

Boekbesprekingen

Redactie:

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

telefoon 013 - 640319

P. McCullagh and J.A. Nelder

Generalized Linear Models; Second edition

Chapman and Hall, London, 1989, xix + 511 pag., ISBN 0-412-31760-5, £ 30.00

De toename in het aantal bladzijden - van 261 naar 511 - maakt direkt duidelijk, dat de tweede editie van *Generalized Linear Models* meer is dan een herdruk met correcties en toevoeging van een enkele paragraaf. De auteurs zelf schrijven:

"In the new edition the following topics have been added: conditional and marginal likelihood methods; estimating equations; and models for dispersion effects and components of dispersion. In addition of the following topics has been substantially revised and expanded: loglinear and related models; log odds-ratio regression models; multinomial response models; inverse linear and related models; quasi-likelihood functions; and model checking. Several new worked examples have been added and the discussion of previous examples has been revised and enlarged. For the benefit of teacher and student, exercises, both theoretical and applied, have been added to all chapters."

De wijze waarop de auteurs zich van hun taak gekweten hebben is indrukwekkend. Het boek is veel leesbaarder geworden; de stof wordt beter gestructureerd aangeboden, cryptische omschrijvingen zijn nu uitgewerkt en technische verhalen worden ondersteund met goed te begrijpen voorbeelden. Dit geldt vooral voor de "oude" hoofdstukken, waarin specifieke GLM's behandeld worden. Recente resultaten zijn volledig geïntegreerd in de tekst.

Als voorbeeld neem ik hoofdstuk 4: Binary data. Het verschil tussen "grouped data" en "ungrouped data" wordt op een andere, veel duidelijker wijze geïntroduceerd. Nieuw zijn de paragrafen "Genisis" en "Moments and cumulants", waarin kort ingegaan wordt op de eigenschappen van de verdeling. Daarna volgen normale en Poisson benadering en transformaties. Deze onderwerpen komen veel uitgebreider aan bod dan in de eerste editie. Grotendeels nieuw zijn ook de paragrafen "Parameter interpretation", "Bias and precision of estimates", "Sparsness" en "Extrapolation" (Bijvoorbeeld naar zeer lage doses voor het vinden van de maximum acceptable dose). Een zeer interessante uitbreiding is het stuk "Over-dispersion".

De hoofdstukken die volledig of voor een groot deel nieuw zijn, komen nog wat minder uit de verf. De onderwerpen zijn lastig - HS 7 "Conditional likelihoods", HS 9 Quasi-likelihood functions, HS 10 "Joint modelling of mean and dispersion" en HS 14 "Components of dispersion" en de theorie nog niet volledig uitgewerkt, maar deze delen zijn moeilijker leesbaar, dan alleen op grond van de moeilijkheidsgraad van de stof verwacht mag worden. In deze hoofdstukken lijken de auteurs hun didactisch streven wat uit het oog verloren te hebben. De behandelde stof is zeer recent of, in ieder geval voor mij, nieuw. Het idee, dat de basis vormt voor HS 10 bijvoorbeeld is in 1984 door Pregibon geopperd en de enige artikelen die ik ken, zijn Nelder & Pregibon (1987, *Biometrika*) en Smyth (1989, *JRSS-B*). Ook de aanpak van spreidingscomponenten door middel van quasi-likelihood (HS 14) is niet eerder zo uitgewerkt opgeschreven. Let wel, deze uitwerking is nog niet meer dan een raamwerk, dat voor een bepaald voorbeeld uitgewerkt wordt.

Conclusie: De tweede editie van *Generalised Linear Models* is een indrukwekkende verbetering van de eerste. De eerste editie was een bestudering al zonder meer waard, wat nog te zeggen van de tweede? Ook eerste editie-kenners en -bezitters raad ik aan deze tweede editie te bestuderen en aan te schaffen.

Peter van Ewijk
Duphar B.V.

Thomas P. Ryan

Statistical methods for quality improvement

John Wiley & Sons, Ltd., 1989, xviii + 446 pag., ISBN 0-471-84337-7, £ 39.30.

Dit boek beschrijft technieken die van belang zijn voor de kwantitatieve aspecten van het kwaliteits verbeteringsproces. De opzet is dat een onderwerp wordt behandeld en dat daarna in korte paragrafen, meestal niet langer dan een halve pagina, nadere aspecten vermeld worden met verwijzingen naar de literatuur. Het gevolg van een dergelijk brede aanpak is onvermijdelijk dat er bij de geïnteresseerde lezer met een statistisch technische achtergrond te weinig diepgang in het boek zit, maar een zeer waardevolle bron van informatie is over de aan het onderwerp gerelateerde aspecten en waar deze informatie te vinden is. De auteur is momenteel consultant op dit gebied en is een tijd lang directeur geweest van het consultatiecentrum van de Universiteit van Iowa.

Het boek is gesplitst in drie delen, welke uiteraard weer zijn onderverdeeld in hoofdstukken. Ieder hoofdstuk bevat opgaven, de uitkomsten van sommige opgaven zijn achterin het boek opgenomen. Tevens wordt na ieder hoofdstuk een opgave gegeven van de literatuur.

In het voorwoord wordt als motivering voor het boek aangegeven dat voor kwaliteitsverbetering meer nodig is dan alleen procesbeheersingskaarten. Om te weten te komen hoe een proces binnen specificaties gekregen kan worden bij buiten de efficiënte methoden van de statistische proefopzetten. In deel 1 van het boek worden de fundamentele concepten van kwaliteitsbeheersing en statistiek behandeld. Het bevat een hoofdstuk over de beknopte beschrijving van de geschiedenis van de toepassingen van de statistische methoden in de kwaliteitszorg, een hoofdstuk waarin de Japanse aanpak wordt beschreven met een historisch overzicht (rol van McArthur en Deming) en enige technieken die een rol spelen (visgraat diagram, pareto diagram, histogram en spreidingsdiagram). Verder worden in een hoofdstuk de basisprincipes van de kansrekening en de statistiek, de 'usual stuff' behandeld.

Deel 2, over proceskaarten en capabiliteits analyses, begint met een algemene verhandeling over de opzet en achterliggende gedachten van de proceskaarten. Korte stukjes over bijvoorbeeld waar, wanneer en hoe gebruik gemaakt kan worden van deze techniek. De volgende twee hoofdstukken gaan over de proceskaarten voor continue grootheden. Een hoofdstuk is gewijd aan de situatie waarbij de waarnemingen worden bemonsterd in subgroepen. Daarin komen de R-kaart, de s-kaart, s^2 -kaart, X-gemiddelde kaart, kaarten gebaseerd op CUSUM procedures en een kaart voor het voortschrijdend gemiddelde aan bod. De voorbeelden waarmee deze technieken 'verkocht' worden zijn niet gebaseerd op uit de praktijk komende data. In de tekst staat 'if the data had been real data, ...'. Voor mij is dat een zwak punt. Omdat deze technieken specifiek uit praktijkomgevingen voortkomen vind ik het veel meer spreken als voor de illustraties ook feitelijke data worden gebruikt. Nu blijf je zitten met de vraag 'hoe het nu in werkelijkheid zou gaan'. Uw recensent is daarvan wel al overtuigd, maar lezers die deze overtuiging zoeken in het boek of handvatten willen om de technieken in te voeren zouden daarmee geconfronteerd kunnen worden. Interessant in de afsluiting van het hoofdstuk, dat handelt over de (statistische) veronderstellingen waarop het gebruik van de regelkaarten is gebaseerd en wat gedaan kan worden als aan deze veronderstellingen niet is voldaan (bij niet-normaliteit bijvoorbeeld transformatie uitvoeren en bij niet-onafhankelijkheid de afwijkingen van een AR(1) model op de kaarten uitzetten). De bijlage van dit hoofdstuk bevat de wiskundige afleiding van de bij de kaarten te gebruiken constanten (voor de ingewijden d_2 , c_4 , D_3 , D_4 , B_3 en B_4).

Het daaropvolgende hoofdstuk handelt over continue grootheden waarbij bemonstering niet plaatsvindt in subgroepen. Kort aan bod komen de X-kaart, een kaart voor waarden met trend (AR(1)), voortschrijdende gemiddelden en CUSUM procedures voor locatie- en variatiematen. Hoofdstuk 7 gaat over capabiliteits analyses. Althans, zo luidt de kop van het hoofdstuk. Behandeld worden alleen maar de capabiliteitsindices C_p en C_{pk} en dan nog kort. Daarmee doet de auteur dit onderwerp naar mijn mening sterk te kort. Capabiliteits Analyses vervullen de rol van een uitgebreidere (informatie-)kennismaking met het proces onder studie en dienen vaak als uitgangspunt voor de toepassing van de proces-beheersingsmethodieken als procesbeheersings kaarten. Deze rol had veel duidelijker naar voren moeten komen.

Hoofdstuk 8 is gewijd aan de controlekaarten voor attributieve grootheden. Aan bod komen de p-kaart, np-kaart, arcsinus transformatie voor proporties, np-kaart met runs-kriterium, CUSUM procedure voor binomiale data als alternatief voor p-kaart of np-kaart (via transformatie van de waarnemingen naar een bij benadering normaal verdeelde grootheid), c-kaart (met transformaties $2\sqrt{c}$, $2\sqrt{(c+3/8)}$, $\sqrt{c} + \sqrt{(c+1)}$), CUSUM procedures voor de c-kaart (weer via transformaties) en de u-kaart. Vooral de CUSUM procedures en de transformatie zijn nogal van academische aard, omdat mijns inziens deze methoden veel te ingewikkeld zijn voor het toepassen onder 'dagelijkse omstandigheden' door de 'eigenaren' van het proces, de operators of de kwaliteits medewerkers. Alleen de mogelijkheid de methoden onder te brengen in software zou deze methoden bruikbaar kunnen maken. Het laatste hoofdstuk van dit deel gaat over multivariate proceskaarten. Daarvoor geldt ook hetgeen hierover is gezegd. De methode wordt netjes en voorzichtig uitgelegd, via de één-dimensionale t-grootheid naar de Hotelling's T-grootheid. Het advies van de schrijver is om univariate kaarten en multivariate kaarten gelijktijdig te gebruiken. Aan de hand van een voorbeeld (met gesimuleerde waarden) laat hij zien dat een multivariate kaart met $p=2$ een 'buiten regelgrenzen' signaal kan geven zonder dat de beide univariate kaarten een afwijking melden. De verklaring daarvoor wordt wel gegeven, maar het blijft vaag of deze extra informatie wel de extra inspanning waard is. In een aparte paragraaf wordt duidelijk gemaakt dat een multivariate kaart alleen iets toevoegt in vergelijking met aparte kaarten (met Bonferroni grenzen) indien de p kwaliteits kenmerken gecorreleerd zijn.

Deel 3 handelt over de in de eerste alinea genoemde methoden voor kwaliteitsverbetering die verder gaan dan de proceskaarten. Het eerste hoofdstuk bevat grafische methoden. Onderwerpen als histogram, stam-en-blad diagram, snorrendoos en normaal waarschijnlijkheids papier komen aan de orde, maar ook de gezichten van Chernoff, de 'draftsman plot' en de 'weathervanse' worden genoemd. Van deze laatsten is alleen een afbeelding van twee Chernoff-gezichten opgenomen, de anderen zijn alleen vermeld met literatuurverwijzing.

Lineaire regressie is het onderwerp van het volgende hoofdstuk. Aan de hand van een ietwat armoedig voorbeeld (X: tijd besteed aan kwaliteitsverbetering, Y: aantal afkeurprodukten) wordt de techniek uitgelegd. Veel onderwerpen worden genoemd, meestal slechts met verwijzing naar literatuur (modelselectie middels Cp, collineariteit, partiële residuen plot). Een goed voorbeeld uit de praktijk had veel illustratiever geweest en had ook veel beter kunnen worden geëxploiteerd voor het behandelen van diverse technieken.

