KM13 (1984)
pag 191-206

BOEKBESPREKINGEN

Redactie: J.J. Dik

Correspondentie adres: Subfaculteit Wiskunde UvA
Roetersstraat 15
1018 WB Amsterdam
(020) 522-3090

G.J.G. Upton

The Analysis of Cross-tabulated Data

John Wiley & Sons, 1978, 148p, £10.75

In het voorwoord schrijft Upton dat zijn boek in de eerste plaats is bestemd voor onderzoekers in de sociale wetenschappen die hun data hebben verzameld en ze vervolgens moeten analyseren. Na een korte inleiding in enkele statistische begrippen, zoals het toetsen van hypothesen, behandelt Upton in hoofdstuk 2 associatiematen en toetsen op onafhankelijkheid voor een 2×2 - tabel en in hoofdstuk 3 voor een algemene twee-dimensionale frequentietabel. In hoofdstuk 4 worden diverse vormen van associatie en onafhankelijkheid gedefinieerd voor tabellen met een hogere dimensie. De overige hoofdstukken gaan alle over het log-lineaire model, waarbij het model geleidelijk wordt uitgebouwd. Hoofdstuk 5 beperkt zich tot de 2×2 - tabel. Hoofdstukken 6 en 7 behandelen resp. het verzadigde en het onverzadigde model voor meer-dimansionale tabellen; de voorbeelden beperken zich ook hier tot dichotome variabelen. De overige drie hoofdstukken behandelen hoofdzakelijk specifieke problemen en toepassingsmogelijkheden zoals interpretatie bij polytome variabelen, de problematiek van lege cellen, onafhankelijkheidstoetsen voor een deel van de tabel (bijvoorbeeld exclusief de hoofddiagonaal), symmetrie-toetsen, het gebruik van log-lineaire modellen bij panel-studies en het latente-klasmodel. Het boek is sterk gebaseerd op het werk van Goodman, hetgeen onder meer blijkt uit 17 literatuurverwijzingen. Vanzelfsprekend ontbreken verwijzingen naar diens artikelen van na 1977.

Mijn voornaamste kritiek betreft de titel van het boek. Upton's boek gaat vrijwel geheel over log-lineaire analyse en zou daarom beter "Introduction to log-linear analysis" hebben kunnen heten. De huidige titel is te pretentieus; het inlassen van enkele hoofdstukken over associatiematen doet hier weinig aan af. Er zijn teveel technieken voor het analyseren van kruistabellen die ongenoemd blijven. Geen aandacht wordt bijvoorbeeld besteed aan allerlei schaaltechnieken die tot doel hebben relaties tussen discrete variabelen te optimaliseren via het toekennen van schaalwaarden (kwantificaties) aan de categorieën. Een aantal behandelde voorbeelden met polytome variabelen zou met deze technieken gemakkelijker zijn te analyseren. Reeds ruim voor 1978 bestond op dit ge-

bied een vrij aanzienlijke literatuur, bijvoorbeeld het werk van Lancaster dat in het geheel niet wordt genoemd. Voor meer recente literatuur hierover, zie bijv. Gifi (1981). Een andere omissie is dat vrijwel alleen in de eerste hoofdstukken wordt ingegaan op ordinale variabelen, waar enkele associatiematen voor twee ordinale variabelen worden behandeld, zoals Kendall's tau. Zodra meer dan twee variabelen tegelijk worden beschouwd, wordt vrijwel geen aandacht meer aan ordinaliteit besteed. Ook aan classificatie-technieken wordt nauwelijks aandacht besteed. Andere hier niet genoemde technieken zijn bijv. te vinden in Andersen (1980) en Goldstein & Dillon (1978). Het is natuurlijk het goed recht van Upton om log-lineaire analyse de techniek bij uitstek te vinden om kwalitatieve data te bestuderen, het gebruik ervan in plaats van andere technieken wordt echter in het geheel niet beargumenteerd. Integendeel, er wordt gedaan alsof log-lineaire analyse de enig bestaande techniek is voor het analyseren van tabellen met meer dan twee dimensies.

De hoofdmoot van het boek, het log-lineaire model, wordt op een prettige manier behandeld. Veel voorbeelden worden gegeven die het de lezer vergemakkelijken om in te zien hoe de techniek werkt en hoe deze te interpreteren is. Bij een enkel voorbeeld wordt iets te veel gekeken naar de waarden van de toetsingsgroothedenen iets te weinig naar de inhoudelijke kant.

De hoofdstukken over associatie in twee-dimensionale tabellen zijn veel minder duidelijk geschreven. Onder andere de volgorde waarin de associatiematen worden behandeld, verdient verbetering. Zo heeft § 2.7 de titel "The odds ratio" (= kruisproduct-verhouding), terwijl in § 2.3 de theorie hiervoor reeds is behandeld, en de tussenliggende paragrafen andere onderwerpen behandelen. Naast wat slordigheden in deze eerste hoofdstukken wordt de meest ernstige fout gemaakt bij het opsplitsen van tabellen in subtabellen (§ 3.3). Upton beweert dat hierbij zowel Pearson's chi-kwadraat als de log-likelihood-ratio (G-kwadraat) kan worden gepartitioneerd en toont dit met een voorbeeld aan. Dit voorbeeld bevat echter een rekenfout bij de berekening van Pearson's chi-kwadraat. Het voorbeeld zou daarom juist moeten tonen dat Pearson's chi-kwadraat zich niet laat partitioneren, althans niet in het huidige geval.

