

STAtOR

periodiek van de VWS jaargang 7 nummer 1, april 2006

Adaptief toetsen

Six Sigma in de zorg

Schijn bedriegt soms, en soms niet

Operations research voor rangeerplanning

Kansrekening op reis

Politiek, cultuur, taal en religie:
een analyse van het stemgedrag tijdens
het Eurovisie Songfestival

STATOR

Jaargang 7, nummer 1, april 2006

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

Goos Kant (hoofredacteur), Wies Akkermans, Martijn Berger, Han Oud, Gerrit Stermerdink (eindredacteur), Fred Steutel, Marnix Zoutenbier.

Kopij en reacties richten aan

Prof. dr. G. Kant (hoofredacteur), Faculteit der Economische Wetenschappen van de Universiteit van Tilburg, Postbus 90153, 5000 LE Tilburg, telefoon 013 - 4668234, mobiel 06-11045089, <G.Kant@uvt.nl>.

Bestuur van de VVS

Voorzitter: A.W. van der Vaart <aad@cs.vu.nl>
Penningmeester: S.J. Koopman <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 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

Uiterlijk vier weken voor verschijnen te zenden aan Pharos, Moeflonstraat 5, 6531 JS Nijmegen, telefoon 024 - 3559214, e-mail <hootegem@xs4all.nl>. STATOR verschijnt in april, juni, september en december.

Ontwerp en opmaak

Pharos / M. van Hootegem, Nijmegen

Uitgever

© Vereniging voor Statistiek en Operationele Research
ISSN 1567-3383

Inhoud

- 3** Nothing changes at New Year's day
- 4** Adaptief toetsen
Wim J. van der Linden
- 11** Six Sigma in de zorg
Jaap van den Heuvel
- 15** Schijn bedriegt soms, en soms niet
Ivo Adan en Fred Steutel
- 17** Operations research voor rangeerplanning
Ramon M. Lentink
- 22** Kansrekening op reis - *column*
Fred Steutel
- 24** Politiek, cultuur, taal en religie: een analyse van het stemgedrag tijdens het Eurovisie Songfestival
Laura Spierdijk en Michel Vellekoop
- 32** Woensdag 14 juni: Dag voor Statistiek en Besliskunde 2006



Nothing changes at New Year's Day

Een bekende regel uit het nummer *New Year's Day* van U2. Het gaat over het feit dat vele mensen aan het begin van het jaar of bij een bijzonder moment plannen maken, maar dat er aan het eind meestal weinig verandert. In dit lied wil de zanger juist niets veranderen, hij wil gewoon bij zijn geliefde zijn.

2006 is op het gebied van OR en Statistiek een bijzonder jaar. Het is het jaar van 50 jaar econometrie in Rotterdam, het jaar waarin 'De Statistische Dag' verandert in 'Dag voor Statistiek en Besliskunde', het jaar van 60 jaar CWI, van 25 jaar ORTEC en het jaar van het 25ste *STATOR*-nummer! Tijd dus voor veranderingen. Onze hoofdredacteur van het eerste uur, Dick den Hertog, heeft besloten om het stokje over te dragen aan mij. Daarnaast stopt ook Wies Akkermans na jarenlange trouwe inzet. Vanaf deze plaats wil ik graag Dick en Wies van harte bedanken voor hun bijdrage. Mede door hen heeft het blad de huidige kwaliteit.

De doelstelling van het blad verandert niet, we willen nog steeds op een vertrouwde wijze de prachtige toepassingen van OR en statistiek onder de aandacht brengen. Daarvan in dit blad weer diverse voorbeelden: hoe te komen tot een efficiënte rangeerplanning van treinen en hoe te komen tot een betrouwbare en gestandaardiseerde manier van toetsen. In dit laatste voorbeeld werken kansberekening, statistiek en OR nauw

samen om zogenaamde schaduwtesten te berekenen. Een mooie kruisbestuiving!

