

Boekbesprekingen

Redactie:

Arend D. Oosterhoorn
Van Doorne's Transmissie bv
Postbus 500
5000 AH Tilburg

telefoon 013 - 640319

Roderick J.A. Little en Donald B. Rubin

Statistical Analysis with Missing Data

John Wiley & Sons, New York, 1987, xiv + 278 pag.,
ISBN 0-471-80254-9, £ 32.05.

Dit is een leuk en nuttig boek. Iedere statisticus die zich met consultatie bezig houdt, komt regelmatig problemen tegen waar bepaalde waarnemingen ontbreken.

Het boek begint met een overzicht over vaak voorkomende situaties met ontbrekende waarnemingen, en over 'klassieke' methoden om daar mee om te springen. Dat zijn enerzijds klassieke en goed bekende methoden (de regressie-methode van Yates en de covariantie-methode van Bartlett), anderzijds 'quick' methoden, zoals het werken met alleen de volledige 'cases' of het invullen van de ontbrekende waarden, die gewoonlijk ook 'dirty' zijn.

Een hoofdstuk is gewijd aan gebruikelijke methoden voor het behandelen van non-response in steekproeven, zoals de hot-deck imputatie (dat is zo'n mooi woord dat niet kan nalaten het even te noemen).

Vervolgens worden methoden besproken die meer de goedkeuring en de interesse van de auteurs kunnen wegdragen, en aan de ontwikkeling waaraan zij veel hebben bijgedragen. Deze zijn op de likelihood gebaseerd; vooral op de marginale likelihood van de wel waargenomen waarden.

Hierbij wordt het EM algoritme uitvoerig besproken en in veel situaties toegepast: voor (multivariaat) normale verdelingen, waaronder lineaire regressie, voor kruistabellen (waar klassifikaties onvolledig kunnen zijn), en voor combinaties van normaal verdeelde en kategorische variabelen. Hierbij wordt steeds verondersteld, dat waarden 'at random' ontbreken: het 'missing data mechanism' wordt dan 'ignorable' genoemd.

Helaas is aan deze voorwaarde vaak niet voldaan. Een hoofdstuk is gewijd aan een aantal modellen voor 'non-ignorable missing data models', o.a. modellen voor afgeronde of gegroepeerde data, stochastische censureringsmodellen en modellen voor kategorische met verschillende a priori respons-klassen in de verschillende cellen. Tot slot wordt een hoofdstuk besteed aan de bayesiaanse benadering van non-response in steekproeven, die Rubin nader uitwerkt in zijn boek 'Multiple imputation for nonresponse in surveys' (John Wiley & Sons, 1986).

Het boek is helder geschreven en bevat veel voorbeelden. De bestaande literatuur wordt niet uitputtend beschreven. Het boek van Y. Dodge (Analysis of Experiments with Missing Data), in 1985 bij dezelfde uitgever verschenen, wordt bijvoorbeeld niet eenmaal genoemd. Dat boek gaat wel een heel andere kant op (kleinste kwadraten oplossingen e.d. in designed experiments), maar een verwijzing was wel op zijn plaats geweest. Ook de techniek van 'data augmentation of streamflow records' zoals hydrologen dat noemen, en die in feite hetzelfde is als enkele van de in hoofdstuk 6 e.v. behandelende technieken, wordt niet genoemd. Dat is een stukje van de literatuur op dit gebied waar ik toevallig van op de hoogte ben, maar als dit ontbreekt, zal er ook wel wat anders ontbreken.

Ondanks deze laatste kritiek is het boek zeker een aanrader voor wie met ontbrekende waarnemingen gekonfronteerd wordt en voor wie in toepassingen van het EM algoritme geïnteresseerd is.

Tom A.B. Snijders
Rijksuniversiteit Groningen

K. van Montfort

Estimating in structural models with non-normal distributed variables: some alternative approaches.

DSWO-press, Leiden, 1989, 151 pag., ISBN 90-6695-030-7, f 28,00.

Strukturele modellen in de economische en sociale wetenschappen gaan meestal uit van numerieke multivariaat normaal verdeelde variabelen. In empirisch onderzoek echter, zijn de variabelen meestal niet normaal verdeeld. Dit boek laat zien hoe enkele gebruikelijke modellen en technieken aangepast moeten worden, wanneer we met variabelen te maken hebben die niet-normaal verdeeld zijn.

