^2}$ chikwadraat verdeeld met f vrijheidsgraden. Hieruit wordt een 95% betrouwbaarheidsinterval afgeleid, waarvan het resultaat eveneens in tabel 2 is vermeld.

		schattingen van de parameter		f	95% betrouwbaarheidsinterval voor σ_y^2	
		D_{10}	σ_y^2		ondergrens	bovengrens
variabele	TC	1.7	0.055	9	0.026	0.183
	TSC	1.5	0.017	7	0.0074	0.070
	TANC	1.8	0.136	7	0.059	0.563
	TANSC	1.4	0.063	6	0.026	0.306

Tabel 2: Conclusies ten aanzien van D_{10} en σ_y^2 op basis van de beschikbare gegevens van peterselie.

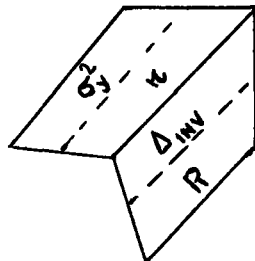
De keuze van het aantal herhalingen

Het optimale aantal herhalingen r wordt vastgesteld door de nauwkeurigheid bij verschillende waarden van r af te wegen tegen de kosten die moeten worden gemaakt om het bij r behorende aantal waarnemingen te verkrijgen. Dit afwegen is een zaak van de onderzoeker. De nauwkeurigheid wordt gekarakteriseerd door de verwachte lengte van het

betrouwbaarheidsinterval voor een D_{10} -waarde (benaderd via [5.5.]) en door het onderscheidingsvermogen van de toets op gelijkheid van twee D_{10} -waarden bij gegeven verschil Δ_{inv} . Proefopzetten die worden vergeleken zijn de combinaties van R (de range van doses waarin het lineaire verband geldt): 4, 8 en 12 en van r (het aantal herhalingen): 2, 3, 5 en 10. α (de onbetrouwbaarheid) wordt op 0.05 gesteld.

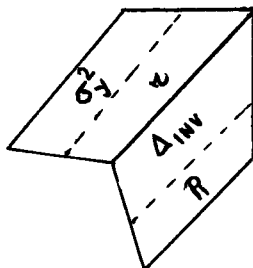
De onzekerheid ten aanzien van σ_y^2 n rekening gebracht door de berekeningen uit te voeren voor zowel de schatting als voor de onder- en bovengrens van het 95% betrouwbaarheidsinterval ervan. Als D_{10} -waarden worden gekozen 1, 1.5, 2, 2.5. Het aantal A_w -waarden dat in de proef wordt vergeleken is gesteld op 3. Dit aantal heeft invloed op het aantal vrijheidsgraden dus op $t_{f,\alpha}$ en moet daarom in de beschouwing worden betrokken.

Voor de berekening van het onderscheidingsvermogen worden voor Δ_{inv} de waarden 0.05, 0.1, 0.2, 0.3, en 0.5 gekozen. Welke waarde voor Δ en \bar{D}_{10} hiermee corresponderen kan worden afgeleid uit [5.6.]. De resultaten zijn weergegeven in de tabellen 3 en 4.



		0.18					0.06					0.03				
		2	3	5	10		2	3	5	10		2	3	5	10	
1 4 2 25	1	0.56	0.39	0.28	0.19	0.31	0.22	0.16	0.11	0.08	0.21	0.16	0.11	0.08		
	1.5	1.38	0.92	0.65	0.44	0.71	0.50	0.36	0.25	0.49	0.35	0.26	0.17			
	2	2.85	1.76	1.20	0.79	1.32	0.92	0.65	0.44	0.89	0.63	0.46	0.31			
	2.5	5.62	3.04	1.96	1.26	2.18	1.47	1.04	0.70	1.43	1.00	0.72	0.49			
1 8 2 2.5	1	0.26	0.19	0.14	0.10	0.15	0.11	0.08	0.05	0.11	0.08	0.06	0.04			
	1.5	0.61	0.43	0.31	0.21	0.34	0.25	0.18	0.12	0.24	0.17	0.13	0.09			
	2	1.11	0.78	0.56	0.38	0.61	0.44	0.32	0.22	0.43	0.31	0.23	0.16			
	2.5	1.82	1.25	0.89	0.60	0.97	0.69	0.50	0.34	0.67	0.49	0.35	0.24			
1 12 2 2.5	1	0.17	0.13	0.09	0.06	0.10	0.07	0.05	0.04	0.07	0.05	0.04	0.03			
	1.5	0.40	0.29	0.21	0.14	0.23	0.16	0.12	0.08	0.16	0.12	0.08	0.06			
	2	0.71	0.51	0.37	0.25	0.40	0.29	0.21	0.15	0.28	0.21	0.15	0.10			
	2.5	1.14	0.81	0.58	0.40	0.64	0.46	0.33	0.23	0.45	0.32	0.24	0.16			

Tabel 3: De verwachte lengte van het 95% betrouwbaarheidsinterval voor D_{10} onder verschillende proefsituaties, berekend met formule [5.5.].



		0.18				0.06				0.03			
		2	3	5	10	2	3	5	10	2	3	5	10
4	0.05	0.06	0.07	0.08	0.12	0.08	0.09	0.14	0.25	0.11	0.15	0.23	0.45
	0.1	0.08	0.12	0.17	0.32	0.14	0.25	0.41	0.73	0.27	0.44	0.69	0.95
	0.2	0.20	0.32	0.52	0.84	0.49	0.72	0.93	1.00	0.77	0.95	1.00	1.00
	0.3	0.38	0.61	0.85	0.99	0.82	0.97	1.00	1.00	0.98	1.00	1.00	1.00
8	0.05	0.08	0.12	0.17	0.32	0.14	0.25	0.41	0.74	0.27	0.44	0.69	0.95
	0.1	0.20	0.32	0.52	0.84	0.49	0.72	0.93	1.00	0.77	0.95	1.00	1.00
	0.2	0.60	0.85	0.98	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.3	0.91	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
12	0.05	0.13	0.19	0.32	0.61	0.31	0.49	0.74	0.97	0.54	0.78	0.96	1.00
	0.1	0.38	0.61	0.82	0.99	0.82	0.97	1.00	1.00	0.98	1.00	1.00	1.00
	0.2	0.91	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Tabel 4: Het onderscheidingsvermogen van de toets op gelijkheid van twee D_{10} -waarden onder verschillende proefsituaties.

Uit tabel 4 is bijvoorbeeld af te lezen, dat er tien herhalingen moeten worden gedaan om met kans 0.74 aan te tonen, dat een verschil Δ_{inv} van 0,05 bestaat, wanneer in werkelijkheid $\sigma_y^2 = 0.06$ en wanneer $R = 8$. Onder die omstandigheden zou de verwachte lengte van het 95% betrouwbaarheidsinterval voor D_{10} ongeveer 0,05 zijn als D_{10} in werkelijkheid 0,5 was en 0,34 als D_{10} 2,5 zou zijn.

De onzekerheid van de conclusies wordt weerspiegeld door de onbetrouwbaarheid α , het onderscheidingsvermogen en ook door onzekerheid omtrent σ_y^2 en D_{10} . Wil men ten aanzien van σ_y^2 tamelijk safe spelen, dan zou men hiervoor bijvoorbeeld de bovengrens van het 95% betrouwbaarheidsinterval voor σ_y^2 kunnen kiezen. Na iedere proef krijgt men meer informatie over

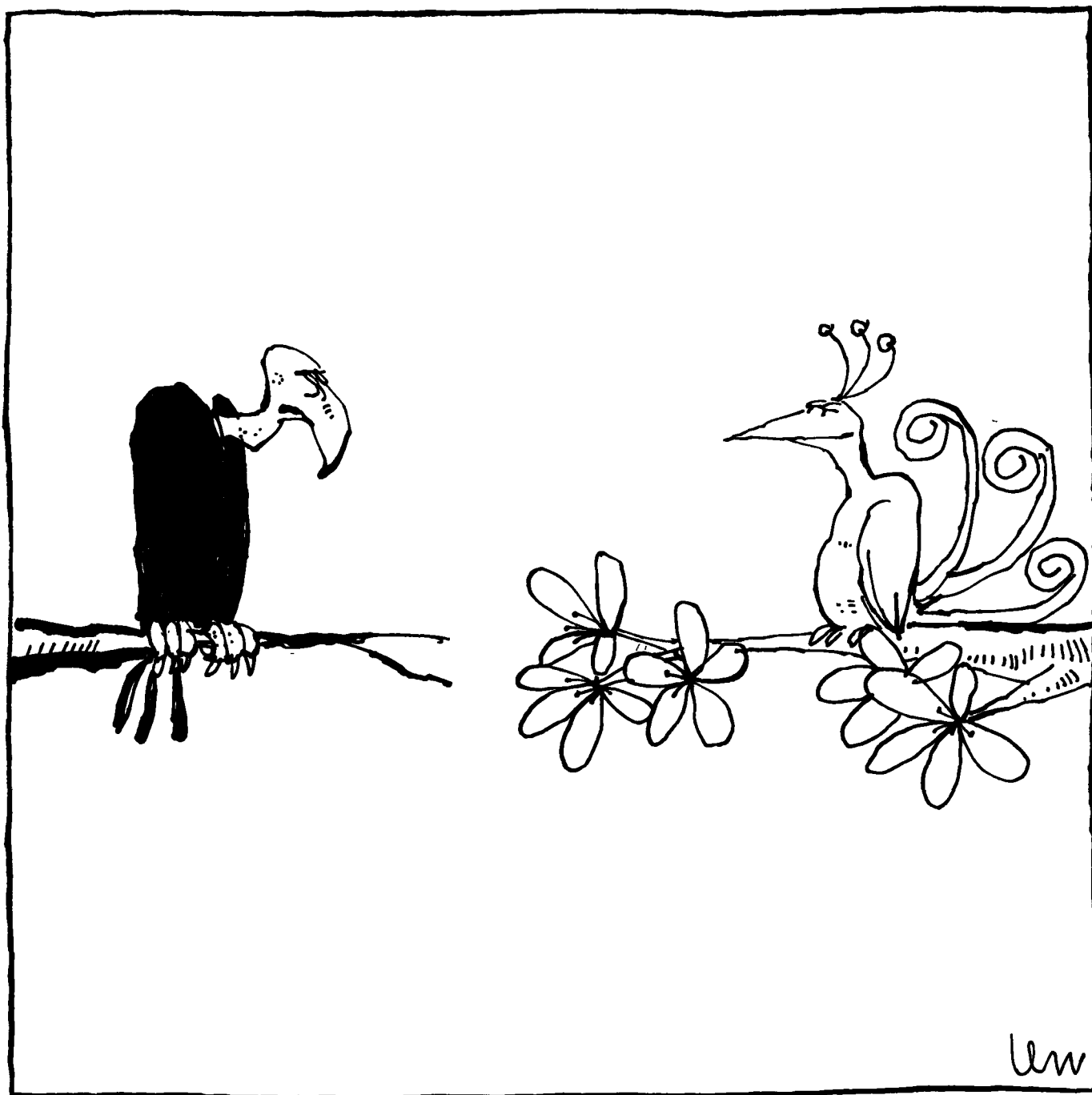
σ_y^2 en ook over de D_{10} -waarden die te verwachten zijn en dus betere mogelijkheden om de proef optimaal op te zetten. Bij de heersende onzekerheden is optimaliteit nog een tamelijk beperkt begrip.

Dit laatste geldt des te meer als men bedenkt dat mogelijk ook het model voor de waarnemingen niet volledig juist is.

Vogelbevolking en vegetatiestructuur

Drs. J. H. Oude Voshaar

Rapport van de statistische begeleiding van de RIN-projecten
nrs. 99 en 144.



Inhoud

1. Inleiding en samenvatting	143
2. De gegevens	144
2.1. <i>De waarnemingen en de definities van de variabelen</i>	144
2.2. <i>Correctie van S voor de oppervlakte van het perceel</i>	147
3. Regressie en selectie van variabelen (theorie)	148
3.1. <i>Regressie</i>	148
3.2. <i>Selectie van variabelen</i>	150
4. De statistische analyse	153
4.1. <i>Verantwoording</i>	153
4.2. <i>Resultaten</i>	154
4.3. <i>Conclusies</i>	157
5. Slotopmerkingen	158
Literatuurverwijzingen	158
Bijlage I	160
Bijlage II	161

1. Inleiding en samenvatting

Van een oud loofbosgebied genaamd "Hasbruch", gelegen in Nieder-Sachsen (W.-Dld), zijn in 1979 broedvogelinventarisaties verricht en tevens zijn er opnamen van de vegetatiestructuur gemaakt. Doel van het onderzoek¹⁾ is het verklaren van kenmerken van de vogelbevolking uit de vegetatiestructuur. (Verwacht wordt dat een bos met een meer gevarieerde vegetatiestructuur ook een rijkere vogelsamenstelling zal bezitten). Hiertoe is het voornoemde bosgebied verdeeld in 17 percelen, die elk vrij homogeen zijn met betrekking tot de vegetatiestructuur en waarvan de oppervlakte varieert van 2 tot 12 hectare (kleinere percelen werden buiten beschouwing gelaten).

Aan elk van de percelen zijn vier vogelvariabelen en tien vegetatievariabelen gemeten. De vogelvariabelen geven elk een bepaald aspect van de vogelbevolking weer en van de vegetatievariabelen hopen we dat ze tezamen een groot deel van de spreiding in de waargenomen vogelvariabelen zullen verklaren. Of dit inderdaad het geval is zal worden onderzocht met behulp lineaire regressieanalyse. Vervolgens zullen de belangrijkste vegetatievariabelen worden opgespoord. Bij deze "selectie van variabelen" vormt het feit dat de variabelen gecorreleerd zijn een complicerende factor.

Uit de analyses zal blijken dat een in dit onderzoek gedefinieerde variabele, die de heterogeniteit van de vegetatiestructuur in de laag van 10 tot 20 meter boven de grond beschrijft, een zeer sterk verband toont met de dichtheid en het aantal soorten van de broedvogels en met de dichtheid van de holenbroeders. De dichtheid van de stamfourageerders blijkt samen te hangen met de gemiddelde omtrek van de bomen.

Behalve op de regressie en de selectie van variabelen wordt in dit rapport nog ingegaan op de correctie van het soortenaantal voor het verschil in oppervlakte tussen de percelen.

1) Het betreft hier de projecten 99 en 144 van het Rijksinstituut voor Natuurbeheer, uitgevoerd onder leiding van Dr. P.F.M. Opdam.

2. De gegevens

2.1. De waarnemingen en de definities van de variabelen

Van elk van de 17 percelen is het aantal broedparen van elke vogelsoort bepaald volgens de karteringmethode (Opdam & Van Bladeren, 1981). Uit deze gegevens zijn vier vogelvariabelen afgeleid waarmee getracht wordt de "rijkdom" van de vogelbevolking te beschrijven.

Deze zijn per perceel als volgt gedefinieerd:

1. D = dichtheid van de broedparen.
2. S = het aantal broedende soorten dat in het perceel werd waargenomen.
3. D_{hol} = dichtheid van de holenbroeders.
4. D_{stam} = dichtheid van de stamfourageerders.

D , D_{hol} en D_{stam} zijn uitgedrukt in aantal paren per 10 ha.

De vegetatie is beschreven op basis van de volgende waarnemingen. Op een rechte lijn (transect), geheel binnen het perceel gelegen, werd op 50 punten, op een onderlinge afstand van telkens 5 meter, de aan- of afwezigheid van vegetatie bepaald voor elk van de volgende 4 lagen:

0.75 - 2 m, 2 - 5 m, 5 - 10 m en 10 - 20 m boven de grond. Bij aanwezigheid werd ook de betreffende plantensoort geregistreerd. Verder werd voor elk van de 50 punten de omtrek (op borsthoogte) van de dichtstbijzijnde boom gemeten.

Hieruit zijn de volgende variabelen afgeleid, waarmee de vegetatiestructuur wordt beschreven.

1. \bar{O} = de gemiddelde omtrek van de 50 gemeten bomen.
In tabel 1 (bijlage 1) is deze uitgedrukt in cm.
2. VCO = de variatiecoëfficiënt van de omtrek van de 50 gemeten bomen (= het gemiddelde gedeeld door de standaardafwijking).
In tabel 1 (bijlage 1) is deze in procenten uitgedrukt.
3. B_i (voor $i = 1, \dots, 4$) = het bedekkingspercentage in de vegetatielagen 0.75 - 2 m, 2 - 5 m, 5 - 10 m en 10 - 20 m.
4. $TOTBED$ = de totale bedekking = $B_1 + B_2 + B_3 + B_4$.

5. H_B = de "lagendiversiteit", gedefinieerd als: $-\sum_{i=1}^4 p_i \ln p_i$,
 waarbij

$$p_i = \frac{B_i}{B_1 + B_2 + B_3 + B_4} \quad (i = 1, \dots, 4).$$

6. H_S = "soortendiversiteit", gedefinieerd als

$$-\sum_{i=1}^4 \sum_j p_{ij} \ln p_{ij},$$

waarbij p_{ij} = relatieve abundantie van plantensoort j in laag i . H_S is dus gelijk aan de som van vier Shannon-Weaver-indices van de verdeling van de soortabundanties per vegetatielaag.

7. HET_i (voor $i = 1, \dots, 4$) = $^{10}\log(N_{i+} \cdot L_{i+} \cdot L_{i-} + 1)$ waarin i op de vegetatielaag slaat en waarin

N_i = het aantal "runs" met vegetatie aanwezig,

L_{i+} = het aantal verschillende runlengtes waarin vegetatie aanwezig is,

L_{i-} = het aantal verschillende runlengtes zonder vegetatie.

HET_i is een maat voor de heterogeniteit van de i^e vegetatielaag. De variabele neemt een hogere waarde aan naarmate de vegetatiestructuur in horizontale richting onregelmatiger is. H_B is een maat voor de heterogeniteit in de verticale richting en H_S geeft de heterogeniteit van de vegetatiesoorten weer.

De variabele \bar{O} wordt alleen van belang geacht voor de verklaring van D_{stam} . Van de variabelen B_1 t/m B_4 wordt alleen B_4 enigszins van belang geacht, maar deze is vrij sterk gecorreleerd met HET_4 , zodat we B_4 buiten beschouwing laten.

Verder is ook het percentage dode bomen bepaald (voor de 50 gemeten bomen), maar omdat het erg klein is ($< 8\%$), vonden we de relatieve fout hierin, ervan uitgaande dat dit percentage representatief is voor het gehele perceel te groot om deze variabele in de analyse op te kunnen nemen.

Er is dus een informele voorselectie van variabelen gepleegd. Voor een uitgebreidere argumentatie van de keuze van de vogel- en vegetatie-variabelen wordt verwezen naar Opdam & Van Bladeren, 1981.

Voorbeeld

Ter verduidelijking van de bovenstaande definities van de vegetatie-variabelen volgt hier een (gefingerd) voorbeeld, waarin voor de 50 punten de aan- of afwezigheid van vegetatie in de desbetreffende laag met * of - wordt voorgesteld.

			B_1
vegetatie- laag	4	****-***-***-***-***-***-***-***-***-***-***-***-***-***	26
	3	**-*	32
	2	-----*****-----*****-----*****	14
	1	*-*	20

In dit voorbeeld geldt:

$$B_1 = 20, \quad B_2 = 14, \quad B_3 = 32, \quad B_4 = 26$$

$$\text{TOTBED} = 20 + 14 + 32 + 26 = 92$$

$$H_B = -\left(\frac{20}{92} \ln \frac{20}{92} + \dots + \frac{26}{92} \ln \frac{26}{92}\right) = 1.34$$

$$\text{HET}_1 = \log (12 \cdot 3 \cdot 4 + 1) = 2.16$$

$$\text{HET}_2 = \log (3 \cdot 2 \cdot 4 + 1) = 1.40$$

$$\text{HET}_3 = \log (13 \cdot 5 \cdot 3 + 1) = 2.29$$

$$\text{HET}_4 = \log (12 \cdot 5 \cdot 5 + 1) = 2.48$$

2.2. Correctie van S voor de oppervlakte van het perceel

Het is bekend dat het soortenaantal S samenhangt met de oppervlakte van het te beschouwen perceel. In de literatuur wordt dit verband veelal als volgt beschreven.

$$S = c \cdot \text{opp}^z$$

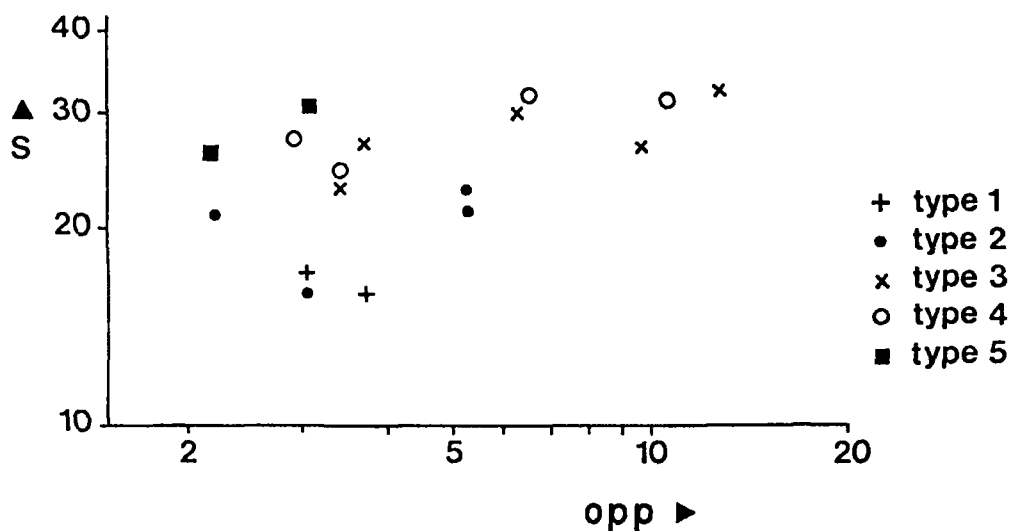
Strikt genomen dient in deze formule voor S de verwachte waarde van S geschreven te worden. Ook voor andere diergroepen dan vogels wordt een dergelijk verband geconstateerd. Voor vogels worden voor z doorgaans waarden tussen 0.15 en 0.30 opgegeven. Men kan c beschouwen als het soortenaantal omgerekend naar een standaardoppervlakte.

Deze correctie (= deling door opp^z) is uiteraard alleen mogelijk als de waarde van z bekend is. Uit onze waarnemingen kan de waarde van z echter worden geschat, omdat de onderzoekers van tevoren de percelen konden onderverdelen in vijf typen waarbinnen de percelen min of meer dezelfde vegetatiestructuur bezitten.

Als we $\log S$ uitzetten tegenover $\log(\text{opp})$, dan gaat het bovengenoemde verband over in:

$$\log S = \log c + z \cdot \log(\text{opp})$$

In figuur 2 zijn S en de oppervlakte beide logaritmisch uitgezet.



Figuur 2: Verband tussen $\log(S)$ en $\log(\text{opp})$.

Uiteraard komt het lineaire verband pas tot uiting als men de vijf verschillende bostypen onderscheidt die elk een verschillende waarde van c hebben. Als binnen elk type c constant is en als voor alle typen de waarden van z gelijk zijn, dan zullen de verwachte soortantallen op vijf evenwijdige lijnen liggen, maar uiteraard is er variatie om deze lijnen (alleen al tengevolge van het feit dat S alleen maar gehele waarden kan aannemen).

Figuur 2 suggereert dat het bovenstaande aardig opgaat voor de typen 2, 3 en 4. Voor de typen 1 en 5 valt weinig te zeggen over de lineariteit en de evenwijdigheid omdat van deze typen slechts twee percelen in het onderzoek zijn opgenomen, waarvan bovendien de oppervlakten bijna gelijk zijn.

Een regressieanalyse waarin vijf evenwijdige lijnen simultaan worden aangepast, levert bij onze waarnemingen als schatting van z op :
 $z = 0.15$. (Het 95%-betrouwbaarheidsinterval heeft als grenzen: 0.03 en 0.28). Deze schatting komt redelijk overeen met de in de literatuur opgegeven waarden.

In paragraaf 4 zal echter blijken dat de conclusies niet erg gevoelig zijn voor de waarden van z , die men voor de correctie gebruikt.

3. Regressie en selectie van variabelen (theorie)

Alvorens op de statistische analyse van dit onderzoek in te gaan zullen we de hierbij benodigde theorie van regressie en selectie van variabelen schetsen.

3.1. Regressie

Bij lineaire regressie probeert men een stochastische grootheid y te verklaren uit p x -variabelen (in ons geval is, behalve voor D_{stam} , p gelijk aan 8).

Hierbij wordt uitgegaan van het volgende lineaire model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \underline{e}$$

waarbij \underline{e} de "stochastische ruis" voorstelt, die normaal verdeeld

verondersteld wordt met verwachting 0 en variantie σ^2 (deze σ^2 heet de restvariantie).

Op basis van n waarnemingen y_1, \dots, y_n , waarvoor bij de i^e waarneming de x -variabelen x_{i1}, \dots, x_{ip} behoren, worden de schattingen van β_0 t/m β_p zodanig gekozen dat de restkwadraatsom KS_r gedefinieerd door

$$KS_r = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2$$

minimaal is.

We zeggen dat het model goed "past" als de geschatte restvariantie

$$\hat{\sigma}_r^2 \left(= \frac{KS_r}{n-p-1} \right) \text{ klein is ten opzichte van de totale variantie van } \underline{y} \text{ (die}$$

gedefinieerd is

$$\text{als } \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n-1} \cdot KS_{\text{tot}}, \text{ waarbij } \bar{y} \text{ gelijk is aan } \frac{1}{n} \sum y_i \text{).}$$

Als maat van de kwaliteit van het model wordt daarom vaak gebruik gemaakt van de grootte V^2 gedefinieerd door:

$$V^2 = 1 - \frac{\frac{1}{n-p-1} \cdot KS_r}{\frac{1}{n-1} \cdot KS_{\text{tot}}}$$

V^2 is de fractie van de totale variantie van \underline{y} die door het model wordt verklaard.

Opmerking

In de literatuur wordt vaak de (ons inziens minder geschikte) grootte R^2 gehanteerd als maat van de kwaliteit van het aangepaste model. R^2 wordt wel eens de hoeveelheid verklaarde variatie genoemd en is als volgt gedefinieerd:

$$R^2 = 1 - \frac{KS_r}{KS_{\text{tot}}}.$$

Als de p aangepaste variabelen in het geheel geen samenhang met \underline{y}

vertonen, dan is R^2 in verwachting niet gelijk aan 0 maar aan $\frac{p}{n-1}$. Als in dit geval p gelijk is aan $n-1$, dan is R^2 dus gelijk aan 1 en zou men kunnen concluderen dat met deze "nonsens"-variabelen alle variatie in y verklaard wordt!

Als p klein is ten opzichte van n , dan is uiteraard R^2 wel ongeveer gelijk aan $\frac{p}{n}$.

3.2. Selectie van variabelen

Het model dat alle gemeten variabelen bevat (plus de van belang geachte kwadraten, produkten en andere transformaties van de variabelen) noemen we het "volledige model". Als het volledige model goed past, dan worden de volgende vragen interessant:

1. Welke variabelen zijn de belangrijkste?
2. Welke variabelen kunnen uit het model worden weggelaten zonder dat het model (veel) slechter wordt, dat wil zeggen zonder dat V^2 (veel) kleiner wordt?

Als de x -variabelen ongecorrleerd zijn (zoals bijvoorbeeld bij een orthogonale proefopzet) dan levert de beantwoording van deze vragen geen problemen op. Immers: weglating van een variabele uit het model resulteert in een toename van de restkwadraatsom, die dan niet afhangt van de overige variabelen in het model. Deze toename noemt men het door deze variabele verklaarde deel van de restkwadraatsom.

Als dit deel groot is ten opzichte van de restvariantie van het volledige model, dan heet deze variabele "significant".

Als de x -variabelen echter gecorreleerd zijn, dan hangt de toename van de restkwadraatsom in het algemeen wèl af van het feit welke andere variabelen in het model aanwezig zijn. De situatie waarin x_i en x_j sterk gecorreleerd zijn is hiervoor zeer illustratief: als x_i in het model is opgenomen, dan heeft weglating van x_j weinig invloed op de grootte van de restkwadraatsom, terwijl deze invloed erg groot kan zijn als x_i niet in het model is opgenomen. Anderzijds kan het ook voorkomen dat een variabele alleen dan veel verklaart als één of meer andere variabelen in het model zijn opgenomen.

Ondanks (of beter: dank zij) deze moeilijkheden is er een aantal methoden ontwikkeld die, hiermee rekening houdend, toch tot een keuze van de "beste" regressievergelijking leiden.

De betrouwbaarste methode is die waarbij alle combinaties van wèl of niet opnemen van de x-variabelen in het model worden onderzocht. Deze methode werkt als volgt.

Als het volledige model p x-variabelen bevat, dan zijn er in totaal 2^p modellen mogelijk. (Elk van de variabelen kan namelijk wèl of niet in het model worden opgenomen). Van al deze modellen wordt de restkwadraat-som KS_r berekend.

Als men het aantal parameters vast neemt, zeg gelijk aan k, dan is de keuze van het beste model op grond van de kleinste $\hat{\sigma}_r^2$ eenduidig. Na eenmaal voor elke waarde van k het beste model te hebben gevonden, moet men uit al deze modellen een keuze maken. Dit komt neer op het zoeken van een compromis tussen:

1. maximalisatie van de voorspellende waarde van de regressiefunctie, hetgeen gerealiseerd wordt door het model met de kleinste rest-variantie. Doorgaans bevat dit model vrij veel variabelen.
2. minimalisatie van het aantal variabelen, bijvoorbeeld ten behoeve van besparing van meetkosten bij toekomstige toepassingen van het model.

Het is de taak van de onderzoeker om hiertussen een afweging te maken. In statistische zin kan echter het volgende worden gezegd over de keuze van het aantal variabelen. Als twee modellen "hiërarchisch" zijn (dat wil zeggen dat alle variabelen van het ene model ook deel uitmaken van het andere), dan kan men deze twee modellen vergelijken door een uitspraak te doen over het feit of de variantiereductie ten gevolge van de extra variabele(n) "significant" is dan wel op louter toeval berust. Deze uitspraak is gebaseerd op een F-toetsingsgrootte, gedefinieerd door

$$F_{n-p-1}^{k_2-k_1} = \frac{(KS_{r2} - KS_{r1}) / (k_2 - k_1)}{\hat{\sigma}_r^2}$$

waarbij k_1 en k_2 het aantal variabelen in de modellen voorstelt, ($k_1 < k_2$), KS_{r1} en KS_{r2} de bijbehorende restkwadraatssommen zijn en waarbij

$\hat{\sigma}_r^2$ de geschatte restvariantie van het volledige model is. (Afggeraden wordt voor $\hat{\sigma}_r^2$ de kleinst mogelijke restvariantie te nemen omdat deze σ^2 systematisch te laag schat).

Als voor oplopende waarden van k het beste model steeds het vorige omvat, dan is een vergelijking in statistische zin dus mogelijk.

Als de modellen niet hiërarchisch zijn, dan worden deze vaak op dezelfde wijze vergeleken, omdat ook dan de bovengenoemde "F-grootheid" een goede indicatie geeft of het grotere model al dan niet te prefereren is.

In de praktijk zal het vaak voorkomen dat alleen voor lage waarden van k een model met k variabelen significant beter is dan elk model met $k-1$ variabelen (in ons geval zal blijken dat dit alleen geldt voor $k = 1, 2$ en 3). Juist voor deze waarden van k zijn de beste modellen vaak wèl hiërarchisch zodat men dan over de belangrijkste variabelen ook in statistische zin een uitspraak kan doen.

Opmerking

De hierboven beschreven procedure, waarbij alle 2^p modellen worden doorgerekend, is nogal bewerkelijk, zodat voor de uitvoering ervan bij niet al te kleine waarden van p een computer onmisbaar is. Vooral in de tijd dat deze nog niet algemeen ter beschikking stond zijn er methoden ontwikkeld die niet alle modellen in beschouwing nemen. (Draper & Smith, 1966, pag 163 e.v.). De bekendste hiervan zijn "voorwaartse selectie", "achterwaartse eliminatie" en enige varianten hierop. Deze hebben echter alle het nadeel dat ze kunnen uitkomen op een model dat suboptimaal is (dat wil zeggen dat er een ander model bestaat met evenveel parameters maar met een kleinere restvariantie).

Een ander nadeel van deze minder bewerkelijke methoden is dat men bij de verschillende waarden van k geen overzicht krijgt van de andere modellen met evenveel variabelen, waarvan de restvariantie nauwelijks groter is dan die van het gekozen model. Sommige van deze suboptimale modellen kunnen voor de onderzoeker om andere redenen zeer aantrekkelijk zijn.

4. De statistische analyse

4.1. Verantwoording

De vraagstelling luidt: leiden de gemeten vegetatievariabelen tot een redelijke voorspelling van de vogelvariabelen en zo ja, welke vegetatievariabelen oefenen hierop de grootste invloed uit?

De eerste vraag kan beantwoord worden door univariaat, dat wil zeggen voor elke vogelvariabele apart, een regressie-analyse uit te voeren op de vegetatievariabelen (het volledige model). Wij vonden een waarde voor V^2 van circa 0.6 groot genoeg om over te gaan tot het beantwoorden van het tweede deel van de vraag. Aangezien de vegetatievariabelen gecorreleerd zijn (zie tabel 2, bijlage) gebruiken we hiervoor de in paragraaf 3.2. beschreven methode van selectie van variabelen. Hierbij is gebruik gemaakt van het computerprogrammapakket genaamd LINWOOD.

Bij de analyse kunnen nog de volgende kanttekeningen worden gemaakt.

1. Het verband wordt lineair in de vegetatievariabelen verondersteld. Het kleine aantal percelen laat niet toe om ook kwadraten, produkten of andere transformaties van vegetatievariabelen in het model op te nemen. Wel is in veel gevallen, door het grafisch uitzetten van vogelvariabele tegen vegetatievariabele, gecontroleerd of de vegetatievariabele niet beter getransformeerd zou kunnen worden (de HET-variabelen zijn op deze wijze gevonden). We zijn ons echter bewust van het feit dat het uitzetten van de y-variabele tegen de x-variabelen afzonderlijk een zeer bedriegelijk beeld van het te onderzoeken verband kan geven (Daniel & Wood, 1971, pag. 53). Daarom hebben we de analyse nog tweemaal uitgevoerd voor transformaties van de HET-variabelen, namelijk door $\log(N_{i+} \cdot L_{i+} \cdot L_{i-} + 1)$ te vervangen door $N_{i+} \cdot L_{i+} \cdot L_{i-1}$ respectievelijk $\sqrt{N_{i+} \cdot L_{i+} \cdot L_{i-} + 1}$. Deze modellen vertoonden echter lagere waarden van V^2 , zodat hierop de resultaten van deze analyses niet verder zullen worden besproken. Niettemin zou het interessant zijn te onderzoeken of er nog betere heterogeniteitsmaten bestaan.
2. Uiteraard zullen er voor de vogelparameters ruimtelijke afhankelijkheden bestaan tussen de verschillende percelen. De betere voedselplaatsen gaan bijvoorbeeld gepaard met een hogere reproductie aldaar,

hetgeen een uitstralingseffect geeft naar de naburige percelen. Bij de in dit rapport beschreven analyses is met deze ruimtelijke afhankelijkheid geen rekening gehouden.