Behandeling van de proefopzetten begint met een voorbeeld (vier rubbersamenstellingen voor autobanden, vier wielposities en vier auto's) om de noodzaak aannemelijk te maken dat er goed moet worden nagedacht over de manier waarop de informatie gestructureerd moet worden om een informatieve proef uit te voeren. In 58 pagina's worden veel onderwerpen behandeld. Aspecten van factoriële 2^k en 2^{k-p} schema's (verstrengeling, aliasing, resolutie), Yate's algorithmen, uitzetten van effecten op normaalwaarschijnlijkheidspapier komen aan bod, waarbij ook nog een uitleg wordt gegeven over de benodigde variantie analyse technieken. Heel veel dus in beperkte ruimte, waardoor de lezer op dit gebied alleen maar 'natte voeten' kan krijgen in de zee van methoden en technieken van de proefopzetten.

De Taguchi methoden zijn onderwerp van hoofdstuk 4. Na een verhandeling over de verlies-functie zoals deze is geïntroduceerd door Taguchi volgt de behandeling van de 'signal-to-noise' ratios en alternatieven (PERMIA's). Duidelijk wordt de doelstelling van deze procedures uitgelegd, eerst de variabiliteit verlagen middels de instellingen van de variabelen die daarop van invloed zijn en daarna het proces instellen naar de gewenste waarde. De recente literatuur wordt aangehaald om de discussie omtrent de waarde van deze ratios en de mogelijke verbeteringen te ondersteunen. Daarna volgen de 'orthogonale arrays', waarbij de hoofdzaak ligt bij het aantonen dat de fractionele factoriële aanpak proefschema's opleveren die efficiënter zijn (hogere resolutie bijvoorbeeld). De nadruk komt in de hoofdstuk daardoor te veel te liggen op de bestrijding van de experimenteermethode volgens Taguchi alleen aan de hand van deze efficiency grootheden.

Hoofdstuk 15 besteedt aandacht aan EVolutionary OPERATION, de methode die door Box in 1957 is geïntroduceerd om middels steeds herhaalde kleine proefschema's de processen te optimaliseren en die uitgevoerd zou moeten kunnen worden door 'operators'. Aan de hand van een voorbeeld wordt de methode kort behandeld. Uit diverse publicaties blijkt dat deze manier van continue verbetering van de processen niet erg aanslaat en de (mogelijke) redenen daarvoor worden ook

genoemd in dit hoofdstuk.

Tenslotte een onderwerp dat ik nog niet in boekvorm ben tegengekomen, namelijk 'ANalysis Of Means' (ANOM). Deze methode is eind jaren vijftig geïntroduceerd door E.R. Ott. Het idee is om de groepsgemiddelden grafisch weer te geven ten opzichte van zogenaamde 'decision lines' welke worden afgeleid uit het algemeen gemiddelde plus of min een aantal maal de (gepoolde) standaard deviatie binnen groepen. Vallen groepsgemiddelden buiten deze grenzen, dan is dat een aanwijzing dat er significante verschillen bestaan. Methodisch zijn er sterke overeenkomsten met regelkaarten. Dat maakt de werkwijze aantrekkelijk omdat het gebruik ervan inzichtelijker kan worden gemaakt aan mensen die niet de kennis hebben van variantie analyse maar wel onderzoekjes doen voor het vergelijken van gemiddelden. Er worden toepassingen gegeven voor zowel waarnemingen op continue schaal als voor afkeurpercentages per groep en aantal afkeur per groep. Vooral de behandeling van deze voorbeelden maakt de relatie met de regelkaarten duidelijk.

Over een algemeen oordeel over het boek zou ik het volgende willen zeggen. De opzet en de inhoud doet academisch aan. Zoals gezegd, voorbeelden zijn hoofdzakelijk gebaseerd op gesimuleerde data. Een ander nadeel vind ik dat er weinig aandacht is besteed aan de presentatie van de grafieken, deze zien er nogal knullig uit. Daar had meer aandacht aan geschonken moeten worden. Het boek is een bron van informatie voor degenen die verder willen zoeken naar (achtergronden van de) methoden en technieken. Door de zeer brede opzet is de behandeling van de onderwerpen kort en statistisch technisch van aard.

Het boek lijkt mij uitstekend geschikt voor mensen met een statistische achtergrond, die het vakgebied van de kwaliteitsbeheersing en de kwaliteitsverbetering binnen willen stappen en voor mensen die daarin werkzaam zijn en op zoek zijn naar een breed overzicht van de voorhanden zijnde methoden. Daarmee voldoet het aan de gestelde doelen.

Arend Oosterhoorn

Van Doorne's Transmissie bv

Glen McPherson

Statistics in Scientific Investigation, Its Basis, Application and Integration

Springer Texts in Statistics, Springer Verlag, New York, 1990, xxvi + 666 pag., ISBN 0-387-97137-8, DM 98.00.

Dit boek is een inleiding in elementaire statistiek. In de inleiding werpt de auteur zelf de vraag op of het nodig is nog een boek toe te voegen aan de reeds lange reeks van dergelijke boeken. Als argumentatie voor een positieve beantwoording van deze vraag wijst hij op twee uitgangspunten. Ten eerste heeft hij gestreefd naar het combineren van een elementaire opzet met beknoptheid in de presentatie. Het elementaire karakter is zeker aanwezig: er is nauwelijks wiskundige voorkennis vereist. De beknoptheid komt wat minder goed uit de verf gezien het aantal pagina's van dit boek en de aankondiging dat de auteur nog twee vervolgdelen op stapel heeft staan. Afgezien van enige wijdloopigheid, bijvoorbeeld in hoofdstuk 22 over het gebruik van computerpakketten, kan de omvang de auteur niet euvel geduid worden. Hij ruilt deze namelijk in voor een uitzonderlijk heldere presentatie verlucht met vele voorbeelden. Het hoge didactische niveau dat de schrijver ten toon spreidt, is ongetwijfeld een gevolg van zijn zeer lange ervaring met het doceren van de stof en met statistisch consultatiewerk.

Een tweede uitgangspunt is dat het boek niet bedoeld is voor wiskundigen doch voor hen die statistische methoden willen toepassen in hun eigen onderzoek. Het boek maakt de lezer wegwijs in de verschillende manieren van denken die gangbaar zijn binnen de statistiek en die meestentijds voor buitenstaanders een ondoorzichtig geheel vormen. Ongetwijfeld is het sterkste punt van het boek de objectieve weergave van de uitgangspunten van de verschillende strategieën. Zo wordt bijvoorbeeld ingegaan op de sterke punten en de beperkingen van een Bayesiaanse aanpak. Deze strategie wordt in het boek overigens weinig toegepast, omdat de auteur er theoretisch wel een

voorkeur voor heeft maar de praktische nadelen te ernstig vindt.

Het tweede uitgangspunt van de auteur tezamen met de heldere schrijfstijl rechtvaardigen deze uitgave meer dan voldoende. Als inleiding en/of naslagwerk is de aanschaf aan te bevelen aan niet wiskundig georiënteerde toepassers van statistiek. De data van alle tabellen in voorbeelden en opgaven mitsgaders de programmatuur, waarmee de opgaven zijn op te lossen met behulp van MINITAB of SAS, zijn bij de auteur op aanvraag verkrijgbaar.

Een laatste sterk punt van het boek is een mededeling op de koft: printed on acid-free paper! Na deze opmerking gelezen te hebben neem ik m'n andere boeken met argwaan en voorzichtigheid ter hand.

J. Molenaar

Instituut Wiskundige Dienstverlening Eindhoven

Faculteit Wiskunde en Informatica, TUE

H.T. Nguyen & G.S. Rogers

Fundamentals of mathematical Statistics, Volume I : Probability for Statistics

Springer Verlag, Berlin Heidelberg New York, 1989, x + 432 pag., ISBN 0-387-97014-2 & ISBN 3-540-97014-2, DM 84.00.

Fundamentals of mathematical Statistics, Volume II : Statistical Inference

Springer Verlag, Berlin Heidelberg New York, 1989, xi + 422 pag., ISBN 0-387-97020-7 & ISBN 3-540-97020-7, DM 84.00.

Dit tweedelig werk vormt een inleiding tot de mathematische statistiek. Het is bedoeld voor studenten wiskunde en verwante studierichtingen. Een grondige kennis van de analyse en enige bekendheid met matrices wordt aanwezig verondersteld.

De delen bevatten elk drie (doorgenummerde) hoofdstukken die zijn opgedeeld in lessen:

- H.I Elementary Probability and Statistics (183 pag.) 21 lessen.
- H.II Probability and Expectation (108 pag.), 13 lessen.
- H.III Limiting Distributions (133 pag.), 15 lessen.
- H.IV Sampling and Distributions (80 pag.), 14 lessen.
- H.V Statistical Estimation (109 pag.), 14 lessen
- H.VI Testing Hypotheses (106 pag.), 12 lessen.
- H.VII Special Topics (124 pag.), 11 lessen.

De opdeling ("colleges") is karakteristiek voor deze inleiding. De auteurs schatten dat voor elke les drie uren zelfstudie nodig zijn. Dit lijkt een redelijke schatting omdat de leerstof in de lessen zo uitvoerig is opgeschreven dat zelfstandige uitwerking van details voor het volgen van redeneringen nauwelijks nodig is. Zelfwerkzaamheid wordt wel gevraagd voor de vraagstukken die in de leerstof geïntegreerd zijn en derhalve moeilijk kunnen worden overgeslagen.

H.I is een zeer eenvoudige, vrij informele inleiding in elementaire kansrekening en statistiek en vormt de motivatie voor een meer formele opbouw in de volgende hoofdstukken.

H.II,III bevatten de gebruikelijke onderwerpen uit de kansrekening, opgezet m.b.v. maattheorie. Deze expositie is vrij grondig. De auteurs streven in hun bewijsvoering niet naar volledigheid. Juist aan eenvoudige zaken wordt uitvoerig aandacht besteed. Geen bewijs wordt gegeven van de echte harde resultaten van de maat- en integratietheorie (uitbreidingsstelling van maten, stelling van Fatou en convergentiestelling van Lebesgue, verwisselingsstelling van Fubini, etc.). De convergentiebegriffen voor stochastische variabelen en hun onderlinge samenhang komen uitvoerig aan bod. In H.IV wordt veel aandacht besteed aan de asymptotische eigenschappen van de ML-methode. In

vergelijking met andere inleidingen met dezelfde doelstelling zijn de gegeven afleidingen werkelijk zeer degelijk.

Van H. V kan het zelfde gezegd worden. Wel vallen hier twee onzorgvuldigheden m.b.t. de informatie-ongelijkheid op: 1) de conclusie op pagina 107 dat in het geval van gelijkheid (voor alle parameterwaarden?) de zuivere schatter UMVUE is, kan niet zomaar getrokken worden; 2) het bewijs op p. 112 m.b.t. "een exponentile familie in het geval van gelijkheid" bevat de (gebruikelijke) onnauwkeurigheid dat uitzonderingsverzamelingen van de onbekende parameter kunnen afhangen, en contrasteert met de elders in het boek zo zorgvuldige betoogtrant.

H. IV bevat de klassieke Neyman-Pearsontheorie. Van stellingen m.b.t. UMP-zuivere toetsen voor tweezijdige toetsingsproblemen worden geen bewijzen gegeven maar wel toepassingen behandeld.

H. VIII vormt een enigszins wilde verzameling van sterk uiteenlopende zeer beknopt behandelde onderwerpen.

Aan de concrete analyse van "real-life" data wordt geen enkele aandacht besteed. Waar al getallen optreden worden deze steeds gebruikt voor invuloefeningen in formules.

Samenvattend mag gesteld worden dat dit boek vanwege zijn lessenopbouw en zijn aangenaam nauwkeurige mathematische betoogtrant een welkome toevoeging is aan de verzameling van inleidingen kansberekening en mathematische statistiek.

Of het ontbreken van cases uit de praktijk als bezwaarlijk ervaren wordt, hangt af van de doelstelling van de gebruiker.

B.B. van der Genugten, KUB

K.C. Peace (editor)

Statistical issues in Drug research and Development.

Marcel Dekker, Inc., New York, 1990, xi + 366 pag. ISBN 0-8247-8290-9, \$ 119.50.

