

STAtOR

periodiek van de VVS jaargang 4 nummer 2 juni 2003

Dynamische visualisatie met meerdimensionele schaling

Tennis en kansrekening

Kwantitatieve modellen en methoden: praktijkgericht
onderwijs en onderzoek aan de TU/e

De zin van statistiek(onderwijs)

Een statisticus is ook een mens

In Memoriam Prof.dr. Freerk Auke Lootsma

Ben ik in beeld? Planning van een observatie satelliet

STATOR

Jaargang 4, nummer 2, juni 2003

STATOR is een uitgave van de Vereniging voor Statistiek en Operationele Research (VVS). STATOR wil leden, bedrijven en overige geïnteresseerden op de hoogte houden van ontwikkelingen en nieuws over toepassingen van statistiek en operationele research. Verschijnt 4 keer per jaar.

Redactie

Dick den Hertog (hoofdredacteur), Wies Akkermans, Martijn Berger, Han Oud, Marc Schuld, Gerrit Stermerdink (eindredacteur), Fred Steutel.

Kopij en reacties richten aan

Prof. dr. ir. D. den Hertog (hoofdredacteur), Faculteit der Economische Wetenschappen van de Universiteit van Tilburg, Postbus 90153, 5000 LE Tilburg, telefoon 013 - 466 2122, <D.denHertog@kub.nl>.

Bestuur van de VVS

A.W. van der Vaart (voorzitter) <aad@cs.vu.nl>; S. J. Koopman (penningmeester) <s.j.koopman@econ.vu.nl>; namens de Bedrijfssectie (BDS) P. Banens <banens@cqm.nl>; namens de biometrische sectie (BMS) A. Stein <alfred.stein@wur.nl>; namens de economische Sectie (ECS) P.H.F.M. van Casteren <casteren@fee.uva.nl>; namens het Ned. Genootschap voor Besliskunde (NGB) H. Fleuren <fleuren@uvt.nl>; namens de Sectie Mathematische Statistiek (SMS) P. Spreij <spreij@science.uva.nl>; namens de Sectie Statistische Programmatuur (SSP) S.H. Heisterkamp <sh.heisterkamp@rivm.nl>; namens de Sociaal Wetenschappelijke Sectie (SWS) C. Glas <c.a.w.glas@edte.utwente.nl>.

Leden- en abonnementenadministratie van de VVS

VVS, Postbus 2095, 2990 DB Barendrecht, telefoon 0180 - 623796, fax 0180 - 623670, e-mail <admin@vvs-or.nl>. Raadpleeg onze website over hoe u lid kunt worden van de VVS of een abonnement kunt nemen op STATOR of op een van de andere periodieken.

VVS-website

<http://www.vvs-or.nl>

Advertenties

Contactpersoon: Rita Oomen, telefoon 0167 - 563401, fax 0167 - 561200, <japm.oomen@worldonline.nl>. Uiterlijk vier weken voor verschijnen te zenden aan Pharos, Moeflonstraat 5, 6531 JS Nijmegen, telefoon 024 - 3559214, e-mail <hooategem@xs4all.nl>. STATOR verschijnt in maart, juni, september en december.

Ontwerp en opmaak

Pharos / M. van Hooategem, Nijmegen

Uitgever

© Vereniging voor Statistiek en Operationele Research
ISSN 1567-3383

Inhoud

3

Quo vadis?

Dick den Hertog

4

Dynamische visualisatie met meerdimensionale schaling.

Patrick Groenen en Pieter Jan Stappers

11

Tennis en kansrekening. Column.

Onno Boxma

13

Kwantitatieve modellen en methoden: praktijkgericht onderwijs en onderzoek aan de TU/e. Wat is het belang voor de praktijk?

Mynt Zijlstra en Leonard Fortuin

19

De zin van statistiek (onderwijs).

Wim van Nieuwenhuysen en Ton Kuylen

24

Een statisticus is ook een mens. Column.

Fred Steutel

26

In Memoriam Prof. dr. Freerk Auke Lootsma.

Roger Cooke en Kees Roos

27

Ben ik in beeld? Planning van een observatie satelliet.

Eelco Kuipers

30

Agenda



QUO VADIS?

De afgelopen maanden waren spannende tijden voor de redactie. Wat zal er uit de enquête komen? Waarderen de lezers het blad *STaTOR*? We zijn blij dat de lezers die de enquête hebben ingevuld over het algemeen erg positief zijn over ons blad. Dat was echt een hart onder de riem voor de redactie.

Dat betekent niet dat er niets te verbeteren valt. De uitkomsten van de enquête bieden genoeg stof voor verbeteringen. Zelf zou ik bijvoorbeeld graag (nog meer) artikelen in *STaTOR* willen zien die de huidige beoefening van statistiek en OR kritisch onder de loep nemen. In *STaTOR* staan voornamelijk succesverhalen op het gebied van statistiek en OR. Op die successen wil ik niets afdingen, maar een kritische analyse blijft vaak achterwege. Sluit het statistiek- & OR-onderwijs aan bij de praktijk en de trends in de maatschappij? Zijn studenten voldoende bekend met de beroepspraktijk en raken ze door de praktijk gemotiveerd? Springen we wat ons onderzoek betreft voldoende in op nieuwe ontwikkelingen? Waar heeft de moderne maatschappij werkelijk behoefte aan? Zijn we wel voldoende zichtbaar? Enzovoorts.

Daarom ben ik erg blij met de bijdrage van Wim van Nieuwenhuysen en Ton Kuylen 'De zin van (statistiek)onderwijs' in het nummer dat voor u ligt. In dit artikel wordt stevige kritiek geuit op het huidige statistiekonderwijs. Als de auteurs gelijk hebben, betekent dat nogal wat. Ik zou het prachtig vinden als hierover verder gediscussieerd zou worden in *STaTOR*. Schroom dus niet om te reageren!

Rest mij nog om u, als u de enquête nog niet hebt ingevuld, aan te moedigen om dit alsnog te doen. Mocht u het formulier kwijt zijn, dan kunt u de enquête ook vinden op de VVS-website of u kunt mij erom vragen. In het volgende nummer zullen we de uitkomsten van de enquête uitvoerig bespreken.

Veel leesplezier!

Dick den Hertog
hoofdredacteur



Foto: Pieter Bosch

Dynamische visualisatie met meerdimensionele schaling

PATRICK J.F. GROENEN EN PIETER JAN STAPPERS

In de statistiek wordt vaak gebruik gemaakt van visualisatie om op een gemakkelijke wijze statistische informatie over te brengen. Met de huidige snelheid van computers is het gemakkelijk geworden om dynamiek en interactie toe te voegen, waardoor een beter inzicht of gemakkelijker inzicht verkregen kan worden in complexe relaties. Hier bespreken Patrick Groenen (Econometrisch Instituut, Erasmus Universiteit Rotterdam) en Pieter Jan Stappers (Industrieel Ontwerpen, TU Delft) dynamische en interactieve meerdimensionele schaling, toegepast op verschillen tussen politieke partijen en gelijkenissen tussen rollerskates. Zij betogen hoe deze techniek tot nieuwe vormen en toepassingen van statistische presentatie kan leiden.

Het is tegenwoordig niet moeilijk om aan gegevens te komen. Wat wel moeilijk blijft is om de belangrijkste relaties in de data op een gemakkelijke manier te ontdekken. Visualisatie van de gegevens kan daarbij een belangrijk hulpmiddel zijn. Daarom is er in de statistiek en daarbuiten meer en meer aandacht voor visualisatietechnieken. Hier richten we ons op een van die technieken, namelijk meerdimensionele schaling (MDS). Deze techniek bestaat al sinds begin jaren zestig van de vorige eeuw en is bijzonder geschikt gebouwd om relaties in een collectie van vergelijkbare elementen weer te geven. Meestal wordt hiervoor een kaart gebruikt waarin elementen dicht bij elkaar afgebeeld worden als ze op elkaar lijken en veraf als ze juist niet op elkaar lijken. MDS is een ruimtelijk plaatje waarin de plaatsing van de elementen hun belangrijkste relaties 'verraadt'. Met de huidige grafische en rekenmogelijkheden van computers kunnen we echter verder gaan. Het plaatje kan bewegen en de gebruiker kan het plaatje manipuleren, en krijgt daarbij feedback waarbij de relaties in de data zich rijker en inzichtelijker manifesteren. Op deze manier wordt de 'lezer' intensiever betrokken bij de data en uitgedaagd om zelf te exploreren. Om dit te realiseren moet de techniek naar de achtergrond treden, zodat de lezer zich beter kan concentreren op de data zelf en de relaties daarin. Naar onze mening leent MDS zich uitstekend voor het inzichtgenererend 'spelen' met data. Aan de hand van twee toepassingen laten we zien hoe kaarten verkregen met MDS gemakkelijk inzicht geven in twee alledaagse toepassingen.

Het politieke landschap voor de 2e-kamer verkiezingen 22 januari 2003

De eerste toepassing vindt zijn oorsprong bij de 2^e-kamer verkiezingen van januari 2003. De vraag is hoe de programma's van politieke partijen van elkaar verschillen. Voor veel kiezers is de stemwijzer (<http://www.stemwijzer.nl>) een belangrijk

hulpmiddel geweest. Ongeveer 2,25 miljoen kiezers hebben van de stemwijzer gebruik gemaakt om hun keuze te bepalen. De stemwijzer bestaat uit dertig stellingen over concrete onderwerpen die op dat moment speelden in de politiek. Zo luidde een stelling: 'Nederland moet de Verenigde Staten steunen bij een militaire aanval op Irak'. De gebruiker van de stemwijzer geeft bij elke stelling aan of hij of zij het eens is, oneens is of geen mening heeft over de stelling. De grootste tien partijen hebben dezelfde dertig stellingen beantwoord aan de hand van hun programma. Om tot een stemadvies te komen vergelijkt de stemwijzer de antwoorden van de kiezer met die van de politieke partijen. Het resultaat is een sortering van de partijen, waarbij de partij wier antwoorden het meest met die van de gebruiker overeenkomen bovenaan staat.

De stemwijzer geeft weliswaar antwoord op de vraag hoe dicht een kiezer bij een partij ligt, maar het geeft geen inzicht in hoeverre de partijen onderling verschillen. De gebruiker wordt dan ook aangeraden zelf verder te kijken, omdat de dertig vragen natuurlijk niet alles perfect konden dekken. De antwoorden bieden echter meer houvast dan alleen het gesorteerde lijstje. We kunnen een kaart maken van het politieke landschap aan de hand van de antwoorden van de partijen op de vragen van de stemwijzer. Om zo'n kaart te maken, moeten we allereerst weten hoe verschillend de partijen van elkaar zijn. Een manier om dat te doen is als volgt. We kunnen elke partij als een punt beschouwen in een ruimte waarin elke stelling een aparte dimensie vormt. De coördinaten van een partij worden dan simpelweg gegeven door het antwoord op elk van de dertig stellingen, waarbij 'mee eens' gecodeerd is als 1, 'geen mening' als 0, en 'mee oneens' als -1. Met andere woorden, elke partij is een punt in een dertigdimensionele ruimte. Nu heeft een computer geen moeite om punten in een 30-dimensionale ruimte te op te slaan, maar 30-dimensionaal visualiseren

is niet te doen. De meeste kaarten zijn twee-dimensionaal. We kunnen wel een 2-dimensionaal plaatje maken dat de afstanden tussen de 10 partijen zo goed mogelijk weergeeft. Die ‘gewone’ Euclidische afstanden tussen de partijen berekenen we in dertig dimensies. De resultaten staan in Tabel 1. Twee partijen die precies dezelfde antwoorden op alle stellingen hebben, zouden een afstand 0 krijgen; partijen die precies tegenovergesteld de stellingen beoordelen een afstand 10.95 (de vierkantswortel van dertig maal een verschil van 2). De tabel toont dat de extremen niet voorkomen: het grootste verschil (tussen VVD en GroenLinks) is 9.64, en de grootste overeenkomst (tussen SP en GroenLinks) is 3.74.