Ondanks de genoemde kritiek vind ik het boek geschikt voor de onderzoeker die in het praktisch gebruik van log-lineaire analyse is geïnteresseerd. Veel referenties worden gegeven voor hen die wat meer theoretische kennis willen hebben. Is de lezer vertrouwd met onafhankelijkheidstoetsen voor twee-dimensionale tabellen dan kan hij desgewenst met hoofdstuk 4 beginnen om sneller aan het log-lineaire model toe te komen. De prijs van het boek is overigens hoog, vergeleken met bijv. Fienberg (1980) dat voor een groot deel dezelfde onderwerpen behandelt.

Literatuur

ANDERSEN, E.B. (1980) Discrete statistical models with social science applications (North Holland, Amsterdam).

FIENBERG, S.E. (1980) The analysis of cross-tabulated data, 2nd ed. (MIT Press, Cambridge, Massachussets).

GIFI, A. (1981) Non-linear multivariate analysis (Department of Data Theory, Leyden University, Leiden).

GOLDSTEIN, M. & W.R. DILLON, (1978) Discrete discriminant analysis (Wiley, New York).

A.Z. Israels, C.B.S.

J.M. Keizer

Economische Wiskunde

Stenfert Kroese, Leiden, 3e herziene druk 1978, xii + 278p, f52,--

De schrijver is docent wiskunde en statistiek bij de H.E.S. (= HEAO) in Rotterdam. Het boek is geschreven "voor diegenen, die menen van wiskunde niets te begrijpen, maar dat vak voor hun economische studie nodig hebben". Daar de studentenpopulatie van het HEAO zowel MEAO-abituriënten met minimale wiskundekennis als VWO-ers met wiskunde II in het examenpakket bevat is het geen geringe opgave voor deze populatie iets te schrijven. Duidelijk is dat de wiskunde (algebra) praktisch van de grond af opgebouwd moet worden. Het boek is niet geschreven voor zelfstudie.

De onderwerpen zijn: rechte lijnen, parabolen, orthogonale en scheve hyperbool; rekenkundige en meetkundige rijen; differentiëren, functies van twee variabelen, (totale) differentialen; lineair programmeren (grafische oplossing maar ook de simplex methode) en logaritmen. Dit alles belicht vanuit de economie met aparte hoofdstukken over elasticiteiten, indifferentiecurven, producentengedrag en macro-economie. Elk hoofdstuk bevat een aantal (meestal economisch getinte) vraagstukken met achterin de antwoorden.

De schrijver komt lof toe voor de poging om de wiskunde aantrekkelijk te maken door de wiskunde vanuit de "economische praktijk" op te bouwen. Schrijver heeft de ervaring dat gebruikers zonder wiskunde vooropleiding niet (al te) veel moeite met het boek hebben.

Ik ben niet erg enthousiast over het boek als leerboek. Ik volsta met enige opmerkingen:

Het eerste hoofdstuk - rechte lijnen - wordt behandeld vanuit kostenfuncties. De wiskunde (en dit geldt voor het hele boek) wordt terloops behandeld, hier: lijn door twee punten en snijpunt van twee lijnen. Het begrip richtingscoëfficiënt krijgt ruime aandacht, maar of het begrip tangens verduidelijkend werkt

betwijfel ik. De algemene formule van de rechte lijn wordt niet gegeven. Toch is de opzet van het eerste hoofdstuk aardig. Veel grafieken (het boek bevat er 128) zijn onoverzichtelijk (veel te veel lijnen en slechte vermelding van getallen bij de assen).

Naast aardige economische opgaven zijn er veel te weinig oefenopgaven. Onderwerpen als de techniek van een tweedemacht wortel trekken en de logaritmen tafel maken het boek gedateerd.

R.J. Köster.

H. Späth

Cluster analysis algorithms for data reduction and classification of objects

John Wiley & Sons, New York, 1980, 226p, £16.50

Dit boekje bevat de Fortran-teksten van 37 door de auteur vervaardigde programma's en subroutines op het gebied van de cluster-analyse. Deze programma's bestrijken de meest gebruikelijk afstands en similariteitsmaten, alsmede de hierarchische en niet hierarchische cluster-analyse. In de begeleidende tekst wordt de werking van de programma's beschreven, en wordt aangegeven in welke omstandigheden zij kunnen worden toegepast. Bovendien worden bij de meeste programma's de resultaten van één of meer toepassingen weergegeven en kort besproken. De auteur verwijst veelvuldig naar literatuur waarin vergelijkbare algorithmen worden beschreven. Merkwaardig genoeg echter ontbreken verwijzingen naar het alom aanwezige en veel gebruikte programma-pakket CLUSTAN van D. Wilshart.
De kwaliteit van Späth's programma's kan ik niet beoordelen. Hoewel in de tekst ook enige aandacht aan de methodiek van cluster-analyse wordt besteed lijkt die behandeling mij te zwak om als leidraad te kunnen dienen bij de keuze van een methode. Meer dan eens laat de auteur het voordeel van een kortere rekentijd prevaleren boven de kwaliteit van het resultaat. Hij had die keuze beter aan de lezer kunnen overlaten. Feitelijk onjuist zijn opmerkingen over het effect van normalisatie van objecten (p.20) en schaalinvariantie in het univariate geval (p.60) en de definitie van "hierarchical cluster algorithms" (p.155).
Jammer dat de figuren niet van onderschriften zijn voorzien. Niettemin is het boekje aan te bevelen bij wie geinteresseerd is in programmatuur voor cluster-analyse.

J.J. de Gruijter.

Gerald A. Fleischer (editor)

Contingency Table Analysis for Road Safety Studies

Nato ASI - series, Sijthoff & Noordhoff, Alphen aan de Rijn, 1981, 300p,
ISBN 90-286-0960-1, f65,--

Op het eerste gezicht lijkt het wat vreemd om een boek te recenceren waaraan men zelf heeft meegewerkt. Ik heb het verzoek dan ook iets ruimer opgevat en zal daarom eerst ingaan op de achtergronden van de NATO-ASI die in 1979 is gehouden.