Dit boek is de neerslag van de inspanningen van een aantal werkgroepen binnen de Biofarmaceutische sectie van de American Statistical Association. De aanpak en samenstelling van deze werkgroepen heeft geleid tot een bundeling van 7, los van elkaar leesbare hoofdstukken waarin op heldere en evenwichtige wijze methodologische aspecten met kwantitatieve illustraties beschreven worden voor een aantal algemeen voorkomende onderwerpen op het terrein van gecontroleerd klinisch geneesmiddelen onderzoek.

Voor elke werkgroep gold het volgende "statement of purpose": "It is not necessarily intended that the work group undertakes original research on the topics. The thrust is to put the problem in perspective and to present and integrated discussion of information relevant to the topic. The purpose is to develop a report that would contain:

- a detailed description of the problem and its history;
- an assessment of the impact of the problem on the applied area of research;
- a (re)solution of the problem, if possible. Otherwise a characterization of the pros and cons of possible (re)solutions;
- conclusions and recommendations from the (re)solution of the problem;
- a bibliography".

De 7 hoofdstukken worden kort besproken:

2. Summarization, Analysis, and Monitoring of Adverse Experiences.

Vrijwel alle literatuur gaat over het effectiviteitsbewijs. De weinige geschriften over analyse van bijverschijnselen zijn dermate versnipperd, dat de auteurs ervoor hebben gekozen een op zich zelf staand document te produceren. Dit heeft geresulteerd in een omvangrijk hoofdstuk (140 pag) dat op zichzelf al de moeite waard is om dit boek voor aan te schaffen. Naast algemene methodologische beschouwingen, komen een aantal statistische technieken van Mantel-Haenszel, overlevingsduur

en gegeneraliseerde lineaire modellen aan de orde, inclusief software-verwijzingen. Het geheel wordt gelardeerd met een aantal uitgebreide praktijkvoorbeelden die het inzicht verdiepen.

3. Two treatment Crossover designs.

Dit hoofdstuk stelt de lezer in staat om zich vertrouwd te maken met de twistpunten rondom het gebruik van dit type proefopzet en een aantal suggesties om deze op te lossen.

4. Active Control Equivalence Studies.

In dit hoofdstuk wordt het betreffende nieuw type clinical trial besproken dat is bedoeld om de gelijkwaardigheid in effectiviteit aan te tonen van een nieuwe behandeling met de standaard behandeling. Dit kan aan de orde zijn als het bijv. niet etisch is om een placebo toe te dienen. In de beschouwingen komt ook de bruikbaarheid van historische controles aan de orde.

5. Optimization in Clinical trials and Combination Development.

De keuze van de "beste" dosering komt hier aan de orde. Voor een combinatie preparaat wordt een toepassing van RSM (response surface methodology) geïllustreerd, ontleend aan de statische theorie voor het optimaliseren van productieprocessen.

6. Dosing in the Elderly

Sinds kort worden ouderen als aparte doelgroep binnen de patiëntenpopulatie onderkend. Het hoofdstuk geeft een motivatie waarom dit zinvol is maar wijst tevens op een aantal interpretatie problemen die hierbij opdoemen.

7. Intention to treat in Clinical Trials.

De RCT (randomized clinical trial) is de hoeksteen van het geneesmiddelenonderzoek. Binnen deze keuze doen zich problemen voor als niet alle gerandomiseerde patiënten het "eindpunt" van de studie bereiken. Het dispuut laat zich samenvatten met 'is the purpose of the experiment to study the drug or to study the randomization process?

Hoewel de werkgroep zelf geen consensus hierover bereikte, is zij er wel ingeslaagd om de diverse posities helder weer te geven.

8. Dual Control Groups in Rodent Carcinogenicity Studies.

Recente suggesties om bij carcinogeniciteits-onderzoek twee controle groepen te gebruiken worden in dit hoofdstuk gemotiveerd verworpen.

Resumerend, kan gesteld worden dat dit praktijkgerichte boek zich prettig laat lezen, met een afwisseling van kwalitatieve en kwantitatieve beschouwingen. De inhoud dient tot het geestelijk eigendom te behoren van iedere statisticus die met klinisch geneesmiddelen-onderzoek van doen heeft. Grote delen van dit boek zijn ook voor de serieuze medische onderzoeker toegankelijk.

Paul Koopman, Duphar B.V.

Marcel F. Neuts

Structured stochastic matrices of M/G/1 Type and their applications

Marcel Dekker, Inc., New York, 1989, 510 pag., ISBN 0-8247-8283-6, \$ 150.00.

De M/G/1 queue is een wachtijsysteem met als aankomstproces, met intensiteit zeg λ , met willekeurige bedieningsduurverdeling H en met één bediende. Beschouw deze queue op achtereenvolgende vertrektijdstippen $T_1, T_2, \dots$. Zij N_k het aantal achterblijvende klanten op tijdstip T_k en definieer

$$Q_{ij(x)} = P \{N_{k+1} = j, T_{k+1} - T_k \leq x \mid N_k = i\}$$

De matrices

$$Q(\cdot) = \{Q_{ij(x)}, x > 0\}$$

vormen de overgangsmatrices van een Markov-vernieuingsproces. Deze matrices zijn van een zeer

speciale vorm, namelijk:

$$Q(x) = \begin{bmatrix} B_0(x) & B_1(x) & B_2(x) & \dots \\ A_0(x) & A_1(x) & A_2(x) & \dots \\ 0 & A_0(x) & A_1(x) & \dots \\ 0 & 0 & A_0(x) & \dots \\ \vdots & \vdots & \vdots & \ddots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (*)$$

met

$$A_k(x) := \int_0^x e^{-\lambda u} \frac{(\lambda u)^k}{k!} dH(u) \quad \text{en} \quad B_k(x) := \int_0^x \lambda e^{-\lambda u} A_k(x-u) du$$

De scalaire grootheid $A_k(x)$ is de kans op k binnenkomsten, bij een bedieningsduur van ten hoogste x tijdseenheden; bij $B_k(x)$ wordt de bedieningsperiode nog vooraf gegaan door een exponentiële tijdsduur gedurende welke de queue leeg is. Uitgaande van de functies $A_k(x)$, $B_k(x)$, $k > 0$, is het mogelijk de M/G/1 queue tot in groot detail te bestuderen. Dit wordt gedaan in Hoofdstuk 1 van bovenstaande tekst, als inleiding tot meer ingewikkelde modellen in de verdere hoofdstukken. In deze verdere hoofdstukken worden de scalaire functies vervangen door matrices $A_k(x)$ en $B_k(x)$. Een voorbeeld hiervan is de semi-Markov queue M/SM/1, waarbij opeenvolgende klanten hun bedieningstijd kiezen volgens een Markov-vernieuingsproces met m toestanden. Voor de M/SM/1 queue worden de bedieningsuren gegeven door de matrix $(H_{ij}(x))$, $i=1, \dots, m$, $j=1, \dots, m$ terwijl $Q(\cdot)$ in matrix-notatie wederom gerepresenteerd wordt door $(*)$.

Hoofdstuk 2 beschrijft welke klasse van matrices $Q(x)$ bestudeerd worden in dit boek en geeft noodzakelijke e voldoende condities opdat het bijbehorende Markov-vernieuingsproces recurrent is. Hiertoe wordt de fundamentele periode (doorgangstijd van "level" $i+1$ naar "level" i) geïntroduceerd en berekend met behulp van de getransformeerde matrix $\tilde{G}(s, z)$. Bij de M/SM/1 queue is $\tilde{G}$ de minimale oplossing van:

$$\tilde{G}(z, s) = z \sum_{k=0}^{\infty} \tilde{A}_k(s) \tilde{G}^k(z, s) = z \tilde{H}(s + \lambda - \lambda \tilde{G}(z, s)), \quad (\tilde{H}(s) = \int_0^{\infty} e^{-sx} dH(x))$$

een formule die het analogon is van de vergelijking voor de Laplace-getransformeerde van een drukke periode in de M/G/1. Overigens geeft de auteur de voorkeur aan vergelijkingen van het type:

$$G = \sum_{k=0}^{\infty} A_k G^k,$$

boven een functionaalvergelijking voor Laplace-getransformeerden, vanwege het numerieke werk (de oplossing G wordt, na afkapping, numeriek bepaald door successieve iteraties $G_{n+1} = \sum A_k G^k$ startend vanuit $G_0 = A_0$).

In hoofdstuk III behandelt de auteur positieve recurrentie van het Markov-vernieuingsproces. gecontroleerd moet nu worden of gemiddelde terugkeertijden van de "levels" eindig zijn. Hier komen dus de afgeleiden

$$\frac{\partial}{\partial z} \bar{G}(z, s) \quad \text{en} \quad \frac{\partial}{\partial s} \bar{G}(z, s)$$

aan de orde. Verder worden in dit hoofdstuk besproken: de stationaire vector x van de matrix $Q(\infty)$ en de momenten van deze kansvector. Het hoofdstuk sluit met een uitstekende paragraaf over de numerieke berekening van de verschillende matrices, hun respectievelijke stationaire vector en de momenten van deze kansvectoren. De verdere hoofdstukken 4, 5 en 6 behandelen specifieke modellen, zoals bijvoorbeeld genoemde M/SM/1 queue, Bailey's queue en de PH/G/1 queue, tot in groot detail, terwijl in hoofdstuk 5 ook meer algemene aankomstprocessen behandeld worden.

Dit boek is een vervolg op een eerder boek van dezelfde schrijver: Matrix-Geometric Solutions in Stochastic Models (zie SIAM REVIEW 1983(25), 127-128 voor een bespreking). Het is niet echt noodzakelijk, maar wel aanbevelingswaardig, dit eerdere boek, dat vooral GI/M/1 type Markov-vernieuingsprocessen behandelt, eerst te bestuderen.

Het boek zit wiskundig uitstekend in elkaar, maar is niet altijd gemakkelijk te lezen. Veel rekendetails, die voor een geroutineerd wachttijdspecialist misschien onmiddellijk duidelijk zijn blijven achterwege (relatie (3.2.18) volgt echt niet trivaal uit (2.1.2), zie p. 141). Met de notatie ben ik ook niet altijd even gelukkig; bijvoorbeeld; de notatie jj' (zie (2.2.3) op p. 79), of de keuze van de letter $\bar{H}$ in (3.2.7), bij het uitwerken van deze laatste relatie ontstaat verwarring daar $\bar{H}$ ook de Laplace getransformeerde van de bedieningsduur aangeeft.

Het aantal referenties (± 1200) is formidabel. Al met al is dit boek een must voor de wachttijdspecialist, maar ontoegankelijk voor anderen, vooral ook door de buitensporig hoge prijs.

Gerard Hooghiemstra
Technische Universiteit Delft

Alan J.B. Anderson

Interpreting Data, A first course in statistics

Chapman and Hall, London 1989, xvi + 223 pag., ISBN-0-412-29570-9, £ 10.95.

Onder mathematisch-statistici is gegevensverzameling en beschrijvende statistiek niet erg populair. In het onderwijs, met name op de tegenwoordige hogescholen, komt het wel expliciet in het programma voor maar de docenten, vaak mathematisch-statistisch geschoold, vinden het niet exact genoeg of zelfs "gezwets". In leerboeken worden deze onderwerpen wel behandeld maar toch wat stiefmoederlijk bedeed. Mijns inziens is dat niet terecht, en ik was verheugd over het boek van Anderson omdat daaruit hetzelfde blijkt.

In dit boek wordt op zeer aantrekkelijke wijze aandacht besteed aan gegevens: hoe ze worden verzameld, gerangschikt, samengevat en "verstaan". Dit alles is van fundamenteel belang voor statistische analyse. Het woord fundamenteel is hier op z'n plaats omdat gegevens de basis vormen voor diepergaand onderzoek. De kwaliteit van het werk van de statisticus, de conclusies die het onderzoek opleveren, staan of vallen uiteindelijk toch met de kwaliteit van de gegevens, iets waarvan ik mij als CBS'er dagelijks bewust ben.

