

Redactie: Arend D. Oosterhoorn
Correspondentieadres: Zie bladzijde 1.

P. Debets, A van Meurs, S.H.J. Veling, A. Verbeek (red.)

Symposium Statistische Software 1985

Technisch Centrum FSW, Universiteit van Amsterdam, 1985, 296 p.,
prijs f 35,--.

De bundel Symposium Statistische Software 1985 is uitgegeven ter gelegenheid van het tweede Symposium Statistische Software dat werd georganiseerd door de Sectie Statistische Programmatuur van de Vereniging voor Statistiek in samenwerking met het Technisch Centrum FSW van de Universiteit van Amsterdam. De - zeer goed verzorgde - bundel omvat een 15-tal bijdragen over statistische programmatuur, daarin gebruikte methoden en methodieken om het gebruik van deze programmatuur te vergemakkelijken.

Bakker presenteert een interactieve generator voor SPSS set-ups, waarbij (detail) kennis van de SPSS commando-taal niet nodig is. Van Blokland-Vogelesang, Vlaander en De Swart bespreken het PROFIT systeem en geven daarbij aanwijzingen voor verbeteringen. De Boer behandelt expertsystemen voor statistische analyse, in het bijzonder voor de keus van de analyse-methode en de interpretatie van de uitkomsten.

Gaaf en De Boer inventariseren statistische programma's voor micro-computers en bepalen daarbij voor elk programma een score op basis van de beschikbare faciliteiten. Bovendien concluderen zij dat in ieder geval de betere pakketten een grote rekennauwkeurigheid hebben.

De Graauw bespreekt criteria voor data-base programmatuur voor de invoer van gegevens.

De Greef, Bakker, Buisman en Van de Vlucht behandelen het kennis-systeem voor statistische consultatie.

Groeneveld presenteert het systeem DAPHNE, een bibliotheek van interactieve programma's voor de invoer, updating en beschrijving van gegevens.

Hilhorst geeft verslag van een poging een data-base systeem te ontwikkelen, waarmee een keus uit diverse pakketten kan worden gemaakt voor het uitvoeren van een statistische analyse, zonder dat de gebruiker hoeft te weten welk pakket moet worden gebruikt en de specificaties hiervan hoeft te kennen.

Keller, Verbeek en Bethlehem presenteren ANOTA, een snelle methode voor de analyse van contingentie-tabellen analoog aan multiple regressie.

Koppelaar en De Jong behandelen methoden voor de berekening van exacte verdelingen van stochastische grootheden door geautomatiseerde manipulatie van formules voor kansverdelingen.

Te Meerman presenteert een nieuw algoritme voor het fitten van MDS modellen; de kwaliteit hiervan wordt getoetst met behulp van

Monte-Carlo methoden.

Van Meurs en Banens analyseren het PC-pakket ABSTAT; hun conclusie is dat het pakket weliswaar zeer vriendelijke data-manipulatie bevat, maar statistisch mager is, terwijl de reken-nauwkeurigheid dubieus is.

Molenaar en Broersma evalueren de perceptie van verschillende presentaties van statistische informatie; zij geven hierbij enige interessante aanbevelingen voor pakket-ontwerpers.

Stokman en Van Veen geven een overzicht van de stand van zaken met betrekking tot hardware, software, programmeer-opvattingen, nieuwe toepassingsgebieden in het licht van een programmatuur-beleid ten behoeve van de gedrags- en maatschappijwetenschappen; zij bevelen hiervoor een landelijk samenwerkingsorgaan aan.

Weesie gaat in op kleinste kwadratenmethoden voor data-analyse van grote steekproeven.

Al met al een zeer uiteenlopend scala van onderwerpen, waarin een ieder die zich bezig houdt met statistische programmatuur en het gebruik ervan wel enige interessante verhalen zal vinden. Ook de methodoloog en de wiskundig-statisticus zal in dit boek zeker iets van zijn gading vinden.

Het boek geeft een uitstekend overzicht van de stand van zaken van statistische programmatuur en de richting waarin verder onderzoek plaats vindt. Met nieuwsgierigheid kan worden uitgekeken naar het volgende symposium dat in 1987 door de Sectie Statistische Programmatuur zal worden georganiseerd.

