

LANDBOUWWETENSCHAPPEN EN BIOMETRIE

door prof.dr. D.A.M.K. Rasch



Inaugurele rede uitgesproken op 29 oktober 1992
bij de aanvaarding van het ambt van hoogleraar in de
Wiskundige Statistiek aan de Landbouwniversiteit
Wageningen

LANDBOUWWETENSCHAPPEN EN BIOMETRIE

*Mijnheer de Rector Magnificus,
geachte toehoorders,*

Inleiding

Menigeen mag denken dat het "Uilen naar Athene dragen" zou zijn als iemand over de samenhang tussen de Biometrie en de Landbouwwetenschappen wil praten. Het is zeker waar, dat de eerste belangrijke resultaten van de Biometrie voortgekomen zijn uit landbouwwetenschappelijke problemen - maar de wortels van de Biometrie en ook de meerderheid van de tegenwoordige toepassingen liggen op het medische en sociale gebied. Eerst wil ik een korte schets van Biometrie geven en daarna zal ik iets over de geschiedenis van dit vak vertellen. Verder zullen de voor de landbouwwetenschappen belangrijkste deelgebieden gekarakteriseerd worden. Aan het eind wil ik een mogelijkheid aangeven om met moderne hulpmiddelen van de informatica voor landbouwwetenschappers een gemakkelijker toegang tot de biometrische resultaten mogelijk te maken.

Wat is Biometrie?

Een antwoord op deze vraag kan bijvoorbeeld gegeven worden indien wij de definitie van het woord Biometrics in het Engelstalige Biometrical Dictionary bekijken, dat uitgegeven zal worden door Elsevier Publ. Comp.¹⁾

"Biometrics is the application of mathematical methods, particularly mathematical statistics, to the life sciences (biology, agricultural science, human and veterinary medicine, pharmacology, etc.). Owing to

the variability of organisms and the corresponding diversity in the expression of their traits, biological relationships can rarely be described in terms of strictly defined functions, and the characters and properties that are studied are almost always influenced by random factors. They can, therefore, only be treated as random variables, and the laws they obey must be studied. Many of the methods used in mathematical statistics were originally developed to investigate biological phenomena.

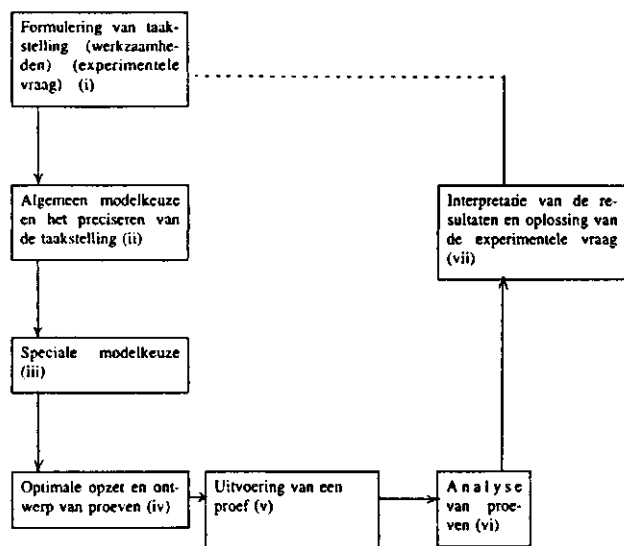
In Biometrics, the suitability of general mathematical statistical methods for studying biological phenomena is considered from the formulation of the problem through the recording and analysis of data to the interpretation of the results. Since the purpose of Biometrics is to find answers to biological problems, the biological aspect and the statistical methods must be compatible. Biometrics has become an important tool in life science research. One of the achievements of Biometrics is that inferences can now be drawn from biological data rationally and with known error probabilities and risks.

The term Biometrics is sometimes used in a more restricted sense as a synonym of Biostatistics, while the non-statistical mathematical methods are grouped together with Biometrics under the blanket term Biomathematics. Usage varies in this respect."

Deze definitie bevat meer dan van de woordelijke vertaling of van de oorspronkelijke zin van het woord Biometrie verwacht wordt. Het woord Biometrie is afgeleid van de griekse woorden βίος (leven) en μέτρον (meten). Meten zou hier staan voor de uitvoering van een proef, maar Biometrie in zijn moderne betekenis is juist niet de uitvoering van een

proef maar bevat alle methodische activiteiten voor en na een proef die in de biologische wetenschappen uitgevoerd moeten worden.

Dat kunnen wij met het volgende schema verduidelijken.

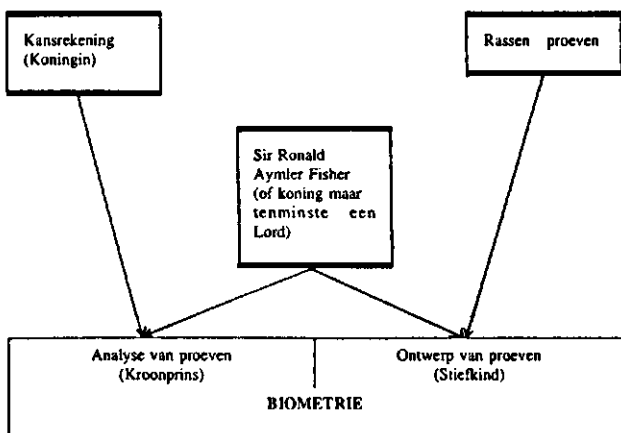


Figuur 1. De positie van de biometrie in het wetenschappelijk proces van winning van inzicht.

Meestal wordt onder Biometrie alleen de analyse van proeven op biologisch gebied verstaan. Hiervoor kunnen wij twee redenen aanvoeren. Een van deze redenen is dat niet alleen in de Biometrie maar heel algemeen in de statistiek het opzetten van proeven als een stiefkind wordt gezien.