Met een MDS-procedure kan deze tabel vertaald worden in een kaart van het politieke landschap in twee dimensies (zie Figuur 1). Deze techniek vertaalt gelijkenissen tussen objecten in een kaart die zeer eenvoudig te interpreteren is: twee partijen die veel gelijkenis vertonen liggen dicht bij elkaar, twee objecten die weinig gelijkenis vertonen liggen ver van elkaar. De extreme verschillen tussen GroenLinks en SP enerzijds en VVD anderzijds komen duidelijk tot uiting. Opvallend is dat de partijen van het kabinet Balkenende I dicht bij elkaar liggen. Op zich is dat niet zo verwonderlijk omdat hun verkiezingsprogramma's



Figuur 1. MDS-Kaart van het politieke landschap voor de 2e-kamer verkiezingen van 22 januari 2003 gebaseerd overeenkomsten en verschillen uit Tabel 1 tussen de partijen op de dertig stellingen van de stemwijzer.

niet zo ver aflaggen van het regeringsbeleid dat ze gevoerd hebben voorafgaande aan de verkiezingen. Verder zien we dat er drie blokken te onderscheiden zijn: een confessioneel blok (CDA, SGP en Christenunie), een progressief blok (PvdA, D66, Groenlinks en de SP) en een conservatief liberaal blok (VVD, LN en LPF). Partijen die het minst van elkaar verschillen zijn Groenlinks en de SP. Ook is

	CDA	LPF	VVD	PvdA	GL	SP	D66	CU	SGP	LN
CDA	o									
LPF	6.08	o								
VVD	5.57	4.24	o							
PvdA	7.14	7.62	8.49	o						
GL	8.37	8.89	9.64	4.80	o					
SP	8.83	8.89	9.54	5.74	3.74	o				
D66	7.42	6.78	7.87	4.90	5.57	6.08	o			
CU	6.71	7.21	7.48	6.16	6.71	6.86	6.32	o		
SGP	5.83	6.71	6.56	6.71	8.12	8.25	7.68	4.58	o	
LN	6.00	4.36	5.39	6.40	7.87	8.12	6.08	7.14	6.93	o

Tabel 1. Ongelijkheden tussen politieke partijen berekend als Euclidische afstanden tussen de partijen op dertig stellingen uit de stemwijzer voor de 2e-kamer-verkiezingen van 22 januari 2003. Grootste en kleinste afstanden zijn vet aangegeven.



Figuur 2. Kaart van het politieke landschap verkregen met dynamische MDS als alleen gekeken wordt naar de verschillen tussen de partijen op financiële stellingen van de stemwijzer.

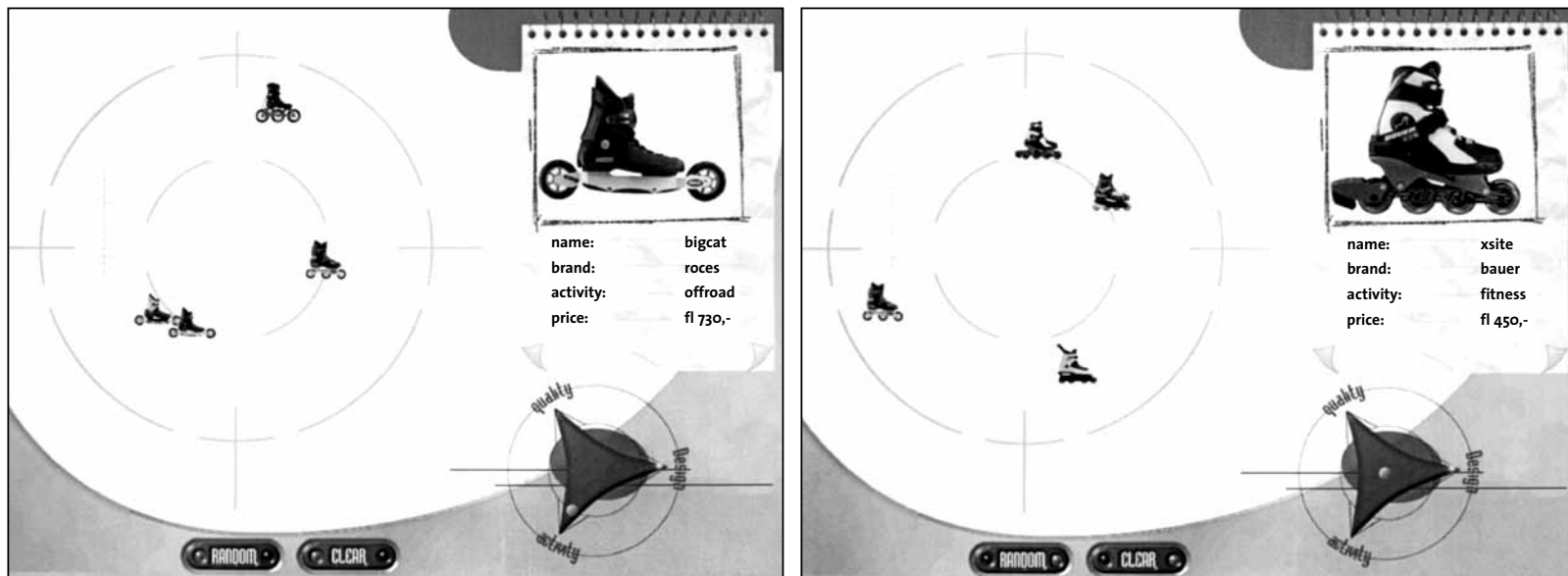
er behoorlijke overeenkomst tussen LPF en VVD en tussen LPF en LN.

Het politieke landschap op financieel gebied

Tot nu toe hebben we niets nieuws gedaan. MDS wordt al zo'n 40 jaar gebruikt om reeksen relaties samen te vatten in een 2D of 3D kaart. De kracht van *dynamische* MDS is dat verschillende aspecten ná elkaar of tegelijk benadrukt kunnen worden. Om dit te demonstreren hebben we de dertig stellingen onderverdeeld in zeven verschillende beleidsterreinen: veiligheid, asielbeleid, financieel beleid, sociaal beleid, buitenland, milieu beleid en overig beleid. Vóór het maken van de kaart kan aan elk van die beleidsterreinen een gewicht toegerekend worden. Dat gewicht wordt bij het berekenen van de afstanden gebruikt voor alle stellingen die bij het betreffende beleidsterrein horen. Met dynamische MDS kan de gebruiker zelf de gewichten wijzigen, en probeert de computer de tabel met gewogen afstanden zo goed mogelijk te

benaderen met de afstanden tussen de partijen op de kaart. Figuur 2 toont het politieke landschap als alleen naar de overeenkomsten en verschillen gekeken wordt op financiële stellingen. We zien dat er een grote afstand is tussen de drie grote partijen (CDA, PvdA en VVD). Opvallend is dat er kennelijk redelijke overeenkomst is in het financieel beleid tussen de progressieve partijen SP, Groenlinks, D66 en PvdA en de Christenunie omdat ze allemaal dicht bij elkaar liggen. Op vergelijkbare wijze kan er voor elk beleidsterrein of combinaties daarvan een politieke kaart gemaakt worden.

Dynamische MDS combineert de eenvoud van een kaart (liggen partijen dichtbij elkaar, dan lijken ze op elkaar, liggen ze ver weg dan lijken ze niet op elkaar) met de dynamiek die optreedt door veranderingen door de gebruiker. Het moedigt aan om de gegevens van verschillende kanten te onderzoeken en geeft daarmee extra inzicht in de data, een inzicht dat anders moeilijk op te sporen is in de tabel of één enkel statisch totaalplaatje.



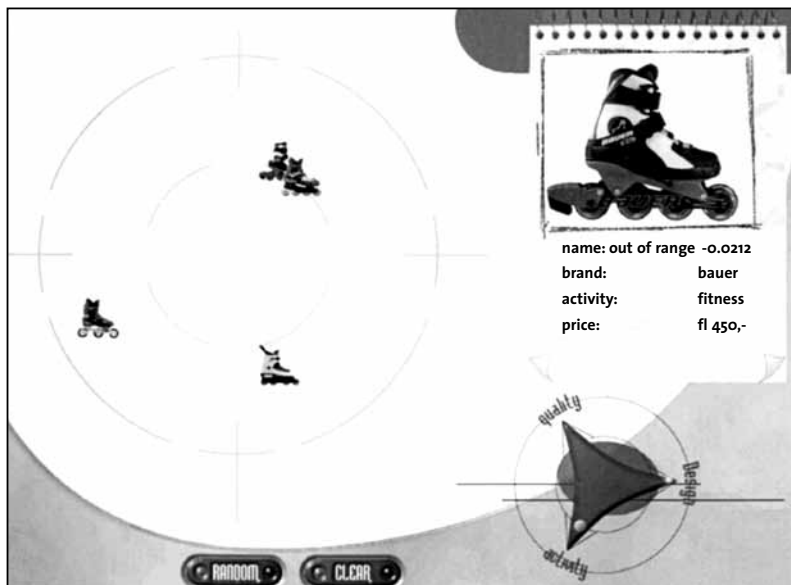
Figuur 3. Kaart van verschillen tussen vier aselekt gekozen rollerskates (links) verkregen met dynamische MDS. In het linker paneel tellen alle eigenschappen even zwaar. Het rechter paneel geeft de kaart van dezelfde vier rollerskates waarbij verschillen die te maken hebben met het type activiteit zwaarder meetellen dan verschillen in kwaliteit of ontwerp.

Dynamische MDS voor de keuze van rollerskates

De inzichtelijkheid van kaarten van collecties elementen kan ook ingezet worden voor alledaagse vragen waar mensen zich op een onbekend terrein wagen. Stelt u zich eens voor dat u rollerskates wil kopen. Er blijkt een enorme hoeveelheid aan verschillende rollerskates beschikbaar te zijn. De rollerskates verschillen in kwaliteit, prijs, design, type gebruik, enzovoorts. Dynamische MDS kan nu gebruikt worden om uw keuzeproces op interactieve wijze te ondersteunen. Figuur 3 laat de interface zien van dynamische MDS voor rollerskates. Het witte gedeelte geeft de kaart weer. Interpretatie van de kaart is als voorheen: rollerskates die dicht bij elkaar liggen lijken op elkaar, zij die ver weg van elkaar geplaatst zijn lijken niet op elkaar. Het notatieblok laat een paar belangrijke gegevens zien van één van de rollerskates die daarvoor geselecteerd is. De gegevens die ten grondslag liggen aan deze kaart vormen een database van 70 rollerskates, waarin voor elke

rollerskate 21 eigenschappen zijn genoteerd. Deze eigenschappen zijn onderverdeeld in drie groepen (activiteit, kwaliteit en ontwerp). Zo'n groep eigenschappen kan benadrukt worden door het kleine balletje in de driehoek rechtsonder naar een van de hoekpunten te bewegen. Zo is de kaart in het rechterpaneel van Figuur 3 verkregen door meer nadruk te leggen op de overeenkomsten en verschillen van de activiteit eigenschappen van de rollerskates.

Om een goede keuze van rollerskates te kunnen maken moeten we verschillende selecties van rollerskates kunnen laten zien. Hiervoor voegen we twee extra faciliteiten toe aan dynamische MDS. Het resultaat heet MDS-Interactive. Ten eerste kan een rollerskate uit de kaart verwijderd worden door hem weg te slepen uit het witte gebied van de kaart. Ten tweede kunnen rollerskates toegevoegd worden aan de kaart. Dit laatste kan aselekt gebeuren door op de knop 'random' de drukken. Een andere mogelijkheid is om op een lege plek in de kaart te klikken. MDS-



Figuur 4. Kaart met drie andere rollerskates voor off-road gebruik verkregen door drie keer te klikken nabij de linker rollerskate in Figuur 3 rechts en vervolgens de drie rechter rollerskates daarin weg te slepen.

Interactive zoekt dan naar een nog niet getoonde rollerskate die ongeveer op die plek zou moeten liggen. Het toevoegen van een nieuwe rollerskate heeft tot gevolg dat de kaart dynamisch verandert totdat een stabiele oplossing is verkregen.

Stel dat u geïnteresseerd bent in offroad rollerskates voor hobbelige wegooppervlakken. U start met de situatie van Figuur 3 rechts. De meest linker rollerskate hierin heeft drie wieltjes en is geschikt voor offroad gebruik. U klikt enkele malen nabij deze rollerskate waardoor dynamische MDS een paar andere offroad rollerskates toevoegt. Vervolgens sleept u de drie rollerskates die al eerder op de kaart stonden weer weg. De aldus verkregen kaart staat in Figuur 4. Op eenvoudige wijze heeft u nu ingezoomd op rollerskates voor offroad gebruik. We zien in Figuur 4 dat de beide rollerskates met twee wieltjes het meest op elkaar lijken en de rollerskate aan de bovenkant het meest verschillend is van de anderen. Om tot een verdere aanscherping van de keuze te komen kan het nodig zijn het proces van selectie, inzoomen en wegslepen enige malen te herhalen. De kracht van dynamische MDS ligt er juist in dat dit proces intuïtief en gemakkelijk is.