De bundel is op het moment uitverkocht maar zal worden herdrukt als er hiervoor voldoende belangstelling is. Te bestellen bij: Technische Centrum FSW, UvA, Herengracht 196, 1016 BS Amsterdam.

L. Th. van der Weele, RUG

Richard J. Brook and Gregory C. Arnold

Applied regression analysis and experimented design

Marcel Dekker, Inc., New York, 1985, 237 p., ISBN 0-8247-7252-0, prijs \$ 39.75.

De auteurs zeggen bij de inleiding van het boek dat de titel ook "Fitting a model to data" had kunnen zijn.

Dit is een zeer ter zake doende opmerking omdat zij zowel de keuze van het juiste model als het verzamelen en analyseren van gegevens uitvoerig bespreken.

Dat de auteurs grote waarde toekennen aan de manier waarop gegevens worden verzameld wordt geïllustreerd met de bespreking van onderwerpen zoals herhaling, blokindeling, experimentele eenheid, verloting, split-plot schema's en dergelijke.

In een appendix zijn een groot aantal waarnemingsuitkomsten gegeven, die zowel bij de voorbeelden als in de opgaven verschillende keren worden gebruikt.

De auteurs besteden in vrijwel elk hoofdstuk veel aandacht aan het analyseren van residuen terwijl er een apart hoofdstuk wordt gewijd aan het bespreken van eigenaardigheden in de waarnemingen. Omdat het boek een groot aantal opgaven bevat, waarvan ook de

antwoorden in druk te verkrijgen zijn, is het behalve voor zelfstudie ook geschikt om in cursussen te worden gebruikt.

Enige basiskennis van statistiek en lineaire algebra is daarbij noodzakelijk.

De titels van de hoofdstukken zijn:

1. Fitting a model to data.
 2. Goodness of fit of the model.
 3. Which variables should be included in the model.
 4. Peculiarities of observations.
 5. The experimental design model.
 6. Assessing the treatment means.
 7. Blocking.
 8. Extensions to the model.
- Appendix A. Review of vectors and matrices.
 Appendix B. Expectation, linear and quadratic forms.
 Appendix C. Data sets.

Een nadeel van het boek is wellicht dat sommige onderwerpen, zoals principale componenten, ridge regression en baysiaanse aanpak in slechts enkele bladzijden worden besproken. Toch is het een prestatie van de auteurs dat zij in ruim 200 bladzijden zoveel aspecten op een begrijpelijke, praktische en fundamentele wijze bespreken.

Samenvattend kan ik dan ook dit boek van harte bij u aanbevelen.

P.M. Upperman, Nederlandse
 Philips Bedrijven BV

C.B. Tilanus, O.B de Gans, J.K. Lenstra (editors)

Quantitative Methods in Management

Case studies of Failures and Successes

John Wiley & Sons, New York, 1986, 272 p, ISBN 0-471-90841-X,
 \$17.50.

Ter gelegenheid van het 25-jarig bestaan van de Sectie Operationele Research (SOR) van de Vereniging voor Statistiek is in 1983 een bundel opstellen in het Nederlands verschenen over Operationele Research in de praktijk. Deze bundel is op ruime schaal verspreid.

Thans is de engelstalige editie verschenen kennelijk gemotiveerd door het eerdere succes. Dat succes is vooral te danken aan het feit dat de doelstelling van het boek is het beschrijven, in brede zin, van het gebruik van kwantitatieve methoden in het management in plaats van voor het management.

Maar al te vaak zijn in het verleden prestigieuze modellen en technieken ontwikkeld die weliswaar uit wetenschappelijk oogpunt interessant waren maar voor de praktijk bitter weinig relevantie hadden.

Dit boek bevat ervaringen van praktijkmensen over het gebruik van kwantitatieve methoden in het management. Het gaat vooral om de vervlechting van modellen in praktijksituaties. Deze laatste betreffen zowel profit- als non-profit organisaties.

Ook negatieve ervaringen worden gerapporteerd. Immers ook uit mislukte toepassingen van Operationele Research modellen kan veel lering worden getrokken.

Om een evenwichtig presentatie van de toepassingen te verkrijgen zijn de volgende sectoren gekozen: landbouw, industrie, transport, bankwezen, adviesbureaus, overheidsinstellingen en universiteiten. In elke bijdrage worden vragen betreffende aanpak en afloop van het project aan de orde gesteld.