In geval van multivariaat normaal verdeelde variabelen bevatten de gemiddelden en de varianties/covarianties alle informatie van de data. Wanneer de variabelen niet normaal verdeeld zijn, is het mogelijk om additionele informatie uit andere statistieken, zoals bijvoorbeeld hogere orde momenten, te halen. Er zijn alternatieven om multivariate verdelingen te specificeren. Na Fourier transformaties van de dichtheden van de variabelen is het mogelijk om gebruik te maken van de empirische karakteristieke funktie. Het doel van deze transformatie is om de 'course of dimensionality' van de verdelingsfuncties van de variabelen te reduceren. Deze 'dimensionaliteit' levert namelijk grote variaties en slechte robuustheid op bij kleine steekproeven.

In dit boek worden de twee bovengenoemde methoden gebruikt om multivariate verdelingen te specificeren: hogere orde momenten en karakteristieke funkties worden gebruikt. Er worden niet parametrische schatters afgeleid voor de volgende strukturele modellen:

- 1) het lineaire regressiemodel met meetfouten in de variabelen (errors-in-variables, EIV)
- 2) een (tweedegraads) polynoom regressie model met EIV
- 3) een regressie model met EIV voor dichotome variabelen
- 4) een faktoranalyse model.

Het boek bestaat uit 3 delen.

In het eerste deel worden derde orde momenten gebruikt om regressie coëfficiënten te identificeren en te schatten. Dit wordt gedaan voor het lineaire regressie model met EIV, het polynome regressie model met EIV, en het regressiemodel met EIV voor dichotome variabelen.

In het tweede deel worden karakteristieke funkties gebruikt om onbekende modelparameters te schatten. Dit wordt gedaan voor zowel het lineaire regressie model met EIV als het faktoranalyse model.

In het derde en laatste deel worden enkele recente resultaten besproken over semi-parametrisch maximum likelihood schatten. Voor het regressie model met EIV wordt deze ML-methode vergeleken met de benadering via hogere orde momenten en met de benadering via karakteristieke funkties.

Vanwege de vele literatuurverwijzingen en besprekingen is het boek goed bruikbaar als introductie tot onder andere de regressie-analyse met meetfouten in variabelen. Sommige delen van het boek, met name de delen die het gebruik van karakteristieke functies beschrijven, vereisen echter wel een gedegen kennis van de mathematische statistiek.

Daarnaast komen in het boek enkele veelbelovende schattingsmethoden aan de orde waaruit mijns inziens blijkt hoe vruchtbaar het gebruik van de mathematische statistiek in de econometrie en psychometrie kan zijn.

Over het geheel genomen is het boek duidelijk en helder geschreven.

Marco Vriens
Vakgroep Sociale Zekerheid
K.U. Brabant

Magnus, J.R. en H. Neudecker

Matrix Differential Calculus with Applications in Statistics and Econometrics

John Wiley & Sons, New York, 1988, xiii + 393 pag.,
ISBN 0-471-91516-5, £ 24.50.

Nederlanders schijnen iets te hebben met matrices. In ieder geval vormt de matrixtheorie een gebied waar landgenoten relatief veel hebben bijgedragen. Het boek van Jan Magnus en Heinz Neudecker in de prestigieuze Wiley-reeks 'Applied Probability and Statistics' is een nieuwe bijdrage aan deze traditie. Hun boek bestaat uit zes delen, bij elkaar 17 hoofdstukken omvattend. Aan de hoofdstukken zij oefeningen toegevoegd, plus geannoteerde verwijzingen naar de relevante literatuur. In het eerste deel worden matrices geïntroduceerd, en wordt de matrixtheorie van de grond af aan opgebouwd. Allerlei klassieke resultaten komen aan de orde, zoals de diverse decomposities, maar ook een aantal resultaten die in de statistiek en econometrie recenter in de mode zijn gekomen, zoals de vec-operator, het Kronecker-produkt, gegeneraliseerde inversen, de 'commutatief'-matrix, die de volgorde van het Kroneckerprodukt van twee matrices omdraait, en de 'duplicatie'-matrix, die een vector van lengte $n(n+1)/2$ uitbreidt tot de 'vec' van een symmetrische $n \times n$ -matrix.

Deel twee van het boek gaat over de theorie van de matrix-differentiaalrekening. Dit deel begint met een stuk klassieke analyse, waarna differentiaten en afgeleiden volgen. In het bijzonder worden de problemen behandeld die aan bod komen bij het ordenen van de vele elementen die (bijvoorbeeld) de tweede afgeleide vormen van een matrix-waardige functie van een matrix-argument. Tevens wordt in dit deel een stuk optimaliseringstheorie besproken.