Het boek is vrij elementair; een VWO-voorbereiding is zeker voldoende. In feite sluit het goed aan bij het huidige Wiskunde A onderwijs. In het eerste gedeelte (mijn indeling) worden begrippen geïntroduceerd. Er zijn hoofdstukken over de populatie en steekproeven, tabellen en grafieken, numerieke grootheden als mediaan, gemiddelde en variantie, over verbanden tussen variabelen (regressie) en het toetsen van hypothesen. Vervolgens komt de ambtelijke statistiek aan de orde. Uiteraard wordt in dit boek uitgegaan van de Britse omgeving maar met het Statistisch Jaarboek van het CBS bij de hand kan dit eenvoudig vertaald worden naar de Nederlandse situatie. Er zijn hoofdstukken over indexcijfers, demografische begrippen als bevolkingspyramide en levensverwachting, en tijdreeksen. Het derde gedeelte gaat over waarneming. Aan de orde komen onder

andere verschillende vormen van steekproefopzet, het opstellen van vragenlijsten, en experimenten. Voorbeelden zijn vooral medisch/biologisch of uit de sociale wetenschappen.

Elk hoofdstuk wordt afgesloten met aansprekende opgaven. Het boek als geheel wordt gecompleteerd door een overzicht van gebruikte symbolen en een index. De vormgeving van het boek is prettig, de schrijfstijl die van een goed verteller. Het ziet er allemaal verzorgd uit en ik beveel het dan ook van harte aan.

Dirk Gorter, CBS

Kenneth D. Lawrence en Jeffrey L. Arthur (eds.)

Robust Regression, Analysis and Applications.

Marcel Dekker Inc., New York, 1990, xiii + 287 pag., ISBN 0-8247-8129-5, \$ 107.50.

Het boek bestaat uit 18 op zichzelf staande artikelen over robuuste regressie. Alle auteurs hebben zich gehouden aan de definitie van Huber (1981): Robuuste statistische procedures zijn ongevoelig voor kleine afwijkingen van de vooronderstellingen.

De laatste jaren wordt in de literatuur veel aandacht besteed aan high-breakdown methoden en de effecten van enkele flinke uitschieters. In dit boek wordt hierover niet gerept. Uitzonderlijk voor een bloemlezing van dit type is dat er wel een alfabetische index is opgenomen. De volgende concepten zal men hierin tevergeefs zoeken: influence function, gross error sensitivity, (optimal) B-robustness, local-shift sensitivity en rejection point.

De hier behandelde onderwerpen hebben betrekking op de meer klassieke opvattingen betreffende robuuste regressie. Ze zijn gerubriceerd in vier delen. Het eerste deel heeft betrekking op nieuwe ontwikkelingen. Er wordt aandacht besteed aan Bayesiaanse schattingen (Chen en Box), Lp-schatters (Sposito) en robuuste niet-lineaire regressie (Lawrence en Arthur). De Bayesiaanse methode betreft gewogen regressie waarbij het gewicht van een waarneming afneemt naarmate deze verder van het model is verwijderd. Zoals te verwachten wordt bij de Lp-schatters voornamelijk aandacht besteed aan de L1-schatter die de som van de absolute waarden van de residuen minimaliseert. In het hoofdstuk over niet-lineaire regressie wordt de robuustheid verkregen door gebruik te maken van iteratief herwogen kleinste kwadraten. De aanpak wordt geïllustreerd met vier echte data-sets die middels SAS zijn geanalyseerd.

Het tweede deel gaat over methoden voor robuuste regressie (in welk opzicht dit zich onderscheidt van het thema voor het eerste hoofdstuk is mij niet duidelijk geworden). De onderwerpen zijn: meervoudige multivariate uitschieters in samenhang met L1-regressie (Armstrong en Beck), iteratief herwogen kleinste kwadraten met afnemende gewichten voor punten met grote residuen (Morgenthaler), de robuustheid van de L1-schatter (Narula en Wellington) en residuen in modellen met variantie-componenten (Fellner). De bijdrage van Armstrong en Beck is zeker lezenswaard. Hun methode spoort ieder k-tal uitschieters op, ook indien deze onderling verstrengeld zijn. Het is een interessante combinatorische puzzel, vergelijkbaar met de Cook en Weisberg problematiek betreffende meervoudige uitschieters bij klassieke regressie.

Het derde deel gaat over de rol van robuuste regressie bij voorspellingen. De volgende onderwerpen worden behandeld: robuuste tijdreeks analyse middels L1-regressie (Cogger), robuuste methoden voor invoer-uitvoer modellen zoals bij productie technologie en goederenstromen tussen industrieën (Lawrence en Gold), het opsporen van uitschieters middels empirische transformaties (Geurts en Tolley), een interessante econometrische case-study waarin de resultaten van klassieke regressie worden vergeleken met die van robuuste alternatieven (Guerard en Lawrence) en de behandeling van uitschieters in op regressie gebaseerde modellen voor de voorspelling van verkoopcijfers (Geurts, Rinne en Lawrence). In dit deel is in het bijzonder de studie van Guerard en Lawrence lezenswaard. Klassieke regressie levert in hun voorbeeld namelijk resultaten op die intuïtief niet acceptabel zijn. Toch zijn in het verleden uit deze resultaten conclusies getrokken en gepubliceerd. Deze conclusies kunnen middels robuuste methoden worden

ontkend.

Het laatste deel gaat over Ridge regressie. Volgens de stelling van Gauss-Markoff resulteert klassieke regressie onder de gebruikelijke vooronderstellingen in de schatters met de kleinste variantie binnen de familie der zuivere schatters. Ridge regressie is een middel om nauwkeuriger schatters te verkrijgen waarbij de eis van zuiverheid is opgeofferd. In het bijzonder kan de schade die multicollineariteit bij klassieke regressie aanricht hierdoor worden beperkt. De eerste bijdrage in dit deel gaat over de simultane behandeling van uitschieters en multicollineariteit (Pfaffenberger en Dielman) en in de tweede worden klassieke en Ridge regressie vergeleken in een case-study (Guerard en Ochsner). Van beide regressie-methoden wordt in dit onderzoek ook nog een robuuste variant betrokken.

Ondanks de vele auteurs bestrijkt dit boek maar een beperkt gedeelte van het terrein dat door de titel gesuggereerd wordt. Hoewel dat nergens genoemd wordt, zou het me niets verbazen als deze bundel is ontstaan uit de proceedings van een conferentie. De verschillende bijdragen per onderdeel bevatten nogal wat overlappende informatie. Elementaire zaken worden op meerdere plaatsen toegelicht en definities herhaald gegeven. Op de omslag staat: This is the leading reference for applied statisticians, operations researchers, econometricians, marketing forecasters, business administration and management scientists, and industrial engineers as well as a text for graduate statistics or economics courses. Deze kwalificatie lijkt me wat overdreven. Er zijn echter een paar onmiskenbaar aardige bijdragen bij en vooral onder de case-studies is voor de geïnteresseerde robuusticus nog wel wat nieuws te vinden.

Jan B. Dijkstra, TUE

Eeke van der Burg

Nonlinear canonical correlation and some related techniques

DSWO Press, Leiden, 1988, 241 pag., ISBN 90-6695-026-9, f 35.00.

Het boek is een verzameling gepubliceerde artikelen waarop Eeke van der Burg in Leiden promoveerde op 15 september 1988. De promotoren waren prof. J. de Leeuw en prof. W. J. van der Linden. Zoals de schrijfster in haar inleiding opmerkt, moet het boek gezien worden als één van de vele publicaties die het werk van Gifi (1981) uitbreiden en verdiepen. Pas onlangs is het boek van Gifi in bijgeschaafde vorm uitgekomen bij Wiley (Gifi, 1990). Bovendien zijn een aantal methoden daaruit nu algemeen beschikbaar als module CATEGORIES van SPSS. Er is daarom ongetwijfeld een vernieuwde vraag naar achtergrond materiaal. Het is zeker geen gemakkelijk leesbaar boek, maar voor de specialist zeer de moeite waard. De hoofdstukindeling met subtitels geeft een goed overzicht van het boek:

1. Introduction
2. CANALS: two-set canonical correlation. Nonlinear canonical correlation.
3. REDUNDALS: redundancy analysis. Nonlinear redundancy analysis.
4. OVERALS: k-set canonical correlation. Nonlinear correlation analysis with k sets of variables.
5. OVERALS: k-sets homogeneity analysis. Homogeneity analysis with k sets of variables: an alternating least squares method with optimal scaling features.
6. HOMALS: homogeneity analysis. Homals classification of whales, porpoises and dolphins.
7. Stability: permutation methods. The permutational limit distribution of generalized canonical correlations.

8. Stability: randomization methods. Use of the multinomial jackknife and bootstrap in generalized nonlinear canonical correlation analysis.

Author Index

Cajo J. F. ter Braak

Groep Landbouwwiskunde, Wageningen

Ida Stamhuis

'Cijfers en Aequaties' en 'Kennis der Staatskrachten', Statistiek in Nederland in de negentiende eeuw

Rodopi, Amsterdam, 1989, xi + 295 pag, ISBN 90-5183-120-X, f 60.00.

Dit geschrift is de neerslag van het promotieonderzoek naar de geschiedenis van de statistiek in Nederland in de negentiende eeuw. Over deze materie is slechts zeer weinig geschreven.

Naaste een inleiding waarin de verandering van de betekenis van de term 'statistiek' wordt geïllustreerd aan de hand van omschrijvingen in woordenboeken en encyclopedia (Woordenboek voor Kunsten en Wetenschappen van G. Nieuwenhuis (1826, 1866), Winkler Prins Encyclopedie (1880), Oosthoek Encyclopedie (1930) en Winkler Prins Encyclopedie (1983)) bestaat het boek uit vijf delen.

Het eerste deel geeft een beschrijving van de politieke rekenkunde en de 'Staatenkunde', de wortels van de statistiek, de bijdrage van de kansrekening en een behandeling van de rol die de Belgische geleerde Quetelet (1796-1874) heeft gespeeld voor de Nederlandse ontwikkeling.

Het tweede deel bespreekt de Nederlandse wiskundige Rehuel Lobatto (1797-1866). Veel nadruk ligt daarbij op sterfgetabellen en de problemen bij het opstellen van een betrouwbare tabel door de geringe beschikbaarheid van cijfers. Lobatto heeft zich veel bezig gehouden met de in die tijd opkomende weduwenfondsen. Hij was degene die, op basis van de overweging van de wet van de grote getallen, aangaf dat een dergelijk fonds uit een bepaald minimum aantal deelnemers zou moeten bestaan. In de eerste jaren was het aantal deelnemers vaak gemajoreerd. Zijn belangrijkste publicatie gaat over levensverzekeringen. Hij heeft zich zeer ingespannen voor overheidsbemoeienis bij de levensverzekeringen. tevens gaf hij een statistisch jaarboekje uit, waarvan de eerste in 1826 verscheen. Het doel daarvan was tweeledig, 'de verspreiding van algemeen nuttige kundigheden onder alle klassen van beschaaftde ingezetenen, alsmede de bekendmaking van belangrijke opgaven, meer bijzonder dit land betreffende, ...'. Het was dus niet speciaal voor wetenschappelijk of statistisch publiek. De eerste volkstelling in 1795 is ook mede door zijn inspanningen verricht.

Deel drie geeft de verandering in opvattingen over statistiek aan de Nederlandse universiteiten en in de Vereniging voor Statistiek (niet de huidige) weer. Deze veranderingen worden geïllustreerd aan de hand van de inzichten van drie hoogleraren, Adriaan Kluit (1735-1807), Simon Vissering (1818-1888) en Anthony Beaujon (1853-1890). Adriaan Kluit, hoogleraar in geschiedenis en oudheidkunde aan de Universiteit van Leiden, was de eerste die statistiekcolleges verzorgde aldaar. Simon Vissering, ook in Leiden, veranderde van inzicht in zijn loopbaan. In het begin zag hij de statistiek nog alleen als hulpwetenschap voor de staathuishoudkunde, om economische wetten op te sporen, later zag hij ook toepassingen voor theologen, historici en juristen. Anthony Beaujon was directeur van het Statistisch Instituut van de vereniging en hoogleraar aan de Gemeentelijke Universiteit van Amsterdam. Bij hem is een duidelijke verschuiving zichtbaar in de richting van de beschrijvende statistiek. Meer dan Kluit en Vissering was hij vertrouwd met kwantitatief denken en wiskunde. Dat blijkt ook uit zijn interpretatie van een grafiek van de 'huwelijksquotiënt van verschillende Europese staten sedert 1873', hij keek meer achter de cijfers en zocht een verklaring voor het fenomeen dat door de cijfers werd weergegeven.

