

BOEKBESPREKINGEN

Redactie: Arend D. Oosterhoorn

Correspondentie-adres: Zie bladzijde 1.

Sir Maurice Kendall, Alan Stuart & J. Keith Ord

The advanced Theory of Statistics

Volume 3 (Design and Analysis, and Time Series), 4e druk

Griffin, Londen, 1983, 780p, ISBN 0-85264-268-7

Kendall en Stuart (waarvan de eerstgenoemde overigens niet meer onder ons is) hebben bij deze herziene druk van het boek waaraan hun tweetal namen is verbonden gezelschap gekregen van J.K. Ord. Het is een bewijs van moed om een dergelijk monumentaal werk als mede-auteur te herzien, en Ord mag van mij daar een compliment voor hebben.

Ik ben gecharmeerd door de opzet van dit boek (samen met de twee andere delen): de min of meer klassieke basis van de wiskundige statistiek wordt er uitvoerig en in goede stijl in beschreven op een wijze die niet onderhevig is aan de mode van de dag. Dat laatste is natuurlijk ook te danken aan het feit dat de eerste uitgave al weer van vrij lang geleden is (1966 voor dit derde deel) terwijl het boek aan een redelijk frekwente en grondige revisie onderhevig is. Die revisies zijn natuurlijk niet volmaakt, maar dat moet je ook niet verwachten van een werk van deze breedte. Het bij de hand hebben van "Kendall & Stuart" is nog steeds de beste manier om eenvoudig toegang te hebben tot het meeste van wat de kern van de statistiek genoemd kan worden, inclusief redelijk veel verwijzingen naar (maar niet zo veel verwerking van) de moderne literatuur.

Dit deel behandelt variantie-analyse, statistische proefopzetten, steekproeftheorie, multivariate analyse en tijdreeksen. Het is bekend dat er over elk van deze onderwerpen geweldig veel materiaal in "Kendall & Stuart" te vinden is, geschreven vanuit een gezonde engelse praktische invalshoek. Afgezien van de algemene revisie, die op veel plaatsen bestaat uit beknopte omschrijvingen van resultaten uit de meer recente literatuur, zijn er substantiële aanvullingen op de derde druk aangebracht op het gebied van tijdreeksen (o.a. modellen en technieken uit de econometrische hoek en ruimtelijke processen, wat van Ord verwacht mocht worden) en op deelgebieden van de multivariate analyse (latente structuurmodellen en multidimensionale schaaltechnieken). Hier en daar worden er wel onderwerpen gemist. Bijvoorbeeld worden MINQUE schatters niet genoemd in het hoofdstuk over Model II Variantie-analyse, en moet het Kalman filter genoegen nemen met een minimale (onvoldoende) verwijzing. Het ene gebied wordt wat meer up to date behandeld dan het andere. In de hoofdstukken over tijdreeksen bv. wordt vrij veel recente literatuur behandeld, en wordt uitgegaan van een verfrissende veelheid van invalshoeken (geen uitputtendheid, dus); in de twee hoofdstukken over steekproeftheorie is de moderne literatuur echter minder goed verwerkt. Als symbool hiervan wordt alleen naar de eerste druk van het standaardwerk van Cochran verwezen, terwijl die toch ook zijn best heeft gedaan op revisies in latere drukken. De hoofdstukken over modelveronderstellingen (en afwijkingen daarvan) bij variantie-analyse, toetsen bij multi variate analyse en over classificatie zijn redelijk in hun verwijzingen naar recente literatuur.

Al met al is *The Advanced Theory* een boek waar je kritiek op kunt hebben maar waar je ontzettend veel aan hebt.

Tom A.B. Snijders, RUG

K. Matusita (Editor)

Recent developments in statistical inference and data analysis.

Proceedings of the International Conference in Statistics in Tokyo.

North-Holland, Amsterdam, 1980, 364p, ISBN 0-444-86014-1, f120,-

De bundel bevat 39 van 67 bijdragen gepresenteerd op de conferentie. Iedere redacteur zoekt een interessante, dekkende titel en deze is zeker niet de slechtste. Hij laat zien dat er geen oude koek verteld werd en dat de voordrachten alle mogelijke gebieden van de statistiek betreffen.