Reeds enige tijd verheugen de kwantitatieve methoden zich in een sterk groeiende belangstelling. Er ontstaat een generatie managers die intensief tijdens de opleiding heeft kennis gemaakt met de kwantitatieve methoden en informatica.

De laatste discipline heeft vooral de condities gecreëerd onder welke de kwantitatieve methoden zo'n hoge vlucht konden nemen.

In deze ontwikkeling past het verschijnen van dit boek goed. Het geeft managers een uitstekend beeld van de mogelijkheden (en ook de beperkingen) van de kwantitatieve methoden. Ook in het onderwijs (zowel universitair als HBO) kan het boek een rol spelen.

Samengevat kan worden gesteld dat het boek een juist beeld geeft van de mogelijkheden van de Operationele Research in de praktijk en zeer goed bruikbaar is voor grote groepen studenten (bedrijfskunde, wiskunde, economie, technische wetenschappen e.d.) en voor managers die met de kwantitatieve methoden kennis willen maken.

P. van Beek, LHW

J. Oosterhaven en B.B.A. Drewes (redactie)

Constructie en actualisering van regionale en interregionale input-output tabellen

Gezamenlijke Economisch Technologische Instituten, Arnhem, 1985, 232p, f 45,--.

Deze publicatie bevat de inleidingen en discussies van een tweedaagse workshop over regionale input-output tabellen, die gehouden werd in oktober 1983 onder auspiciën van de gezamenlijke Economisch Technologische Instituten. Een belangrijke reden voor het organiseren van de workshop was de op handen zijnde CBS-publicatie van regionale saldo-tabellen betreffende het jaar 1975. Omwerking van deze tabellen tot (semi-)integrale tabellen en/of het actualiseren ervan zou daarna ruime aandacht krijgen.

Door het zorgvuldig op elkaar afstemmen van de onderwerpen en door het uniformeren van de gebruikte terminologie en notatie,

waartoe de redacteuren voor, tijdens en na de workshop intensief overleg voerden met alle auteurs en voorzitters van de sessies, is een waardevol boek ontstaan dat de volgende bijdragen bevat:

1. Algemene inleiding: regionale input-output analyse op een kruispunt
(J. Oosterhaven)

Deel I Constructie van half geregionaliseerde tabellen.

2. Constructie van regionale input-output tabellen volgens de saldo-methode door het CBS (H.M.T. Lunter)
3. Constructie van regionale afzettabellen (E.C. van der Knijff)

Deel II Constructie van volledig geregionaliseerde tabellen

4. Discrete constructie van interregionale input-output tabellen (E.C. van der Knijff)
5. Discrete constructie van een stelsel van regionale input-output tabellen (H.C.J. Koning, H.E.C.M. Thomaes en J.C.H. Strankinga)
6. Discrete constructie van een regionale tabel zonder CBS-input-output gegevens (H.E.C.M. Thomaes en J.C.H. Strankinga)

Deel III Directe methoden versus indirecte methoden ter constructie van regionale en interregionale tabellen

7. Overzicht van indirecte methoden en een vergelijking met de directe methode (J. Spijker, M. Walsweer en B.B.A. Drewes)
8. Determinanten van de interregionale handel (J.M. Meijering en J.A.M. Wesseling)
9. Het gebruik van de graviteitsbenadering als gemengde methode
(J. Oosterhaven)
10. Een combinatie van de directe en de indirecte methoden
(J. Spijker)

Deel IV Actualisering van regionale en interregionale tabellen

11. Actualisering van saldo-tabellen (J.A.M. Wesseling en J.M. Meijering)
12. Actualisering van de input-output tabel voor het Noorden
(J. ten Welle)
13. Het actualiseren van interregionale tabellen (J. Oosterhaven en E.C. van der Knijff)

Discussieverslagen per hoofdstuk

De voorzitters van de workshop-zittingen waren in volgorde van de delen

W. Althuyzen, H.P.H. Brouwer, F. Muller en J.H.P. Paelinck.

De redacteuren zijn erin geslaagd het gestelde doel, stimulans voor en coördinatie van regionaal economisch onderzoek, te bereiken. De bundel levert een nuttige bijdrage voor theorie en praktijk. Dit blijkt bijvoorbeeld uit zeer recente publicaties over regionaal onderzoek:

- De macro-economische betekenis van de Olympische Spelen voor Amsterdam (NEI, Rotterdam, 1985, 68p)
- De economische betekenis van de professionele kunsten in Amsterdam (SEO, Amsterdam, 1985, 246p)

Doordat de auteurs hun bijdragen aan de bundel vanuit verschillende achtergrond geleverd hebben, is een interessant totaal beeld van het maken en gebruiken van regionale input-output tabellen ontstaan. Het boek is dan ook aan te raden voor allen die zich met regionaal economisch onderzoek bezighouden.

J.A.M. Wesseling, UvA

S.W. Bergman & J.W. Gittins

Statistical methods for pharmaceutical research planning

Marcel Dekker inc., New York, 1985, 257 p., ISBN 0-8247-7146-X, f 212,--.

Dit boek richt zich hoofdzakelijk op de statistiek rond de eerste fase van pharmaceutisch onderzoek, gekarakteriseerd door een groot aantal potentieel interessante stoffen, waaruit enkele geselecteerd moeten worden voor verder, intensief onderzoek. Zoals de titel al suggereert wordt tevens ingegaan op methoden die beslissingen verband houdende met dit onderzoek, kunnen ondersteunen.

Het boek kent 4 hoofdstukken

1. Quantitative structure-activity relationships (QSAR) (104 pag);
2. Screening procedures for discovering active compounds (36 pag);
3. Allocation of resources between related screens (26 pag);
4. R & D project selection methods (80 pag).

Hoofdstuk 1 behandelt de technieken die van dienst kunnen zijn bij het relateren van de structuur van verbindingen aan de chemische eigenschappen om op die wijze verbeteringen te kunnen aanbrengen. Aan de orde komen:

- lineaire regressie (schatten, toetsen, goodness of fit, selectie van variabelen, multicollineariteit);
- logistische regressie;
- classificatie (clusteranalyse & discriminanten-analyse, met enkele varianten).

Hoofdstuk 2 is gewijd aan de problemen die optreden, als men uit een groot aantal verbindingen de actieve wil selecteren. Centraal staan de OC-curve - Operation Characteristic curve -, welke bij iedere "echte" activiteit θ de kans $\rho(\theta)$ weergeeft om een bepaald screeningsschema te passeren en het "average sample number" $\eta(\theta)$, dat het verwachte aantal observaties benodigd om een beslissing te nemen, aangeeft. Op grond van verschillende

criteria (misclassificatie, opbrengst onder bepaalde randvoorwaarden, netto-opbrengst) worden diverse schema's, al of niet "multi-stage", bekeken en de beste gezocht.

Hoofdstuk 3 bouwt verder op de ideeën van hoofdstuk 2. Nu wordt echter ingegaan op allocatie over verschillende screens (bijv. dierproeven naast klinisch onderzoek) en op de volgorde van (onderdelen van) screens (bijv. werkzaamheid-screens naast toxiciteit-screens).

Hoofdstuk 4 gaat in op de hulp die statistiek en meer, operationele analyse kunnen bieden bij beleidsbeslissingen op directie-niveau. Het traject van simpele "check-lists", "scoring-index" en "profile charts" via "profitability indices" tot "integer, lineair, stochastisch en dynamic programming" wordt besproken. Ook een decision support system wordt beschreven (zonder gebruik van die term overigens). Dit alles op een elementair niveau.

De opbouw van het boek is prettig. Ieder onderwerp wordt eerst in een inleidende uiteenzetting oppervlakkig behandeld, waarna voor de statistisch beter geschoolden een verdergaande uitwerking volgt. De lay-out is goed verzorgd, het aantal drukfouten is klein. Soms is het lastig dat stellingen, bewijzen, afleidingen en voorbeelden niet duidelijk als zodanig herkenbaar zijn, maar gewoon in de tekst zijn opgenomen.