De oorzaak kan zijn dat - en dat brengt ons terug tot de titel van het verhaal - een deel van de Biometrie niet alleen in de landbouwwetenschappen gebruikt wordt maar ook daar is ontstaan en daarmee een niet-wiskundige wortel heeft.

Dat maakt de volgende figuur duidelijk:



Figuur 2. De biometrische familie.

Behalve de eerste reden dat het ontwerpen van proeven wegens zijn landbouwwetenschappelijke wortel binnen de wiskunde niet bon ton is, ligt de tweede reden in de psyche van de onderzoekers zelf.

Het is veel moeilijker het preciese doel van een onderzoek samen met de nodige nauwkeurigheid van de resultaten vóór een proef te bepalen als na de uitvoering van een proef te zeggen "dat is precies dat wat ik zou willen".

Vaak kunt U de vraag vernemen "Hoeveel dieren, planten enz. zijn voor een bepaald onderzoek nodig?". De biometricus antwoordt dan steeds met een tegenvraag "Hoe nauwkeurig moet het resultaat zijn en hoe groot is de variatie van dat materiaal?". Om een goed antwoord daarop te geven moet een onderzoeker nadenken en veel van het object van zijn toekomstig onderzoek weten. Hier ligt een psychologische drempel voor de algemene geliefdheid van het ontwerpen van proeven.

Dames en heren,

In 1993 blikken wij met z'n allen 75 jaar terug in de LU historie. Ook ik wil nu terugkijken maar dan in mijn vakgebied Biometrie en graag iets over de geschiedenis van een voor de landbouwwetenschappen belangrijk gebied zeggen.

Ontwikkeling van Biometrie als een wetenschap:

In het begin van de Biometrie stond het meten (of tellen) op de voorgrond - het ontwerpen maar ook de analyse van de verzamelde getallen was minder belangrijk.

Zo weten wij dat Koning David de hoogte van de mensen van zijn volk heeft laten meten en dat in de geslachtslijn (stam) Levi het aantal linkshandige soldaten geteld is. (Ludwig^{2b}). Uit het kerstverhaal (in de bijbel) weten wij dat de Romeinen ("In die dagen

ging er een gebod uit van Keizer Augustus ...) het aantal mensen in hun rijk geteld hebben.

In de 14^e eeuw is het eerste register over veranderingen (geboorten, huwelijken, sterfgevallen) in Bourgondië ontstaan. De eerste moderne volkstelling gebeurde in 1666 in Canada. De eerste analyse methoden waren heel gemakkelijk. Sinds de zware pest in 1592 in London is het aantal (absolute frequentie) geboorten en sterfgevallen gepubliceerd (bills of mortality). In het begin iedere week, later verschenen deze lijsten onregelmatig maar sinds 1603 weer regelmatig en sinds 1629 zijn ze onderverdeeld naar het geslacht van de personen. De eerste methodische analyses van deze gegevens heeft J. Graunt³⁾ gedaan.

Als één van de grondleggers van de moderne Biometrie kan Johann Peter Süßmilch genoemd worden. In zijn hoofdwerk "Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechts aus der Geburt, Tod und Fortpflanzung erwiesen"⁴⁾ wordt de vraag gesteld of het toevallig kan zijn wanneer onder 14.000 geboren en 7200 jongens en 6800 meisjes zijn. Dat komt in principe neer op de vraag naar een toets voor de hypothese

$$H_0: p = 0.5$$

als p voor de kans staat dat een geborene een jongen is. Volgens Todhunter⁵⁾ heeft een medicus namens J. Arbuthnot al een soort tekentoeets (sign test) gebruikt om de Londense geboorten frequenties tussen 1629 en 1700 H_0 te toetsen. Over de geldigheid van zo'n toets hebben s' Gravesande⁶⁾ en de Moivre gekorrespondeerd, deze briefwisseling was door de

Montmort⁷⁾ gepubliceerd. Het was W. Gosset, die misschien als eerste de samenhang tussen het aantal herhalingen en de nauwkeurigheid van een gevolgtrekking onderzocht heeft. Van 3000 mensen heeft hij de lengte en de lengte van de linker middelvinger gemeten, en iedere kenmerkwaarde op een papiertje geschreven. Uit deze 3000 papiertjes heeft hij zonder teruglegging 750 aselechte steekproeven met omvang 4 en later met andere omvangen getrokken. Op deze wijze kwam hij empirisch tot de t-verdeling, ook Student-verdeling genoemd omdat Gosset zijn resultaten onder het pseudoniem Student^{8,9)} gepubliceerd heeft.

In het jaar 1901 werd het wetenschappelijke tijdschrift "Biometrika" opgericht als "A journal for the statistical study of Biological problems" door W.F.R. Weldon, Karl Pearson en C.B. Davenport. De uitgevers hebben inleidend geschreven "It is intended that Biometrika shall serve as a means not only of collecting under one title biological data of a kind not systematically collected or published in any other periodical, but also of spreading a knowledge of such statistical theory as may be requisite for their scientific treatment".

De definitie van Biometrie gegeven door F. Galton in het eerste nummer van "Biometrika"¹⁰⁾ kan tegenwoordig niet meer worden gebruikt.

"The primary object of Biometry is to afford material that shall be exact enough for the discovery of incipient changes in evolution which are too small to be otherwise apparent."

Onafhankelijk van deze definitie verschijnen al in de eerste jaargang van Biometrika artikelen die in samenhang met landbouwwetenschappen kunnen worden gezien als er "Mendel's laws of alternative

inheritance in pea¹¹⁾ en "On the inheritance of coat color of thoroughbred horses"¹²⁾ zijn.

Intussen hebben wij een groot aantal van biometrische tijdschriften bijvoorbeeld:

Biometrics - the journal of the Biometric Society (1946)

Biometrical Journal (1959)

Biometrie - Praximetrie (1960)

Deelgebieden van de Biometrie

De volgende biometrische richtingen zijn belangrijk voor de landbouwwetenschappen. Zoals wij in de beide artikelen in de eerste band van Biometrika kunnen zien, was de genetica één van de eerste toepassingen van de statistiek voor de landbouwwetenschappen.

