

BOEKBESPREKINGEN

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

H.A. David

Order Statistics, second edition

John Wiley & Sons, New York, (1981), 360 p, ISBN 0-471-02723-5,
19.90 pound.

Dit boek is de tweede aanzienlijk uitgebreidere editie van de bekende monografie van H.A. David over order statistics. Het boek houdt het midden tussen een niet al te diepgravend leerboek en een naslagwerk met veel verwijzingen naar de omvangrijke literatuur. De eerste editie van het boek, welke in 1970 verscheen, gaf een goed leesbaar overzicht van een aantal uiteenlopende onderwerpen op het terrein van de statistiek, waarbij order statistics van belang zijn. Min of meer uitvoerig kwamen onder meer aan de orde : het efficiënt schatten van locatie en schaalparameter van een verdeling m.b.v. lineaire combinaties van order statistics, de behandeling van uitbijters en grafische schattingsmethoden bij het gebruik van waarschijnlijkheidspapier. Verder werden een aantal van oudsher bekende, niet-efficiënte maar wel - ook zonder een computer - bijzonder eenvoudig uit te voeren statistische procedures gebaseerd op een of slechts enkele order statistics, zoals bijvoorbeeld de steekproefmediaan en de steekproefwijdte, behandeld. Ook vinden we in de eerste editie van David's boek enkele meer probabilistisch getinte hoofdstukken, onder meer over de exacte verdelingstheorie voor functies van order statistics, en over ongelijkheden en benaderingen voor momenten van order statistics.

In de tweede editie van het boek van H.A. David wordt, zonder de opzet van het boek wezenlijk te wijzigen, naast de hiervoor genoemde onderwerpen ook aandacht besteed aan enkele meer recente ontwikkelingen op het terrein van de statistiek voor order statistics. De sectie over robuuste statistiek is uitgebreid, secties over order statistics, gebaseerd op onafhankelijke, maar niet identiek verdeelde, stochastische grootheden en over toetsen voor uitbijters in een lineair regressie model zijn toegevoegd, terwijl de sectie over de asymptotische theorie voor lineaire combinaties van order statistics is verrijkt met een aantal resultaten van meer recente datum. Het aantal verwijzingen naar de snel groeiende literatuur is gestegen van omstreeks 700 in de eerste editie tot een duizendtal in de tweede editie.

Evenals de eerste editie uit 1970 is zonder meer duidelijk dat ook de tweede editie van H.A. David's monografie een goed bruikbare introductie is tot de literatuur op het terrein van statistiek voor order statistics.

R. Helmers, CWI

D.C. Hoaglin, F. Mosteller en J.W. Tukey (editors)

Exploring data tables, trends and shapes

John Wiley & Sons, New York, 1985, 527 p, ISBN 0-471-09776-4, 40.85 pound.

Enige jaren geleden verschenen drie boeken op de markt die inleidingen bevatten omtrent robuuste en exploratieve data analyse.

De boeken waren geschreven door Tukey (1977), Mosteller en Tukey (1977) en Velleman en Hoaglin (1981). De boeken hadden tot doel de varia van technieken op het gebied van de robuuste en exploratieve data analyse te presenteren en te illustreren aan de hand van voorbeelden. Omtrent het hoe en waarom bleef de lezer veelal in het ongewisse.

In 1983 publiceerde Hoaglin, Mosteller en Tukey een boek waarin een gedeelte van de technieken op een meer wiskundige wijze werd behandeld. Met name werden de stem-and-leaf displays, boxplots, resistant lines, median polish method en verscheidene locatie schatten uitvoerig besproken. De auteurs kondigden in de inleiding een tweede boek aan waarin de overige technieken aan bod komen.

Inmiddels is het tweede boek verschenen. Evenals het eerste deel heeft een groot aantal auteurs bijdragen geleverd. De belangrijkste onderwerpen die beschreven worden zijn :

Twee en drie weg analyses, robuuste (multiple) regressie en een aantal grafische methoden voor het bestuderen van de vorm van een verdeling. De theorie wordt geïllustreerd met voorbeelden. Zowel voor het onderwijs als voor de praktijk is aanschaf van het boek aan te bevelen.

Referenties

- Hoaglin, D.C., F. Mosteller & J.W. Tukey (1983), Understanding robust and exploratory data analysis, Wiley, New York.
- Mosteller, F. & J.W. Tukey (1977), Data analysis and regression, Addison Wesley, Reading.
- Tukey, J.W. (1977), Exploratory data analysis, Addison Wesley, Reading.
- Velleman, P.F. & D.C. Hoaglin (1981), Applications, basics and computing of exploratory data analysis, Duxbury, Boston.