De stof wordt in het algemeen zonder veel diepgang behandeld. Veel resultaten uit andere literatuur worden aangehaald. Dit heeft als voordeel het verschaffen van een grote hoeveelheid informatie, maar als nadeel een soms erg grote gedachtendichtheid en een wat "springerig" karakter. De doelgroep wordt omschreven als "all scientists engaged in the speculative end of new-product chemical research, particularly those with a responsibility for any kind of planning, together with the statisticians and operational researchers who advise them". Dit lijkt me een goede omschrijving van de doelgroep, alhoewel van de eerstgenoemden een redelijke kennis van kansrekening en statistiek (combinatoriek, verdelingen) en wiskunde (matrix-rekening) verwacht wordt.

Voorbeelden en toepassingen zijn schaars, zeker praktijkvoorbeelden. Computertoepassingen komen nauwelijks voor (van de standaardprogrammatuur bijvoorbeeld wordt alleen BMD (zonder P) genoemd).

Van dit boek zijn de hoofdstukken 2 en 3 het lezen zeker waard. Hoofdstuk 1 kan de lezer de indruk geven van de bij het farmaceutisch onderzoek gebruikte technieken en van de probleemstelling die men daarbij ontmoet. De behandeling van de technieken is echter beperkt in omvang en diepgang. Over hoofdstuk 4 kan hetzelfde gezegd worden. Gelet op de inhoud van dit hoofdstuk kan men zich afvragen of de titel van het boek niet te eng is.

P.H. van Ewijk, Duphar BV

K.W. Kemp

Dice, data and decisions, introductory statistics

Ellis Horwood Ltd., Chichester, 312p, ISBN 0-85312-671-2, \$12.--

Het boek is volgens de auteur geschreven voor twee doelgroepen. De eerste groep is die van studenten in wiskunde of andere wetenschappen, die enig inzicht in de principes van de basisgedachten van de statistische methoden nodig hebben. De tweede groep is die van mensen in het onderwijs, die in de afgelopen jaren ertoe moesten overgaan om ook statistiek te onderwijzen, maar daarvoor zelf niet de benodigde (formeel) instructie hebben ontvangen.

Volgens mij is het de schrijver goed gelukt om voor deze groepen een waardevol boek te produceren. In het boek behandeld de auteur veel onderwerpen op formele, wiskundige wijze. Een breed spectrum van wiskundig statistische onderwerpen passeren de revue.

Het boek begint met een hoofdstuk "statistics and society", waarin de schrijver aangeeft dat omgaan met onzekerheid en natuurlijke variatie in feite dagelijks gebruik is en ook door bijna iedereen (gevoelsmatig) ingeschat kan worden. Het tweede hoofdstuk bouwt daarop voort, doch daarin worden meer de mogelijkheden aangedragen om met onzekerheid en natuurlijke variatie om te gaan vanuit het (wiskundig) statistisch denkraam.

Deze techniek gebruikt de schrijver door het gehele boek, de theorie wordt niet slechts geponereerd, maar steeds wordt er aan de hand van inleidende stukjes als het ware een behoefte gecreëerd om ideeën formeel te regelen.

Het derde hoofdstuk behandeld inleidende theorie van kansrekening en in het vierde hoofdstuk worden dichtheidsfuncties, verwachting, variantie en covariante ingevoerd.

In het vijfde hoofdstuk worden een paar kansverdelingen behandeld (binomiale, multinomiale, negatief binomiale, hypergeometrische, poisson), meestal aan de hand van of naar aanleiding van voorbeelden, zodat de noodzaak voor het bestaan van dergelijke verdelingen wordt aangegeven.

Hoofdstuk 6 gaat in zijn geheel over de normale verdeling, het zevende hoofdstuk over sommen van stochastische variabelen (convolutie, kans- en momentgenererende functies, centrale limietstelling).

Hoofdstuk 8 geeft toepassingen van de normale verdeling met onder andere een paragraaf over toepassen van eigenschappen van normaal verdeelde grootheden in de industriële kwaliteitsbeheersing (slechts 3 pagina's) en een stuk over transformeren van variabelen, zodat zij beter middels een normale verdeling beschreven kunnen worden.

Hoofdstuk 9 handelt over de theorie van het toetsen van hypothesen, waarop in hoofdstuk 10 wordt doorgedaan met onderwerpen als "most powerful tests" en het lemma van Neyman en Pearson.

Hoofdstuk 11 legt de link tussen populatie parameters en de schattingen daarvan op basis van een steekproef (zuiver en

onzuivere schatters, ongelijkheid van Cramer-Rao), hoofdstuk 12 bouwt hier weer op door met behandeling van "sufficient estimators", maximum likelihood en Baysiaanse methoden.