Duidelijk wordt in de behandeling de ontwikkeling van de rol van 'het getal'. Dat was in het geheel niet zo, de 'staatenkunde' bevatte ook besprekingen van moraal en zeden, naast het aantal weerbare mannen en cijfers over criminaliteit. Het beeld veranderde in de loop van de eeuw in de richting van 'het getal' (er was een tijd dat mensen die tabellen produceerden geringschattend

'tabellenstatistiker' werden genoemd).

De Vereeniging voor Statistiek (1857-1892) is het onderwerp van het vierde deel van het boek. Het ontstaan, de ontwikkeling en de betrokkenheid bij het overheidsbeleid wordt beschreven. Van een mogelijke oprichting van zo'n vereniging was al in 1846 besproken in de Sectie Regtsgeleerdheid en Staatswetenschappen van het Provinciaal Utrechtsch Genootschap voor Kunsten en Wetenschappen en in 1847 bij het Lanbouwkundig Genootschap. De feitelijke oprichting is er toen niet van gekomen omdat de overheid aanstalten maakte om initiatieven te ontplooiën in de richting van de statistiek. De oprichting van de vereniging is voortgekomen uit de uitgave van een Staatkundig en Staathuishoudkundig jaarboekje door de Bosch Kemper, waarvan de eerste uitgave kwam in 1849. De redactievergadering stelde in 1857 voor om een Nederlands Statistisch Genootschap op te richten. In 1862 werd de vereniging officieel een rechtspersoon. Directe aanleiding daarvoor was dat de Rijksc commissie voor statistiek, opgericht in 1858, in 1861 alweer ter ziele was. Dat de vereniging zich zeer verantwoordelijk voelde ten opzichte van de taak die de statistiek zou kunnen uitvoeren in het landsbelang blijkt uit het feit dat zij de leemte opvulde van het wegvallen van het statistisch bureau van het ministerie van Binnenlandse Zaken (1848-1876). Daarvoor werd het Statistisch Instituut opgericht (1884-1892). Toen de overheid middels de Centrale Commissie en later met de oprichting van het Centraal Bureau voor de Statistiek (1899) die taak weer naar zich trok, was bij de vereniging de aandacht voor de statistiek ook snel verdwenen (deze vereniging bestaat nu nog onder de naam Vereniging voor Staathuishoudkunde).

In het laatste deel wordt de Nederlandse situatie vergeleken met die in Engeland, waar in 1834 de London Statistical Society ontstond en welke, in de vorm van de Royal Statistical Society, nog steeds bestaat. Dat deze vereniging nog wel als statistische vereniging bestaat wordt toegeschreven aan het feit dat in Nederland geen duidelijke band werd gelegd tussen de statistiek en de wiskunde, wat in Engeland, na veel verwoede discussies, wel gebeurde. Aan Lobatto wordt toegeschreven dat hij dat verband wel zag. Er waren in het begin van de twintigste eeuw opnieuw een aantal wetenschappers die zich verdiepten in het toepassen van de kansrekening op de statistiek, maar het is vooral dankzij de inspanningen van Van Dantzig dat de mathematische statistiek ingang vond in Nederland (maar die geschiedschrijving komt eraan naar ik heb begrepen).

De bijlagen bevatten het voorstel van Simon Vissering in 1857 tot oprichting van een statistisch genootschap en de statuten van de VVS.

Hoe moet je een dergelijk boek wegen. De promotiecommissie zal voor de borging van de wetenschappelijke kwaliteit hebben gezorgd. De vraag is dus waarom ik het lezen van dit boek aan zou raden. Wel, het is gewoon leuk om deze geschiedenis te lezen. Een aantal zaken vallen door het opnemen van deze kennis op zijn plaats. Zo kom je bekende namen tegen uit de geschiedenislessen van vroeger, Domela Nieuwenhuis, bijvoorbeeld, heeft als minister meerdere malen gepleit voor de oprichting van een soort CBS. Ook het bestaan van een eerdere Vereeniging voor Statistiek en de rol die deze vereniging heeft gespeeld bij de oprichting van het CBS was mij niet bekend. Dat de wortels van de statistiek ergens bij beschrijving van de toestand van een staat (Staatenkunde) lagen, was mij wel bekend, maar nu is het plaatje duidelijk.

Een ander aardig gevolg van het lezen van deze informatie is, dat het vertellen over de geschiedenis van de statistiek, de oorsprong en de ontwikkeling, een gewillig oor vindt bij collega's niet-statistici en van dergelijke gewillige oren moeten velen van ons het toch hebben in onze beroepsuitoefening.

Als goedgeaard Nederlander heb ik toch iets op te merken. Ten eerste is Ida Stamhuis van huis uit geen statisticus, want ze maakt de klassieke fout bij het omschrijven van een 'betrouwbaarheids-interval'. Ten tweede, maar dat is inherent aan het feit dat dit een proefschrift is, wordt het lezen gestoord door de vele noten, die in de tekst zijn opgenomen. Gelukkig staan alle noten onderaan de pagina waar zij voorkomen uitgewerkt. Wel heb ik respect voor het enorme snuffelwerk in ongetwijfeld stoffige omgevingen, dat aan het schrijven ten grondslag heeft gelegen. In de vorm als proefschrift, maar zeker als er nog een 'gepopulariseerde' versie uitkomt, kan ik het zeker aanraden het boek te lezen.

Arend Oosterhoorn
Van Doorne's Transmissie bv

Rien van der Leeden

Reduced rank regression with structured residuals

DSWO Press, Leiden, 1990, x + 148 pag., ISBN 90-6695-045-5.

Bij mijn statistiek opleiding was het toch wel een schok te horen dat multivariate multiële regressie in feite neerkomt op het uitvoeren van een reeks afzonderlijke univariate multiële regressies. Door echter beperkingen aan te brengen in het multivariate regressie model krijgen we een model waarvoor deze equivalentie niet meer geldt. Rien van der Leeden bespreekt in zijn interessante boek twee typen beperkingen, die afzonderlijk en in combinatie kunnen worden toegepast:

- (1) restricties op de rang van de matrix van regressie coëfficiënten. Als geldt dat de rang van deze matrix $\leq r$ is, is deze matrix te schrijven als AB waarbij het aantal kolommen van A gelijk is aan r .
- (2) restricties op de covariantie matrix van de fouten. Deze matrix wordt gemodelleerd door structurele relaties op te stellen à la LISREL.

Het boek is helder geschreven. Na een korte inleiding (hoofdstuk 1) wordt in hoofdstuk 2 het model geformuleerd en worden methoden voor schatten en toetsen aangegeven. Daarna volgen vier hoofdstukken waarin bijzondere gevallen nader onder de loep worden genomen. Hoofdstuk 3 belicht als speciale gevallen onder meer canonische correlatie analyse en redundantie analyse. Hoofdstuk 4 richt zich op het LISREL aspect van het model. Het model geeft de extra mogelijkheid om versturende variabelen als regressors op de nemen zodat het LISREL model vervolgens de residuele variatie modelleert. Zuinigheid in het model wordt bevorderd door een rang restrictie op te leggen aan de matrix van regressie coëfficiënten behorende bij de versturende variabelen. In hoofdstuk 5 wordt het model toegespitst op factor analyse met vaste individu scores en met externe verklarende variabelen. Hierbij is de covariantie matrix diagonaal. Tenslotte wordt in hoofdstuk 6 verkend hoe het model kan worden gebruikt voor de analyse van groeicurven en andere herhaalde waarnemingen. De hoofdstukken 4-6 worden elk aangevuld met één of twee voorbeelden.

Nog een paar kanttekeningen. Bij het onderdeel toetsen wordt de indruk gewekt dat geen van de standaard toetsen bij MANOVA van het aannemelijkheidsquotiënt λ gebruik maakt. Zonder enige reserve in acht te nemen wordt ervan uitgegaan dat $-2 \log \lambda$ een chikwadraat verdeling volgt. Voor het vergelijken van rangrestricties (bijvoorbeeld $r = 0$ versus $r = 1$) geldt dit in elk geval niet (ondanks genommerde bronnen die het tegendeel beweren). Hiervoor hebben we de verdeling van de grootste eigenwaarde nodig (Corsten & van Eijnsbergen, 1972). Ik zou niet weten of de chikwadraat benadering geldt voor de voorgestelde toetsen op de bijdrage van extra regressors in modellen met vaste rang, zoals toegepast in tabel 5.2 met $r=2$. De asymptotische verdeling van schatters voor A en B wordt voor een speciaal geval gegeven door Velu et al. (1986). In het LISREL deel, wordt het effect van alcoholgebruik, roken en voeding op bloeddruk en vetheid gemodelleerd. Hierbij wordt gecorrigeerd voor de versturende variabelen leeftijd en hoeveelheid lichamelijke activiteit. Uit de resultaten van de analyse kan ik niet concluderen dat deze variabelen 'effect modifiers' zijn (p.75; misschien wel uit de verdere, afzonderlijke LISREL analyses). Is er rangrestrictie toegepast in het voorbeeld? Bij het narekenen van het aantal parameters in de twee voorbeelden van hoofdstuk 4 blijkt dat $r=2$. Aan de keuze van deze rang worden niet veel woorden vuil gemaakt. In het eerste voorbeeld gaat het in elk geval over meningen over twee onderwerpen, namelijk doodstraf en abortus, elk vertegenwoordigd met een aantal vragen. In het tweede voorbeeld suggereert de theorie misschien dat er twee basale concepten zijn. In beide gevallen zijn er een aantal verklarende variabelen. Een aantal modellen wordt aangepast, waarbij het meest ingewikkelde model (met alle interacties erin) uiteindelijk wordt gekozen (op grond van de chikwadraat benadering van de toetsingsgrootheid). Nu valt een model met driefactor interacties niet gemakkelijk uit te leggen. Daartoe wordt bij de presentatie gemiddeld over één factor. Toch valt het moeilijk de meest saillante interacties te achterhalen. Mijns inziens zou een vergelijking met het model met alleen hoofdeffecten hierbij hebben kunnen helpen. Bij het onderdeel

groeicurven wordt van parallelle profielen gesproken waar proportionele profielen bedoeld worden. Ik mis referenties naar de verwante aanpak van Snee (1972) en Snee et al. (1979) en naar de vergelijking van methoden (met analyses van de 'rat data') door Keen et al (1986). De aanpak van dit boek biedt aantrekkelijke extra mogelijkheden omdat de covarianties tussen tijdstippen er ook goed mee gemodelleerd en aangepast kunnen worden. In dit hoofdstuk blijkt dat in de voorgaande hoofdstukken het gemiddelde van elke response variabele steeds is verwijderd. Dit had kunnen worden voorkomen door een net iets algemener model te bekijken waarin voor een deel van matrix van regressie coëfficiënten een rang restrictie geldt ('partial reduced rank regression'). Een eventueel op te nemen constante vector of een versturende variabele zou dan buiten dat deel vallen.

De spreuk voorin het boek luidt "it is better to see once than to hear a hundred times". Dit boek is het waard bekeken te worden door een ieder die klassieke multivariate analyse gebruikt of doceert. De spreuk is tevens bijzonder van toepassing als de beperkte rang grafisch wordt uitgebuit door 'biplots' (ter Braak, 1990) te maken.

- Corsten, L. C. A. and van Eijnsbergen, A. C. (1972). Multiplicative effects in two-way analysis of variance. *Statistica Neerlandica* 26, 61-68.
- Keen, A., Thissen, J. T. N. M., Hoekstra, J. A. and Jansen, J. (1986). Successive measurement experiments. *Statistica Neerlandica* 40, 205-223.
- Snee, R. D., Acuff, S. K. and Gibson, J. R. (1979). A useful method for the analysis of growth studies. *Biometrics* 35, 251-260.
- Snee, R. D. (1972). On the analysis of response curve data. *Technometrics* 14, 47-62.
- ter Braak, C. J. F. (1990). Interpreting canonical correlation analysis through biplots of structural correlations and weights. *Psychometrika* 55, september issue.
- Veľu, R. P., Reinsel, G. C. and Wichern, D. W. (1986). Reduced rank models for multiple time series. *Biometrika* 73, 105-118.