3. De door ons uitgevoerde correctie van S voor de oppervlakte zou ook op een andere wijze kunnen worden gerealiseerd, namelijk door meteen het volgende niet-lineaire model aan te passen (waarin z een extra parameter vormt):

$$S = \text{opp}^z \cdot (\beta_0 + \beta_1 \cdot \bar{0} + \beta_2 \cdot \text{VCO} + \dots + \beta_8 \cdot \text{HET}_4)$$

Aanpassing van dit niet-lineaire model is echter minder eenvoudig uit te voeren en voor de selectie van variabelen is geen computerprogramma beschikbaar. De thans gevolgde werkwijze vonden we zeker niet minder inzichtelijk. Een mogelijk alternatief, dat wij echter niet hebben beproefd, lijkt het aanpassen van een lineair model in log S:

$$\log S = \alpha_0 + \alpha_1 \cdot \log(\text{opp}) + \alpha_2 \cdot \log \bar{0} + \dots + \alpha_9 \log(\text{HET}_4)$$

hetgeen neerkomt op:

$$S = e^{\alpha_0} \cdot \text{opp}^{\alpha_1} \cdot \bar{0}^{\alpha_2} \cdot \dots \cdot \text{HET}_4^{\alpha_9}$$

4. In het beginstadium is overwogen "canonieke correlatieanalyse" toe te passen op alle vogelvariabelen en vegetatievariabelen tezamen. Het feit dat er dan bijna evenveel waarnemingen als variabelen in het spel zouden zijn en het feit dat de vogelvariabelen niet alle van even groot belang zijn, heeft ons hiervan weerhouden. Voorts is de interpretatie van de uitkomsten van zo'n analyse meestal niet eenvoudig. (Kendall & Stuart, vol. III, 1968, pag. 305).

4.2. Resultaten

In deze paragraaf komen aan de orde de resultaten van de analyse zoals genoemd in paragraaf 4.1. Voor D, S en D_{ho1} bestaat het volledige model uit VCO, TOTBED, H_B H_S, HET₁, HET₂, HET₃ en HET₄. Voor D_{stam} is

daarnaast ook de gemiddelde omtrek \bar{O} opgenomen. Voor de met betrekking tot de oppervlakte gecorrigeerde S is de analyse uitgevoerd voor de volgende waarden van z: 0.10, 0.125, 0.15, 0.175, 0.20, 0.225 en 0.25. De resultaten voor z = 0.15 en z = 0.20 zijn hier weergegeven. De overige waarden van z gaven gelijklopende resultaten, zodat men mag stellen dat voor z tussen 0.10 en 0.25 de conclusie niet afhangt van de bij de correctie gebruikte waarde van z.

Voor elke vogelvariabele apart wordt in een tabel voor de relevante waarden van k (= het aantal variabelen) het "beste" model gegeven en ook de eventuele modellen waarvan de restvariantie nauwelijks groter is. Een model met k variabelen noemen we significant beter dan alle modellen met k-1 variabelen als $F_8^1 > 5.32$ en als $F_7^1 > 5.59$ (bij $\alpha = 0.05$).

De dichtheid D

k	model	KS_r	$\hat{\sigma}_r^2$	v^2	F_8^1
0		64350	4022		
1	HET ₄	25146	1676	.58	29.4
2	HET ₄ , H _S	19844	1418	.65	3.97
	HET ₄ , H _B	19950	1425	.65	3.90
3	HET ₄ , H _S , TOTBED	15020	1155	.71	3.62
	HET ₄ , H _B , HET ₂	15080	1160	.71	3.57
4	HET ₄ , H _S , H _B , HET ₂	11416	951	.76	3.27
8	volledig model	10672	1334	.67	

S/opp¹⁵

k	model	KS _r	$\hat{\sigma}_r^2$	v ²	F ₈ ¹
0		243.5	15.22		
1	HET ₄	110.8	7.39	.51	18.3
2	HET ₄ , H _S	88.8	6.34	.58	3.04
3	HET ₄ , H _S , TOTBED	78.4	6.03	.60	1.43
	H _S , H _B , HET ₁	79.4	6.11	.60	1.29
4	HET ₄ , H _S , H _B , HET ₁	61.4	5.12	.66	2.34
8	volledig model	58.1	7.26	.52	

S/opp²⁰

k	model	KS _r	$\hat{\sigma}_r^2$	v ²	F ₈ ¹
0		213.7	13.35		
1	HET ₄	90.5	6.03	.55	21.9
2	HET ₄ , H _S	69.6	4.97	.63	3.69
3	HET ₄ , H _S , TOTBED	60.7	4.67	.65	1.59
	H _S , H _B , HET ₁	63.6	4.89	.63	1.07
4	HET ₄ , H _S , H _B , HET ₁	47.8	3.98	.70	2.29
8	volledig model	45.1	5.63	.58	

D_{hol}

k	model	KS _r	$\hat{\sigma}_r^2$	v ²	F ₈ ¹
0		30261	1891		
1	HET ₄	16987	1132	.40	15.65
2	HET ₄ , VCO	12043	860	.55	5.83
3	HET ₄ , VCO, HET ₁	10667	820	.57	1.62
	HET ₄ , VCO, HET ₃	10796	830	.56	1.47
4	HET ₄ , VCO, HET ₁ , HET ₃	8724	727	.62	2.29
8	volledig model	6786	848	.55	

D_{stam}

k	model	KS _r	$\hat{\sigma}_r^2$	v ²	F ₇ ¹
0		975.3	61.0		
1	$\bar{0}$	674.8	45.0	.26	11.6
2	$\bar{0}$, TOTBED	547.5	39.1	.36	4.91
	$\bar{0}$, VCO	550.5	39.3	.35	4.79
3	$\bar{0}$, VCO, H _G	401.0	30.8	.49	5.66
4	$\bar{0}$, VCO, H _B , HET ₁	347.0	28.9	.53	2.08
5	$\bar{0}$, VCO, H _G , H _B , HET ₁	274.0	24.9	.59	2.82
9	volledig model	181.5	25.9	.57	

Als men in de modellen alleen de significante (of eventueel bijna significante) variabelen opneemt, dan vindt men de volgende schattingen van de regressievergelijkingen:

$$\begin{aligned}
 D &= 53.8 + 67.5 \text{ HET}_4 \\
 S/\text{opp} \cdot 15 &= 15.16 + 3.93 \text{ HET}_4 \\
 S/\text{opp} \cdot 20 &= 13.91 + 3.78 \text{ HET}_4 \\
 D_{\text{hol}} &= -23.2 + 35.4 \text{ HET}_4 + .94 \text{ VCO} \\
 D_{\text{stam}} &= 1.9 + .21 \bar{0} \\
 & (= -26.5 + .39 \bar{0} + .26 \text{ VCO} - 3.0 \text{ H}_G)
 \end{aligned}$$

De standaarddeviatie van de regressiecoëfficiënten bedraagt meestal 20 à 40% van de schatting ervan. Als men meer variabelen aan het model toevoegt, dan worden de schattingen van de regressiecoëfficiënten minder nauwkeurig als gevolg van de afhankelijkheden tussen de variabelen.

4.3. Conclusies

- De dichtheid D hangt sterk samen met HET₄. Verder is er een aantal factoren (H_G, H_B, HET₂ en eventueel TOTBED) die ook enige samenhang met D vertonen maar niet significant zijn.
- Het soortenaantal S toont een sterk verband met HET₄ en ook enigszins, hoewel niet significant, met H_G.

- Ook D_{hol} hangt sterk samen met HET_4 en vertoont daarnaast een significant verband met VCO.
- D_{stam} hangt sterk samen met \bar{O} . De overige factoren geven gezamenlijk nog een vrij grote variantiereductie. Hiervan zijn VCO en H_3 nagenoeg significant.

5. Slotopmerkingen

- 1) In andere onderzoeken zijn veel sterkere verbanden gevonden tussen vegetatiestructuur en vogelvariabelen (onder andere MacArthur & MacArthur, 1961). Dit is vermoedelijk een gevolg van het feit dat in die onderzoeken ook andere vegetatietypen zoals heide, weiland en jonge bossen werden betrokken. Daar werd een veel breder traject onderzocht, hetgeen meestal sterkere verbanden oplevert. In het onderhavige onderzoek werd echter gepoogd de variatie binnen één vegetatietype, namelijk de oudere bossen te verklaren. Het was niet te verwachten dat binnen de typen de relatie gelijk is aan de relatie die over het bredere traject geldt.
- 2) Wellicht ten overvloede wordt opgemerkt dat hier onderzocht is hoe de vogelvariabelen samenhangen met de gemeten waarden van de vegetatievariabelen en niet de samenhang met de populatiewaarden (zoals bijvoorbeeld de werkelijke gemiddelde omtrek van alle bomen in het perceel). Overigens is het voor de HET-variabelen slechts mogelijk om een operationele definitie van de populatiewaarde te geven, namelijk de verwachting van de uitkomsten bij deze meetprocedure.

Literatuurverwijzingen

- Daniel, C. and F.S. Wood, (1971), Fitting equations to data. Wiley, New York.
- Draper, N.R. and H. Smith, (1966), Applied regression analysis. Wiley, New York.

Kendall, M.G. and A. Stuart, (1968), The advanced theory of statistics.

Vol. 3, 2nd edition, Griffin, London.

MacArthur, R.H. and J.W. MacArthur, (1961), On bird species diversity.

Ecology 42, 594-598.

Opdam, P.F.M. en G.J. van Bladeren, (1981), De vogelbevolking van
beheerde en onbeheerde delen van het Forstamt Hasbruch.(Oldenburgerland,
BRD) in relatie tot de bosstructuur. RIN-rapport 81-2, Leersum.

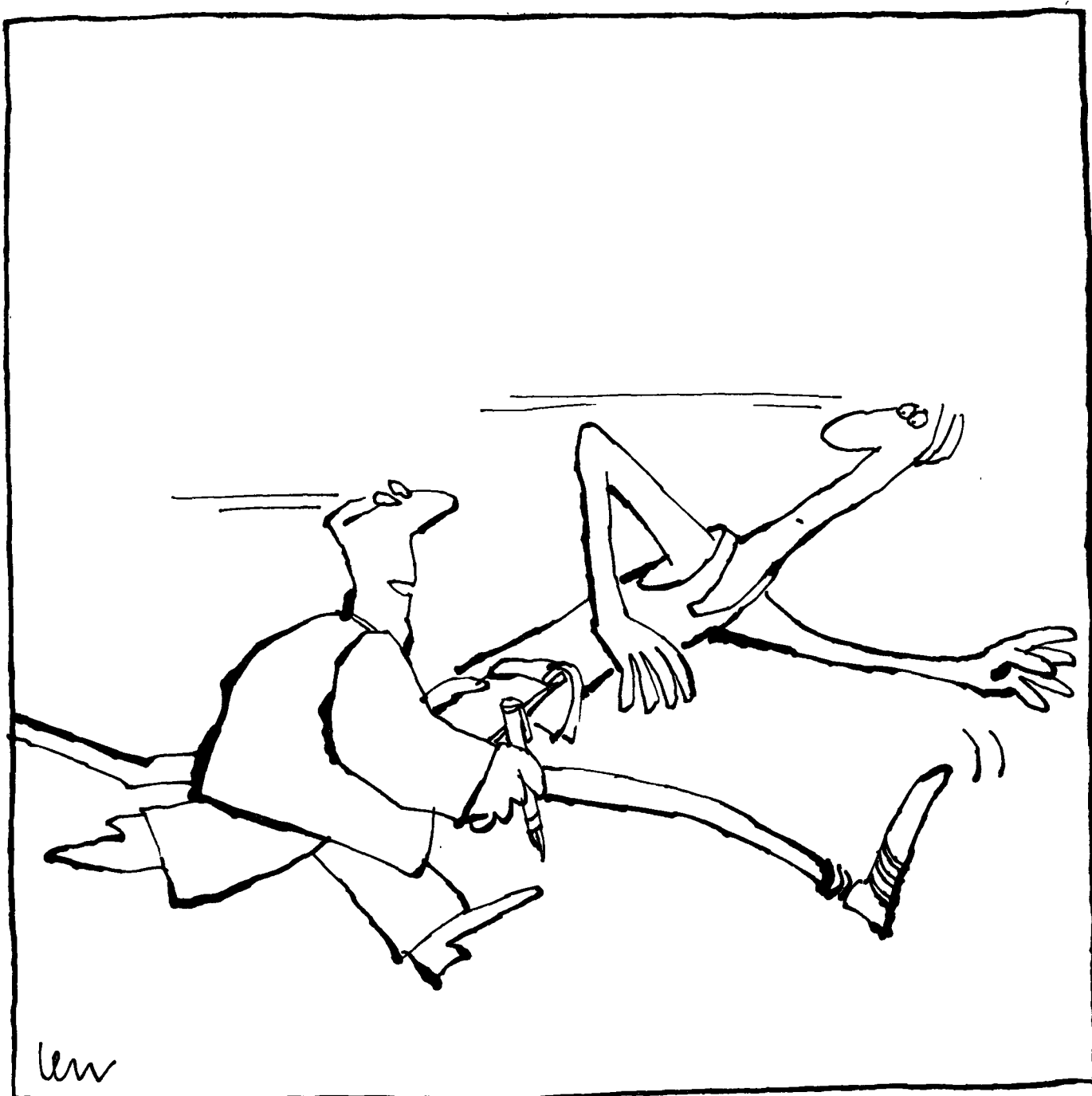
perceel nr.	opp	D	S	D _{hol}	D _{stam}	\bar{O}	VCO	TOT- BED	H _B	H _S	HET ₁	HET ₂	HET ₃	HET ₄
35	3.76	55.9	16	23.0	7.1	44.2	87.8	152	0.99	1.87	1.71	2.18	1.56	0.00
9	3.03	79.2	17	19.8	2.2	51.1	56.0	242	1.25	0.47	1.69	2.05	1.91	0.00
23	3.05	73.1	16	28.4	12.0	99.8	41.7	152	0.85	0.99	0.00	1.56	2.22	0.85
29	5.28	62.5	21	43.0	7.6	94.1	45.1	136	0.88	2.38	0.48	1.99	1.76	1.23
32	5.27	75.9	23	37.3	12.0	75.8	51.7	160	0.98	1.93	0.48	2.05	2.26	1.00
34	2.10	114.3	21	68.2	22.2	97.3	43.7	158	1.05	1.08	1.00	2.26	2.38	1.23
33	3.71	122.4	27	68.4	21.5	97.2	54.6	182	1.20	3.38	1.41	2.30	2.38	1.85
14	12.61	162.2	32	75.1	20.7	93.3	43.1	166	0.87	1.91	0.00	1.40	2.21	1.28
31	3.42	137.3	23	118.9	28.2	123.3	45.2	166	0.87	2.31	0.00	1.76	2.21	0.70
19	9.64	71.6	26	49.7	11.7	88.4	52.8	168	0.98	2.38	0.48	1.93	2.38	0.00
4	6.22	111.5	30	35.3	22.4	89.8	38.1	222	1.01	2.30	0.00	2.26	1.23	1.00
17	6.36	207.0	31	111.1	22.0	115.2	55.7	168	1.21	3.66	1.61	2.10	2.42	2.16
28	10.78	139.8	31	111.6	19.2	67.3	95.2	264	1.27	2.08	1.93	2.18	1.74	1.32
12	2.87	212.5	28	162.6	29.1	96.8	90.4	234	1.28	2.25	1.81	2.21	2.16	1.63
11	3.51	189.0	24	143.4	22.7	78.8	84.9	248	1.36	2.06	2.22	1.96	2.08	2.26
18	3.06	222.2	31	86.1	14.2	111.2	52.8	180	1.26	3.44	1.96	2.05	2.25	2.04
13	2.17	256.5	27	98.3	9.2	97.6	72.1	224	1.36	5.90	2.30	1.96	2.38	1.86

Tabel 1 : De in de 17 percelen waargenomen waarden van de vogelvariabelen en de vegetatievariabelen.

Onderzoek naar het effect van vitaminesuppletie op het prestatievermogen van CIOS-studenten

Ir. J. Th. N. M. Thissen

Dit verslag is geschreven ten behoeve van het Instituut CIVO-Toxicologie en Voeding van de Hoofdgroep Voeding en Voedingsmiddelen TNO



Inhoud

1.	Inleiding en doel van het onderzoek	167
2.	Opzet van het onderzoek	167
3.	Analyse	169
3.1.	<i>Gevolgen voor de analyse door het meten vooraf</i>	169
3.2.	<i>Covariantie-analyse</i>	170
3.3.	<i>Uitgebreidere covariantie-analyse</i>	173
4.	Resultaten	174
4.1.	<i>Resultaten met behulp van covariantie-analyse</i>	174
4.2.	<i>Resultaten met behulp van uitgebreidere covariantie-analyse</i>	176
	Tabel 1	175
	Tabel 2	178
5.	Conclusies	179
	Bijlage 1 t/m 3	

1. Inleiding en doel van het onderzoek

"Is het gewenst en mogelijk de voeding van de Nederlandse sportman/vrouw te verbeteren door meer vitamine en ijzerrijke produkten erin op te nemen of preparaten te verstrekken? Suppletie zou een mogelijk verhoogde behoefte kunnen dekken en zou wellicht tevens een hogere sportprestatie mogelijk kunnen maken".

Zo luidde de vraag die aanleiding gaf tot dit onderzoek.

Voor de uitvoering van het onderzoek heeft men zich gewend tot jongens en meisjes op de CIOS-instituten te Overveen en Arnhem. Aanvankelijk was het de bedoeling om uit de tweede klas van elk der instituten ongeveer 50 vrijwilligers te werven. Het aantal vrijwilligers bleek echter dermate klein, dat besloten is ook eerstejaars bij het onderzoek te betrekken.

2. Opzet van het onderzoek

Bij het begin van het onderzoek waren vier groepen beschikbaar:

Sexe	CIOS	Jaar	Aantal
Jongens	Arnhem	1	21
	Overveen	2	16
Meisjes	Arnhem	1	19
	Arnhem	2	27
		totaal	83

Om de in paragraaf 1 genoemde vraagstelling te kunnen beantwoorden werd elk der vier groepen aselekt verdeeld in twee helften, waarbij de ene helft tijdens de onderzoeksperiode een vitaminepil slikte en de andere helft een placebo.

Er werden metingen verricht voordat het onderzoek van start ging en aan het einde van de onderzoeksperiode, die duurde van half november 1979 tot half maart 1980, een periode van ongeveer 16 weken.

De metingen, voorafgaande aan het onderzoek, waren bedoeld om het onderscheidingsvermogen van de toets op het bestaan van een effect van het gebruik van een vitaminepil te vergroten. In paragraaf 3 zullen we hier nader op ingaan.

De volgende variabelen zijn gemeten:

Antropometrie

- ① lengte
- ② gewicht
- 3 kniebreedte
- 4 polsbreedte
- 5 schouderbreedte breedtematen
- 6 borstbreedte
- 7 bekkenbreedte
- 8 armomtrek
- 9 borstomtrek
- 10 buikomtrek
- 11 bekkenomtrek omtrekmaten
- 12 bovenbeenomtrek
- 13 kuitomtrek
- ⑭ huidploidikte biceps
- ⑮ huidploidikte triceps
- ⑯ huidploidikte scapula
- ⑰ huidploidikte crista
- 18 huidploidikte subcostaal
- 19 huidploidikte para-umbilicaal

Inspanningsfysiologie ⑳ VO_2 max in l/min

21 Cooper-test

Bloeddruk

22 fase I

23 fase IV

24 fase V

Biochemie

- 25 Hb
- 26 Ht
- 27 ery
- 28 serumijzer
- 29 TYBC
- 30 ijzerverzadigingspercentage
- 31 ferritine
- 32 α -ETK
- 33 α -EGR
- 34 tot. cholesterol
- 35 HDL-cholesterol

Uit deze variabelen werden de volgende nieuwe variabelen berekend met behulp van de tussen haakjes genoemde variabelen:

- 36 gecorrigeerde armomtrek (variabelen 8 en 15)
- 37 som van vier huidploidikten (variabelen 14 t/m 17)
- 38 vetpercentage (variabele 37)
- 39 vetmassa (variabelen 2 en 38)
- 40 vetvrije massa (variabelen 2 en 38)
- 41 VO_2 max in $\frac{ml}{min/kg}$ (variabelen 2 en 20)
- 42 VO_2 max in $\frac{ml}{min/VVM}$ (variabelen 20 en 40)
- 43 MCV (variabelen 26 en 27)
- 44 MCH (variabelen 25 en 27)
- 45 MCHC (variabelen 25 en 26)
- 46 gecorrigeerde armdiameter (variabelen 8 en 15)

De omcirkelde variabelen zijn geanalyseerd. Vanwege de scheefheid van de verdeling werd de variabele ferritine logaritmisches getransformeerd.

3. Analyse

3.1. Gevolgen voor de analyse door het meten vooraf

De gegevens zijn per (in de vorige paragraaf omcirkelde) variabele geanalyseerd voor jongens en meisjes apart.

Door het uitvallen van enkele jongens en meisjes bleven uiteindelijk de volgende groepen over:

	Suppletie	Placebo	Pil	Totaal
Sexe	Jaar			
Jongens	1	9	9	18
	2	8	7	15
Meisjes	1	8	10	18
	2	12	14	26

Indien we alleen de beschikking zouden hebben over de metingen op onderzoekdatum 2 (aan het einde van het onderzoek) zou de analyse per sexe bestaan uit een variantieanalyse met twee criteria, namelijk de factoren jaar (1 en 2) en suppletie (placebo en pil).

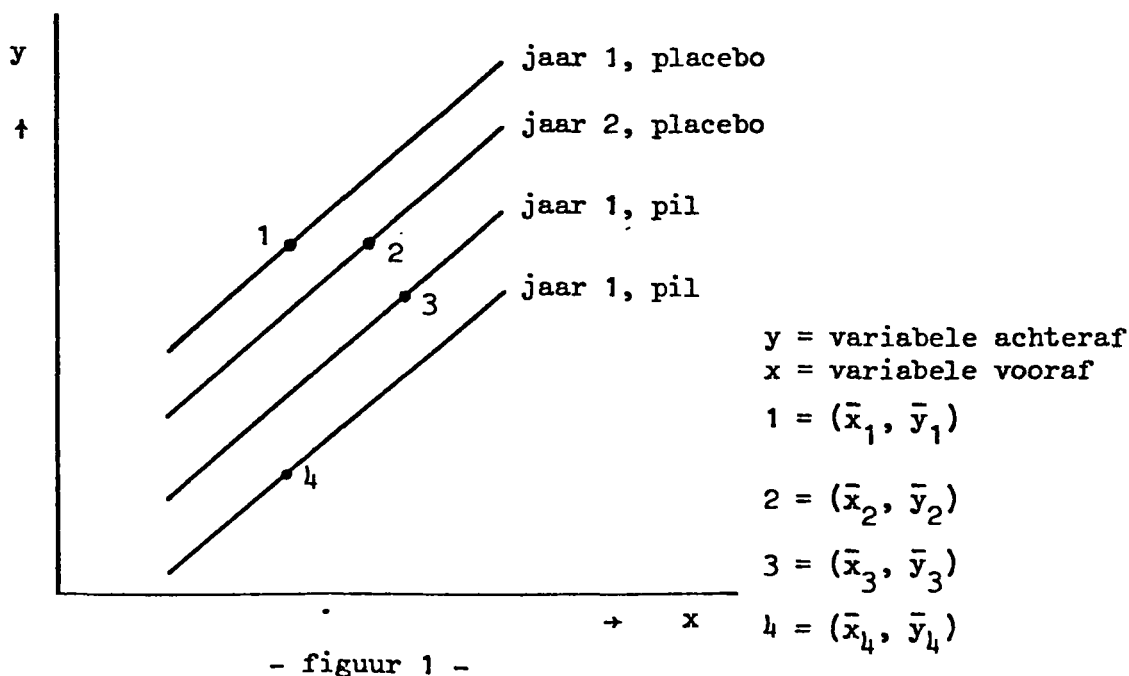
In dat geval zou in de restvariantie, die gebruikt wordt om een mogelijk suppletie-effect te toetsen, ook de variabiliteit tussen personen opgenomen zijn.

Doordat nu vóór het onderzoek, op onderzoekdatum 1, dezelfde variabelen gemeten zijn, is het mogelijk een deel van de tussenpersoonsvariantie uit de restvariantie te elimineren. Dit kan gebeuren met behulp van een techniek die bekend staat onder de naam covariantieanalyse. Zie hiervoor paragraaf 3.2.

Een tweede voordeel van het meten van de variabelen op onderzoekdatum 1 is: het kunnen nagaan of een mogelijk suppletie-effect afhankelijk is van het niveau van de meting. Dat wil zeggen dat voor mensen met een lage waarde voor bijvoorbeeld de VO_2 max, het suppletie-effect anders is dan voor mensen met een hoge waarde voor de VO_2 max. Hierop komen we terug in paragraaf 3.3.

3.2. Covariantie-analyse

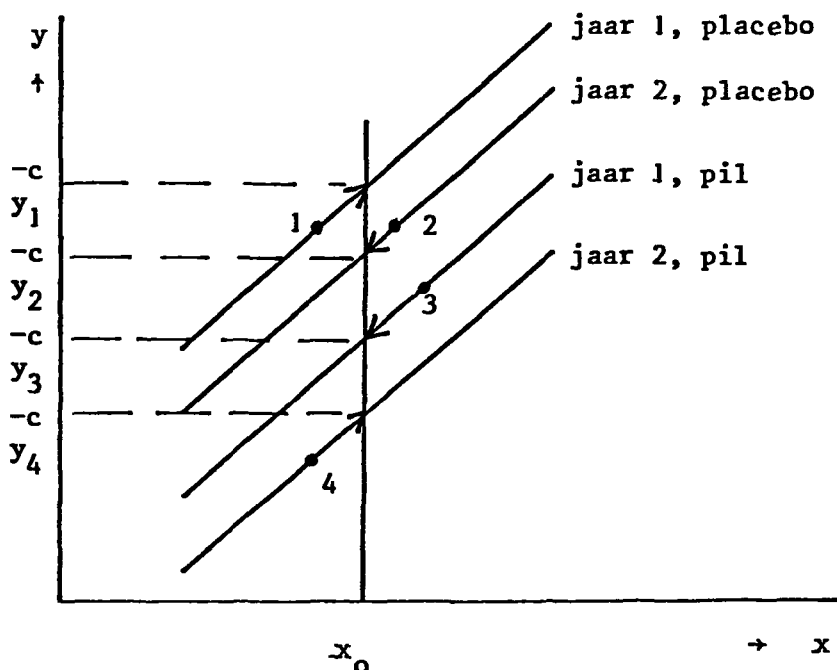
Het eenvoudigste covariantieanalyse-model staat beschreven in bijlage 1. Dit model past per sexe voor de vier groepen (combinaties van jaar en suppletie) lijnen aan met dezelfde richtingscoëfficiënt. In figuur 1 hebben we geprobeerd dit weer te geven.



Met bijvoorbeeld \bar{x}_1 wordt bedoeld het werkelijk gemeten gemiddelde van de variabele vooraf voor de groep "jaar 1, placebo" en met bijvoorbeeld \bar{y}_4 het werkelijk gemeten gemiddelde van de variabele achteraf voor de groep "jaar 2, pil".

Indien het lineaire verband tussen y en x echt bestaat, dan kunnen de gemiddelden achteraf (\bar{y}_1 t/m \bar{y}_4) scherper met elkaar worden vergeleken door rekening te houden met het verschil in de gemiddelden vooraf (\bar{x}_1 t/m \bar{x}_4).

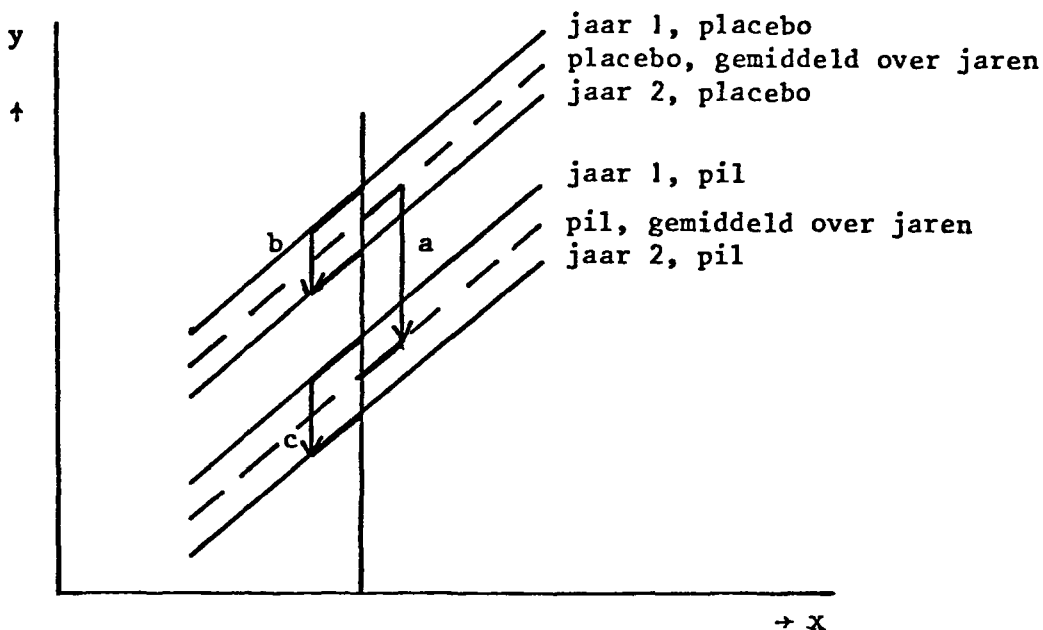
Met behulp van de berekende richtingscoëfficiënt corrigeert de covariantieanalyse als het ware de werkelijk berekende gemiddelden achteraf (\bar{y}_1 t/m \bar{y}_4) door bij een vaste waarde van x , zeg x_0 , nieuwe gemiddelden y , zeg \bar{y}^c , te berekenen. De keuze van x_0 is in wezen arbitrair, maar doorgaans wordt hiervoor het gemiddelde van alle x 'en genomen. In figuur 2 hebben we dit weergegeven.



- figuur 2 -

De nieuwe gemiddelden \bar{y}^c worden nu gebruikt om de verschillende effecten te toetsen. Het suppletie-effect bijvoorbeeld is de mate waarin de bewuste variabele verandert, als een persoon gesuppleerd wordt die afkomstig is uit een met de onderzochte groep CIOS leerlingen te vergelijken populatie.

De grootte van het suppletie-effect is in figuur 3 aangegeven met behulp van de lengte van pijl a, terwijl de grootte van de interactie jaar x suppletie berekend kan worden door de lengten van de pijlen b en c van elkaar af te trekken met inachtneming van de richting van de pijlen.

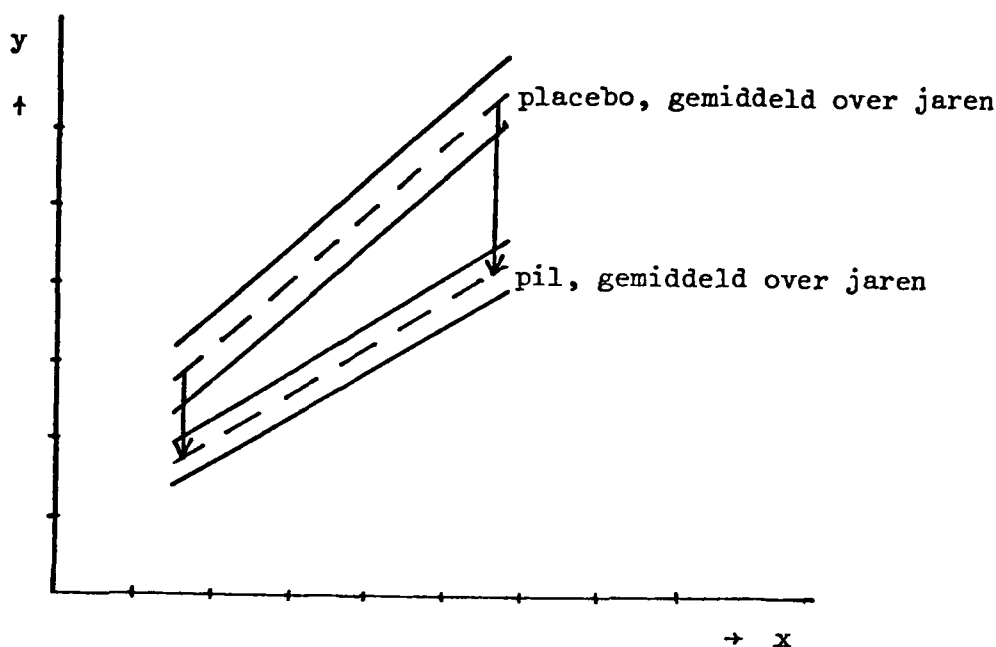


- figuur 3 -

3.3. Uitgebreidere covariantie-analyse

In paragraaf 3.1. is als mogelijk voordeel van het meten vooraf genoemd het kunnen nagaan of een mogelijk suppletie-effect afhankelijk is van het niveau van de meting. Bij de analyse in paragraaf 3.2. is hiermee geen rekening gehouden. We hebben namelijk verondersteld dat de richtingscoëfficiënt γ voor de vier combinaties dezelfde was.

Nemen we deze mogelijke afhankelijkheid tussen het suppletie-effect en het niveau van de meting (de waarde van de variabele vooraf) in het model op, dan moeten we γ vervangen door γ_i , $i=1,2$. We veronderstellen dan dat voor de placebogroepen van jaar 1 en jaar 2 richtingscoëfficiënt γ_1 geldt en voor de pilgroepen van jaar 1 en jaar 2 richtingscoëfficiënt γ_2 . We hebben een voorbeeld weergegeven in figuur 4.



- figuur 4 -

In deze figuur is te zien dat het suppletie-effect bij lage waarden van x ongeveer één eenheid y bedraagt (lengte pijl a), terwijl dit effect bij hoge waarden van x ongeveer twee eenheden y bedraagt (lengte pijl b). Door te toetsen of γ_1 gelijk is aan γ_2 kunnen we besluiten of we in plaats van het eenvoudige covariantieanalyse-model het uitgebreidere model gebruiken voor de beschrijving van de waarnemingen.

4. Resultaten

4.1. Resultaten met behulp van covariantie-analyse

In paragraaf 3.1. zijn de aantallen jongens en meisjes vermeld, die uiteindelijk meegedaan hebben in de analyse. ¹⁾

In tabel 1 (zie volgende pagina) zijn door middel van symbolen overschrijdingskansen gegeven bij het toetsen op het al dan niet aanwezig zijn van de verschillende effecten. Er is ook een kolom opgenomen voor het effect van de covariabele, de meting vooraf. Is de overschrijdingskans voor dit effect erg klein, dan wil dit zeggen dat het voor de bewuste variabele zinvol is geweest de extra meting vooraf uit te voeren.