Wondermiddel?

Met dynamische MDS is een inzichtelijk zoekproces mogelijk. De gebruiker start het zoeken met concrete voorbeelden van mogelijke oplossingen. Sommige eigenschappen (bijvoorbeeld kleur en vorm) zijn visueel te herkennen, andere vereisen dat de gebruiker kijkt naar de lijst eigenschappen in het notitieblok. Wanneer de keuze verder is toegespitst, zal het soms nodig zijn dat de gebruiker zich verder verdiept in de betekenis van de eigenschappen. Dynamische MDS helpt deze kennisvraag te doseren. Als bijvoorbeeld alle offroadschaatsen hetzelfde type lagers gebruiken, dan hoeft de gebruiker zich niet verder te verdiepen in de voor- en nadelen van verschillende lagers: het heeft immers geen relevantie voor de keuze die hij of zij wil maken.

Uiteraard kunnen dezelfde ideeën ook gebruikt worden voor andere productgroepen. Dynamische MDS is een veelzijdige techniek die meer inzicht kan geven. De techniek steunt op het visualiseren van gelijkenisgegevens, interactie met de gebruiker en dynamische weergave van de aanpassingen door de gebruiker. Op de website <http://studiolab.io.tudelft.nl/mdsi/> staat een reeks voorbeel-

den van MDS-Interactive interfaces, variërend van online TV-gidsen tot jaarverslagen. Sommige hiervan zijn ook interactief uit te proberen of als demo te downloaden.

Natuurlijk blijft de kwaliteit van de dataset een eerste vereiste. Als belangrijk aspecten, zoals de prijs van de rollerskate, ontbreken dan werken dynamische en interactieve MDS niet inzichtelijk, of wellicht zelfs misleidend. Zo zou natuurlijk ook de stempijzer misleiden als er enkel vragen over financieel beleid in waren opgenomen en we hier de resultaten zouden presenteren als een representatief totaaloordeel. Wel willen we opmerken dat MDS ongevoelig is voor kleine veranderingen in de meegenomen eigenschappen. Bovendien zijn de visualisaties tamelijk voorzichtige suggesties in vergelijking tot bijvoorbeeld tabellen met numerieke waarden zoals Tabel 1. De ervaring leert ons dat gebruikers vergevingsgezind zijn als er een enkele keer een onverwacht of onbegrepen resultaat komt. Het is vergelijkbaar met de verkoper in een schoenenzaak die met een paar schoenen komt waarvan hij dacht dat ze u zouden bevallen. Omdat de gebruiker zijn vragen niet expliciet hoeft te beschrijven tolereert hij of zij ook een enkele verrassing.

Naar een nieuw soort statistische visualisatie

We kunnen tegenwoordig meer doen met computers. Ging het vroeger vooral om het rekenwerk, nu kan de computer ook ingezet worden voor animatie en interactiviteit. Juist deze aspecten zijn goed te gebruiken in de statistiek. Naast de onderliggende statistische technieken bestaat de visualisatie-uitdaging erin om het geheel in een consistente en inzichtelijke vormgeving te brengen.

Typerend is ook dat dynamische en interactieve MDS tot *vragen* leidt, niet alleen tot antwoorden op reeds precies uitgekristalliseerde vragen. Het vinden van de juiste vraag is vaak een moeilijker probleem dan het beantwoorden van die

vraag. Dynamische en interactieve technieken met een op de gebruiker toegespitste vormgeving kunnen hierbij een waardevolle bijdrage leveren. Interactieve statistische visualisatie maakt het mogelijk voor de 'lezers' om te 'spelen' met de data, en zo gevoel te krijgen voor de relaties in de data. Deze vormen van exploratieve interactie zijn tot voor kort voorbehouden gebleven aan experts, die bijzonder vertrouwd moesten zijn met de statistische rekensoftware. De voorbeelden die we hier besproken hebben geven aan hoe dergelijke exploratie ook binnen het bereik van een breder publiek kan komen dat tot voor kort slechts passief de visualisaties op papier tot zich kon nemen. Het succes van spreadsheetprogramma's als Bricklin's VisiCalc, waarmee 'gewone' gebruikers zonder complexe programmeertalen voor het eerst kleine rekenexperimenten konden uitvoeren, geeft aan hoe een beetje interactiviteit tot meer begrip kan leiden. Interactieve vormen van MDS bieden op een vergelijkbare manier mogelijkheden om inzichten te verkrijgen uit data zonder de toch complexe tussenkomst van geformaliseerde statistiek-pakketten. De vormgeving en het ontwerp van dergelijke interactieve visualisatie-hulpmiddelen is echter niet triviaal. Het gereedschap moet ervoor zorgen dat de gebruiker niet misleid wordt door de visualisaties die hij of zij maakt vanuit de data. Gezien de enorme vloed aan informatie en keuze die de moderne burger over zich uitgestort krijgt bestaat er een maatschappelijke vraag, en ook een markt, voor het ontwikkelen van dergelijke interactieve visualisaties. Statistiek staat op de kaart, en zij beweegt.

PATRICK J. F. GROENEN is hoogleraar statistiek bij het Econometrisch Instituut aan de Faculteit Economische Wetenschappen, Erasmus Universiteit Rotterdam.
E-mail: <groenen@few.eur.nl>.

PIETER JAN STAPPERS is hoogleraar ontwerptechnieken, Faculteit Industrieel Ontwerpen, TU Delft.
E-mail: <P.J.Stappers@IO.TUdelft.nl>.

Tennis en kansrekening

ONNO BOXMA

De hoofdredacteur van *STAtOR* belde me op met de vraag of ik columnist van zijn blad wilde worden. Eindelijk, eindelijk! Ik onderdrukte met moeite een jubelkreet, en mijn ogen werden vochtig. Daar doe je het immers allemaal voor. Ik belde mijn moeder, die het direct aan mijn stem hoorde: 'Jongen, nee toch, heeft *STAtOR* je gevraagd?'

Twee dagen later kwam de bezinning. De hoofdredacteur, bedreven in acquisitie, had aan het eind van het gesprek fijntjes opgemerkt dat de deadline voor het volgende nummer al over ruim een week was, en dat hij me dat toch eigenlijk niet aan kon doen - waarop ik de uitdaging gretig was aan-

gegaan. Maar wat zou de inhoud van de column worden? Een diepzinnig betoog over besliskunde? Een vlammende polemiek tegen die paar middelbare scholieren die nog niet voor een wiskundestudie hebben gekozen? Het papier voor mij werd steeds leger. Uiteindelijk besloot ik terug te vallen op een paar onderwerpen die me heel dierbaar zijn: kansrekening, tennis en onderwijs. Bij colleges over Markovketens probeer ik de studenten uit te leggen hoe de intreekansen in de diverse absorberende toestanden kunnen worden berekend. Als illustratie licht ik dan een tipje op van de sluier over mijn ongeëvenaarde successen als tennisser. Dat gaat ongeveer als volgt.

Het geheim

De mensen vragen me vaak: 'Wat is toch het geheim van je tennissucces?' En dan worden legendarische prestaties opgerakeld zoals het bereiken van de kwartfinale van de sterk bezette 'Gerwense kampioenschappen' (en begin nu niet weer te zeuren over die bye en walk-over). Welnu, het geheim van mijn succes is niet alleen een buitengewone techniek – zo buitengewoon dat tegenstanders altijd weer verrast worden als de bal het net nog haalt. Het is ook meer dan een wedstrijdmentaliteit, vergeleken waarmee die van Edgar Davids bijna meelijwekkend aandoet, of een mate van zelfoverschatting die (mogen we in alle bescheidenheid stellen) fenomenaal is. Nee, er is meer. En omdat altijd maar winnen ook eens gaat vervelen, zal ik u deelgenoot maken van de speciale kennis achter 'het geheim'.

Stel dat u gemiddeld 1 van de 10 slagenwisselingen van mij wint - ongetwijfeld een optimistische inschatting - en dat u toch een keer in een game op 40-40 komt. Wat is dan de kans dat u die game wint? Het teleurstellende antwoord is: die kans is slechts één op de tweeëntachtig. Immers, nadat het misschien nog één of meer keren deuce is geworden, eindigt de game of doordat u twee punten op rij wint (kans $1/10$ keer $1/10$, dus één op de honderd) of doordat ik twee punten op rij win (kans $9/10$ keer $9/10$, dus 81 op de honderd). Slechts één op de tweeëntachtig gevallen waarbij de game eindigt is gunstig voor u! In het onwaarschijnlijke geval dat u gemiddeld één op de drie slagenwisselingen van me wint (doordat het gips om de arm me nog teveel hindert), en het weer deuce is, wint u met kans $1/5$ de game. Algemener: als u een fractie p van de rally's tegen mij pleegt te winnen, en het staat 40-40, dan valt uit het voorgaande snel in te zien dat u een kans $p^2/(p^2 + (1-p)^2)$ hebt om de game winnend af te sluiten.

Dit soort verschijnselen is kenmerkend voor de tennistelling. Niveauverschillen worden flink vergroot – althans per service game; natuurlijk zijn

de kansen in werkelijkheid verschillend voor het geval u serveert of ontvangt. Op Wimbledon wint een serveerder ongeveer 60% van de punten, wat winst in bijna driekwart van de servicegames oplevert. Als u in elke rally winstkans 0.55 zou hebben, dan zou u in een tiebreak een winstkans van ruim 0.65 hebben, en zou u ruim 62% van de games winnen en meer dan 80% van de sets!

Voor de didactici onder de lezers kan ik er nog aan toevoegen, dat de bijbehorende Markovketen als toestandsruimte de punten -2, -1, 0, +1 en +2 heeft, waarbij 0 correspondeert met 40-40, en -2 en +2 absorberend zijn ('game over'). Het is niet moeilijk een stelsel van drie vergelijkingen op te stellen voor de absorptiekansen in +2, startend in de toestanden -1, 0 en +1. Overigens is eenvoudig na te gaan dat de gemiddelde intreetijd vanuit 0 in een absorberende toestand gelijk is aan $2/(p^2 + (1-p)^2)$.

Voor de tennissers onder de lezers geef ik nog aan hoe het voorgaande hen kan helpen om eenvoudig eens een wedstrijd te winnen. Heel belangrijk is het mentale aspect. Loop na 6-2 verlies van de eerste set niet rond met zo'n gezicht als op de poster 'ongewenst zwanger, wat nu?', maar bedenk dat u toch nog wel 20 van de 50 punten hebt gewonnen. Met een beetje meer inspanning en overleg worden dat er 27 van de 50 in de volgende sets en is de winst binnen handbereik.

Verder is er de tactiek tegen de sterkere tegenstander. Rommel zonder al te veel inspanning lekker aan op zijn service, enkele bewonderende kreten slakend. Maar speel punten die u écht moet winnen (uw service games, een tiebreak, of die enkele keer dat het op zijn service 40-40 wordt) extra geconcentreerd, met de slagen die u het beste liggen. Komt in die games uw kans p ruim boven de helft uit, dan gaat u winnen. Reken maar!

Onno Boxma is hoogleraar Stochastische Besliskunde bij de Faculteit Wiskunde en Informatica van de TU Eindhoven. E-mail: <boxma@win.tue.nl>.



KWANTITATIEVE MODELLEN EN METHODEN

Praktijkgericht onderwijs en onderzoek aan de TU/e

Wat is het belang voor de praktijk?

MYNT ZIJLSTRA EN LEONARD FORTUIN

TU/e, de Technische Universiteit Eindhoven, verricht academisch onderzoek en biedt academisch onderwijs dat gericht is op de praktijk. Die praktijkgerichte doelstelling heeft de TU/e vooral bij het onderwijs over en onderzoek aan kwantitatieve modellen en methoden (KM). Wat komt daarvan terecht? Wat zijn de jongste ontwikkelingen? Wat mogen we in de toekomst verwachten? Wat betekenen de ICT-ontwikkelingen voor de toekomst van ons vakgebied? Hoe gaat de academische wereld daar mee om? De antwoorden zijn interessant voor beoefenaars van KM in de praktijk, dus voor adviseurs en stafmedewerkers in bedrijven met een KM-achtergrond. Immers, zij proberen meerwaarde te realiseren voor hun opdrachtgevers door het creatief toepassen van KM. Een goede reden om vanuit de wereld van die toepassers op bezoek te gaan bij de TU/e en daar een vijftal hoogleraren vragen voor te leggen over zaken van wederzijds belang.

De vijf gesprekspartners

PROF.DR. E.H.L. AARTS is lid van de directie van Philips Research en hoogleraar aan de faculteit Wiskunde en Informatica van de TU/e. Tevens is hij adviseur bij CQM.