In het volgende deel wordt de praktijk van het differentiëren besproken. Differentiaten en afgeleiden (eerste en tweede orde) van een aantal in de praktijk veel voorkomende functies zoals spoor, determinant, (één) eigenwaarde van een matrix of een functie van matrices worden behandeld. In het bijzonder zijn de behandeling van het differentiëren van eigenwaarden en eigenvectoren, en van Moore-Penrose-inversen interessant.

Deel vier bevat een grote verzameling ongelijkheden, te beginnen met Cauchy-Schwarz, om via een aantal min of meer exotische ongelijkheden te eindigen bij

kleinste kwadraten. Overigens gaat het steeds om ongelijkheden in een scalaire zin en komen matrix-ongelijkheden (in de Löwner-zin) alleen in een paar oefeningen aan bod.

Het lineaire regressiemodel is onderwerp van deel vijf. Ter introductie is een eerste hoofdstuk gewijd aan de invoering van een aantal elementaire begrippen uit de statistiek. Vervolgens worden onder meer OLS, GLS, schatbaarheid, lineaire restricties, singuliere variantiematrices van de storingenvector behandeld, BQUE voor σ^2 , BLUS en BLUF behandeld. Daarna komen aan bod (deel zes) ML-schatting, in bijzonder van simultane modellen (FIML en LIML). Een paar thema's samengevat onder de noemer psychometrie sluiten het boek af. Het gaat hierbij om een aantal varianten van componentenanalyse, en (kort) om factoranalyse. De nadruk ligt steeds op de algebra van eerste- en tweede- orde voorwaarden.

Uit deze opsomming blijkt dat er veel onderwerpen aan de orde komen in het boek. Zoals we gewend zijn van deze auteurs is de uitleg steeds helder, worden resultaten steeds met bewijzen gegeven en worden afleidingen gewoonlijk volledig, met alle tussenstappen, gegeven, en is de presentatie uiterst nauwkeurig.

Die veelheid aan behandelde onderwerpen vormt tevens het zwakste punt van het boek. Door de breedheid van de inhoud kan onmogelijk alles even gedetailleerd worden behandeld, wat tot een aantal onevenwichtigheden leidt. Sommige thema's worden zeer diepgaand besproken. Dit geldt bijvoorbeeld voor het lineaire model in hoofdstuk 13, dat een elegante uitwerking bevat van allerlei gevallen van restricties op de regressorenmatrix in samenhang met mogelijke singulariteit van de variantie-matrix van de storingen. Een dergelijk soort nauwgezette en gedetailleerde uitwerking vindt men bijvoorbeeld ook in hoofdstuk 16 bij de afleiding van tweede-orde-aspecten van simultane modellen. Hier staat tegenover dat andere zaken er bekaaid afkomen. Zo is de hoeveelheid geïntroduceerde statistische begrippen zeer beperkt gehouden, wat er bijvoorbeeld toe leidt dat op blz. 296/7 de BQUE voor de variantie-matrix van de multidimensionale normale verdeling in anderhalve bladzijde wordt afgeleid door middel van differentiëren, wat via de theorie van de 'afdoende' statistics en de stelling van Lehmann-Scheffé in een paar regels had gekund.

Ook binnen deel een, over matrices, is de evenwichtigheid niet altijd bevredigend. Zo wordt wel de duplicatiematrix D_n besproken (inclusief het wat exotische verband tussen D_n en D_{n+1}) maar niet zijn tegenhanger, de 'eliminatie-matrix', die redundant elementen uit de 'vec' van een symmetrische matrix weglaat. Dat is vreemd want beide nuttige begrippen zijn door de auteurs zelf geïntroduceerd, en beide zijn vaak tegelijk dienstig. Ook is het merkwaardig dat bij gegeneraliseerde inversen steeds de Moore-Penrose-inverse wordt genomen. De algemene g-inverse brengt het niet verder dan een voetnoot en de opgaven. Over de rang van gepartitioneerde matrices wordt weinig vermeld en het begrip 'projector' komt in het geheel niet aan de orde. De buitengewoon nuttige formule voor de inverse van de som van twee matrices komt alleen als speciaal geval in een oefening voor, evenals de bijbehorende determinantvergelijking voor een nog specialer geval.

Een merkwaardige plaats in het boek wordt ingenomen door de matrixdifferentiaal-rekening (MDR). MDR is het belangrijkste onderwerp dat het eigen karakter van dit boek bepaalt, in het bijzonder door de uitgebreide aandacht voor differentiaties. De tweede auteur heeft op het gebied van het introduceren van het gebruik van differentiaal in de multivariate statistiek en econometrie een pioniersrol gespeeld. De delen twee en drie van het boek zijn aan MDR gewijd. In deel vier, over ongelijkheden, wordt MDR vrijwel niet gebruikt, terwijl in deel vijf het gebruik van MDR zich voornamelijk beperkt tot het eenvoudige geval van een differentiaal van een scalaire functie met een matrixargument.