Je kunt nog zulke prachtige kansen hebben met OR of met statistiek, de kunst is vervolgens om dit ook daadwerkelijk toe te passen en te implementeren. In deze *STATOR* daarom een beschrijving van de Six Sigma implementatiemethode in twee ziekenhuizen. Six Sigma is een uit de Verenigde Staten afkomstige integrale methode voor kwaliteitsverbetering. De conclusie is dat door middel van deze methode grote resultaten kunnen worden geboekt op het gebied van kostenbesparing en kwaliteitsverbetering. Wellicht is de methode naast de eventuele implementatie van OR- en statistiek toepassingen ook toepasbaar bij je nieuwjaarswensen.

Wat goed is hoeft echter niet veranderd te worden, ontdekte de U2-zanger al. Vandaar ook in dit nummer naast diverse andere artikelen de bekende columns.* Opdat zij anderen weer mogen inspireren tot baanbrekend werk.

Veel leesplezier!

Goos Kant

* Door een technische fout ontbreekt in dit nummer de column van Onno Boxma. Onze excuses daarvoor aan de auteur en de lezers.

Figuur 1. Alfred Binet (1857-1911)



ADAPTIEF TOETSEN

Vorig jaar vierden we het eeuwfeest van de psychologische test. Deze begon met de publicatie van de Binet-Simon intelligentietest in 1905. Eerder werd er ook al volop getest, maar Binet was de eerste die zijn test volledig standaardiseerde.

Merkwaardig genoeg met één uitzondering:

tijdens de test werd de selectie van de opgaven aangepast aan het tussentijds geschatte intelligentieniveau. Toch moesten we tot na 1990 wachten alvorens deze als *adaptief toetsen* bekend geworden methode op grote schaal kon worden ingevoerd. Eerst moest de testtheorie zich als statistische discipline ontwikkelen. Daarna was het wachten op het beschikbaar komen van lokale rekenkracht in de vorm van personal computers.

WIM J. VAN DER LINDEN

Standaardiseren

Waaruit bestonden precies de verdiensten van Alfred Binet (figuur 1) en zijn medewerker Théodore Simon? Deze kunnen alleen maar duidelijk worden tegen de achtergrond van voorgangers en tijdgenoten als Fechner, Wundt, Ebbinghaus, Quetelet en Galton. Deze pioniers bewogen zich op gebieden die al tot wasdom waren gekomen, zoals de antropometrie en de psychofysica met hun meting van lichaamskenmerken en van gewaar-

wordingen van fysische stimuli zoals licht en geluid. Dankzij het werk van Fechner bezitten we bijvoorbeeld de wet van Weber-Fechner, die het verband tussen de sterkte van een prikkel en een sensatie met een logaritmische functie beschrijft en die we gebruiken als we aan de knop van onze stereoversterker draaien.

Binet kreeg van de stad Parijs de opdracht om een test te ontwikkelen waarmee – het probleem is herkenbaar – scholen konden differentiëren

tussen leerlingen met een mentale achterstand en leerlingen die gewoon lui waren. Voor de eerste categorie zou special onderwijs komen. Binet had eerder al over intelligentiemeting nagedacht en was zich van de problemen ervan bewust. Al in 1898 schreef hij (*Revue psychologique*; Wolf, 1973, pag. 149, op cit.):

'There is no difficulty in measurement as long as it is a question of experiments on... tactile, visual, or auditory sensations. But if it is a question of measuring the keenness of intelligence, where is the method to be found to measure the richness of intelligence, the sureness of judgment, the subtlety of mind?'

Binet's oplossing kenmerkte zich door drie vernieuwingen. In de eerste plaats bestond zijn test uit een groot aantal taken, die verschillende mentale functies, zoals geheugen, redeneren, beoordelen en abstraheren, moesten meten. Hiermee deed hij recht aan de hierboven geciteerde *richness of intelligence*. Maar daardoor verschilde hij ook van de psychofysici, die een vaste taak aanboden en dan één stimulus in sterkte varieerden.

Vervolgens was de test volledig gestandaardiseerd – het testmateriaal, de afname, de interpretatie van de resultaten en de scoring, alles werd nauwkeurig door Binet geprotocolleerd. De test werd daarmee een onafhankelijk, objectief meetinstrument; iedere proefleider moest bij dezelfde leerling op dezelfde resultaten uitkomen.