Uit de overschrijdingskansen blijkt de invloed van suppletie slechts op een gering aantal variabelen. Deze variabelen, TYBC, log ferritine, α -ETK, α -EGR, MCV en MCHC, bekijken we in bijlage 2 voor zowel jongens als meisjes, ook als de variabele voor slechts één sexe significant bleek. Tevens bekijken we in deze bijlage de drie variabelen VO_2 max achtereenvolgens in l/min, ml/min/kg en ml/min/VVM, alhoewel het suppletie-effect niet significant is.

In deze bijlage 2 zijn per sexe en per variabele de "metingen achteraf" uitgezet tegen de "metingen vooraf" in vier aparte grafieken:

linksboven: jaar 1, placebo	rechtsboven: jaar 2, placebo
linksonder: jaar 2, pil	rechtsonder: jaar 2, pil

Verder zijn in deze grafieken de lijnen aangegeven die gevonden zijn door het covariantieanalyse-model van bijlage 1 te gebruiken. Ook zijn de gecorrigeerde gemiddelden aangegeven, waarmee de effecten van jaar, suppletie en de interactie jaar x suppletie berekend zijn.

¹⁾ Het schema is voor zowel jongens als meisjes niet geheel orthogonaal, wat gevolgen kan hebben voor de manier van toetsen van de effecten. Aangezien echter de inorthogonaliteit in dit geval klein is, maakt het niet veel uit welke manier gekozen wordt. De volgende keus is gedaan:
 jaareffect niet gecorrigeerd,
 suppletie-effect gecorrigeerd voor jaareffect,
 interactie jaar x suppletie gecorrigeerd voor jaar- en suppletie-effect.

Tabel 1 Overschrijdingskansen behorende bij het toetsen van de verschillende effecten van het eenvoudige covariantieanalyse-model; voor de betekenis van de symbolen zie onderaan de tabel.

	meisjes				covariabele	jongens			
	jaar	suppletie	jaar x suppletie	effect		jaar	suppletie	jaar x suppletie	effect
lengte	-	-	-	***		-	0	-	***
gewicht	-	-	0	***		-	-	-	***
gecorr.armontrek	-	-	-	***		-	-	-	***
som 4 hpl	**	-	-	***		-	-	-	***
vetpercentage	*	-	0	***		-	-	-	***
vetmassa	*	-	0	***		-	-	-	***
vetvrije massa	***	-	-	***		-	-	-	***
biceps	***	-	-	***		-	-	-	***
triceps	-	-	-	***		-	-	-	***
scapula	**	-	-	***		-	-	-	***
crista	*	-	-	***		-	-	-	***
VO ₂ max l/min	-	-	-	***		-	-	-	***
VO ₂ max ml/min/kg	*	-	-	***		-	-	-	**
VO ₂ max ml/min/VVM	-	-	-	***		-	-	-	***
Hb	***	-	-	***		-	-	-	**
Ht	-	-	-	***		-	-	-	***
ery	**	-	-	***	*	-	-	-	**
serumijzer	-	-	-	*	**	-	-	-	-
TYBC	0	**	-	***	**	-	-	-	***
ijzerverzadigingsperc.	-	-	-	*	*	-	-	-	-
log ferritine	-	***	-	***		-	-	-	***
α-ETK	-	***	-	0		-	**	-	-
α-EGR	-	***	**	**		-	***	-	*
MCV	**	0	-	***		-	0	-	**
MCH	-	-	-	***		-	-	-	**
MCHC	*	-	*	-		-	***	-	***
gecorr.armdiam.	-	-	-	***		-	-	-	***
tot.cholesterol	-	-	-	***		-	-	-	***
HDL-cholesterol	*	-	-	***		-	-	-	***

- 10% < p
0 5% < p ≤ 10%
* 1% < p ≤ 5%
** 0.1% < p ≤ 1%
*** p < 0.1%

Deze effecten zijn voor alle onderzochte variabelen weergegeven in tabel 2 (zie pagina 178). Het jaareffect is het verschil tussen jaar 2 en jaar 1, het suppletie-effect het verschil tussen de pilgroepen en de placebo-groepen en het interactie-effect het verschil tussen het suppletie-effect in jaar 2 en dat in jaar 1. Tevens is vermeld de standaardafwijking s_e , een maat voor de spreiding van een punt rond de lijn.

4.2. Resultaten met behulp van uitgebreidere covariantie-analyse

Zoals vermeld in paragraaf 3.3., is het mogelijk te toetsen of een uitgebreider covariantieanalyse-model beter past bij de waarnemingen dan het covariantieanalyse-model van bijlage 1.

We hebben dit voor alle onderzochte variabelen gedaan met de volgende resultaten (toetsingen uitgevoerd met een onbetrouwbaarheid van 5%):

- bij jongens een significante modelverbetering voor de variabelen triceps, log ferritine, α -EGR, MCV en MCH;
- bij meisjes een significante modelverbetering voor de variabelen som 4 hpl, vetvrije massa, crista, $VO_2 \max \frac{ml}{min/VVM}$ en α -ETK

Wat als eerste opvalt bij deze resultaten is dat geen enkele variabele zowel bij jongens als bij meisjes voorkomt. Bekijken we de overschrijdingskansen en de grafieken van de hier vermelde variabelen bij jongens en bij meisjes, dan blijkt alleen de variabele α -EGR voor beide sexen consequent hetzelfde te reageren (de overschrijdingskans voor de variabele α -EGR bij meisjes was circa 10%). Dit is de reden dat we alleen voor α -EGR het uitgebreidere model presenteren.

De grafieken uit bijlage 2 voor α -EGR zijn nogmaals weergegeven in bijlage 3 met dit verschil dat in bijlage 3 voor de placebo- en pilgroepen lijnen met verschillende richtingscoëfficiënten zijn getrokken.

De vergelijkingen van de diverse lijnen, inclusief de vermelding van bijbehorende standaardafwijkingen, zijn

α -EGR jongens	: y =	a +	b	x	s_b	s_e
jaar 1, placebo	: y =	-0.059	+ 1.203	x	.323	.088
jaar 2, placebo	: y =	-0.179	+ 1.203	x	.323	.088
jaar 1, pil	: y =	0.781	+ 0.170	x	.226	.088
jaar 2, pil	: y =	0.835	+ 0.170	x	.226	.088

 α -EGR meisjes

jaar 1, placebo	: y =	0.388	+ 0.716	x	.225	.089
jaar 2, placebo	: y =	0.432	+ 0.716	x	.225	.089
jaar 1, pil	: y =	0.676	+ 0.259	x	.175	.089
jaar 2, pil	: y =	0.621	+ 0.259	x	.175	.089

Uit de grafieken en de vermelde regressielijn blijkt dat voor zowel jongens als meisjes met een α -EGR vooraf rond de 0.8 het suppletie-effect in de buurt ligt van 0.1, terwijl voor jongens en meisjes met een α -EGR vooraf rond de 1.4 het suppletie-effect een orde van grootte van 0.4 heeft.

Tabel 2 Grootte van de effecten en de standaardafwijking van het eenvoudige covariantieanalyse-model.

	meisjes				jongens			
	jaar	suppletie	jaar x suppletie	s _e	jaar	suppletie	jaar x suppletie	s _e
lengte (cm)	-0.05	0.03	-0.10	0.51	-0.19	0.41	-0.35	0.57
gewicht (kg)	0.20	0.86	-2.06	1.81	0.10	-0.20	0.02	2.15
gecorr.armontrek (cm)	-0.12	0.08	0.11	0.54	-0.10	0.23	-0.83	0.79
som 4 hpl (mm)	6.0	2.5	-7.0	6.9	0.3	-0.1	-3.5	4.4
vetpercentage (%)	1.63	0.81	-2.29	2.04	0.08	0.10	-1.24	1.55
vetmassa (kg)	1.19	0.74	-1.79	1.64	0.05	0.14	-1.14	1.39
vetvrije massa (kg)	-0.96	0.19	0.06	0.85	0.2	-0.3	1.0	1.70
biceps (mm)	1.37	0.27	-0.78	1.06	0.01	-0.03	-0.15	0.40
triceps (mm)	0.80	0.75	-2.41	2.59	-0.34	-0.23	-0.13	1.14
scapula (mm)	1.22	0.09	-1.12	1.28	-0.07	-0.21	-0.35	0.76
crista (mm)	2.67	1.57	-2.75	3.8	0.71	0.48	-2.18	3.5
VO ₂ max (l/min)	-0.099	0.045	-0.069	0.20	0.076	0.011	0.070	0.36
VO ₂ max (ml/min/kg)	-1.86	0.14	0.06	2.78	0.8	1.1	-1.8	5.2
VO ₂ max (ml/min/kg VVM)	-1.3	0.9	-1.3	4.1	1.1	1.2	-2.2	5.5
Hb (mmol/l)	-0.41	-0.17	-0.04	0.34	-0.18	0.17	-0.17	0.44
Ht (%)	-0.96	-0.89	-1.35	1.78	-0.59	-0.15	0.13	1.82
ery (10 ¹² /l)	-0.22	-0.02	-0.00	0.22	-0.16	0.06	-0.20	0.21
serumijzer (μmol/l)	1.0	-0.1	5.5	7.7	-6.4	1.1	-2.8	5.7
TYBC (μmol/l)	-2.8	-4.5	-1.0	4.8	-4.5	-2.3	-2.9	4.3
ijzerverzadigingsperc. (%)	4.0	2.7	11.1	12.7	-8.2	3.8	-3.1	9.1
log ferritine (mol/l)	0.11	0.41	0.16	0.29	-0.05	0.16	0.31	0.33
α-ETK	0.027	-0.146	-0.071	0.067	0.039	-0.113	-0.014	0.087
α-EGR	-0.001	-0.289	-0.167	0.090	0.008	-0.203	0.025	0.097
MCV (fl)	2.81	-1.78	-2.55	3.2	1.36	-2.17	4.10	3.5
MCH (amol)	0.016	-0.024	-0.031	0.059	0.005	0.007	0.006	0.081
MCHC (mmol/l)	-0.33	0.05	0.66	0.43	-0.10	0.48	-0.38	0.35
gecorr.armdiam. (cm)	-0.04	0.03	0.09	0.19	-0.04	0.06	-0.31	0.25
tot.cholesterol (mmol/l)	0.02	-0.01	-0.14	0.55	-0.03	0.07	-0.04	0.35
HDL-cholesterol (mmol/l)	-0.15	0.07	0.02	0.20	-0.09	0.03	0.06	0.18

5. Conclusies

5.1. Voor de volgende variabelen is er bij toepassing van het eenvoudige covariantieanalyse-model een significant suppletie-effect gevonden (onbetrouwbaarheid 5%).

	Grootte van suppletie-effect	95% betrouwbaar- heidsinterval
meisjes: TYBC (mol/1)	-4.5	(-7.5 , -1.5)
log ferritine (mol/1)	0.41	(0.23 , 0.59)
α -ETK	-0.146	(-0.189, -0.103)
α -EGR	-0.289	(-0.344, -0.234)
jongens: α -ETK	-0.113	(-0.176, -0.050)
α -EGR	-0.203	(-0.273, -0.133)
MCHC (mmol/1)	0.48	(0.23 , 0.73)

5.2. Voor de variabele MCHC bij meisjes bestaat er een significante interactie jaar x suppletie: het suppletie-effect voor jaar 2 is positief (0.32) en voor jaar 1 negatief (-0.34).

5.3. Voor de variabele α -EGR blijkt het suppletie-effect voor zowel jongens als voor meisjes af te hangen van de waarde van α -EGR (zie 4.2.).

Het in 5.1. gegeven effect voor α -EGR is te beschouwen als een "gemiddeld effect", gemiddeld over de in het onderzoek voorkomende lage en hoge waarden van α -EGR.

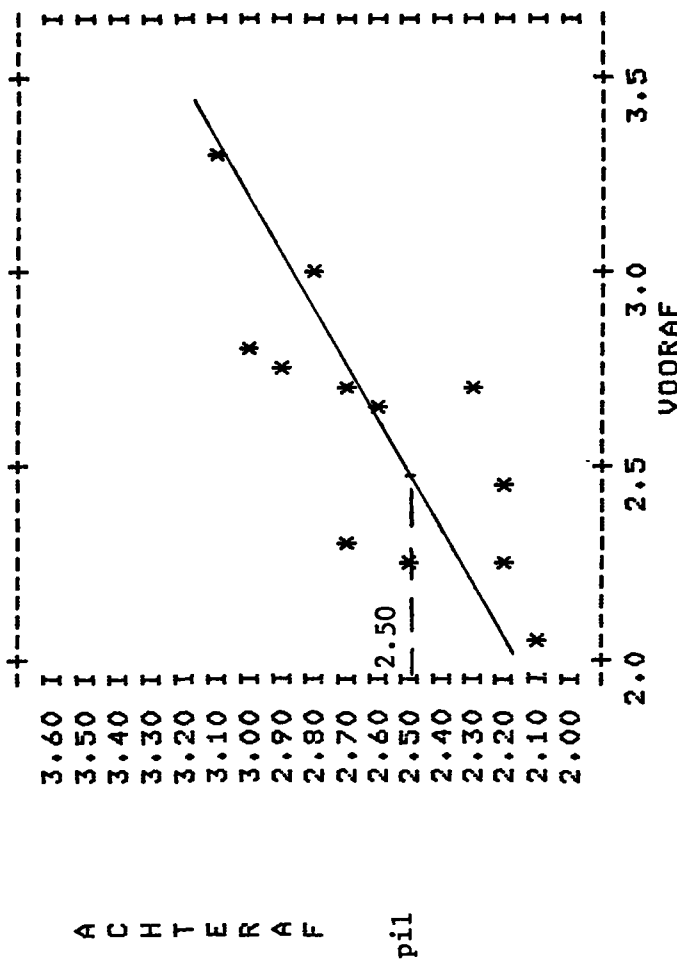
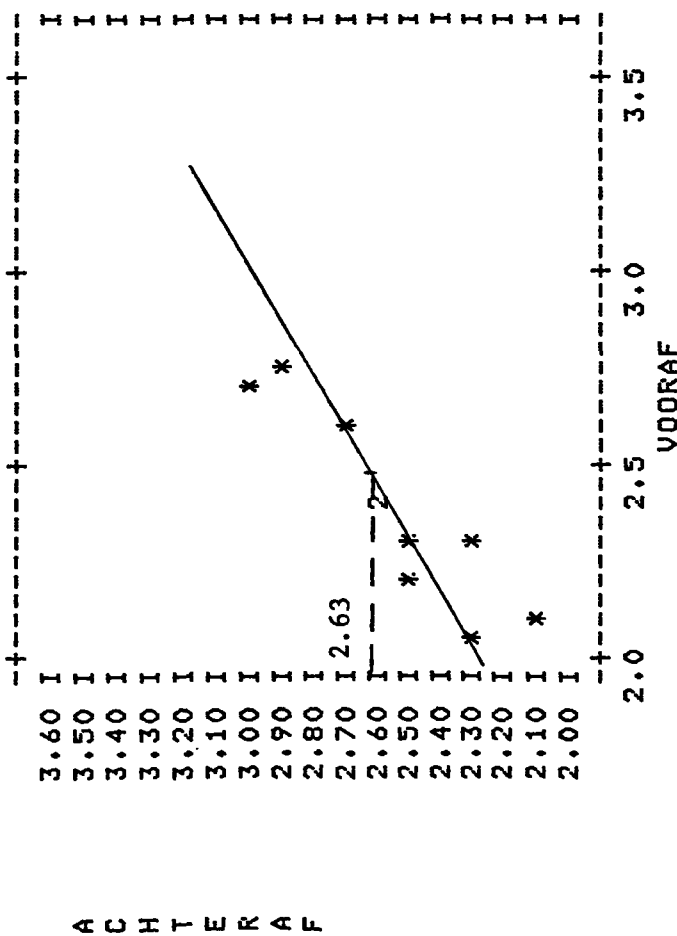
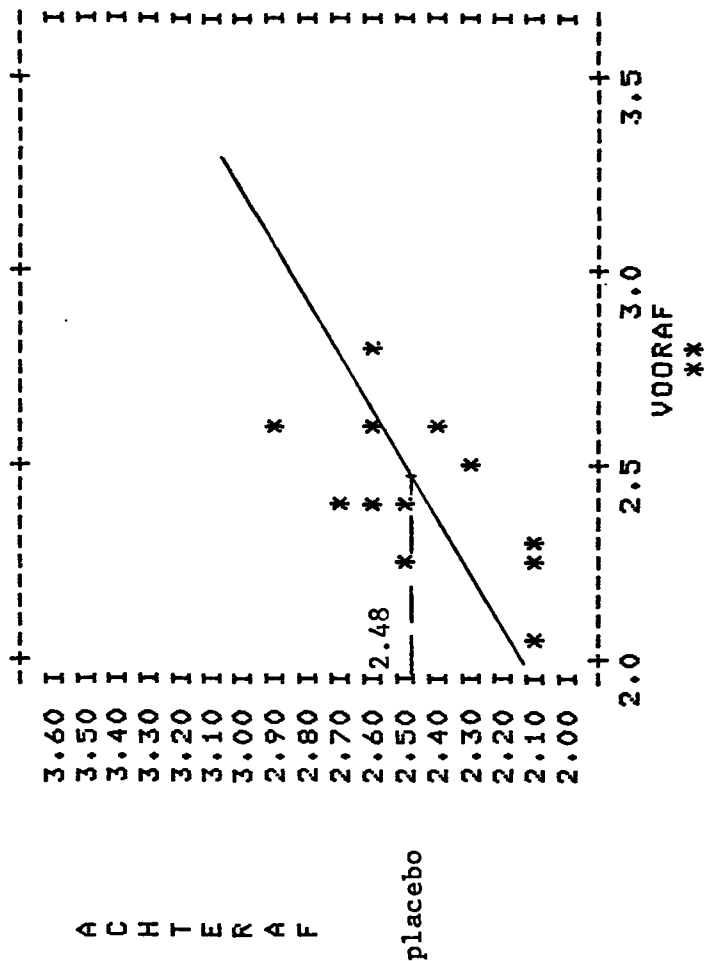
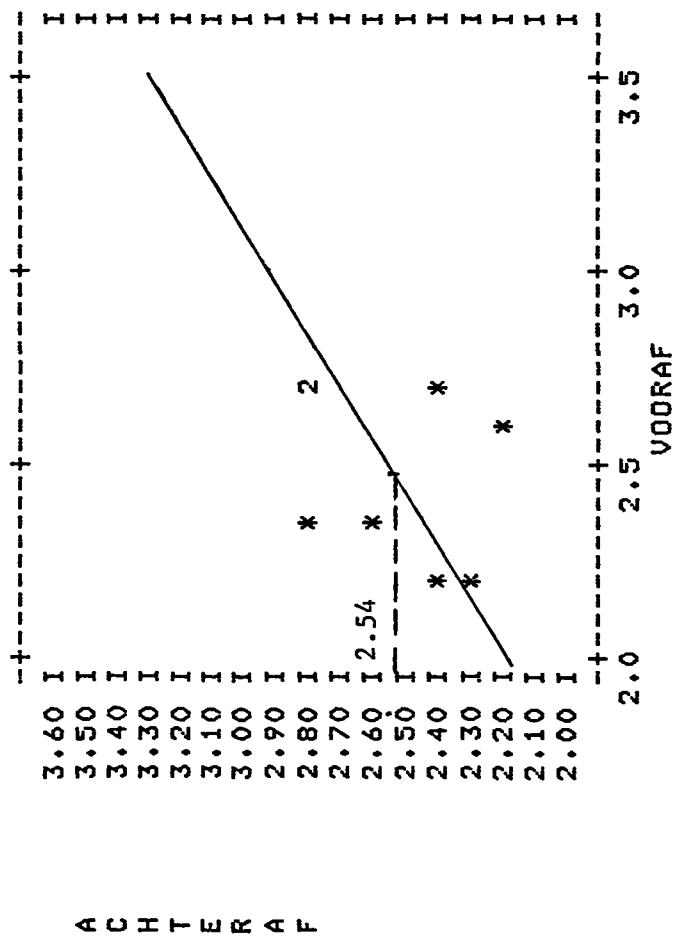
Bijlage 1

Voor de te beschouwen variabele y wordt het volgende mathematische model ondersteld:

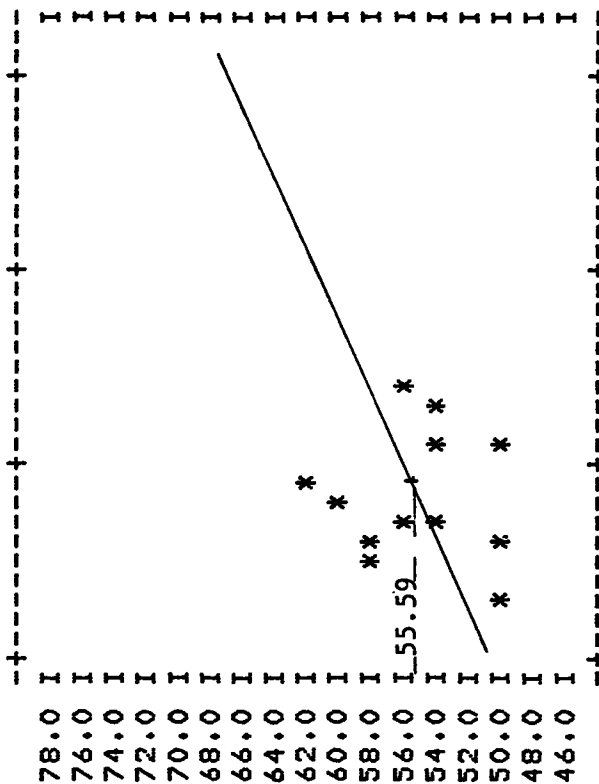
$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma x_{ijk} + e_{ijk}, \quad i=1,2; j=1,2; k=k(i,j)$$

- waarbij y_{ijk} = de te beschouwen variabele van persoon k in jaar i en suppletiegroep j gemeten op onderzoekdatum 2
- x_{ijk} = analoog maar dan gemeten op onderzoekdatum 1
- α_i = een constante grootte die het effect van jaar i representeert ($i = 1$ jaar 1; $i = 2$ jaar 2)
- β_j = een constante grootte die het effect van suppletie representeert
($j = 1$ placebo; $j = 2$ pil)
- $(\alpha\beta)_{ij}$ = een constante grootte, die de interactie tussen jaar i en suppletie j representeert
- γ = een constante, die de richtingscoëfficiënt representeert bij de regressie van y op x
- e_{ijk} = een stochastische grootte, die de variabiliteit van y_{ijk} beschrijft voor zover niet reeds bevat in voorgaande termen.

Veronderstellingen: e_{ijk} onderling onafhankelijk normaal verdeeld met verwachtingswaarde nul en variantie σ_e^2 , en voorts de gebruikelijke normeringsveronderstellingen.



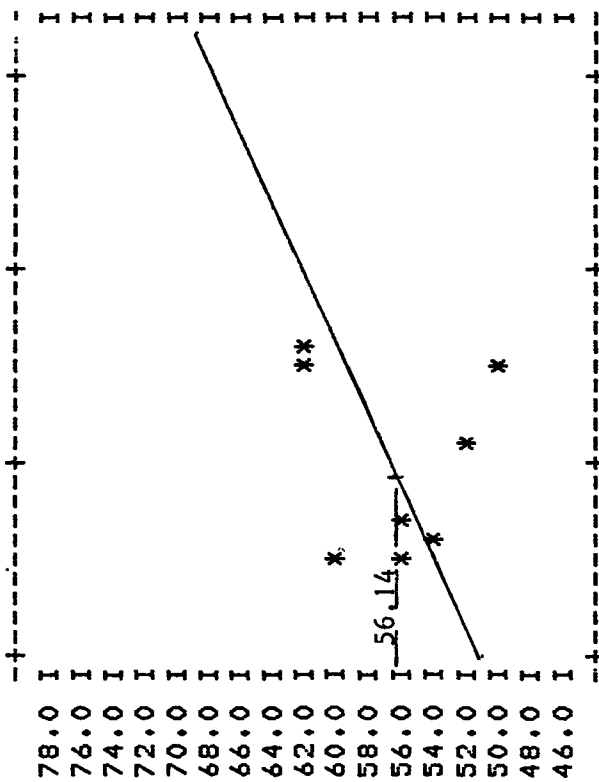
jaar 2



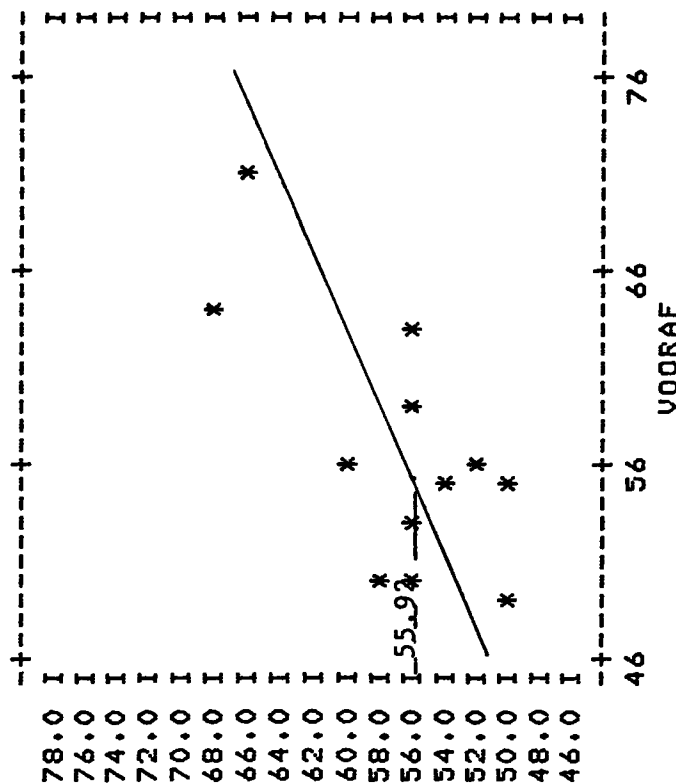
A C H T E R A F

placebo

jaar 1

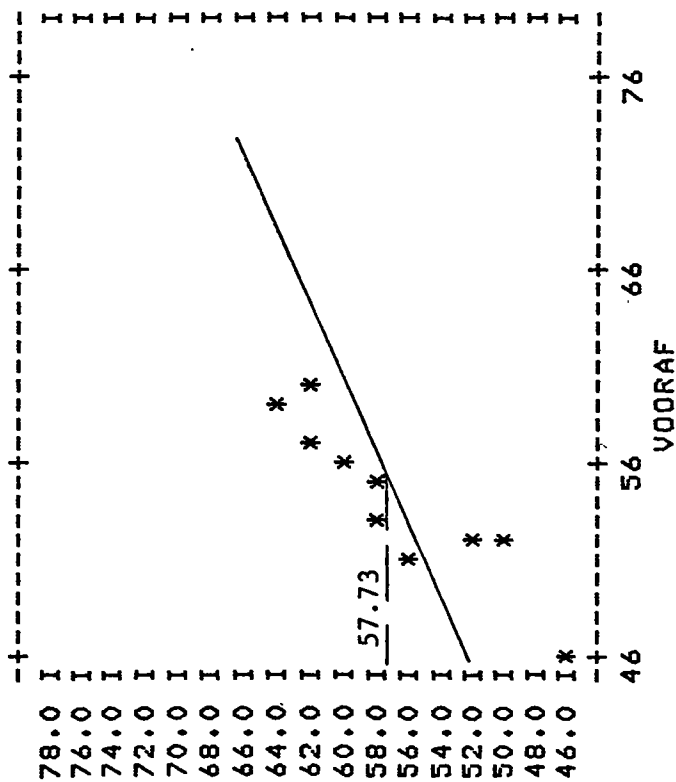


A C H T E R A F



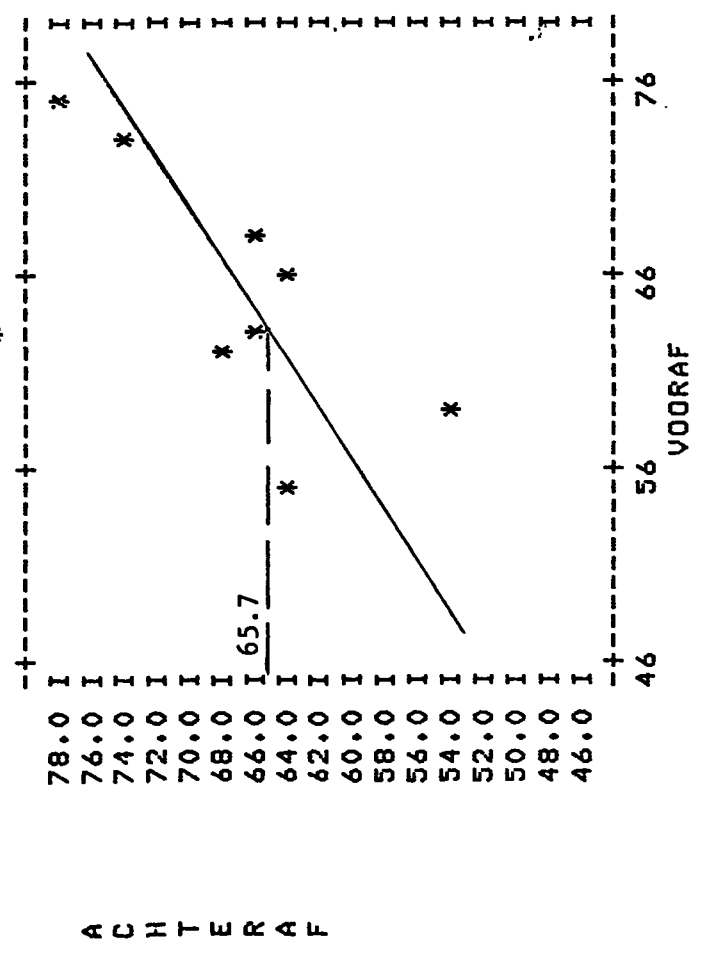
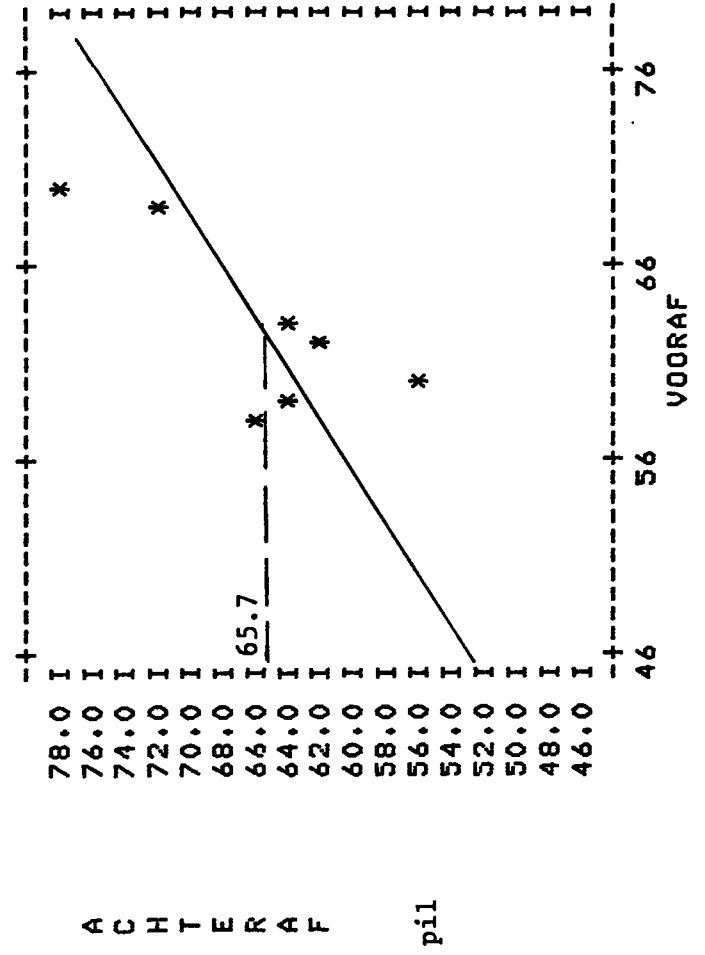
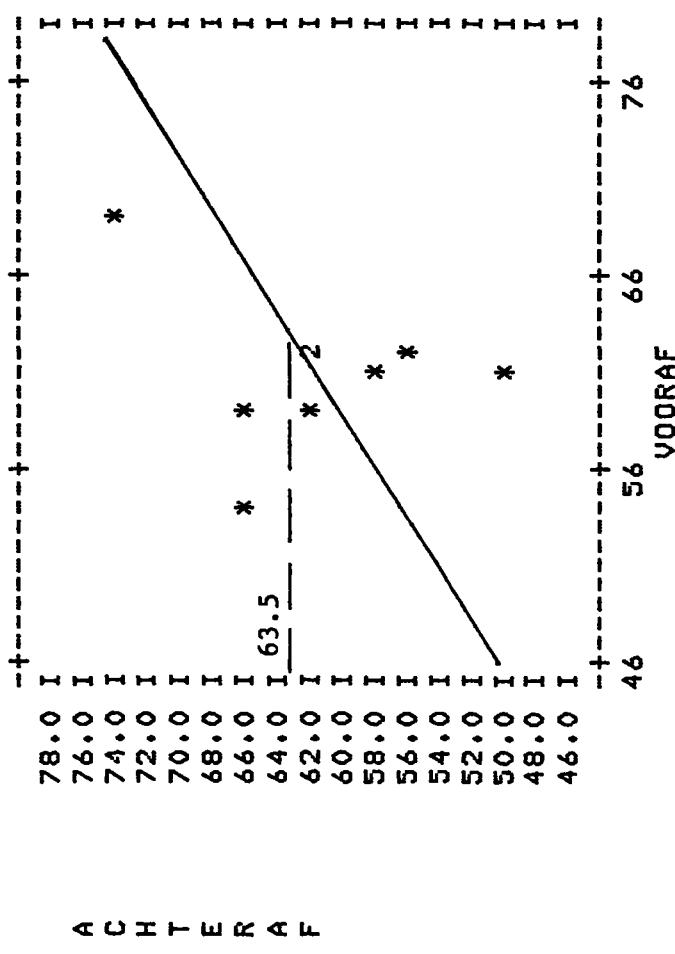
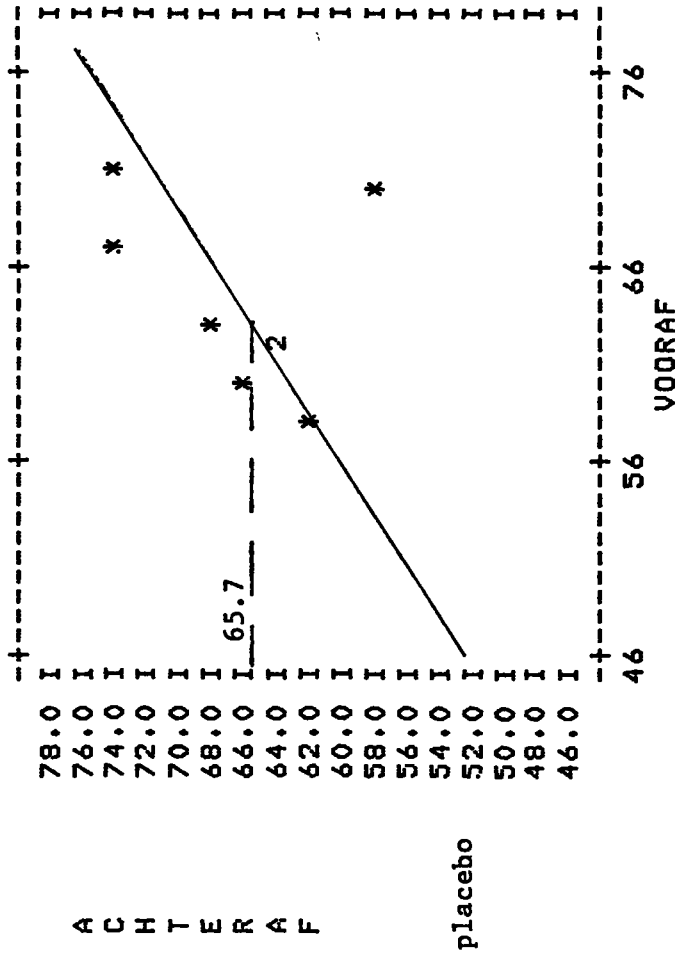
A C H T E R A F