Een algemenere kwestie is de plaats zelve van MDR: het boek kiest in het algemeen, en in hoofdstuk 13 over het lineaire model in het bijzonder, voor deductieve bewijzen, waarbij MDR goede diensten bewijst, ook in gevallen waar korter 'inductieve' bewijzen, niet gebruikmakend van MDR, mogelijk zijn. Wat met preferenceert is een kwestie van smaak.

T.J. Wansbeek.
RU Groningen

P.E. Caines

Linear Stochastic Systems

John Wiley & Sons, New York, 1988, xiv + 875 pag.,
ISBN 0-471-08101-9, £ 52.50.

Dit boek beschrijft de voornaamste theorie en resultaten op het grensgebied van systeem- en regeltheorie, stochastiek en statistiek. Het behandelt voorspelling, realisatie, parameterschatting en adaptieve regeling van stochastische processen in discrete tijd.

Het boek kan worden verdeeld in drie delen, nl. over structuur en voorspelling van lineaire stochastische systemen (hoofdstuk 1-4), identificatie (hoofdstuk 5-10), en regeling (hoofdstuk 11 en 12).

Het eerste deel bevat een beschrijving van klassieke resultaten als de Wold decompositie en spectrale representaties, de werking van lineaire systemen op stochastische processen, ARMAX systemen, en kleinste kwadraten voorspelling, gebaseerd op het Wiener filter of, voor het geval van rationele spectra, op het Kalman filter. Voort wordt een recente geometrische theorie van stochastische realisatie beschreven.

Het tweede deel is gewijd aan systeemidentificatie. Een belangrijk onderscheid is dat tussen exacte systeemmodellen (ESM), gebaseerd op de klassieke vooronderstelling dat de modelklasse een model bevat dat een exacte beschrijving geeft van het waargenomen proces, en approximatieve systeemmodellen (ASM), waarbij het benaderen van een waargenomen proces door een eenvoudig model centraal staat. Er wordt uitgebreid aandacht geschonken aan consistentie, efficiëntie en asymptotische normaliteit voor kleinste kwadraten schatters (voor ASM) en voor identificatie d.m.v. maximale aannemelijkheid schatters (voor ESM). Verder wordt voor ASM de algemene methode van identificatie d.m.v. minimale voorspellingsfouten beschreven. Voorts worden resultaten gepresenteerd over cross-sectie analyse en over causaliteit, terugkoppeling en identificatie van geregelde systemen.

Het derde deel behandelt optimale regeling van stochastische systemen. Voor lineaire stochastische systemen met bekende verdelingsfunctie wordt het centrale resultaat beschreven door het LQG-Filter, waarbij optimale regeling wordt verkregen op een eenvoudige, recursieve wijze, m.b.v. toestandsschatting gegeneerd door het Kalman filter. Tot slot wordt aandacht besteed aan adaptieve regeling, d.w.z. het regelen van ten dele onbekende systemen, i.h.b. minimum variantie regeling van tijdsinvariante en van tijdsvariërende systemen.

Het voorgaande overzicht geeft aan dat dit boek een brede behandeling geeft van theorie en toepassingen van lineaire stochastische systemen. Het boek geeft een sterk samenhangend, goed gemotiveerd, en wiskundig rigoureuze beschrijving van klassieke resultaten op het gebied van filteren, voorspellen, identificatie en regeling, en van meer recente resultaten als realisatietheorie, identificatie met de methode van minimale predictiefouten, en adaptieve regeling. Het boek kan van grote waarde zijn voor onderzoekers op het gebied van systeem- en regelingstheorie, waarschijnlijkheidsrekening en statistiek, tijdreeksanalyse, en econometrie. De meeste resultaten worden geformuleerd onder zeer algemene condities. Hierdoor zijn de bewijzen soms wat langdradig en minder inzichtelijk. Ook met het oog op onderwijs zou het wellicht beter zijn minder algemene resultaten te bewijzen, en een indicatie te geven van mogelijke generalisaties.

Het boek is van hoog niveau en uitnemende kwaliteit. Het biedt een evenwichtige combinatie van de probleemstelling van de ingenieur, gebaseerd op praktische problemen als identificatie en regeling, en van de analyse van de wiskundige, waarbij deze problemen zodanig worden geformuleerd dat ze aansluiten bij bestaande theorie of leiden tot nieuwe theorievorming. Dit laatste is zeker het geval voor de methoden van identificatie en regeling beschreven in dit boek.