Biometrische genetica

Een model dat bekeken wordt is het waarschijnlijkheidsmodel van natuurlijke selectie.

Als wij bijvoorbeeld uitgaan van diploïde individuen met q allelen A_1, A_2, \dots, A_q op een locus A . Dan kunnen wij zeggen dat er een stochastische variabele \underline{A} bestaat die de waarden A_1, A_2, \dots, A_q met de kansen p_1, p_2, \dots, p_q aanneemt. Een stochastische variabele is kort gezegd een variabele die waarden op een toevallige wijze aanneemt zoals bijvoorbeeld het aantal ogen die een geworpen dobbelsteen aan kan geven.

Onder verschillende voorwaarden kunnen die kansen van de mogelijke genotypen, de selectiecoëfficiënten of de "fitness" van deze genotypen en de dynamica in een populatie berekend worden.

In het geval van polygene kenmerken zijn kwantitatieve modellen in gebruik, bijvoorbeeld de phenotype (waarde van een kenmerk) $P=y$ is opgebouwd uit een erfelijk bepaalde (G) en een door het milieu bepaalde component (u) dat is:

$$P = G + u.$$

Deze formule vinden wij al in een artikel van Yule (1902)¹³⁾. Op dit eenvoudige model zijn de meeste resultaten van de moderne biometrische genetica (of populatie genetica) van vandaag gebaseerd. Dat geldt voor de methoden voor het schatten van de "heritability coefficient" of erfelijkheidsgraad (aandeel van de erfelijk veroorzaakte variantie aan de totale (phenotypische) variantie) en voor formules voor het bepalen van steekproefomvang voor zulke schattingen, formules die er vaak op wijzen dat enkele duizenden waarnemingen voor een nauwkeurige schatting nodig zijn.

Met behulp van de biometrische genetica kan men ook aangeven op het feit hoe gemakkelijk fouten kunnen worden voorkomen.

Bijvoorbeeld is de gestandaardiseerde selectiewaarde d_s gedefinieerd als de gestandaardiseerde (d.w.z. gedeeld door de standaardafwijking σ) verschil tussen de verwachting μ_s van een kenmerk in het door afknotting geselecteerde deel van een populatie en de oorspronkelijke verwachting μ .

$$d_s = \frac{1}{\sigma}(\mu_s - \mu)$$

In het geval van een normaal verdeelde variabele kan d_s met behulp van de kansdichtheidsfunctie en de kansverdelingsfunctie van de standaard normale

verdeling berekend worden. Maar de op deze wijze ontstane formule wordt ook vaak voor niet-normale verdelingen gebruikt en dat kan aanzienlijke fouten opleveren^{14,15}.

Voorspelling van fokwaarden

In de veefokkerij is het heel belangrijk te weten welke vaderdieren (bijvoorbeeld fokstieren of fokhengsten) goed in de fokkerij gebruikt kunnen worden. Door de biometrici zijn verschillende methoden ontwikkeld, om met speciale lineaire modellen fokwaarden te bepalen. Als de fokwaarden in zo'n model vast zijn, dan zijn dat de methoden van fokwaarden-schattingen (vergelijking van dochter en moeder of "contemporary comparison test"). Maar als de fokwaarden in een lineair model stochastische variabelen zijn, zijn de methoden van fokwaarden voorspelling in gebruik, speciaal de BLUP methode (Best Linear Unbiased Prediction).

Constructie van selectie indexen

Veefokkers en plantenveredelaars zijn er vaak niet alleen in geïnteresseerd één kenmerk te verbeteren maar gelijktijdig een groot aantal kenmerken y_1, y_2, \dots, y_r te verbeteren. Omdat tussen de kenmerken fenotypische en genotypische afhankelijkheden bestaan is het niet optimaal voor ieder kenmerk een eigen selectiegrens (ondergrens) te bepalen en alleen individuen te selecteren die boven alle grenzen liggen. Er zijn optimale selectieprocedures voor zulke gevallen ontwikkeld, één daarvan is de constructie van selectie-indexen

$$I = c_1y_1 + c_2y_2 + \dots + c_ry_r$$

waar de waarden c_i (gewichten) afhankelijk zijn van het economische belang van y_i , maar ook van de erfelijkheid en van de genetische correlatiecoëfficiënten tussen de kenmerken.

Rassenproeven

In de plantenveredeling worden vaak nieuwe rassen ontwikkeld. Deze rassen moeten zorgvuldig getoetst worden voordat zij in de plantenteelt gebruikt kunnen worden. Het was Sir Ronald Aymler Fisher die tijdens zijn werkzaamheden op Rothamsted Experimental Station de biometrische methoden voor deze toetsen ontwikkelde¹⁶⁾.

Om de invloed van de grond op de resultaten te verminderen zijn bij Fisher technieken ontwikkeld die tegenwoordig heel algemeen en ook buiten de landbouw gebruikelijk zijn. Dat zijn bijvoorbeeld onvolledige blokkenproeven (BIBs) en Latin Squares. Hoe groot de invloed van de landbouwwetenschappen op de Biometrie was dat kunnen wij zien aan het aantal behandelingen in een blokkenproef, dat in de wiskundige theorie nog steeds met v wordt aangegeven ($v = \text{variety} = \text{ras}$).

Dames en heren,

In deze samenhang kan ik een anekdote vertellen die de relatie tussen Biometrie en Landbouwwetenschappen weerspiegelt.