Volgens zeggen zijn de artikelen beoordeeld en zo is het boek een gebonden uitgave van een eenmalig algemeen statistiek-tijdschrift met artikelen ter lengte van gemiddeld zo'n negen pagina's getypt A4.

De redacteur wist geen betere volgorde voor de artikelen te vinden dan de achternamen van de eerste auteurs. Enige hulp wordt wel geboden door het achterin vermelde programma van de conferentie, maar ook daar zijn de hoofdcategorieën niet met name genoemd.

Uit dit programma blijkt dat aan 'bod zijn gekomen: experimentele proefopzetten, enkele medisch-biologische statistische toepassingen, lineaire modellen m.n. regressie- en variantie-analyse, verdelingstheorie, (univariate) testen, Bayesiaanse statistiek, stochastische processen en diverse onderwerpen. De nadruk ligt zeer sterk op de mathematisch statistische kant (31 artikelen), meer dan op de toegepaste statistiek en data-analyse (8 artikelen). Deze laatste behandelen de volgende onderwerpen: analyse van zinsstructuren (Chan), Simultane analyse van meer dan twee matrices (Escoufier), ruimtelijke patronen bij territoria van dieren (Hasegawa & Tanemura), grootte van dierpopulaties (Hayashi), keuze van aantal clusters (Matusita & Ohsumi), programma's voor niet-lineaire kleinste kwadraten (Nakagawa & Oyanagi), kwantificeren en effecten van kwalitatieve factoren (Tanaka), programma's voor medische data analyse (Tango & Kariya), scheve verdelingen en veilige doses (Yanagimoto & Hoel). Een aardig aspect van deze bundel is, dat het een overzicht geeft van de (mathematische) statistiek bedreven in Japan, aangenomen dat de bundel hiervan een redelijke afspiegeling is. Zelf zou ik de voorkeur geven aan een overdruk van enkele artikelen.

P.M. Kroonenberg, R.U. Leiden

Gerhard Meinlschmidt

Statistische Methoden zur Berechnung van Preisniveauindizes integrierter Wirtschaftsräume

Angewandte Statistik und Okonometrie Heft 28

Vandenhoeck & Ruprecht, Göttingen, 1985, 142p, ISBN 3-525-11262-9

Dit boek behandelt het probleem van de berekening van (intertemporele) prijsindices voor een zgn. "geïntegreerde economie". Men denke in dit verband aan de Europese Gemeenschap. Daarbij is sprake van tweeërlei aggregatieproblematiek: waardebepalingen van de diverse landen moeten worden geaggregeerd tot een waardebedrag voor de geïntegreerde economie, en waardebedragen voor consumptie, investeringen, e.d. moeten worden samengeteld tot waardebedragen voor het bruto binnenlands produkt zowel voor de landen als voor de geïntegreerde economie. Voor de interregionale aggregatie zijn

interregionale prijsindices, ook wel koopkrachtpariteiten genoemd, nodig. Brengen we vervolgens de tijdsdimensie in rekening, dan zal duidelijk zijn dat er allerlei consistentieproblemen ontstaan, bijvoorbeeld tussen de koopkrachtpariteiten van land j t.o.v. land h gemeten op tijdstippen t en $t+1$ en de intertemporele prijsindices van de landen j resp. h . Een en ander wordt door de auteur dan ook nog gecompliceerd doordat hij niet, zoals gebruikelijk, een waardeverhouding (tussen twee regio's op één tijdstip, of tussen twee tijdstippen voor één regio) wenst te ontbinden in twee componenten, t.w. een prijs- en een hoeveelheidscomponent, doch in vier, t.w. een prijsniveauindex, een hoeveelheidsniveauindex, een prijsstructuurindex en een hoeveelheidsstructuurindex.

