Boekbesprekingen

Redactie

Alex J. Koning

postadres: Vakgroep Econometrie en Besliskunde
Erasmus Universiteit Rotterdam
Postbus 1738
3000 DR Rotterdam
telefoon: 010-4081268/59
fax: 010-4527746
e-mail: koning@few.eur.nl

Besproken boeken in Kwantitatieve Methoden nr. 54

J.K. Lindsey
Parametric statistical inference

Eelko Huizingh
Inleiding SPSS voor Windows

Kleinbaum, D.G.
Survival analysis. A self-learning text

D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen (editors)
Time Series Models In econometrics, finance and other fields

N.J.A. Sloane and Simon Plouffe
The encyclopedia of integer sequences

Gregory F. Lawler
Introduction to stochastic processes

David B. Orr
Fundamentals of applied statistics and surveys

Adam Shwartz and Alan Weiss
Large deviations for performance analysis: queues, communications, and computing

Mark J. Schervish
Theory of statistics

Samprit Chatterjee, Mark S. Handcock, Jeffrey S. Simonoff
A casebook for a first course in statistics and data analysis

Hall, P.
The bootstrap and Edgeworth expansion

M. Ishaaq Bhatti
Testing regression models based on sample survey data

Hox, J.J.
Applied multilevel analysis (incl. diskette)

Barnett, Vic and K. Feridun Turkman (eds.)
Statistics for the environment 2: water related issues

Mann, P.S.
Introductory statistics

J.K. Lindsey
Parametric statistical inference
1996, Clarendon Press, Oxford,
ix + 490 blz, £ 35,00, ISBN 0-19-852359-9.

De auteur richt zich tot mathematisch georienteerde statistici, gevorderde studenten in de mathematische statistiek, die erin geinteresseerd zijn hoe statistiek in de praktijk kan worden toegepast, en toegepaste statistici die de wens hebben het inferentie proces te begrijpen.

Zoals de titel (mij) suggereert, neemt de auteur aan dat de geinteresseerde lezer een of andere familie kansmechanismes in gedachte heeft, die geparametrizeerd kan worden door een eindig stel parameters. Het zijn eigenschappen van deze parameters waarop de inferentie betrekking heeft. De kansmechanismes waar de auteur aandacht aan besteedt zijn steekproeven van onafhankelijke, identiek verdeelde waarnemingen (univariaat of multivariaat), eventueel met een stochastische omvang, zoals bij sequentiele procedures, maar ook series van onafhankelijke waarnemingen met variabele regressor waarden zoals gebruikelijk bij (gegeneraliseerde) lineaire regressie, en ten derde, waarnemingen van stochastische processen. Bij inferentie wordt gedacht aan puntschatting, dat is schatten van parameters en het aangeven van betrouwbaarheidsgebieden, evenals het toetsen van hypotheses en zelfs modelvorming. Bovendien wordt in een tweetal hoofdstukken een overzicht gegeven hoe een statisticus het nemen van beslissingen zou kunnen begeleiden, enerzijds door middel van de frekwentistische denkwijze van het optimalizeren van het te verwachten nut, anderzijds door een Bayesiaanse denkwijze. In een tweetal hoofdstukken wordt de theorie min of meer volledig uitgewerkt, en wel betreffende Poisson regressie, repectievelijk binomiale regressie. Voor zoiets toepasselijks als mutinomiaal verdeelde grootheden, kun je door het hele boek allerlei opmerkingen vinden, die je in staat stellen om doelgericht methodes uit te werken maar het lijkt me voor een toegepast statisticus toch beter om in deze situatie een tekst te zoeken die dit geval meer expliciet behandelt en waarin wellicht ook gewag gemaakt wordt van allerlei ervaringen en speciale omstandigheden die bij betreffende situaties optreden (zoals McCullagh en Nelder, zie verderop).

Het centrale thema is nu al het werk te doen met likelihood funkties, dat wil zeggen, de kansdichtheid geevalueerd in de waarneming, gezien als funktie van de parameters (welgedefinieerd tot op een scalaire vermenigvuldiging, afhankelijk van de referentiemaat). Op die manier krijgen we een inferentieproces, dat niet afhangt van de manier waarop de kansruimte zelf beschreven is, d.w.z. niet afhangend van de referentiemaat.

Bovenstaande contekst is natuurlijk nog veel te algemeen, om over iedere geparametrizeerde familie kansmechanismes iets zinvols te kunnen zeggen. Voor de gepresenteerde methodes dienen de kansmaten binnen een familie equivalent (wederzijds absoluut continu) te zijn. Verreweg de meeste aandacht wordt besteed aan exponentiele families. Daarnaast verschijnen her en der nog exponentiele dispersie families en transformatie families. In een transformatie familie hangen de kansmaten samen via een groep van symmetrieen van de kansruimte, zoals bijvoorbeeld in een locatie-schaal

familie van univariate kansmaten. Wat mij bijzonder deugd deed is een pleidooi van de auteur voor het gebruik van andere stabiele families, dan de klassieke met eindige varianties (d.w.z. de familie der normale kansdichtheden).

Meer algemene opmerkingen. De schrijver beoogt begrip bij te brengen voor het inferentie proces. Voor mij zie ik een tekst waar de theorie in grote lijnen uitgezet wordt, voorzover die goed werkt voor de bekende exponentiele families, gepresenteerd in een algemene contekst. Hierbij wordt het niet duidelijk wat de graad van algemeenheid nu precies is. Men zal vergeefs zoeken naar gedetailleerde formele bewijzen. Het blijft dus open in hoeverre de algemeenheden van toepassing zijn op iemands specifieke probleemstelling. Daartegenover wordt er prima uiteengezet wat de beweegredenen kunnen zijn om bepaalde methodes of principes te gebruiken (zoals het likelihood principe). Om deze kern heen worden zeer geavanceerde onderwerpen aangesneden.

In het boek vindt men naast dit alles nog diverse passages over de praktische kant van de statistiek. Meermalen wordt er ingegaan op het probleem(pje) dat waarnemingen aan continu verdeelde grootheden pas na een numerieke voorbehandeling verwerkt kunnen worden. Er wordt ingegaan op wat men zoal kan doen als er gegevens ontbreken bij een regressieprobleem. De tekst staat boordevol met literatuurverwijzingen, de literatuurlijst bevat 363 items. Ieder hoofdstuk bevat vele opgaven, die meest uit de literatuur gehaald zijn, en voorzien zijn van verwijzingen ernaar. In mijn ogen een bijzonder leerzaam, vormend boek. Ik wil nog graag een aantal kanttekeningen maken, die enigszins kritisch van toon zijn.

De auteur brengt menig idee over in een reeks van algemene opmerkingen. Zo'n opmerking wordt vaak toegelicht met een aantal voorbeelden. Er zijn opmerkingen die mij uit het hart gegrepen zijn, er zijn er die ik niet begreep, maar waar de voorbeelden duidelijk maakten wat er bedoeld werd, er zijn er die ik begreep, maar waar ik zeer oneens mee was, er zijn er die ik wel begreep maar waar ik de grondslag niet van kon inzien terwijl er ook geen literatuurverwijzing bij was. Er worden ideeen aangedragen om technieken robuust te maken, bijvoorbeeld door het opvangen van uitbijters door het wijzigen van het kansmodel. Dit lijkt me zo'n controversieel idee. Als hij ons probeert aan te sporen een beter, juister, kansmodel te gebruiken, men bestudere dit boek. Maar het lijkt mij nogal verkeerd om op een onjuist kansmodel over te stappen om uitbijters op te vangen. Met name als de uitbijters foutieve irrelevante metingen zijn, lijkt me dat er een gezonder alternatief is in de robuuste statistiek. Zie de ontbrekende referentie P.J.Huber (1980), Robust Statistics, John Wiley, New York. Confronterend ook is de manier waarop hij het wijdverbreide gebruik van Fisher-ratio toetsen bij modelvorming (in het lineaire model) aan de kaak stelt en daarvoor 'eigenlijke' model selectie criteria in de plaats stelt, zoals AIC en BIC. In mijn ogen maakt de schrijver het wel erg bont met zijn theorie van bijna 5 pagina's over het probleem van het toetsen van de adequaatheid van een enkele kansverdeling op grond van een enkele waarneming (pag. 148-153).