R.J.M.M. Does, RUL

J.G. Odink

Inkomensherverdeling, Enkele aspecten van de inkomensherverdeling door de overheid in Nederland.

Wolters-Noordhoff, Groningen, 1985, 254 p, ISBN 90-01-661017, f. 39,75.

Reeds geruime tijd houdt J.G. Odink zich, aan de Universiteit van Amsterdam, bezig met het bestuderen van de inkomensverdeling in Nederland. Zijn eerste artikel op dit terrein dateert uit 1977. De lange reeks publicaties die sindsdien is gevolgd, heeft hij, bijgewerkt en voorzien van inleiding en samenvatting, gebundeld tot een boek, dat eerder verscheen als proefschrift.

Het boek bevat naast deze inleiding en probleemstelling (hoofdstuk 1) en samenvatting met conclusies (hoofdstuk 7) vijf hoofdstukken waarin een aantal aspecten van het Nederlandse beleid van inkomensherverdeling door middel van belastingen, premies en uitkeringen worden beschreven.

Hoofdstuk 2, over het meten van de inkomensongelijkheid, begint met een bespreking van de meetproblemen die ontstaan wanneer men de inkomensongelijkheid tussen n personen wil samenvatten in één scalaire grootheid. Alhoewel er een groot aantal maatstaven ter beschikking staan, kan men toch op basis van enkele algemeen aanvaarde eigenschappen (bijv. dimensieloosheid, symmetrie etc.) het aantal reduceren. Op grond van zo'n axiomatische aanpak kiest Odink voor de in Nederland vaak gebruikte Theil-coëfficiënt. In dit hoofdstuk gaat hij dan ook dieper in op de meetproblemen verbonden aan het hanteren van de Theil-coëfficiënt.

In 1976 verscheen in de Economisch Statistische Berichten het intussen beroemde artikel van J. Pen en J. Tinbergen over de vermindering van de inkomensongelijkheid in de tijd. Tussen 1983 en 1976 zou volgens hen de inkomensongelijkheid in Nederland met 50 % zijn afgenomen. In hoofdstuk 3, over inkomensongelijkheid in Nederland in historisch perspectief, wordt deze berekening geactualiseerd en bevestigd. Hierbij onderstreept Odink dat de nivelleringstendens in de periode tussen 1966 en 1976 zeer duidelijk is versneld.

Hoofdstuk 4 gaat over de invloed van het schijventarief van de Nederlandse inkomstenbelasting en van de sociale verzekeringen op het bruto-netto traject. Het gecombineerd effect van premies en belastingen in Nederland is duidelijk nivellerend en veroorzaakt dan ook bij een inkomensgroei een afname van de inkomensverschillen. In hetzelfde hoofdstuk wordt ook het tekortschieten van de inflatiecorrectie van het belastingstelsel besproken en worden de bruto-netto trajecten van de ambtenaren en de werknemers van de particuliere sector vergeleken. Hier blijkt reeds dat hogere ambtenaren relatief slechter en lagere ambtenareninkomens relatief beter uit de bus komen dan vergelijkbare inkomens in de particuliere sector.

Waar in hoofdstuk 4 de herverdelende invloed van de overheid op de beloningsverschillen tussen actieven onderling aan de orde was, wordt in hoofdstuk 5 de inkomensverhouding tussen actieven en niet-actieven ontleed. Meer in het bijzonder komt de relatieve inkomenspositie van de AOW-ers ter sprake. Hun reëel netto

inkomen blijkt sinds 1957 in een veel hoger tempo te zijn gestegen dan het netto reëel gemiddeld loon van nijverheidsarbeiders. Ook het rendement van studeren bij verschillende vormen van studietoelagen wordt in dit hoofdstuk behandeld.

In hoofdstuk 6 tenslotte wordt gepoogd om het totale herverdelingseffect van overheidsoverdrachten te meten. Geschat wordt, op basis van de Theil-coëfficiënt, dat ongebonden overdrachten (voornamelijk sociale zekerheid en belasting) de ongelijkheid met ca 3/4 doen afnemen. Deze herverdelende werking van de ongebonden overdrachten wordt nog in aanzienlijke mate versterkt indien rekening wordt gehouden met gebonden overdrachten zoals bijv. subsidies op openbaar vervoer, huursubsidies etc.

Het boek is geen monografie waarin een compleet overzicht van alpha tot omega van de inkomensherverdeling wordt gegeven. Het is de weergave van de onderzoekinteressen van de auteur. Desondanks zit er een duidelijke lijn in het betoog en is het een goed startpunt voor kennismaking met de fundamentele problemen en resultaten van het meten van herverdeling in de praktijk.