De auteur pakt de problemen op een in sommige kringen gemeengoed geworden manier aan. Hij begint met het specificeren van een aantal criteria, "tests" genaamd, waaraan de diverse indices zouden moeten voldoen. Vervolgens gaat hij van een aantal formules en methoden na in hoeverre ze aan de diverse criteria voldoen. Het resultaat is niet spectaculair. Voor de niveau-indices beveelt hij de Törnqvist-indices (door hem Walsh-indices genaamd) aan, en voor de structuur-indices de quotienten van Fisher- en Törnqvist-indices. Het consistentieprobleem, tussen intertemporele en interregionale indices, wordt opgelost door het construeren van geschikte gemiddelden van combinaties van beide typen indices.

De abundante wiskundige notatie (bijv. door het gebruik van een veelheid van super- en subscripten) en het minimum aan intuïtief-economische argumentatie wekken de indruk dat er geen speld tussen het betoog te krijgen is.

Helaas is dat slechts schijn. Op een essentieel punt in zijn betoog, t.w. de constructie van waardebedragen voor de geïntegreerde economie, zou de auteur er beter aan gedaan hebben de interpretatie van zijn formules eens wat nader te bekijken. Wat er nu staat in formule (4.1) in combinatie met (2.3.2) betekent dat in de constructie het te construeren object verondersteld wordt bekend te zijn. Dezelfde fout manifesteert zich in paragraaf 5.4.2.8 en in het schema op pagina 120. Het gehele betoog komt hierdoor op losse schroeven te staan.

Daarnaast heb ik diverse kleinere punten van kritiek:

- a. De definitie van de Paasche-index (pag. 55/56) is verkeerd. Wat gegeven wordt is de index van Palgrave. Extra vervelend is daarbij dat de Fisher index gedefinieerd wordt in termen van de Laspeyres en de Paasche-index. De lezer kan daardoor gemakkelijk op een dwaalspoor geraken.
- b. De uitgebreide literatuurlijst functioneert niet in het betoog. Voor de niet ingevoerde lezer is daardoor niet duidelijk wat nu de eigen bijdragen van de auteur op het onderhavige terrein zijn. Zo is bijvoorbeeld de tweede karakterisering van de prijsindex van Fisher (pag. 62/63) reeds te vinden in een, overigens in de literatuurlijst niet opgenomen, artikel van Van IJzeren uit 1952.
- c. In tegenstelling tot het gestelde op pag. 79 voldoet de index van Törnqvist (volgens de auteur: Walsh) niet aan de circulariteitseis.
- d. Van de vele slordigheden die mijn oog troffen vermeld ik die welke te vinden zijn in de definities van koopkrachtpariteiten volgens Van IJzeren en volgens Geary-Khamis (pag. 90/91/92).

Konklusie: Als inleidend werk is het boek, gezien de vele feilen, ongeschikt, en of de op het onderhavige terrein werkzame statisticus er verder mee komt betwijfel ik in hoge mate.

D.V. Lindley

Making Decisions

John Wiley & Sons, London, 1985, xii+207p, ISBN 0-471-90808-8

Dit is een tweede editie van het boek. Ten opzichte van de eerste, uit 1971, zijn enige correcties aangebracht, is hier en daar het werk meer didactisch verantwoord gemaakt (enigszins ten koste van formele correctheid), en is een hoofdstuk IX toegevoegd, op grond van nieuwe ontwikkelingen op het gebied van procedures voor het bepalen van subjectieve kansen en nutten. De bedoeling van dit boek, en van het hele werk van de schrijver, wordt goed duidelijk uit het begin van paragraaf 4.5:

The recommended procedure for making decisions is to

- (1) List the possible decisions $(d_1, d_2, \dots, d_m)$;
- (2) List the uncertain events $(\theta_1, \theta_2, \dots, \theta_n)$;
- (3) Assign probabilities to the latter $(P(\theta_1), P(\theta_2), \dots, P(\theta_n))$;
- (4) Assign utilities $u(d_i, \theta_j)$ to the consequences (d_i, θ_j) ;
- (5) Choose that decision of maximum expected utility

$$\bar{u}(d_i) = \sum_{j=1}^n u(d_i, \theta_j) P(\theta_j)$$

This procedure follows from basic ideas of coherence in judging both uncertain events and consequences. It is the simple message of this book. If we were to adopt the style of a religious text, I would say to you 'Go ye forth and maximise your expected utility'.