Ongeveer tien jaar geleden heb ik een lezing gehouden voor plantenveredelaars over methoden van de wiskundige statistiek die goed geschikt zijn om in de

plantenveredeling toegepast te kunnen worden. Dat waren moderne selectiemethoden ontwikkeld door Bechhofer¹⁷⁾. Selectie in deze samenhang betekent selectie van de beste populatie uit een aantal k van gegeven populaties (bijvoorbeeld de populatie met de grootste verwachting) en niet de kunstmatige selectie van individuen uit een populatie. Ik heb toen voorgesteld niet verder de niet op het probleem toegepaste methoden van de simultane vergelijkingen (multiple t-toets, Tukey toets enz.) te gebruiken maar wel deze nieuwe selectiemethoden, die het ras met de hoogste opbrengst als het beste ras kiest. Tijdens de discussie stond een deelnemer op en zei dat de biometrici toch wel erg onduidelijk waren. Tot 1930 hebben de rassenveredelaars altijd de rassen met de hoogste opbrengst gekozen. Toen kwamen de biometrici die zeiden dat dat niet zo eenvoudig gedaan kan worden maar dat men moet toetsen of er significante verschillen tussen de rassen zijn. Sinds deze tijd zijn de rassenveredelaars begonnen simultane vergelijkingen te gebruiken en nu is er weer een biometricus die zegt dat dat niet kan.

Wat kunnen wij vanuit het biometrische standpunt daarop antwoorden?

Het gaat altijd om de nauwkeurigheid van resultaten. Een ras kan toevallig het hoogste gemiddelde in een proef hebben ook als het niet de beste ras (het ras met de grootste verwachting) is. In 1930 waren significantie toetsen bekend en de biometrici hebben aanbevolen te toetsen of het beste significant van de overige afwijkt omdat in deze tijd de Bechhofer-theorie niet bekend was. Nu de theorie van selectiemethoden ontwikkeld is en wij met deze

methoden de kans van een verkeerde selectie kunnen berekenen, is het beter de selectiemethoden te gebruiken, die beter op het probleem van rassen ontwikkeling toegepast zijn. In de Hegel'se zin doen wij achter de "Negation der Negation" hetzelfde als vroeger maar nu op een hoger niveau. Vanuit het oogpunt om het benodigde aantal herhalingen voor een rassenproef te bepalen is het ook beter een selectie methode te kiezen.

In tabel 1 zijn de aantallen herhalingen aangegeven die nodig zijn om voor verschillende methoden bepaalde nauwkeurigheden te krijgen.

Methode	Aantal n
"Comparisonwise" kans op fout van de eerste soort $\alpha = 0.05$	n = 16
"Experimentwise" kans op fout van de eerste soort $\alpha_E = 0.05$	n = 32
Selectie van de beste soort met een kans 0.05 voor een foute selectie	n = 12

Tabel 1. Aantallen herhalingen per ras voor simultane vergelijkingen en selectie methoden met 9 rassen ($\sigma = 16$, $d = 4$, $\beta_0 = 0,20$).

Hierbij hebben wij de kans op een foute selectie gelijkgesteld aan de kans op een fout van de eerste soort (0.05). Voor de simultane vergelijkingen moet de kans op een fout van de tweede soort kleiner dan $\beta_0 = 0.20$ zijn, als het grootste verschil tussen de

verwachtingen groter dan $d = 4$ is. De kans op een foute selectie is kleiner dan 0.05, als de grootste verwachting tenminste 4 (dat betekent een standaard afwijking) groter dan de volgende is.

Analyse van groeikrommen

In de landbouwwetenschappen komt het vaak voor dat het groeien van planten, dieren of van economische ontwikkelingen onderzocht moet worden. Een empirische beschrijving van een groeiproces kan door een groeikromme gegeven worden. Daarbij worden de waarden van het groeikenmerk (bijvoorbeeld de hoogte van planten) tegen de tijd (tijdvak of leeftijd) in een coördinatenstelsel afgezet. Door de zo ontstane punten trekken wij een kromme die zo goed mogelijk door deze punten past en noemen deze kromme een groeikromme. De vraag is of wij dit proces kunnen algoritmiseren en of wij van de kromme een functie kunnen afleiden, die het groeiproces zo goed mogelijk beschrijft. Meestal is een beschrijving niet voldoende. Als het alleen om een beschrijving gaat kunnen wij dat altijd heel precies doen met een polynoom van een voldoende hoge graad. De wetenschappers die met een groeifunctie moeten werken zijn meestal geïnteresseerd in een functie waarvan de parameters geïnterpreteerd kunnen worden. In dat geval is een polynoom geen goede keuze. Verder is het met een polynoom ook niet mogelijk een extrapolatie voor toekomstige tijdstippen te doen. Daarom is een belangrijke stap in de analyse van groeikrommen de modelkeuze. Er moet een functie worden gekozen die goed met de gegevens overeenstemt en liefst met zo min mogelijk parameters. Er zijn criteria voor de modelkeuze die bij

een gelijkwaardige aanpassing van twee functies, dié functie kiezen die het laagste aantal parameters heeft. (Schwartz-criterium, Mallows-criterium, Akaike-criterium¹⁸⁾.)

Als er een speciale groeifunctie gevonden is, zijn methoden voor het schatten van de parameters nodig. De ontwikkeling van zulke methoden gebeurde aanvankelijk binnen de Biometrie maar later meestal in de numerieke wiskunde. De moeilijkheden zijn veroorzaakt door de niet-lineariteit van de meeste groeikrommen. Voor het schatten van de parameters zijn beginwaarden nodig. Een landbouwwetenschapper en een biometricus kunnen vaak op eenvoudige wijze bepalen hoe de parameters van zulke groeifuncties geïnterpreteerd kunnen worden.

Veel functies waarbij dat mogelijk is zijn meestal meer dan 100 jaar oud en binnen biometrisch onderzoek ontwikkeld (Verhulst¹⁹⁾ Gompertz²⁰⁾, Quetelet²¹⁾, Malthus²²⁾, Mitscherlich²³⁾, Richards²⁴⁾).