J. Theeuwes, VU

Frank R. Hampel, Elvezio M. Ronchetti, Peter J. Rousseeuw & Werner A. Stahel

Robust Statistics, The Approach Based on Influence Functions.

John Wiley & Sons, New York, 1986, 502 p, ISBN 0-471-82921-8, 40.90 pound.

Door groeiende alomtegenwoordigheid van de computer is het aan de meeste algemeen bekende statistische technieken verbonden rekenwerk zelden meer een belemmering om ze toe te passen. Dank zij "vriendelijke" programmapakketten, talen als BASIC en veel stimulerende literatuur is de drempel zelfs zeer laag geworden. De winst van deze ontwikkeling is dat meer mensen dan ooit met statistische toepassingen vertrouwd raken. Een verlies is echter dat vaak blindelings een recept gevolgd wordt en het gegevensmateriaal te weinig kritisch beschouwd wordt. Doordat veel parametrische technieken erg gevoelig zijn voor uitbijters of anderszins invloedrijke waarnemingen, kunnen uitkomsten ontstaan die een onvolledige of misleidende weergave zijn van het gegevensmateriaal.

In het afgelopen decennium heeft dit probleem, dat men eerder te lijf ging met non-parametrische technieken, veel aandacht gekregen onder de noemer van "robuste en exploratieve statistische analyse". De recente door Hoaglin, Mosteller en Tukey geredigeerde bundels geven een goed overzicht van deze technieken.

Het boek van Hampel e.a. kiest een andere invalshoek, die meer op de confirmatieve analyse gericht is. Een opsomming van de titels van de hoofdstukken duidt dit al aan :

1. Introduction and Motivation
2. One-dimensional Estimators
3. One-dimensional Tests
4. Multi-dimensional Estimators
5. Estimation of Covariance Matrices and Multivariate Location
6. Linear Models : Robust Estimation
7. Linear Models : Robust Testing
8. Complements and Outlook

Het grootste deel van het boek is gewijd aan het systematisch beschrijven van diverse schattingsmethoden die de invloed van extreme waarnemingen reduceren of zelfs elimineren. De invloeds-functie, die grof gezegd weergeeft welke invloed een extra waarneming of een verandering in een aanwezige waarneming heeft op de waarde van een geschatte parameter, blijkt daarbij een krachtig gereedschap. De theoretische eigenschappen van invloeds-functies en aanverwante functies worden uitgebreid onderzocht. In tegenstelling tot wat de auteurs in hun voorwoord stellen, vraagt de theoretische behandeling gedegen mathematische en statistische kennis van de lezer. De gebruikte stijl doet aangenaam aan : de mathematische strengheid wordt gelardeerd met ongedwongen taalgebruik. De eerste zeven hoofdstukken worden elk afgesloten met een groot aantal oefeningen en opgaven van sterk wisselende zwaarte. Antwoorden en uitwerkingen zijn niet opgenomen. Het aantal numerieke voorbeelden is gering.

In de hoofdstukken over robuuste toetsen worden eigenschappen van bestaande parametrische en non-parametrische toetsen onderzocht. Ook wordt een robuuste versie van de F-test ontwikkeld die blijkens een voorbeeld erg effectief is . Het is waarschijnlijk voor het eerst dat robuuste toetsen zo uitgebreid in boekvorm behandeld worden.

Het eerste en het laatste hoofdstuk steken bij de andere af, doordat zij juist zeer weinig theoretische vaardigheid vragen. De inleiding beschrijft op zeer lezenswaardige wijze de noodzaak van robuuste methoden. Daar vindt men ook een schets van hun nog korte geschiedenis en een overzicht van de doelen en werkwijzen van robuuste technieken. Het slothoofdstuk heeft een wat wonderlijke samenstelling. Daar vindt men een beschouwing over tijdreeksen met "olifantengeheugens", waarvan de seriële correlatie veel minder snel naar nul tenderen dan men met autoregressieve modellen kan beschrijven. Daar wordt verwezen naar de door Mandelbrot uitvoerig beschreven "fractals". Dit zijn zeer lezenswaardige bladzijden, als men bedenkt hoe gebrekkig natuurlijke processen in lucht en water zich vaak door ARMA-modellen laten beschrijven. Het is echter de vraag of veel tijdreeksanalisten dit boek zullen raadplegen.