pil



A C H T E R A F

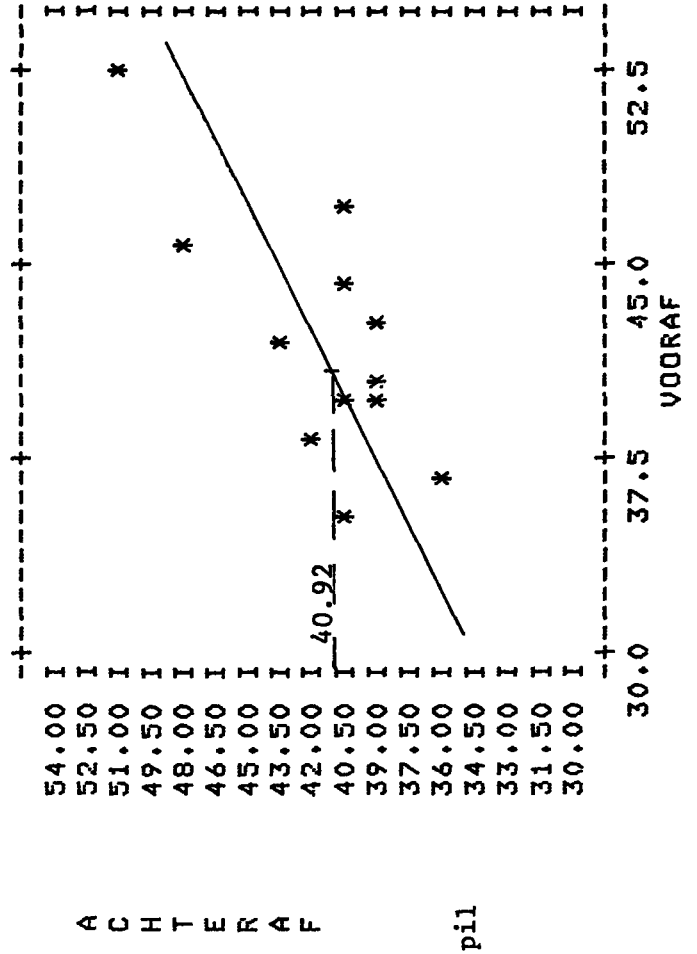
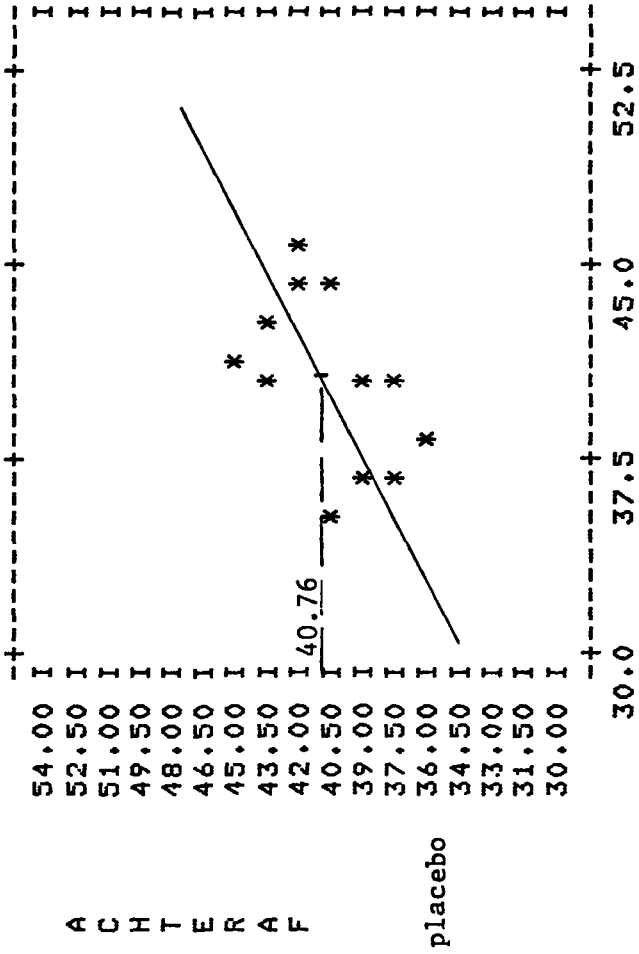
jaar 2



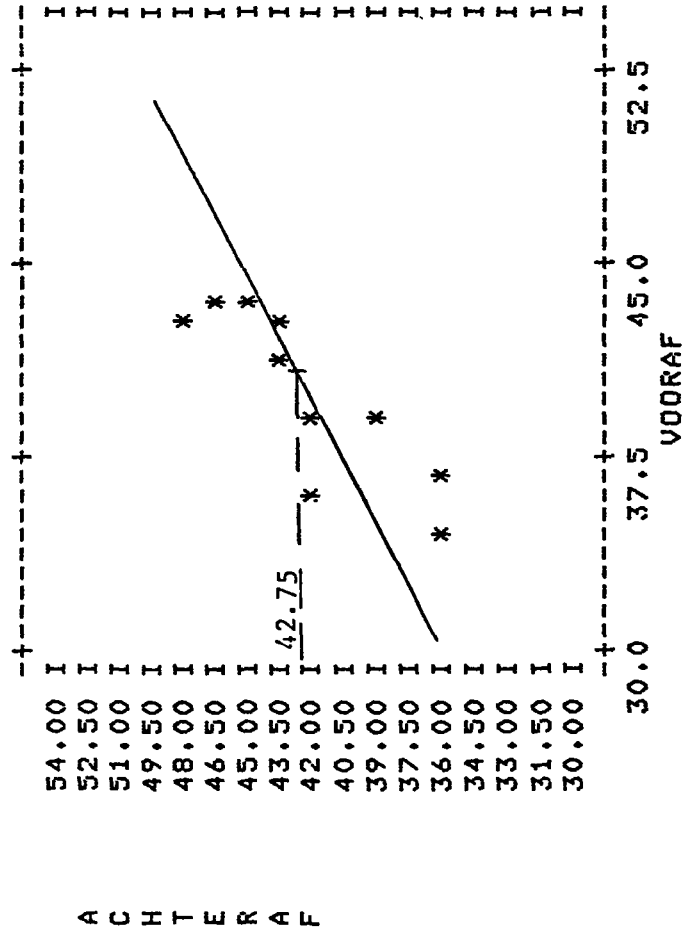
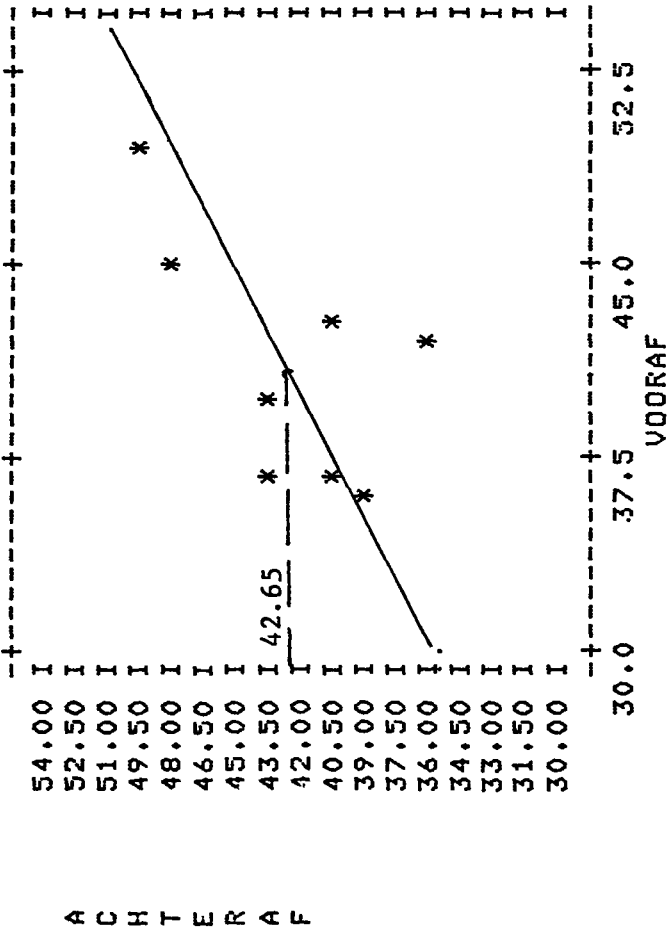
jongens

VO2 MAX IN ML/MIN/VUUM

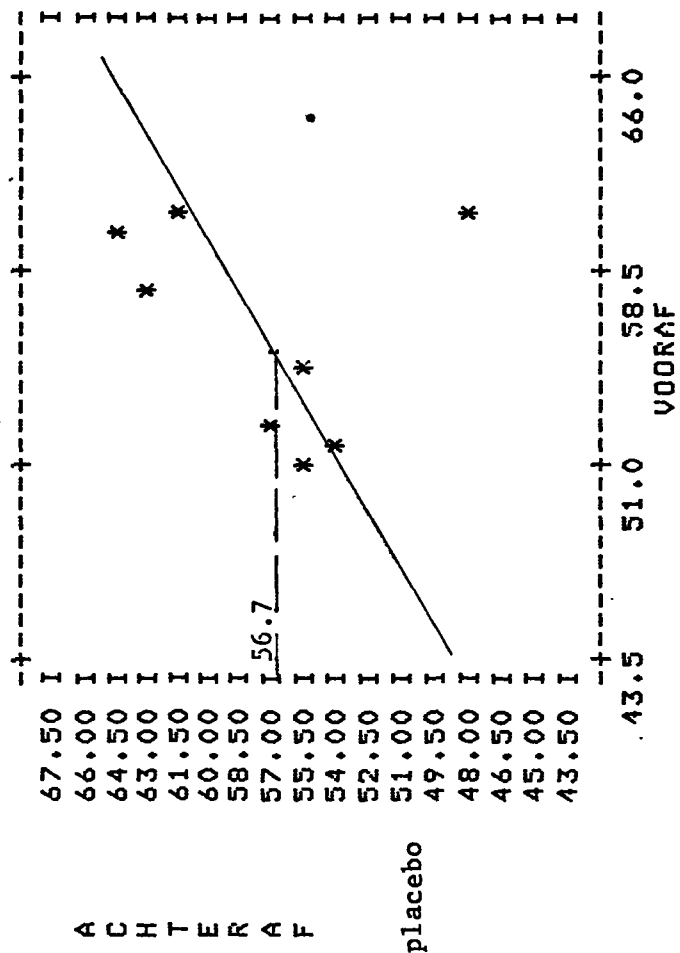
jaar 2



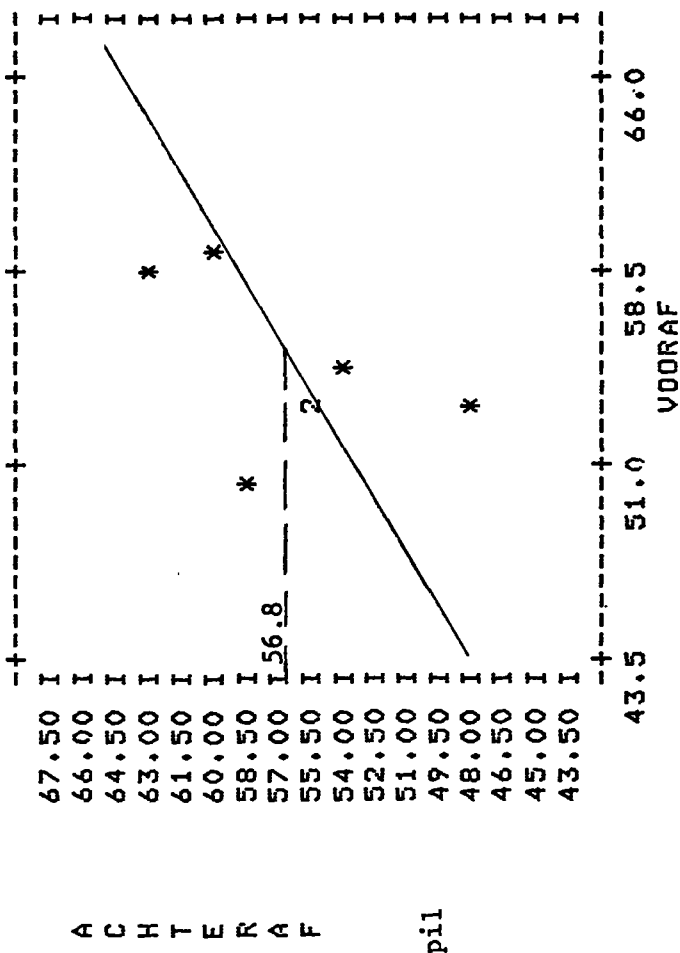
jaar 1



jaar 2

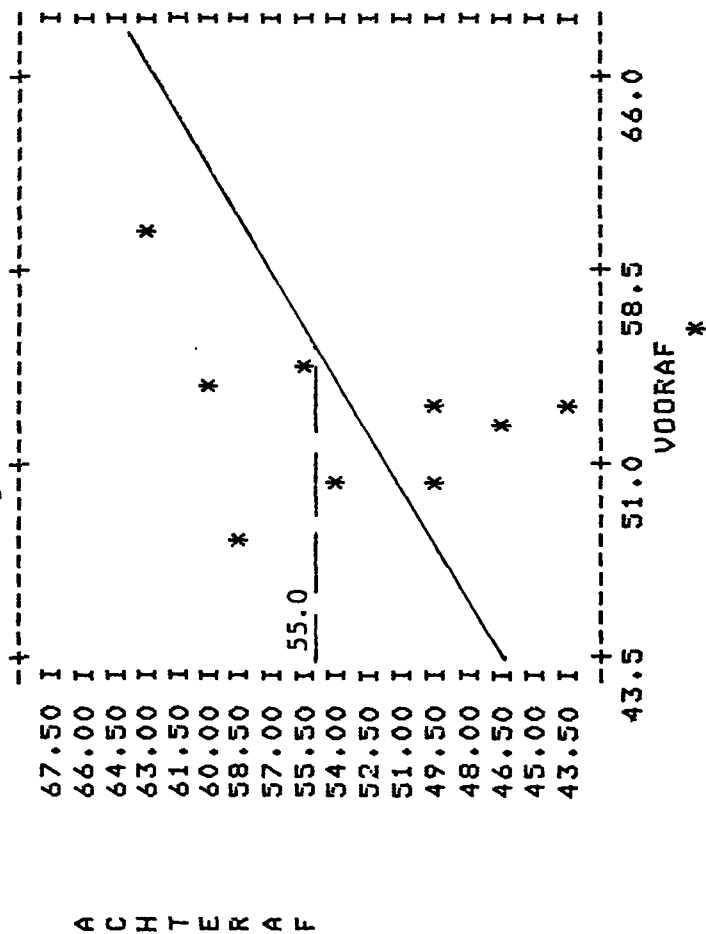


placebo

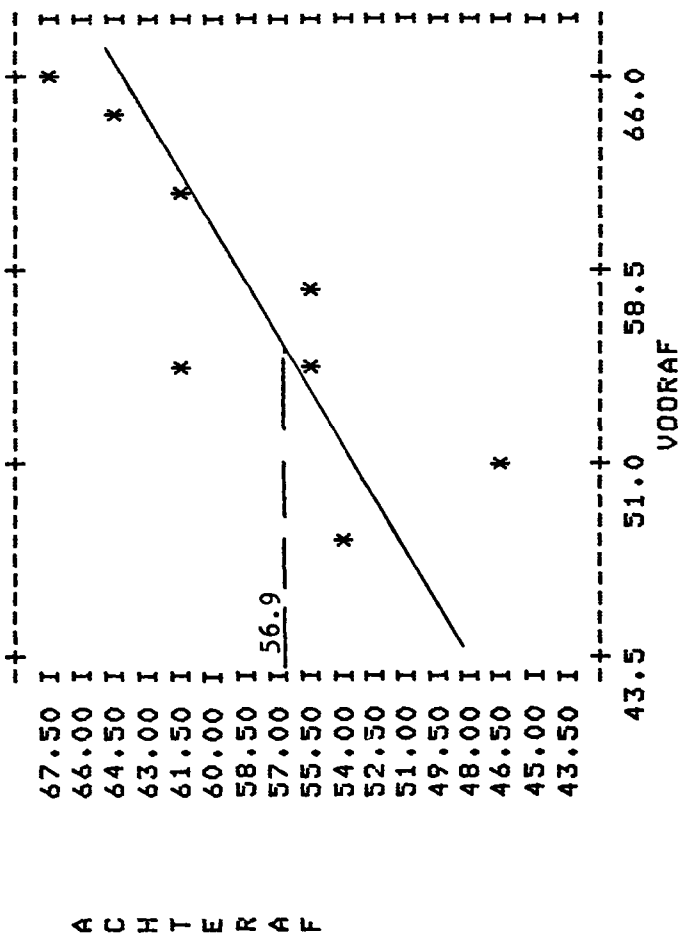


pil

jaar 1



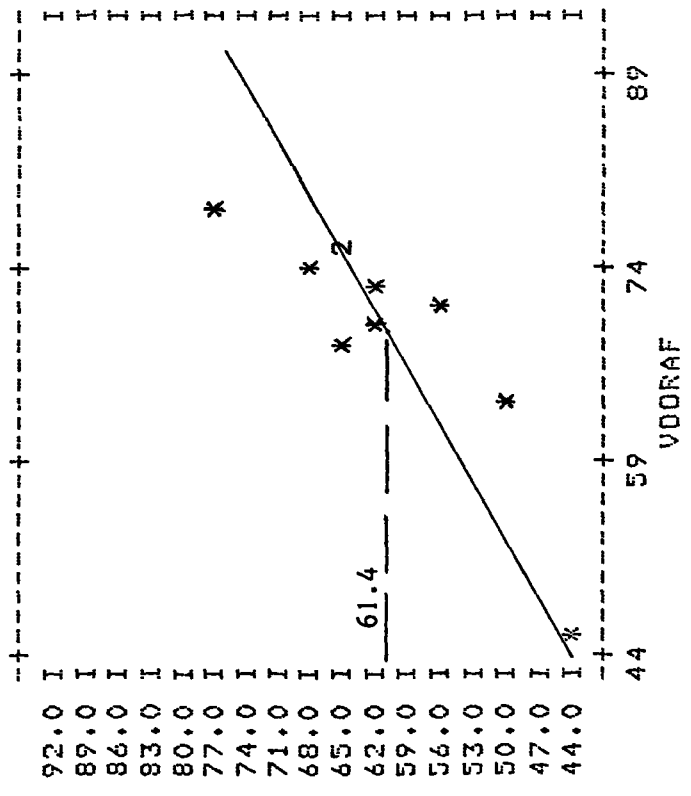
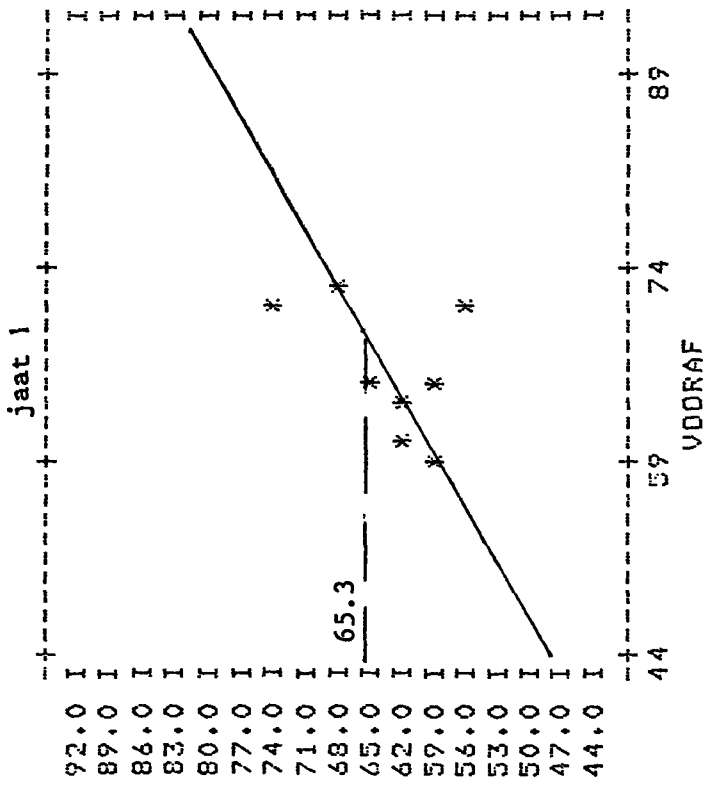
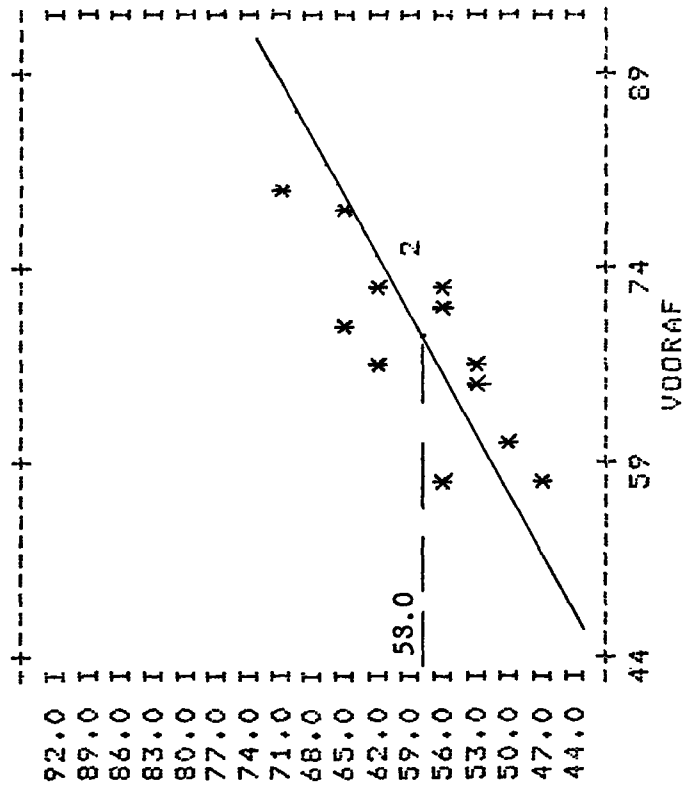
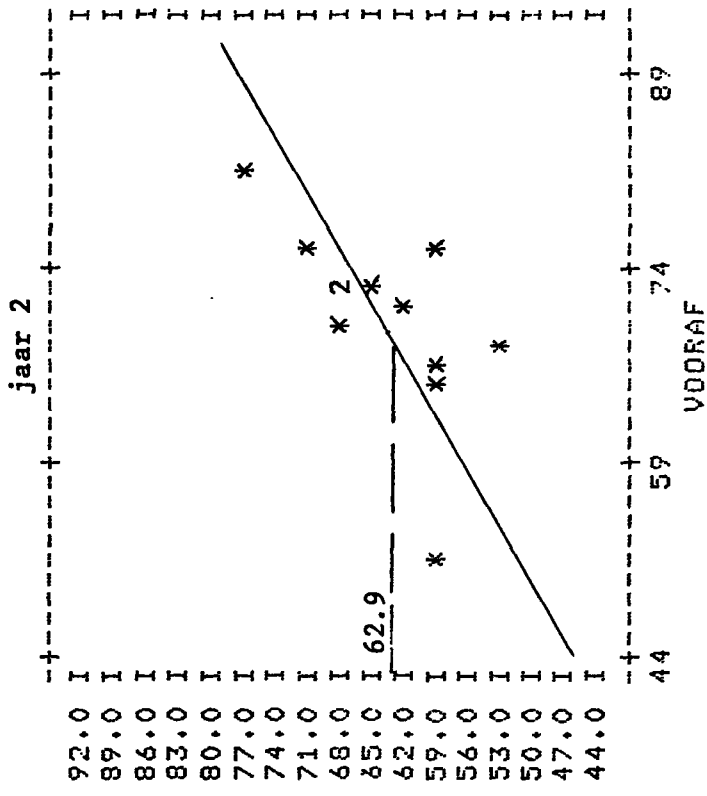
placebo



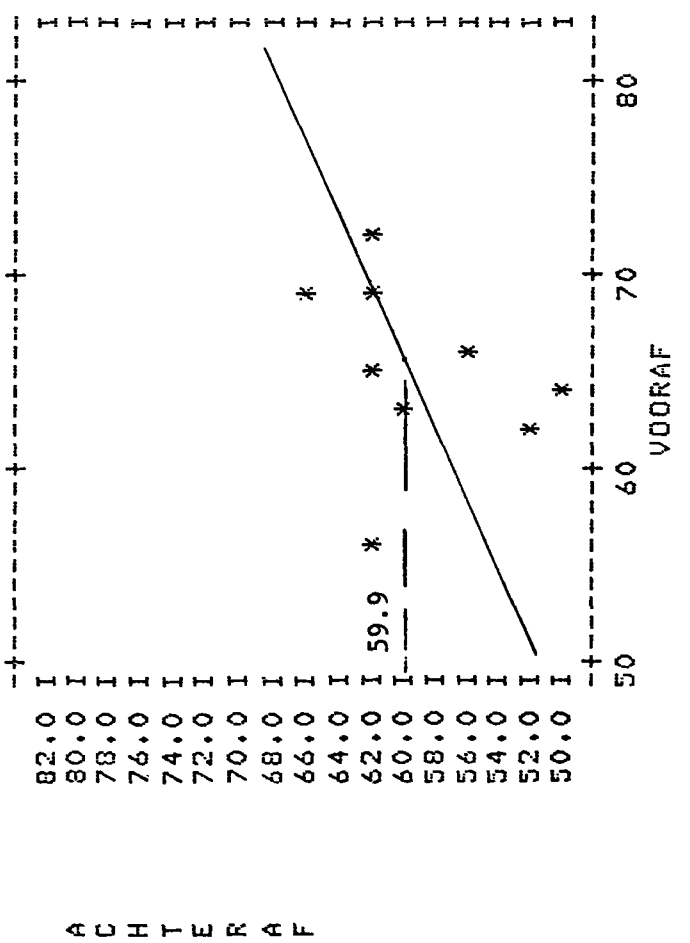
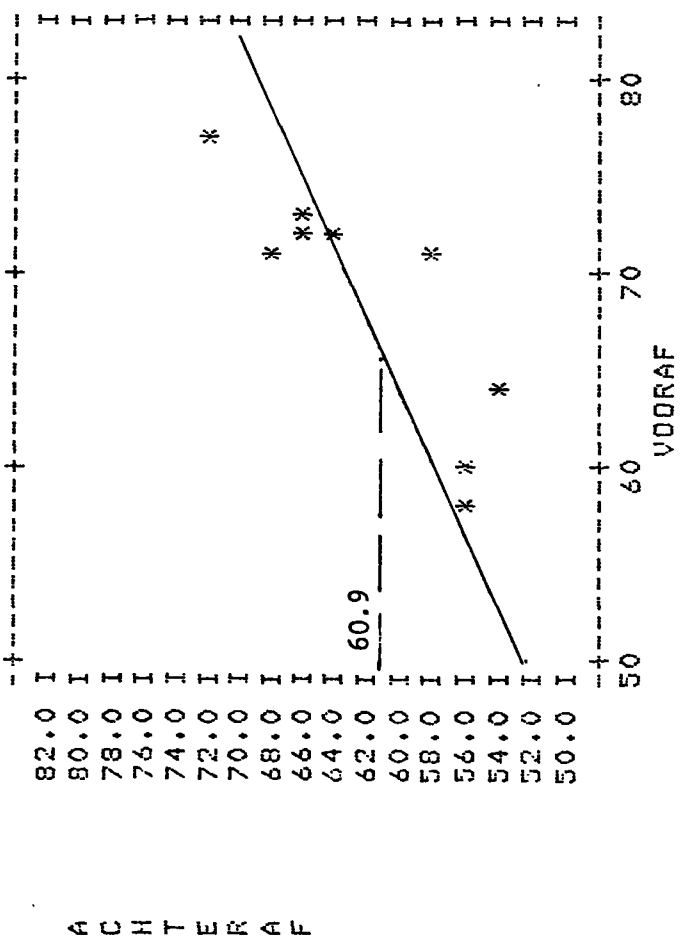
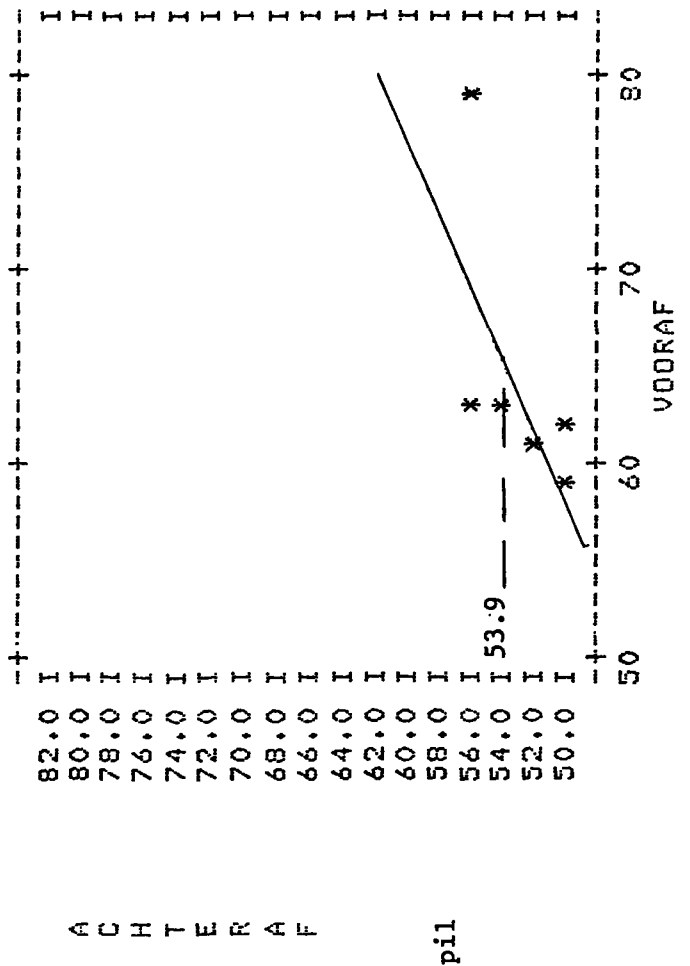
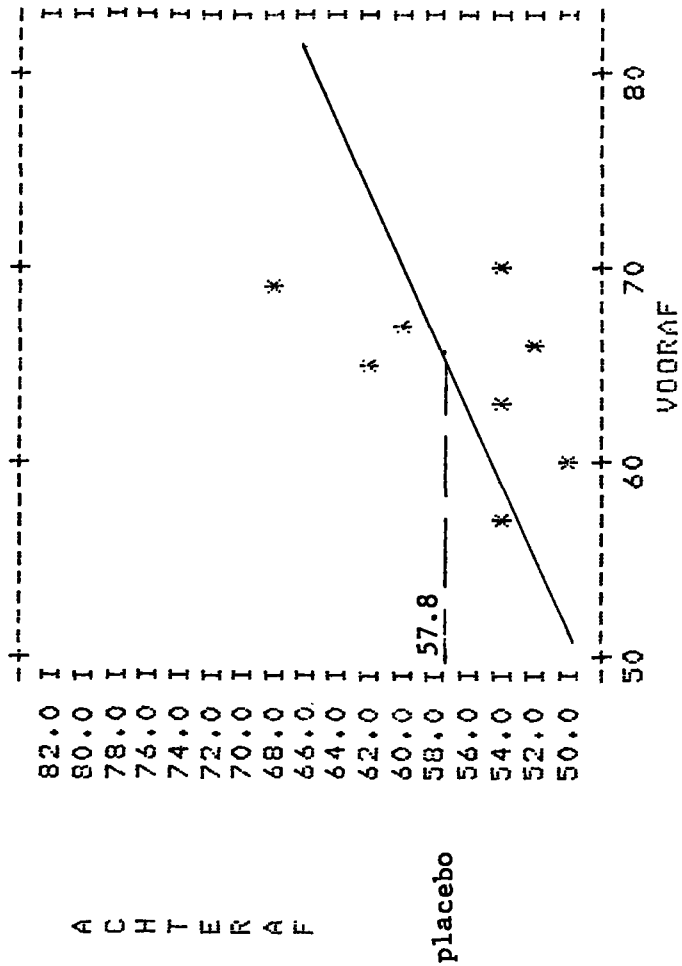
pil