Einde citaat.

Het zal duidelijk zijn dat het doel van het boek maar één aspect van beslissingstheorie betreft: het modelleren van onzekerheid in beslissingssituaties door middel van subjectief verwacht nut. En dit niet voor descriptieve toepassingen, waar wordt bekeken hoe in de praktijk beslissingen worden genomen, maar voor prescriptieve toepassingen, waar wordt bekeken hoe het beste beslissingen genomen kunnen worden. Lindley baseert het gebruik van verwacht nut op het "coherentie-principe". Oorspronkelijk is dit een idee van Bruno de Finetti, zie bijvoorbeeld De Finetti (1931, 1974). Het betreft beslissingssituaties waarin er onzekerheid is betreffende de financiële consequenties van handelingen, en het houdt in dat een beslisser moet voorkomen dat een zogenaamd "Dutch book" tegen hem gemaakt kan worden. Een Dutch book is een serie van handelingen, die ieder voor zich door de beslisser als gunstig worden beoordeeld, maar die, in combinatie, met zekerheid een negatieve opbrengst leveren. Natuurlijk zal geen enkele Nederlander zo een combinatie van handelingen ambiëren, dus de terechtheid van de term "Dutch book" kan in dit verband slechts beaamd worden.

Het boek van Lindley poogt een en ander helder, en zeer grondig, uiteen te zetten, en richt zich op een niet-wiskundig-geschoold publiek. Het besteedt daarbij zeer uitgebreid aandacht aan de elementaire intuïtieve achtergronden, en maakt alle afwegingen tussen formele correctheid en didactische belangen ten gunste van de didactische belangen. Misschien ook daarom worden er niet een exacte definitie van coherentie, en een formele afleiding van de implicaties van coherentie, gegeven, maar wordt de term coherentie losjes gebruikt, soms als synoniem voor rationeel, dan weer als synoniem voor transitief, monotoon, enz.

Nadat in de eerste vier hoofdstukken op informele wijze is uiteengezet hoe het subjectieve verwachte nut model afgeleid kan worden, volgt in hoofd-

stuk 5 een behandeling van nutsfuncties voor geld en risico-afkerigheid, in hoofdstuk 6 de formule van Bayes en het gebruik ervan o.a. in de statistiek, in hoofdstuk 7 een manier om informatie te waarderen, in hoofdstuk 8 het gebruik van de beslissingsbomen, in hoofdstuk 9 volgen procedures om subjectieve kansen en nutten te bepalen, en tenslotte volgt in hoofdstuk 10 een "appreciation" van het werk welke vooral ingaat op de complicaties van groepsbeslissingssituaties, en speltheorie. Aan het eind van de meeste hoofdstukken volgen opgaven, waarvan de antwoorden achter in het boek zijn opgenomen.

Het boek is prettig leesbaar en goed verzorgd, en kan de niet-wiskundig-gechoolde lezer bewust maken van de verre draagwijdte van het werk van vooral Bruno de Finetti. Wel is het soms wat ongenueanceerd, coherentie wordt voorgesteld als vanzelfsprekend juist, i.p.v. discussiabel, en alle andere opzetten worden als onjuist beoordeeld (zie bijvoorbeeld paragraaf 3.7 over "fuzzy sets"). De visie (bijvoorbeeld in de laatste regel van de twee-voorlaatste alinea van paragraaf 10.13) dat het schenden van coherentie het enige is wat men fout kan doen, is als de visie van de logicus die zegt dat het enige wat een astronoom fout kan doen is het schenden van de regels van de logica. Overigens kan deze visie ook in het werk van De Finetti gevonden worden.

Kortom, een prettig leesbaar, informeel boek, "breed geschreven", met een vurig, soms ongenueanceerd, pleidooi voor het gebruik van het subjectief verwacht nut model.