Dr. C. Heij
vakgroep Wiskunde
Faculteit der Algemene
Wetenschappen RUL

H.F.M. Crombag, L.J.Th. Van Der Kamp en C.A.J. Vlek (eds)

De Psychologie Voorbij, ontwikkelingen rond Model, Metriek en Methode in de Gedragswetenschappen

Swets & Zeitlinger, Lisse, 1987, 284 pag., ISBN 90-265-0891-3, f 49.50.

Bij het afscheid van Prof. dr J.P. van de Geer als hoogleraar methodenleer, in het bijzonder de datatheorie, aan de sociale faculteit van de R.U. Leiden, is hem in december 1987 een feestbundel aangeboden. Deze bevat zeventien artikelen van oud-leerlingen en vrienden, benevens een ten geleide, een biografie en een bibliografie van de scheidende hoogleraar.

John van de Geer studeerde van 1947-1953 psychologie in Leiden, werd er in 1960 lector en in 1963 hoogleraar. Naast A.D. de Groot in Amsterdam was hij ongeveer de eerste psycholoog in Nederland die zich voluit bezig hield met de methodenleer en de bewerking en modellering van psychologische data; de bundel accentueert nog eens dat veel prominenten op dat vakgebied onder zijn leiding hebben gewerkt. Evenals De Groot is hij ook altijd psycholoog gebleven, al heeft hij vooral in de laatste twintig jaar vooral aan formele modellen voor multivariate analyse gewerkt.

De naam 'Datatheorie' voor zijn facultaire vakgroep is een eerbetoon aan de onlangs overleden Clyde H. Coombs, en weerspiegelt tevens dat Van de Geer de inzichtelijke afbeelding van data en hun structuur nastreefde en weinig op had met de wiskundige statistiek. Evenals zijn leerling De Leeuw gaf hij prioriteit aan het gebruik van matrixrekening, numerieke analyse en meetkunde bij het ontwerpen van geavanceerde vormen van gegevensverwerking. Uw recensent

heeft in een debat met De Leeuw zijn bezwaren tegen de daaruit voortvloeiende nadruk op exploratie ten koste van generalisatie geuit, zie *Statistica Neerlandica* van juni 1988.

Dat neem niet weg dat ik de bundel met veel plezier heb gelezen. Het niveau is goed, het spectrum van onderwerpen weerspiegelt de veelzijdigheid van van de Geer. De bundel bevat natuurlijk weinig statistiek in de nauwe zin des woords, maar veel datatheorie, en daarnaast o.m. cognitieve en mathematische psychologie. Ik zal de artikelen noemen die voor KM-lezers van speciaal belang zouden kunnen zijn.

Bij Wagenaar (Logisch voorwaarts en intelligent terugwaarts: modellen voor het stellen van diagnoses), Vlek & Hendrickx (Statistisch risico versus persoonlijke controle als grondslag voor het beoordelen van veiligheid) en Moraal & Koornstra (Over een risicovolle alledaagse verrichtingen: autorijden en verkeersongelukken) staat de mens als gebrekkige informatieverwerker te kijk. Roskam evalueert de axiomatische meet-theorie; De Leeuw & Mooijaart doen dat met lineaire structurele modellen en Ten Berge met gewichten van principale componenten.

Van der Kamp & de Gruijter presenteren een heldere inleiding 'Item response theorie in de psychometrie', met speciale aandacht voor het Raschmodel. Heiser & Meulman (Afstandsmodellen voor multivariate analyse) evenaren die helderheid in een voortreffelijke inleiding tot een kernthema van de 'Leidse school'. Daarna illustreren Van der Kloot & Van Knippenberg resp. Daalder & Wolters de toepassing van meetkundige afbeeldingen in de sociale psychologie resp. de politicologie. Visser besluit de bundel met een verkenning van de Sociaal Wetenschappelijke Informatica. Deze vijf bijdragen lijken mij voor een breed publiek van KM-lezers toegankelijk.

Daarnaast bevat de bundel werk van Jaspars, Frijda, Michon, Levelt & Flores d'Arcais en Van der Heijden, dat meer over psychologie dan over kwantitatieve methoden gaat.

Ivo W. Molenaar
RU Groningen

Edward J. Dudewicz & Satya N. Mishra

Modern Mathematical Statistics

John Wiley & Sons, New York, 1988, xix + 838 pag.,
ISBN 0-471-81472-5, f 38.50.