Bij verkeersveiligheidsonderzoek wordt meestal gebruik gemaakt van ongevallen studies. Kenmerken van ongevallen worden gerelateerd aan kenmerken van de weg, de verkeersdeelnemer, het voertuig en de omstandigheden. De analyse van dit soort gegevens is in feite vaak gebaseerd op meerdimensionale kruistabellen, waarin de ongevallen naar een aantal kenmerken zijn uitgesplitst. Dit verklaart waarom er vanaf het begin belangstelling was voor de ontwikkelingen op het gebied van de log-lineaire analyse.

De ASI-bijeenkomst kan worden gezien als een confrontatie tussen onderzoekers op het gebied van de verkeersveiligheid en statistici die zich bezig hielden met log-lineaire analyse. Deze confrontatie is voortgekomen uit een aantal bestaande kontakten, zowel in Europa als in de Verenigde Staten. De proceedings van deze studiegroep zijn neergelegd in het hier beschouwde boek, dat bestaat uit een meer theoretisch deel en een deel waarin toepassingen van diverse technieken op verkeersveiligheidsgegevens zijn beschreven.

Het eerste deel begint met een uiteenzetting van het inmiddels klassieke log-lineaire analyse model door E.B. Andersen. Andersen had zich eerder zelf bezig gehouden met het vinden van een beschrijving van gewogen celaantallen ("unequal cell rates") ten behoeve van veiligheidsanalyses. Hij gaat uit van het verzadigde model en laat zien dat het log-lineaire model in feite een "herparametrisering" geeft. Vervolgens beschrijft hij de ML-schatters en hun betrouwbaarheid, het toetsen van hypotheses, de modelvorming bij twee-weg en drieweg tabellen en de mogelijkheid om de totale Chi-kwadraat op te splitsen in onafhankelijk te interpreteren delen.

Dan volgt een bijdrage van mijzelf, waarin ik getracht heb de diverse elementen van de recente ontwikkelingen die van belang zijn voor de analyse van kruistabellen te beschrijven, zoals de generalisering naar meerdimensionale tabellen, de partitionering van interactietermen in relatie tot de klassen binnen de kenmerken, de grotere nadruk op het schatten van parameters en het aangeven van de betrouwbaarheid i.p.v. uitsluitend te toetsen met een totale Chi-kwadraatwaarde en de mogelijkheid om de celaantallen te wegen alvorens ze te analyseren. Tenslotte wordt ingegaan op de redelijkheid van diverse assumpties voor bepaalde verkeersveiligheidsgegevens (verdelingsassumpties, onafhankelijk-

heidsassumptie).

Dan volgen twee papers waarin specifieke onderzoeksproblemen op het gebied van de verkeersveiligheid worden besproken. Een paper van Lassarre, waarin uitgaande van de Poisson-verdeling en diverse mogelijkheden van conditionering test-statistics worden afgeleid voor kleine steekproef aantallen en waarin voor grote steekproeven wordt aangegeven hoe dit uitmondt in de klassieke Chi-kwadraat toets. De voorbeelden zijn ontleend aan een veel voorkomende situatie waarin het effect van veiligheids maatregelen dient te worden vastgesteld. De relatie met het log-lineaire model is evident maar wordt niet expliciet gemaakt. In een paper van Bui Quoc die uitgaat van Shannon's entropie-maat, wordt een procedure beschreven om uit een aantal random variabelen die variabele(n) te kiezen die de meeste informatie levert. De methode is door hem toegepast om te komen tot een verantwoorde datareduktie bij een veelheid van kenmerken. De methode kan worden gezien als een eerste stap, waarin bepaald wordt welke meerdimensionale tabellen het meest relevant lijken voor nadere analyse.

Daarna volgt een toepassing van de tot dan nog niet behandelde MDI-techniek door Bancroft en Riemersma (van IZF-Soesterberg). De toepassing, getiteld "Measuring Exposure" (blijkens de conclusie is de titel "measuring accident-proneness" correcter), beschrijft een heranalyse van literatuurgegevens m.b.v. een log-lineaire analyse. Een fit met uitsluitend de expositie faktor in het model geeft een redelijke oplossing en wordt niet signifikant verbeterd door toevoeging van de proneness parameters. De MDI-methode levert ook een betere beschrijving met dezelfde parameters als de oorspronkelijke analyse.

In Gokhale's (spreek uit: gókelies) paper wordt ingegaan op de minder bekende log-lineaire analyse variant, die afkomstig is van Kullback, de Minimum Discrimination Information (MDI). De discriminatie informatie wordt gedefinieerd als:

$$I(p:\pi) = \sum_{\omega} p(\omega) \ln \frac{p(\omega)}{\pi(\omega)},$$

waarbij $\omega \in \Omega$, Ω de set van elementen waarover het probleem gedefinieerd is (hier de cellen van de kruistabel) en $p(\omega)$ en $\pi(\omega)$ twee waarschijnlijkheids verdelingen. $I(p:\pi)$ geeft dan de gelijkenis tussen p en π weer, waarbij π een vaste verdeling is en p een aantal constraints heeft.