P.P. Wakker, KUN

Referenties

- de Finetti, B. (1931), "Sul Significato Soggettivo della Probabilità", *Fundamenta Mathematicae* 17, 298-329.
de Finetti, B. (1974), "Theory of Probability", Volumes I and II. Wiley, New York.

Aloke Dey

Orthogonal Fractional Factorial Designs

Wiley Eastern, New Delhi, 1985, 133p, ISBN 0-85226-165-9

De theorie van proefopzetten heeft tot doel proeven zo in te richten dat ze een maximum aan informatie leveren met betrekking tot een gegeven vraagstelling. In de jaren 20 is deze theorie revolutionair veranderd, vooral door de introductie van het loten. In 1926 waren de belangrijkste principes van de huidige theorie zo ongeveer wel geformuleerd.

De 60 jaren die sindsdien zijn verlopen stonden in het teken van de consolidatie: er zijn geen wezenlijk nieuwe begrippen toegevoegd. Er is voortdurend aan het onderwerp gewerkt, bijvoorbeeld aan de theoretische onderbouwing van de lotingstheorie. Ook is veel aandacht besteed aan het systematisch in kaart brengen en op bruikbaarheid ziftten van alle mogelijkheden. Hiermee is niet gezegd dat er in deze jaren niets interessants is gebeurd op het gebied van de opzet en analyse van proeven. Recentelijk zijn bijvoorbeeld, door het beschikbaar komen van steeds betere rekenfaciliteiten, veel schema's praktisch analyseerbaar geworden die het vroeger niet waren, hetgeen tot zeer bruikbare resultaten heeft geleid. Ook bleek de theorie allerlei verrassingen te bieden en doorkijkjes naar geheel andere delen van de wiskunde. Zo bleek bijvoorbeeld, tot genoegen der zuivere wiskundigen, dat abstracte stukjes algebra uiterst toepasbaar waren bij het systematisch afzoeken van allerlei mogelijke proefopzetten. En het is ook niet onaardig

te zien dat de statistisch voor de hand liggende manier om een zogeheten 2ⁿ proef te analyseren eenvoudig kon worden uitgebreid tot een zeer efficiënte numerieke Fourier-transformatie. Om maar niet te spreken van het feit dat de fractionele proeven waarover het voorliggende boekje gaat in een handomdraai te veranderen zijn in error-correcting codes.

Het besproken boekje is een monografie over de opzet van fractionele proeven. Dit zijn experimenten, die vooral in de industrie populair schijnen te zijn, waarbij in 1 proef simultaan meerdere factoren worden onderzocht zonder dat alle mogelijke factor-combinaties worden gerealiseerd. De analyse van zulke proeven moet dan wel gebaseerd worden op de veronderstelling dat bepaalde interacties verwaarloosbaar klein zijn: bijvoorbeeld dat het effect van zeg factor A niet afhangt van het niveau van zeg factor B.

De oudste bekende fractionele proef stamt uit 1934: een proef van Tippet, bestaande uit 25 eenheden, waarin 6 factoren, ieder met 5 niveau's worden onderzocht, zodat slechts 25 van de 5⁶ mogelijke factor-combinaties worden gerealiseerd. Het zal duidelijk zijn dat dergelijke proefopzetten grote voor- en nadelen hebben, en dat in de vermelde proef de nadelen misschien wel overwegen: onder de altijd wat speculatieve veronderstelling dat er geen interacties zijn kan men inderdaad de hoofdeffecten van de 6 factoren schatten, maar dan nog kan men op grond van de proef niets zeggen over de nauwkeurigheid van die schatters. Dit is geen uitzondering: in veel fractionele proeven is over de nauwkeurigheid weinig of niets te zeggen. Tussen haakjes: doordat in het bovengenoemde schema slechts een heel bepaalde en zeer kleine fractie van alle mogelijke factor-combinaties wordt gerealiseerd kan men een eventuele fout die in het schema is geslopen veelal feilloos onderkennen (en herstellen als de proef nog niet is uitgevoerd): men kan het proefschema dus eenvoudig omzetten in een error-correcting code, dat nog goed is ook.