PROF.DR. P.L. CIJSOUW is programmadirecteur van de TU/e-Postgraduate School (voorheen EUFORCE), een bv voor postacademisch onderwijs aan de TU/e.

PROF.DR. A.G. DE KOK is hoogleraar aan de faculteit Technologie Management van de TU/e. Voordien werkte hij als adviseur bij CQM en als Logistics Innovation Manager bij Philips Consumer Electronics.

PROF.DR. P. VAN DER LAAN was tot september 2001 hoogleraar aan de faculteit Wiskunde en Informatica van de TU/e.

PROF. DR. J. WESSELS was tot februari 2000 hoogleraar aan de faculteit Wiskunde en Informatica van de TU/e.

TU/e verricht academisch onderzoek en biedt academisch onderwijs gericht op de praktijk. Hoe kunnen consultants, met hun contacten in die praktijk, daarbij behulpzaam zijn?

Aarts: 'Adviesbedrijven zie ik als etalages waar studenten kunnen zien wat er in de praktijk te koop is. Studenten zijn probleemoplossend ingesteld en moeten nog leren onderkennen dat de kracht van goede consultants ligt in het *formuleren* van een probleem, omdat daarmee vaak al 80 procent van het probleem is opgelost.'

De Kok: 'Gastcolleges bevorderen de motivatie van de studenten. Praktijkverhalen zijn belangrijk. Maar hun aantal moet beperkt blijven, bij het onderwijs ligt het accent op de theoretische basisvorming. Uiteraard zijn consultants zeer behulpzaam als ze samen met ons afstudeerprojecten begeleiden. Ook zie ik voor consultants een rol bij

het vinden van de juiste bedrijven voor contract-research.'

Wessels: 'Met wetenschappelijk onderzoek aan een universiteit los je geen (praktijk)problemen op. Als je dat toch probeert duurt zo'n onderzoek, zeg, tien jaar. Het probleem is een eigen leven gaan leiden en het onderzoek wordt meer door de methode gestuurd. Ondertussen heeft de klant allang een andere "oplossing" gevonden, bijvoorbeeld met behulp van simulatie. Voor ons is de wisselwerking met de praktijk een bron van inspiratie. Er is de laatste 15 jaar veel onderzoek gedaan op het terrein van de wachttijdtheorie, onder andere met probleemstellingen uit de consultancypraktijk.'

Welke gereedschappen en technieken zijn er aan het KM-arsenaal methoden toegevoegd als gevolg van onderzoek en onderwijs aan de TU/e?

Van der Laan: 'Recente promotie-onderzoeken aan de TU/e hebben resultaten opgeleverd waar de statistische consultant wat mee kan, zowel op het terrein van DOE (*Design of Experiments*) als SPC (*Statistical Process Control*) op niet-parametrische basis. *Experts opinions* modelleren is een recente ontwikkeling waaraan de TU/e heeft gewerkt. Met snelle computers en algebra worden dingen mogelijk die voorheen niet konden.'

De Kok: 'Workflow-managementsystemen. Wij werken aan systematiek en algoritmen voor het valideren van processen die interessant kunnen zijn voor de praktijk. Veel mensen in de praktijk hebben een verkeerd beeld van het aansturen van een *supply chain*. Wij ontwikkelen *blue prints* voor de beheersing van de *supply chain*. Modellen optimaliseren we met behulp van snelle, robuuste algoritmen. De resultaten die nu in de praktijk getoetst worden zijn veelbelovend. Zoals blijkt uit een klantenproject dat TU/e en CQM samen uitvoeren.'

Aarts: 'Het gebied van *intelligent agents*, dat is intelligentie ingebakken in gedistribueerde

omgevingen. Denk aan een *broker agent* die voor jou bepaalt op welk moment welk energiebedrijf voor welke prijs aan jou stroom levert. Hier ligt een vruchtbaar terrein voor optimaliseringstechnieken. Ook zie je in de technische wereld de opkomst van *early prototyping* en *design space exploration*. Je kunt daar veel winnen met specifieke ontwerpgereedschappen waar je compacte modellen mee kunt maken. Het door CQM ontwikkelde gereedschap DOT is daarvan een fraai voorbeeld.'

Wessels: 'De laatste jaren zijn echt vorderingen gemaakt met *statistical learning*. Dat vindt in technische omgevingen makkelijk acceptatie doordat het grenst aan de informatica. Ons onderzoek in de wachttijdtheorie (of breder in de stochastische OR) levert in het algemeen geen makkelijk overdraagbare standaardoplossingen. Daarentegen is er wel veel inzicht en ervaring opgebouwd. We weten nu veel beter dan 10-15 jaar geleden hoe we complexe vraagstukken, bijvoorbeeld in de logistiek en de telecommunicatie, moeten aanpakken, moeten modelleren door ze op te splitsen in deelproblemen. Dat zijn inzichten waarmee de consultants hun voordeel doen. Zo kunnen we sneller en gericht ontwerpoplossingen bedenken. Ook zie ik een uitdaging in het

geïntegreerd toepassen van statistische kennis en kennis uit de OR.'

Hoe kan het bedrijfsleven van die nieuwe gereedschappen en technieken profiteren?

Van der Laan: 'Consultants spelen een sleutelrol om het bedrijfsleven te laten profiteren van de nieuwe gereedschappen en technieken. Naar mijn mening heeft ook de TU/e daarin een taak. Wij organiseren samen met consultants symposia waarin we mensen uit het bedrijfsleven laten zien welke ontwikkelingen en trends er zijn. Door dat samen te doen wek je vertrouwen bij de klanten in de nieuwe zaken.'

De Kok: 'Bedrijven kunnen zeker profiteren van wat wij hier uitvinden in ons onderzoek. Maar ik denk dat ze vooral kunnen profiteren als de consultants een kritisch filter zijn tussen wetenschap en praktijk. Zij zijn de 'inkoper' namens de klant en moeten adviseren wat al wel rijp is voor implementatie bij betrokken klant en wat nog een brug te ver is.'

Cijsouw: 'Om geavanceerde *tools* uit de Statistiek of de OR te laten werken moet je de klanten ermee leren om te gaan en hen enthousiast maken. Het is een leuk idee dat je met geavanceerde methoden van wiskundige aard iets kunt



Prof.dr. E.H.L. Aarts



Prof.dr. P.L. Cijsouw



Prof.dr. A.G. de Kok

doen, maar als je niet kunt uitleggen waarom het werkt, vergeet het dan maar. Goede consultants OR en Statistiek besteden daar veel aandacht aan, ze scheppen een vertrouwensrelatie met de klant.'

Aarts: 'Technologische ontwikkelingen veranderen de economische modellen waarmee zaken gedaan worden. Televisietoestellen zouden net zo in de wereld gezet kunnen worden als nu mobiele telefoons. Het apparaat kost dan bijna niks, je betaalt voor het abonnement. KM stellen je in staat om te analyseren wat je als bedrijf moet doen in zo'n nieuwe wereld. Vroeger was het product zelf de oplossing, maar tegenwoordig gaat het steeds om het oplossen van het integrale probleem van een klant. Philips levert niet alleen lampen aan een voetbalclub voor het verlichten van het stadion. Nee, daar hoort een berekening bij voor de optimale opstelling van lichtmasten en armaturen. De klant krijgt een systeem, een totaaloplossing. Daar liggen enorme mogelijkheden voor modellenbouwers. Dat is reeds bewezen en zal alleen maar toenemen.'

Wessels: 'Voor wiskundigen in de praktijk (consultants) is het lastig om de literatuur bij te houden, omdat ze er nauwelijks tijd voor kunnen vrijmaken. Voor ons aan de TU/e is het *bread and butter* om de literatuur bij te houden en te verwerken in onze colleges en ons onderzoek. Consultancyfirma's krijgen de nieuwe kennis alleen als ze mensen aannemen die pas bij ons vandaan komen en die vervolgens contact blijven houden via gezamenlijke projecten, afstudeerders en dergelijke. Ik ken veel praktijkvoorbeelden, in de logistieke sfeer en in de telecommunicatie, waarmee de betrokkenen absoluut geen raad wisten. Door een frisse inbreng vanuit de KM-hoek kreeg men *grip* op het vraagstuk. Wiskundigen hebben geleerd van de buitenkant naar problemen te kijken en worden niet te zeer belast door de inhoudelijke kant. De kunst is dan om snel ideeën te vinden die wel werken.'

Hoe kan het gebruik van het Internet met KM ondersteund worden?

Van der Laan: 'De *performance* van Internet kan verbeterd worden. Dit is iets voor OR. Ook taalkundig onderzoek is nodig ten behoeve van zoeken, op een bepaald woord, maar ook op combinaties en associaties.'

Wessels: 'Hoe kan je Internet zo ontwerpen dat de klanten een goeie *performance* krijgen? Dat het net niet verstopt raakt of plat gaat, dus de wachtrijproblemen. Wij zijn sterk bezig met ontwerp-problemen rond Internet.'

Hoe kunnen KM-werkers anticiperen op de mogelijkheden van het Internet?

Aarts: 'Belangrijk zijn de behoeften van de klant, dus het Internet gebruiken om de klant dat te leveren wat hij wil. Bij een *supply chain*-ontwikkeling zonder gebruik van het Internet moet je vraagtekens zetten. De KM-werker dient zich af te vragen of hij voldoende kennis in huis heeft van het Internet en van de ontwikkelingen op dat gebied, om de vragen van de klant te kunnen beantwoorden. Dit zou een toets van kwaliteit kunnen zijn!'

De Kok: 'Internet kan overal binnendringen, iedereen kan overal gegevens vandaan plukken. Mogelijk worden *data warehousing* en *data mining*, dat is op zoek gaan naar verbanden die je van tevoren niet kent. Maar welke gegevens heb je echt nodig? Als je niet weet wat je wilt en niet modelleert, dan kom je nooit tot een zinnige uitspraak.'

Van der Laan: 'Voor de gebruikers van KM is Internet een plaats waar ze info kunnen krijgen. Een adviesbureau doet er goed aan daarop in te spelen. Bijvoorbeeld door aardige toepassingen, zonder formules, kort geformuleerd, op het Internet te zetten. Een nieuwe wereld gaat open met het Internet. Maar je verdrinkt gemakkelijk in de gegevens. Wat zou je ermee willen? Een vorm van *data mining* lijkt wenselijk, om wat structuur aan te brengen. Het lijkt zinvol eerst een doel te formuleren en dan te experimenteren. De tijd zal leren of hier nieuwe technieken nodig zijn.'

Wessels: 'Nieuwe KM zijn onmisbaar, want ook de wiskunde ziet allerlei vragen op zich af komen. Bijvoorbeeld deze: hoe organiseren we nu met Internet onze logistiek? Voor Internet beschikbaar kwam, waren er al bedrijven die zeiden: 'Als we onze bestellingen direct kunnen doorgeven aan onze leveranciers, en zelf onze leveranciers in ons productieplan en in onze voorraden laten kijken, hoe moeten we dan de zaak opzetten en wat kunnen we ermee verdienen?' Met Internet kan dat en dat betekent leuk werk voor modelbouwers.'

Zelfmedicatie via het Internet wordt steeds meer een reëel alternatief voor een bezoek aan de dokter. Zullen mutatis mutandis ook probleemeigenaars in het bedrijfsleven hun problemen zelf oplossen door gereedschappen en 'oplossingen' van het World Wide Web te plukken?

Aarts: 'In de KM-wereld verwacht ik net zo'n ontwikkeling als in de medische wereld: de KM-werker zet producten die hij zelf heeft ontwikkeld op het Internet, voor vrij gebruik. Maar de KM-adviseur biedt nog steeds ook specialistische

expertise aan, waarbij de klant contact opneemt vóór er een consult plaats vindt.'

Cijsouw: 'Zelf iets van Internet halen wordt steeds gemakkelijker. Maar meestal gaat dat zo: je gebruikt een pakket en na enige tijd krijg je uitkomsten die niet kloppen. Dan heb je een KM-deskundige nodig. Het Internet zie ik als een prachtig communicatiemiddel, meer niet. Hoogstens leidt het tot meer begrip voor jullie werk. Ze komen wel terug, die klanten, ook als ze eerst zelf iets proberen.'