Ik zou op grond van de achterflap denken een boek aan te treffen dat een soort van mengelmoes zou zijn van een hoogst theoretisch boek, maar daarom niet minder interessant, als O.E.Barndorff-Nielsen (1975), Information and Exponential Families in

Statistical Theory, John Wiley, New York, en het hoogst praktische boek, maar daarom niet minder interessante P. McCullagh & J.A. Nelder (1983,1989), Generalized linear models, 2nd ed. Chapman and Hall, London. Merkwaardig genoeg wordt niet eens verwezen naar dit laatste boek. Hoewel het niet tot het onderhavige onderwerp zou behoren, bevreemdt het mij dat iemand op zo'n globale, methodische, wijze een techniek kan bespreken, zonder op te merken dat dezelfde methode ook toepasbaar kan zijn bij het optimaliseren van een andere nutsfunktie dan de log-likelihoodsfunktie, zoals bij M-schatting (merk de ontbrekende L op), zie Huber, l.c.

dr H.W.M. Hendriks
Vakgroep Wiskunde
Katholieke Universiteit Nijmegen

Eelko Huizingh
Inleiding SPSS voor Windows
1995, Academic Service, Schoonhoven,
346 blz, fl 49,90, ISBN 90-5261-186-6.

Het boek is in twee delen gesplitst:

Deel I Leren werken met SPSS

Deel II Werken met SPSS

In het eerste deel wordt in hoofdstuk 1 o.a. iets gezegd over de benodigde en gewenste hardware en wat u van Windows dient te weten. Het boek is overigens zodanig geschreven dat eventuele 'Windows' leemtes zonder enig probleem vanzelf aangevuld worden. Hoofdstuk 2 gaat over verschillende fasen in een onderzoek en hoe en waar SPSS daarin nuttig kan zijn. Tevens passeren allerlei procedures en opties in SPSS de revue aan de hand van allerlei statistische 'vragen' die in een onderzoek aan de orde kunnen komen. Hoofdstuk 3 gaat over het coderen en omschrijven van variabelen en er wordt een zeer zorgvuldig gekozen datasetje geïntroduceerd.

Voor de hoofdstukken 4, 5 en 6 is het de bedoeling dat de lezer achter een computer plaatsneemt en zelf alles probeert, te beginnen met het zelf invoeren van het datasetje. De presentatie is zeer gedetailleerd en zodanig dat iedereen hier zonder verdere hulp uit zou moeten kunnen komen. Het lijkt me dat een gebruiker na deze hoofdstukken al aardig zijn weg binnen SPSS weet te vinden en zijn output ook nog wel weet te printen of naar zijn tekstverwerker kan sturen.

In het tweede deel wordt de kennis wat uitgebreid en is meer als naslagwerk bedoeld hoewel het ook geschikt is om 'snel' te lezen.

Hoofdstuk 7 is gewijd aan het maken en controleren van het databestand. Het belangrijkste hierin is wat mij betreft het inlezen van data vanuit andere programma's. Vanuit spreadsheets en database programma's zijn er nauwelijks problemen. Er wordt

met geen woord gerept over de mogelijkheid om zogenaamde 'Open DataBase Connectivity' bestanden te verwerken. Bij het inlezen van ASCII bestanden in 'Freefield' formaat staat echter dat de gegevens door komma's gescheiden dienen te zijn: dit is onjuist daar spaties (gelukkig) ook prima werken. Een voor de beginner bijna onoplosbaar probleem kan ontstaan als Windows ingesteld staat op de decimale komma in plaats van de decimale punt; ik denk dat de auteur er goed aan doet om in de volgende druk expliciet te vermelden hoe dit via Windows ingesteld kan worden en dat dit dan ook gevolgen heeft voor andere Windows programma's.

De hoofdstukken 8, 9 en 10 gaan over zaken als nieuwe variabelen berekenen, indelen in klassen, tellen, wegen, selecteren en sorteren en verder het samenvoegen, aggregeren en 'kantelen' van bestanden. Over het algemeen wordt het nut van dergelijke bewerkingen direct aan de hand van een voorbeeld duidelijk gemaakt.

De hoofdstukken 11 tot en met 18 gaan in op concrete analyses, grofweg de functies die in het menu onder 'Statistics' en 'Graphs' vallen en inclusief allerlei opties. Het lijkt me het handigst om dergelijke analyses met al hun opties gewoon vanuit het menu te proberen en het boek pas te raadplegen als er iets echt onduidelijk is. Ik zou de vertaling van 'M-estimator' naar 'Maximum-likelihood estimator' in hoofdstuk 11 overigens graag anders zien (!), ondanks dat ook de SPSS help functie een dergelijke omschrijving geeft. Bij sommige statistics wordt in het boek kort uitgelegd wat er (ongeveer) berekend wordt, bij de wat minder gangbare statistics wordt nauwelijks iets aan de informatie op het scherm toegevoegd. Het zou handig zijn als de auteur literatuurverwijzingen opneemt waarin de betreffende grootheden duidelijk beschreven worden.

Hoofdstuk 19 is het laatste hoofdstuk en gaat in op het afstemmen van SPSS op specifieke wensen. Tenslotte volgt nog een register, een handig overzicht van statistische analyses ingedeeld naar doel(stelling) en meetniveau en op dezelfde manier een overzicht van grafieken.

Wat ik mis in het boek is vermelding van de mogelijkheid die SPSS biedt om achteraf in de uitvoer getallen af te ronden. Een overzicht van handige toetsaanslagen zou prettig zijn, als alternatief voor grote muisbewegingen. Ik vind ook dat de auteur duidelijk moet aangeven dat er een tutor in SPSS zit en dat het geen kwaad kan die tutor eens uit te proberen (de tutor maakt een boek niet overbodig). Verder kan ik me indenken dat bijvoorbeeld de data editor van SPSS niet echt handig is om grote bestanden in te voeren; in dergelijke gevallen mag de auteur gerust zijn mening geven over beperkingen van SPSS.

Mijn belangrijkste conclusie is dat iemand met behulp van dit boek betrekkelijk gemakkelijk op eigen kracht SPSS kan leren beheersen.

dr J.M. Sneek
Vakgroep Econometrie
Vrije Universiteit Amsterdam

Kleinbaum, D.G.
Survival analysis. A self-learning text
1996, Statistics for biology and health,
Springer-Verlag, Berlin,
250 blz, DM 58.00, ISBN 0-387-94543-1.

This is a good introduction to survival analysis. It is well written and easy to read. It bridges the gap between the basic statistical knowledge and the modern literature on survival analysis. The book contains the following six chapters:

1. Introduction to Survival Analysis
2. Kaplan-Meier Survival Curves and the Log-Rank Test
3. The Cox Proportional Hazards Model and Its Characteristics
4. Evaluating the Proportional Hazards Assumption
5. The Stratified Cox Procedure
6. Extension of the Cox Proportional Hazards Model for Time-Dependent Variables

The book can be used for university courses or can serve as a self- learning text. I regret that not much attention has been paid to the use of the Cox model in other sciences. Though the Cox model was developed for medical statistics it is now also widely used in other sciences like labour economics, engineering and environment studies. It would have been interesting to include examples of applications of the Cox model in other sciences as well so that the book would also become attractive for people working in those fields.

drs E. Schulte Nordholt
Divisie Sociaal Economische Statistieken
Centraal Bureau voor de Statistiek

D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen (editors)
Time Series Models In econometrics, finance and other fields
1995, Monographs on Statistics and Applied Probability 65,
Chapman and Hall, London,
225 blz, £ 30.00, ISBN 0-412-72930-X.

Uit de opleidingsactiviteiten van het Netwerk Kwantitatieve Economie, voorloper van het huidige NAKE, is twee keer een verzamelbundel van artikelen voortgekomen welke een beeld vormen van 'the state of the art' van het economisch onderzoek en de

daarin gebruikte econometrische onderzoeksmethoden. Die bundels, beide met de titel 'Advanced Lectures in Quantitative Economics', verschenen in respectievelijk 1990 en 1993, en kenden als primaire doelgroep studenten in de tweede fase van hun studie. Het voorliggende boek is zeer vergelijkbaar in opzet. Het is eveneens voortgekomen uit tweede-fase onderwijs, en wel het tweede 'Séminaire Européen de Statistique', SEMSTAT, dat in december 1994 te Oxford plaatsvond. Zoals de titel van het boek al aanduidt, richtte het seminar zich vooral op de tijdreeks-analyse. Een gelukkige keuze is geweest slechts een vijftal bijdragen in het boek op te nemen: deze zijn daardoor alle een kleine 50 pagina's lang, wat de auteurs toestaat tot een afgeronde en op zich zelf staande introductie van hun onderwerp te komen. Het boek onderscheidt zich daardoor in positieve zin van de vele verzamelwerken waarin de beknoptheid van de beschouwing in vergaande mate afbreuk doet aan de leesbaarheid ervan. En het stelt daarenboven mij in staat de afzonderlijke bijdragen kort aan te geven.

Het eerste hoofdstuk is van de hand van Neil Shephard: 'Statistical aspects of ARCH and stochastic volatility'. Wat deze bijdrage doet verschillen van andere recente overzichtswerken over het modelleren van systemen met tijdsvariërende varianties is de introductie, naast de bekende klasse van ARCH- en GARCH-modellen (de 'observation-driven models' volgens Shepard), van een klasse van modellen waarin volatiliteit middels niet waarneembare variabelen wordt gemodelleerd: de 'parameter-driven models'.