Cajo J. F. ter Braak
Groep Landbouwwiskunde, Wageningen

Edward J. Dudewicz, Pinyuen Chen en Baldeo K. Taneja

Modern Elementary Probability and Statistics

American Science Press, Inc., Syracuse New York, 1989, ix + 372 pag., ISBN 0-935950-19-2, \$ 34.50.

Eigenlijk mag je dit geen boek noemen, het is meer een syllabus. Zachte kaft, bijna A4 formaat en gewoon getypte vellen met handmatig aangebracht commentaar en tekeningen. De tekst is ontwikkeld voor eerste-jaars cursussen in statistiek en kansrekening. Volgens de auteurs is het materiaal uitvoerig getest op bruikbaarheid. Voor cursussen met een kortere tijdsduur dan nodig zou zijn voor de gehele stof is het mogelijk bepaalde onderdelen weg te laten en deze zijn gemerkt in de tekst. In drie delen wordt de stof aangeboden.

In het eerste deel zijn dat onderwerpen als randomisatie, beschrijvende statistiek, kansrekening en voorwaardelijke kansen, Bayes' theorema, stochastische grootheden, kansverdelingen, binomiale verdeling en normale verdeling met de centrale limietstelling. In het tweede deel worden onderwerpen als punt- en intervallschattingen, het toetsen van hypothesen, lineaire regressie en correlatie, chi-kwadraattoetsen en variantieanalyse behandeld. Het laatste deel is gewijd aan drie pakketten, SAS, Minitab en BMDP, waarmee verschillende problemen worden aangepakt. Persoonlijk vind ik dit een instructief deel van de tekst, omdat ten eerste dezelfde problemen met de drie pakketten worden bekeken en omdat er verder nog opgaven in staan waarbij wordt gevraagd naar interpretatie van de uitvoer van de pakketten. Dit idee lijkt mij goed bruikbaar voor toepassingen in het

onderwijs.

Verder is er weinig opwindends te melden over deze uitgave. Alleen als iemand geïnteresseerd zou zijn in de wijze van aanpak van het cursusmateriaal is het aan te bevelen om ook een blik te werpen in deze uitgave, maar verder kan ik mij geen reden bedenken waarom dit boek zou moeten worden aangeschaft.

Arend Oosterhoorn
Van Doorne's Transmissie bv

A.P. Barten e.a.

Econometrische lessen

Academic Service, Schoonhoven, 1989, vi + 404 pag., ISBN 90-5261-002-9, f 49.50.

Zoals de titel doet vermoeden is dit boek in het Nederlands of eigenlijk in het Vlaams geschreven. Het boek bestaat uit 12 lessen en is een begeleidende tekst voor het college Econometrie aan de Katholieke Universiteit Leuven. De schrijver geeft aan dat het theorieeldeel van dat college ongeveer 45 uur omvat.

Het boek is op doctoraal niveau geschreven en is daardoor uitermate geschikt voor econometrie-studenten en voor een ieder die wat meer wil weten van toepassingen van regressie-analyse. Voor HEAO-studenten zal het boek te moeilijk zijn.

De lessen bevatten veel voorbeelden om de theorie te illustreren. De data zijn ontleend aan de Belgische economie. Elke les wordt afgesloten met een serie opgaven en een opsomming van verdere literatuur. Het boek bevat geen antwoorden. Bij de literatuur wordt vaak naar Belgische uitgaven verwezen.

Als voorkennis voor het boek achten de schrijvers enige kennis van analyse, lineaire algebra, waarschijnlijkheidsrekening en statistiek noodzakelijk. (Mij lijkt enige kennis van economie ook gewenst.) Toch wordt vaak van weinig parate kennis uitgegaan en worden veel zaken voorzichtig geïntroduceerd. Naast vele formules worden een heleboel zaken ook gewoon in schrijftaal uitgelegd. Aan het eind van een les is vaak nog een resumé van de behandelde stof gegeven. Hierdoor is een zeer helder en goed toegankelijk boek ontstaan.

De schrijvers wisten mij het meest te boeien in de veel voorkomende vaak filosofische passages - in erg mooi proza! - over bijvoorbeeld wat wil je met een model of waarom kies je een model. Hierdoor is het ook een echt econometrie boek geworden.

De inhoud van de lessen komt voor het merendeel neer op toepassingen van regressie-analyse en in het bijzonder het schatten van de coëfficiënten van regressievergelijkingen.

Les 1 is te beschouwen als een inleiding.

In les 2 wordt de kleinste kwadraten methode behandeld (35 pagina's), eerst voor het geval van enkelvoudige lineaire regressie zonder matrixrekening, daarna, als brug naar het vervolg, met behulp van matrixrekening. Vervolgens wordt het meervoudige lineaire regressiemodel besproken. Gelukkig wordt ook aandacht besteed aan de residuen en er worden twee grafische methoden besproken om de residuen in beeld te brengen. (In les 9 komt de autocorrelatie van storingstermen aan bod.)

Les 3 behandelt het statistische model en les 4 de eigenschappen van de kleinste kwadratenschatters.

In les 5 wordt voor het eerst uitgegaan van normaalverdeelde storingstermen en wordt de schattingsmethode van de grootste aannemelijkheid behandeld.

Les 6 is geheel gewijd aan voorspellen en les 7 aan toetsen. (Helaas wordt in een deel van de les voor de toetsingsgrootheid de letter θ gebruikt).

Les 8 heeft de titel specificaties en behandelt de moeilijkheden bij een overbodige verklarende variabele in de regressie-vergelijking en bij het ten onrechte weglaten van een verklarende variabele in de regressievergelijking. Daarnaast multicollineariteit en het gebruikmaken van

dummyvariabelen. Ook worden er enkele opmerkingen gemaakt over niet-lineaire regressie.

Les 9 handelt over storingspatronen. Hier wordt 'generalised least squares' besproken, het geval dus waar de storingstermen onderling gecorreleerd zijn, met als bijzonder geval de heteroskedastische en de autocorrelatie.

Les 10 gaat in op het geval dat de verklarende variabelen eigenlijk verklarende kansvariabelen zijn.

Les 11 handelt over vertragingen (lagged variables) en voortschrijdende gemiddelden.

In de laatste les (45 pagina's) worden stelsels van vergelijkingen (wij spreken meestal over simultane vergelijkingen) behandeld, de herleide vorm en de identificatie-problemen.

Dan bevat het boek ook nog twee nomogrammen voor de Durbin-Watson toets, een tabel van t en F verdeling, een lijst van statistische termen met het engelse franse en duitse equivalent (daarbij is ook het Vlaamse(?) geslacht van de term opgenomen) en een uitgebreid trefwoordenregister.

Het boek bevat erg weinig drukfouten.

Warm aanbevolen.

R.J. Köster.

Robert M. Groves

Survey Errors and Survey Costs

John Wiley & Sons Ltd., New York, 1990, xxi + 590 pag., ISBN 0-471-61171-9, £ 47.50.

Literatuur over foutenbronnen in survey onderzoek is zeer heterogeen. In verschillende vakgebieden bestaan verschillende indelingen in soorten van fouten. Behalve in de steekproeftheorie bestaat er bovendien weinig literatuur over de wijze waarop tegen hogere kosten de fouten in survey onderzoek gereduceerd kunnen worden. In "Survey Errors and Survey Costs" doet Groves een poging om de literatuur over foutenbronnen te integreren en dit bovendien te integreren met de kostenaspecten.

Hoofdstuk 1 bevat een indeling in foutenbronnen en een gemakkelijke discussie over de wijze hoe binnen verschillende vakgebieden als psychometrie, econometrie en survey statistiek tegen deze indeling wordt aangekeken. Hoofdstuk 2 is een inleidend hoofdstuk over kosten. Hierin wordt met name het stochastische en het discontinue karakter van kosten benadrukt. Overschrijdt de steekproefomvang een bepaalde grens, dan moet een extra supervisor of administrateur worden ingezet.

De volgende hoofdstukken behandelen alle een onderwerp uit het eerder gedefinieerde schema van foutenbronnen. Hoofdstuk 3 bevat een discussie over de gevolgen en de bestrijding van dekkingsfouten. Daarbij gaat het onder andere om het goed boven tafel krijgen van de samenstelling van het huishouden.

Hoofdstukken 4 en 5 gaan over non respons. In hoofdstuk 4 worden vooral de bekende statistische methoden om met non respons om te gaan behandeld. Hoofdstuk 5 behandelt de oorzaken van non respons en methoden om non respons te reduceren, zoals beloningen, introductiebrieven en interviewertraining. Aan de psychologische literatuur worden suggesties ontleend om de non respons te reduceren.

In hoofdstuk 6 wordt de steekproeffout behandeld die ook onderwerp is van een grote hoeveelheid literatuur over steekproeftheorie. Hierin komen uitgebreide modellen voor interviewervariantie aan de orde, met daarbij de kostenconsequenties. Het hoofdstuk wordt voor een belangrijk deel gevuld door een denkbeeldige discussie tussen een beschrijvend statisticus en een "modellenbouwer" die wordt afgesloten met het commentaar "only the publisher's constraint of making this book of finite length forces the ending to this debate".

Hoofdstuk 7 behandelt de respondent als variantiebron. Hierin komen test-hertest correlaties en multi-trait multi method modellen aan de orde. Hoofdstuk 8 heeft de interviewer als foutenbron als onderwerp. Aangevoerd wordt onder andere dat zowel een goede training van als een goede

supervisie over interviewers vruchten kan afwerpen. Hoofdstuk 9 behandelt de bias die het gevolg is van fouten op respondentniveau, zoals geheugeneffecten.

Hoofdstuk 10 heeft de rol van vragenlijsten als onderwerp. Aan de orde komen onder andere het verwoorden van vragen, open versus gesloten vragen, randomized response, volgorde effecten en kostenmodellen om concepten optimaal te meten. Hoofdstuk 11, tenslotte, gaat over het effect van de wijze van data verzameling en met name de tegenstelling tussen telefonisch en face to face enqueteren. Het boek maakt zijn pretentie om een goed geïntegreerde behandeling van verschillende vormen van fouten in survey onderzoek te geven zeker waar. Uiteraard is het vakgebied zo groot dat een keuze uit vele onderwerpen moet worden gemaakt. Naar mijn smaak zou de (relatief goedkope) statistische behandeling van fouten, zoals poststratificatie of empirisch Bayesiaanse methoden, meer aandacht hebben verdiend. De lezer vindt in het boek echter voor elk onderwerp een goede introductie. De integratie met kostenmodellen is minder overtuigend. Bij een aantal fouten worden kostenmodellen zeer expliciet behandeld; sommige van deze modellen zijn niet erg bekend en kunnen goed als eye opener dienen. Bij andere onderwerpen wordt deze modelmatige behandeling achterwege gelaten. Er worden ook geen voorbeelden gegeven van de kosten die door slechte schatters worden veroorzaakt. Tenslotte wordt één kostenpost geheel over het hoofd gezien. Die is het gevolg van de schaarste aan personen die de combinatie van wetenschappelijke en management kwaliteiten hebben om complexe kosten modellen met succes te kunnen implementeren. Daarnaast vereist implementatie een organisatorische souplesse die niet door elk onderzoeksbureau opgebracht kan worden. Ondanks deze kleine bedenkingen is "Survey Errors and Survey Costs" een prettig geschreven boek waaruit door velen veel valt te leren.

Dirk Sikkkel

Research International Nederland BV

Mary Jackson

Advanced spreadsheet modelling with Lotus 1-2-3™

John Wiley & Sons, Ltd., New York, 1989, ix + 177 pag., ISBN 0-471-91989-6, £ 17.95.