Het boek is inderdaad geschreven vanuit de formele behandeling van de basisbegrippen, er staan heel veel formules in het boek en weinig getallen.

Geen behandeling van onderwerpen als regressie-analyse, variantie-analyse, toepassen van toetsen (de goodness of fit toets wordt wel in een voorbeeld gebruikt, maar nergens behandeld), maar dat lag denk ik ook niet in de bedoeling van de auteur.

Ik denk dat de schrijver geslaagd is in zijn opzet een breed gebied van de basisbegrippen van de statistiek formeel te behandelen. Een boek wat goed gebruikt kan worden door de twee beschreven doelgroepen en een boek waarnaar de praktiserend statisticus ook nog wel kan teruggrijpen op zoek naar de formele fundamenten van de statistiek.

Een klein minpuntje vind ik de slordige presentatie van de tekeningen, die niet altijd even minitieuus zijn uitgevoerd en dat is zonde, want het boek ziet er verder zeer goed verzorgd uit.

A.D. Oosterhoorn, Van Doorne's
Transmissie BV

F.R. Joliffe

Survey Design and Analysis

Ellis Horwood Ltd., Chichester, 1986, 178 p., ISBN 0-85312-599-6 \$28.50.

Dit boek is bedoeld voor hen die zich in toenemende mate met steekproefonderzoek bezig houden en met de analyse van de vergaarde gegevens. Bij dit laatste wordt er vanuit gegaan dat de onderzoeker wel een of ander statistische pakket tot zijn beschikking heeft. Het boek kan ruwweg in twee delen worden opgesplitst. Het eerste deel (hoofdstukken 1 tot en met 5) beschrijft een aantal aspecten van opzet, uitvoering en verwerking van steekproefonderzoek. Het tweede deel (hoofdstukken 6 en 7) richt zich op de analyse van gegevens.

Hoofdstuk 1 geeft een globaal overzicht van de gang van zaken bij opzet en uitvoering van steekproefonderzoek. In vogelvlucht worden zaken behandeld als planning, steekproeftrekking, steekproefkaders, soorten variabelen, het opstellen van vragenlijsten en allerlei soorten fouten die op kunnen treden. Het is jammer dat bij de bespreking van de kwaliteit van steekproefkaders zaken als onderdekking (under-coverage) en overdekking (over-coverage) niet duidelijk uit de verf komen. Ook het gedeelte over het opstellen van vragenlijsten is wat summier. Er zou veel meer kunnen worden gezegd over problemen die zich kunnen voordoen bij

het formuleren van vragen.

Hoofdstuk 2 geeft een overzicht van een aantal manieren van het trekken van steekproeven. Het verschil tussen trekken met en zonder teruglegging komt niet uit de verf. De enkelvoudige aselechte steekproef wordt zeer uitgebreid behandeld. Daarnaast komen aan de orde gestratificeerde steekproeven, clustersteekproeven en systematische steekproeven. Bij de gestratificeerde steekproef blijkt pas in een zeer laat stadium waarom deze wijze van steekproeven zo aantrekkelijk kan zijn: de schatters hebben vaak kleinere variantie. Bij de schattingsprocedure wordt wel aandacht besteed aan de quotientschatter (ratioschatter), maar de regressieschatter (die nooit slechter is dan de quotientschatter) wordt in het geheel niet genoemd. Terecht wordt de nodige aandacht besteed aan stratificatie achteraf (post-stratificatie). Deze schattingstechniek komt er in de algemene werken over steekproeftheorie vaak bekaaid af. Bij stratificatie achteraf wordt wel ingegaan op het verschil tussen conditionele en onconditionele variantie, maar de formule voor de onconditionele variantie wordt niet gegeven.

Hoofdstuk 3 behandelt een aantal wat meer praktische aspecten van steekproefonderzoek. Het gaat met name om niet-steekproeffouten, onzekerheden in de uitkomsten, die niet het gevolg zijn van de steekproeftrekking. Een van de belangrijkste bronnen van fouten is non-respons. Hieraan wordt de nodige aandacht besteed, maar er wordt weinig gezegd over technieken om non-respons en de kwalijke gevolgen daarvan tegen te gaan. Randomized response wordt behandeld als een techniek om gevoelige vragen te stellen. Andere belangrijke problemen die aan de orde komen, zijn herinneringseffecten, interviewereffecten en verwerkingsfouten.