In het voorwoord schrijven de auteurs dat de inhoud van het boek de basis is voor een moderne cursus mathematische statistiek. Om dit doel te verwezenlijken presenteren zij een boek van 838 bladzijden. Een cursus van deze omvang zal echter een vol studiejaar 40 weken à 4 uren) moeten beslaan en is met name geschikt voor wiskundige studenten.

De onderwerpen die behandeld worden laten zich ruwweg in drie categorieën opdelen t.w. grondbeginselen van de waarschijnlijkheidsrekening, toetsings- en schattingstheorie en geselecteerde (multivariate) methoden. De eerste categorie (grondbeginselen van de waarschijnlijkheidsrekening) komt qua stof en niveau zeer goed overeen met het boek van Fabius en van Zwet (zie referentie onderaan). Gelet op de bekendheid van dit boek binnen Nederland

lijkt een uitvoerige beschrijving van de onderwerpen in deze boekbespreking overbodig.

De tweede categorie (toetsings- en schattingstheorie) kan het beste worden vergeleken met de corresponderende hoofdstukken uit het boek van Bickel en Doksum (zie referentie onderaan). De onderwerpen die worden behandeld zijn puntschatters (w.o. maximum likelihood schatters, momentenschatters, Cramér-Rao ongelijkheid, voldoendeheid en volledigheid van schatters, zuiverheid), toetsen van hypothesen (w.o. Neyman-Pearson lemma, t-toetsen, Behrens-Fischer probleem, likelihood ratio toetsen, chi-kwadraat toetsen) en betrouwbaarheidsintervallen.

De geselecteerde methoden hebben betrekking op rangorde en selectie procedures, beslissingstheorie, verdelingsvrije methoden, (multiple) regressie en variatie (1-weg en 2-weg) analyse en robuuste statistische procedures (w.o. M-schatters, jackknife schatters en bootstrap schatters).

De theorie wordt op een elementair wiskundig niveau behandeld, veelvuldig afgewisseld met (klassieke) voorbeelden en opgaven (waarvan een deel wordt uitgewerkt). Speciale aandacht hebben de auteurs besteed aan de statistische tabellen van de binomiale, hypergeometrische, Poisson, beta, gamma, chi-kwadraat, normale, F en t verdelingen.

Het moderne van het boek wordt door de auteurs mede gemotiveerd door het opnemen van enkele Fortran programma's en BMDP en SAS routines. Echter gelet op de marginale wijze waarop de computer bij de stof betrokken wordt kan dit argument niet serieus genomen worden. De beoordelaar is van mening dat het moderne van dit boek is dat het een geslaagde combinatie is van de grondbeginselen van de waarschijnlijkheidsrekening en de mathematische statistiek. Hij beveelt het boek dan ook van harte aan.

R.J.M.M. Does
RU Limburg

Referenties:

- J. Fabius en W.R. van Zwet, Grondbegrippen van de waarschijnlijkheidsrekening, MC Syllabus 10, CWI, Amsterdam (1970)
- P.J. Bickel and K.A. Doksum, Mathematical Statistics: Basic Ideas and Selected Topics, Holden-Day, San Francisco (1977).

Sufi M. Nazem

Applied Time Series Analysis for Business and Economic Forecasting

Marcel Dekker, Inc., New York and Basel, 1988, x + 431 pag.,
ISBN 0-8247-7913-4, \$ 107.50.

Volgens de auteur is dit boek geschreven voor studenten in de (bedrijfs-)economische wetenschappen met als doel hen kennis bij te brengen op het terrein van de tijdreeksanalyse en de voorspellingstechnieken. Men zou hieruit kunnen concluderen dat uitvoerig wordt ingegaan op diverse methoden en technieken die in de economische wetenschappen bij het analyseren en voorspellen van tijd-

reeksen worden gehanteerd. Niets is echter minder waar, immers het boek beperkt zich slechts tot de alom bekende enkelvoudige autoregressieve moving average (ARMA) tijdreeksmodellen. Daarbij worden de basisideeën, vastgelegd in het 'klassieke' boek van Box en Jenkins, nog eens 'dunnetjes' overgedaan. De auteur streeft bij dit alles naar een minimale wiskundig/statistische uiteenzetting, hetgeen de duidelijkheid en nauwkeurigheid van bepaalde begrippen niet bevordert.