Werken met Lotus 1-2-3 en het overzichtelijk opzetten van een spreadsheet zijn de feitelijke onderwerpen van dit boek. Als toepassingen worden financiële onderwerpen gehanteerd. Het is ook bedoeld voor mensen die willen starten met Lotus 1-2-3, enige affiniteit hebben met kwantitatieve analyse en betrokken zijn bij besluitvorming in de financiële sector. Het maken van plaatjes, het doen van voorspellingen en het maken van budgetten komen aan de orde. Voor een overzicht kunnen we de inhoudsopgave weergeven

- Introduction
- 1 Estimation Models
- 2 Graphs for Data Presentation and Modelling
- 3 Financial Modelling - Planning
- 4 Financial Modelling - Budgeting
- 5 Business Forecasting
- 6 Decision Making
- 7 Introductory Macro Models
- 8 Simulation Models
- Appendix
- Index

Statistische technieken komen niet veel aan de orde en worden zeker niet behandeld als zodanig. Bij het onderwerp 'schatten van kosten' ligt de nadruk op het berekenen van deze schatting op basis van getallen in de spreadsheet en niet op de manier van schatten, dus de

statistische aspecten daarvan. Over grafieken wordt niets verteld maar er wordt aangegeven hoe je bepaalde grafieken (de 'standaard set') kunt maken met behulp van Lotus opdrachten. Bij het onderwerp 'Business Forecast' komen wel onderwerpen als trendbepaling uit tijdreeksdata en regressie analyse aan de orde, maar op de techniek zelf wordt niet ingegaan. Voor het bepalen van de mate van aanpassing van de regressielijn aan de waarnemingen wordt wel gesteld dat de standaard deviatie van de residuen gebruikt kan worden als maat, maar een formule is niet te vinden. Voor het berekenen van de t-waarde van de regressiecoëfficiënten wordt wel de formule gegeven, maar voor de interpretatie wordt verwezen naar boeken over statistiek.

Al met al is dit boek een beetje een vreemde eend in deze bijt. Misschien niet een echte uitschieter omdat er wel degelijk statistisch geïntereerde onderwerpen aan bod komen, maar er wordt niets uitgelegd over statistiek, het gaat om het toepassen van de spreadsheet techniek.

Vanuit dat oogpunt is het wel een goed en informatief boek. Heel rustig en overzichtelijk worden de technieken aangedragen en behandeld. Ik denk dat het zeer bruikbaar is voor mensen die een begin willen maken met het spreadsheet programma. Dat hoeft niet persé Lotus te zijn omdat de meeste spreadsheet programma's toch ongeveer hetzelfde doen. Als je bij het lezen van dit boek een vertaalslag kunt maken naar de opdrachten van het gebruikte programma, dan is het zeer bruikbaar om werken daarmee te leren. Misschien hoort daarmee dit boek meer thuis in een rubriek over computerboeken, maar dan toch bij deze.

Overigens is, zeer netjes, een diskette bijgevoegd met alle gebruikte voorbeelden in de tekst, dus je kunt zo aan de slag.

Arend Oosterhoorn
Van Doorne's Transmissie bv

D.G. Rees

Essential Statistics

Chapman and Hall Ltd., London, New York, 1989, xiii + 258 pag., ISBN 0-412-32030-4, £ 9.95.

Deze tweede uitgave is een update en uitbreiding van wat Rees zelf noemt 'the successful first edition'. Hij wordt hierbij gesteund door een recensie in *Journal of Quality Technology* waarin staat dat 'Anyone who wants to begin learning statistics should be delighted with this book'. Het doel van het boek wordt omschreven als het opwekken en aanmoedigen van belangstelling voor het vak statistiek voor beginners op dit terrein. Getracht is deze doelstelling te verwezenlijken door :

- toepassing van statistische methoden op een grote verscheidenheid aan vakgebieden
- een minimum aan mathematische theorie te gebruiken
- erg veel opgaven in het boek te verwerken.

Het boek bestaat in totaal uit 16 hoofdstukken, elk hoofdstuk wordt afgesloten met een zogenaamd worksheet bestaande uit een serie opgaven (ca 15 opgaven per hoofdstuk). Deze worksheets zijn een essentieel onderdeel van het boek waarmee de opgedane kennis direct getoetst kan worden. Het maken van veel opgaven is de beste manier om het vak te leren. Prettig is dat achterin het boek oplossingen worden vermeld, wat bij eventuele zelfstudie erg handig is. Een multiple choice test sluit het boek tenslotte af.

Hoofdstuk 16 is een losstaand onderdeel van het boek en behandelt de toepassing van het interactieve komputerpakket Minitab. Dit pakket is een hulpmiddel bij het statistiek onderwijs.

Deze veelheid aan onderwerpen wordt op een prettig leesbare wijze beschreven en met een veelheid aan voorbeelden steeds toegelicht zodat het geen dorre en droge opsomming is geworden. Bij het behandelen van begrippen en technieken gaat de schrijver snel op zijn doel af. Hierbij gaat hij voorbij aan alle mathematische afleidingen en ook worden gepresenteerde formules niet of

nauwelijks toegelicht. De lezer wordt zo snel mogelijk met het eindresultaat geconfronteerd zodat hij er sommen over kan maken. Deze opzet tot beknoptheid en eenvoud leidt soms tot een al te summier uitleg en soms is meer uitleg zelfs dringend gewenst.

De hoofdstukken 1 t/m 4 behandelen onderwerpen uit de beschrijvende statistiek, zoals tabellen en grafieken maken, berekenen van centrale waarden en maten voor de spreiding. De in deze tweede uitgave toegevoegde 'stem and leaf plots' en 'box and whisker plots' zijn zeker een aanwinst, omdat deze plots horen bij de exploratory data analysis technieken en veel in computerprogramma's worden gebruikt.

Jammer is het dat de auteur geen regels geeft voor het opstellen van frekwentietabellen. Terecht wijst hij de lezer er op hoe klassegrenzen eruit moeten zien, zonder overlap enz., maar hij geeft geen regels voor:

- het aantal klassen; het aantal klassen dat hij in het boek hanteert (blz. 12) is kleiner dan volgens de $\sqrt{n}$ regel vereist is,
- de ondergrens voor de laagste klasse.

Zonder deze noodzakelijke informatie is het niet mogelijk om over dit stuk van de theorie vraagstukken te maken.

Op blz. 31 wordt vermeld dat verderop in het hoofdstuk de reden om bij de standaardafwijking door (n-1) te delen nader zal worden bediscussieerd. Deze discussie (blz. 32) is wel erg oppervlakkig want deze blijkt dan in te houden dat vermeld wordt dat op deze wijze betere schattingen voor de standaardafwijking van de populatie verkregen worden.

De hoofdstukken 5 t/m 9 hebben betrekking op elementaire kansrekening, binomiale en Poisson verdeling, normale verdeling en (heel kort) de uniforme verdeling. Vervolgens komen steekproef, populatie en betrouwbaarheidsintervallen voor μ (normale verdeling) en p (binomiale verdeling) aan de orde. Een veel gebruikte techniek als controlekaarten wordt hier niet behandeld, terwijl bij het betrouwbaarheidsinterval wel de formule voor het voorspellingsinterval wordt gegeven, zonder deze zo te noemen.

Op blz. 81 wordt een zeer belangrijk stuk theorie op een al te korte wijze uit de doeken gedaan. Na een korte verbale inleiding wordt in enkele regels vermeld, zonder bewijs, dat $\mu(\text{gemiddelde}) = \mu$ en dat $\sigma(\text{gemiddelde}) = \sigma/\sqrt{n}$. Voor een dergelijk belangrijk stuk gereedschap, waarop in de rest van het boek steeds teruggevallen wordt, is meer uitleg nodig. Met de methode van het combineren van onafhankelijke trekkingen tot een gemiddelde zijn bovenstaande formules eenvoudig af te leiden en voor de lezer beter te begrijpen.

De hoofdstukken 10 t/m 15 behandelen de onderwerpen toetsen, zowel parametrische als niet-parametrische, correlatie en regressie. Met name het toetsingsonderwerp wordt zeer receptmatig behandeld waardoor de lezer deze techniek al snel zelf kan toepassen. Alle formules voor toetsingsgrootheden in de verschillende situaties worden zonder commentaar gepresenteerd. Stap voor stap wordt aangegeven wat er bij een toetsingsprobleem van de lezer verwacht wordt en zo leidt de toets naar een resultaat. Hopelijk wordt het de lezer duidelijk wat hij hierbij aan het doen is. Opvallend is dat alle toetsingen met een kritieke waarde worden uitgevoerd en niet met een overschrijdingskans. Deze laatste wijze van toetsen wordt beduidend vaker gehanteerd, ook in computerprogramma's.

Samenvattend kan ik zeggen dat het boek naar mijn indruk zal beantwoorden aan zijn doel, het gebruik door de beginnende student op het gebied van de statistiek. Het boekje is gemakkelijk leesbaar en zal door het vermijden van teveel theorie de lezer niet snel afstoten. Het bevat veel technieken, maar verschaft weinig inzicht om de lezer echt vertrouwd te maken met die verschillende statistische technieken. Een goede begeleiding kan hier mogelijk in voorzien. De grote hoeveelheid oefenstof spreekt mij erg aan.

Wim Jansen, AKZO Research Laboratories Arnhem

James R. Thompson

Empirical Model Building

John Wiley & Sons Ltd., New York, 1989, xiv + 242 p., ISBN 0-471-60105-5, £ 28.85.

Uw recensent is enige keren opnieuw aan dit boek begonnen, omdat hij meende het onvoldoende systematisch te lezen. Dit moet echter aan het boek zelf liggen, want ook nu nog komt het als los zand over.

De schrijver zegt ergens: "De poging om te begrijpen wat er werkelijk gebeurt is de inspanning waard." Dat moet dan ook maar voor dit boek zelf gelden.

"Empirical Model Building" stelt niet de methodologie maar het probleem centraal en bouwt vanuit de beschrijving van dat probleem een model. Alle modellen worden dientengevolge uit cases ontwikkeld. Een beschrijving met woorden wordt langzaam aan vertaald naar cijfermatige afleidingen. Deze worden op hun beurt gecombineerd en waar mogelijk omgezet in recursieve formules. Als er dan een gesloten vorm algoritme ontstaat is dat erg mooi, maar daar mag men niet altijd van uit gaan

De samenhang tussen de opeenvolgende problemen wordt uitsluitend weergegeven door de titels van de eerste twee hoofdstukken. Het boek omvat vijf hoofdstukken en twee appendices.

- Chapter one. Models of Growth and Decay.
- Chapter two. Models of Competition, Combat, and Epidemic.
- Chapter three. Simulation and the Coming Qualitative Change in Scientific Modeling.
- Chapter four. Some Techniques of Non Standard Data Analysis.
- Chapter five. Paradoxes and False Trails.
- Appendix 1. Brief Primer in Stochastics.
- Appendix 2. A simple Optimization Algorithm that Usually Works.

Elk hoofdstuk wordt gevolgd door een List of References, waar men nadere details over de behandelde onderwerpen kan vinden.

De onderwerpen komen voornamelijk uit de financiële, sociale en medische hoek. De meeste ervan bevatten het element tijd/ periode en worden aangepakt met differentiaal- en differentie vergelijkingen.

Bij de hoofdstukken 3 en 4 vroeg uw recensent zich af, wat hij daarmee in het kader van het boek moest beginnen en bij hoofdstuk 5 raakt hij het spoor bijster. Dit wordt vermoedelijk veroorzaakt, doordat de schrijver afzonderlijke problemen en niet hun methodologische samenhang centraal stelt.

Appendix 1 moet men niet gebruiken voor een eerste kennismaking met stochastiek. Het algoritme dat meestal werkt - in appendix 2 - omvat anderhalve pagina en komt voort uit enkele eerdere opmerkingen over data.

Wellicht vraagt u zich thans af, of dit boek volgens uw recensent dan helemaal geen verdiensten heeft. Die zijn er toch wel. Als men het cijfermatige en wiskundige deel laat voor wat het is, blijven er vlot geschreven beschrijvingen over - doorspekt met de eigen interpretaties van de schrijver - die beslist de moeite van het lezen waard zijn. Alleen daarvoor moet men het echter niet kopen.

Al met al een boek, dat ik toch niet echt kan aanbevelen.