Hoofdstuk twee, 'Likelihood-based inference for cointegration of some nonstationary time series', is een prachtige samenvatting van Søren Johansen van diens in 1995 bij Oxford University Press verschenen leerboek over cointegratie met dezelfde titel. Maar het is meer dan een samenvatting: op sommige plaatsen voegt het aan intuïtie toe, en het gebruikt een empirisch voorbeeld dat veel overtuigender overkomt dan de twee in z'n leerboek.

Ook hoofdstuk drie is te karakteriseren als een samenvatting van een leerboek, maar dan van het nog te verschijnen leerboek over de theorie van het voorspellen van Michael Clements en David Hendry. In hun bijdrage 'Forecasting in macro-economics' benadrukken deze twee auteurs vooral het belang van trendbreuken en andere vormen van parameter variaties in een voorspeltheorie.

Hoofdstuk vier, 'Longitudinal panel data: an overview of current methodology' van Nan Laird is gemotiveerd door de karakteristieke probleemstelling in biomedische toepassingen: de analyse van data met zowel een tijdreeks- als cross-sectie dimensie, waarbij het aantal waarnemingen in de tijdsdimensie gering is ten opzichte van de omvang van de panel.

Een iets wat vreemde eend in de bijt is het laatste hoofdstuk: 'Pricing by no arbitrage' door Bjarne Jensen en Jørgen Nielsen over de equivalentie van lineaire prijsvorming, het concept van 'geen arbitrage' en de martingaal representatie van de prijsvorming van financiële activa. Deels omdat dit hoofdstuk in wat mindere mate een overzichtsartikel is. Maar bovenal omdat het zich als enige minder goed schikt in de titel van het boek: de tijdreeks-analyse heeft hier plaats gemaakt voor de analyse van dynamische modellen in de meer wiskundig-economische traditie.

Het boek is zorgvuldig geredigeerd en fraai uitgegeven. Dat laatste brengt onvermijdelijk de enige negatieve kanttekening met zich mee: door de prijsstelling zal dit boek minder snel z'n weg vinden naar de persoonlijke bibliotheek van aio of onderzoeker.

D. Tempelaar
Vakgroep Kwantitatieve Economie
Rijksuniversiteit Limburg

N.J.A. Sloane and Simon Plouffe
The encyclopedia of integer sequences
1995, Academic Press, San Diego,
585 blz, $ 44.95, ISBN 0-12-558630-2.

The book is, as the title says, an Encyclopedia of Integer Sequences. The present book replaces the first author's Encyclopedia of Integer Sequences published in 1973. The main table of the present Encyclopedia contains over 5000 integer sequences.

Since there are an infinite number of sequences, the authors restrict attention to sequences of nonnegative integers which are well-defined, infinite and known or published i.e. it should have appeared somewhere in the scientific literature.

The sequences in the book are arranged in numerical order up to a starting (or first significant) entry of 945. A typical entry contains:

- the sequence identification
- the sequence itself
- a name or explanation (if any)
- a recurrence relation or generating function
- further references

If the reader encounters an integer sequence "at work or at play", he can use the search strategies described in Chapters 2 and 3. Using differences, recurrences, transformations, etc. most sequences can be identified. For other cases the authors have developed software to find a sequence or to find an explanation for a sequence.

To use a simple look-up service, readers can send e-mail to the address sequences@research.att.com. For advanced look-up service, readers can send e-mail to the adress superseeker@research.att.com. In both cases the message should be of the form

lookup 5 7 7 2 1 5 7

As a test of the book, I took the sequence "5772157...". Using the guidelines of the book, it was easy to find that this is sequence M3755, the decimal expansion of Euler's constant. Another sequence I tried was the sequence

$$105, 1260, 9450, 56980, 302995, \ldots$$

I was however unable to identify this sequence using the book.

Also other sequences were hard to identify using the Encyclopedia. It seems that characterisations or identifications of sequences with a large starting entry are rather scarce in the Encyclopedia. I challenge the reader of this section to find an explanation for the sequence mentioned before!

All by all, this Encyclopedia is the result of a huge amount of work and it should be in the library of anyone who encounters sequences "at work or at fun".

prof. dr E. Omey
EHSAL

Gregory F. Lawler
Introduction to stochastic processes
1995, Probability series,
Chapman and Hall, London,
176 blz, ISBN 0-412-99511-5.

Readership: students with a background in analysis, linear algebra and probability

Stochastic processes play an important role in the scientific world. During their education, students and scientists in e.g. computer sciences, economics, finance, psychology, biology etc. somehow come across stochastic models. The present book provides an introduction to stochastic processes not only for students with a mathematical but also for students with a different undergraduate background.

In the book the author assumes the students are familiar with undergraduate analysis, linear algebra and probability. Also the author encourages the use and the writing of small computerprogrammes in order to solve the exercises and to extend the examples in the text.

The sequence of chapters is standard. The proofs of classical theorems are not given but are well motivated and illustrated. As to the examples and applications, it is a pity that these are not very real and not really practically oriented. The author motivates this by "true applications require a good understanding of the field being studied and it is not a goal of this book to discuss many different fields in which stochastic processes are used".

Positive in the book is that in each chapter the basic definitions are followed by some classical results and one or more examples. Each chapter closes with a set of interesting exercises and questions. Solutions are not given.

A point of regret is that there is no list of references or "further reading". This book is one of the few books without any bibliography or bibliographic notes. During and after reading this book, students should receive at least some guidelines as to where to find additional materials.

To conclude, I would not really recommend this book. The book is not suitable for distance learning as, in my opinion, students need some help in reading it. As a reference book or a book for researchers, it contains not enough information. It could however be used as a handbook during a (lectured) course in stochastic processes.

prof. dr E. Omey
EHSAL

David B. Orr
Fundamentals of applied statistics and surveys
1996, Chapman and Hall, London,
346 blz, £ 27.50, ISBN 0-412-98821-6.

'Fundamentals of Applied Statistics and Surveys' van David Orr is volgens het voorwoord bedoeld als achtergrondtekst voor leken en wetenschappers die statistiek en methodologie gebruiken, maar daar niet zelf in gespecialiseerd zijn. Het boek belooft zich meer op concepten dan op rekentechnische snufjes te richten, en zou moeten dienen als verheldering en achtergrondinformatie voor consumenten en huis-, tuin- en keukengebruikers van statistische informatie. Voor de review heb ik me dan ook voornamelijk gebaseerd op de vraag of de geboden onderwerpen relevant zijn voor het beoogde publiek, of de diepgang adequaat is, en of de uitleg van de stof helder is.

Het boek bestaat uit 4 delen. Deel I behandelt de fundamenten van de statistiek, te weten wat algemeenheden over getallen, data en statistieken. Deel II behandelt verdelingen, gemiddelden en variabiliteit, elk in een apart hoofdstuk. Deel III behandelt in drie hoofdstukken covariantie, relaties tussen variabelen en inferentiele statistiek. In deel IV ten laatste wordt in één hoofdstuk het een en ander over surveys - als exemplarisch onderdeel van statistiek en onderzoeksmethodologie - behandeld, in het bijzonder sampling, sampling error en nonresponse, alsmede een algemeen overzicht in vogelvlucht van de verschillende fasen in een onderzoek.

Om verschillende redenen ben ik minder enthousiast over dit boek dan ik op grond van de inleiding en de auteur verwacht had. De voornaamste reden is dat het boek ongebelanceerd is. Zelfs als men zich de doelgroep voor de geest haalt, behandelt deel I de geboden stof tergend langzaam en op een wel zeer laag niveau. Kwesties als: 'wat zijn getallen?' zijn aardig als ze in een smakelijk filosofisch sausje behandeld worden, maar hier blijft het allemaal in tamelijk kinderlijke regeltjesproduceerderij hangen (bijv. Sxy - SxSy). Waar dan bijvoorbeeld de geschiedenis wel aangehaald wordt, is het op obligate en onvolledige wijze (de 'Golden Age of Greece' wordt wel genoemd, maar niet de

oude Egyptenaren, de Arabieren etc.) Al met al heeft deel I naar mijn smaak te weinig diepgang om te boeien. Het had allemaal behoorlijk wat korter en gedegener gekund.