De derde innovatie was subtieler. Er bestaat geen natuurlijke intelligentieschaal; evenmin bestonden er in Binet's tijd conventies op dit punt. Zijn oplossing was echter even simpel als geniaal: hij koos *leeftijd* als schaal. In een vooronderzoek probeerde hij al zijn opgaven empirisch uit en werden voor iedere leeftijds-groep van 3-11 jaar die opgaven als representatief geselecteerd die door 75% van de leerlingen

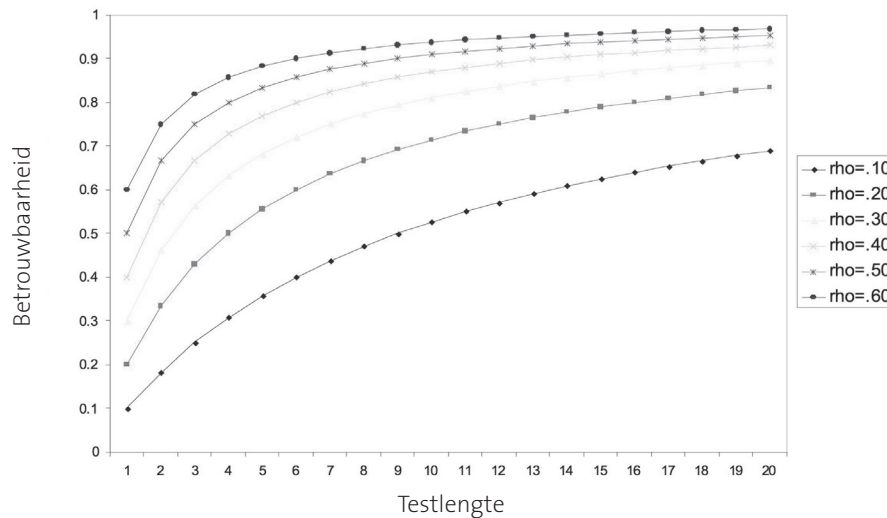
correct werden gemaakt. Deze leeftijd was dus de *schaalwaarde* van de opgave. Tijdens de test werden ze gebruikt om de leeftijd te schatten waarvoor de prestaties van de leerlingen representatief waren. Deze *mentale leeftijd* was de score van een Parijse leerling op de test. (Zes jaar later stelde William Stern voor om de verhouding tussen mentale en chronologische leeftijd als intelligentiequotiënt te definiëren. Hieraan danken wij het IQ. Weer een paar jaar later introduceerde Lewis Terman de gewoonte om het IQ met 100 te vermenigvuldigen. Sindsdien wordt het gemiddelde IQ steeds op 100 gesteld.)

Al was Binet een vernieuwer, hij was zeker ook een kind van zijn tijd. Zo stond hij in verbinding met Spearman, die ongeveer gelijktijdig het begin van de klassieke testtheorie formuleerde en al snel met zijn Spearman-Brown formule het effect van testverlenging op de betrouwbaarheid van de scores voor een gegeven betrouwbaarheid van de testitems vastlegde (zie figuur 2). Binet's idee om een lange test te kiezen zal ongetwijfeld ook door de wens tot betrouwbaarheid zijn gemotiveerd.

Ook de idee van standaardisatie was niet helemaal nieuw, maar paste precies in de opkomende traditie van het experiment met zijn standaardisering en randomisering dat, sinds Wundt in 1897 zijn laboratorium in Leipzig opende, greep op de psychologie had gekregen.

Parametriseren

Misschien was Binet's grootste vernieuwing dus wel de idee om zowel de testopgaven als de leerlingen op eenzelfde schaal af te beelden. Daardoor kon hij iets doen wat eigenlijk tegen de eis van standaardisering inging. In de protocollen voor zijn test bevinden zich precieze voorschriften die de selectie van de volgende opgave laten afhangen van de antwoorden op de vorige. Let wel, zo iets is dus alleen mogelijk bij een vaste schaal; voor



Figuur 2. Spearman-Brown formule voor de betrouwbaarheid van een homogeen verlengde test met testitems met betrouwbaarheden van .10, .20, ..., .60

een informele test, waarbij de score bijvoorbeeld wordt berekend als het aantal goede antwoorden, zou onmiddellijk de klassieke vraag ontstaan of een lagere score nu ligt aan een slechtere prestatie of aan opgaven die moeilijker waren.