In het boekje wordt zo systematisch mogelijk opgesomd welke fractionele schema's er bekend zijn en hoe men die kan maken. Voor de specialisten: de auteur beperkt zich tot orthogonale fractionele proeven, symmetrisch en asymmetrisch, van resolutie 3, 4 en 5. In het boekje zijn veel recente, over de tijdschriften verspreide, ontwikkelingen samengevat: De schrijver streeft naar volledigheid, maar slaagt daar, buiten zijn schuld, steeds minder in naarmate de stof complexer wordt: de lijst van symmetrische schema's met hooguit 81 eenheden is compleet, dezelfde lijst voor asymmetrische schema's bevat alles wat er tot nog toe bekend is, en daarna gaat het steeds verder bergafwaarts met de volledigheid.

De schrijver zegt weinig over de situaties waarin met vrucht fractionele schema's kunnen worden gebruikt, en nog minder over de situaties waarin men dit beslist niet moet doen. De mogelijkheid blokken in deze proeven aan te brengen wordt al te summier besproken. De problemen die voortvloeien uit de soms slechte schatbaarheid van nauwkeurigheden worden 'opgelost' door een kort stukje over een grafische procedure die bekend staat als de half-normal-plot techniek van Daniel. Voor dit soort statistische zaken kan men beter in een boek als COCHRAN & COX terecht, dat overigens een veel minder volledige lijst van fractionele proeven bevat.

De in het boek gecatalogiseerde schema's worden niet expliciet uitgeschreven, wat jammer is, zeker voor de incidentele gebruiker. In plaats daarvan wordt beschreven hoe de genoemde schema's kunnen worden gemaakt. Maar er wordt weer niet uitgeduid waarom deze constructiemethoden leiden tot het gewenste resultaat.

Al met al lijkt het boekje mij vanwege zijn volledigheid wel bruikbaar voor degenen die vaak fractionele proeven opzetten. Omdat het boekje weinig combinatorisch of statistisch inzicht geeft zou het op den duur eigenlijk

vertaald moeten worden in een computerprogramma dat helpt bij het vinden van proefschema's, en dat deze schema's ook volledig uitschrijft.

Michiel Jansen, ITI-TNO

M.L. Eaton

Multivariate Analysis, A Vector Space Approach

John Wiley & Sons, 1983, 512p, ISBN 0-471-02776-6

The completion of this review has been delayed because the reviewer could not make up his mind. On the hand he was impressed by the display of knowledge. He was grateful for the beautiful mathematics presented and for the nice explicit way in which this was done. Students seemed to like the book. They were able to start reading almost everything. Thus the book is beautiful, useful, etc. One should not hesitate: buy the book !

On the other hand the reviewer was annoyed by the complete lack of reference to actual applications. Multivariate analysis has its roots in anthropometry, biometry, psychometry, etc. Its fruits should germinate on that soil. Eaton presents his topic as a piece of art, a branch of pure mathematics. He introduces new terminology to obtain more esthetic formulations and to perform computations in a concise and beautiful manner.

Those who are involved in actual applications know that respect for the mathematical argument, elegance of formulations, etc. is one thing, but respect for the data and for the people involved in their processing and interpretation is another indispensable affair. A more complete understanding of multivariate models and techniques by the statisticians -Eaton's motivation for writing this book- does not always go together with a more complete understanding by those who have to employ and interpret the results. In this respect, new terminology and emphasis on normality, spherical symmetry, invariance, hypothesis testing, etc. may be irrelevant. Eaton's book is completely mathematics-oriented. It does nothing to convince the reader that an application-oriented attitude is more appropriate, at least for those who want to make mathematical statistics useful.

Eaton's book shows the reader how beautiful multivariate analysis, at least a part of it, can be. The emphasis is on vector space methods and invariance considerations. The first part of the book is a high-quality introduction to the required mathematics: vector space theory, random vectors (with values in inner product spaces), the normal distribution on a vector space, linear statistical models, matrix factorizations and Jacobians, topological groups, and invariant measures. The second part of the book contains high-quality applications of invariance, e.g. to the Wishart distribution, inference for means in multivariate linear models, and canonical correlations. Throughout the book many (mathematical) problems are formulated. At the end corresponding solutions are indicated.