Vanaf deel II wordt het aardiger: frequentieverdelingen en allerlei soorten van verdelingen worden behandeld, gelardeerd met flink wat plaatjes en weinig formules. Scheefheid en kurtosis komen aan bod. De auteur behandelt vervolgens de modus, mediaan, het gemiddelde alsmede een aantal voor- en nadelen van elk van deze maten (bijv. het effect van extreme waarden op het gemiddelde, resp. de mediaan). Daarna passeren percentielscores (zeer nuttig) en het kwadratisch, geometrisch en harmonisch gemiddelde (wellicht weer iets te veel van het goede) de revue. In het laatste hoofdstuk van deel II komt variabiliteit aan bod. Dit stuk komt echter veel minder uit de verf, de uitleg is ook beduidend minder helder. Ook zijn her en der definities naar mijn smaak te wazig, zeker gezien de doelgroep die behoefte zal hebben aan vastigheid. De definitie van modus ('Statisticians call a score that occurs very frequently a mode') is helemaal krukkig, ook al wordt dat verderop wel weer rechtgezet. Verder wordt er meer dan een hele bladzij uitgewijd over het feit dat de som van de afwijkingen altijd 0 is, terwijl een simpel formuletje dit in één klap laat zien. Naarmate het over méér gaat, wordt het boek interessanter, maar wordt de didactiek minder. Het duidelijkst is dit in deel III. Eerst worden achtereenvolgens variantie, covariantie, standaard scores, correlatie en regressie behandeld. Het volgende hoofdstuk over samenhang bevat hopeloos onoverzichtelijke stukken, een simpel tabelletje had echt niet misstaan bij de behandeling van alle soorten maten voor samenhang tussen variabelen. Ook wordt partiële correlatie wel erg snel geintroduceerd, en worden daar geen goede voorbeelden bij gegeven. In het laatste hoofdstuk van dit deel komt de inferentiele statistiek aan bod: in één hoofdstuk krijgt de lezer kansverdelingen, hypothesetoetsing, vrijheidsgraden, eenvoudige en meer gecompliceerde ANOVA's en betrouwbaarheidsintervallen te verstouwen. De conceptuele basis van ANOVA had een stuk beter uitgelegd kunnen worden. Het laatste deel is weer een stuk aardiger. Een heleboel aspecten van surveys worden hier besproken op een heldere en overzichtelijke wijze. Dit hoofdstuk kan ook gebruikt worden om later nog eens iets na te slaan: wat was ook al weer multistage sampling, wat waren de voor- en nadelen etc.

Al met al is het boek dus onevenwichtig, zowel in het tempo waarmee onderwerpen behandeld worden als in de diepgang van de stof, als in de kwaliteit van de uitleg. Het boek bevat inderdaad niet veel formules (hoewel ook dit wisselt). Het boek zou denk ik veel effectiever zijn met veel meer illustraties en veel meer voorbeelden, inhoudelijke zowel als kleine getallenvoorbeeldjes. De schrijfstijl van de auteur is daarnaast soms wat omslachtig, veel omhaal van woorden, wat het voor (Nederlandse) studenten zeker minder aantrekkelijk maakt. Het is niet onaardig het boek eens gelezen te hebben, maar of ik het zou aanraden aan iemand die eens wil weten van waar statistiek nou eigenlijk over gaat, blijft de vraag.

dr C.C.J.H. Bijleveld
Vakgroep Methoden en Technieken van Psychologisch Onderzoek
Rijksuniversiteit Leiden

Adam Shwartz and Alan Weiss
Large deviations for performance analysis: queues, communications, and computing
1995, Stochastic Modelling Series,
Chapman and Hall, London,
556 blz, £ 39.95, ISBN 0-412-06311-5.

Gedurende de laatste jaren is gebleken dat Large Deviations (LD) de belangrijkste methode is om zeldzame gebeurtenissen in stochastische processen te analyseren, met name in wachtrijprocessen. In de laatste decennia is een aantal wiskundig zware boeken verschenen [5], [6], [4], [3], maar daarnaast ook boeken die gericht zijn op toepassingen (in engineering en informatietheorie), bijvoorbeeld [2]. Het is duidelijk dat er behoefte bestond aan een boek dat de gecompliceerde theorie combineert met specifieke toepassingen, in het bijzonder die op het gebied van telecommunicatiesystemen. Het boek van Adam Shwartz (Technion, Haifa, Israel) en Alan Weiss (AT&T Bell Labs, USA) voorziet hierin: ze brengen een gedegen analyse van LD samen met toepassingen op het gebied van netwerkperformance.

Ruwweg gesproken onderzoekt LD kansen die naar nul gaan in een zekere parameter. Het vindt asymptoten (op een exponentiële schaal) van deze kansen, maar het leert je ook de bijbehorende zeldzame gebeurtenissen te begrijpen. In feite, gegeven dat een bepaalde zeldzame gebeurtenis plaatsvindt, verschaft LD inzicht in het *pad* van het stochastisch proces naar die zeldzame gebeurtenis. Preciezer: conditioneel op de zeldzame gebeurtenis, is er een meest aannemelijk pad daar naar toe (vanuit het evenwicht), dat gevonden kan worden met LD. Denk bijvoorbeeld aan de bufferinhoud van een wachtrijproces: dan vindt LD (i) de exponentiële snelheid van de verlieskans in de buffergrootte en (ii) het meest aannemelijke pad naar een volle buffer (startend met een leeg systeem).

Het eerste deel van het boek is theoretisch van aard. Shwartz en Weiss ontwikkelen de theorie vanaf de eenvoudigste resultaten voor het steekproefgemiddelde van onafhankelijke, gelijkverdeelde stochasten: de bekende stellingen van Cramér en Chernoff. Dan wordt het algemene concept van het LD principe uit de doeken gedaan, en wordt er aangetoond dat de empirische verdeling van onafhankelijke, gelijkverdeelde trekkingen aan een dergelijk principe voldoet. In de andere hoofdstukken van het eerste deel wordt aandacht besteedt aan recente resultaten op het gebied van continue-tijds stochastische processen op een discrete toestandsruimte (Freidlin/Wentzell). Naast alle formele bewijzen wordt er nadruk gelegd op heuristieken en de intuïtie achter de theorie. Daardoor is de tekst, ondanks de technische moeilijkheden, goed leesbaar. Dit komt ook door de uitgebalanceerde collectie instructieve vraagstukken. Kortom: hoewel het onderwerp bijzonder technisch van aard is, slagen de auteurs erin – tenminste naar mijn smaak – een goed evenwicht te vinden tussen zwaardere wiskundige technieken en de toegankelijkheid van de tekst. Hoofdstuk 6 bijvoorbeeld geeft een elegante behandeling van de bijzonder lastige Freidlin-Wentzell theorie.

Het tweede deel van de tekst is gewijd aan toepassingen uit het gebied van de telecommunicatie. Onder andere de volgende modellen worden bekeken:

- Analyse van zeldzame gebeurtenissen in de M/M/1 wachtrij (volle buffers, extreem lange busy periods, etc.)
- Een gedegen analyse van het transiënte gedrag van het bekende Erlang verliesmodel. Voor licht, gematigd en zwaar verkeer wordt het optimale pad naar blokkering afgeleid. Ook wordt een aantal nuttige uitbreidingen beschouwd, bijvoorbeeld multi-rate modellen en trunk reservation schema's.
- LD resultaten over het fundamentele Anick-Mitra-Sondhi (ATM) model: aan-uit bronnen die een gemeenschappelijke buffer hebben. Asymptotiek van de verlieskans wordt behandeld, voor buffergroote 0, voor kleine en grote buffers. Ze beschouwen ook de uitbreiding naar het model met heterogene bronnen, zoals typisch het geval is in de ATM context (beeld, spraak, etc.).
- Analyse van het (continue zowel als discrete-tijds) ALOHA toelatingsprotocol.
- Analyse van wachtrijdisciplines met prioriteiten, zoals 'bedien-de-langste-rij' en 'ga-naar-de-kortste-rij'. Aandacht wordt besteed aan de kansverdeling van de ene rij, gegeven dat de andere een extreme waarde aanneemt.

De sterke kant van het boek is de combinatie van redelijk zware wiskunde en relevante toepassingen. Het is daardoor interessant voor praktijkgerichte ingenieurs, maar ook voor meer fundamenteel-georiënteerde onderzoekers. Het boek is geschreven in informeel Engels, hetgeen ten goede komt aan de toegankelijkheid.

Een aantal aanmerkingen kunnen worden gemaakt. De belangrijkste is dat een groot aantal relevante concepten uit het gebied van LD niet zijn opgenomen, hoewel ze bijzonder nuttig zijn in de context van telecommunicatietoepassingen. Bijvoorbeeld het effectieve-bandbreedte-concept [7] wordt niet behandeld. Een andere omissie is efficiënte simulatie [1] van kleine verlieskansen door importance sampling, waarbij de (optimale) alternatieve verdeling gevonden kan worden met LD technieken.

Verwijzingen

[1] S. Asmussen and R. Rubinstein [1995]. Steady state rare event simulation in queueing models and its complexity properties. *Advances in Queueing: theory, methods and open problems*, J. Dshalalow, ed., p. 429-462.