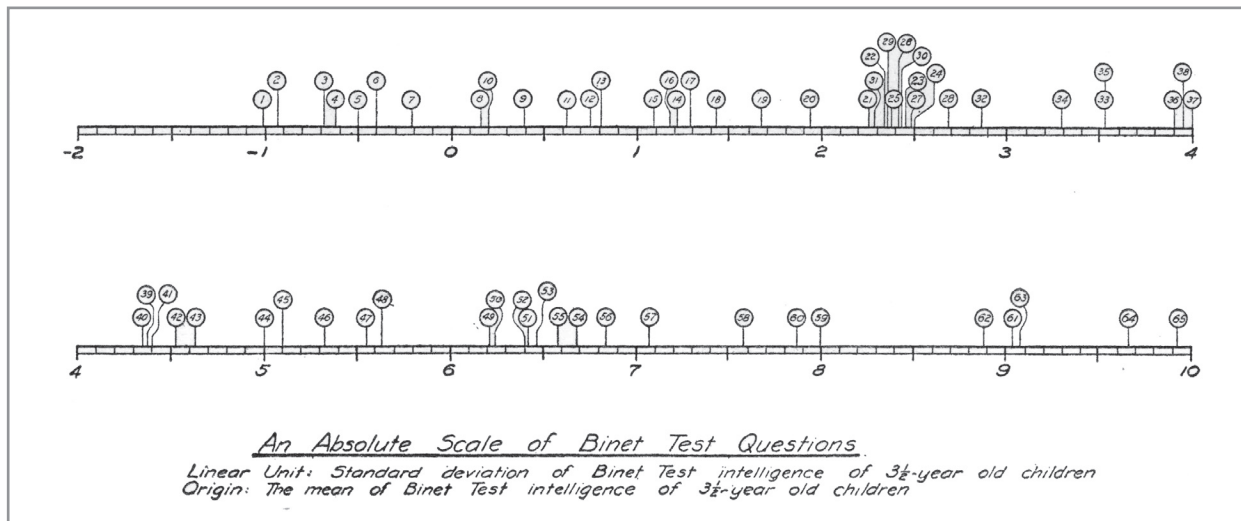
De reden waarom Binet zijn test adaptief maakte weet iedereen die mondelinge examens afneemt: efficiëntie. Het heeft geen zin om iemand die functioneert op de leeftijd van een 10-jarige opgaven voor te leggen die geschikt zijn voor een 6-jarige.

In het werk van Binet lagen diverse noties die een statisticus onmiddellijk zou oppikken. De manier waarop hij zijn opgaven schaalde bracht de noodzaak van het schatten van populatieverdelingen met zich mee. En met zijn intuïtieve ideeën over de efficiëntie van zijn test begaf hij zich op een gebied dat later onder de noemer adaptieve methoden in de *optimal design* theorie furore zou gaan maken.

De eerste die in dit gat sprong was – wonderlijk genoeg – niet Spearman. Hij bleef bij zijn lineaire model voor een ware score en een meetfout en werd de vader van de factoranalyse. (En hij had in deze periode waarschijnlijk veel van zijn tijd nodig voor een hoogoplopend conflict met Karl

Pearson over zijn correctie van de productmoment correlatie voor meet-onbetrouwbaarheid.) Degene die dat wel deed was Louis Thurstone, die in 1925 een schaalmethode introduceerde waarmee hij de opgaven uit de Binet test op een nieuwe schaal plaatste. Hij deed dit door voor iedere opgave een onderliggende normale verdeling te postuleren. Met enkele aanvullende aannamen lukte het hem om alle opgaven op een gemeenschappelijke schaal te plaatsen. Ook liet hij zien hoe zijn aannamen empirisch gecontroleerd konden worden. In feite maakte hij daarmee intelligentie los van leeftijd en introduceerde iets hij wat nu bekend staat als een *latente variabele*. Figuur 3 geeft de plaatsing van een aantal van Binet's opgaven op Thurstone's schaal aan (met een arbitraire eenheid en nulpunt gefixeerd aan de hand van een genormaliseerde verdeling van de scores van de kinderen van 3,5 jaar).