De laatste tijd zijn er ook methoden ontwikkeld om betrouwbaarheidsinterval schatters en toetsen voor de parameters van zulke functies aan te geven²⁵⁾. Maar ook het "stiefkind" ontwerp van proeven is op dit gebied ontwikkeld. In deze samenhang is niet alleen het aantal waarnemingen belangrijk maar ook wanneer er moet worden gemeten. Om de meetpunten optimaal te bepalen is een criterium nodig. Vaak wordt het criterium van de D-optimaliteit gebruikt, dat van de covariantie-matrix van de asymptotische verdeling van de schattingen van de parameters afhankelijk is.

Met een voorbeeld wil ik laten zien hoe groot het ver-

schil van een optimale proefopzet met het gewone meten op gelijke afstand (equidistant meten) is.

tijd t (maanden)	hoogte (cm)	tijd t (maanden)	hoogte (cm)
0	77,2	36	129,2
6	94,5	42	129,9
12	107,2	48	130,4
18	116,0	54	130,8
24	122,4	60	131,2
30	126,7		

Tabel 2. Schofthoogte in cm van runderen in afhankelijkheid van de leeftijd in maanden.

Voorbeeld 1

Groei van de schofthoogte van runderen. In tabel 2 vinden wij de resultaten van metingen tussen geboorte en 60 maanden indien zij gewoon worden vastgesteld.

Indien wij een exponentiële functie van de vorm

$$f(t) = \alpha + \beta \exp(\gamma t)$$

aanpassen weten wij dat voor negatieve γ -waarden α de schofthoogte voor volwassen runderen is en $\alpha + \beta$ (β negatief) de schofthoogte bij de geboorte is. Met γ wordt de snelheid van het groeien gekarakteriseerd.

In figuur 3 kunnen wij de 11 paren van waarden (t,y) van tabel 2 en de aangepaste functie f zien waarbij de

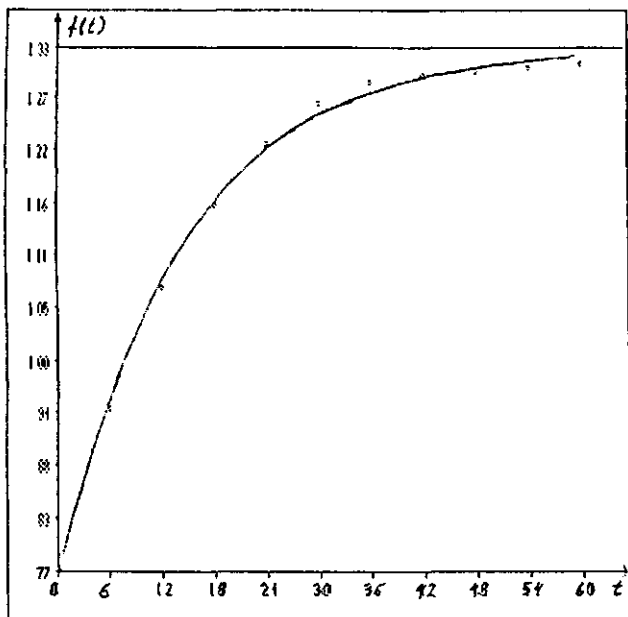
schattingen $\hat{\alpha}$, $\hat{\beta}$ en $\hat{\gamma}$ van α , β en γ door

$$\hat{\alpha} = 132,96$$

$$\hat{\beta} = -56,42$$

$$\text{en } \hat{\gamma} = -0,0677$$

gegeven zijn.



Figuur 3. Waarden van tabel 2 en de aangepaste functie $132,96 - 56,42 \exp(-0,0677t)$.

Indien nu de optimale proefopzet met het criterium van de D-optimaliteit en 11 metingen bepaald wordt,

geeft dat de volgende proefopzet

4 keer meten bij de geboorte (0 maanden)

4 keer meten met 13,72 maanden

3 keer meten met 60 maanden

De waarde van het criterium is $0,2960 \cdot 10^6 \sigma^6$.

Als wij de waarde van dit criterium voor de geanalyseerde proefopzet in tabel 2 berekenen is de waarde $0.87559 \cdot 10^6 \sigma^6$, en dat is ongeveer 3 keer zo groot.

Dat betekent dat wij met een optimale proefopzet in dit geval ongeveer dezelfde nauwkeurigheid met 8 metingen krijgen als met die opzet waarbij 11 metingen op gelijke afstand van elkaar bepaald zijn

$$(8 \cdot \sqrt[3]{3} \approx 11) \text{ .}$$

Dit voorbeeld geeft ook een indruk van de bruikbaarheid van de discipline "Optimale opzet van proeven".

Dames en heren,

Er zijn nog enkele gebieden van de Biometrie die belangrijk zijn voor de landbouwwetenschappen zoals "BIOASSAY". Dat zijn alle biometrische methoden die voor biologisch onderzoek van stoffen met bepaalde werking als farmaceutische middelen of middelen in plantenmest worden gebruikt. Verder is er een nieuw gebied dat statistische ecologie genoemd kan worden en de biometrische methoden in de milieuwetenschappen toepast enz. Maar ik denk dat de

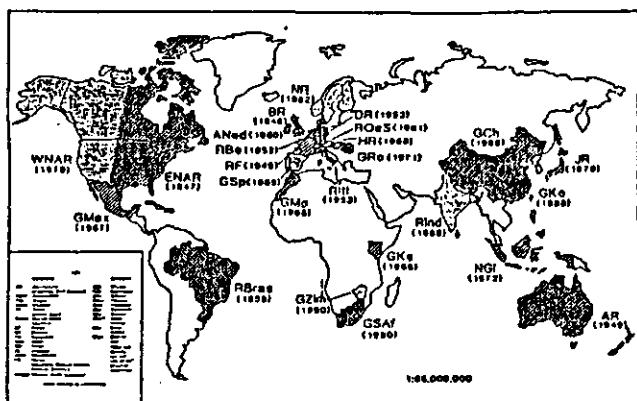
gegeven voorbeelden voldoende zijn om een algemene indruk te geven.