[2] J.A. Bucklew [1990]. *Large Deviation techniques in Decision, Simulation, and Estimation.* Wiley, New York.

[3] A. Dembo and O. Zeitouni [1993]. *Large Deviations Techniques and Applications.* Jones and Bartlett, Boston.

[4] J.D. Deuschel and D.W. Stroock [1989]. *Large Deviations.* Academic Press, Boston.

[5] M.D. Donsker and S.R.S. Varadhan [1975]. Asymptotic Evaluation of Certain Markov Process Expectations for Large Time, I and II. *Communications on Pure and Applied Mathematics* 28, p. 1-47 and p. 279-301.

[6] M.I. Freidlin and A.D. Wentzell [1984]. *Random Perturbations of Dynamical Systems.* Springer Verlag, New York.

[7] G. Kesidis, J. Walrand, and C.S. Chang [1993]. Effective bandwidths for multiclass markov fluids and other ATM sources. *IEEE/ACM Transactions on Networking,* Vol. 1, p. 424-428.

drs M.R.H. Mandjes
vakgroep Econometrie
Vrije Universiteit Amsterdam

Mark J. Schervish
Theory of statistics
1995, Springer series in statistics,
Springer Verlag Berlin,
699 blz, DM 88,00, ISBN 0-387-94546-6.

De schrijver van dit boek beoogt aan gevorderde studenten (graduates) een begrijpelijk en geavanceerd overzicht te geven van de statistiek. Juist die onderwerpen worden behandeld die voor een beginnende PhD-student nodig zijn om aan een promotie in de statistiek te beginnen, zoals estimation theory, testing, large sample theory etc. Om de lezer een breed perspectief aan te reiken worden de onderwerpen wiskundig rigoreus, vanuit de traditionele kant, de frequentistische kant, en vanuit het Bayesiaanse gezichtspunt belicht. Hiervoor is wel een basiskennis van maattheorie en kanstheorie vereist, wat tevens uitgebreid in de appendices aan de orde komt.

In het eerste hoofdstuk worden parametrische statistische modellen ingevoerd met behulp van DeFinetti's representatie stelling voor 'exchangeable random variables'. Tevens wordt er ook aandacht besteed aan oneindig dimensionale parameters, met als voorbeeld Dirichlet processen en andere 'tailfree' processen. Ter verduidelijking: een steekproef van oneindig veel random variabelen $(X_n)_{n=1}^{\infty}$ is exchangeable dan en slechts dan als er een random grootheid $\mathbf{P}$ is zodanig dat de X_i conditioneel i.i.d. zijn gegeven $\mathbf{P}$. In het tweede hoofdstuk wordt uitgebreid ingegaan op sufficient statistics, exponentiële families en het concept informatie, belicht vanuit de Bayesiaanse en de 'frequentist' gezichtspunten. Het derde hoofstuk is gewijd aan decision theory, waaronder de begrippen admissibility en minimaxity. Dit onderwerp komt al zo vroeg in het boek aan de orde omdat in het vierde hoofdstuk, wat over hypothesen testen gaat, de traditionele aanpak van testen d.w.z. de 'uniformly most powerful' aanpak vergeleken wordt met de belisvormige aanpak. In het volgende hoofstuk komt het schatten van punten en

verzamelingen aan de orde zoals zuivere schatters, MLE schatters, betrouwbaarheidsintervallen, voorspellingen en tolerantie verzamelingen, robuuste schatters en de bootstrap. Hoofdstuk 6 is gewijd aan equivariantie. Vervolgens wordt in hoofdstuk 7 asymptotische schattingstheorie behandeld, waarin o.a. de asymptotische eigenschappen van quantielen, MLE schatters, robuuste schatters en posterior verdelingen. De laatste twee hoofdstukken gaan over situaties waarin random variabelen niet gemodeleerd kunnen worden als 'exchangable'. Hoofdstuk 8 gaat over hierarchische modellen, zoals bv. normale lineaire modellen als one-way ANOVA, two-way mixed ANOVA; niet normale modellen met poisson en bernouilli proces data; mixture models; Gibbs sampling, Markov chain Monte Carlo methoden of successieve substitutie methoden. In hoofdstuk 9 staan enkele onderwerpen uit de sequentiële analyse.

Alhoewel dit boek vrij geavanceerd is en theoretisch, is het prettig leesbaar. Er wordt, naast stellingen, definities en bewijzen ook uitgebreid aandacht besteed aan de achterliggende ideëen. Er staan veel voorbeelden in de tekst en er is een uitvoerige collectie van opgaven. Jammer genoeg zijn er geen antwoorden van de opgaven.

Vooral het feit dat de schrijver parametrische modellen introduceert met behulp van exchangeable random variabelen geeft een ander en mijn inziens wel een duidelijker en consistenter gezichtspunt op de theorie. Daarnaast is het een duidelijk voordeel van het boek dat het de traditionele en de Bayesiaanse theorie naast elkaar zet en vergelijkt. Voor mensen die geïnteresseerd zijn in theoretische statistiek, die de preciese (maattheoretische) achtergronden willen weten is dit boek dus een aanrader.

Dr C. Oudshoorn
Instituut voor Bedrijfs- en Industriële Statistiek
Universiteit van Amsterdam

Samprit Chatterjee, Mark S. Handcock, Jeffrey S. Simonoff
A casebook for a first course in statistics and data analysis
1995, John Wiley and Sons, Chichester,
314 blz, £ 19.95, ISBN 0-471-11030-2.

"Statistiek leren is statistiek doen". Deze uitspraak, afkomstig van mijn wiskundeleraar op de middelbare school, is de basis van het boek. De auteurs zijn van mening dat een cursus "inleiding in de statistiek" de studenten vaak niet meer biedt dan het leren van een aanzienlijke hoeveelheid stof waarbij er nauwelijks tijd is om naar de toepassingen te kijken. De student leert technieken en formules maar heeft geen idee van het belang van de statistiek in onze maatschappij.

Dit boek bevat 61 cases gebaseerd op 36 datasets. Deze datasets staan in ASCII formaat op een bijgeleverde diskette. De cases omvatten een breed scale aan onderwerpen uit de landbouw, industrie, economie, sport, financien etc. Enkele bijzondere cases zijn die over erupties van de "old faithful" geiser, de rampzalig vlucht van de space shuttle

"Challenger", air bags en autotype's, een onderzoek naar condoomgebruik voor AIDS preventie en de geboorte van walviskalf "Hudson".

Het boek bestaat uit vier hoofdstukken: exploratieve data analyse, toegepaste kansrekening, statistische inferentie en regressie analyse. Naast de standaard onderwerpen voor zo'n inleidende cursus komen bijvoorbeeld ook al praktische zaken als overdispersie en leverage aan de orde.

Elke case bestaat uit achtergrondinformatie, keywords en een vraag. Sommige cases zijn reeds uitgewerkt, anderen deels of helemaal niet. Het feit dat datasets meerdere malen gebruikt worden laat zien dat er vaak meerdere analyse mogelijkheden voor bepaalde gegevens zijn.

Hoewel de bronnen ook algemeen toegankelijk zijn, en je dus in principe je eigen cases kunt maken, zijn de auteurs er in geslaagd om er een aantrekkijke verzameling van te maken. De cases zijn duidelijk beschreven, de onderwerpen zijn vaak van Amerikaanse origine (metro van New York, baseball). Voor een inleidende cursus statistiek kan dit boek zeker nuttig zijn. Via internet worden door de auteurs nieuwe cases verzameld en verspreid. Ook de prijs van dit werkje mag de docent niet van aanschaf weerhouden.

Drs A.F. Huele
Instituut voor Bedrijfs- en Industriële Statistiek
Universiteit van Amsterdam

Hall, P.
The bootstrap and Edgeworth expansion
1995, Springer series in statistics,
Springer-Verlag, Berlin,
xiii+352 blz, DM 48.00, ISBN 0-387-94508-3.

Met de opkomst van zeer krachtige desktop computers heeft ook de interesse in statistische methoden die veel rekenwerk vereisen een grote vlucht genomen. Sterker nog, momenteel worden methoden toegepast die begin jaren tachtig slechts in theoretische bijdragen aan de literatuur werden besproken. Eén van die methoden die aan populariteit heeft gewonnen is de bootstrap, een zogenaamde 'resample'-methode waarbij de eindige steekproefeigenschappen van schatters kan worden verbeterd. De methode is geïntroduceerd door Efron in 1979 en Peter Hall is één van de statistici die met papers en boeken heeft bijgedragen aan een beter begrip van de bootstrap.