V02 MAX IN ML/MIN/KG

jongens



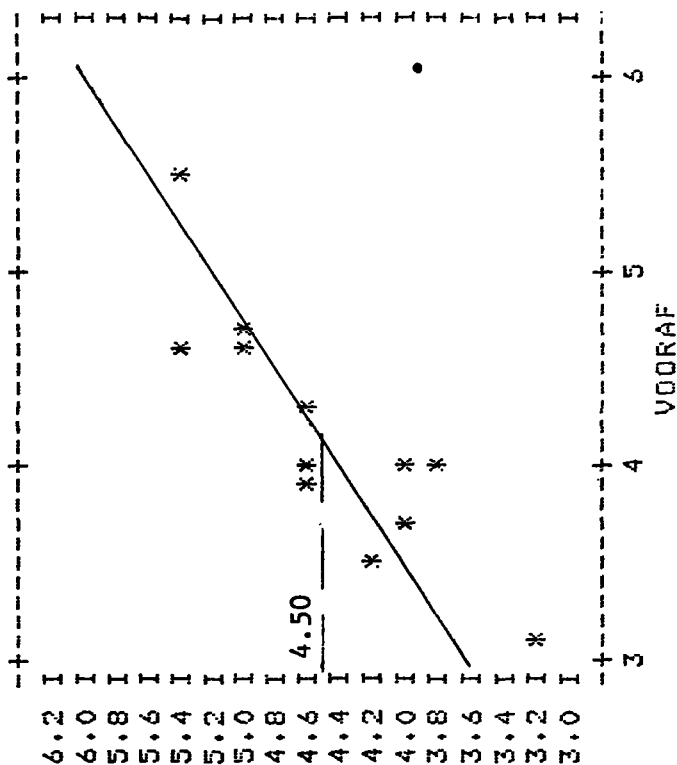
Jaar 2



jongens

TYBC

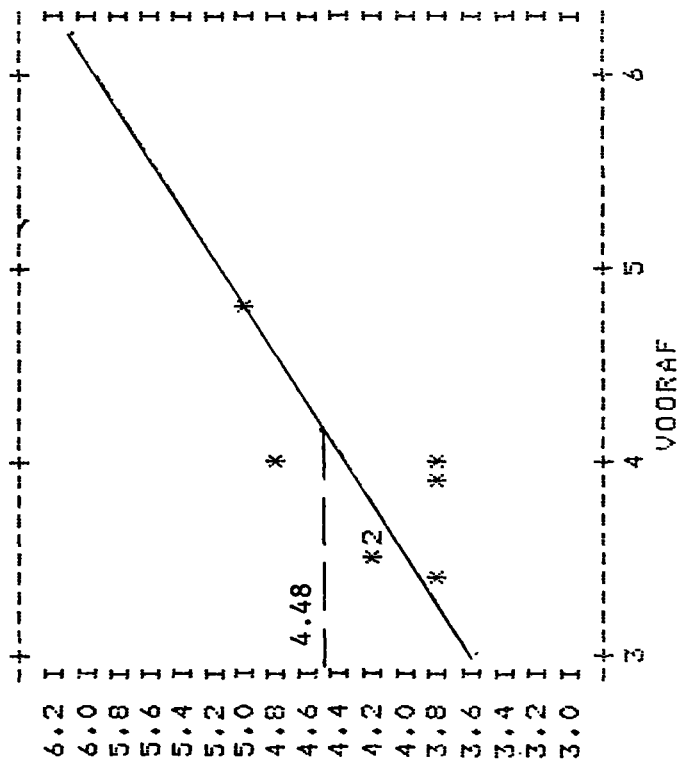
jaar 2



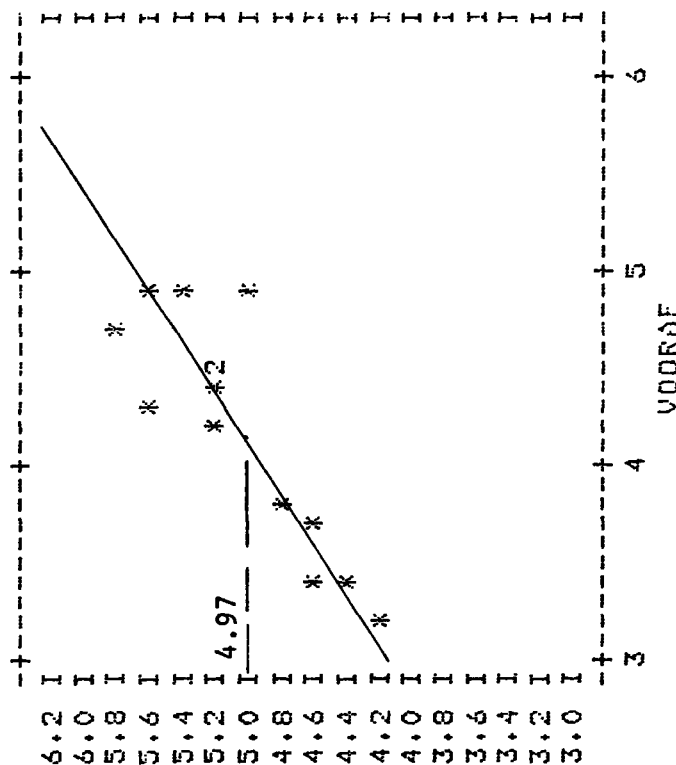
ACHTERAF

placebo

jaar 1

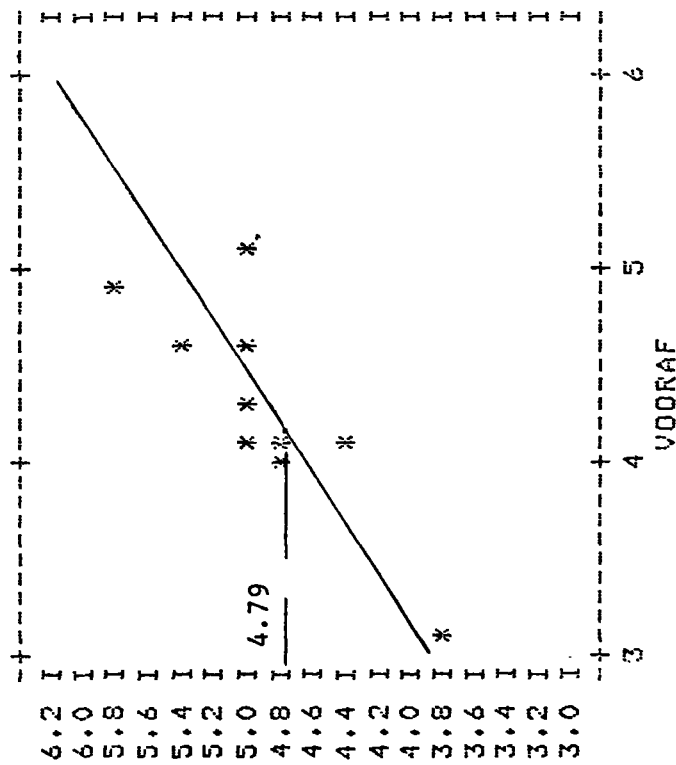


ACHTERAF



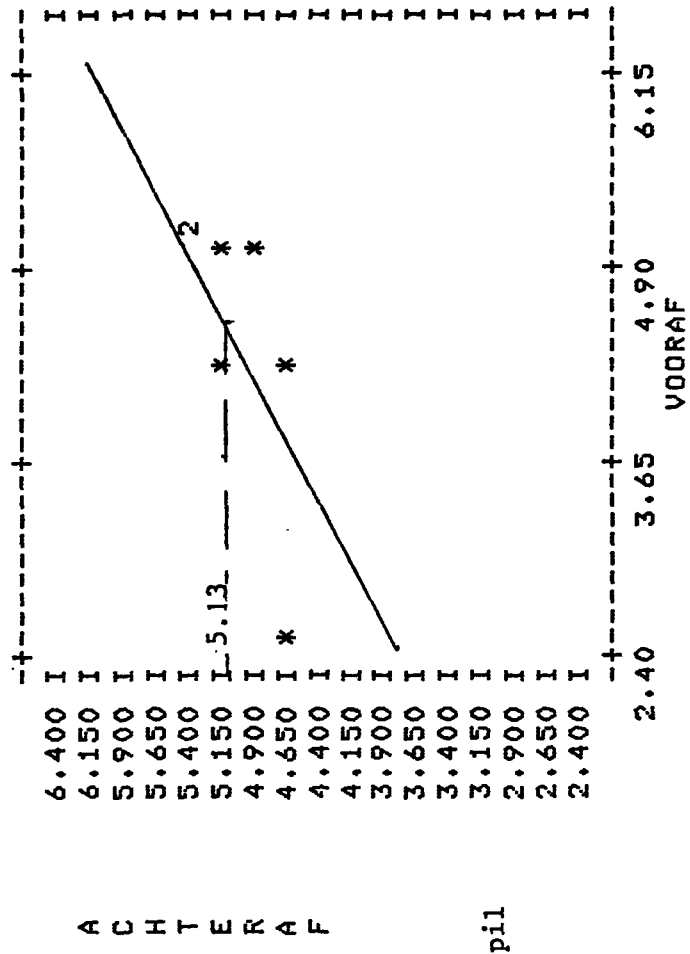
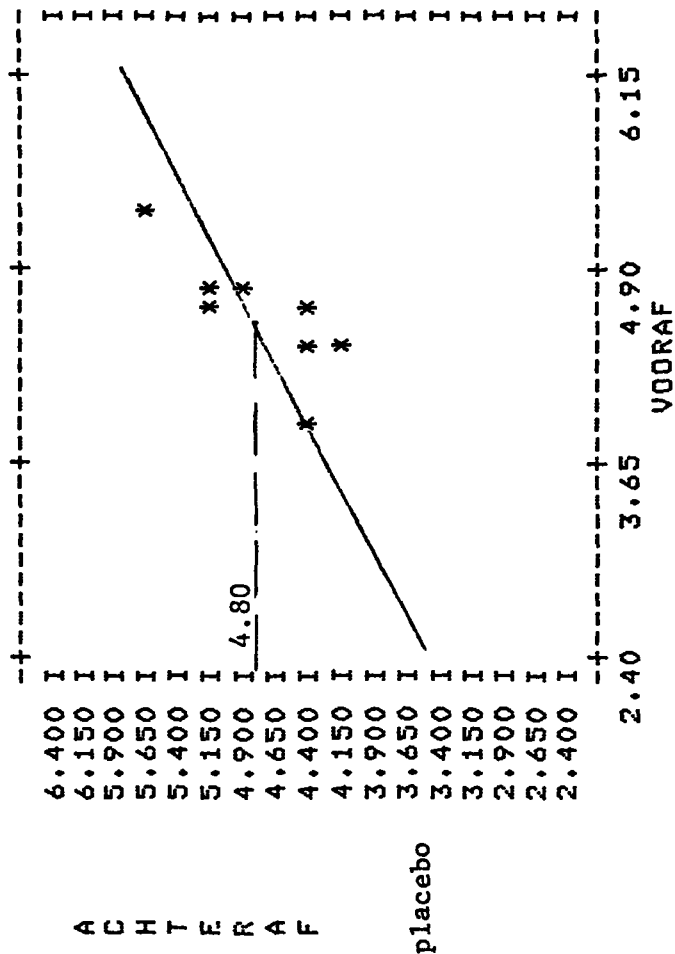
ACHTERAF

pil

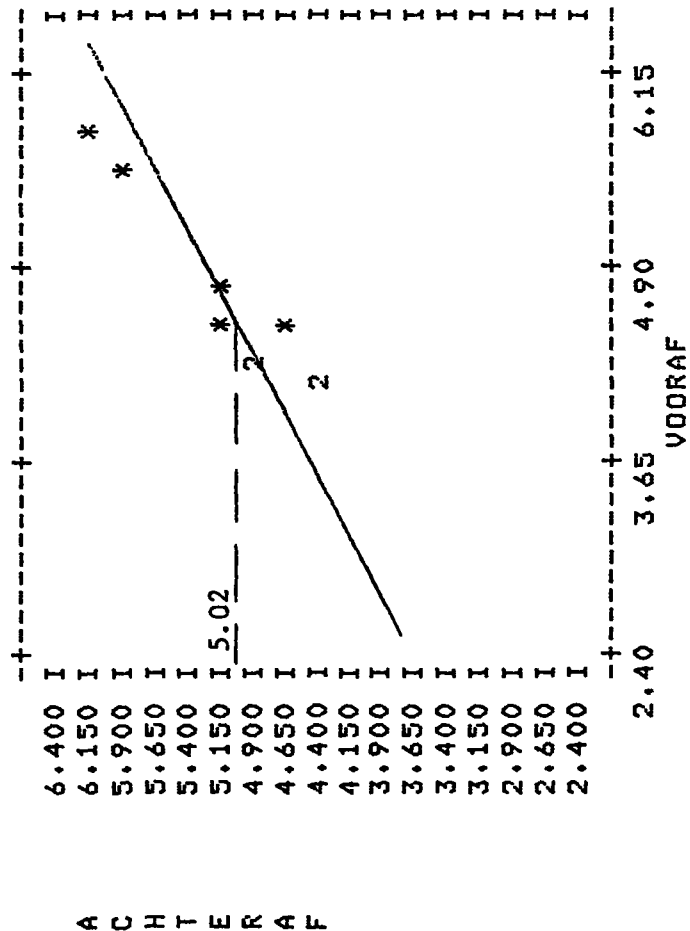
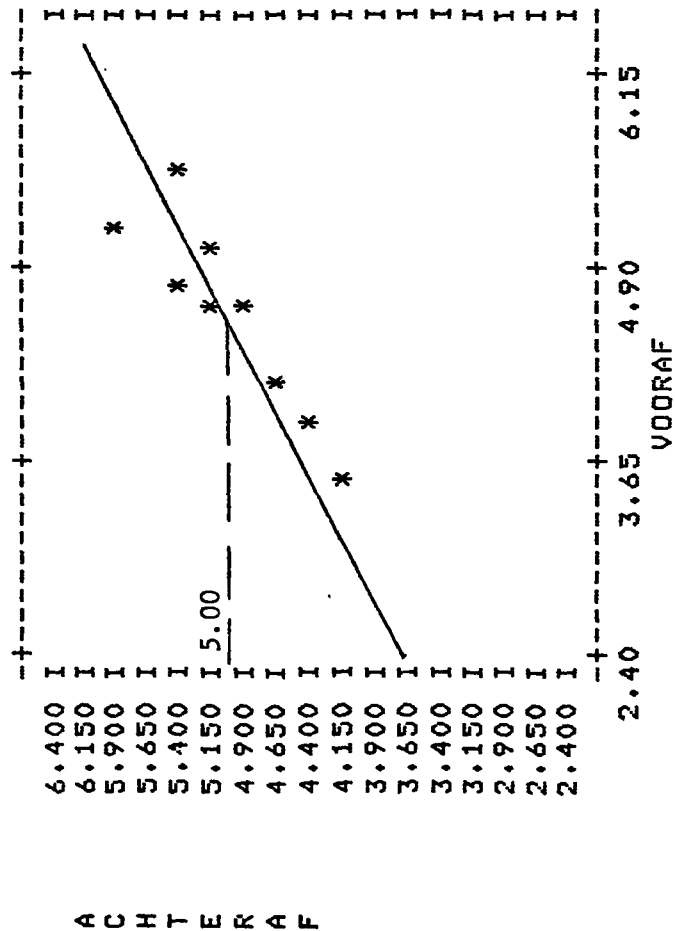


ACHTERAF

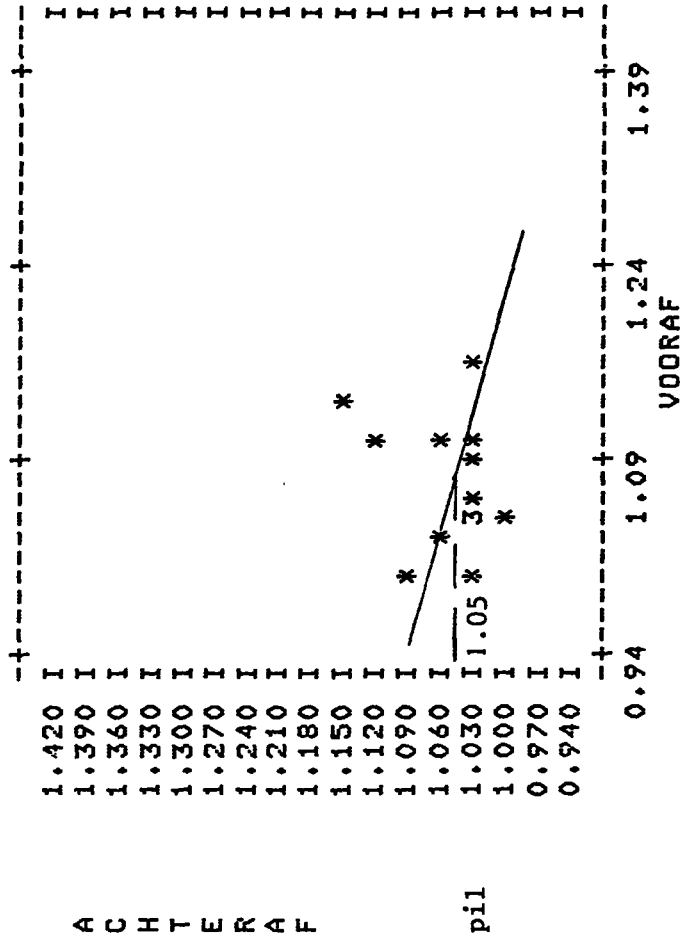
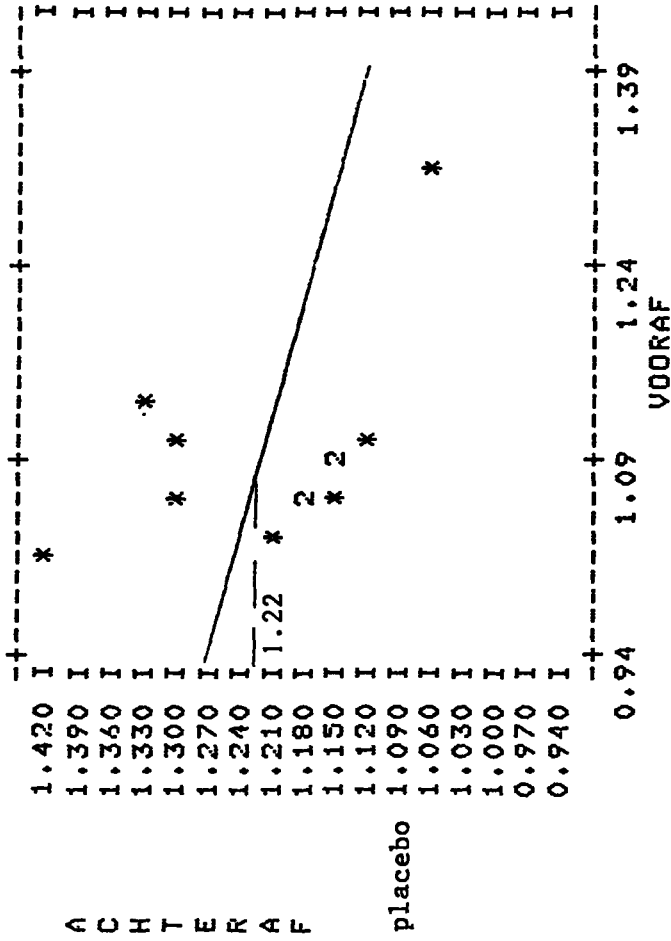
jaar 2



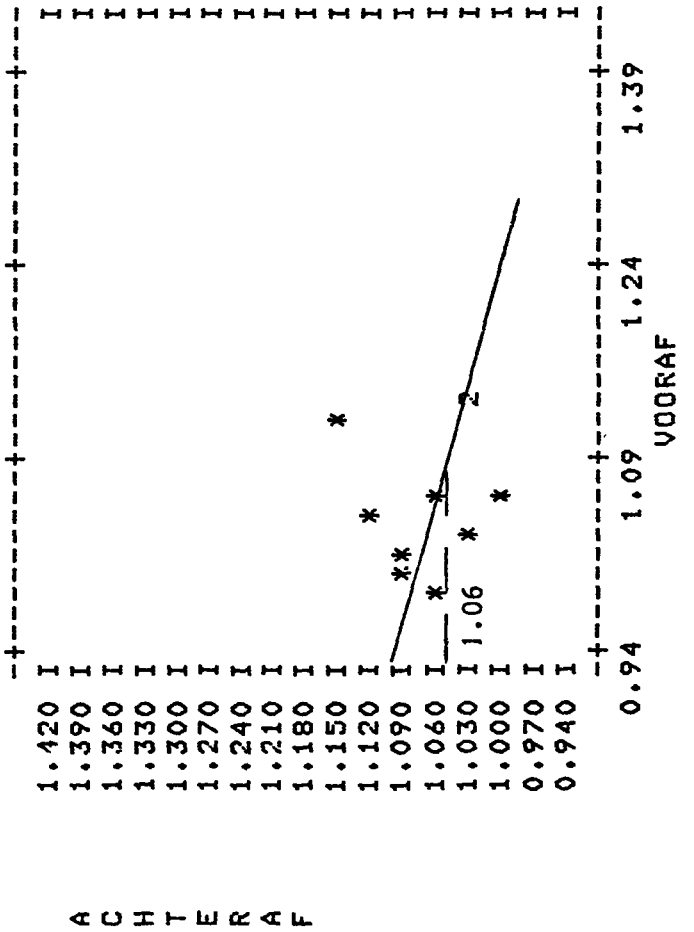
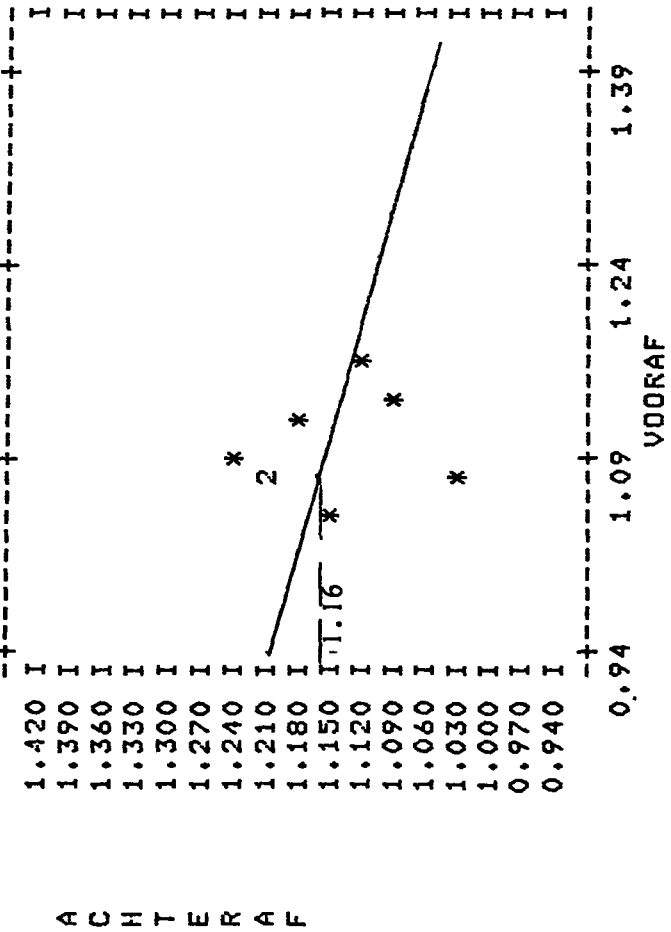
jaar 1



jaar 2



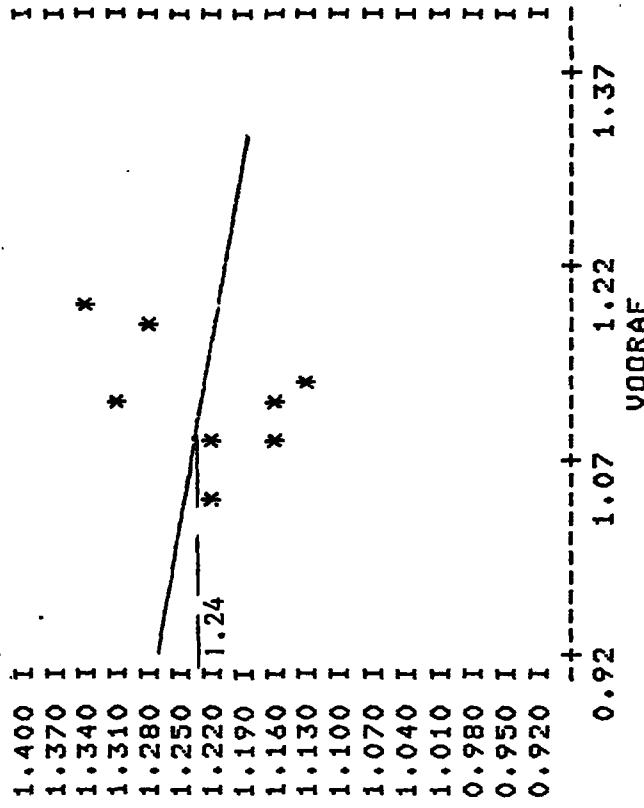
jaar 1



meisjes

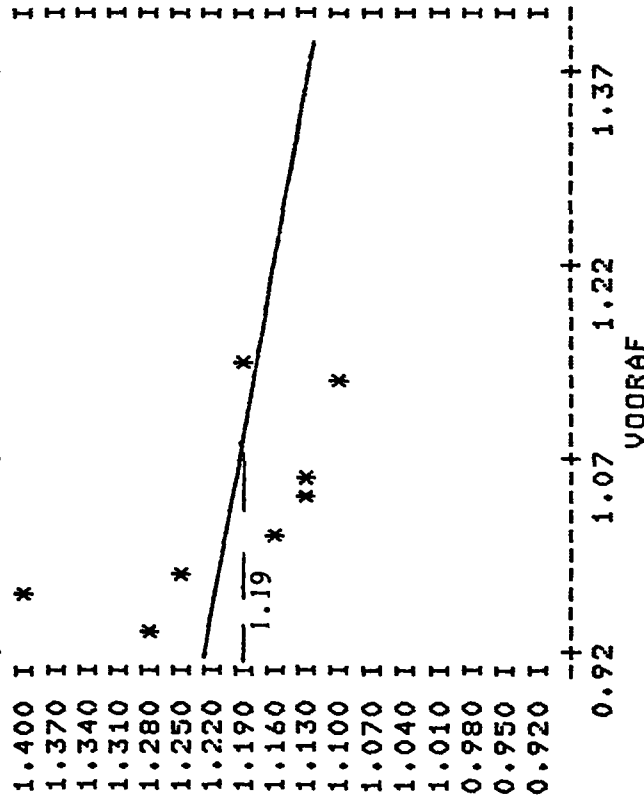
ETK

Jaar 2

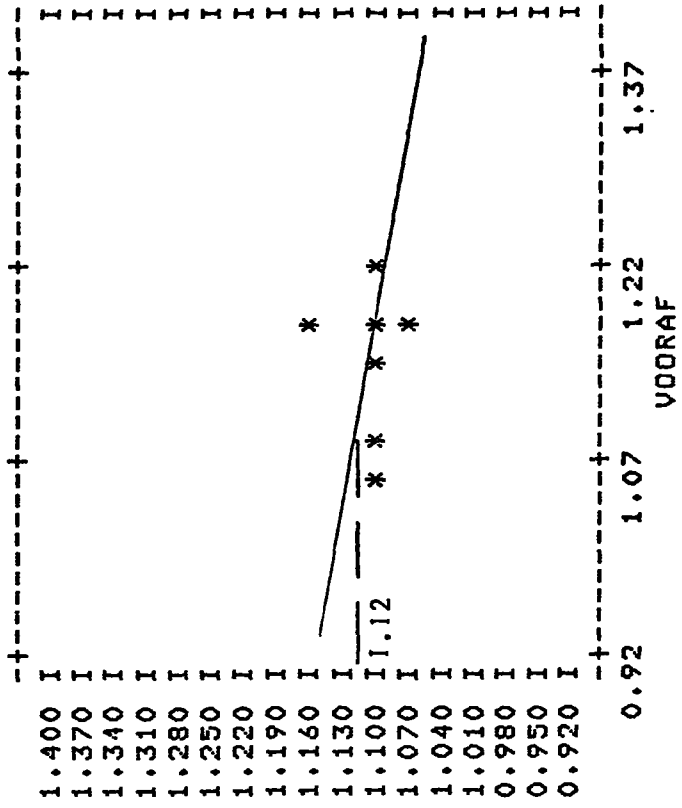


A C H T E R A F

placebo

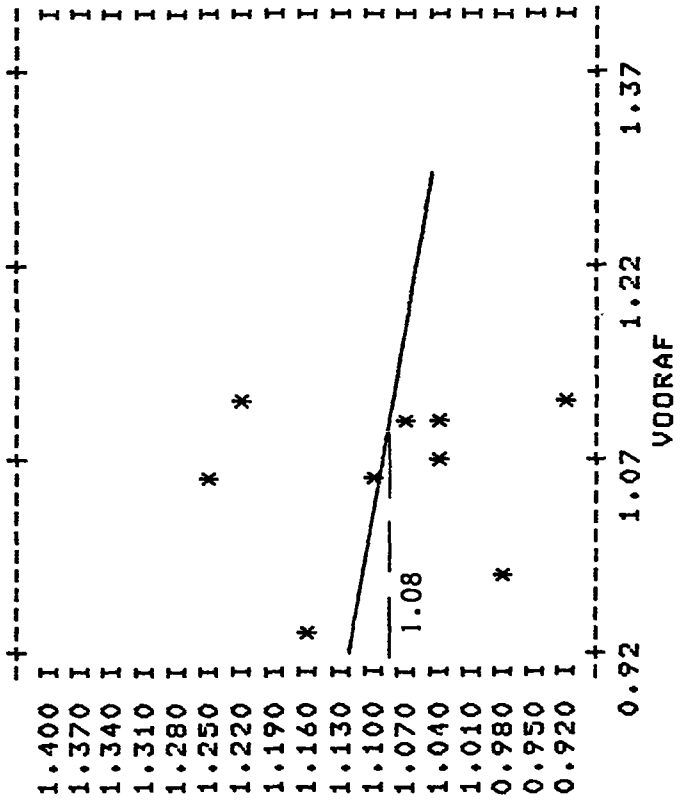


A C H T E R A F



A C H T E R A F

pil

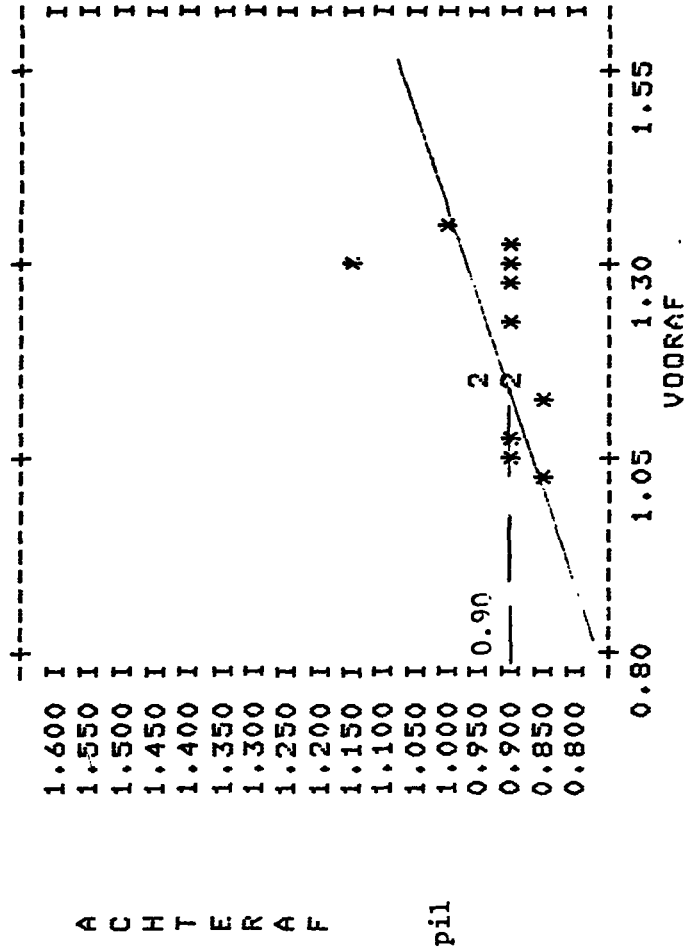
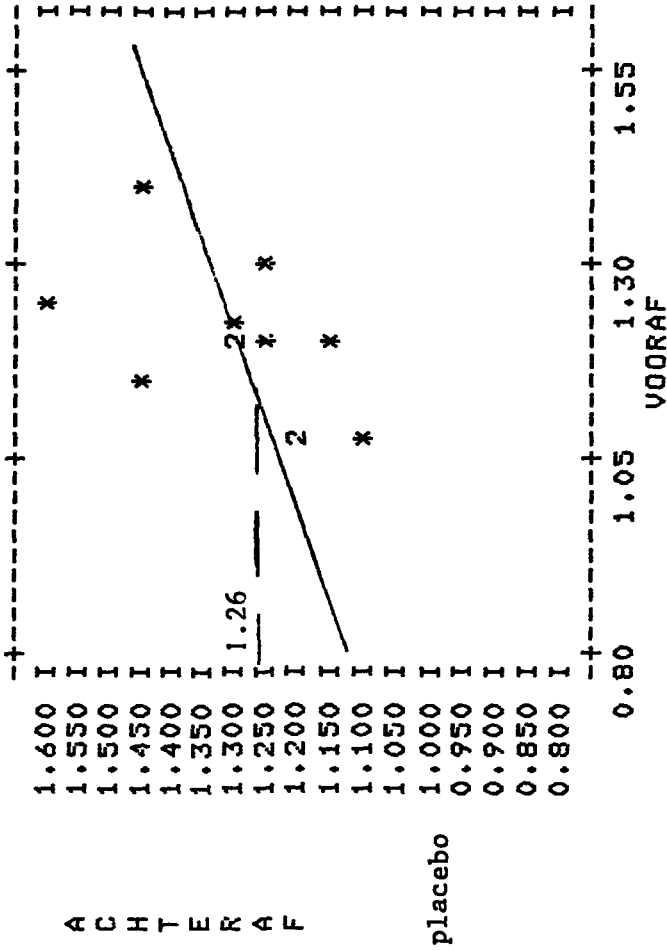


A C H T E R A F

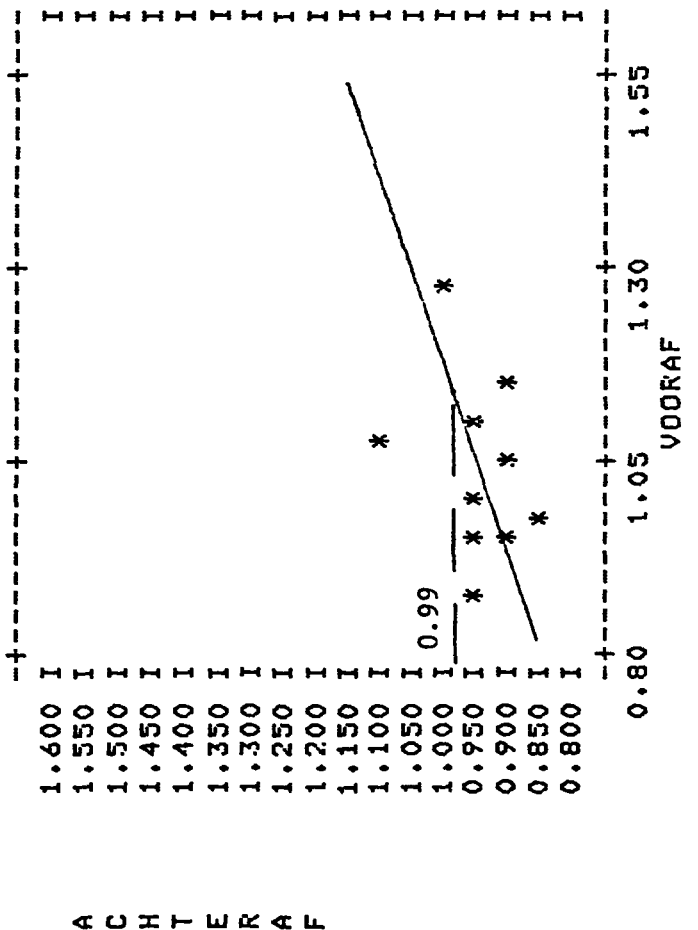
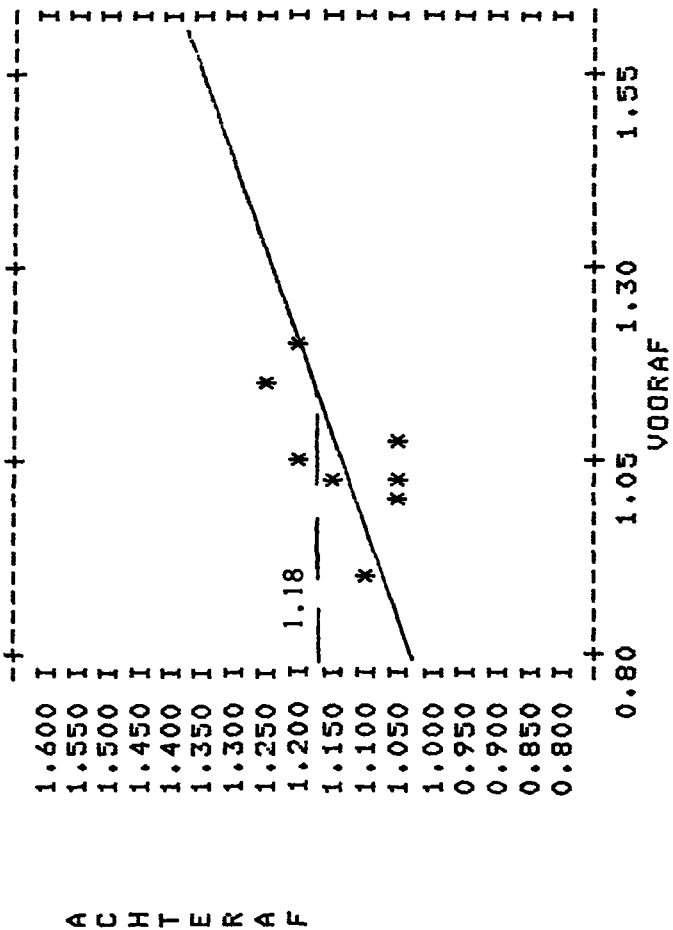
jongens