Eaton's book is a masterpiece, at least as a work of art and mathematics.

W. Schaafsma

A: Stuart

The Ideas of Sampling

Ch. Griffin & Company, High Wycombe, 1984, vii+91p, £ 4,95

Dit boek is een kleine revisie van het boekje dat in 1962 voor het eerst verscheen als 'Basic ideas of scientific sampling' en waarvan in 1976 de 2e editie verscheen (recensie hiervan in VVS-bulletin 10 (1977) nr. 1). Het boek verklaart de hoofdgedachten die betrokken zijn bij het steekproef-trekken uit een eindige populatie om het populatiegemiddelde of populatie-totaal te schatten.

Het boek is verdeeld in 44 paragrafen die niet in hoofdstukken gerangschikt zijn. De behandelde hoofdonderwerpen zijn:

1. De noodzaak van een steekproeftheorie;
2. Enkelvoudige aselechte steekproef en betrouwbaarheidsintervallen voor het populatiegemiddelde (of totaal) berustend op asymptotische resultaten als de steekproefomvang groot wordt; verder een korte discussie over de ratioschatter;
3. Gelaagde steekproeven en de steekproef allocatie over de lagen;
4. Tros- (= 'cluster') steekproeven met en zonder deelsteekproeven; steekproeven getrokken met en zonder gelijke kansen.

De auteur demonstreert de verschillende steekproefmethoden aan een populatie van 6 leerlingen van een klas. De steekproefverdeling van de verschillende schatters wordt dan expliciet afgeleid door alle mogelijke steekproeven te tabelleren en met een minimum aan formules de schatters en de variantie ervan te berekenen. De benodigde technische uitdrukkingen en termen worden tijdens de uiteenzetting gedefinieerd.

Toch zou de opzet van het boek verbeterd kunnen worden door na de expliciete afleiding van de steekproefverdeling van een schatter en de berekeningen van de variantie bij die verdeling, dit samen te vatten in een formule. Nu moet de lezer onthouden dat tot pagina 21 de 'population variance' beschreven is via een berekening op pagina 13 en 14 als

$\sum_{i=1}^N (y_i - \bar{y})^2 / N$ en de 'sample variance' op pagina 18 als $\sum_{j=1}^n (y_j - \bar{y})^2 / n$ en dat dezelfde begrippen maar nu tussen aanhalingstekens zijn

"population variance" is $\sum_{i=1}^N (y_i - \bar{y})^2 / (N-1)$ en

"sample variance" is $\sum_{j=1}^n (y_j - \bar{y})^2 / (n-1)$,

en vanaf deze herdefinitie op pagina 21 meteen de aanhalingstekens weer weggelaten worden. Didactisch is het toch beter om gebruik te maken van de symbolen σ^2 en $\bar{\sigma}^2$ voor de eerste twee begrippen en S^2 en s^2 voor de laatste twee, zoals Cochran dit doet.

De argeloze lezer moet er aan denken, dat vanaf het begin (pagina 5) onder een 'sample' een steekproef getrokken zonder teruglegging wordt verstaan. Pas op pagina 26 wordt losjes opgemerkt:

'In our discussion sofar, we have made one assumption concerning the sampling procedure which we must now examine', en wordt 'sampling with replacement' besproken, waaruit blijkt dat ervoor steeds 'sampling without replacement' was gedaan.

Storende drukfouten zijn

op pagina 73, bovenaan 'sampling variance = 32/12' dat 38/12 moet zijn
op pagina 81, in 'Table 44 sample no. 11' is de 'chance of selection' niet 1/36 maar 1/16;

op pagina 85, in 'Scheme D, cluster III' is de 'chance of selection' niet 1/2 maar 1/3;

Samenvattend kan dit boekje gode diensten doen voor een onderzoeker of beginnend statisticus om dit te lezen naast een echt leerboek over steekproeftechnieken.