Jaap Duisterwinkel, Turmac bv.

A.M. Mathai (editor)

Distributions of test statistics, exact & asymptotic, null & non-null, methods/comparisons/research frontiers

American Science Press, Syracuse, 1989, 182 pag., ISBN 0-935950-20-6, \$ 89.75.

Deze tot boekvorm omgevormde uitgave van het tijdschrift *American Journal of Mathematical and Management Sciences*, Volume 9, Nos. 1 & 2, 1989, ISSN 0196-6324, bevat een achttal artikelen omtrent exacte of asymptotische verdeling van toetsingsgrootheden onder de nulhypothese en onder de alternatieve hypothese betreffende één of meer multivariate normale populaties. De discussie is vooral gericht op de verdelingen van likelihood ratio toetsings grootheden voor het toetsen van hypothesen omtrent deze multivariate normale verdelingen.

De artikelen zijn

Distributions of Lawley-Hotelling's T^2 related statistics: a review

Minoru Siotani

Test of hypotheses for covariance matrices and distributions under multivariate normal populations

Hisao Nagao

Power function studies

E.M. Carter

Invariant polynomials and related tests

Takesi Hayakawa

Asymptotic distributions of test statistics for covariance matrices concerning elliptical distributions

C.G. Khatri & C.D. Bhavsar

Some common structures of distribution problems in geometric probabilities and multivariate statistical analysis

A.M. Mathai & G. Pederzoli

Wilks' integral equations in multivariate distribution theory

Marietta J. Tretter & William Walster

Generalized hypergeometric functions and exact distributions of test statistics

P.N. Rathie

>>>>>>>>>> CWI publicaties <<<<<<<<<<<<<<

Polderman, J.W.

Adaptive control & identification: Conflict or conflux?

CWI, Amsterdam, tract 67, 1989, 116 pag., ISBN 90-6196-377-X, f 17.80.

B.G. Hansen

Monotonicity properties of infinitely divisible distributions

CWI, Amsterdam, tract 69, 1990, ii + 101 pag., ISBN 90-6196-380-X, f 38.50.

ILO, Geneve, 1990, xiv + 90, ISBN 92-2-107276-2, Sfr 17.50

ILO, Geneve, 1990, viii + 116, ISBN 92-2-105559-0, Sfr 25.00

ILO, Geneve, 1989, x + 131, ISBN 92-2-106507-3, Sfr 20.00

ILO, Geneve, 1989, xii + 181, ISBN 92-2-106436-0, Sfr 27.50

ILO, Geneve, 1990, xvi + 161, ISBN 92-2-106511-1, Sfr 30.00

ILO, Geneve, 1990, vi + 182, ISBN 92-2-106446-8, Sfr 27.50

ILO, Geneva, Occupational safety and health series 61, 1990, viii + 57, ISBN 92-2-106509-X, Sfr 12.50

>>>>>>>> Binnengekomen tijdschriften <<<<<<<<<<

Dit is een speciaal nummer met onderwerpen uit de toegepaste econometrie. Het bevat zeven artikelen, elk gepresenteerd ten tijde van de Australian Meeting of the Econometric Society, gehouden van 18 augustus tot 2 september 1988.

Specification testing in censored regression models: parametric and semi-parametric methods

Bell, G.M. en Lavis, D.A.

Statistical mechanics of lattice models, vol. 1

Ellis Horwood Ltd., 1989, 368, 0-85312-822-7, £ 55.00

Heiberger, R.M.

Computation for the analysis of designed experiments

John Wiley & Sons Ltd., 1990, xv + 683, 0-471-82735-5, £ 47.50

Kotz, S. & Johnson, N.L.

Encyclopedia of statistical sciences, supplement volume

John Wiley & Sons, Ltd., 1990, xv + 283, 0-471-81274-9, £ 49.95

Yarmolik, V.N. & S.N. Demidenko

Generation and application of pseudorandom sequences for random testing

John Wiley & Sons, Ltd, 1988, vii + 167, 0-471-91999-3, £ 25.50

Bollen, K.A.

Structural equations with latent variables

John Wiley & Sons Ltd., 1989, xiv + 514, 0-471-01171-1, £ 39.30

Bhattacharya, R.N. & Waymire, E.C.

Stochastic processes with applications

John Wiley & Sons, Ltd., 1990, xiii + 672, 0-471-84272-9, £ 55.15

Skinner, C.J., Holt, D. en Smith, T.M.F. (eds.)

Analysis of complex surveys

John Wiley & Sons Ltd., 1989, xv + 309, 0-471-92377-X, £ 44.30

Gifi, A.

Nonlinear multivariate analysis

John Wiley & Sons, Ltd, 1990, xx + 579, 0-471-92620-5, £ 34.95

Whittaker, J.

Graphical models in applied multivariate statistics

John Wiley & Sons, Ltd., 1990, xiv + 448, 0-471-91750-8, £ 39.95

Witzling, L.P. en Greenstreet, R.C.

Presenting statistics, a manager's guide to persuasive use of s

John Wiley & Sons, Ltd., 1990, xiii + 241, 0-471-80307-3, £ 12.95

Hald, A.

A history of probability and statistics

John Wiley & Sons, Ltd., 1990, xiii + 586, 0-471-50230-8, £ 55.15

Kaufman, L. en Rousseeuw, P.J.

Finding groups in data, an introduction to cluster analysis

John Wiley & Sons, Ltd, 1990, xiv + 342, 0-471-87876-6, £ 39.30

Bhattacharya, R.N. en Waymire, E.C.

Stochastic processes with applications

John Wiley & Sons, Ltd., 1990, xiii + 672, 0-471-84272-2, £ 55.15

De Finetti

Theory of probability, volume 1

John Wiley & Sons Ltd., 1990, xix + 300, 0-471-92611-6, £ 19.95

De Finetti

Theory of probability, volume 2

John Wiley & Sons Ltd, 1990, xviii + 375, 0-471-92612-4, £ 19.95

Conrad, S. (ed.)

Assignments in applied statistics

John Wiley & Sons Ltd., 1989, xv + 273, 0-471-92281-1, £ 28.50

Hosmer, D.W. and Lemeshow, S.

Applied logistic regression

John Wiley & Sons, Ltd., 1990, xiii + 307, 0-471-61553-6, £ 36.00

Ochi, M.K.

Applied probability and stochastic processes

John Wiley & Sons, Inc., 1990, xvi + 499, 0-471-85742-4, £ 54.65

Bunke, H. & Bunke, O. (editors)

Nonlinear regression, functional relations and robust method

John Wiley & Sons, 1989, 449, 0-471-91239-5, £ 33.00

Upton, G.J.G. & Fingleton, B.

Spatial data analysis by example, volume 2

John Wiley & Sons, Ltd., 1989, xi + 416, 0-471-92086-X, £ 49.95

Bowman, K.O. & L.R. Shenton

Continued fractions in statistical applications

Marcel Dekker, Inc., 1989, x + 330, 0-8247-8120-1, \$ 107.50

Berry, D.A. (ed.)

Statistical methodology in the pharmaceutical sciences

Marcel Dekker, Inc., 1990, viii + 578, 0-8247-8117-1, \$ 150.00

Ratkowsky, D.A.

Handbook of nonlinear regression models

Marcel Dekker, Inc., 1990, ix + 241, 0-8247-8189-9, \$ 83.50

Scheaffer, R.L., Mendenhall, W. en Ott, L.

Elementary survey sampling

PWS-Kent Publishing group, 1990, x + 390, 0-534-92185-X, £ 16.95

Bowerman, B.L. & O'Connell, R.T.

Linear statistical models, an applied approach

PWS-Kent Publishing Company, 1990, xiv + 1024, 0-534-98152-6, £ 18.95

Ott, L. en Mendenhall, W.

Understanding statistics

PWS-Kent publishing group, 1990, xvii + 727, 0-534-98169-0, £ 15.95

Bain, L.J. en Engelhardt, M.

Introduction to probability and mathematical statistics

PWS-Kent Publishing Company, 1989, ix + 565, 0-534-98145-3, £ 16.95

Scheaffer, R.L.

Introduction to probability and its applications

PWS-Kent Publishing Company, 1990, x + 363, 0-534-91970-7, £ 16.95

Scheaffer, R.L. en McClave, J.T.

Probability and statistics for engineers

PWS-Kent publishing group, 1990, viii + 696, 0-534-92184-1, £ 15.95

Nijdam, B. en H. van Buuren

Statistiek voor de economische wetenschappen

Samson Uitgeverij b.v., 1989, xiii + 403, 90-14-03696-5,

Lindsey, J.K.

The analysis of categorical data using GLIM

Springer-Verlag Berlijn Heidelber, 1989, v + 168, 3-540-97029-0, DM 36.00

Gleser L.J., Perlman M.D., Press S.J., Sampson A.R

Contributions to probability and statistics for Ingram Olkin

Springer-Verlag, Berlijn Heidelberg, 1989, x + 505, 3-540-97076-2, DM 142.0

Shah, K.R. en Sinha, B.K.

Theory of optimal designs

Springer-Verlag, Berlijn Heidelberg, 1989, viii + 171, 3-540-96991-8, DM 36.00

Tong, Y.L.

The multivariate normal distribution

Springer-Verlag, Berlijn Heidelberg, 1990, xiii + 271, 3-540-97062-2, DM 98.00

Gupta, S. en Mukerjee, R.

A calculus for factorial arrangements

Springer-Verlag, Berlijn Heidelberg, 1990, vi + 126, 3-540-97172-6, DM 30.00

Fomby, T.B. en Seo, T.K.

Studies in economics of uncertainty, in honor of Josef Hadar

Springer-Verlag New York Inc., 1990, viii + 231, 0-387-97047-9, DM 98.00

Emery, M.

Stochastic calculus in manifolds

Springer-Verlag New York Inc., 1990, x + 151, 0-687-51664-6, DM 48.00

West, M. en Harrison, J.

Bayesian forecasting and dynamic models

Springer-Verlag New York Inc., 1989, xxi + 704, 0-387-97025-8, DM 118.0

Györfi, L., Härdle, W., Sarda, P. en Vieu, P.

Nonparametric Curve Estimation from time series

Springer-Verlag, Berlijn Heidelberg, 1989, viii + 153, 3-540-97174-2, DM 36.00

Brickman, L.

Mathematical introduction to linear programming and game theory

Springer-Verlag New York Inc., 1990, ix + 130, 0-387-96931-4, DM 68.00

Santner, T.J. en Duffy, D.E.

The statistical analysis of discrete data

Springer-Verlag New York Inc., 1989, xii + 367, 0-387-97018-5, DM 88.00

Reiss, R.-D.

Approximate distributions of order statistics

Springer-Verlag New York Inc., 1990, xii + 355, 0-387-96851-2, DM124.00

Barndorff-Nielsen, O.E., Blaesild, P, Eriksen, P.S

Decomposition and invariance of measures, and statistical transformation models

Springer-Verlag, Berlijn Heidelberg, 1990, i + 147, 3-540-97131-9, DM 36.00

Abrahamse, A.P.J., van Driel, G.J., van Ravenzwaaij, C.

Statistiek (derde herziene druk)

Van Gorkum, 1990, x + 238, 90-323-2560-0, f 45.00

Keller, G., Bartel, H. en Warrack, B.

Statistics for management and economics, 2nd edition

Wadsworth Publishing Company, Inc., 1990, xxiv + 1050, 0-534-12678-2, £ 19.95

Healey, J.F.

Statistics: a tool for social research

Wadsworth Publishing Company, Inc., 1990, xiii + 434, 0-534-12450-X, £ 15.95

Maxwell, S.E. en Delaney, H.D.

Designing experiments and analyzing data

Wadsworth Publishing Company, Inc., 1990, xvi + 902, 0-534-10374-X, £ 19.95

Hamilton, L.C.

Modern data analysis, a first course in applied statistics

Wadsworth Publishing Company, Inc., 1990, xv + 684, 0-534-12846-7, £ 15.95

Berry, D.A. en Lindgren, B.W.

Statistics, theory and methods

Wadsworth Publishing Company, Inc., 1990, xii + 763, 0-534-09942-4, £ 14.95