Ik wil graag nog zeggen dat de rol van de Biometrie niet eindigt met het wetenschappelijk onderzoek in de Landbouw. Statistische methoden kunnen ook bij de kwaliteitsbeheersing van landbouwprodukten worden gebruikt (statistische kwaliteitsbeheersing).

Ik zal nog enige opmerkingen maken om te beschrijven hoe de computer voor de toepassing in de Biometrie gebruikt kan worden.

Biometrie en computer

Toen de Biometric Society was opgericht in het jaar 1946 speelden computers nog geen grote rol in de Biometrie. Het aantal leden van deze vereniging groeide snel van ongeveer 200 in 1946 tot 6123 in 1992 in 25 regio's en nationale groepen zie figuur 4.



Figuur 4. Landen, regio's en groepen van de Biometric Society.

Gelijktijdig zijn rekenmachines, later hollerith-machines, grote centrale computers, personal computers en nu workstations steeds belangrijker voor de biometrici geworden. Fisher heeft in Rothamsted uren en uren verdelingstabellen en analyses van waarnemingen berekend die tegenwoordig in enige minuten gedaan kunnen worden. Voor het analyseren van gegevens hebben wij nu een groot aantal programma-pakketten zoals bijvoorbeeld GENSTAT (ontwikkeld bij biometrici in het landbouwonderzoek in Rothamsted), SAS, SPSS, BMDP, PSTAT, SYSTAT, STATGRAPHICS, MINITAB en vele andere. Zoals wij in figuur 1 kunnen zien bestaat de Biometrie niet alleen uit het analyseren van gegevens, veel belangrijker is het vastleggen van de stappen voor het uitvoeren van experimenten. Ook hier wordt de stiefkind situatie van het opzetten van proeven gereflekteerd, omdat alleen 3 programma-(pakketten voor dit gebied genoemd kunnen worden (PLAN, N, CADEMO).

De Biometric Society heeft op zijn "XV-th International Biometric Conference" 1990 in Budapest een "Invited paper Session" over "Biometrical Workstations" opgenomen met drie bijdragen^{26,27,28}. In deze sessie zijn de volgende minimale eisen voor een biometrisch workstation geformuleerd:

" A biometrical workstation should support the following research activities:

- precise formulation of the problem... including choice of a rough model
- selecting a special model
- selecting an experimental design (optimum sample size, optimum allocation...construction of block

- and other designs)
- common methodology based on normal distributions
 - basic methodology for frequency tables
 - basic procedures for comparison of distributions
 - other methods (nonparametric)...and...
 - all graphical displays should allow elementary transformation and
 - histograms with automatic and user-defined parameters
 - Box & Whisker plots
 - distribution functions
 - normal probability plot
 - scatterplot
 - quantile - quantile plot".

De mogelijkheid bestaat om CADEMO (voor de stappen 1-4) en GENSTAT (voor stap 6) als componenten van zo'n biometrisch werkstation te kiezen.

CADEMO (Computer Aided Design of Experiments and MOdelling) is een menu-gestuurd statistisch computerprogramma. Het programmapakket heeft als doel om bij experimenten de optimale proefopzet en modelkeuze te vinden. Dit wil zeggen dat door middel van CADEMO een passend statistisch model uitgekozen wordt en voor dit model de optimale proefopzet wordt bepaald.

CADEMO is ontwikkeld door verschillende instituten en universiteiten (Rostock, Berlin, Halle, Dresden), de ontwikkeling werd gecoördineerd door het instituut voor vee fokkerij in Dummerstorf naast Rostock.

CADEMO bestaat uit een stuurprogramma en

verschillende modulecomplexen, die zelf ook weer verschillende modules omvatten als kleinste bouwstenen van het systeem. Vanuit het stuurprogramma (module LPRO) kan men in de volgende modulecomplexen komen.

- WIBI Bioassay;
- LEDA Vernieuwingsproblemen, statistische kwaliteitsbeheersing;
- MIWA Schatten en toetsen van gemiddelden en kansen;
- WACH Analyse van groeikrommen;
- REA1 Regressie-analyse;
- ANLA Keuze en opbouw van proefschema's;
- AUWA Selectiemethoden;
- POPG Populatiegenetica;
- VARZ Schatten en toetsen van varianties;
- FEVE Veldproeven

Dames en heren,

Biometrici moeten als zij onderzoek willen doen of als zij advies willen geven verschillende wiskundige methoden gebruiken. Bijvoorbeeld: voor de constructie van proefschema's zijn er methoden van de discrete wiskunde belangrijk zoals

- Galois-theorie
- Grafentheorie
- Differentie stelsels
- Eindige projectieve of euclidische meetkunde
- "Orthogonal arrays".

Zo kunnen bijvoorbeeld volledig evenwichtige onvolledige blokkenproeven worden geconstrueerd

indien men een eindige projectieve meetkunde (geometrie) opbouwt, hierin deelruimten van een zekere dimensie bepaalt en de deelruimten gelijk stelt aan de blokken van een blokkenproef. Indien men verder de punten van de meetkunde gelijk stelt aan de behandelingen volgt uit de eigenschappen van een eindige projectieve meetkunde dat

- ieder blok (deelruimte) een gelijk aantal r van behandelingen heeft
- iedere behandeling gelijk vaak in een blok aangetroffen wordt
- ieder paar van behandelingen even vaak in één van de blokken aangetroffen wordt.

Daarmee is de zo opgebouwde onvolledige blokkenproef evenwichtig.

Om het aantal waarnemingen te bepalen, die voor een zekere nauwkeurigheid nodig is moeten wij de theorie en vooral de numerieke berekening van niet-centrale kansverdelingen beheersen.

Voor optimale proefopzetten in lineaire modellen zoals regressie analyse of variantie analyse is de kennis van niet-lineaire discrete en geheeltallige optimalisering nodig.