Wessels: 'Deze ontwikkeling is allang aan de gang. Ooit zaten statistici achter grote elektromechanische rekenmachines om kwadraatsommen uit te rekenen. Statistici verdienden hun brood niet doordat ze zo veel inzicht hadden, maar doordat ze bereid waren dagenlang sommetjes te maken. Computers namen die sommetjes over. Daarna kwamen er standaardpakketten voor statistische analyses. Probleemeigenaars kunnen nu uitzoeken welke methode voor hen geschikt zou zijn. Dan gaat Internet een rol spelen. De statisticus is een adviseur geworden bij complexere problemen en



Prof.dr. J. Wessels



Prof.dr. P. van der Laan

niet meer voor de standaardproblemen. OR volgt dezelfde weg. Bij management-achtige problemen gaan bedrijven vaker zelfstandig dingen proberen, maar uiteindelijk zoeken ze toch weer een adviseur. In de logistiek zie je dat heel duidelijk. Maar de weg is soms moeilijk. Bij veel bedrijven heerst het idee, vooral bij het middenkader, dat inschakelen van een adviseur een brevet van onvermogen is. Het hogere kader daarentegen maakt vaak gebruik van adviseurs om het middenkader te doorbreken.'

Van der Laan: 'Vijfendertig jaar geleden, bij Philips, vroeg ik me af: 'Als je een statistisch computerprogramma maakt, moet je dat dan vrijgeven zodat de klant er ook zelfstandig gebruik van kan maken?' Ik zag het bezwaar van toepassen zonder te weten wat de voorwaarden zijn. Dat speelt hier ook. Liever zou ik goede voorbeelden van toepassingen op het Internet zetten dan de methoden zelf. Aldus werkt Internet drempelverlagend.'

De toekomst van de kwantitatieve methoden, is die 'verleden tijd', zoals Ackoff beweerde, of 'schitterend', volgens Rinnooy Kan?

Aarts: 'Ackoff en Rinnooy Kan hebben allebei gelijk. De toekomst is schitterend als je bereid bent het strijdtoneel van KM te verleggen. De eerste 50 jaar was OR gebaseerd op de ontwikkeling van rekenkracht. Computer en OR zijn onlosmakelijk geworden. Maar de computer ontwikkelt zich extreem snel, die wordt geminiaturiseerd in alle opzichten. Onze KM zullen daarop moeten inspelen. Voor NP-complete problemen verwacht ik nog wel grote ontwikkelingen, zoals met de Kachiyan-methode. Maar dan uit een ander hoek, de natuurkunde. De Turingmachine wordt als basismodel achterhaald, de quantumcomputer komt er voor in de plaats.'

Cijsouw: 'Ik sluit me aan bij Rinnooy Kan. Er zijn zo veel gegevens, die liggen daar maar, kunnen we er iets verstandigs mee? Ook is er een groeiend

begrip voor mathematisch modelleren, en voor wiskunde in het algemeen: als je iets wilt onderbouwen dan moet dat kwantitatief. Beide dingen zitten in de lift, beide hebben met KM te maken.'

De Kok: 'De toekomst van KM is schitterend. Maar zoals wij nu met problemen bezig zijn, dat gaat veranderen. Het wordt allemaal pure casuïstiek.'

Van der Laan: 'De mathematisering van de maatschappij is in volle gang. Er zal meer vraag komen naar integrale modellen, voor allerlei problemen. Er is veel info uit allerlei soorten onderzoek. Hoe breng je daar structuur in aan? Nee, ik geloof niet dat het einde van de mathematisering nadert, integendeel. Ik ben dus aanhanger van Rinnooy Kan.'

Wessels: 'Beide auteurs hebben gelijk. Ackoff ging provocerend in tegen de overheersende manier van denken over KM. Rinnooy Kan zag mogelijkheden voor het toepassen van KM, omdat de wereld complexer wordt en de informatisering toeneemt. Alles moet zorgvuldiger, goedkoper, sneller, in telecommunicatie, verkeer en vervoer, en met meer kwaliteit.'

LITERATUUR

R.L. Ackoff (1979) The future of OR is past, *Journal of the Operational Research Society* (30) 93-104.

A.H.G. Rinnooy Kan (1989) The future of operations research is bright, *European Journal of Operational Research* (38) 282-5.

L. Fortuin & M. Zijlstra (1989) Operational Research in practice: Experiences of an OR group in industry, *European Journal of Operational Research* (41) 108-121.

L. Fortuin & M. Zijlstra (2000) Operational Research in practice: consultancy in industry revisited, *European Journal of Operational Research* (120) 1-13.

MYNT ZIJLSTRA was 10 jaar directeur van CQM, een adviesbureau voor kwantitatieve methoden en modellen te Eindhoven. E-mail: <Zijlstra@cqm.nl>.

LEONARD FORTUIN is oud-medewerker van CQM en directeur van Studio ForTekst, een bureau voor redactionele ondersteuning van bedrijven en andere organisaties, te Roermond. E-mail: <studio.fortekst@planet.nl>, zie ook <www.studio-fortekst.nl>.

Zij schreven dit artikel op persoonlijke titel.

De wereldwijde groei van analyse-systemen wordt vooral bepaald door ondernemingsgewijze installaties. Paradoxaal genoeg krimpt de markt voor stand-alone systemen omdat gegevensanalyse, *data mining* en kwaliteitscontrole (SPC) systemen een steeds belangrijker rol gaan spelen in het zakenleven en het productieproces.

De zin van statistiek(onderwijs)

Illustratie uit: Hans Magnus Enzensberger, *De Telduivel* (1997).

*54.42. $\vdash :: \alpha \in 2. \supset :: \beta \subset \alpha. !\beta. \beta \neq \alpha. \equiv. \beta \in l''\alpha$
Dem.
 $\neg. *54.4. \supset \vdash :: \alpha = l'x \cup l'y. \supset ::$
 $\beta \subset \alpha. \exists !\beta. \equiv : \beta = \wedge. v. \beta = l'x. v. \beta = l'y.$
 $[*24.53.56.*51.161] \equiv : \beta = l'x. v. \beta = l'y. v. \beta = \alpha : \exists !\beta$
 $\vdash. *54.25. \text{Transp.} *52.22. \supset \vdash : x \neq y. \supset. l'x \cup l'y$
 $[*13.12] \supset \vdash : \alpha = l'x \cup l'y. x \neq y. \supset. \alpha \neq l'x. \alpha \neq l'y. (2)$
 $\vdash. (1). (2). \supset \vdash :: \alpha = l'x \cup l'y. x \neq y. \supset ::$
 $\beta \subset \alpha. \exists !\beta. \beta \neq \alpha. \equiv : \beta = l'x. v. \beta = l'y :$
 $[*51.235] \equiv : (\exists z). z \in \alpha. \beta = l'z :$
 $[*37.6] \equiv : \beta \in l''\alpha \quad (3)$
 $\vdash. (3). *11.11.35. *54.101. \supset \vdash. \text{Prop.}$
 $*54.43. \vdash :: \alpha, \beta \in 1. \supset : \alpha \cap \beta = \wedge. \equiv. \alpha \cup \beta \in 2$
Dem.
 $\vdash. *54.26. \supset \vdash :: \alpha = l'x. \beta = l'y. \supset : \alpha \cup \beta \in 2. \equiv x \neq y.$
 $[*51.231] \equiv. l'x \cap l'y = \wedge.$
 $[*13.12] \equiv. \alpha \cap \beta = \wedge \quad (1)$
 $\vdash. (1). *11.11.35. \supset$
 $\vdash :: (\exists x, y). \alpha = l'x. \beta = l'y. \supset : \alpha \cup \beta \in 2.$
 $\equiv. \alpha \cap \beta = \wedge \quad (2)$
 $\vdash. (2). *11.54. *52.1. \supset \vdash. \text{Prop.}$

De groeiende belangstelling voor gegevensanalyse en kwaliteitscontrole (Six Sigma) heeft er voor gezorgd dat statistische programma's zich ontwikkeld hebben van 'gespecialiseerde software voor relatief exotische toepassingen' tot een van de belangrijkste componenten van het centrale zenuwstelsel van een organisatie.

WIM VAN NIEUWENHUYSEN EN TON KUYLEN

Vanuit vele invalshoeken (inkoop, verkoop, productie, klantencontact) worden gegevens verzameld. Complete gegevenspakhuizen worden gevuld met terabytes aan gegevens. Maar wat te

doen met al deze gegevens? De moeizaam gerealiseerde maar vaak weinig succesvolle systemen bieden meestal zeer beperkte mogelijkheden tot (online) analyse waardoor de verzamelde data



Inspelen op acute problemen.



Zoeken naar nieuwe mogelijkheden.

niet effectief worden gebruikt. Veel teleurstellingen uit het verleden kunnen verklaard worden uit een gebrek aan integratie, analytische kennis en projectmanagement. In het verleden werden de beslissingen om niet-kritische software als statistische of SPC-toepassingen aan te schaffen veelal overgelaten aan individuele specialisten. Dit resulteerde vaak in systemen met exotische programmeertalen en databasestructuren, een gebrek aan standaardisatie, integratieproblemen en hoge kosten per gebruiker voor onderhoud, training, consultancy en support. Dit is iets wat ondernemingen zich niet kunnen veroorloven wanneer deze software centraal in de organisatie terechtkomt en routinematig wordt gebruikt door veel verschillende medewerkers met sterk uiteenlopende achtergronden en interesses.

Inspelen op acute problemen

Het snel kunnen verzamelen, analyseren, interpreteren en communiceren van gegevens vormt de basis voor *Business Intelligence*, *Enterprise Intelligence*, *Supply Chain Intelligence* (SCI) en *Enterprise Performance Management* (EPM). Het gaat er om dat organisaties interne (inkoop, productie, verkoop) en externe (marktontwikkelingen, concurrentieverhoudingen) gegevens in onderling verband analyseren en de verworven inzichten gebruiken om op flexibele wijze te reageren op ontwikkelingen op de markt en in de organisatie zelf.

De nieuwe generatie 'analytische' software moet een gedifferentieerd en uitgebreid aanbod van technieken combineren met een goede kop-

peling met meetapparatuur, databases en andere software. De software moet gebruikersvriendelijk zijn, maar ook krachtig genoeg om gebruikers zonder veel statistische achtergrond te begeleiden door het woud van geavanceerde analytische technieken. In tegenstelling tot de standaard statistische pakketten (ontworpen voor meer technische gebruikers met een behoorlijke statistische achtergrond) kunnen de nieuwe analyseproducten gebruikt worden door een breed publiek (iedereen die regelmatig analyses wil uitvoeren en rapporten wil maken als onderdeel van het dagelijkse werk). Integratie en uitwisselingsmogelijkheden met andere in de organisatie aanwezige systemen spelen daarbij een belangrijke rol (*Enterprise Application Integration*, ofwel EAI). Organisaties willen de vruchten plukken van 'groupware' en 'Collaborative Business Intelligence'.

Was het vakgebied in het verleden voorbehouden aan specialisten, door de ontwikkelingen op soft- en hardwaregebied zien we dat de toepassing van de statistiek als het ware in de organisatie doordringt en beschikbaar komt voor mensen zonder specifieke statistische kennis. Vrijwel iedereen krijgt tegenwoordig, ook al is men zich dat vaak niet bewust, te maken met vraagstukken waarbij een statistische benadering op zijn plaats is:

- het beslissen op basis van cijfers;
- het interpreteren van tabellen;
- het grafisch presenteren van resultaten;
- het omgaan met onzekerheid;
- het onderscheiden van oorzaken van verschijnselen;
- het verband leggen tussen verschijnselen;
- etc.

Elementaire kennis van het statistisch denken is hierbij onontbeerlijk. Wanneer we echter de personeelsadvertenties volgen, blijkt dat er steeds minder vaak specifiek naar een statisticus wordt gevraagd. Wanneer we kijken naar de markt zien we de volgende gebruikersgroepen ontstaan:

<i>Statisticians</i>	2 %
<i>Knowledge Workers</i>	15 %
<i>Information Consumers</i>	82 %
<i>Executives (managers)</i>	1 %

Zoeken naar nieuwe mogelijkheden

Na de spreadsheet, de tekstverwerker en de database willen organisaties nu meer geavanceerde analysemiddelen beschikbaar stellen aan zoveel mogelijk medewerkers met verschillende achtergronden en kennisniveau's. Dit impliceert dat de nadruk steeds meer verschuift van de formules naar het praktische gebruik zelf. Veel eindgebruikers blijken niet meer geïnteresseerd in de achterliggende formules, maar willen weten:

- hoe ze een beleidsprobleem kunnen vertalen naar een analysebaar probleem;
- hoe ze de gegevens (uit meerdere bronnen) kunnen verzamelen (en samenvoegen) voor de analyse;
- welke techniek gezien de aard van de vraag of plaats in het rapport het meest voor de hand ligt;
- hoe de uitvoer moet worden gelezen;
- hoe de resultaten geïnterpreteerd dienen te worden in het licht van de originele vraagstelling.