Hoofdstuk 4 gaat in op de verdeling van schatters. Het design-effect wordt behandeld als een middel om de kwaliteit van verschillende steekproefopzetten met elkaar te vergelijken. Ook de doorwerking van verschillende typen steekproeven en effecten als non-respons op betrouwbaarheidsintervallen komt aan de orde. Tevens wordt in dit hoofdstuk uitgelegd hoe de steekproefomvang moet worden bepaald. Met name ten aanzien van de enkelvoudige aselechte steekproef was de stof wat inzichtelijker geweest als dit al was behandeld in hoofdstuk 2. Voor het schatten van varianties in complexe steekproeven worden de methode van de herhaalde steekproeven en jack-knife schatters behandeld. Aan het einde van het hoofdstuk wordt heel kort (te kort om interessant te kunnen zijn) ingegaan op de twee verschillende benaderingswijzen van de steekproeftheorie: de 'design-based' benadering waarin de stochastiek afkomstig is van het lotingsmechanisme van de steekproef en 'model-based' benadering waarin de stochastiek afkomstig is van een verondersteld superpopulatiemodel.

Hoofdstuk 5 gaat over de verwerking van de verzamelde gegevens. Terecht wordt hieraan de nodige aandacht besteed, want een goede verwerking kan bijdragen tot een aanzienlijke kwaliteitsverbetering van de resultaten. Aspecten die aan de orde komen zijn controle en correctie, imputatie-technieken en coderen van open en gesloten vragen. Als belangrijke nieuwe ontwikkeling wordt in dit hoofdstuk opgevoerd CATI (computer assisted telephone interviewing). Deze technieken bestaan al reeds enige tijd. Er

zou inderdaad van nieuwe ontwikkeling in dit recente boek sprake zijn als bijvoorbeeld aandacht werd besteed aan CAPI (computer

assisted personal interviewing).

Hoofdstuk 6 is het eerste hoofdstuk over analyse. Dit hoofdstuk gaat met name over de analyse van tabellen. De schrijver hecht een groot belang aan een goede representatie van tabellen. Hij geeft echter geen richtlijnen of voorbeelden. Op globale wijze wordt het bekende rijtje associatiematen voor tabellen behandeld en verder komen loglineaire modellen aan de orde.

Hoofdstuk 7 betreft de analyse van kwantitatieve gegevens. In een algemene beschrijving zonder diepgang komen achtereenvolgens aan de orde regressie-analyse, principale componentenanalyse, clusteranalyse en factoranalyse. Welgeteld een pagina is ingeruimd voor grafische technieken (overigens zonder een plaatje). Met uitzondering van een enkele alinea bij regressieschatter, wordt in het geheel niet duidelijk hoe al deze technieken moeten worden toegepast op complexe steekproeven.

Samenvattend dan worden gesteld dat het boek een overzicht geeft van de theoretische en praktische aspecten van steekproefonderzoek en van de analyse van de daarmee vergaarde gegevens. Met name ten aanzien van de analysetechnieken is het boek zeer oppervlakkig. Het boek beoogt 'self-contained' te zijn. Voor het kiezen van een analysetechniek, het uitvoeren ervan en de interpretatie van de uitkomsten is dit zeker niet waar.

Duidelijk komt de grote kloof tot uiting die bestaat tussen enerzijds de steekproeftheorie en anderzijds de multivariate analyse. Deze analysetechnieken kunnen alleen worden uitgevoerd op eenvoudige en nette enkelvoudige aselecte steekproeven. Zodra de echte steekproeftheorie om de hoek komt kijken met complexe steekproeftrekkings- en schattingsprocedures en problemen als non-respons, moet de multivariate analyse afhaken. Als gevolg daarvan is het boek niet een geheel, maar valt uiteen in twee verschillende stukken die weinig met elkaar te maken lijken te hebben.

Het boek biedt een aardige inleiding voor hen die nog niet eerder in aanraking zijn geweest met de mogelijkheden en problemen van steekproefonderzoek. Een onderzoeker die dit werk wil gebruiken voor een kennismaking met en een beter begrip van multivariate analyse kan beter verwezen worden naar een van de vele inleidende andere boeken op dit gebied.