ETK

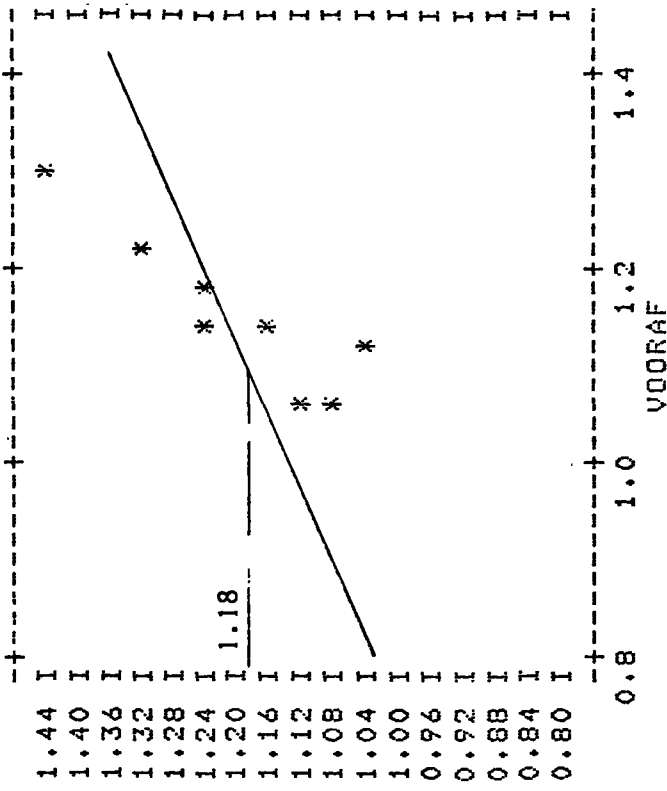
jaar 2



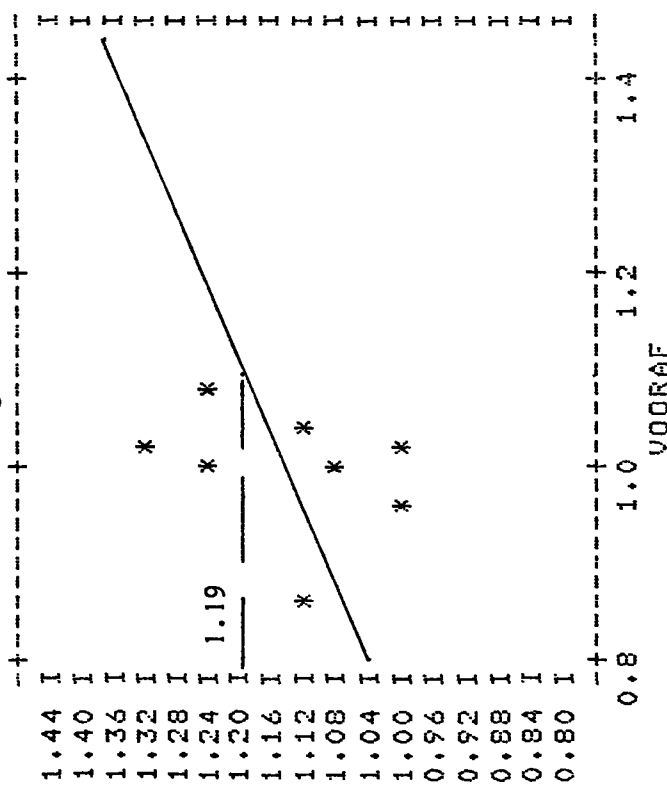
jaar 1



jaar 2

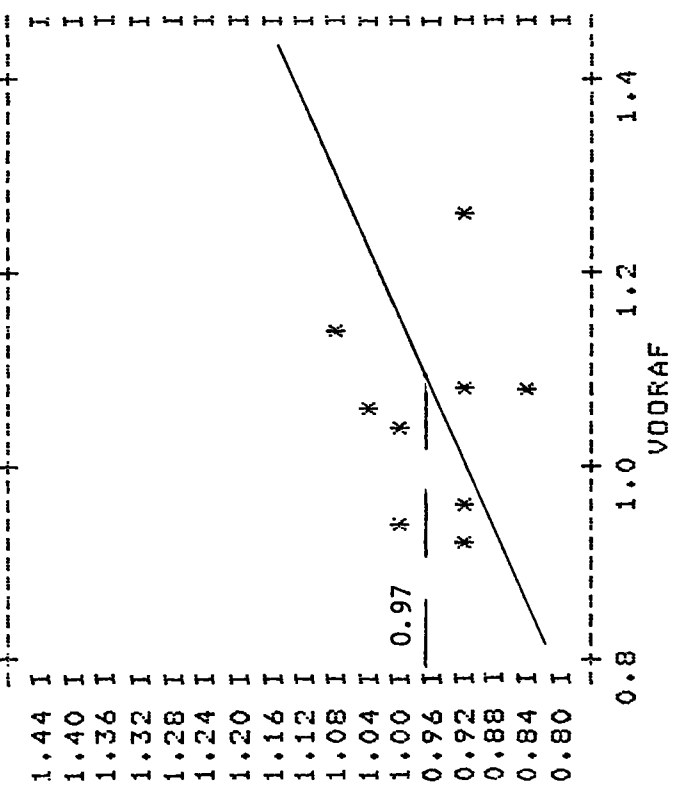
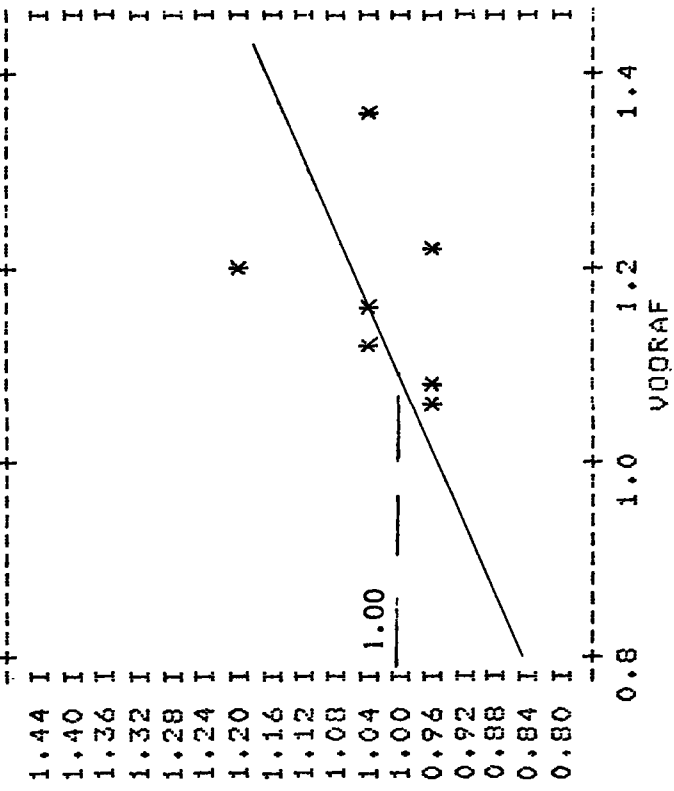


jaar 1

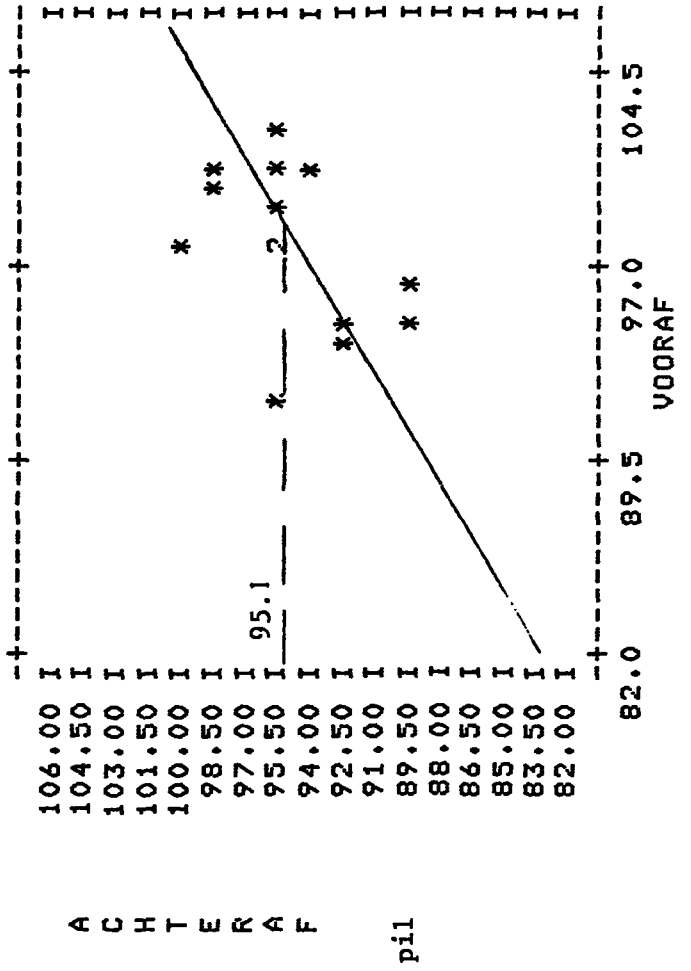
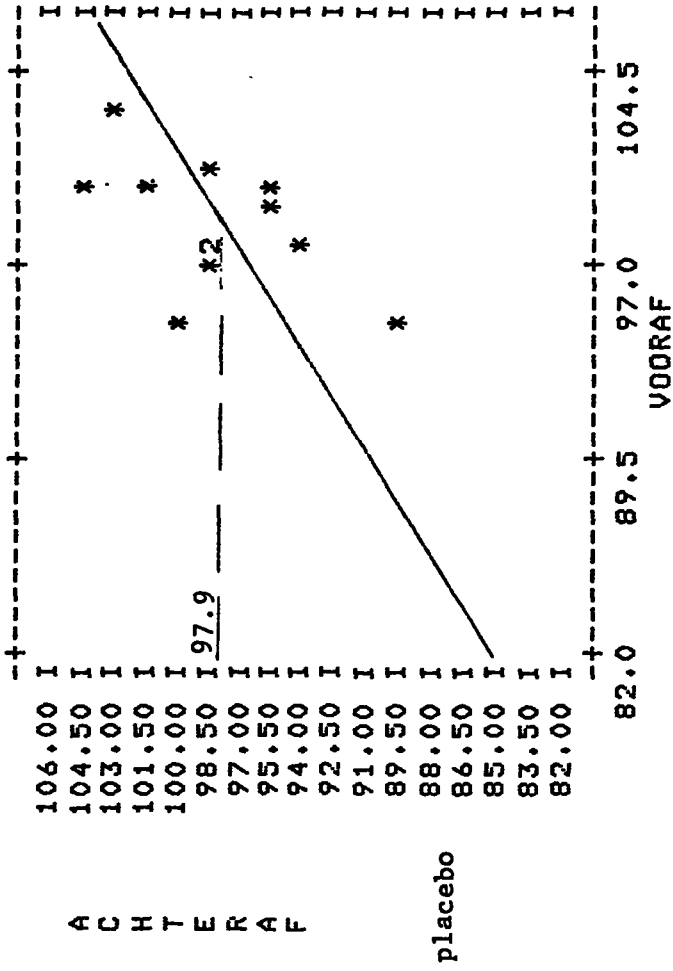


jongens

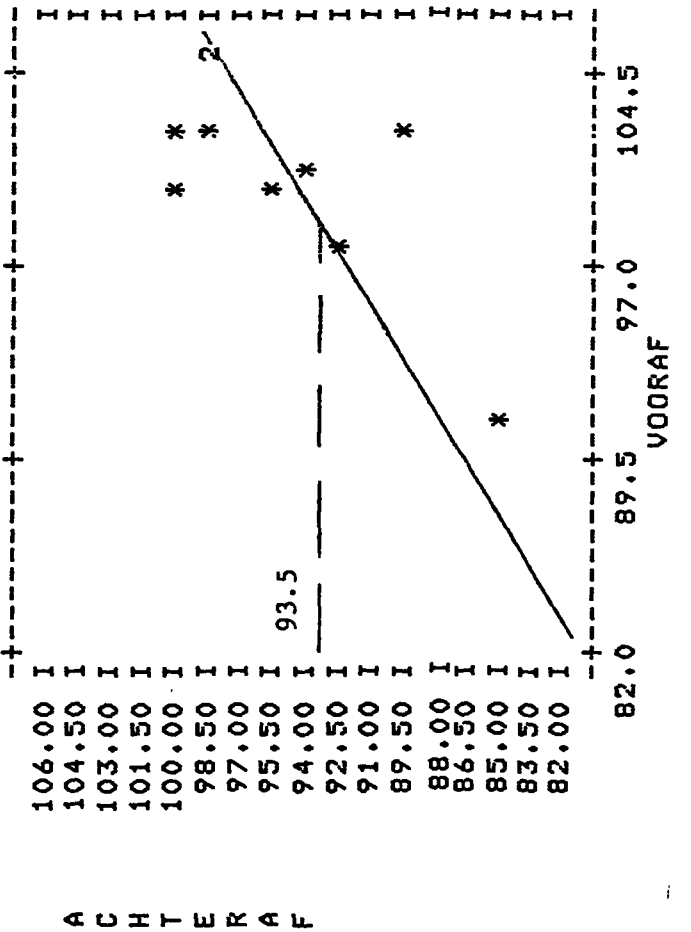
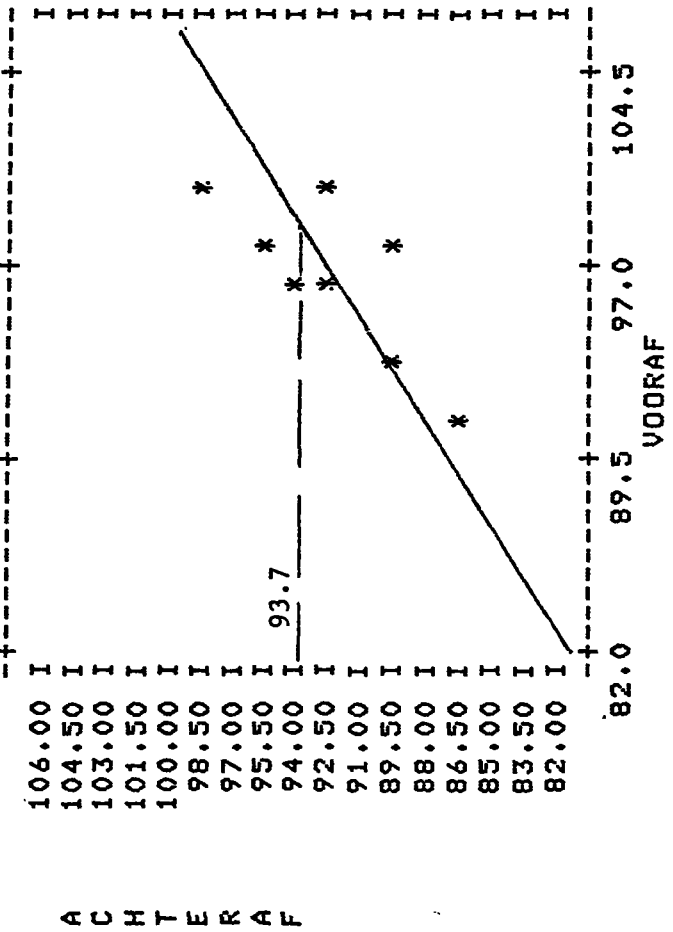
EGR



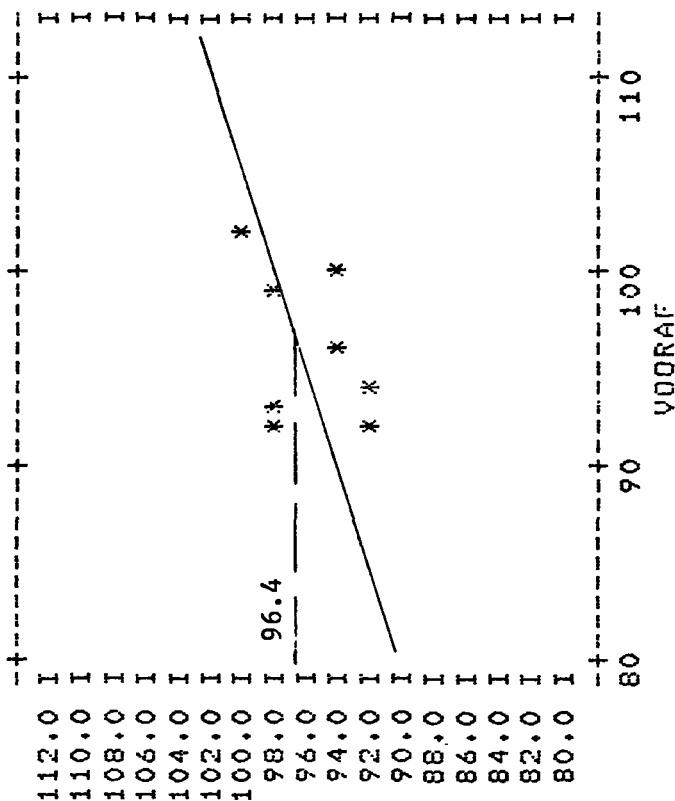
jaar 2



jaar 1



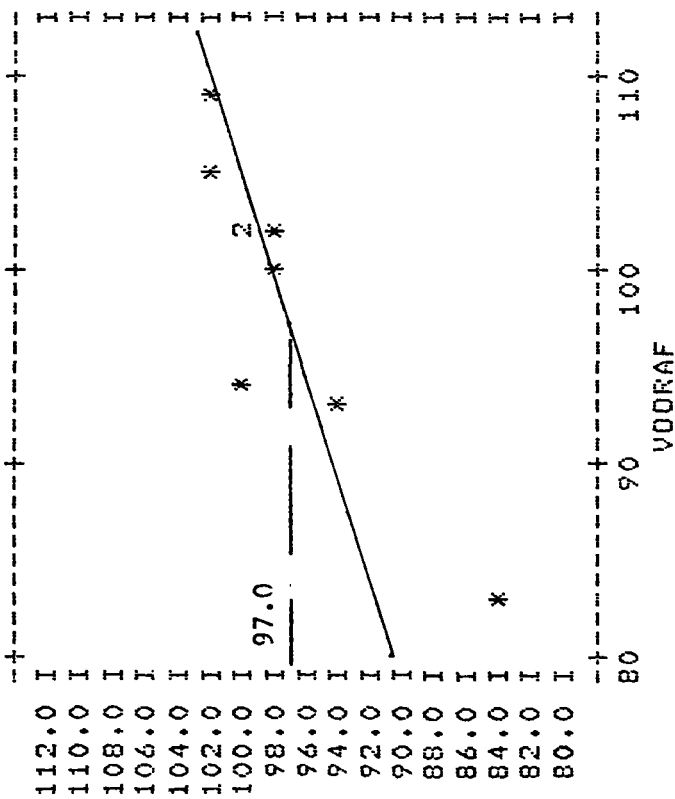
jaar 2



A C H T E R A F

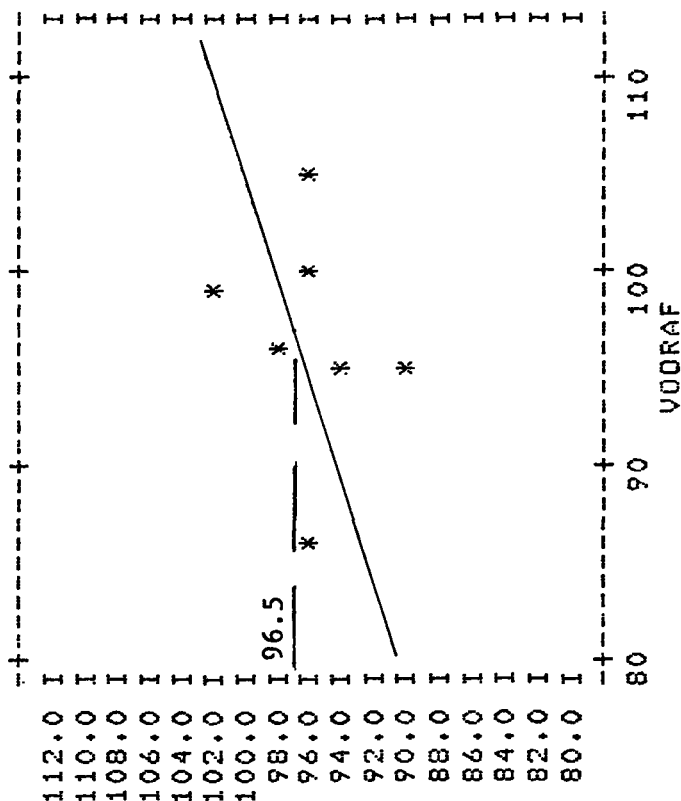
placebo

jaar 1



A C H T E R A F

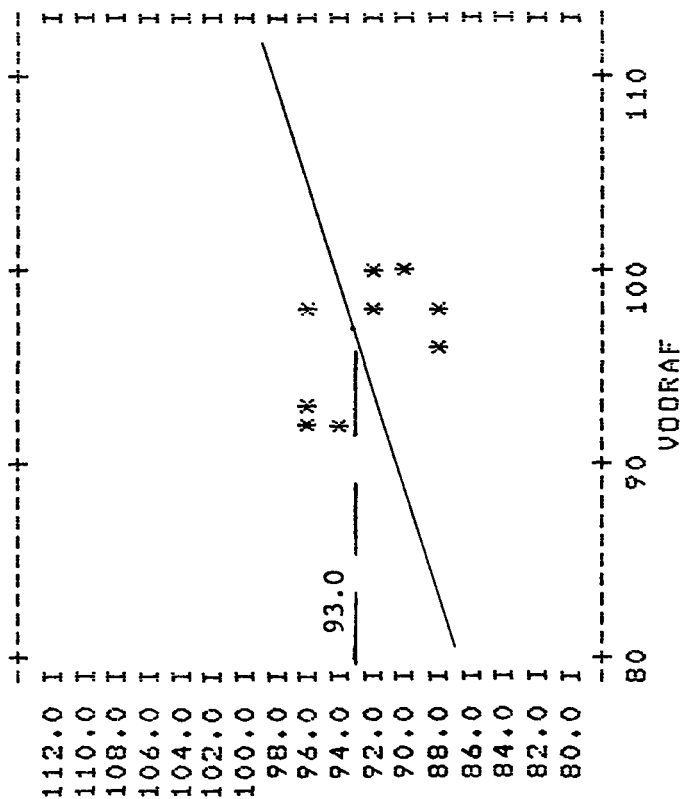
MCV



A C H T E R A F

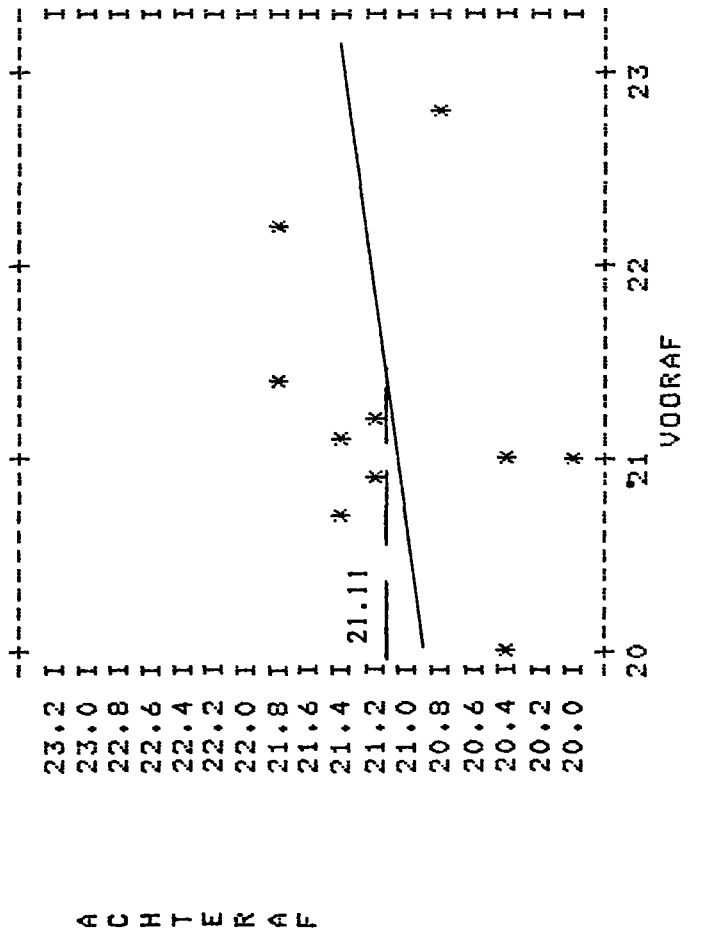
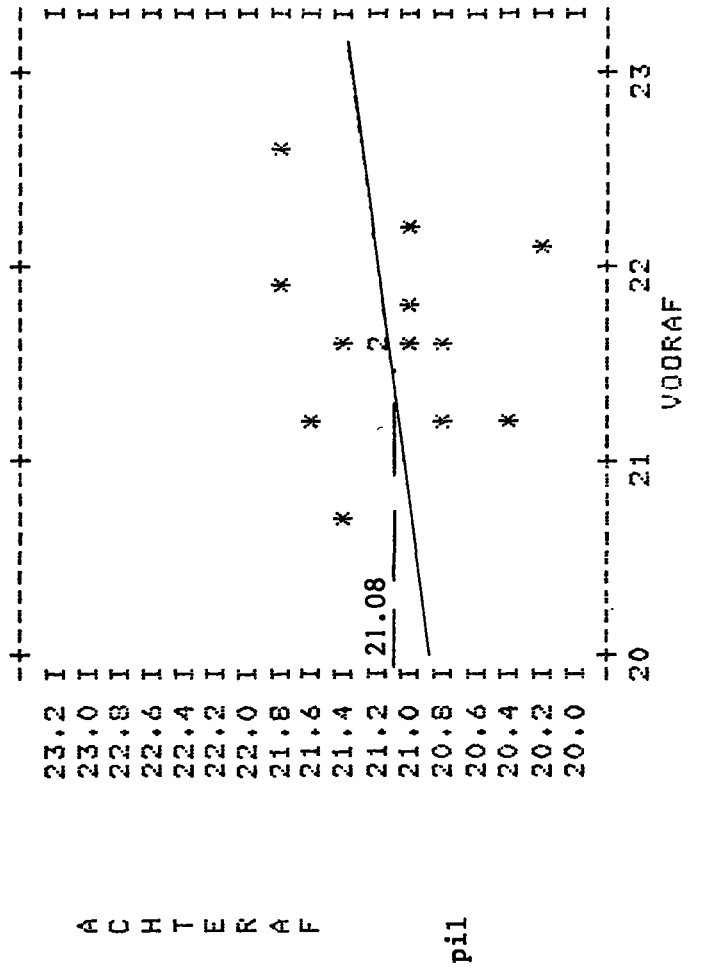
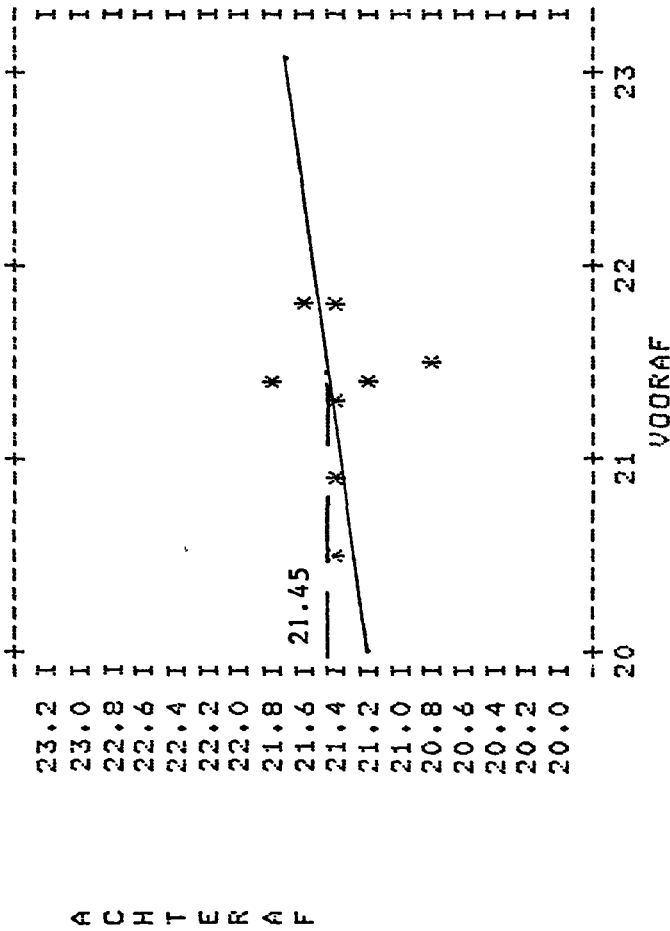
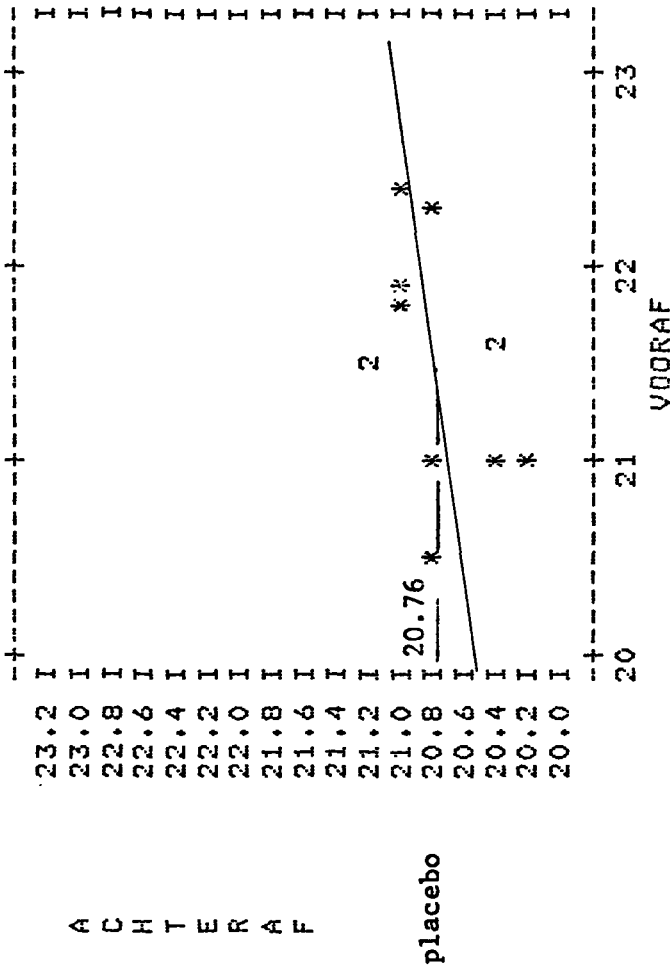
pil