Met de notie van een latente variabele opende zich een geheel nieuw gebied aan toepassingen. Een paar jaar later liet Thurstone zien hoe zoiets vaags als een attitude gemeten kon worden. Alles wat nodig was waren een schaalmethode en de reacties 'eens' en 'oneens' op een aantal uitspraken die het object van de attitude beschrijven.

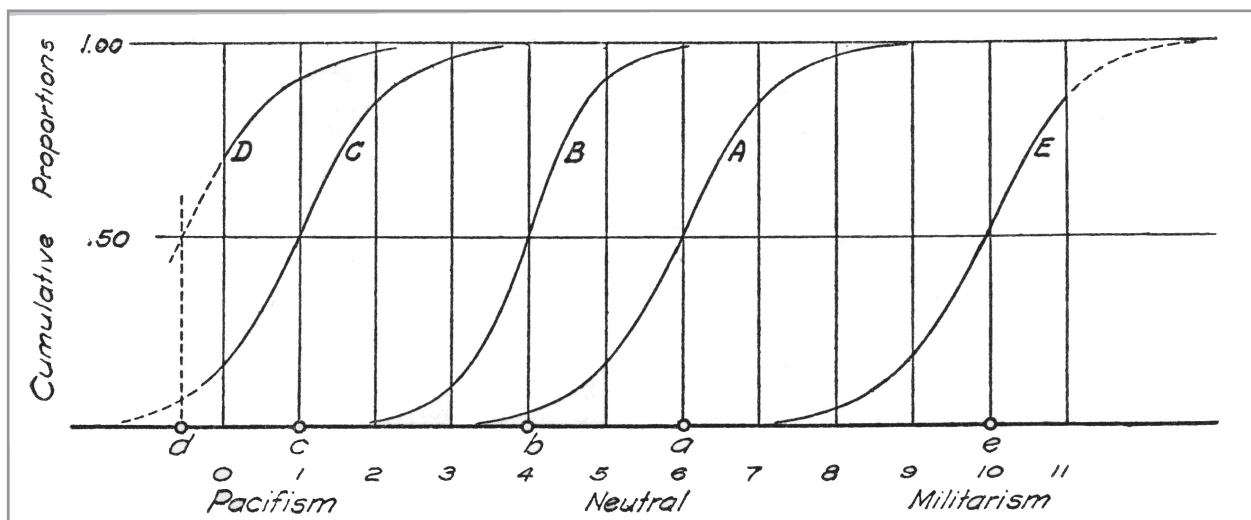


Figuur 3. Thurstone's schaling van de Binet opgaven (uit: Journal of Educational Psychology, 1925, 16, 433-451)

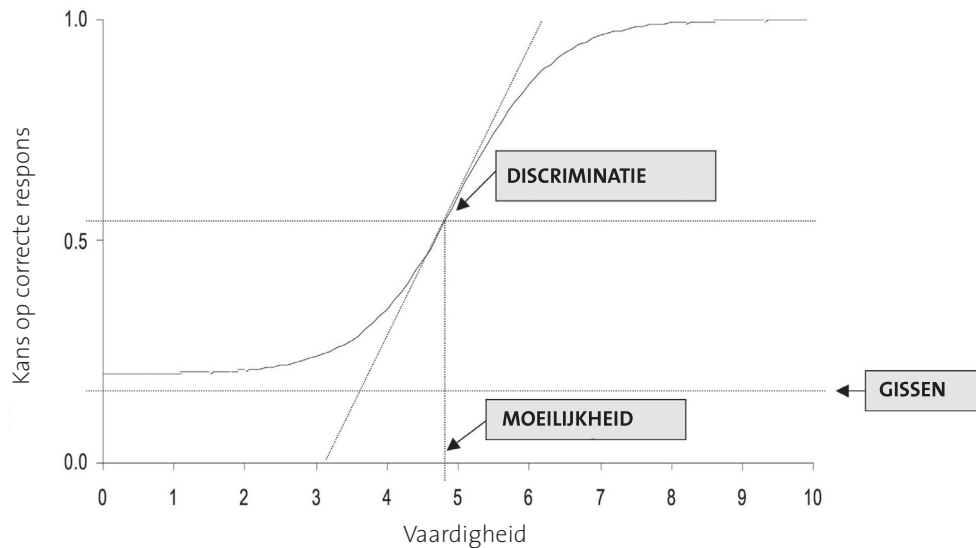
Figuur 4 geeft de cumulatieve verdelingsfuncties aan voor de opgaven op een schaal die Thurstone in 1928 voor de attitudedimensie pacifisme-militarisme ontwikkelde. Thurstone was buitengewoon creatief en heeft meer vindingen op het gebied van de schaalmethoden op zijn naam staan. Buiten de psychologie is de bekendste hiervan wellicht zijn methode voor de analyse van paarsgewijze vergelijkingen.