De constraints, behorende bij elke hypothese, worden verdeeld in ICP's (Internal Constraints Problem) en ECP's (External C.P.). Tot de eerste groep behoren de diverse marginal constraints, tot de laatste groep de constraints die een gekozen model definieren. Een designmatrix C bevat het totaal van de constraints voor een bepaalde hypothese. $I(p:\pi)$ geeft dan de MDI-oplossing voor p die "het meest lijkt op" de vast gekozen π. De statistic $2I(p:\pi)$ is Chi-kwadraat verdeeld. Hierarchische modellen kunnen worden getoetst door decompositie:

$$2I(p:\pi) = 2I(p:x) + 2I(x:\pi).$$

Deze decompositie is in het eerder genoemde voorbeeld gebruikt om expositie en proneness van elkaar te onderscheiden.

Keegel geeft daarna een beschrijving van de procedure waarbij voor een gegeven set van constraints de optimale MDI-oplossing kan worden gevonden. Nolan geeft vervolgens een beschrijving van het MDI-programma 'CONTAB', geschreven in PL/I.

Bij de toepassingen vinden we een bijdrage van Lassarre, met een voorbeeld van toepassing van "analyse des correspondences", als tegenhanger van de log-lineaire kruistabellen analyse. De techniek wordt door hem gebruikt als exploratieve techniek, wanneer er veel kenmerken zijn. Dan een voorbeeld van MDI-analyse toegepast op fietsongevallen in Californië. Het laat de mogelijkheden en beperkingen zien van log-lineaire analyse op databestanden met veel kenmerken.

Daarop volgen drie minder op de analyse en meer op de probleemstelling gerichte onderzoeksverslagen op het gebied van de verkeersveiligheid.

Het geheel overziende constateer ik dat het boek een aardige hoeveelheid informatie bevat over praktische data-analyse, zowel voor data-analisten als voor veiligheidsonderzoekers. De theorie van de data-analyse staat hierbij op de achtergrond. Het bevat niet veel informatie die niet al ergens anders te vinden is en is nogal eenzijdig gericht op de MDI-techniek. Voor de confrontatie tussen theorie en praktijk is dit geen echt bezwaar. Voor data-analisten die geinteresseerd zijn in het oplossen van problemen van onderzoekers en onderzoeksgegevens niet uitsluitend gebruiken als voorbeeld van de eigen analysetechniek, is het misschien aardig dit boek eens te bekijken. Voor de deelnemers aan het ASI was dit aspekt in elk geval een positieve ervaring.

S. Oppe, S.W.O.V.

Ross L. Prentice & Alice S. Whittemore (editors)
Environmental Epidemiology: Risk Assessment
Proceedings of a conference sponsored by SIAM Institute for mathematics and society, and supproted by the Department of Energy.
Siam, Philadelphia, 1982, ix + 299p, £19.00

Het boek is een verslag van een symposium dat gehouden werd in het jaar van uitgave. Er staan 15 lezingen/artikelen in, verdeeld over 5 secties:

1. Risk assessment among geographically exposed populations. Bijvoorbeeld: de gevolgen van incidentele blootstelling aan radioactieve straling (atoombom) of aan een op grote schaal rondgesproeid pesticide (malathion).
2. Risk assessment among occupationally exposed populations. De voorbeelden

hier zijn ontleend aan uranium mijnwerkers, arbeiders in de asbest-, koper- en kernindustrie.

3. Risk assessment among populations screened for disease. Deze 'populaties' betreffen vrouwen betrokken bij preventief onderzoek naar borstkanker.
4. Non-standard data sources for risk assessment.
5. Data analytic methods for risk assessment.

In de laatste twee secties wordt geprobeerd bestaande werkwijzen te verbeteren door uit te gaan van gegevens zoals zich die in de praktijk nu eenmaal zonder tevoren geraadpleegde statisticus voor doen, respectievelijk door modellen te modificeren.

Al doende wordt een goed inzicht gegeven in de moeilijkheden die zich voordoen bij bevolkingsonderzoek. Waarom er nog altijd discussie over de vraag of lage doses radioactiviteit gevaarlijk zijn of niet, waarom toch niet zeker weten of het in Vlagtwedde gezonder is voor CARA-patiënten dan in het Nieuwe Waterweggebied.

Menig artikel bevat een flinke hoeveelheid 'understatements':

"Naturally some rather strong assumptions are necessary ..." (p 135).

"One encounters challenging difficulties when analyzing occupational cohort data..." (p 65)

"Unfortunately the errors in exposure estimation...are large enough to distort seriously..." (p 78).

Een helder overzicht van de problemen waar de onderzoeker van dosis-effect relaties mee kampt geeft het artikel over de besproeing van ongeveer een miljoen mensen met malathion bij een kampagne om het Californische fruit te vrijwaren voor een vreemd vliegje. Er werd gekeken naar het effect op de reproductie, maar er doen zich zeven moeilijkheden voor. De bespreking daarvan leidt tot de conclusie dat dit soort onderzoek als het maar enigszins mogelijk is vermeden dient te worden. Daarbij wordt de kanttekening geplaatst dat als de onderzoeker het niet weet, de politiek, het beleid de beslissing neemt om door te gaan met de mogelijk gevaarlijke stof.

Het onderzoek naar de gevolgen op het milieu en de mens van lage en wisselende concentraties, die vaak slecht gemeten zijn, betreft een uitdaging. Zoals het voorwoord zegt: "...one needs sensitive, robust and sophisticated methods..."