jongens

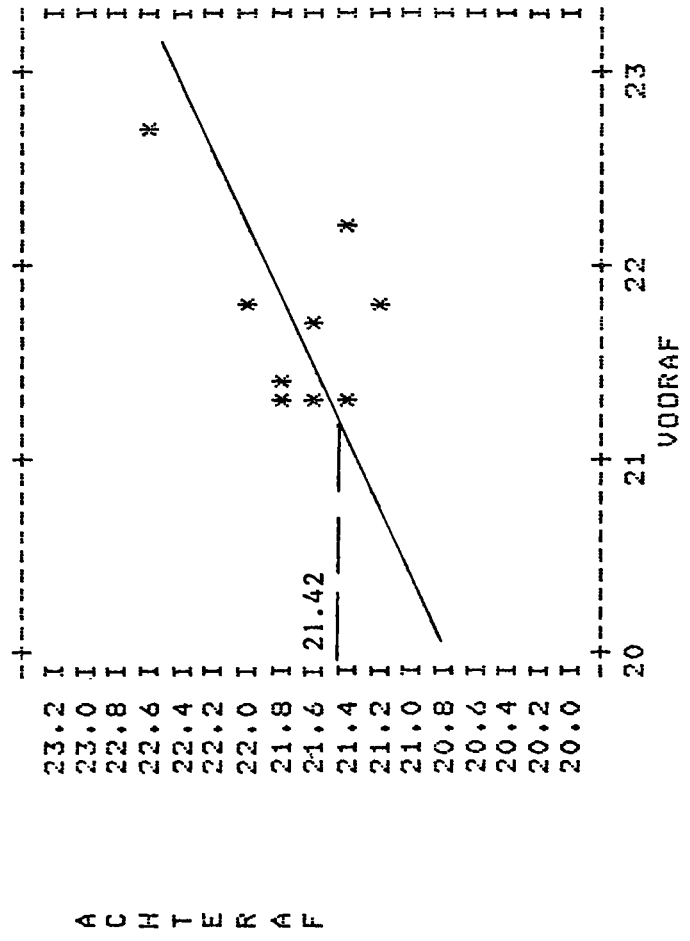
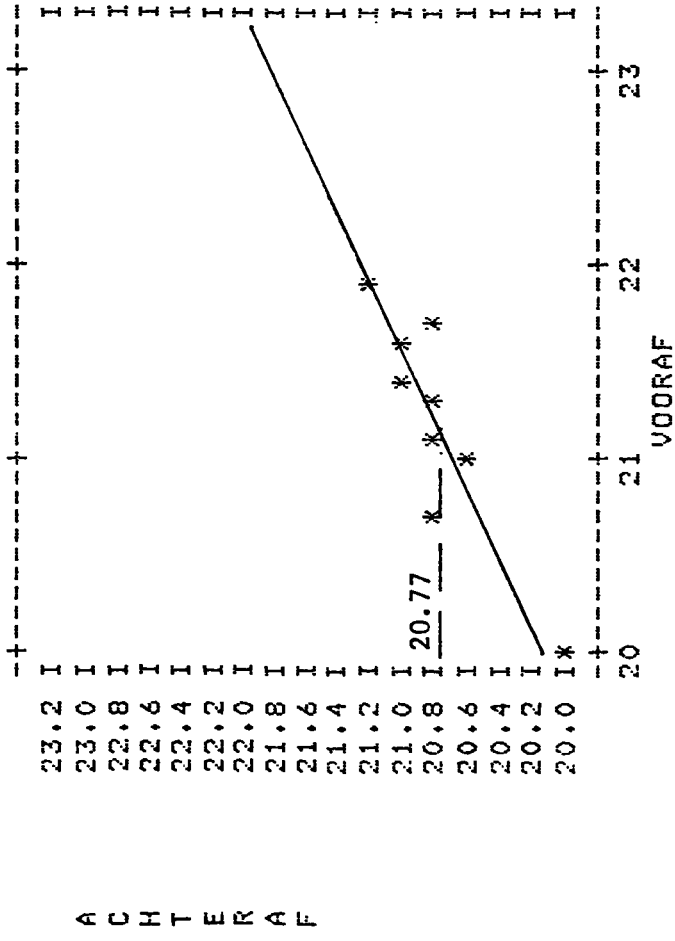
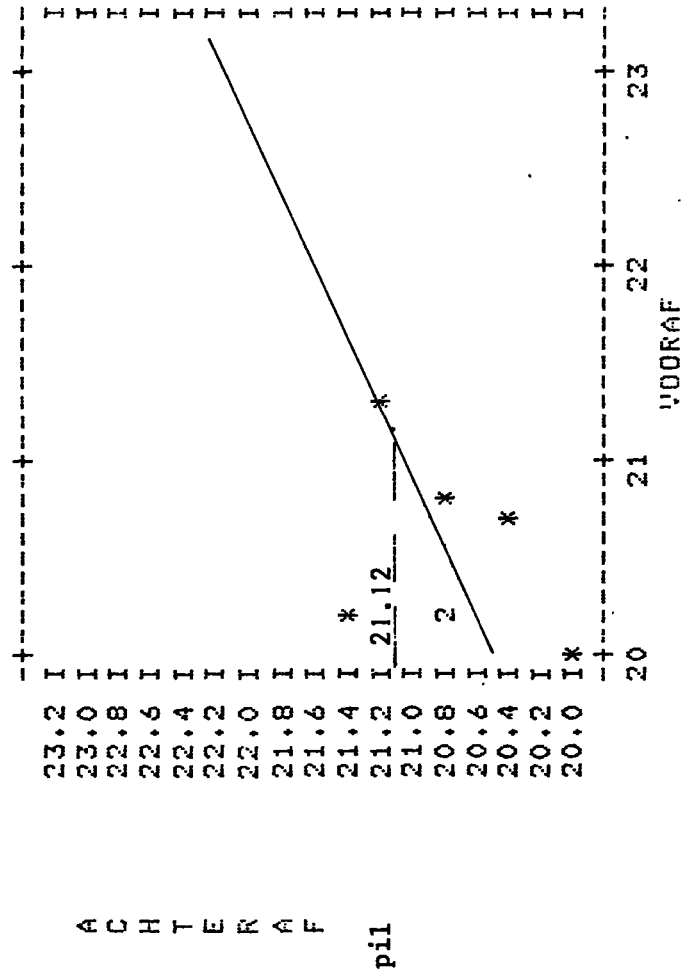
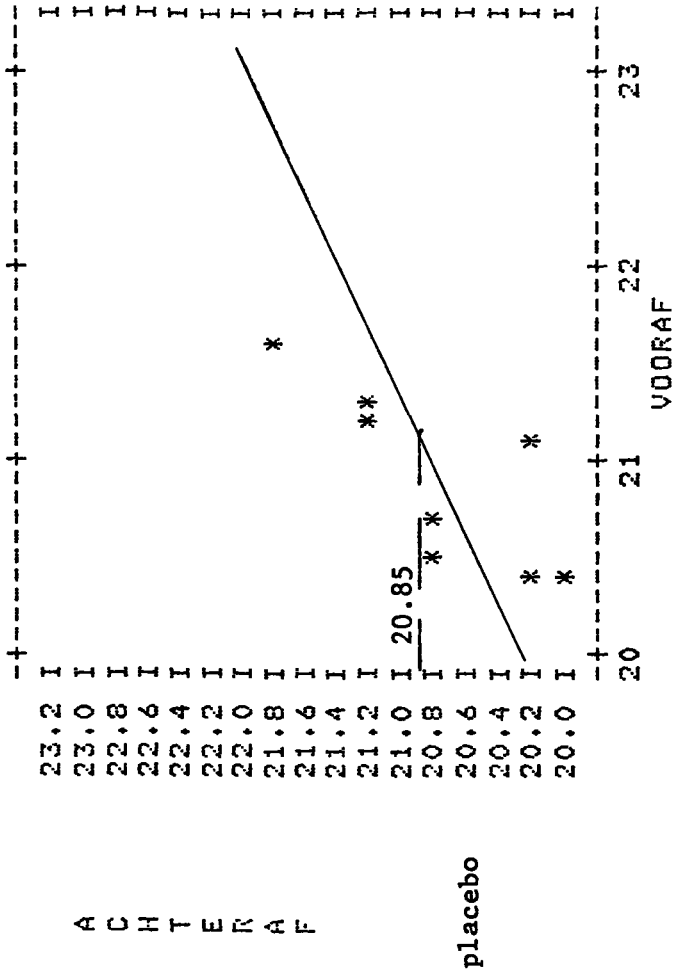


A C H T E R A F

jaar 2



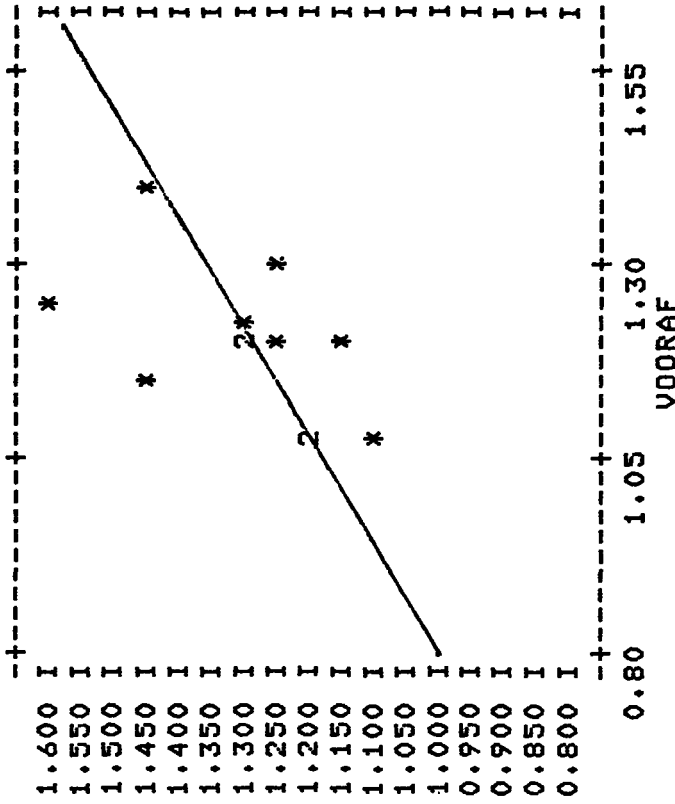
jaar 2



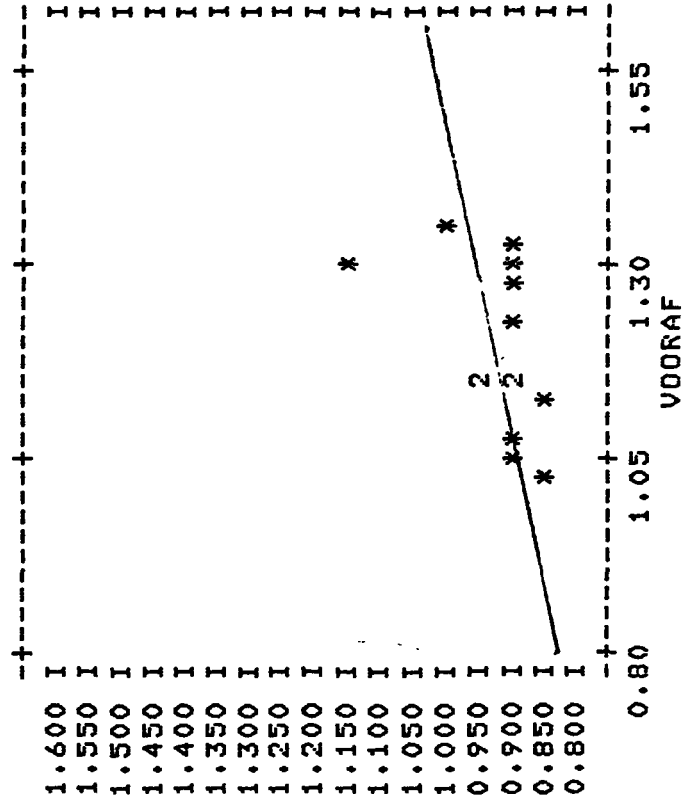
jongens

MCHC

jaar 2

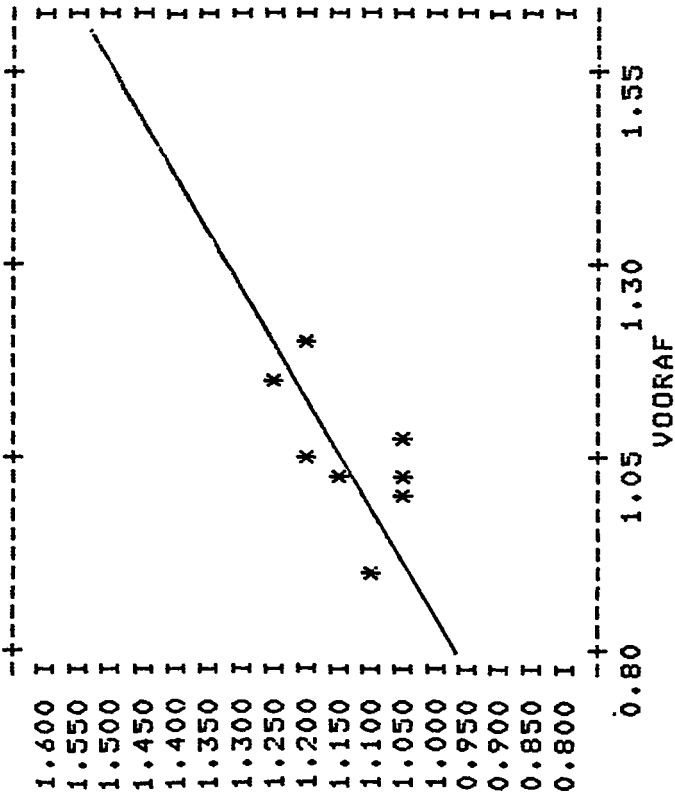


placebo

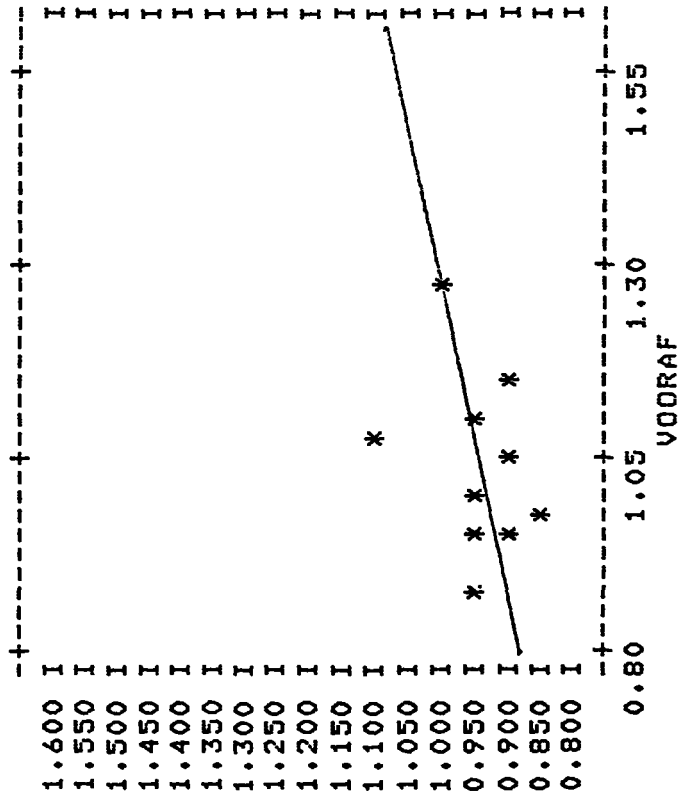


pil

jaar 1

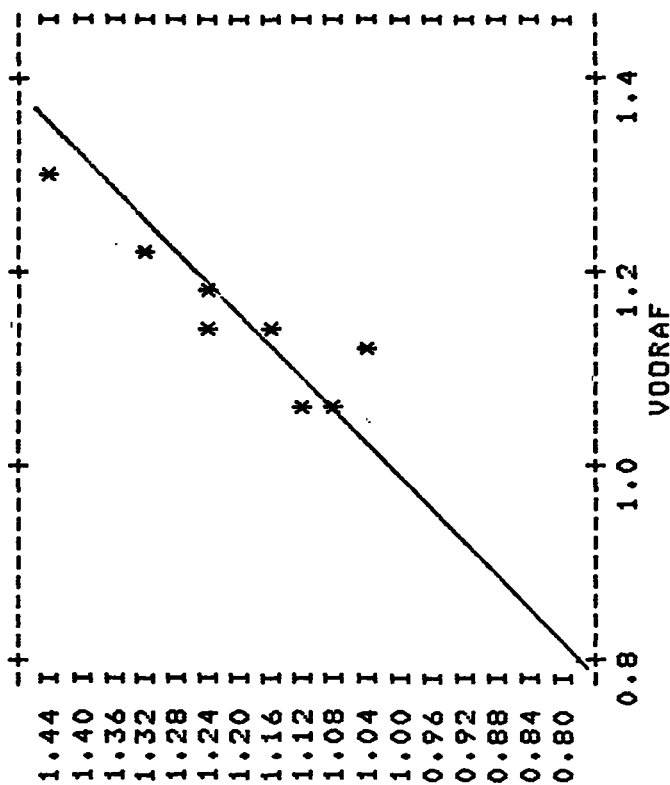


A C H T E R A F



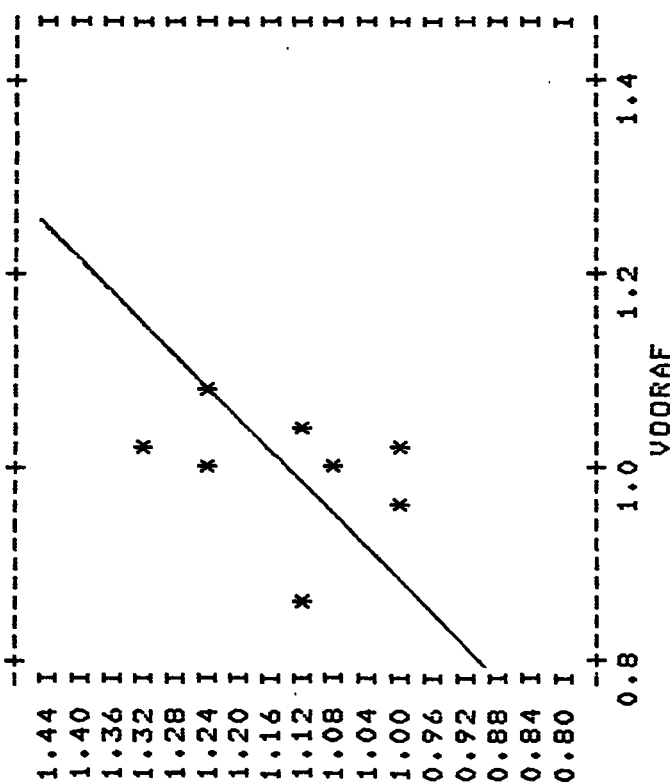
A C H T E R A F

jaar 2

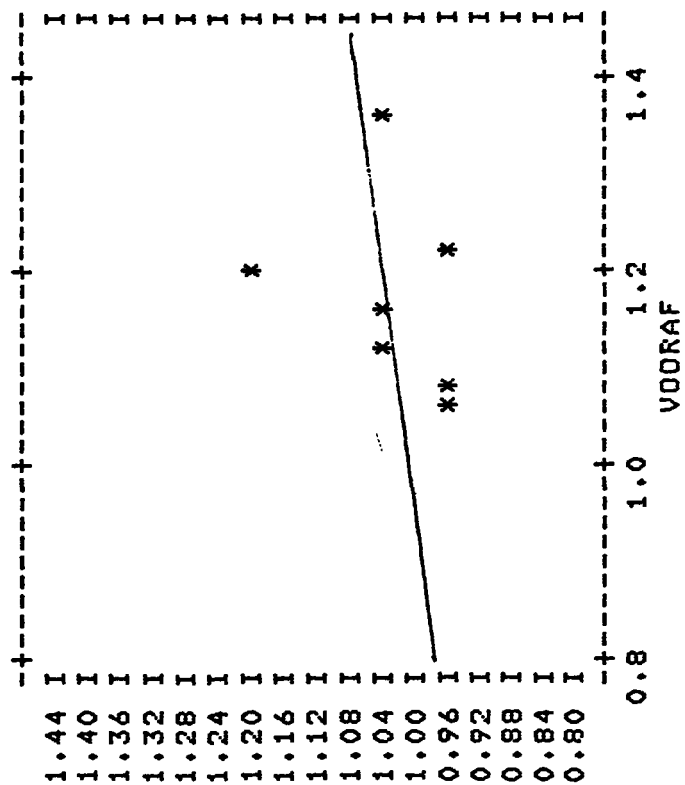


A C H T E R A F
placebo

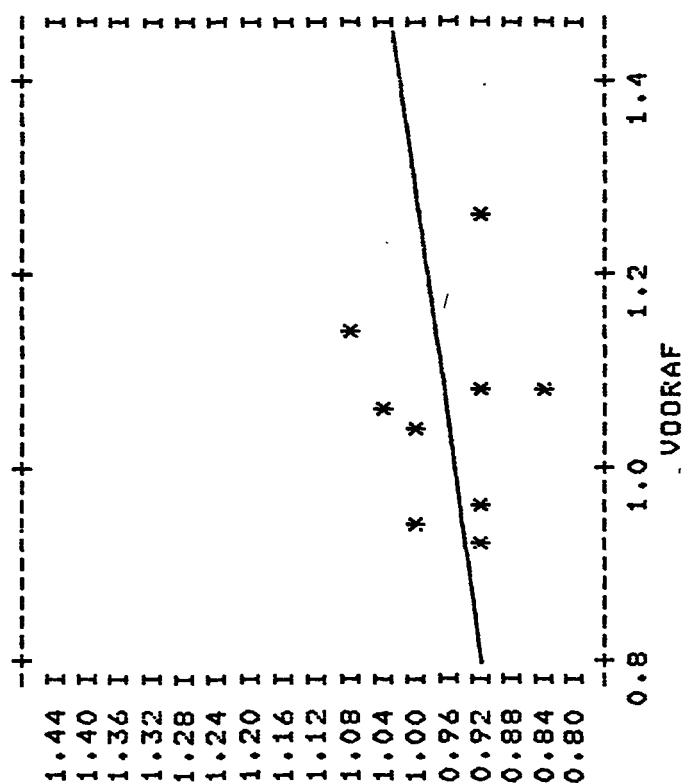
jaar 1



A C H T E R A F



A C H T E R A F
pil



A C H T E R A F

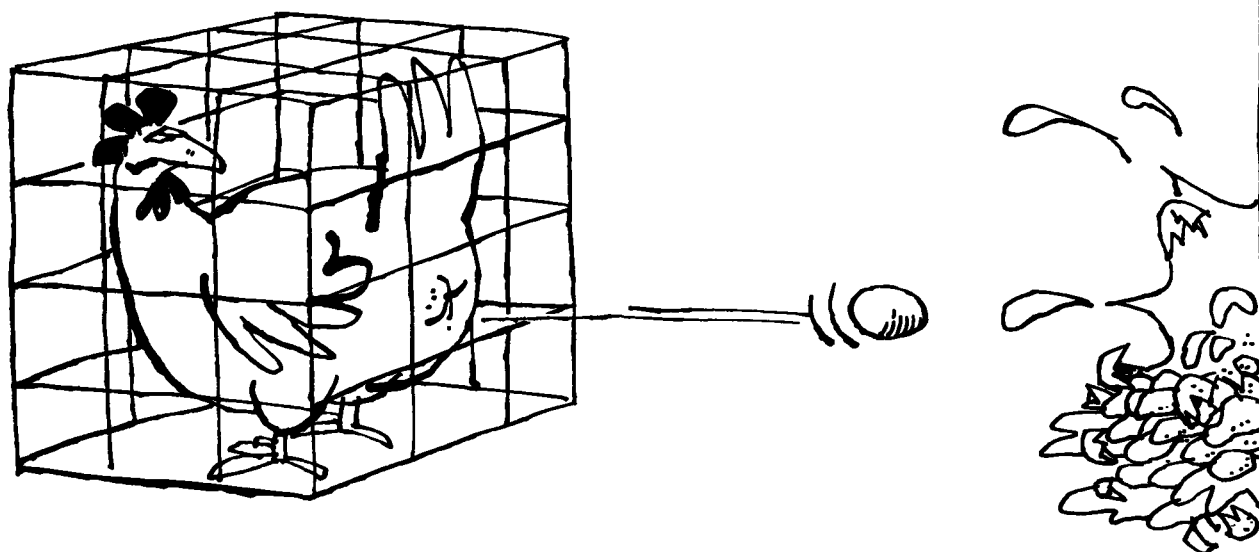
jongens

EGR

Eischaalbeschadigingen op legbatterijen

Ir. P. F. G. Vereijken

Rapport van de statistische begeleiding van project nr. 622
van het Instituut voor Pluimveeonderzoek "Het Spelderholt".



Inhoud

1.	Inleiding	205
2.	Te meten variabelen	206
3.	Proeffactoren	207
4.	De experimentele eenheid	209
4.1.	<i>Keuze van de experimentele eenheid</i>	209
4.2.	<i>De opstelling van de experimentele eenheden</i>	211
4.3.	<i>De experimentele eenheid in de tweede proef</i>	213
4.4.	<i>De plaatsing van de experimentele eenheden in de tweede proef</i>	214
5.	Proefopzetten	216
5.1.	<i>De omvang van de proef</i>	216
5.2.	<i>Te onderzoeken factoren</i>	218
5.3.	<i>Randvoorwaarden</i>	218
5.4.	<i>Opzetten in geval de kooibodems gekoppeld zijn</i>	219
5.4.1.	<i>Fractionele proeven</i>	220
5.4.2.	<i>2ⁿ proeven</i>	220
5.5.	<i>Opzetten in geval de kooibodems niet gekoppeld zijn</i>	222
5.5.1.	<i>Fractionele proeven</i>	222
5.5.2.	<i>2ⁿ proeven</i>	223
5.6.	<i>Keuze van de opzet</i>	225
6.	Bemonstering	225
7.	Verbeteringen ten opzichte van de opzet voor de eerste proef	226

1. Inleiding

De fractie beschadigde eieren van leghennen op batterijkooien kan aanzienlijk variëren. Daar een beschadigd ei gemiddeld 2.5 à 3 cent opbrengstverlies betekent, is de fractie beschadigde eieren voor de pluimveehouder een economisch zeer belangrijke grootheid. Daarom wordt getracht de fractie beschadigde eieren te verkleinen door de eischaal te verbeteren en door te onderzoeken hoe de eieren op de kooibodem voorzigtiger behandeld kunnen worden. Bij bestudering van het batterijsysteem is namelijk gebleken dat het merendeel van de beschadigingen plaatsvindt op de kooibodem en in de eierlade. Voorts kunnen eieren worden beschadigd tijdens het verzamelen en het verpakken op de pluimveebedrijven, tijdens het transport en op de pakstations waar de eieren gesorteerd worden. Batterijkooien zijn opgebouwd uit gaasconstructies, die zijn opgehangen in metalen frames.

De beschadiging op de kooi ontstaat bij de val van het ei op de kooibodem en tijdens het afrollen van het ei naar de eierlade. In de eierlade ontstaan de beschadigingen als gevolg van het botsen tegen de rand van de eierlade en het botsen van de eieren onderling.

Op het Instituut voor Pluimveeonderzoek "Het Spelderholt" wordt in samenwerking met Big Dutchman (een fabrikant van legbatterijen) de fractie beschadigde eieren bij een aantal verschillende kooitypen onderzocht. Deze kooitypen verschillen van elkaar met betrekking tot de ingestelde waarden van een aantal factoren welke direct of indirect met de kooiconstructie te maken hebben, zoals de draaddikte en de maaswijdte van de kooibodem. Op "Het Spelderholt" is in een eerder uitgevoerd onderzoek een aantal kooitypen vergeleken met het doel inzicht te verkrijgen in de effecten van een aantal kooiconstructie-factoren op de fractie beschadigde eieren over een proefperiode van een jaar. Deze proefperiode begon op het moment dat de produktie, gedefinieerd als het gemiddeld aantal eieren per hen per dag \times 100 procent, boven de 50 procent lag. In deze proef varieerde de fractie beschadigde eieren over de totale proefperiode voor de verschillende factorcombinaties van 0.056 tot 0.149. Aangezien in deze proef slechts een beperkt aantal factoren is onderzocht, wordt momenteel een vervolgonderzoek gepland. Daarin worden wederom de effecten van een aantal factoren onderzocht.

Getracht wordt in samenwerking met de onderzoekers tot een opzet te komen, waarin de instelling van zoveel mogelijk factoren wordt gevarieerd; van elke factor zullen twee instellingen worden beproefd. Sommige afhankelijke variabelen, te weten het aantal geraapte eieren (hierin begrepen de beschadigde en de vuilschalige eieren) en de aantallen beschadigde-, vuilschalige- en door de kooibodem getrapte eieren, zullen niet dagelijks worden gemeten. Dit heeft tot gevolg, dat bij de gekozen opzet een bemonsteringsschema dient te worden gemaakt.

Op de problemen die vastzitten aan de keuze van de experimentele eenheid en aan de plaatsing van de experimentele eenheden van de diverse kooitypen in de proefstal wordt nader ingegaan. Verder wordt besproken welke proefopzetten voor het onderzoek zijn overwogen en in welk opzicht deze opzetten een verbetering zijn, vergeleken met de opzet van de reeds eerder uitgevoerde proef. Tenslotte wordt globaal het bemonsteringsschema beschreven en wordt aangegeven welke verbeteringen dit schema heeft, vergeleken met het schema van de reeds uitgevoerde proef. In het vervolg wordt het eerder uitgevoerde onderzoek aangeduid met "de eerste proef" en de nog uit te voeren proef met "de tweede proef".

2. Te meten variabelen

In het onderzoek gaat de interesse uit naar de volgende variabelen:

- 2.1. het aantal geraapte eieren (x), hierin begrepen (y) en (v),
- 2.2. het aantal beschadigde eieren (y),
- 2.3. het aantal vuilschalige eieren (v),
- 2.4. het aantal door de kooibodem getrapte eieren (d); deze variabele kan alleen bij de Compactbatterij worden gemeten,
- 2.5. het aantal ten gevolge van ziekte en ongevallen gestorven dieren.

Voor de tweede proef start de proefperiode op het moment dat de productie (zie paragraaf 1) boven de 50% ligt en de totale proef zal wederom ongeveer een jaar duren. Aan elk van de experimentele eenheden zullen op 30 tot 40 dagen verspreid over de totale proefperiode de variabelen 2.1. tot en met 2.4. worden gemeten aan de gehele dagproducties. Uit de steekproefuitkomsten voor deze variabelen zal de waarde over de totale proefperiode worden geschat. Op het steekproefschema wordt in paragraaf

6 ingegaan. De verkregen schattingen \hat{X} , \hat{Y} , \hat{V} en \hat{D} voor de waarden van de variabelen 2.1. t/m 2.4. over de totale proefperiode: X, Y, V en D worden gebruikt om de volgende grootheden te schatten:

- de fractie beschadigde eieren over de totale proefperiode ($\frac{Y}{X}$),
- de fractie vuilscalige eieren over de totale proefperiode ($\frac{V}{X}$) en
- de fractie doorgetrapte eieren over de totale proefperiode ($\frac{D}{X}$)

De effecten van de beproefde factoren op bovengenoemde afgeleide variabelen zullen worden geanalyseerd. Voor het vaststellen van de gewenste omvang van de proef is de precisie van $\frac{Y}{X}$ bepalend.

3. Proeffactoren

In de eerste proef waren zeven factoren (= in te stellen variabelen) opgenomen, die invloed kunnen hebben op de hoogte van de fractie beschadigde eieren. Deze zeven factoren en de beproefde niveaus (= in te stellen waarden) zijn:

a) kooiconstructie-factoren

- 3.1. helling van de kooibodem (7° , 9° of 11°),
- 3.2. draaddikte van de bodem (2.05 mm of 2.45 mm),
- 3.3. maaswijdte van de bodem (1x1.5, 1x2 of 1x3 inch),
- 3.4. type scheidingswand (gaas of plaat),
- 3.5. coating van de bodem (epoxy of verzinkt),

b) de factor ras

- 3.6. lichte (witte) of middelzware (bruine) dieren,

c) de factor kooibezetting

- 3.7. aantal dieren in elk van de 24 kooien van een kooiblok (5 of 6), zie figuur 1 in paragraaf 4.

In de eerste proef zijn van de in totaal $3 \times 2 \times 3 \times 2 \times 2 \times 2 \times 2 = 288$ mogelijke combinaties van factorniveaus er 28 daadwerkelijk beproefd. De geselecteerde combinaties zijn de 18 mogelijke combinaties van de niveaus van de factoren: helling van de kooibodem, draaddikte van de bodem en maaswijdte van de bodem bij een geselecteerde instelling van elk van de overige factoren. Daarboven werden nog tien factorcombinaties genomen met het doel op grond van de proefuitkomsten de beste van deze tien (dat wil zeggen de combinatie(s) met de kleinste fractie beschadigde eieren) aan te wijzen. Van elk van deze 28 kooitypen is een viertal kooiblokken beproefd. Op de gevolgde proefopzet wordt nog nader ingegaan.

In de tweede proef zijn naast de reeds genoemde proeffactoren de volgende vier factoren overwogen:

d) kooiconstructiefactoren

3.8. ophanging van de bodem; in de eerste proef was er naar later bleek, ten aanzien van de bevestiging van de kooibodem een verschil tussen kooien met een scheidingswand van plaat en die met een scheidingswand van gaas. Bij de eerste soort was de bodem door middel van twee plastic clips aan elke scheidingswand bevestigd, terwijl bij de tweede soort de bodem met zeven draadverbindingen aan elke scheidingswand werd bevestigd. Dit betekent dat een scheidingswandeffect gestrengeld was met een effect als gevolg van de wijze van ophanging. De onderzoekers zijn van mening dat het gevonden totaaleffect voornamelijk wordt veroorzaakt door de wijze van ophanging. Dit is de reden dat in de tweede proef twee verschillende wijzen van ophanging zullen worden onderzocht.

3.9. ondersteuning van de bodem (wel of geen beugel).

3.10. wel of niet aanbrengen van een eiersluis (mechaniek om de snelheid waarmee de eieren in de eierlade aankomen te verminderen).

e) de factor batterijsysteem

3.11. Universa-batterij of Compact-batterij; de belangrijkste reden voor het onderzoeken van twee batterijsystemen is het vermoeden van de onderzoeker dat de in de eerste proef voor de Universa-batterij gevonden effecten van de kooiconstructiefactoren wel eens specifiek voor dat batterijsysteem kunnen gelden. Daarnaast kunnen bij de Compact-batterij voor de diverse kooitypen gemakkelijk de door de kooibodem getrapte eieren worden geteld, hetgeen bij de Universa-batterij onmogelijk is.

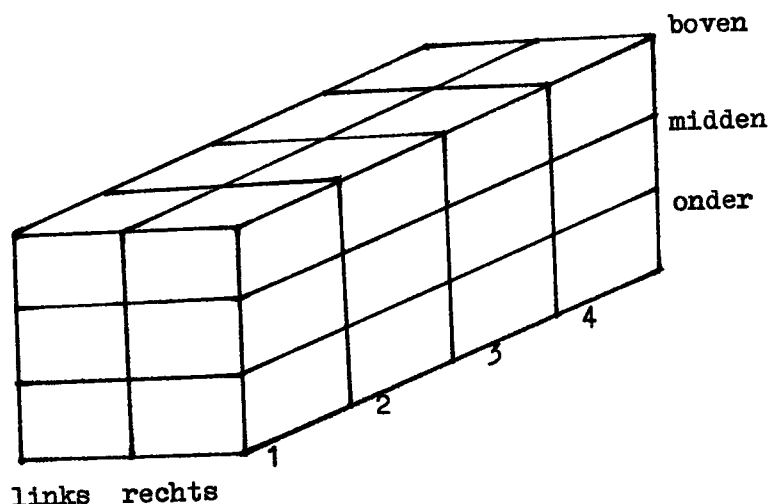
Voor de tweede proef zijn de niveaus van 11^0 voor de bodemhelling en die van 1×3 inch voor de maaswijdte op technische gronden niet langer interessant. Verder zullen in de helft van de kooien vier dieren en in de andere helft vijf dieren worden geplaatst. Dit houdt in dat voor de tweede proef in totaal $2^{11} = 2024$ verschillende combinaties van factorniveaus mogelijk zijn; een aantal dat hoog is, vergeleken met de omvang van de proef, zoals later zal blijken.

4. De experimentele eenheid

4.1. Keuze van de experimentele eenheid

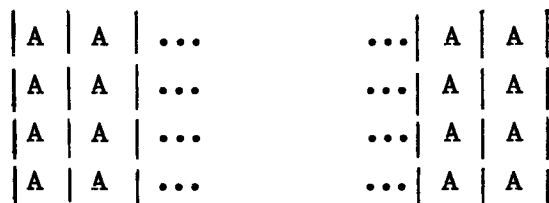
In het onderzoek naar schaalbreuk wordt onder de experimentele eenheid verstaan de kleinste groep van vijf-henskooien waaraan een willekeurige combinatie van proeffactoren kan worden toegewezen, waaraan afzonderlijke metingen worden gedaan en waarop de berekeningen worden gebaseerd. Een belangrijk facet, dat mede door de keuze van experimentele eenheid wordt bepaald, is de mate waarin de verkregen resultaten overdraagbaar zijn naar de praktijk. In dit kader is het belangrijk te weten of schaalbreuk op één enkele vijf-henskooi alleen veroorzaakt wordt door de kooi zelf en de dieren in de kooi, of tevens door de invloed van buurkooien. De keuze van een geschikte experimentele eenheid is bepalend voor het generalisatieniveau van het onderzoek.