L.R. Verdooren, LHW

P. Griffiths & I.D. Hill (Editors)

Applied Statistics Algorithms

Ellis Horwood Ltd (Wiley), Chichester, 1985, 307p, ISBN 0-85312-772-7, £180 (voor onderwijs £120)

In *Applied Statistics Algorithms* wordt een dertigtal subroutines die zijn te gebruiken voor statistische programma-ontwikkeling gepresenteerd. De subroutines vormen een deelverzameling van de collectie die is gepubliceerd in het tijdschrift *Applied Statistics*, maar zijn voor het boek opnieuw getest, geactualiseerd en geuniformeerd. Behalve de code, wordt ook een beschrijving van het algoritme gegeven, die varieert van een literatuurverwijzing tot een tamelijk minitieuze beschrijving. Bovendien wordt de nauwkeurigheid van het algoritme aangegeven. Vijf subroutines zijn geschreven in Algol 60; hiervan wordt echter ook een versie in Fortran 66 en 77 gepresenteerd. Van vier subroutines wordt slechts een Fortran 66 versie gegeven, terwijl de overige 24 zowel in Fortran 66 als Fortran 77 programma's zijn te gebruiken. De subroutines zijn van uitgebreid commentaar voorzien, zodat ze goed leesbaar zijn. Voorts worden de formele parameters nauwgezet beschreven en worden de besprekingen en eventueel gebruik in dubbele precisie aangegeven. De verzameling subroutines is ruwweg onder te verdelen in de volgende klassen:

- . matrix-bewerkingen (inverse, eigenwaarden en -vectoren)
- . verdelingsfuncties (centrale en niet-centrale t-verdeling, normale verdeling, χ^2 -verdeling)
- . cluster-analyse ("single linkage", "k-means")
- . diversen (updating van steekproef-gemiddelden en dispersie-matrix, incomplete beta en gamma integraal, Fourier transformatie, (psuedo) random generator, Kalman filter, log-lineaire fit voor contingentie-tabellen, minimalisatie volgens een Simplex methode)
- . data representatie (half-normale plot, histogram, multi-dimensionale tabellen, scatter-diagram).

Al met al dus een zeer diverse verzameling van toepassingsgebieden. Van een aantal subroutines, met name voor matrix-bewerkingen en verdelingsfuncties, is uiteraard een gelijkwaardige versie te vinden in de bekende programmatheken zoals IMSL en NAG, die op vrijwel alle (grotere) computersystemen beschikbaar zijn. Van andere subroutines is een implementatie niet of slechts incidenteel te vinden. De beschreven verzameling is overigens via de uitgever van het boek in machine-leesbare vorm te verkrijgen. De prijs is £180 (voor onderwijs: £120) + £5 verzendkosten.

De vraag wie wat aan dit boek heeft, is moeilijk te beantwoorden, enerzijds omdat een groot deel van een verzameling subroutines zoals hier wordt gepresenteerd al kant en klaar in het computersysteem beschikbaar is of is inbegrepen in een standaardpakket, anderzijds door de grote diversiteit van de verzameling. Een werker op een specifiek vakgebied zal daardoor weinig van zijn gading vinden. De grote diversiteit maakt het boek aan de andere kant geschikt voor diegenen, die verantwoordelijk zijn voor het beschikbaar stellen van programmatuur, zoals rekencentra. Verder kan het om dezelfde reden worden aanbevolen voor bibliotheken.

Ook als referentie voor de werking van een (geïmplementeerd) algoritme kan het boek een functie hebben, bijvoorbeeld in een cursus in één van de deelgebieden die door de verzameling algoritmen wordt bestreken.

Tenslotte wil ik nog opmerken, dat het boek opvalt doordat het zo goed verzorgd is: niet alleen zijn de afzonderlijke subroutines zorgvuldig gedocumenteerd, maar ook het boek als geheel maakt door een systematische presentatie en overzichtelijkheid een prettige indruk.

L.Th. van der Weele, RUG