In het onderhavige boek worden bootstrap schatters voor 'gladde' functies van een steekproef besproken. De eigenschappen van deze schatters worden afgeleid met behulp van Edgeworth-expansies. Het boek beslaat vijf hoofdstukken en evenzoveel appendices. In hoofdstuk 1 bespreekt Hall de bootstrap methode en in hoofdstuk 2 de Edgeworth-expansie, zonder ze met elkaar in verband te brengen. In de hoofdstukken 3 en 4 worden de eigenschappen van bootstrap schatters afgeleid aan de hand van

Edgeworth-expansies. Wiskundige details worden tenslotte in hoofdstuk 5 behandeld. Van de appendices is met name Appendix II over Monte Carlo de moeite van het lezen waard.

In het voorwoord geeft de auteur niet expliciet aan voor welke groep lezers het boek is bedoeld. Het is allicht te technisch en abstract om het te gebruiken in doctoraal colleges, maar het kan zeker een basis vormen voor een AiO-cursus. In dat geval zou het boek moeten worden gecomplementeerd met enige toegepaste problemen: het boek is nogal theoretisch van aard.

De hier besproken paperbackversie uit 1995 is identiek aan de originele hardcover editie uit 1992. Hoewel het boek een boeiend overzicht geeft van de bootstrap zal men de meest recente ontwikkelingen er niet in vinden. Desalniettemin komen de kracht en de mogelijkheden van de bootstrap voldoende naar voren om het een interessant boek te laten zijn. Aan de andere kant kan iemand met minder tijd misschien beter het hoofdstuk van Hall in het *Handbook of Econometrics* Vol. 4 over de bootstrap lezen.

Dr R.H. Koning
Vakgroep Econometrie
Vrije Universiteit Amsterdam

M. Ishaaq Bhatti
Testing regression models based on sample survey data
1995, Avebury, Aldershot UK,
xvi + 200 blz, £ 35.00, ISBN 1-85628-642-8.

A linear regression model is often fitted to sample survey data. When, as is frequently the case, these data were collected in clusters (or blocks), the residuals from the regression are usually positively correlated within the clusters. If one ignores this correlation ρ in estimating the standard errors of the regression coefficients, one underestimates these standard errors. Moreover, even if ρ is small, the underestimate may be considerable, since a positive ρ may inflate the square of such a standard error by a factor as large as $1 + (k - 1)\rho$, where k is the maximum cluster size [2].

In the 1980s Maxwell L. King of Monash University introduced the notion of point optimality (PO). A PO test of a one-dimensional parameter maximizes the power against an alternative value of this parameter for which the power p is neither close to 0 nor close to 1 but has a reasonable value such as 0.5 or 0.8. The notion of beta optimality (BO) was first introduced by Robert B. Davis in 1969. For a BO test the power function is a monotone nondecreasing function of the parameter and reaches a predetermined value p_1 most quickly. For the case of a d-dimensional parameter King and his student P.X. Wu in 1990 suggested using LMMP tests which are locally most powerful on the average (mean) over the d components of the parameter. King (in single-authored and joint papers) applied these three notions in particular to time series.

In 1990 King and the author of the book here reviewed derived the PO tests (which are BO) with $p_1 = 0.5$ and $p_1 = 0.8$ under normality with ρ being the only unknown parameter. They found that, at least at the 5% level, they are approximately uniformly most powerful for the small and medium sizes of sample considered, in contrast to the locally best tests. The present volume reproduces this result. It also contains the derivations of the LMMP tests for the case in which the intracluster correlation coefficient may vary from cluster to cluster; of the PO and LMMP tests for the case in which the sample design includes clusters and subclusters, each with its own intracluster correlation coefficient (ρ_1 and ρ_2, respectively); and of the exact or approximate PO tests for testing the hypothesis that $\rho_2 = 0$, whatever be the value of ρ_1 in $(0, \frac{1}{2}]$. The volume also includes the corresponding invariant tests for the case in which also the scale parameter is unknown. Here numerical comparisons are made between the derived tests and the Lagrange multiplier tests.

Unfortunately the book is not free from disturbing printing errors, such as in (3.8) and in line 3 on page 36; (iv) on page 58 is quite unintelligible. My main criticism of the book is that the question of how to proceed when $(k-1)\rho$ is not small compared to unity is not taken up at all, although this would be very frequently the case (one would often conclude this even without test, viz. on the basis of knowledge of the nature of the material and of the size of k). Of course, the title of the book does not indicate that it would include this matter. On the other hand, the title does not indicate that some of the material in the book is of considerable relevance to the analysis of certain data which do not necessarily arise from sample surveys, as is indicated in the first paragraph on page 14 and the last paragraph of page 41 of the book. The publication is therefore of interest not only to those interested in analyzing sample surveys but also to psychometricians, biometricians analyzing experimental results and some others, as well as to mathematical statisticians who may be challenged by questions such as the one raised above. (There is a reference in the book to a relevant paper [1] by the author of the book in a journal which unfortunately is not available to me at the present time.)

Verwijzingen

[1] Bhatti, M.I. (1993). Efficient Estimation of Random Coefficient Models Based on Survey Data. Journal of Quantitative Economics, vol.9, 99-110.

[2] Scott, A.J. and Holt, D. (1982). The Effect of Two-Stage Sampling on Ordinary Least Square Methods. Journal of the American Statistical Association, vol. 77, 848-854.

Prof. dr H.S. Konijn
Department of Statistics
The Hebrew University of Jerusalem

Hox, J.J.
Applied multilevel analysis (incl. diskette)
1994, TT-publikaties,
x + 112 blz, ISBN 90-801073-2-8.

Het boek is bedoeld als een niet-technische inleiding van de multiniveau analyse. De term multiniveau analyse wordt in dit boek gebruikt als algemene term voor modellen voor geneste data, ofwel data waarin we meerdere niveaus kunnen onderscheiden waarbij de lagere niveaus genest zijn binnen de hogere niveaus. Voorbeelden van geneste data zijn er legio. Een bekend voorbeeld vinden we in onderwijsresearch waarin we gegevens verzamelen over scholieren en waarbij we variabelen hebben op leerlingniveau, op klasniveau en op schoolniveau. In dit geval zijn de scholieren genest in de klassen en zijn de klassen genest binnen scholen. Andere voorbeelden van geneste datastructuren treffen we aan bij longitudinale data, bij meta-analyse en conjuncte analyse.

Het boek bespreekt twee algemene multiniveau modellen (het multiniveau regressie model en multiniveau structurele modellen), illustreert drie computerprogramma's voor multiniveau analyse en geeft enkele toepassingen. De volgende hoofdstukken treffen we in het boek aan:

Chapter 1: Introduction

Chapter 2: Multilevel Regression Models

Chapter 3: Working with HLM, VARCL and ML3

Chapter 4: Special Applications of Multilevel Regression Models

Chapter 5: Structural Models for Multilevel Data

In hoofdstuk 1 worden voorbeelden gegeven van situaties waarin we multiniveau modellen nodig hebben en de auteur geeft aan dat wanneer multiniveau data geanalyseerd worden op een niveau dit twee nadelen geeft: (1) een statistisch nadeel (bij aggregatie door afname van het aantal waarnemingen verlies van power en bij disaggregatie wordt de aanname van onafhankelijke waarnemingen geschonden en tevens door toename van het aantal waarnemingen toename van Type I fout) en (2) een conceptueel nadeel (b.v. het analyseren van data op een hoger niveau maar de resultaten interpreteren op een lager niveau). Hoofdstuk 2 bespreekt het meest gebruikte multi level model: het multiniveau regressie model. Vergeleken met een regressie model op een niveau heeft een volledig multiniveau regressie model veel parameters. De auteur bespreekt in hoofdstuk 2 een procedure, die bij de afwezigheid van een theorie om het aantal parameters te reduceren, stapsgewijs nagaat welke parame- ters wel en niet in het model opgenomen zouden moeten worden. In hoofdstuk 3 worden aan de hand van een uitgewerkt praktijkvoorbeeld de drie bestaande software programma's voor multi-niveau analyse, (HLM, VARCL en ML3), geïllustreerd en met elkaar vergeleken. Hoofdstuk

4 behandelt twee speciale toepassingen van multiniveau regressie-modellen: de meta-analyse en multiniveau analyse voor gegevens met niet-normaal verdeelde residuele termen. Voor dit probleem wordt ingegaan op de multiniveau variant van gegeneraliseerde lineaire modellen. Hoofdstuk 5, tenslotte, gaat in op de wijze waarop structurele modellen voor multiniveau gegevens geschat kunnen worden.

Het boek kan beschouwd worden als een beknopte maar praktische inleiding in de multiniveau analyse. Het boek is helder geschreven en bevat vele voorbeelden die verschillende modellen en toepassingen duidelijk illustreren.