Big Dutchman past de te onderzoeken kooi toe in zogenaamde kooiblokken, waarvan er één in figuur 1 schematisch is weergegeven.



Figuur 1: Schematische weergave van een kooiblok, bestaande uit $2(\text{links/rechts}) \times 3(\text{etages}) \times 4(\text{kooien}) = 24$ kooien.

In de praktijk worden de kooiblokken naast elkaar in rijen opgesteld in een stal. In figuur 2 is dit schematisch weergegeven voor kooiblokken A; hierbij stellen de verticale streepjes tussen twee kooiblokken frames voor, waartussen de kooiblokken zijn bevestigd.



Figuur 2: Schematische praktijkopstelling van kooiblokken van één type in een leghennenstal

In de eerste proef is als experimentele eenheid één kooiblok genomen. Overwogen dient te worden of een kooiblok de meest geschikte experimentele eenheid is. Omdat niet bekend is of de breuk in vijf-henskooien uitsluitend door de kooi zelf of tevens door buurkooien bepaald wordt, moet hierover eerst meer informatie worden vergaard. Door experimentele eenheden van een groot aantal kooiblokken te nemen,

zal het effect van buurkooien niet of nauwelijks worden gevoeld. Het probleem is echter dat, ter beperking van de proefomvang, de experimentele eenheden niet groter moeten worden genomen dan strikt noodzakelijk is. Door te onderzoeken of er buurkooi-effecten optreden en zo ja, tot hoever deze reiken, kan de aanvankelijk als "veilig" gekozen experimentele eenheid mogelijk aanzienlijk worden verkleind. Doordat bij een vooraf gegeven aantal vijf-henskooien het aantal experimentele eenheden groter wordt naarmate de experimentele eenheid kleiner wordt, kunnen meer combinaties en/of meer proeffactoren worden onderzocht dan bij onnodig grote experimentele eenheden het geval is. Als verscheidene proeven met dezelfde kooien zouden worden gedaan, kan bij de opzet van elke proef ruim aandacht worden geschonken aan het vergaren van informatie om in volgende proeven op verantwoorde wijze tot kleinere experimentele eenheden te komen. Een andere mogelijkheid is om door middel van onderzoek in de praktijk meer te weten te komen over buurkooi-effecten en wellicht op die manier tot kleinere experimentele eenheden te kunnen concluderen. Omtrent het laatstgenoemde punt zijn in de discussies met de onderzoekers mogelijkheden naar voren gebracht, waarop hier niet nader wordt ingegaan.

4.2. De opstelling van de experimentele eenheden

In de eerste proef zijn de verschillende kooiblokken (= experimentele eenheden) naast elkaar in rijen in de proefstal geplaatst (zie figuur 3).



Figuur 3: Opstelling van de kooiblokken in de eerste proef

Een probleem waar destijds niet aan gedacht is, is het feit dat als gevolg van het koppelen van verschillende kooitypen de fractie bescha-

digde eieren voor een kooiblok van type A, dat direct verbonden is met kooiblokken van typen B en C, zoals in de proefsituatie, wel eens anders zou kunnen zijn dan wanneer datzelfde kooiblok van type A direct verbonden zou zijn met kooiblokken eveneens van type A, zoals in de praktijk het geval is. Wanneer het bovengenoemde effect optreedt kan de feitelijk gemeten fractie beschadigde eieren afwijken van de fractie beschadigde eieren die men had willen meten. Feitelijk dient te worden onderzocht of deze zogenaamde overgangseffecten, als gevolg van het koppelen van verschillende kooitypen, al dan niet bestaan en zo ja, welke maatregelen kunnen worden genomen om deze ongewenste effecten zoveel mogelijk te vermijden dan wel te elimineren in de analyse van de fractie beschadigde eieren.

In de discussie met de onderzoekers zijn de volgende mogelijkheden genoemd om de hinder van ongewenste overgangseffecten voor de analyse te verminderen.

- modelleren van de overgangseffecten

Omdat elk kooitype een combinatie is van ingestelde waarden van een groot aantal factoren, zijn er vermoedelijk veel parameters vereist om deze overgangseffecten te modelleren. In dat geval biedt deze mogelijkheid geen bruikbare oplossing.

- de experimentele eenheden voorzien van een aantal "randkooien"

Het is mogelijk om aan linker- en rechterzijde van elke experimentele eenheid groepen "randkooien" aan te brengen, bestaande uit 2 (links/rechts) x 3 (etages) = 6 kooien. De aldus uitgebreide eenheden worden dan naast elkaar in een rij geplaatst. De idee is dat de overgangseffecten van twee gekoppelde uitgebreide eenheden niet verder reiken dan de aangebrachte groepen randkooien. Op de randkooien worden weliswaar leghennen geplaatst, doch de waarnemingen aan de randkooien worden niet gebruikt bij de analyse van de proefgegevens. Wel kunnen zij op zich interessante informatie leveren.

- het tussenplaatsen van buffers op bepaalde plaatsen

De onderzoekers verwachten dat niet elk tweetal gekoppelde kooitypen overgangseffecten geeft. Vanuit die gedachte worden de

proeffactoren opgedeeld in twee groepen.

- 1) Een groep van factoren waarvoor kooitypen, die verschillen in de ingestelde waarde van één van deze factoren, bij koppeling géén overgangseffecten geven.
- 2) Een groep van factoren waarvan kooitypen, welke verschillen in de ingestelde waarde van één van deze factoren, bij koppeling vermoedelijk wél overgangseffecten geven (bijvoorbeeld de helling van de bodem).

Voorgesteld is om de kooitypen welke geacht worden geen overgangseffecten te geven bij onderling koppelen, zoveel mogelijk bij elkaar te plaatsen. Daar waar twee kooitypen naast elkaar staan die vermoedelijk wel overgangseffecten geven, zullen dan buffers worden geplaatst. Uiteraard beperkt dit de vrijheid van het toewijzen van de diverse kooitypen aan de experimentele eenheden.

- het zonodig gescheiden ophangen van de kooiblokken

Op plaatsen waar overgangseffecten worden verwacht, worden de kooiblokken van elkaar gescheiden opgehangen; door middel van technische voorzieningen zal eventueel worden gecorrigeerd voor het ontbreken van buurblokken (bijvoorbeeld corrigeren voor de bodembelasting van de ontbrekende buurblokken, zoals in de praktijk). Voor het geval overgangseffecten, als gevolg van het aan elkaar koppelen van verschillende kooitypen inderdaad optreden, is de enige correcte oplossing elke experimentele eenheid te voorzien van een "rand", die groot genoeg is. Omdat dit vermoedelijk nogal wat extra vijf-henskooien gaat kosten, is gepoogd andere oplossingen aan te dragen. Van deze oplossingen mag echter hoogstens worden verwacht dat de invloed van de overgangseffecten op de te meten fracties beschadigde eieren wordt verkleind.

4.3. *De experimentele eenheid in de tweede proef*

In de tweede proef zal wederom een kooiblok als experimentele eenheid worden genomen omdat:

- a) een kooiblok door Big Dutchman als technische eenheid wordt gemaakt,

- b) een kooiblok als eenheid tussen twee frames wordt opgehangen,
- c) de overwogen alternatieven minder geschikt leken.

Als alternatieven zijn overwogen een experimentele eenheid bestaande uit twee of drie aaneengekoppelde kooiblokken of experimentele eenheden bestaande uit halve kooiblokken; daarbij is gedacht aan zowel een linker- of een rechterhelft, als aan een voorste of achterste helft van een kooiblok.

Experimentele eenheden bestaande uit twee of drie kooiblokken zijn niet genomen. Weliswaar zou men hierbij per experimentele eenheid een betere schatting voor de fractie beschadigde eieren krijgen, maar bij het vooraf gegeven aantal vijf-henskooien voor het vervolgonderzoek zou het aantal experimentele eenheden te klein worden. Evenmin zijn experimentele eenheden bestaande uit halve kooiblokken genomen. De redenen hiervoor waren dat enerzijds de onderzoekers bevreesd zijn voor overgangseffecten binnen een kooiblok en anderzijds het feit dat in de praktijk niet met halve kooiblokken wordt gewerkt. De onderzoekers vrezen dat de aan halve kooiblokken verkregen resultaten niet overdraagbaar zijn naar de praktijk.

4.4. De plaatsing van de experimentele eenheden in de tweede proef

De onderzoekers hebben besloten in de tweede proef niet op de in paragraaf 4.2. beschreven punten in te gaan. Aangenomen wordt dat de richting en de grootte van de diverse effecten niet wezenlijk beïnvloed worden door het optreden van een aantal overgangseffecten. Er zullen daarom geen kooien van het vooraf gegeven aantal vijf-henskooien worden gebruikt voor het aanbrengen van randen rond de experimentele eenheden of voor het plaatsen van een aantal buffers.

Om het effect van eventuele overgangseffecten, als gevolg van het koppelen van twee verschillende kooitypen, te verkleinen is echter het volgende overwogen.

4.4.1. Teneinde de spanning in de constructie te houden zouden in eerste instantie de kooibodems van naast elkaar opgestelde kooiblokken worden doorverbonden. Omdat het koppelen van een kooiblok met een bodemhelling van 7° aan een kooiblok met een bodemhelling van 9° technische problemen geeft, dienen binnen elke batterijrij kooiblokken met eenzelfde

bodemhelling zoveel mogelijk bij elkaar te worden geplaatst.

4.4.2. In tweede instantie zijn de onderzoekers teruggekomen op het aanvankelijk genomen besluit de kooibodems van naburige kooiblokken te verbinden. In plaats daarvan zal elk kooiblok worden opgehangen aan twee frames en zullen de kooiblokken met frames op een onderlinge afstand van 10 cm worden geplaatst in een batterijrij.

In figuur 4 is de opstelling van de kooiblokken in een rij schematisch weergegeven in de onder 4.4.1. beschreven situatie en in de nieuwe situatie.



Figuur 4: Plaatsing van de kooitypen in een kooibatterijrij. Links de situatie van punt 4.4.1. en rechts de situatie van punt 4.2.2. Ieder dwarsstreepje stelt een frame voor.

Aangezien de kooiblokken niet direct gekoppeld zijn, verwachten de onderzoekers geen overgangseffecten meer. In verband met de overdraagbaarheid van de proefresultaten naar de praktijk dienen de gemiddelde fracties beschadigde eieren voor de diverse kooitypen, in de situatie waarin kooiblokken van één kooitype onderling verbonden zijn (de praktijk) en in de situatie waarin de kooiblokken van hetzelfde kooitype tussen dubbele frames worden geplaatst en in een rij op een onderlinge afstand van 10 cm worden opgesteld (de situatie van 4.4.2.), niet qua volgorde te verschillen. In de volgende paragraaf wordt nader ingegaan op de overwogen proefopzetten en de situaties waarin deze opzetten kunnen worden gebruikt.

5. Proefopzetten

In deze paragraaf wordt nader ingegaan op de overwegingen voor de opzet van de tweede proef. Deze overwegingen geven de mogelijkheid om te kiezen uit een aantal proefopzetten.

Het aantal factoren, dat voor beproeving in aanmerking komt, is groot in verhouding tot de maximale proefomvang. In oriënterend onderzoek met een groot aantal proeffactoren kan vaak met vrucht gebruik worden gemaakt van 2^n -proeven om hoofdeffecten en lagere orde interacties van de proeffactoren te schatten. Van elke factor worden twee instellingen genomen en in de proef worden alle 2^n verschillende combinaties van factorinstellingen toegepast. Is dit aantal te groot dan kan vaak worden volstaan met een deel van de 2^n combinaties: een "zogenaamde" fractionele herhaling, bijvoorbeeld $\frac{1}{4} \times 2^n$ of wel een "kwartherhaling". In fractionele herhalingen zijn niet alle hoofdeffecten en interactie-effecten onafhankelijk van elkaar te schatten. Door echter interessante hoofdeffecten en interacties te strengelen met hogere orde interacties, die naar verwachting niet bestaan of te verwaarlozen zijn, kunnen deze interessante effecten toch worden geschat. Omdat in de tweede proef één van de uitgangspunten bij de keuze van een opzet is, dat zoveel mogelijk hoofdeffecten en lagere orde interacties (in elk geval de interacties tussen twee factoren) van de 11 in paragraaf 3 genoemde proeffactoren moeten kunnen worden geschat, wordt de keuze beperkt tot 2^{11} -proeven of een fractionele herhaling daarvan.

5.1. De omvang van de proef

De omvang van de proef dient dusdanig te zijn dat, bij toetsing, effecten die groter zijn dan een relevante waarde, met een grote waarschijnlijkheid als significant zullen worden aangemerkt. Voor het geval de per experimentele eenheid gemeten fracties beschadigde eieren normaal verdeeld zijn, zijn de toetsen voor de verschillende effecten in een 2^n -proef of een fractionele herhaling daarvan de bekende t-toetsen. Het onderscheidingsvermogen van de t-toets, dat wil zeggen de kans dat een effect ter grootte van de relevante waarde bij toetsing significant wordt bevonden, hangt onder meer af van de grootte van die relevante waarde, de werkelijke waarde van de standaardafwijking (per

experimentele eenheid) van de te onderzoeken variabele en het aantal vrijheidsgraden van de onderliggende t-verdeling.

Aanvankelijk was door de directie van Het Spelderholt gesteld dat, bij het maken van proefopzetten voor de tweede proef, uitgegaan moest worden van maximaal 7000 hennen. Dit aantal maakt een proef met 64 kooiblokken met vier dieren per kooi of een proef met eveneens 64 kooiblokken, maar ditmaal in de helft van de kooiblokken vier dieren per kooi en in de andere helft vijf dieren per kooi, mogelijk. Omdat de te beschouwen proefopzetten 2^n -proeven of een fractionele herhaling daarvan zijn, wordt een schatting van het onderscheidingsvermogen van de toetsen gegeven van zowel een proef met 32 als één met 64 experimentele eenheden. Uit de resultaten van het eerder uitgevoerde onderzoek is voor de variatiecoëfficiënt van de fractie beschadigde eieren per kooiblok over de totale proefperiode een betrouwbaarheidsinterval met een betrouwbaarheid van ongeveer 0.95 berekend van 7 tot 11.3 procent. Op grond hiervan is het onderscheidingsvermogen bij tweezijdige toetsing met een onbetrouwbaarheid $\alpha = 0.05$ berekend voor waarden van de variatiecoëfficiënt van respectievelijk 7, 9 en 11 procent en bij een relevant effect ($\frac{\Delta}{F}$) van 5 respectievelijk 10 procent van de gemiddelde fractie beschadigde eieren. De resultaten zijn samengevat in tabel I.

Tabel I. Onderscheidingsvermogen voor de t-toets bij tweezijdig toetsen met $\alpha = 0.05$ en oneindig veel vrijheidsgraden.

aantal experimentele eenheden	$\frac{\Delta}{F}$ (%)	variatiecoëfficiënt fractie beschadigde eieren (%)		
		7	9	11
32	5	0.55	0.35	0.25
	10	0.98	0.88	0.74
64	5	0.82	0.60	0.44
	10	1.00	0.99	0.95

Hierbij is Δ de relevante waarde van een effect, dat men wil detecteren en F de gemiddelde fractie beschadigde eieren over alle kooitypen.

Hoewel de getallen in tabel I berekend zijn bij oneindig veel vrijheidsgraden, geven zij reeds bij tien of meer vrijheidsgraden een redelijke schatting van het werkelijke onderscheidingsvermogen.

Is het aantal vrijheidsgraden echter beduidend kleiner dan tien, dan liggen de werkelijke waarden van het onderscheidingsvermogen een stuk lager. Bij het bespreken van de diverse opzetten wordt eveneens het te verwachten onderscheidingsvermogen aangegeven.

5.2. *Te onderzoeken factoren*

Met 64 experimentele eenheden is een volledige proef met zes factoren mogelijk. Ingeval de hogere orde interacties niet bestaan of verwaarloosbaar zijn, behoren onvolledige proeven met zeven of meer factoren tot de mogelijkheden. In eerste instantie zijn zeven factoren, welke door de onderzoekers het belangrijkste worden geacht, uit de 11 factoren aangewezen. Deze factoren zijn: batterijsysteem, bodemhelling, maaswijdte, draaddikte, bodemophanging, kooibezetting en coating van de bodem. Van deze zeven factoren dienen voor zover mogelijk de hoofdeffecten en lagere orde interacties met betrekking tot de fractie beschadigde eieren te worden geschat.

5.3. *Randvoorwaarden*

De praktijk stelt voor iedere opzet van onderzoek naar de fractie beschadigde eieren de volgende twee randvoorwaarden.

5.3.1. Alle kooiblokken dienen naast elkaar in een rij in de stal te worden geplaatst. In de stal kunnen vier rijen worden geplaatst, elk van ten hoogste 28 kooiblokken.

5.3.2. Binnen elke rij mogen uitsluitend kooiblokken van één batterijsysteem worden geplaatst.

Bij het bespreken van de verschillende opzetten worden tevens de specifieke randvoorwaarden genoemd.

5.4. *Opzetten in geval de kooibodems gekoppeld zijn*

In deze subparagraaf worden de overwogen opzetten besproken voor het geval dat de kooibodems van naast elkaar staande kooiblokken worden gekoppeld. Daarbij worden de technische aspecten van de opzetten zoveel mogelijk achterwege gelaten. Het aan elkaar koppelen van kooibodems van verschillende kooitypen is zinnig als het koppelen van verschillende kooitypen, de richting of de grootte van de te onderzoeken hoofd- en interactie-effecten van de proeffactoren niet wezenlijk verandert. Door de onderzoekers wordt in eerste instantie aangenomen (hoewel hiervoor geen bewijzen zijn) dat de grootte van de diverse effecten in de proef-situatie niet wezenlijk verschilt van die in de situatie, waarin geen overgangseffecten zouden optreden, doordat bijvoorbeeld naast elk kooiblok randen zijn aangebracht, die groot genoeg zijn.

Aangezien de overgang van een bodemhelling van 7° naar 9° en omgekeerd technische problemen geeft, zullen binnen elke rij de kooiblokken met eenzelfde bodemhelling zoveel mogelijk bij elkaar worden gehouden. Deze voorwaarde en de algemene randvoorwaarden (5.3.1. en 5.3.2.) leiden tot een getrapt systeem voor de opzetten. Als eerste trap wordt het batterijsysteem onderscheiden en als tweede trap de bodemhelling.

Elk van de twee batterijsystemen wordt door loting aan twee rijen toegewezen en in elke rij worden groepjes gekoppelde kooiblokken met dezelfde bodemhelling geplaatst. Binnen elke combinatie van batterijsysteem en bodemhelling worden vervolgens de geselecteerde combinaties van instellingen van de overige factoren aan de kooiblokken toegewezen. De voorwaarde dat kooiblokken met dezelfde bodemhelling in een rij zoveel mogelijk bij elkaar moeten worden geplaatst, beperkt de vrijheid van toewijzen van dergelijke kooiblokken aan de experimentele eenheden in een rij. In verband met de grootte van het onderscheidingsvermogen van de toets voor het bodemhellingseffect, leek het ondanks het technische bezwaar, het gunstigste om groepjes van niet meer dan vier kooiblokken met eenzelfde bodemhelling te nemen.

5.4.1. Fractionele proeven

Ingeval hogere orde interacties niet bestaan, kunnen fractionele proeven worden overwogen. Met groepjes van vier gekoppelde kooiblokken met eenzelfde bodemhelling, is een $\frac{1}{2} \times 2^7$ proef het best bereikbare, met betrekking tot de strengeling van de hoofdeffecten en twee-factor-interacties met hogere orde interacties. Doch zelfs in dat geval is nog een aantal twee-factor-interacties gestrengeld met drie-factor-interacties. Omdat de onderzoekers niet weten welke drie-factor-interacties bestaan, is het niet mogelijk om op verantwoorde wijze een keuze te doen voor het strengelen van de interacties. Er wordt daarom afgezien van een $\frac{1}{2} \times 2^7$ proef.

5.4.2. 2^n proeven

Bij een proefomvang van 64 experimentele eenheden is een proef met zes factoren mogelijk. Dit hield in dat behalve de vier niet geselecteerde proeffactoren ook nog één van de zeven door de onderzoekers geselecteerde proeffactoren op een vaste waarde moest worden ingesteld. Het keuzeprobleem is opgelost door in een hoofdproef aan experimentele eenheden met een niet-gecoate bodem de effecten van de overige factoren te onderzoeken, en in een deelproef het effect van coaten van de bodem na te gaan, wanneer eveneens de factoren bodemhelling, bodemophanging, kooibezetting en maaswijdte worden gevarieerd.

Bij de 2^6 opzet voor de hoofdproef kunnen alle hoofdeffecten, twee-factor-interacties en drie-factor-interacties worden geschat, terwijl een aantal hogere orde interacties niet geschat kan worden omdat deze gestrengeld zijn met effecten tussen groepjes van vier gekoppelde kooiblokken met een zelfde bodemhelling. Als vier-factor-interacties en hogere interacties niet bestaan, worden de kwadraatsommen van deze effecten samen genomen tot een gepoolde restkwadraatsom met tien vrijheidsgraden, hetgeen een redelijk aantal is in verband met het onderscheidingsvermogen van de toetsen voor hoofdeffecten en de twee- en drie-factor-interacties.

Voor het geval de werkelijke variatiecoëfficiënt (per kooiblok) voor de fractie beschadigde eieren tussen de 7 en 11 procent ligt (de beste schatting op grond van eerder verricht onderzoek), geven de in tabel I in paragraaf 5.1. gegeven getallen voor 64 experimentele eenheden een

redelijke indruk van het te verwachten onderscheidingsvermogen voor het toetsen van hoofdeffecten, twee- en drie-factor-interacties, echter met uitzondering van de hoofdeffecten batterijsysteem en bodemhelling en hun interactie. Het onderscheidingsvermogen voor een batterijsysteem-effect is zeer gering vanwege slechts twee vrijheidsgraden, doch dit is niet bezwaarlijk omdat de onderzoekers toch niet in een hoofdeffect van het batterijsysteem geïnteresseerd zijn. Omdat de standaardafwijking (per groep van vier kooiblokken met dezelfde bodemhelling) van de fractie beschadigde eieren over de totale proefperiode niet bekend is, kan het onderscheidingsvermogen van de toetsen van het bodemhellingseffect en de interactie tussen bodemhelling en batterijsysteem niet worden berekend. Voor het geval dat effecten tussen groepjes zich niet manifesteren, lijkt het alleszins redelijk als schatting voor deze standaardafwijking tweemaal de standaardafwijking (per kooiblok) te nemen. In dat geval geven de in tabel I voor 64 eenheden gegeven getallen een redelijke indruk van de grootte van het onderscheidingsvermogen.

Aangezien de onderzoekers eveneens het effect van het coaten van de bodem in een aantal gevallen willen onderzoeken, is de aanvankelijk gestelde proefomvang nog wat vergroot. Bij de oorspronkelijke 64 kooiblokken zijn nog eens 16 kooiblokken genomen. Het effect van het coaten van de bodem wordt onderzocht in een deelproef, waarin naast de factor coaten van de bodem eveneens de factoren bodemhelling, bodemophanging, kooibezetting en de maaswijdte zullen worden gevarieerd. Daarentegen zal het effect van coaten alleen bij de Compact-batterij en alleen bij draaddikte 2.05 mm worden nagegaan. Daar in de deelproef de helft van de $2^5 = 32$ combinaties waaraan het effect van het coaten van de bodem wordt nagegaan, te weten de niet gecoate bodems met een draaddikte van 2.05 mm, reeds deel uitmaken van de 2^6 proef, is de totale proefomvang 80 (64+16). In deze proef kunnen alle hoofdeffecten en twee-factor-interacties worden geschat. Voor het geval dat drie-factor-interacties en hoger niet bestaan, worden de kwadraatsommen van deze effecten samengenomen tot een gepoolde restkwadraatsom met tien vrijheidsgraden. Met behulp van een grafische methode, de zogenaamde half-normal plot, kan worden nagegaan of de veronderstelling dat drie-factor-interacties en hoger niet bestaan redelijk is. Ingeval de werkelijke variatiecoëfficiënt (per kooiblok) voor de fractie beschadigde eieren tussen de 7 en

11 procent ligt, geven de in tabel I gegeven waarden bij 32 eenheden een redelijke indruk van het onderscheidingsvermogen. Voor het onderscheidingsvermogen van de toets van een hellingseffect zijn de in tabel I gegeven waarden slechts grove schattingen, omdat de t-verdeling van de toetsingsgrootte op slechts vijf vrijheidsgraden berust. Bovendien speelt nog een rol het feit dat de gebruikte benadering van de standaardafwijking (per groep van vier kooiblokken met dezelfde hellingsinstelling) door tweemaal de standaardafwijking (per kooiblok) een onderschatting is.

5.5. *Opzetten in geval de kooibodems niet gekoppeld zijn*

In deze situatie wordt elk kooiblok aan twee frames opgehangen en worden de kooiblokken met hun frames op een onderlinge afstand van 10 cm in de rij geplaatst. Door de onderzoekers wordt aangenomen dat deze vorm van opstellen van kooiblokken beter is, dat wil zeggen de richting en de grootte van de te meten effecten minder beïnvloedt dan bij directe koppeling van kooibodems van de verschillende kooitypen. Bij deze wijze van plaatsen van de diverse kooiblokken worden de combinaties van één batterijsysteem toegewezen aan twee batterijrijen. Binnen elke batterijrij kunnen zonder enige beperking de diverse kooitypen aan de experimentele eenheden worden toegewezen. Hieronder worden de overwogen opzetten besproken.

5.5.1. *Fractionele proeven*

Als vier-factor-interacties en hogere niet bestaan, kan voor de zeven geselecteerde proeffactoren een $\frac{1}{2} \times 2^7$ proef worden overwogen. Omdat in een rij uitsluitend kooitypen van één batterijsysteem kunnen worden geplaatst, komt deze proef in feite neer op een $\frac{1}{2} \times 2^6$ proef voor elk batterijsysteem. Daar voor elk batterijsysteem slechts 32 van de 64 combinaties worden genomen, is ieder effect van de zes factoren binnen een batterijsysteem met een ander effect gestrengeld. Het is mogelijk deze 32 combinaties zo te kiezen dat ieder hoofdeffect, iedere twee-factor-interactie en iedere drie-factor-interactie gestrengeld wordt met respectievelijk een vijf-factor-, een vier-factor- en een drie-factor-interactie. In dat geval zijn de twee-factor-interacties met batterijsysteem en de drie-factor-interacties met batterijsysteem gestrengeld met respectievelijk een zes-factor-interactie en een vijf-factor-inter-

actie. Dus in geval 4-factor-interacties en hogere niet bestaan, worden bij deze opzet schattingen verkregen voor alle hoofdeffecten, twee-factor-interacties en drie-factor-interacties met batterijsysteem. De overige drie-factor-interacties zijn gepaard gestrengeld en er kunnen dus alleen schattingen voor het totaal effect van elk paar worden verkregen. Daarbij kan van één paar het totaal effect niet worden geschat, omdat dit gestrengeld is met een rij-effect als gevolg van het feit dat per batterijsysteem de 32 kooitypen in twee rijen van 16 kooiblokken worden opgesteld. Door de kwadraatsommen van de vier-factor-interacties met batterijsysteem samen te nemen, wordt een gepoold gemiddeld restkwadraat met negen vrijheidsgraden verkregen hetgeen acceptabel is in verband met het onderscheidingsvermogen van de toetsen.

In geval de variatiecoëfficiënt (per kooiblok) tussen 7 en 11 procent ligt (hetgeen niet zo hoeft te zijn, omdat de variatiecoëfficiënt bij niet-gekoppelde kooibodems kan verschillen van die bij gekoppelde kooibodems) geven de in tabel I gegeven getallen voor 64 eenheden een redelijke schatting van het onderscheidingsvermogen voor de toetsen van alle effecten tot en met drie-factor-interacties, uitgezonderd het hoofdeffect voor batterijsysteem. Het onderscheidingsvermogen voor een batterijsysteemeffect is gering vanwege slechts twee vrijheidsgraden voor de relevante restkwadraatsom. Dit is echter niet bezwaarlijk, omdat de onderzoekers in feite niet in een batterijsysteemeffect geïnteresseerd zijn.

5.5.2. 2^n proeven

Evenals voor de situatie waarbij de kooibodems gekoppeld zijn is gedaan, zijn een 2^6 en een 2^5 proef met dezelfde factoren overwogen. Het verschil is echter dat ditmaal de kooiblokken met dezelfde bodemhelling in een rij niet bij elkaar geplaatst hoeven te worden. Hieronder worden de hoofdpunten van de 2^6 proef vermeld. De 2^6 proef is in feite een 2^5 proef per batterijsysteem als gevolg van het feit dat in elke rij slechts kooitypen van hetzelfde batterijsysteem kunnen worden geplaatst. Per batterijsysteem worden de $2^5 = 32$ combinaties neergelegd in twee rijen, elk ter lengte van 16 kooiblokken. Daardoor is per batterijsysteem een effect gestrengeld met een rij effect. Bij een speciale keuze van de diverse kooitypen in de rijen wordt de vijf-factor-inter-

actie tussen de factoren bodemhelling, bodemophanging, maaswijdte, draaddikte en kooibezetting gestrengeld met een rij-effect. In dat geval kunnen alle overige effecten, uitgezonderd de zes-factor-interactie, worden geschat. Als vier-factor-interacties en hogere niet bestaan, worden hun kwadraatsommen samengenomen tot een restkwadraatsom met 20 vrijheidsgraden. Ingeval de fracties beschadigde eieren per kooiblok normaal verdeeld zijn en de variatiecoëfficiënt van de fractie beschadigde eieren tussen de 7 en 11 procent ligt, geven de in tabel I voor 64 eenheden gegeven getallen een redelijke schatting van het onderscheidingsvermogen van de toetsen van de diverse effecten.

De belangrijkste punten van de 2^5 proef zijn dat:

- alle effecten uitgezonderd de vijf-factor-interactie, welke gestrengeld wordt met een rij-effect, geschat kunnen worden,
- ingeval vier-factor-interacties en een aantal drie-factor-interacties niet bestaan een gepoolde restkwadraatsom met een redelijk aantal vrijheidsgraden voor de toetsing van effecten kan worden verkregen,
- in bovengenoemd geval de getallen in tabel I voor 32 eenheden een redelijke indruk geven van het onderscheidingsvermogen voor de toetsen van hoofdeffecten, twee- en drie-factor-interacties bij een waarde van de variatiecoëfficiënt tussen 7 en 11 procent.

Vergelijken van de $\frac{1}{2} \times 2^7$ proef en de combinatie van de 2^6 en de 2^5 proeven leert dat:

- in de $\frac{1}{2} \times 2^7$ proef, elk effect gestrengeld is met een ander effect,
- in geval vier-factor-interacties en hogere niet bestaan bij de $\frac{1}{2} \times 2^7$ proef, alle hoofdeffecten, twee-factor-interacties en drie-factor-interacties met batterijsysteem kunnen worden geschat. Elk van de overige drie-factor-interacties met een andere drie-factor-interactie is gestrengeld. Omdat niet bekend is welke drie-factor-interacties bestaan, is niet bekend aan welke van de twee drie-factor-interacties van een gestrengeld paar een effect dient te worden toegeschreven. In de 2^6 en 2^5 proef kunnen alle effecten, uitgezonderd degene welke met rij-effecten werden gestrengeld, worden geschat en zijn er geen problemen bij het toewijzen van een gevonden effect. Bij deze opzetten worden echter geen schattingen verkregen voor eventuele interacties tussen het coaten van de bodem

- met het batterijsysteem of de draaddikte,
- in de $\frac{1}{2} \times 2^7$ en de 2^6 proef het onderscheidingsvermogen voor de diverse effecten nagenoeg hetzelfde zal zijn. Doch het onderscheidingsvermogen voor de effecten in de 2^5 proef is beduidend lager dan dat voor de corresponderende effecten in de $\frac{1}{2} \times 2^7$ proef,
 - de omvang van de $\frac{1}{2} \times 2^7$ proef 64 eenheden bedraagt tegen 80 voor de combinatie van de 2^5 en de 2^6 proef. Dit houdt ondermeer in dat, bij eenzelfde bemonsteringsfrequentie van de kooiblokken, bij de combinatie van de 2^6 en 2^5 proef $1.25 = \left(\frac{80}{64}\right)$ maal zoveel werk moet worden verricht.

5.6. Keuze van de opzet

Aangezien de onderzoekers van mening zijn dat het niet koppelen van de kooibodems beter is dan het direct koppelen van de bodems, wordt de keuze beperkt tot de opzetten in paragraaf 5.4. De onderzoekers hebben gekozen voor de combinatie van een 2^6 en een 2^5 proef. In de overwegingen die tot deze keuze hebben geleid, is het argument dat de resultaten uit de tweede proef direct toegankelijk dienen te zijn voor de praktijk zwaarder gewogen dan de nadelen van de gekozen opzet, namelijk het niet kunnen schatten van een aantal mogelijk interessante effecten en een aanzienlijke reductie van het onderscheidingsvermogen van de toetsen voor een aantal effecten.

6. Bemonstering

De diverse kooitypen zullen gedurende de proefperiode 30 tot 40 maal bemonsterd worden. Daarbij gaan de gedachten uit naar een gestratificeerde aselechte steekproefname. Bij de strata wordt gedacht aan perioden van twee opeenvolgende weken. Omdat de variatie van het aantal beschadigde eieren in een dagproductie toeneemt in de tijd, zal eventueel de bemonsteringsfrequentie voor de strata in de tijd worden vergroot. Dit wordt gedaan om de totale variatie, de som van de variatie tussen experimentele eenheden en de steekproefvariatie, voor de fractie beschadigde eieren per kooiblok te verkleinen.

7. Verbeteringen ten opzichte van de opzet voor de eerste proef

De belangrijkste verbeteringen ten opzichte van de opzet en uitvoering van de eerste proef zijn dat:

- in plaats van de effecten van drie factoren op de fractie beschadigde eieren te onderzoeken, nu de effecten van vijf en zes factoren worden onderzocht bij nagenoeg eenzelfde aantal experimentele eenheden ($18 \times 4 = 72$ in de eerste proef tegen 80 in de tweede proef). Daarbij blijft het onderscheidingsvermogen gelijk, mits de variatiecoëfficiënt in de situatie waarin de kooibodems niet gekoppeld zijn niet verschilt van die in de situatie waarin de kooibodems direct gekoppeld zijn.
- de bemonstering niet meer systematisch geschiedt (de proefresultaten van de eerste proef konden niet zonder extra veronderstellingen geanalyseerd worden).