A.A.J. Jacobs

K.L. Chung (editor)
Pao-Lu Hsu:Collected Papers
Springer, Berlijn, 1982, DM 128,--

Pao-Lu Hsu (1910-1970) was een van de meest vooraanstaande statistici van deze

eeuw. Onder redaktie van K.L. Chung zijn in deze uitgave alle 40 artikelen die Hsu geschreven heeft bijeengebracht. Een aantal ervan, geschreven na de terugkeer van Hsu naar China in 1947, zijn nu voor het eerst beschikbaar voor Westerse lezers. Zeven van deze artikelen zijn speciaal voor deze uitgave uit het Chinees vertaald. In onderstaande inhoudsopgave zijn ze van een "*" voorzien:

- Permissions
- Pao-Lu Hsu 1910-1970, *by* T.W. Anderson, K.L.Chung and E.L. Lehmann
- Hsu's Work on Inference, *by* E.L. Lehmann
- Hsu's Work in Multivariate Analysis, *by* T.W. Anderson
- Hsu's Work in Probability, *by* K.L. Chung
- In Memory of Professor Pao-Lu Hsu, *by* Jiang Ze-han and Duan Xue-fu
- On the Limit of a Sequence of Point Sets
- A Note on the Indices and Numbers of Nondegenerate Critical Points of Biharmonic Functions (*with* Kiang Tsai-Han)
- Contribution to the Theory of "Student's t-Test as Applied to the Problem of Two Samples
- On the Best Unbiassed Quadratic Estimate of the Variance
- Notes on Hotelling's Generalized T
- On the Distribution of Roots of Certain Determinantal Equations
- A New Proof of the Joint Product Moment Distribution
- On n-Fold Iterated Limits
- An Algebraic Derivation of the Distribution of Rectangular Coordinates
- On Generalized Analysis of Variance
- On the Limiting Distribution of Roots of a Determinantal Equation
- On the Limiting Distribution of the Canonical Correlations
- Analysis of Variance from the Power Function Standpoint
- Canonical Reduction of the General Regression Problem
- On the Problem of Rank and the Limiting Distribution of Fisher's Test Function
- The Limiting Distribution of a General Class of Statistics
- Some simple Facts About the Separation of Degrees of Freedom in Factorial Experiments
- The Approximate Distribution of the Mean and Variance of a Sample of Independent Variables
- On the Approximate Distribution of Ratios
- On the Power Functions of the E^2-Test and the T^2-Test
- On the Factorization of Pseudo-Orthogonal Matrices
- Sur un Théorème de Probabilités Dénombrables (*with* K.L. Chung)
- On the Asymptotic Distributions of Certain Statistics Used in Testing the Independence Between Successive Observations from a Normal Population
- Complete Convergence and the Law of Large Numbers (*with* Herbert Robbins)
- A General Weak Limit Theorem for Independent Distributions
- The Limiting Distribution of Functions of Sample Means and Application to Testing Hypotheses
- Absolute Moments and Characteristic Function
- A Lemma on the Coefficient of Reduction of a Sum of Independent Variates
- On Symmetric, Orthogonal and Skew-Symmetric Matrices
* On Characteristic Functions which Coincide in a Neighborhood of Zero
* On a Kind of Transformations of Matrices
* On a Kind of Transformations of Matrix Pairs
* Simultaneous Transformation of a Hermitian Matrix and a Symmetric or Skew-Symmetric Matrix
- The Absolute Continuity of the Distribution Functions in the Class L
* The Differentiability of the Probability Transition Function of a Purely Discontinuous Stationary Markov Process on the Euclidean Space
- An Association Scheme $M_3(6)$ Which is Not an L_3-Scheme
- The Limiting Distributions of Order Statistics (*by* Ban Cheng)
* Partially Balanced Incomplete Block Designs (*by* Ban Cheng)
- On the Coincidence Property of Stochastic Matrices (*with* Cheng Jia-Ding and Zheng Zhong-guo)

* BIB Matrices, Simple Codes and Orthogonal Codes
- Appendix: The Jacobians of Certain Matrix Transformations Useful in Multivariate Analysis (*by* Walter L. Deemer and Ingram Olkin)

Zoals uit deze opgave blijkt, worden de artikelen van Hsu voorafgegaan door een herdenkingsartikel geschreven door E.L. Lehmann, T.W. Anderson en K.L. Chung, vergezeld van een korte bespreking door elk van hen van één van de drie gebieden waarop Hsu bijdragen geleverd heeft. Deze besprekingen, overigens ook terug te vinden in Annals of Statistics 7, 467 - 483 (1979), zijn bijzonder plezierig als leidraad voor de bestudering van de rest van het boek. Storend is alleen dat de verwijzingen niet kloppen. In het oorspronkelijke herdenkingsartikel in de Annals was een genummerde lijst van 37 artikelen opgenomen, waarnaar in de drie daarop volgende besprekingen verwezen werd door vermelding van het betreffende nummer. In het nu uitgegeven boek is, zoals in een voetnoot wordt opgemerkt, nog een drietal artikelen in de lijst verwerkt en de nummering is dus aangepast. De in de drie besprekingen gebruikte nummers zijn echter hetzelfde gebleven, zodat eigenlijk altijd naar een verkeerd artikel verwezen wordt. Soms is dat wel vermakelijk zoals wanneer Chung in zijn stukje opmerkt dat artikel [30] een aparte plaats inneemt in het werk van Hsu op het gebied van de kansrekening omdat het niets te maken heeft met karakteristieke funkties. Volgens de lijst echter is artikel [30] "On characteristic functions which coincide in a neighborhood of zero"!

Lehmann behandelt de bijdragen van Hsu op het gebied van de "inference". Deze bijdragen zijn geinspireerd door het verblijf van Hsu op University College in Londen van 1936 - 1940 en betreffen o.a. het Behrens-Fisher probleem en de variantieschatting in het Gauss-Markov model. Het gaat hierbij om "small-sample inference", dus exacte, niet-asymptotische analyses. Een van de gevolgen van het werk van Hsu op dit terrein was het invoeren van het begrip "volledigheid" door Lehmann en Scheffé.