In industriële situaties willen eindgebruikers weten:

- draait het proces (op termijn) binnen de gestelde normen;
- welke acties moeten worden ondernomen bij problemen;
- hoe worden de problemen veroorzaakt;
- hoe kunnen we de processen verder fijnregelen;
- hoe komen we met een beperkt aantal metingen



Online gegevensanalyse noodzakelijk.

achter de optimale instellingen?

In de academische omgeving verschuift de aandacht steeds meer naar het inhoudelijke vlak (theorievorming en kennisontwikkeling).

Uitgaande van het profiel van het gros van de gebruikers blijkt dat formules (en de vaardigheid om deze te kunnen narekenen met een calculator) in veel praktijksituaties nauwelijks gebruikt worden. De reactie van gebruikers is meestal: 'Ik hoef geen automonteur te zijn om een auto te kunnen besturen. Het gaat er om dat ik snel van A naar B kan reizen.' Anders gezegd, men wil antwoord op vragen zonder geïnteresseerd te zijn in de diepere achtergronden van de berekeningswijze.

Het onderwijs

'Voor vakken als statistiek of marktonderzoek wordt bij veel opleidingen minder tijd uitgetrokken dan voor een gemiddelde kookcursus voor beginners.'

We zien nog steeds dat het onderwijs vooral gericht is op het aanleren en interpreteren van een beperkt aantal formules. De ervaring leert helaas dat veel studenten daarmee nauwelijks aan de slag kunnen in de praktijk.

Vanzelfsprekend ontkennen we niet dat er altijd behoefte blijft bestaan aan een beperkte groep van ontwikkelaars en specialisten (statistici) die enerzijds de ontwikkelingen in de markt weten om te zetten naar nieuwe formules en algorithmes en die anderzijds de grote groep van eindgebruikers kunnen begeleiden en adviseren in het verantwoord gebruik van analysetechnieken. Bij deze groep gebruikers is kennis van formules een must.

We kunnen ons afvragen of we - gezien de markt - alleen 'pure' statistici moeten opleiden. Is er naast een kern van specialisten niet veel meer behoefte aan breed inzetbare 'onderzoekers' ofwel 'organisatiebrede gegevensanalisten'? Moet niet elke student uit elk vakgebied kennis hebben van elementaire gegevensanalyse? Wanneer dat het geval is kunnen we ons afvragen over welke vaardigheden deze laatste (grote) groep eigenlijk dient te beschikken? Kort samengevat moet een (eind-) gebruiker antwoord kunnen geven op vragen als:

- welke gegevens zou je gebruiken en op welke manier;
- hoe kunnen we deze gegevens verwerven en samenvoegen tot een analyseerbaar bestand;
- hoe analyseren we de gegevens zodat voorzien wordt in de informatiebehoeften?

Organisatorisch inzicht

Dit lijken louter 'technische' competenties, maar er komt meer om de hoek kijken. De gegevensanalist heeft inzicht en kennis van gegevensverzameling en -manipulatie en het gebruik van de juiste analyse-technieken voor het beantwoorden van veelal onduidelijk afgebakende beleidsvragen. Daarbij speelt organisatorisch inzicht een belangrijke rol. Een goede analist is iemand die inhoudelijk inzicht weet te combineren met kennis (van de beperkingen) van hard- en software, gegevensverzameling en -manipulatie:

- hij moet in staat zijn een probleem te vertalen naar een onderzoekbaar probleem;
- hij moet weten hoe en waar de benodigde gegevens te verzamelen (in de juiste hoeveelheid en samenstelling);
- de aanpak om de gegevensselectie aan de software over te laten kan tot het generaliseren van toevalligheden leiden. Tevens bestaat het gevaar dat verbanden waarvoor toevallig geen gegevens voorhanden zijn (onterecht) buiten beschouwing worden gelaten;
- het opschonen en aanpassen van de gegevens

veronderstelt achtergrondkennis en ervaring;

- het maken van een analysebestand, het uitvoeren van aanvullende berekeningen en het toevoegen van nieuwe variabelen is een zaak van geduld en inzicht;
- organisatorisch inzicht en kennis van bedrijfsprocessen is hierbij een must;
- het uitvoeren van analyses veronderstelt kennis en inzicht in het gebruik en de beperkingen van statistische en andere procedures. Daarbij is niet alleen het meetniveau van de variabelen doorslaggevend, maar vooral de inhoud van de (beleids)vraag zelf. Kennis van de processen in de organisatie is daarbij onontbeerlijk;
- de rapportage dient een antwoord te geven op de eerder geformuleerde vragen en moet worden aangepast aan het niveau van de lezer.

Het bovenstaande impliceert dat het onderwijs voor het gros van de gebruikers moet worden afgestemd op het aanleren van deze vaardigheden. Het kunnen narekenen van een kruistabel met behulp van een calculator of het kunnen produceren van een syntaxbestandje voor een of ander statistisch softwarepakket is niet voldoende om als analist het hoofd boven water te houden in het nieuwe millennium.

Inzicht

De in de academische wereld gehanteerde databestanden zijn vooral afgestemd op de onderbouwing van de theorie en het ontwikkelen van statistische modellen. Men heeft nauwelijks kennis van datawarehouses en relationele databases zoals die in het bedrijfsleven gebruikt worden. In deze laatste sector gaat het vooral om het vinden van oplossingen voor (strategische) vraagstukken en het op grote schaal inzetten van analyses voor het bewaken en optimaliseren van bedrijfsprocessen. Het gros van de activiteiten in deze setting (soms wel 80 %) bestaat uit het verzamelen en pre-processen van gegevens. Standaardisatie en groupwarefunc-

tionaliteit (het delen van kennis, rapporten en queries) spelen daarbij een belangrijke rol.

Het is maar de vraag of de benodigde vaardigheden via het onderwijs (kunnen) worden aangeleerd. In het traditionele onderwijs vindt men nog steeds een **harde kern** van docenten die menen dat leerlingen/studenten alle gebruikte formules moeten kunnen narekenen. Het gaat echter niet om de berekeningswijze (dat doet de software tegenwoordig wel voor u), maar om inzicht, keuze en verantwoorde toepassing zodat een antwoord kan worden gegeven op vragen die in de organisatie of wetenschap spelen.

Electronische leeromgeving

Veel gebruikers in de praktijk zeggen weinig baat te hebben bij datgene wat ze hebben geleerd op school. *Een vak als statistiek roept niet voor niets bij velen een negatieve associatie op.* Men zegt een kunstje te hebben geleerd, zonder te begrijpen waarvoor het dient. De student kan na een semester zwoegen een kruistabel maken met een statistisch pakket, maar heeft geen idee hoe hij of zij de uitkomsten moet verwoorden of op welke vraag hij nu eigenlijk een antwoord heeft gevonden.

‘Een kritieke waarde opzoeken achter in het boek? Welk boek?’

Vraagstukken als het gebruik van statistische methoden in de praktijk en het plaatsen van analysevraagstukken in een bredere organisatorische context komen in het onderwijs meestal niet aan de orde. Veel docenten beperken zich tot de formules die redelijkerwijs nog met een calculator na te rekenen zijn of bespreken slechts een paar menu-onderdelen van een ‘academisch’ pakket. Moderne alternatieven en onderzoeksvaardigheden komen niet of nauwelijks aan de orde. Er wordt weinig tot geen aandacht besteed aan een goede probleem-analyse gerelateerd aan de businesscontext. Men stimuleert de studenten niet om met gegevens en modellen te experimenteren. Toch is het erg belangrijk dat deze leren omgaan met situaties waarbij

een hoogst significante variabele uit een theoretisch model in de praktijk niet instelbaar blijkt of slechts tegen hoge kosten zodat er naar alternatieven moet worden gezocht. De bereidheid om het onderwijsmateriaal en didactische methode aan te passen aan actuele ontwikkelingen is verbijsterend klein. Het curriculum van het vak ligt veelal voor jaren vast. De bereidheid tot verandering of integratie met andere vakgebieden is verbazingwekkend gering.

‘Er wordt getentamineerd op syntax. Het proefwerk en de samenvatting van het boek staan al jaren op het internet...’

Wanneer we kijken naar de ontwikkelingen op de markt pleiten we voor meer praktijkgericht onderwijs. We vragen ons daarbij af of een vak als statistiek ‘pur sang’ nog wel toekomst heeft. We hebben bij gegevensanalyse te maken met cyclische processen waarin per fase andere technieken, kennis en kunde vereist zijn. De gegevensanalist moet inzicht en kennis hebben van gegevensverzameling en -manipulatie en het gebruik van de juiste analytische technieken voor het beantwoorden van veelal onduidelijk afgebakende beleidsvragen. Daarbij speelt organisatorisch inzicht en systeemkennis een belangrijke rol. Een goede analist is iemand die inhoudelijk inzicht weet te combineren met kennis (van de beperkingen) van hard en software, gegevensverzameling en manipulatie. Last, but not least, de analist moet ook nog de rest van de organisatie weten te overtuigen.

Als arbeid- en organisatie-socioloog/methodoloog is WIM VAN NIEUWENHUYSEN (1959) - na een loopbaan in het onderwijs, de wetenschap en het softwarecircuit - thans werkzaam als algemeen directeur van StatSoft Benelux. Daarnaast opereert hij als publicist, docent en consultant. E-mail: <wim@statsoft.nl>.

Als bijzonder hoogleraar aan de Katholieke Universiteit Brabant doceert TON KUYLEN op het gebied van CRM (Marketing Engineering en Databasemarketing). Daarnaast is hij werkzaam als partner in Cmotions (consultancy met betrekking tot afzetmarkt, klantbehoefte en klantwaarde, risicoanalyse en fraudedetectie, modellenbouw, datamining en business analyse). E-mail: <a.a.a.kuijlen@uvt.nl>.



Foto: Pieter Bosch

Een statisticus is ook een mens

Als de statisticus bij 'het grote publiek' al een naam heeft, dan zeker niet die van romanticus of dichter. De statisticus wordt veeleer gezien als een dorre Droogstoppel, om geen minder vleierende of minder oirbare kwalificaties te gebruiken. Toch zijn er ook aanwijzingen voor het tegendeel. Eén daarvan is het feit dat leerboeken over kansrekening en statistiek schatkamers zijn van poëzie en romantiek: opdrachten aan geliefden, kinderen en ouders, dankbetuigingen aan leermeesters en soms zeer poëtische beschouwingen op het schutblad.

FRED STEUTEL

Aan de muur van één van de zaaltjes in het oergebouw van het Mathematisch Centrum (CWI) hing een ingelijst vers over De Statisticus. Het betrof een zelfmoordstatistiek, en een belangrijke regel over deze statistiek luidde: 'die klopte niet, die klopte niet!' Om de zaak toch kloppend te krijgen wist de statisticus in kwestie niet beter te doen dan de hand aan zich zelf te slaan, en de tekening die het vers kracht bijzette toonde een wanhopige ambtenaar die zich met een kroontjespen in de borst stak. Scheidende medewerkers kregen soms een kopie van De Statisticus mee naar huis, niet direct het ideale rolmodel. Aan de beschrijvende statistiek lijkt dus weinig poëtische eer te behalen.

Hoe anders is het in de theoretische statistiek en de kansrekening. Opdrachten aan verwanten en beminden: '*to Dalma, to Marta, to Pradnya and Chitra, to my parents, to the memory of my father and mother*', maar ook opdrachten aan leermeesters en poëtische inleidingen op het schutblad. Heinz Bauer schrijft in zijn boek *Wahrscheinlichkeitstheorie*: '*Meinem verehrten Lehrer Herrn Professor Dr. Dres. h.c. Otto Haupt in Dankbarkeit gewidmet*.' De meest aandoenlijke opdracht vind ik die in de beide boeken van de kansrekenaar William Feller: '*to O.E. Neugebauer: o et praesidium et dulce decus meum*' en 'O, mijn beschermner en aangename eer'. Weliswaar geleend van Horatius (vertaling van Kox), maar toch.