In de hoofdstukken 1 tot en met 12 komen achtereenvolgens de volgende onderwerpen aan de orde: inleiding in de tijdreeksanalyse; autocorrelaties en partiële autocorrelaties; het MA(1) model; het AR(1) model; het ARMA(1,1) model; enige discussie m.b.t. het schatten van AR en MA modellen; hogere orde MA modellen; hogere orde AR modellen; niet-stationaire tijdreeksen; inleiding seizoen-modellen; het seizoen MA-model; het seizoen AR-model. Ieder van deze hoofdstukken wordt afgesloten met enkele opgaven. Vaak zijn deze opgaven evenwel gespeend van enige praktische betekenis en leiden ze tot recht-toe-recht-aan rekenpartijen die geen verdieping of verduidelijking van de behandelde theorie bij de lezer tot stand brengt. In hoofdstuk 13 wordt onder het kopje 'recente ontwikkelingen en daaraan gerelateerde zaken' stilgestaan bij door de auteur ten onrechte vermeende controversies over de Box-Pierce toets op residuele autocorrelatie. In krap twee pagina's wordt vervolgens iets verteld over meervoudige tijdreeksanalyse hetgeen veel te beperkt is voor dit complexe onderwerp. Hoofdstuk 13 wordt afgesloten met een naar mijn mening overbodige lijst van leveranciers die PC-software leveren voor enkelvoudige tijdreeksanalyse. Onderwerpen die voor economen van belang zijn zoals aggregatie van tijdreeksvariabelen, uitschieters en kalendereffecten in tijdreeksen alsmede causale verbanden tussen tijdreeksen komen hier niet aan de orde. In het laatste hoofdstuk wordt, aan de hand van acht zogenaamde 'real-live' cases, de behandelde theorie nog eens kort toegelicht.

Naar mijn mening is dit boek slechts van beperkte waarde voor de beoogde doelgroep van lezers. Dit komt met name doordat de auteur een receptmatige bespreking geeft van enkelvoudige ARMA tijdreeksmodellen. Alternatieve tijdreeksmodellen en voorspellingstechnieken komen niet aan de orde. Verder is dit boek minder interessant voor economen door het nagenoeg ontbreken van economisch getinte voorbeelden en oefeningen. Ten slotte bevordert het groot aantal typfouten in diverse formules en uitdrukkingen niet erg de leesbaarheid.

Jan G. de Gooijer
Universiteit van Amsterdam

J.P. van de Geer

Analyse van kategorische gegevens

Van Loghum Slaterus, Deventer, 1988, 213 pag., ISBN 90-368-0110-9, f 75.00

Nieuwe analysetechnieken worden doorgaans voorgesteld met een algebraïsche uiteenzetting. Het nadeel van deze aanpak is dat een grote groep van potentiële gebruikers niet wordt bereikt. Een manier om de techniek voor een breder publiek toegankelijk te maken, is de wiskundige aspecten zoveel mogelijk terzijde te schuiven en de lezer met grafische afbeeldingen te introduceren in de nieuwe mogelijkheden.

Deze geïllustreerde aanpak volgt ook Van de Geer in zijn boek over de methoden voor analyse van categorische gegevens volgens het GIFI-systeem.

Het boek bestaat uit twee delen. In deel I wordt de lezer eerst aan de hand van ruimtelijke voorstellingen vertrouwd gemaakt met twee klassieke multivariate analyses, te weten principale componenten analyse en canonische correlatie analyse. Vervolgens geeft Van de Geer een uiteenzetting van de overeenkomsten en verschillen tussen deze methoden en de niet-lineaire multivariate analyse. In dit kader behandelt hij het principe van optimale categorie kwantificatie dat ten grondslag ligt aan de in Leiden ontwikkelde alternating least-squares programma's. Speciale aandacht is er voor de behandeling van ontbrekende gegevens.

In deel II volgt een bespreking van de kwantificatietechnieken volgens het GIFI-systeem, te weten PRIMALS, HOMALS, ANACOR, PRINCALS, CANALS en OVERALS. Met eenvoudige voorbeelden uit de onderzoekspraktijk worden de technieken toegelicht. De analysetechnieken hebben gemeen dat ze een optimale kwantificatie van categorische gegevens opleveren. Door deze kwantificatie worden de correlaties tussen variabelen lineair en maximaal.

Van de Geer heeft bij de keuze van de voorbeelden niet gestreefd naar 'ideale' oplossingen. Terecht heeft hij een te rooskleurige voorstelling van de onderzoekspraktijk vermeden. Bovendien maakt hij de lezer attent op een aantal toevallig- en eigenaardigheden van de computeruitvoer. Toch is het de vraag of onderzoekers praktisch veel voordeel bij het boek zullen hebben. Omdat Van de Geer de weg van de grafische voorstelling bewandelt, moet hij zich noodzakelijk beperken tot eenvoudige rekenvoorbeelden die in de praktijk weinig zullen voorkomen. In dit verband is het ook jammer dat het boek geen overzicht van publikaties bevat waarin de optimale kwantificatietechnieken in onderzoek zijn toegepast.