J.G. Bethlehem, CBS

Mr. Osborne

Finite algorithms in optimization and data analysis

John Wiley & Sons, Chichester, 1985, ISBN 0-4719-0539-9, \$32,--

Volgens de flaptekst is het boek van belang voor studenten en onderzoekers op het terrein van operationeel onderzoek, wiskunde, econometrie, numerieke analyse en statistiek, toe maar. Waarop apprecieerde ik het boek? Niet omdat ik mij op een van bovengenoemde gebieden exclusief thuisvoel, maar juist omdat ik als een omnivoer van alle bordjes de lekkere hapjes waardeer. Een

pur sang statisticus of data-analist zal niet tevreden zijn met het boek. Stochastiek wordt weinig en soms onjuist beschreven. Een pure OR-man zal over grote delen van het boek - met Osborne verzuchten: "The basic reference is Rockafellar and there is really very little more to be said".

Hoofdstuk 1 vat enkele resultaten uit de convexe analyse samen. Van sommige stellingen wordt een bewijs gegeven, andere worden zonder bewijs vermeld. Voor de lezer is niet altijd duidelijk welk beleid de schrijver hierbij heeft gekozen. Zo vermeldt hij de stelling van Klee zonder bewijs en bewijst wel de volgende stelling die Rockafellar, zonder het overbodige tweede "closed", als correlaat vermeldt.

Theorem 2.2. A closed bounded convex set S in R^p is the closed convex hull of its extreme points.

De algorithmes - ook voor statistische problemen - zijn sterk geënt op de convexiteitstheorie van hoofdstuk 1. In dit verband had ik graag gezien dat bij de uitgebreidere samenvatting van convexiteitsstellingen een voor de statistiek - asymptotiek-zeer nuttige stelling was vermeld. Dat is stelling 25.7 van Rockafellar over de uniforme convergentie van rijen afgeleiden van convexe functies :

Theorem 25.7

Let C be an open convex set, and let f be a convex function which is finite and differentiable on C . Let f_1, f_2, \dots , be a sequence of convex functions finite and differentiable on C such that $\lim_{i \rightarrow \infty} f_i(x) = f(x)$ for every $x \in C$.

Then $\lim_{i \rightarrow \infty} \nabla f_i(x) = \nabla f(x)$, for all $x \in C$.

In fact the mappings ∇f_i converge to ∇f uniformly on every closed bounded subset of C .

In hoofdstuk 5 behandelt Osborne "least squares and related methods":

5.1 Introduction

5.2 Least squares method

5.3 Least squares subject to linear constraints

5.4 Iteratively reweighted least squares

5.5 Sensitivity of estimation procedures

5.6 Algorithms for M-estimation

en zoals ieder hoofdstuk eindigt hoofdstuk 5 met een interessante literatuurbeschrijving. Veel meer kennis van statistiek dan het begrip verwachting wordt van de lezer niet vereist, of het zou moeten zijn om vraagtekens te zetten bij de volgende omschrijving op pagina 216:

"To determine T we seek the unbiased estimator for which the variance of $\|\bar{x} - \bar{x}\|_2$ is minimized (the norm being the Euclidean norm)".

Hierbij gaat Osborne uit van het lineaire model

$$f = M\bar{x} + \epsilon,$$

$$E\epsilon^T = 0, E\epsilon\epsilon^T = J \gg 0, M : R^p \rightarrow R^n, \text{ met rang } p < n.$$

De BLUE-schatter voor \bar{x} is de GLS-schatter $\hat{x} = (M^T J^{-1} M)^{-1} M^T J^{-1} f$ en \hat{x} minimaliseert de variantie over de klasse van lineaire zuivere schatters Tf . Nee, de statistiek is er bij gehaald om

leuke algorithmes te illustreren.

Het boek bestrijkt een breed scala aan problemen en is derhalve een interessant startpunt voor b.v. een werkcollege. Op terreinen van operationeel onderzoek, wiskunde, econometrie, numerieke analyse en statistiek zal de docent wel een verdieping moeten aanbrenge.

A.E. Ronner, Nederlandse
Philips Bedrijven BV.