Kendall en Stuart

De mooiste ontboezeringen vinden we in de bekende boeken van Kendall (en Stuart), *The Advanced Theory of Statistics*, delen 1 en 2. Eerst deel 1. Ik zal proberen om het uit het Engels te vertalen, maar de tekst van O. Henry - ooit inwoner van Austin in Texas; daar is een O. Henry House - is niet makkelijk te vertalen. Ik probeer het toch. 'Laten we op deze boomstam langs de weg gaan zitten en even wegdenken van de onmenselijkheid en lichtzinnigheid van dichters. De ware schoonheid steekt in de glorieuze kolommen van vaststaande feiten en wettelijke maten. De boomstam waar we nu op zitten bevat statistiek die mooier is dan het schoonste gedicht.' Dan volgt een exposé over jaarringen en de tijd die het zou vergen om van de boomstam steenkool te maken. Als de spreker even stil valt, wordt hij door zijn metgezellin aangemoedigd met de woorden 'Ga door Mr. Pratt, uw ideeën zijn zo origineel en troostend. *I think statistics are just as lovely as they can be.*' Het laatste zinnetje heb ik maar onvertaald gelaten. Ik raad overigens iedereen aan de originele tekst te lezen.

Op het schutblad van deel 2 lezen we de volgende gepareerde aanval op een statisticus: 'Je hebt me nog niet verteld', zei Lady Nuttal, 'wat je verloofde doet voor de kost'. 'Hij is statisticus', antwoordde Lamia, die zich tot haar ergernis enigszins in het defensief voelde. Lady Nuttal was duidelijk verrast. Ze had niet verwacht dat statistici normale maatschappelijke betrekkingen aangingen. Ze had vermoed dat de soort op een indirecte manier in stand werd gehouden, zoals bij muil dieren. 'Maar tante Sara, zei Lamia met warmte, 'het is een heel interessant beroep'. 'Ik twijfel er niet aan', zei haar tante, die er ten duidelijkste zéér aan twijfelde. 'Om iets van enige betekenis louter in getallen uit te drukken is zo evident onmogelijk, dat er ruime gelegenheid moet zijn voor het geven van goedbetaald advies over hoe dat zou moeten.' Lamia was even stil. Ze had geen zin om te spre-

ken over de verrassende emotionele diepten die ze had ontdekt onder Edwards numerieke vernis. Lamia eindigt met de opmerking 'Het zijn niet de getallen zelf die van belang zijn, maar wat je er mee doet.'

Een deel van deze ontboezeming werd onlangs in Leuven geciteerd door een afscheid nemende statisticus. Het 'citaat' uit *The undoing of Lamia Gurdleneck* wordt toegedicht aan de niet-bestaande schrijver H.A.C. Manderville.

Wiskundigen, statistici en politiek

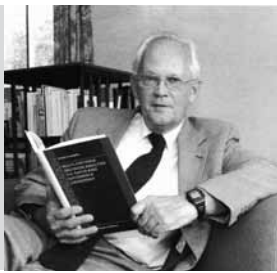
Lewis Carroll, auteur van *Alice in Wonderland*, was weliswaar geen statisticus, maar wel een wiskundige die grote populariteit genoot onder statistici; met name de statisticus Van Dantzig citeerde hem graag - zie *Statistical Priesthood I en II in Statistica Neerlandica*.

De meest satirische statisticus is ongetwijfeld Tom Lehrer, opgeleid in Harvard, en behalve theoretisch statisticus (wetenschappelijke artikelen in 1957 en '58), dichter, zanger en voortreffelijk pianist. Beroemd geworden met liederen als *Poisoning the pigeons in the park*, het zeer actuele *Send the Marines* en de hartverscheurende *Irish Ballad*, één van de lievelingsliederen van de statisticus Van Zwet.

Dat een statisticus het niet altijd makkelijk heeft weten we al uit het Statisticus-vers. Het volgende citaat (uit de mond van Winston Churchill) bevestigt dit: '*The first lesson that you must learn is, when I call for statistics about the rate of infant mortality, what I want is proof that fewer babies died when I was Prime Minister than when anyone else was Prime Minister. That is a political statistic.*'

Uit het bovenstaande blijkt overduidelijk dat een statisticus niet alleen een mens is, maar mogelijk zelfs een bijzonder mens.

FRED STEUTEL is emeritus hoogleraar kansrekening aan de TU Eindhoven; hij is redacteur van *Stator* en van *Statistical Theory & Method Abstracts* van het International Statistical Institute.
E-mail: <f.w.steutel@tue.nl>.



In Memoriam

prof.dr. Freerk Auke Lootsma

ROGER COOKE EN KEES ROOS

Onlangs overleed Freerk Lootsma. Zijn vele vrienden en collega's zijn diep getroffen. Graag willen wij stilstaan bij Freerks vele bijdragen aan de wiskundegroep in de TU Delft en aan de operationele research.

Freerk werd in 1936 te Midlum (Friesland) geboren. Hij studeerde wiskunde en theoretische natuurkunde aan de universiteit van Utrecht. Na een periode van tien jaar te hebben gewerkt in het bedrijfsleven promoveerde hij in 1970 op een proefschrift *Boundary Conditions of Penalty Functions for Constrained Minimization* bij prof. J.F. Benders aan de TU Eindhoven. In 1974 is hij benoemd op de leerstoel Operationele Analyse aan de faculteit wiskunde van de TU Delft. Uit zijn lange lijst van professionele activiteiten noemen we het organiseren van meerdere NATO-workshops, associate editor van de tijdschriften *Optimization* en *Journal of Multi-Criteria Decision Analysis*. Hij zette zich ook in voor buitenlandse bezoekers als voorzitter van de Delft University Visitors Foundation.

In die tijd was de operationele research een tamelijk nieuw gebied in Nederland. Zijn boek *Numerical Methods for Non-Linear Optimization* (1972, Academic Press) gold als baanbrekend en heeft veel internationale aandacht getrokken. Ook zijn proefschrift heeft diepe sporen achtergelaten op dit gebied en wordt nog steeds geciteerd. In contractresearch o.a. voor het ministerie van Economische Zaken en de Europese Commissie werden linear, mixed-integer en non-linear programming toegepast op problemen zoals onderzeeboot-bestrijding, verkeersregeling, productie-allocatie, productieplanning en energieplanning. Later, geïnspireerd door de club van Rome, raakte Freerk er van over-

tuigd dat de rol van operationele analyse moest worden verbreed om een bijdrage te kunnen leveren aan belangrijke maatschappelijke vraagstukken. Hij verdiepte zich in problemen zoals het energievraagstuk, waarin verschillende belanghebbenden met conflicterende doelstellingen tot rationele beslissingen moesten komen. Multicriteria analyse werd zijn hoofd aandachtsgebied. Een veel geciteerde studie van zijn hand ('Saaty's priority theory and the nomination of a senior professor in operations research' *European Journal of Operational Research*, 4, 380-388, 1980) had betrekking op de beraadslagingen in een benoemingsadviescommissie voor een hoogleraar in de wiskunde. Met scherpzinnige analyse en stijlvolle humor liet hij zien hoe de multicriteria methodes de verborgen beweegredenen van de deelnemers aan het licht brachten en zodoende tot een beter begrip leidden. In het verlengde hiervan werd Freerk in de laatste jaren gefascineerd door het menselijk oordeel en het wiskundig modelleren daarvan. Hij zocht wiskundige methodes om vage kwalitatieve oordelen om te zetten in kwantitatieve uitspraken. Vele onderzoekers op dit gebied zijn geboeid door zijn recente briefwisseling met Thomas Saaty.

Charmant en edelmoedig zijn de woorden die bij ons opkomen als we denken aan ons verlies. In de laatste jaren kenden Freerk en Riekje intens persoonlijk verdriet dat zij beiden met bewonderenswaardige moed droegen. In de laatste moeilijke weken kwam Freerks edelmoedigheid zeer sterk naar voren. Dit blijft een onuitwisbare inspiratie voor ons als we terugdenken aan Freerk Lootsma.

De auteurs zijn verbonden aan de TU Delft, respectievelijk als hoogleraar Risico Analyse en hoogleraar Optimaliserings-technieken. E-mail: <R.M.Cooke@ewi.tudelft.nl> en <C.Roos@ewi.tudelft.nl>.



Ben ik in beeld?

Planning van een observatie-satelliet

In 2002 heeft het Franse OR-genootschap ROADEF een internationale OR-competitie georganiseerd. De uitdaging bestond uit een OR-probleem dat was aangedragen door ONERA en CNES, Franse bedrijven die actief zijn op het gebied van ruimtevaart research. Het betrof een vraagstuk omtrent het selecteren en roosteren van observatie-aanvragen door een satelliet. In dit artikel beschrijft Eelco Kuipers (TNO Fysisch en Elektronisch Laboratorium) het probleem en de algemene ideeën achter de oplossing. Hij won de eerste prijs.

EELCO KUIPERS

Met de komst van de nieuwe generatie *Earth Observation Satellites* hebben de Franse Research bedrijven ONERA (*Office National d'Etudes et de*

Recherches Aérospatiales) en CNES (*Centre National d'Etudes Spatiales*) er een complex probleem bij gekregen. De huidige generatie (SPOT)

satellieten is niet in staat om, terwijl zij in een baan om de aarde draaien, naar een andere locatie te kijken dan recht naar beneden. De nieuwe generatie satellieten, Pléiades, daarentegen kunnen hun oriëntatie in hun baan om de aarde wel veranderen. Ze kunnen een stukje voorwaarts of achterwaarts kijken maar ook zijwaarts. Het complexe probleem is nu dat er hierdoor veel meer mogelijke tijdstippen zijn waarop je een foto zou kunnen maken. Hoe kies je het gunstigste tijdstip?

Orderboek

In de praktijk plaatsen bedrijven of personen een verzoek tot het maken van een satellietfoto van een stukje van de aarde. Deze verzoeken worden verzameld in een orderboek. Van dag tot dag wordt bekeken welke verzoeken zullen worden uitgevoerd. Er zijn over het algemeen veel meer verzoeken dan er binnen een dag uitgevoerd kunnen worden. Van ieder verzoek is bekend hoeveel het 'opbrengt' wanneer het op die dag wordt gehonoreerd (uitgevoerd). De vraag die CNES en ONERA zichzelf stellen is: gegeven de verzoeken in mijn orderboek, welke verzoeken kan ik vandaag honoreren zodat mijn opbrengst zo hoog mogelijk is, en in welke volgorde moet de satelliet deze gekozen verzoeken afwerken?

Daarbij spelen de randvoorwaarden van de satelliet natuurlijk een cruciale rol. Ten eerste maakt de satelliet een foto door middel van het strooksgewijs scannen van stukjes aarde; ieder verzoek wordt daarom van tevoren opgehakt in stroken die in beide richtingen gescand mogen worden. Alle stroken van een verzoek hoeven niet binnen één dag te worden gescand; de niet gescande stroken blijven gewoon in het orderboek staan voor de volgende dag. Een ingeplande strook moet echter wel in zijn geheel worden gescand. Door middel van een vastgelegde strafmaat wordt zoveel mogelijk voorkomen dat verzoeken teveel

versnipperd over veel verschillende dagen worden gescand.

Ten tweede is iedere strook gedurende een beperkte tijd zichtbaar. Dit heeft te maken met de eis dat de kijkhoek een minimale waarde moet hebben. Daarnaast kost het tijd om na het scannen van een strook het begin van de volgende geplande strook te vinden. Wanneer de volgende strook verder bij de gescande strook vandaan ligt kost dit meer tijd.

Verder is het mogelijk dat de aanvrager een verzoek doet voor een driedimensionale opname. Dit wordt opgelost door alle stroken van een dergelijk verzoek twee keer te scannen, waarbij een bepaald minimaal verschil moet zitten tussen de kijkhoek van de satelliet tussen de twee scans. Door dit verschil in kijkhoek kan achteraf een driedimensionaal beeld worden geconstrueerd uit de twee afzonderlijke foto's.

OR biedt uitkomst

Nadat ONERA en CNES zelf een tijdje aan het probleem hadden gerekend, hebben zij de hulp ingeroepen van het Operations Research genootschap ROADEF (*Société Française de Recherche Opérationnelle et d'Aide à la Décision*) om een prijsvraag voor dit probleem uit te schrijven. Het doel was te bekijken of het door hen geconstrueerde algoritme nog voor veel verbetering vatbaar was. De opdracht was om een algoritme te programmeren dat in staat is om binnen 5 minuten een rooster uit te rekenen dat de opbrengst van een satelliet maximaliseert. Er deden 29 teams uit 10 landen mee bestaande uit universiteitsdocenten en studenten (o.a. universiteiten van Grenoble, Toulouse, Nîmes, Montréal, Poznan, São Paulo), maar ook professionals uit het bedrijfsleven (o.a. Thales, British Telecom, TNO).

Het aantal mensen in een team varieerde van 1 tot 5. In de kwalificatieronde stuurden 18 teams

een algoritme in en op basis van de resultaten gingen 11 teams door naar de finaleronde, waarin zij nog een aantal weken kregen om het algoritme te verbeteren. Het algoritme dat ik in mijn vrije tijd heb gemaakt kwam uiteindelijk als beste uit de bus.