M. Vriens
Sectie Marktkunde FEW
Rijksuniversiteit Groningen

Barnett, Vic and K. Feridun Turkman (eds.)
Statistics for the environment 2: water related issues
1994, John Wiley and Sons, New York,
xiv + 389 blz, $ 49.95, ISBN 0471-95048-3.

"Statistics for the Environment 2" is het tweede deel in de Wiley serie met de proceedings van de SPRUCE conferenties (Statistics in Public Resources, Utilities and in Care of the Environment). Dit deel bevat 19 uitgewerkte lezingen van de tweede SPRUCE conferentie gehouden in 1993 aan het Rothamsted Experimental Station in Engeland en heeft als ondertitel "Water Related Issues".

Het bestaat uit de volgende vier samenhangende onderdelen:

1. Rainfall and Climate
2. Sea Levels and Wave Energy
3. Water Quality, Supply and Management
4. Hydrological Modelling

Het doel van dit tweede deel van Statistics in the Environment is volgens het voorwoord tweeledig:

1. het formuleren van goede statistische modellen en methoden
2. het gebruik ervan in specifieke, praktische toepassingen.

De doelgroep lijkt me op de eerste plaats de deelnemers te zijn aan SPRUCE II, en voor deze mensen lijkt me het boek waardevol. Verder zou het boek interessant kunnen zijn voor statistici die zich met de genoemde onderwerpen bezig houden en op zoek zijn naar nieuwe inzichten of modellen.

In een aantal artikelen wordt op heldere wijze nader ingegaan op de tijd-ruimte modellering van o.a. neerslaggegevens en meteorologische gegevens. Veel van de voorgestelde modellen staan nog in de kinderschoenen en de toepassingen tonen aan dat er nog veel te doen is op dit gebied. Een drietal artikelen sprongen er voor mij uit, omdat deze zeer goed de dubbele doelstelling van het boek halen.

Castro en Perez-Abreu uit Mexico modelleren in "Some Statistical Analyses of the Cluster Point Processes of Hurricanes on the Coasts of Mexico" de clustergrootte in de tijd van orkanen die de kust van Mexico bereiken. De clustergrootte wordt gemodelleerd m.b.v. een getrunceerde negatief binomiale verdeling en vervolgens worden twee modellen opgesteld via grafische informatie. De parameters van de gedefinieerde niethomogene Poisson processen blijken vervolgens goed te fitten met de gegevens over de jaren 1966–1992.

In "Statistical Methods for the Evaluation of Hydrological Parameters in Landuse Planning" van Busch, Brechtel en Urfer uit Duisland worden resultaten van lange termijn monitoring in de Rijn-Main vallei geanalyseerd, waarbij vele ecologische parameters worden meegenomen. Doel van het onderzoek is te laten zien dat enkele statistische methoden goed geschikt blijken te zijn om veranderende patronen in tijd en ruimte te analyseren en te beschrijven. Zij tonen dit aan voor de analyses m.b.v. herhaalde waarnemingen en voor tijdreeksanalyse met interventie en input variabelen.

Nirel uit Israel behandelt in "Bootstrap Confidence Intervals for the Estimation of Seeding Effect in an Operational Period" het effect op het uitregenen van wolken m.b.v. "inzaaingen" in Israel. Hiertoe dient hij betrouwbaarheidsintervallen te simuleren. In zijn artikel worden drie verschillende bootstrapmethoden met elkaar vergeleken.

De SPRUCE conferenties zijn een spinn-off van de samenwerking van de statistische vakgroepen van de universiteiten van Lissabon en Sheffield. Dat is in dit tweede deel opvallend vanwege het groot aantal auteurs uit Engeland dat ook in het eerste deel publiceerde. Het geeft een indruk van "deja vu", van onderzoek dat in één jaar verricht is en nog niet voldoende uitgediept is, van beperktheid omdat alleen bepaalde technieken worden gebruikt. Met name in het extreme waarden onderzoek leidt dit tot eenzijdig gebruik van de Gegeneraliseerde Pareto Verdeling waarbij waardevol verdelingsvij onderzoek in Amerika en Europa niet eens vermeld wordt. Waarschijnlijk is de korte opeenvolging van de eerste twee SPRUCE congressen daar debet aan.

dr A.L.M. Dekkers
Rijksinstituut voor Volksgezondheid en Milieu
Bilthoven

Mann, P.S.
Introductory statistics
1994, John Wiley and Sons, New York,
xx + 774 blz, £ 24.95, ISBN 0-471-01017-0.

Het boek Introductory Statistics is bedoeld als inleiding voor studenten van verschillende disciplines. Het boek bestaat uit 13 hoofdstukken. De volgende onderwerpen worden behandeld: introductie basisbegrippen (beschrijvende en toetsende statistiek, populatie en steekproef, typen variabelen (kwantitatieve en kwalitatieve variabelen)); frequentieverdelingen; grafieken; histogrammen; 'stem and leaf displays'; centrummaten en spreidingsmaten; kwartielen, percentielen en boxplots; kansberekening (product- en somregel); discrete kansverdelingen (binomiale verdeling en Poissonverdeling); normale verdeling; t-verdeling; steekproevenverdelingen van gemiddelden en proporties; schatten van gemiddelden en proporties (betrouwbaarheidsintervallen); toetsen van hypothesen over (twee) gemiddelden en (twee) proporties; de CHI-kwadraat; toets enkelvoudige lineaire regressie; de eenweg-variantie-analyse.

Het boek "Introductory Statistics" is zorgvuldig opgebouwd: kleine stukjes stof die onmiddellijk gevolgd worden door voorbeelden en veel opgaven. Belangrijke termen en formules die in de tekst worden behandeld, worden ook apart in tekstboxen afgedrukt en aan het eind van elk hoofdstuk is er een lijst van belangrijke termen (Glossary) en van de voornaamste formules (Key Formulas). Nadat een stukje stof is behandeld, volgt er vaak een voorbeeld (Example: in totaal ongeveer 200). Behalve de voorbeelden zijn er ook 32 Case-Studies (gebaseerd op bestaande gegevens) die de stof ook nog weer eens toelichten. Aan het eind van elke paragraaf staan een aantal opgaven en aan het eind van elk hoofdstuk staan ook nog weer opgaven die op het hele hoofdstuk betrekking hebben. In totaal bevat het boek bijna 1100 opgaven. Daarnaast bevat het boek een groot aantal tabellen en figuren. Aan het eind van elk hoofdstuk staat een zelftoets (in totaal 172 opgaven). Elk hoofdstuk bevat tenslotte een aantal 'grijze bladzijden' waarin gedetailleerde beschrijvingen gegeven worden van de MINITAB-commando's die men nodig heeft voor het verrichten van de statistische analyse die in het betreffende hoofdstuk aan de orde is geweest. SPSS komt niet aan de orde.

Men kan een aantal handleidingen aanvragen: docent (bevat de volledige oplossingen voor alle opgaven, ook die van de zelftoetsen); student (oplossingen van de oneven genummerde opgaven en van de zelftoetsen); leeswijzer (hoofdpunten uit elk hoofdstuk); itembank (meerkeuze-vragen en open opgaven, geordend per hoofdstuk, is ook op diskette beschikbaar).

Zoals bij elk boek, blijft er ook bij dit boek hier en daar wel wat te wensen over. Zo wordt er wel een onderscheid gemaakt tussen kwantitatieve variabelen (discrete vs continue) en kwalitatieve variabelen (sexe, haarkleur), maar het onderscheid ordinale variabelen en variabelen op intervalniveau ontbreekt, m.a.w. er wordt geen aandacht besteed aan meetniveaus.

Hoe men hypothesen moet toetsen omtrent proporties wordt duidelijk behandeld, maar er wordt niet bij verteld wat te doen in het geval van kleine steekproeven (kleiner

30). Nonparametrische toetsen (behalve dan de CHI-kwadraat) komen niet aan de orde.

Alleen de enkelvoudige regressie-analyse wordt behandeld; wat beta-gewichten zijn, komt niet aan de orde en ook niet de mogelijkheid dat er meerdere onafhankelijke variabelen kunnen zijn. Bij variantie-analyse wordt alleen de eenwegvariantie-analyse behandeld; de tweewegvariantie-analyse komt niet aan de orde.

De conclusie is dat complicaties vermeden worden, mogelijke bredere verbanden worden niet aangegeven. De vraag is of dat voor een eerste jaars-statistiek boek erg is: studenten kunnen door een teveel aan mogelijke verbanden ook in de war raken. De genoemde tekortkomingen kunnen eventueel ook goed door een docent worden opgevangen.