Anderson bespreekt het werk op het gebied van de multivariate analyse. Hsu leverde hier ondermeer bijdragen aan het vinden van de verdelingen van de steekproef covariantie matrix en van de wortels van bepaalde determinantvergelijkingen.

Chung tenslotte beschouwt de bijdragen van Hsu aan de kansrekening. Deze worden gekenmerkt door een virtuoos gebruik van karakteristieke funkties. Een mooi voorbeeld is het artikel (uit 1945!) waarin Hsu Berry-Esseen grenzen en asymptotische ontwikkelingen behandelt voor zowel het steekproefgemiddelde als de steekproefvariantie.

Samenvattend kan gezegd worden dat het boek een fraai beeld geeft van de wijze waarop allerlei belangrijke ontwikkelingen in de statistiek op gang gebracht zijn. Verder is het nuttig dat dit waardevolle materiaal nu goed toegankelijk gemaakt is en niet langer over het hoofd gezien kan worden, zoals volgens Chung

bijv. met bovengenoemd artikel over asymptotische ontwikkelingen gebeurt is in een recent verschenen standaard boek over dit onderwerp.

W. Albers, T.H.T.

J.T. Buchanan
Discrete and Dynamic Decision Analysis
John Wiley & Sons, New York, 1982, £7.90

Dit boek behandelt zowel beslissingsanalyse ("statistical decision theory and sequential decisionmaking") als dynamische programmering. Het is geschreven voor beginnende studenten in verschillende vakgebieden, zoals wiskunde en operations research, maar ook economie en bedrijfskunde. Met het oog op die studenten heeft de schrijver afgezien van het gebruik van analyse en zich goeddeels beperkt tot discrete methoden.

De leerstof wordt behandeld aan de hand van een kleine verzameling voorbeelden, die in de loop van het boek ingewikkelder gemaakt worden. Deze aanpak heeft ongetwijfeld didactische voordelen, maar brengt ook nadelen met zich mee: bij het opzoeken van een oplosmethode moet voor het bijbehorende voorbeeld nogal eens door het boek teruggebladerd worden.

Op het gevaar af van luiheid beticht te worden geef ik de titels van de belangrijkste hoofdstukken onvertaald, omdat zo de beste indruk van de inhoud van het boek verkregen wordt:

Uncertainty and its Measurement
Criteria for Decision
Decision Trees
Markov Chains
Utility and its Measurement
Decision Problems with Information
Sequential Decision Problems
Infinite Stage Markovian Decision Processes

Ieder hoofdstuk is opgebouwd volgens een vrijwel vast patroon: een inleiding gevolgd door de behandeling van de leerstof, afgesloten met een sectie "Further Reading" en opgaven. Helaas worden noch antwoorden noch uitwerkingen gegeven.

Gezien het relatief geringe aantal leerboeken op het gebied van dynamische programmering en aanverwante onderwerpen, is dit boek een welkome aanvulling.

A. Volgenant, U.v.A. (Act. & Econ.)

A.S.C. Ehrenberg
A Primer in Data Reduction
John Wiley & Sons, New York, 1982, £6.95

Het betreft hier een boek dat bedoeld is als allereerste kennismaking met de statistiek, en duidelijk voor niet-wiskunde studenten bestemd is.

Daar er in dit genre al het nodige gepubliceerd is, lijkt de vraag naar het waarom gewettigd. De schrijver zelf zegt hierover:

"It differs from most other introductory texts in two ways. First it covers a wider range of relevant issues. In particular, it not only describes traditional statistical methods but also notes their limitations and discusses alternatives. Secondly, students have found it easier to understand: it avoids most mathematics but without oversimplifying the underlying concepts, it explains more and it is briefer."

Nadere bestudering bracht mij tot de conclusie dat de schrijver in zijn opzet toch niet is geslaagd. Weliswaar komen veel begrippen ter sprake, maar ze worden niet of niet scherp gedefinieerd (bijv. kurtosis; een maat voor scheefheid ontbreekt geheel). Veel storender is echter nog het ontbreken van een behoorlijke inleiding in de kansrekening. Zo zal men vergeefs zoeken naar de termen "random variable" en "expectation". Ook het feit dat alle resultaten met veel tekst, maar zonder enige vorm van bewijs meegedeeld worden, moet de meer geinteresseerde lezer toch gaan irriteren en dat kan nooit de bedoeling zijn. Voorts draagt de zeer vage en summiere bespreking van technieken als bijv. componenten analyse en multidimensional scaling nauwelijks bij tot meer inzicht en had wellicht beter vervangen kunnen worden door een wat grondiger behandeling van meer elementaire zaken.

De bijgevoegde tabellen, die het boek moeten completeren, hadden wel wat meer aandacht mogen krijgen. Slechts de meest gebruikte kwantielen zijn er in te vinden, terwijl in een tijd dat practisch iedereen over een zakrekenmachine kan beschikken, een tabel van decadische logaritmen en vierkantswortels van samen één pagina mij een volstrekt overbodige toevoeging lijkt.

Al met al waag ik het te voorspellen dat dit boek geen standaard-werk zal worden. Voor gebruik als studieboek op universiteiten en hogescholen lijkt het niveau mij aan de lage kant, mede gezien de uiterst gering veronderstelde voorkennis. Veel meer dan wat encyclopedische kennis zonder begrip kan men uit dit boek niet putten, ondanks de redelijke verzameling uitgewerkte vraagstukken.

H.H. Tigelaar, K.H.T.