Indien wij nieuwe toepassingsgebieden van de Biometrie bekijken, als bijvoorbeeld

- de milieuwetenschappen
of
- de kwaliteitsbeheersing in de voor de landbouw belangrijke industrie
of
- de optimalisering van oogstprocedures van

natuurlijke populaties (vissen, bomen) op die wijze dat de winst maximaal is en gelijktijdig de populatie goed kan groeien,

hebben wij andere gebieden van de wiskunde nodig zoals:

- stochastische processen (b.v. punt processen) en tijdreeksen
- ruimtelijke (twee- en driedimensionale) statistiek
- parameter-vrije methoden
- niet-lineaire statistische modellen.

De belangrijke vragen van de insensitiviteit (robustheid) van biometrische methoden voor het geval dat bepaalde voorwaarden van de statistische theorie in de praktijk niet gegeven zijn, worden meestal met methoden van de moderne analyse (functionaalanalyse) beantwoord.

Mijnheer de Rector Magnificus, leden van het College van Bestuur,

Gaarne wil ik U dank zeggen voor mijn benoeming tot hoogleraar in de Wiskundige Statistiek. Een gebied dat zoals het op de Landbouwniversiteit bestaat voor mij synoniem is aan Biometrie. Ik ben bijzonder verheugd te constateren dat U mij de mogelijkheden heeft willen geven het vakgebied Biometrie aan de LU verder te ontwikkelen. Ik denk ook dat dit vakgebied kan bijdragen tot de van U in enkele brieven gevraagde verbetering van de kwaliteit van proefschriften en in het algemeen tot verhoging van de efficiency van het onderzoek. Dat kan in de eerste

plaats gebeuren door tijdens de fase van de opzet van een proef te streven naar de hoogst mogelijke precisie. Vanuit andere vakgroepen gezien bestaan de taken van een biometricus aan de LU voornamelijk uit onderwijs en advies. Om deze taken echter op een hoog kwalitatief niveau te kunnen uitvoeren is het absoluut noodzakelijk dat biometrici - of wiskundigen in het algemeen - ook blijven bijdragen aan de verdere ontwikkeling van hun eigen vakgebied, zoals dat bij andere fundamentele vakgroepen zoals b.v. fysische en organische scheikunde het geval is. Ik vraag U om uw uiterste best te doen om de wel zeer krappe onderzoeksruijnte binnen onze vakgroep uit te breiden.

Ik dank U, en de leden van de benoemingsadviescommissie voor het vertrouwen dat U in mij gesteld heeft.

Geachte medewerkers van Personeelszaken,

Voor mij als buitenlander ontstonden vele persoonlijke en organisatorische problemen. Ik wil een persoonlijk dankwoord uitspreken tot de heer Van Kamp (directeur PZ) en zijn medewerkers; Sulzle en Doppen die telkens bereid waren deze problemen op te lossen.

Hooggeleerde Brascamp,

Bij mijn komst vorig jaar werd ik getroffen door de hartelijke sfeer in de vakgroep Veefokkerij. Een biometricus kan niet zonder enige relaties uit zijn toepassingsgebieden werken. Ik dank U zeer voor Uw bereidheid tot samenwerking.

Hooggeleerde Struik,

Ik ben U en uw medewerkers zeer dankbaar voor de mogelijkheid bij 5 voorbeelden uit de vakgroep Landbouwplantenteelt en graslandkunde (in de toekomst "Agronomie") uitleg te geven hoe biometrici en speciaal het dialogue system CADEMO het opzetten van proeven kunnen ondersteunen. Ik denk dat dat een goed voorbeeld is van samenwerking tussen verschillende vakgroepen.

Geachte leden van de vakgroep Wiskunde,

Ik ben jullie dankbaar voor de manier waarop jullie mij in de vakgroep hebben opgenomen. De collega's van de sectie Wiskundige en Toegepaste Statistiek wil ik in het bijzonder danken omdat zij veel geduld met mij moeten hebben, omdat ik niet alleen anderstalig ben maar omdat ik ook een heel andere manier van werken in onderwijs en onderzoek gewend was. Mijn dank wil ik ook aan onze secretaresses uitspreken die niet alleen Duitse en Engelse teksten met veel inzet en betrokkenheid typen, maar altijd bereid zijn mijn pseudo-nederlands -niet zonder moeite- in goed Nederlands te transformeren.

*Hooggeleerde Van Beek, hooggeleerde Grasman,
Beste Paul en Johan,*

In het jaar, dat ik nu in Wageningen en in Nederland werkzaam ben, heb ik jullie leren kennen als voortreffelijke collega's, die me in deze nieuwe en in het begin ook vreemde omgeving altijd met nuttige raadgevingen en veelvuldige ondersteuning de weg hebben helpen vinden.

We delen de wens dat er binnen de vakgroep meer ruimte voor onderzoek zou moeten komen en dat de vakgroep Wiskunde in een aantal onderzoeksscholen gaat participeren. Op deze wijze wordt ook in het onderzoek de werkelijke toegevoegde waarde van de wiskunde gerealiseerd. De wijze waarop jullie trachten wiskunde niet alleen te ontwikkelen maar ook binnen en buiten de landbouwwetenschappen toe te passen steun ik van harte.

Lieve Renate,

Ik wil je hartelijk danken voor je liefdevolle steun en aandacht, dat een fijne thuisbasis voor mijn werk geeft.

Hartelijk dank voor de bereidheid onze vertrouwde omgeving en je werk op te geven om naar Nederland te verhuizen.