Opmerkelijk is ook de tirade tegen denkbeldige opponenten van robuuste statistiek. Aan de hand van een tiental tegenwerpingen wordt de goede zaak nog eens extra beargumenteerd. Dit is boeiende stof. Het valt alleen op dat de tegenwerpingen kennelijk nergens op schrift staan, want referenties ontbreken. Zo lijkt het op een spiegelgevecht. Traagheid in het accepteren van robuuste methoden zal meestal het gevolg zijn van praktische

problemen, die deels neerkomen op het ontbreken van geschikte software.

De auteurs besteden enkele bladzijden aan de beschikbaarheid van programma's. Die blijkt niet overdadig te zijn. In dit verband valt het extra op dat zo weinig aandacht besteed wordt aan L_1 -schatters, die de som van de absolute waarden van residuen minimaal maken. Daarvoor zijn vrij veel algoritmen gepubliceerd, deels met programma. Convergentie is vrijwel altijd verzekerd en nevenvoorwaarden zoals niet-negatieve parameters zijn goed te verwerken. Zie bijvoorbeeld de artikelen en vele referenties in Zanakis en Rustagi. Ook lijkt het idee van "regressie-quantielen" van Koenker en Bassett, waarmee men de spreiding in waarnemingen en parameterschattingen kan onderzoeken rijp voor meer algemene toepassing. Gebruikelijk is dat een invloedsfunctie oneven is, dat wil zeggen $\psi(-x) = -\psi(x)$. Een extra parameter p en $\psi(-x) = -p\psi(x)$ geven echter interessante mogelijkheden.

Samenvattend kan men stellen dat dit boek zeer geschikt is voor wie een grondige theoretische verdieping van zijn/haar kennis van robuuste statistiek wil. Als algemene inleiding en voor schetsen van achtergronden zijn het eerste en het laatste hoofdstuk zeer geschikt. Op het tussenliggende niveau vindt men weinig stof; daarvoor geschikte hoofdstukken staan in de boeken van Hoaglin e.a. De druk-technische uitvoering doet de hoge prijs recht. Het boek bevat veel referenties en een zeer uitgebreide index van auteurs en onderwerpen.

Referenties

Hoaglin D.C., Mosteller F., Tukey J.W. (eds.) Understanding Robust and Exploratory Data Analysis, Wiley, 1983.

Hoaglin D.C., Mosteller F., Tukey J.W. (eds.) Exploring Data Tables, Trends and Shapes, Wiley, 1985.

Koenker R.W., Bassett G.W., "Regression quantiles", *Econometrica* 50 (1978), 43-61.

Zanakis S.H., Rustagi J.S. (eds), Optimization in Statistics, North-Holland, 1982.

P.H.C. Eilers, DCMR

S.H. Moolgavkar & R.L. Prentice (eds.)

Modern statistical methods in chronic disease epidemiology

John Wiley & Sons, New York, 1986, xiii+282 p,
ISBN 0-471-83904-3, 31.40 pound

SIMS (SIAM Institute for Mathematics and Statistics) werd in 1973 opgericht met als doelstellingen het ontwikkelen, stimuleren en ondersteunen van onderzoek naar de toepassing van wiskunde bij de bestudering en oplossing van sociale problemen.

Dit boek bevat voordrachten gehouden op het 103 SIMS congres "Modern Statistical Methods in Chronic Disease Epidemiology" in 1985.

Zoals uit de volgende opsomming van de bijdragen blijkt waren veel vooraanstaande biostatistici daar aanwezig :

- M.H. Gail : Adjusting for Covariates That Have the Same Distribution in Exposed and Unexposed Cohorts.
- A.S. Whittemore and S. Grosser : Regression Methods for Data with Incomplete covariates.
- S. Greenland : Partial and Marginal Matching in Case-Control Studies.
- R.L. Prentice, S.G. Self and M.W. Mason : Design Options for Sampling Within a Cohort.
- E. Arjas : Stanford Heart Transplantation Data Revisited : A Real-Time Approach.
- P.K. Andersen : Time-Dependent Covariates and Markov Processes.
- S.H. Moolgavkar and D.J. Venzon : Confidence Regions for Case-Control and Survival Studies with General Relative Risk Functions.
- E.D. Lustbader : Relative Risk Regression Diagnostics.
- C.E. Davis, J.E. Hyde, S.I. Bangdiwala and J.J. Nelson : An Example of Dependencies Among Variables in a Conditional Logistic Regression.
- D. Oakes : A Model for Bivariate Survival Data.
- S.G. Self and R.L. Prentice : Incorporating Random Effects into Multivariate Relative Risk Regression Models.
- N. Breslow : Use of the Power Transformation to Discriminate Between Additive and Multiplicative Models in Epidemiologic Research.
- D.C. Thomas : Use of Auxiliary Information in Fitting Nonproportional Hazard Models.
- R.C. Elston : Modern Methods of Segregation Analysis.
- J.P. Rice : Genetic Epidemiology : Models of Multivariate Inheritance and Path Analysis Applied to Qualitative Traits.
- N.E. Day and S.D. Walter : Screening for Cancer of the Breast and Cervix-Estimating the Duration of the Detectable Preclinical Phase.
- D. Krewski, D. Murdoch and A. Dewanji : Statistical Modelling and Extrapolation of Carcinogenesis Data.