M.L. Davidson

Multidimensional Scaling

John Wiley & Sons, 1983, xiv + 242p, £24,65

MDS is de verzamelnaam voor een aantal technieken om bij gegeven afstanden (in letterlijke of overdrachtelijke zin) tussen een aantal objecten, deze zodanig in een (meestal Euclidische) ruimte te representeren dat de afstanden overeenstemmen (metrische MDS) of in rangorde overeenstemmen (nonmetrische

MDS). Vooral inde laatste twintig jaar heeft dit geleid tot analyses van onder meer gelijkenis-oordelen en voorkeursuitspraken binnen de gedragswetenschappen en het marktonderzoek, maar ook in vele andere vakgebieden.

De ontwerpers van iteratieve algorithmen en computerprogramma's voor MDS (in Nederland vooral Roskam en De Leeuw) hebben tot op heden veel gespecialiseerde tijdschriftartikelen en gebruikershandleidingen geproduceerd en weinig leerboeken. Het hier besproken boek van Davidson is daarom erg welkom. Het staat qua niveau tussen het wel erg elementaire boekje van Kruskal en Wish (1978, Sage Publ.) enerzijds en de klassieke delen van Romney, Shepard en Nerlove (1972, Seminar Press) of Davies en Coxon (1982, Heinemann) anderzijds. Het meest verwant zijn nog Schiffman, Reynolds en Young (1981, Academic Press), veel meer op specifieke programma's gericht en volgens mij toch meer een naslagwerk dan een leerboek, al heet het "Introduction to MDS", en Coxon (1982, Heinemann), even "conceptual and applications oriented", met meer details en meer amusement, maar mogelijk daardoor iets minder duidelijkheid over de hoofdzaken.

De negen hoofdstukken heten achtereenvolgens 1. Introduction; 2. Matrix Concepts; 3. Proximity Measures (met tevens aanwijzingen voor proefopzetten); 4. Torgerson's Metric Group Method; 5. Nonmetric Group Solutions; 6. Individual Differences Models; 7. Preference Models; 8. Examining Prior Hypotheses with MDS; 9. MDS and its Alternatives (vergelijking met het Thurstone-Chaves schaalmodel, met hierarchische clusteranalyse en met factoranalyse). Elk hoofdstuk heeft een aantal opgaven (met antwoorden). Een literatuurlijst van bijna 200 titels, een uitvoerige lijst van gebruikte symbolen en een korte auteurs- en onderwerpindex vullen de rest van de 242 pagina's.

De auteur verwacht bij de lezer enige statistische basiskennis, tot en met multiple regressie, maar gebruikt geen differentiaalrekening. In de dertig pagina's over matrices wordt vooral rotatie uitvoerig uitgelegd; eigenwaarden of orthogonalisatie zijn voor de behandelde stof niet vereist. Ik vind dat de aldus omschreven doelgroep een heldere inleiding voorgeschoteld krijgt, waarin de toepassingen en de interpretatie van begrippen terecht prevaleren boven de specifieke algorithmen en programma's.

Ik vind persoonlijk dat op pp. 77 en 200 wel erg koel wordt gereageerd op de modellen van Ramsey die een statistische foutentheorie en meest aannemelijke schattingen bieden. Ook mis ik in hoofdstuk 8 een verwijzing naar de simulaties van Spence en Graef die een soort ijkingspunt voor de stressmaat opleveren en daarmee voor het vereiste aantal dimensies. Maar dit zijn details die weinig afdoen aan mijn waardering voor het boek.

I.W. Molenaar, R.U.G.

C.J. Verhoeven

Techniques in Corporate Manpower Planning - methods and applications

Kluwer/Nijhoff Publishing, Boston, 1982, 186p, f 90,--

(International series in Management Science/Operations Research).

In dit boek, gebaseerd op het proefschrift van de auteur (zie [1]), wordt een aantal, voornamelijk kwantitatieve methoden behandeld voor middellange en lange termijn personeelsplanning voor groepen personeel, welke bruikbaar zijn voor grotere bedrijven en overheidsorganisaties.

In hoofdstuk 1 wordt personeelsplanning gedefinieerd als een verzameling activiteiten die ten doel hebben het noodzakelijk aantal personen met de vereiste kwalificaties in de organisatie beschikbaar te houden, zodat de doelen van de organisatie worden verwezenlijkt en waarbij rekening wordt gehouden met de belangen van de individuele personeelsleden. Als activiteiten worden onderscheiden: wervingsplanning, loopbaanplanning, opleidings- en vormingsplanning en verplaatsingsplanning. Binnen het personeelsplanningsproces worden 3 fasen onderscheiden, te weten het voorspellen van respectievelijk de behoefte aan en de beschikbaarheid van personeel en de onderlingen afstemming van behoefte en beschikbaarheid. Tenslotte wordt voor een aantal typen organisaties aangegeven welke personeelsplanningsactiviteiten belangrijk zijn.

In de volgende twee hoofdstukken worden methoden behandeld om de behoefte aan en de beschikbaarheid van personeel te voorspellen op middellange en lange termijn. Er worden statistische en subjectieve methoden onderscheiden om een voorspelling van de personeelsbehoefte te verkrijgen. Bij de statistische methoden worden modellen genoemd die gebaseerd zijn op extrapolatie van de trends in de personeelsbehoefte (univariate modellen) en modellen waarbij een relatie wordt verondersteld tussen de personeelsbehoefte en een aantal verklarende variabelen (multivariate modellen). Als subjectieve methoden worden de Delphimethode en de work-study methode genoemd. Bij beide methoden wordt het management van de betreffende organisatie direct betrokken. Omdat de personeelsbehoefte door vele factoren wordt beïnvloed, zijn algemene statistische methoden vaak slecht bruikbaar.