Liebe frühere Kollegen aus dem Forschungsbereich Biometrie des Instituts für die Biologische Landwirtschaftlicher Nutztiere in Dummerstorf und aus dem Fachbereich Mathematik der Universität Rostock,

Mit den meisten von Ihnen habe ich auch nach meinem Wechsel nach Wageningen noch engen Kontakt. Das betrifft sowohl die gemeinsame Forschung auf dem Gebiet der Robustheit biometrischer Methoden aber auch die Arbeiten zur Fertigstellung der englischen Ausgabe des Biometrischen Wörterbuches. Ich möchte Ihnen sagen, daß ich gerne an meine Tätigkeit in beiden Institutionen zurückdenke. Ich freue mich, daß der Fortbestand beider Gruppen unter den derzeitigen schwierigen Bedingungen gesichert ist.

*Dames en heren studenten van de richting
Agrosysteemkunde en van andere studierichtingen,*

Ik heb geprobeerd aan te tonen dat biometrische methoden in de meeste landbouwwetenschappelijke onderzoekrichtingen belangrijk zijn. Ik verwacht dat niet alleen veel studenten van de richting Agrosysteemkunde de speciale oriëntatie Biometrie kiezen maar ook dat studenten van andere richtingen de Biometrie als een afstudeervak kiezen en daardoor een bijdrage leveren tot samenwerking met andere vakgroepen.

Geachte toehoorders,

In het voorgaande heb ik een beeld geschetst van de rol van de Biometrie in de Landbouwwetenschappen maar ook van de rol die de Landbouwwetenschappen voor de Biometrie speelt. Het zou mij een genoegen zijn als wij in de toekomst aan de LU een nog diepere samenwerking tussen de vakgroepen en de biometrici kunnen krijgen.

Ik dank U allen hartelijk voor uw aanwezigheid en aandacht.

Referenties

- 1 Rasch, D. and Tiku, M.L. (eds). Biometrical Dictionary. Elsevier Publishing Co (in print)
- 2 Ludwig, W. Über die Anfänge der Statistik und Biometrik, *Biom. Zeitschr.* 1 (1959), 71-80.
- 3 Graunt, J. Natural and political observations made upon the bills of mortality, London, 1661.
- 4 Süssmilch, J.P. Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechts, aus der Geburt, Tod und Fortpflanzung erwiesen usw. Berlin, 1741.
- 5 Todhunter, J. A history of the mathematical theory of probability from the time of Pascal to that of Laplace. Cambridge and London, 1865.
- 6 s'Gravesande, G.J. *Introductio ad philosophiam; Methaphysicam et Logicam continens*. Leiden, 1733.
- 7 de Montmort, P.R. *Essai d'analyse sur le jeux de hazards*, Paris 1708.
- 8 "Student" The probable error of a mean. *Biometrika* 6 (1908) 1-25.
- 9 "Student" New tables for testing the significance of observations. *Metron* 5 (1925) 105-120.
- 10 Galton, F. Biometry. *Biometrika* 1,1 (1901) 7-10.
- 11 Weldon, W.F.R. Mendel's laws of alternative inheritance. *Biometrika* 1,2 (1902) 228-264.
- 12 Blanchard, B.A. On the inheritance of coat-colour of thoroughbred horses (Grandsire and Grandchildren). *Biometrika* 1,2 (1902) 361-364.
- 13 Yule, R. Mendel's laws and their probable relations to intra-racial heredity. *New Phytol.* 1 (1902) 193-207, 222-238.

- 14 Rasch, D. Further investigations on the robustness of the standardised selection difference. *Biom. Journ.* 28 (1986) 407-416.
- 15 Rasch, D. und Herrendörfer, G. *Handbuch der Populationsgenetik*. Deutscher Landwirtschaftsverlag, Berlin, 1990.
- 16 Fisher, R.A. *The design of experiments* Oliver 2and Boyd, 1935.
- 17 Bechhofer, R.E.A. single sample multiple decision procedure for ranking the means of normal populations with known variances. *Ann. Math. Statist.* 25 (1954) 16-39.
- 18 Rasch, D., Guiard, V. und Nürnberg, G. *Statistische Versuchsplanung. Einführung in die Methoden und Anwendung des Dialogsystems CADEMO*. Fischer Verl. Stuttgart-Jena-New York, 1992.
- 19 Verhulst, P.F. Notice sur la loi que la population suit dans son accroissement. *Corr. Math. et Phys. Publ. A. Quetelet* 10(1838) 113-121.
- 20 Gompertz, B. On the nature of the function expressive of the law of human mortality and on an new method of determining the value of life contingencies. *Phil. Trans. Roy.Soc.* (1825), 513-585.
- 21 Quetelet, A. *Recherches sur le poids de l'homme aux différents âges*. *Mem. de Acad. Bruxelles* 7 (1831).
- 22 Malthus, T.R. *Essay on the principle of population*. London 1798.
- 23 Mitscherlich, E.A. *Das Gesetz des Pflanzenwachstums*. *Landw. Jahresb.* 53(1919), 167-182.
- 24 Richards, F.J. A flexible growth function for empirical use. *Jour. Exper. Bot.* 10 (1959), 290 - 300.

- 25 Rasch, D. Optimum experimental design in nonlinear regression. *Comm. Statist. - Theory & Math.* 19(1990) 4789 - 4806.
- 26 Rasch, D. and Sawitzki, G.: What is a Biometrical Workstation? Joint preface at the invited paper session "Statistical Workstations" at the XVth International Biometric Conference, Budapest 2.7 - 6.7.1990. *Proceedings vol 1*, 141-144.
- 27 Sawitzki, G. Statistical workstations for data analysis and biometry. Invited paper at the XVth International Biometric Conference Budapest 2.7 - 6.7.1990, *Proceedings vol 1*, 145 - 152.
- 28 Rasch, D. Experimental Design using a Biometric Workstation. Invited paper at the XVth International Biometric Conference Budapest 2.7 - 6.7.1990, *Proceedings vol 1*, 153 - 156.
- 29 Engels, T. en Rasch, D. Ontwerp van proeven met behulp van het dialoogsysteem CADEMO geïllustreerd aan vijf gevallen op het gebied van de plantenteelt. Technical Note 92-07, vakgroep Wiskunde, LU Wageningen, juli 1992.