Van de Geer heeft een boek geschreven waarin hij met heldere formuleringen de lezer introduceert in de mogelijkheden van de GIFI-methoden voor de analyse van categorische gegevens. Ofschoon het boek is geschreven voor onderzoekers, lijkt het mij ook geschikt voor onderwijs aan studenten als aanvulling op de klassieke multivariate analyse.

R. Eisinga
KU Nijmegen

***** CWI-publicaties *****

R. van der Horst, R.D. Gill (eds.)

STATAL: statistical procedures in Algol 60, part 1

CWI, Amsterdam, 1988, xvii + 200 pag., ISBN 90-6196-358-3, f 33.00.

R. van der Horst, R.D. Gill (eds.)

STATAL: statistical procedures in Algol 60, part 2

CWI, Amsterdam, 1988, 198 pag., ISBN 90-6196-359-1, f 30.40.

R. van der Horst, R.D. Gill (eds.)

STATAL: statistical procedures in Algol 60, part 3

CWI, Amsterdam, 1988, 227 pag., ISBN 90-6196-360-5, f 35.40.

Deze drie syllabi uit de CWI reeks (respectievelijk nummer 20, 21 en 22) bevatten de beschrijving van de statistische bibliotheek STATAL.

STATAL, ontwikkeld door medewerkers van het Mathematisch Centrum te Amsterdam, bevat procedures voor statistische berekeningen, geschreven in de taal ALGOL 60.

De broncodes voor de procedures zijn opgenomen in de syllabi, waardoor het voor programmeurs in andere computertalen toch goed mogelijk is om deze procedures te gebruiken als leidraad voor het schrijven van eigen programma's. Alle procedures zijn apart beschreven, telkens is een korte (soms wel zeer korte) verklaring gegeven van doel en werking van de betreffende procedure.

Deel 1 bevat procedures voor berekeningen bij verschillende verdelingen. Verdelingsfuncties, inverse verdelingsfuncties en kansen van discrete verdelingen, waaronder ook de verdeling van de Wilcoxon twee-steekproeven grootheid en de Kendall's grootheid worden hier beschreven.

Voor de continue verdelingen worden ook de verdelingsfuncties, inverse verdelingsfuncties en de kansdichtheden behandeld, waaronder ook van de studentized range, de steekproefcorrelatie coëfficiënt en de Kolmogorov-Smirnov grootheid.

Deel 2 bevat de procedures voor berekeningen van diverse toetsingsgrootheden en overschrijdingskansen van toetsingsgrootheden. Aan bod komen bijvoorbeeld bij de niet-parametrische theorie de Wilcoxon grootheid (één en twee steekproeven), de Cramér-von Mises grootheid, Kendall's tau, Spearman's rho en Smirnov's d. Voor de parametrische methoden worden onder andere vermeld de Student grootheid voor één- en twee steekproeven en voor multivariatie methoden de produkt-moment correlatie matrix en een factor analyse volgens het model van Joreskog (1963).

Verder bevat dit deel procedures voor sorteren en rangordenen van waarnemingen in rijen en kolommen en van rijen en kolommen binnen matrices en berekeningen voor permutaties en combinaties.

In deel 3 worden procedures beschreven voor het berekenen van (pseudo) ascellecte getallen uit diverse discrete en continue verdelingen.

Tevens staan in dit deel de procedures voor tabellen en grafieken, waaronder bij de laatste categorie waarschijnlijkheidsplots voor diverse verdelingen, spreidingsdiagrammen en histogrammen vallen.

Al met al een zeer nuttige verzameling, niet alleen voor de gebruikers van de bibliotheek in ALGOL 60, maar zeker ook voor iedereen met programmerings aspiraties.

Arend Oosterhoorn

F. den Hollander, H. Maassen (eds.)

Mark Kac seminar on probability and physics, syllabus 1985-1987

CWI, Amsterdam, 1988, iv + 162 pag., ISBN 90-6196-350-8, f 25.30.

In deze syllabus (nummer 17) zijn rapporten en voordrachten gebundeld van het 'Mark Kac seminar on probability and physics', welk is gehouden in de periode 1985-1987 in Amsterdam.

De bundel bevat een 28 tal bijdragen, waarvan 12 van Nederlanders.