Datgene wàt uitgelegd wordt, wordt over het algemeen goed en helder uitgelegd, hoewel het tempo traag is. Dit maakt dat het boek mogelijk geschikt is voor zelfstudie en/of voor studenten in de sociale wetenschappen met weinig affiniteit voor statistiek.

dr A.A.J. van Peet
Faculteit Pedagogische Onderwijskundige Wetenschappen
Universiteit van Amsterdam

Binnengekomen boeken

Hieronder volgt een overzicht van de recente uitgaven die sinds de vorige aflevering van Kwantitatieve Methoden bij de redactie zijn binnengekomen. Mocht er belangstelling bestaan voor het bespreken van een boek, dan kan contact opgenomen worden met de boekbesprekingsredacteur. Indien het boek dan nog niet vergeven is, krijgt de kandidaat het boek gratis thuisgezonden. Hiervoor dient dan binnen een half jaar een recensie te worden teruggezonden. Een aantal titels is niet fysiek toegezonden, maar kan door ons besteld worden bij de uitgever. Dit heeft consequenties voor de levertijd van deze titels.

John I. Marden
Analyzing and modeling rank data
1996, Monographs on statistics and applied probability,
Chapman & Hall, London,
xiii+329 blz, £ 32.00, ISBN 0-412-99521-2.

Hayakawa, T., Aoshima, M., Shimizu, K. (editors)
Multivariate statistical analysis in honor of Professor Minoru Siotani on his 70th birthday. Volume I of the proceedings of MSI-2000
1995, American Sciences Press, Syracuse, New York,
238 blz, $ 125, ISSN 0196-6324.

Tanner, M.A.
Tools for statistical inference. Methods for the exploration of posterior distributions and likelihood functions
1996, Springer series in statistics,
Springer-Verlag, Berlin,
200 blz, DM 78,00, ISBN 0-387-94688-8.

Haberman, S.J.
Advanced statistics. Volume I: description of populations
1996, Springer series in statistics,
Springer-Verlag, Berlin,
515 blz, DM 88,00, ISBN 0-387-94717-5.

Bruce, A., Gao, H.-Y.
Applied wavelet analysis with S-plus
1996, Springer-Verlag, Berlin,
430 blz, DM 78,00, ISBN 0-387-94714-0.

Huet, S., Bouvier, A., Gruet, M.-A., Jolivet, E.
Statistical tools for nonlinear regression. A practical guide with S-plus examples
1996, Springer series in statistics,
Springer-Verlag, Berlin,
154 blz, DM 65.00, ISBN 0-387-94727-2.

Tanizaki, H.
Nonlinear filters. Estimation and applications
1996, Springer-Verlag, Berlin,
254 blz, DM 128.00, ISBN 3-540-61326-9.

Small, C.G.
The statistical theory of shape
1996, Springer series in statistics,
Springer-Verlag, Berlin,
229 blz, DM 78.00, ISBN 0-387-94729-9.

Saville, D.J., Wood, G.R.
Statistical methods. A geometric primer
1996, Springer-Verlag, Berlin,
275 blz, DM 68.00, ISBN 0-387-94705-1.

Polyak, I
Computational statistics in climatology
1996, Oxford University Press, Oxford,
372 blz, £ 49.50, ISBN 0-19-509999-0.

Borst, S.C.
Polling systems
1996, CWI Tract **115**,
Centrum voor Wiskunde en Informatica, Amsterdam,
vi+232 blz, fl 50.00, ISBN 90-6196-467-9.

Varian, H.
Computational economics and finance. Modelling and analysis with mathematica
1996, Springer-Verlag, Berlin,
480 blz, DM 88,00, ISBN 0-387-94518-0.

Galambos, J., Simonelli, I.
Bonferroni-type inequalities with applications
1996, Springer-Verlag, Berlin,
270 blz, DM 88,00, ISBN 0-387-94776-0.

Christensen, R.
Plane answers to complex questions. The theory of linear models
1996, Springer series in statistics,
Springer-Verlag, Berlin,
465 blz, DM 88,00, ISBN 0-387-94767-1.

Stine, R., Fox, J. (eds)
Statistical computing environments for social research
1997, Sage, Thousand Oaks, California,
250 blz, £ 16.95, ISBN 0-7619-0270-8.

Krickeberg, K.
Petit cours de statistique
1996, Springer-Verlag, Berlin,
149 blz, DM 29,00, ISBN 3-540-61609-8.

Montgomery, D.C.
Introduction to statistical quality control
1996, John Wiley & Sons Ltd., Chichester,
700 blz, £ 21.95, ISBN 0-471-30353-4.

Does, R.J.M.M., Roes, K.C.B., Trip, A.
Statistische procesbeheersing in bedrijf. Implementeren en verankeren van SPC
1996, Kluwer quality handboeken,
Kluwer, Dordrecht,
263 blz, ISBN 90-267-2189-7.

Anderson, T.W., Finn, J.D.
The new statistical analysis of data
1996, Springer-Verlag, Berlin,
640 blz, DM 88, ISBN 0-387-94619-0.

Mareels, I., Polderman, J.W.
Adaptive systems: an introduction
1996, Birkhäuser, Basel,
362 blz, DM 108.00, ISBN 3-7643-3877-6.

Feuer, A., Goodwin, G.C.
Sampling in digital signal processing and control
1996, Birkhäuser, Basel,
560 blz, DM 138.00, ISBN 3-7643-3934-9.

Borodin, A., Salminen, P.
Handbook of Brownian motion - facts and formulae
1996, Birkhäuser, Basel,
476 blz, DM 168.00, ISBN 3-7643-5463-1.

Holden, H., Oksendal, B., Uboe, J., Zhang, T.
Stochastic partial differential equations: a modeling, white noise functional analysis approach
1996, Birkhäuser, Basel,
231 blz, DM 118.00, ISBN 3-7643-3928-4.

Ogden, T.
Essential wavelets for statistical applications and data analysis
1996, Birkhäuser, Basel,
300 blz, DM 76.00, ISBN 3-7643-3864-4.

Fisher, L.D., van Belle, G.
Biostatistics. A methodology for the health sciences
1993, Wiley series in probability and mathematical statistics: applied probability and statistics section,
John Wiley & Sons Ltd., Chichester,
xxii+991 blz, £ 37.50, ISBN 0-471-16609-X.

Gibbons, J.D.
Nonparametric methods for quantitative analysis (third edition)
1997, American series in mathematical and management sciences,
American Sciences Press, Syracuse, New York,
xvi+537 blz, $ 70.95, ISBN 0-935950-37-0.

Brockwell, P.J., Davis, R.A.
Introduction to time series and forecasting
1996, Springer texts in statistics,
Springer-Verlag, Berlin,
430 blz, DM 98,00, ISBN 0-387-94719-1.

Kijima, M.
Markov processes for stochastic modeling
1996, Stochastic modeling series,
Chapman & Hall, London,
x+341 blz, £ 35,00, ISBN 0-412-60660-7.

Beirlant, J., Teugels, J.L., Vynckier, P.
Practical analysis of extreme values
1996, Leuven University Press, Leuven,
vii+137 blz, BEF 950,-, ISBN 90-6186-768-1.

Oproep voor boekbesprekers

De redactie van Kwantitatieve Methoden is steeds op zoek naar mensen die bereid zijn een pas verschenen boek te lezen met het doel daarover een oordeel te geven. Zo'n recensie verschaft de leden van de VVS en overige lezers van Kwantitatieve Methoden inzicht in de kwaliteiten van het boek en verschaft nuttige informatie bij een overwogen aanschaf. Voor de schrijvers biedt een kritisch oordeel van het publiek waarvoor het boek geschreven een mogelijkheid tot verdere verbetering. De uitgever heeft uiteraard meer baat bij een lovende recensie. Als tegenprestatie mogen de recensenten de besproken boeken behouden.

Behalve de algemene boeken over statistiek of besliskunde, waar veel mensen zinnige dingen over kunnen zeggen, behandelen veel boeken een specialistisch onderwerp. Als wij, soms op de gok, iemand een dergelijk boek toesturen, blijkt een enkele keer dat 'het boek precies aansluit bij het onderzoek' van de recensent. Dat is prettig, omdat veel mensen daarbij voordeel hebben. De recensent, omdat hij/zij het nieuw verschenen boek op het onderzoeksgebied ter beschikking heeft, en de auteur(s) omdat er een terzake kundig oordeel wordt gegeven. Om nu de succeskans van het toesturen van een boek te vergroten, willen wij graag in contact komen met zoveel mogelijk collega statistici, kansrekenaars en besliskundigen. Stuur het onderstaande antwoordstrookje naar het adres van de boekbesprekingsrubriek als u interesse heeft om in de toekomst een boek te bespreken. Als er dan in de toekomst wellicht een gloednieuw boek op uw onderzoeksgebied binnenkomt hebben wij een grotere kans op een perfecte match.

Ik ben graag bereid een boek op mijn interessegebied te bespreken.

Naam	
Adres	
Telefoon	
e-mail	
Interessegebieden (graag nauwkeurig omschrijven)	