Bijna alle bijdragen houden zich bezig met problemen die naar voren komen bij de toepassing, uitwerking en generalisatie van modellen en technieken die de laatste 15 jaar zijn ontwikkeld in het kader van biomedisch/epidemiologisch onderzoek naar risico factoren; in het bijzonder de relatief risico regressie modellen voor de analyse van case-control, cohort en klinische trail data.

Er is sprake van oorspronkelijke bijdragen, overzichten en "case-studies".

Zoals van deze auteurs verwacht mocht worden zijn de papers stimulerend en van een behoorlijk niveau. Een individuele bespreking van elke bijdrage is natuurlijk onmogelijk.

Zonder de andere bijdragen te kort te doen zou ik de aandacht willen vestigen op de bijdragen van Self & Prentice (een interessant model voor multivariate overlevingsduren waarbij de afhankelijkheid door middel van multiplicatieve random effecten wordt ingebouwd en waarvan het recente model van Clayton & Cuzick een bijzonder geval is), Andersen (tijdsafhankelijke covariabelen in

Cox regressie modellen worden geïnterpreteerd in de kontekst van Markov processen met discrete toestandsruimte) en van Prentice, Self & Mason (een efficiënt "case-cohort" opzet voor de analyse van grote cohort onderzoeken).

De enige kritische noot geldt de twee papers met betrekking tot genetische epidemiologie. In de inleiding staat dat het congres mede is opgezet om de interactie tussen onderzoekers geïnteresseerd in genetische factoren en onderzoekers geïnteresseerd in milieufactoren te bevorderen. Uit deze twee nogal summiere overzichten blijkt dat de bij de genetische onderzoekers gehanteerde modellen en invalshoeken nogal afwijken van die der overige bezoekers. Het is dan ook jammer dat de uitvoerige "stimulating discussions" niet in de proceedings zijn opgenomen.

Dit boek is een bijzonder nuttige aanvulling voor de biomedische sectie van elke afdelings/instituutsbibliotheek. Statistici die belangstelling hebben voor een of meer der behandelde onderwerpen hoeven het boek natuurlijk niet aan te schaffen : bijna alle interessante bijdragen zijn reeds (geheel of gedeeltelijk) recent verschenen of zullen nog opgenomen worden in de bekende tijdschriften en in het algemeen in een meer uitvoerige en gestroomlijnde versie.

A. Volovics, RUL

Rectificatie

Bij de bespreking van

J. Oosterhaven en B.B.A. Drewes, Constructie en actualisering van regionale en interregionale input-output tabellen

door J. Wesseling is per abuis de laatste regel van deze recensie weggefallen. Deze regel bevat de informatie over de manier van bestellen van het boek.

Deze laatste regel luidt:

"Tot slot kan nog vermeld worden dat het boek besteld kan worden bij het ETI voor Utrecht, Pieterskerkhof 17, 3512 JR, Utrecht."

Waarvoor excuses.

Onderstaande werken zijn bij de redacteur van deze boekbesprekingsrubriek binnengekomen. Indien een lezer er prijs op stelt één of meer van deze werken te bespreken, wordt hij/zij verzocht dit aan de redacteur kenbaar te maken. Het adres staat vermeld op de eerste bladzijde van dit blad

CWI-publicaties :

B.F. Schriever

Order dependence

CWI Tract 20.

Centrum voor Wiskunde en Informatica, 1985, iii+115 p.
ISBN 90 6196 294 3, f. 17.60

D.P. van der Vecht

Inequalities for stopped Brownian motion

CWI Tract 21

Centrum voor Wiskunde en Informatica, 1985, iii+88 p.
ISBN 90 6196 296 X, f. 13.90

A.J.W. Kolen

Tree network and planar rectilinear location theory

CWI Tract 25

Centrum voor Wiskunde en Informatica, 1985, iii+85 p.
ISBN 90 6196 300 1, f. 12.50