De toepasbaarheid van deze methoden is sterk afhankelijk van het type organisatie en de personeelsgroep, waarvoor wordt gepland. Als afsluiting van dit onderwerp wordt een procedureschets gegeven van de personeelsbehoeftebepaling in een organisatie, die gebaseerd is op een aantal praktijksituaties. Voor het voorspellen van de beschikbaarheid van personeel worden als wiskundige methoden Markov modellen, renewal modellen en optimaliseringsmodellen genoemd. De laatste 2 typen modellen zijn geen zuivere voorspellingsmethoden maar combinenren het voorspellen met het afstemmen van beschikbaarheid op behoefte. Er wordt aangegeven aan welke eisen modellen moeten voldoen teneinde toepasbaar te zijn voor

personeelsbeschikbaarheidsanalyse. Op grond hiervan, de gegeven definitie van personeelsplanning en de gekozen indeling van het personeelsplanningsproces in 3 fasen, wordt het Markov model gekozen als basis voor een personeelsplanningssysteem.

In hoofdstuk 4 wordt een op een Markov model gebaseerd personeelsplanningssysteem FORMASY (FOrecasting and Recruitment in MAnpower SYstems) besproken. Dit systeem is gebaseerd op een interactief computerprogramma waarmee een analyse gemaakt kan worden van de huidige en toekomstige personeelsbeschikbaarheid. Met dit systeem kunnen ook de gevolgen van een alternatief personeelsbeleid worden doorgerekend. Essentiële gegevens voor FORMASY zijn de huidige personeelsbezetting ingedeeld in categoriën, de overgangsfracties voor iedere categorie en de toekomstige werving per categorie. De overgangsfracties kunnen daarbij gebaseerd zijn op historische gegevens, maar kunnen ook bepaald zijn op basis van een alternatief beleid. Als resultaat van FORMASY verkrijgt men de voorspelde bezetting per categorie, leeftijdsverdelingen, verwachte salariskosten, stationaire verdelingen en diverse andere opties.

In hoofdstuk 5 komt de onderlinge afstemming van personeelsbehoefte en beschikbaarheid aan de orde. Het ontwikkelen van een alternatief beleid met betrekking tot werving, loopbaanpatronen, opleiding en vorming en verplaatsing wordt hier genoemd en in het bijzonder de rol die een personeelsplannings systeem als FORMASY hierbij kan vervullen. Ook wordt in dit hoofdstuk ingegaan op de categorie-indeling van het personeel en de geschikte lengte van de plannings horizon. Bij beide keuzen is de flexibiliteit van het personeel in de betreffende organisatie een belangrijke factor.

In het laatste hoofdstuk wordt een case study beschreven over de toepassing van FORMASY bij de Koninklijke Nederlandse Luchtmacht.

Als sterke punten van dit boek zou ik willen noemen:

* het boek geeft een goed overzicht van relevante (kwantitatieve) methoden die in de personeelsplanning gebruikt kunnen worden;
* iedere methode wordt op een heldere manier beschreven waarbij ook voor- en nadelen van de betreffende methode genoemd worden;
* de ordening van de behandelde stof per hoofdstuk is overzichtelijk en goed.

Als minder sterk punt van het boek zou ik willen noemen het ontbreken van een duidelijke visie op de integratie van personeelsplanning in het totale besluitvormingsproces in een organisatie. Daardoor is ook in mindere mate gekeken naar de eisen die vanuit een dergelijke integratieproblematiek aan beleidsondersteunende systemen gesteld moeten worden. Verder komt in het boek de relatie tussen de globale behoeftebepaling, die in hoofdstuk 2 behandeld wordt, en het opstellen van het takenpakket in een organisatie niet aan de orde.

Sterke en minder sterke punten afwegend, moet toch geconstateerd worden dat

dit boek als overzichtswerk op het gebied van de (kwantitatieve) personeelsplanning vrij uniek is. Het boek zal dan ook geschikt zijn voor onderwijsdoeleinden als het gaat om voorbeelden van de toepassing van kwantitatieve methoden (in de personeelsplanning) en als het gaat om met name kwantitatieve personeelsplanning. Bovendien zal het boek goed bruikbaar zijn voor mensen uit het bedrijfsleven, die zich bezig houden met onderzoek naar en ontwikkeling van personeelsplanningssystemen.

[1] Verhoeven, C.J. (1980) 'Instruments for Corporate Manpower Planning-applicability and applications', dissertatie T.H. Eindhoven.

J.D. van der Bij

David Freedman

Brownian Motion and Diffusion

Springer Verlag, Berlin, 1983, x + 231p, ISBN 3-540-90805-6, DM 68,--

Het boek is het tweede, in de trilogie, Markov Chains, Brownian Motion and Diffusion, Approximating Countable Markov Chains, geschreven door Freedman rond 1970 en oorspronkelijk verschenen in de "Holden-Day" series in probability and statistics. Springer Verlag verzorgt een heruitgave, het onderhavige boek is het eerste hiervan. Het verschilt niet van de eerste uitgave afgezien van kleine correcties en vier toevoegingen aan de omvangrijke literatuurlijst.

Het eerste hoofdstuk, 98 pagina's verdeeld over 11 paragrafen behandeld de Brownse Beweging; het tweede hoofdstuk 76 pagina's verdeeld over 15 paragrafen betreft de Diffusie. De appendix, 42 pagina's, in 17 paragrafen, resumeert de grondslagen van de kansrekening en enkele begrippen uit de analyse. Het boek bood geen vraagstukken wel veel voorbeelden.

Het is een geschikte inleiding voor het vertrouwd raken met de theorie van Brownse Beweging en Diffusie, bruikbaar als handleiding bij een seminarium, vereiste voorkennis:
elementaire kansrekening en grondslagen kansrekening. Voor een meer gedetailleerde bespreking zie de review #6074 in Math. Rev.

Een aanbevelingswaardig boek.

J.W. Cohen