Lokale zoekmethode

Het algoritme is gebaseerd op een *lokale zoekmethode*. Het principe van een lokale zoekmethode is dat je begint met een (in dit geval) simpel rooster dat bijvoorbeeld handmatig geconstrueerd is maar je kunt ook met een leeg rooster beginnen. Daarna ga je stapsgewijs het rooster veranderen door er steeds een hele kleine willekeurige wijziging in aan te brengen en te kijken wat dit voor gevolgen heeft voor de opbrengst. De kunst zit hem dan enerzijds in het bepalen wat te doen wanneer je weet wat de gevolgen zijn voor de opbrengst; accepteer je de wijziging omdat je denkt dat de opbrengst er uiteindelijk het hoogste van wordt, of wijs je de wijziging af en ga je een andere wijziging proberen? Het is zeker niet verstandig om alleen wijzigingen te accepteren die de opbrengst doen stijgen. Denk maar aan het pakken van alle koffers in de vakantie-auto; wanneer je nooit een verlaging van de opbrengst accepteert (oftewel een koffer uit de auto haalt) zal je waarschijnlijk nooit de indeling kunnen vinden die de koffers optimaal indeelt zodat zo weinig mogelijk ruimte ingenomen wordt. Er bestaan standaard methoden in de mathematische beslis-kunde die je vertellen wanneer je een wijziging wel of niet moet accepteren, bijvoorbeeld *simulated annealing* of *tabu search* (zie referentie).

Anderzijds is de keuze van het type wijziging een belangrijke factor in de kans van slagen van de lokale zoekmethode. In de literatuur luidt het advies over het algemeen alleen zo klein mogelijke wijzigingen te proberen. Daar ben ik het niet

helemaal mee eens; complexe wijzigingen die de opbrengst licht veranderen zijn vaak minstens zo waardevol. Het is in het bijzonder belangrijk dat je met je set van mogelijke wijzigingen theoretisch gezien in iedere mogelijke situatie (op iedere plek in de oplossingsruimte) uit kan komen. Anders sluit je wellicht bij voorbaat al de optimale oplossing (rooster) uit!

Flexibiliteit

Met alleen het toepassen van een standaard (lokale) zoekmethode ben je er helaas meestal niet, en dat bleek ook bij dit probleem het geval te zijn. Nu begon de uitdaging pas echt! Een standaard zoekmethode geeft een goede basis, maar om tot echt goede oplossingen (roosters) te komen is meer vernuft vereist. Dit wordt veroorzaakt door de structuur van de oplossingsruimte die voor ieder probleem weer anders is. Deze structuur is bijna nooit homogeen (regelmatig) waardoor er doorheen zoeken lastig en tijdrovend kan zijn: daarom zijn er additionele mechanismen aan het algoritme toegevoegd. Deze mechanismen beïnvloeden vooral de willekeur van een wijziging, ik stuur als het ware een beetje bij in de zoektocht door de oplossingsruimte richting gebieden waarvan ik weet dat er betere oplossingen te vinden zijn. Hierbij moet je heel goed in de gaten houden dat je de eerder genoemde wetmatigheid dat alle oplossingen in de oplossingsruimte bereikbaar moeten blijven niet overtreedt. Dit kan bijvoorbeeld worden bereikt door sommige soorten wijzigingen met grotere kans te laten optreden dan andere.

Naast de lokale zoekmethode basis (hier is gekozen voor simulated annealing) zijn de twee belangrijkste mechanismen die zijn toegevoegd een tweede optimalisatie criterium (opbrengst is natuurlijk nummer één) voor de mate van flexibiliteit van het rooster en een 'bibliotheek' van stuk-

jes rooster. Na vele uren van proberen te doorgronden waarom het algoritme niet naar de beste (bekende) oplossing wilde gaan kwam ik erachter dat de mate van flexibiliteit van een rooster gebruikt kon worden om betere oplossingen te vinden. Hoe flexibeler een rooster is, hoe gemakkelijker het bleek te zijn dit rooster nog te verbeteren. Dit ligt misschien voor de hand, toch was niemand anders op hetzelfde idee gekomen. Ook was het nog een hele klus om de flexibiliteit uit te drukken in een getal zodat het algoritme ermee kan rekenen. Hoe definieer je wat flexibiliteit precies inhoudt voor een rooster?

Het tweede mechanisme werd geboren uit de observatie dat het algoritme vaak delen van het rooster optimaal indeelde tijdens het zoekproces, maar dat deze optimale stukjes ook weer verdwenen in het stapsgewijs verder wijzigen van het rooster. Door nu delen te onthouden door ze in een soort van bibliotheek te stoppen en ze te

gebruiken bij het evalueren van wijzigingen, kon dit worden voorkomen.

Verbetering van 32%!

Uiteindelijk wist het winnende algoritme de opbrengst voor een satelliet te verhogen met bijna 32% ten opzichte van de nu gehanteerde methode. Zeker als je bedenkt dat een satelliet vele tientallen miljoenen moet opbrengen om rendabel te zijn is dit een aanmerkelijke verbetering. In het oplossen van dit roosterprobleem is de kracht van lokale zoekmethoden maar weer eens aangetoond!

LITERATUUR

E.H.L. Aarts, J.K. Lenstra, *Local Search in Combinatorial Optimization*. Chichester, John Wiley & Sons, 1997.

EELCO KUIPERS is wetenschappelijk medewerker bij TNO Fysisch en Electronisch Laboratorium.
E-mail: <kuipers@fel.tno.nl>.

— NIEUW —



D. Neeleman en J. van Bolhuis

Compendium hypothesetoetsen

Niet altijd zijn standaardtechnieken als t-toets, variantieanalyse en correlatie de beste keuze. Vaak is het mogelijk om met minder waarnemingen en andere toetsen, beter gefundeerde conclusies te trekken. Deze toetsingstechnieken zijn vaak niet beschikbaar in de standaard statistische programmapakketten zoals SPSS, SAS en S-PLUS.

Het zojuist verschenen *Compendium hypothesetoetsen* is dan een uitkomst. Het kan onderzoekers voor een aantal ernstige misstappen behoeden. Met een rekenmachine of met enig programmeerwerk zijn de toetsen zelfstandig uit te voeren voor iedereen die de basisbegrippen uit de statistiek in voldoende mate beheerst. Iedere toets is voorzien van een getallenvoorbeeld.

ISBN 90 5352 764 8, paperback, 494 pagina's, €40,-, ook verkrijgbaar in de boekhandel

www.uitgeverijboom.nl

A G E N D A

Zie voor meer nieuws, conferenties, studiedagen, mededelingen van de VVS en cursussen de site van de VVS <<http://www.vvs-or.nl>>.

Belangrijke mededeling

Het bestuur van de VVS heeft samen met de sectievoorzitters in een aantal brainstormsessies gesproken over de toekomst van de Vereniging. Dit heeft geleid tot een discussie nota die op de website <www.vvs-or.nl> van de VVS is te vinden.

Statistische expertise gezocht

Het Central Statistics Office in Warschau wil methodologische ondersteuning bij het verbeteren van bepaalde statistieken. Het gaat om praktische concrete vragen rond steekproeftrekken, imputatie, tijdreeksanalyse en presentatie van gegevens. Het is vrijwilligerswerk. Reis- en onkosten worden betaald. Informatie: PUA bij Erlijn Logtenberg, 070 - 3490548, <elogtenber@nmcp.nl>.

28 augustus 2003

Eerste Bijeenkomst Werkgroep Statistiek en Micro-Arrays (Universiteit van Nijmegen). Doel van de werkgroep is om tot een uitwisseling van kennis en ervaring te komen ten aanzien van de statistische aspecten van de Micro-array technologie. Informatie: <www.umcn.nl/scientist/departments/epidemiology> (onder courses and workshops).

13-17 februari 2004

Het thema van de 14e Study door de International Commission on Mathematical Instruction (ICMI) is 'Applications and Modelling in Mathematics Education'. Plaats: Dortmund. Nadere informatie: <www.brocku.ca/mathematics/ICMI/study14>.

21-25 juni 2004

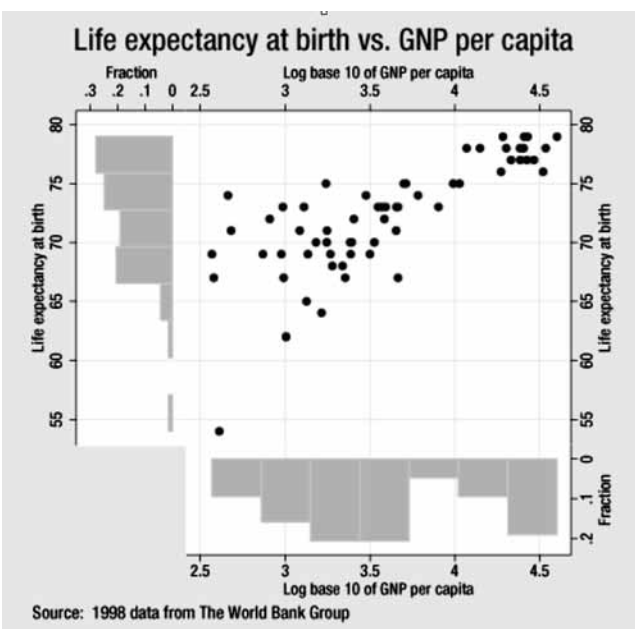
De 13e ECMI (European Conference on Mathematics for Industry) conferentie zal plaatsvinden in Eindhoven. Informatie: <www.ecmi2004.tue.nl>.



Stata Release 8

NIEUW IN STATA 8

- » Graphics: een nieuwe look, publication quality
- » Grafische gebruikersinterface: alle commando's via een dialoog uitvoeren
- » Database management: ODBC support, meer missing values
- » Meer statistische functies: MANOVA, MANCOVA en VAR, SVAR meer...
- » Programmeerfuncties: object georiënteerd programmeren, GUI
- » Stata 8 is sneller



post Postbus 220, 5150 AE Drunen

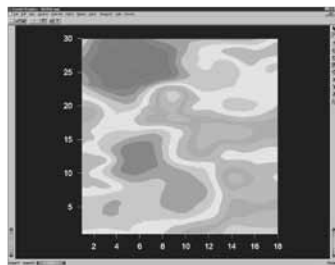
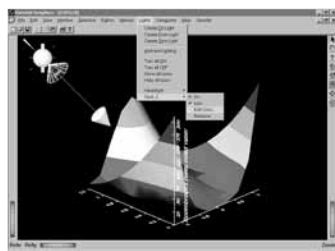
bezoek Grotestaat 401a, 5142 CB Waalwijk

telefoon 0416 - 378 125 (Stata) of 0416 - 543 444,

fax 0416 378 385, www.smitconsult.nl

Smit Consult is ondergebracht bij Cosinus Computing BV

PROBABLY THE BEST STATISTICS PACKAGE CURRENTLY ON THE MARKET!



The new **GenStat® for Windows® 6th Edition**, available now, boasts a long list of new features and improvements.

GenStat's reputation for the quality and accuracy of its statistical functionality is unparalleled.

GenStat is developed at Rothamsted Experimental Station where many of the standard statistical techniques still in use were first discovered, and it has served the needs of the statistical community worldwide for over 30 years!

The package is developed by statisticians who understand the real needs of users and have first hand experience of dealing with challenges in the 'real world'.

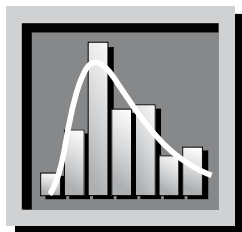
GenStat has the broadest range of statistical tools, all available in the standard package with no additional modules to install and pay for.

With its own, powerful and flexible high-level language, GenStat allows you to extend and customise its functionality virtually without limits. (Should you prefer to work with the user-friendly and familiar Windows® menu structure without writing a single line of 'code' that is, of course, also possible.)

If you have not tried GenStat recently this would be a good time to take another look! GenStat has been completely revised based on major market research and user feedback initiatives. In fact, in a recent series of focus group studies the respondents (consisting of professional users of statistical software from a variety of sectors) ranked GenStat as the best package currently on the market! Contact us to find out why...

For more information please contact:

Cosinus Computing, Postbus 220, 5150 AE Drunen, The Netherlands
tel. +31 (0) 416 543 444, fax +31 (0) 416 378 385, info@cosinus.nl
<http://www.cosinus.nl>



GenStat®

With GenStat you know you can!