Na Thurstone werd het lange tijd stil, totdat Frederic Lord en Alan Birnbaum in 1950-1960 de draad weer oppakten. Zij schaalden hun testopgaven niet met populatieverdelingen, maar met probabilistische modellen voor de kans op een respons door individuele respondenten. Birnbaum introduceerde in 1968 het gebruik van het 3-parameter logistische responsmodel

$$p_i(+|\theta) = c_i + (1 - c_i) [1 + \exp\{-a_i(\theta - b_i)\}]^{-1}$$



Figuur 4. Enkele opgaven op Thurnstone's schaal voor pacifisme-militarisme (uit: American Journal of Sociology, 1928, 23, 529-554)



Figuur 5. Response functie volgens het 3-parameter logistische model.

dat de kans op een goed antwoord beschrijft als een functie van een latente persoonsvariabele θ en een drietal parameters voor de opgave (zie figuur 5). Vervolgens ontstond er een stormachtige ontwikkeling waarin een groot aantal responsmodellen voor verschillende toepassingen werd ontwikkeld en statistisch hanteerbaar werd gemaakt. Een overzicht van deze ontwikkeling, die nu bekend staat als de item-responsstheorie of IRT ('item' is de standaardterm voor een opgave in een test of attitudeschaal) is te vinden in Fischer en Molenaar (1995) en van der Linden en Hambleton (1997).

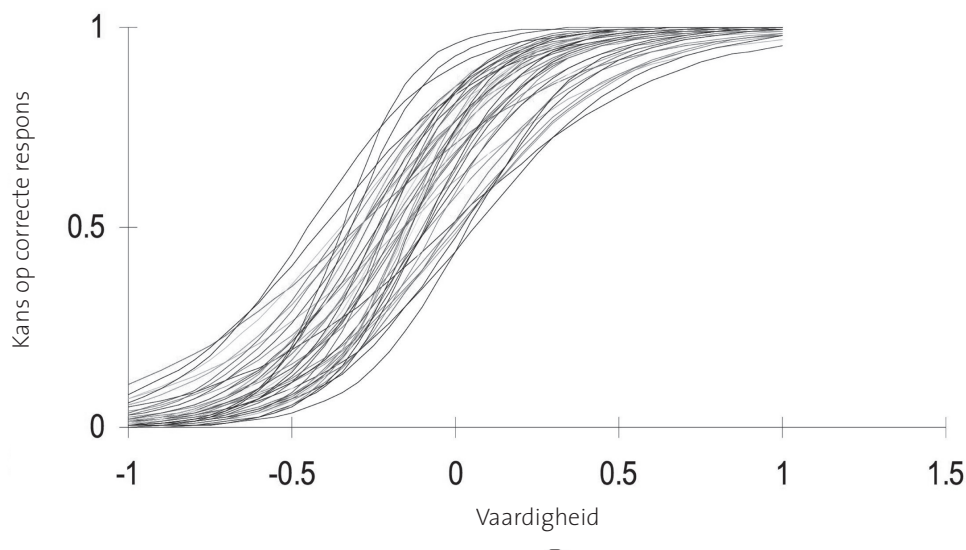
IRT modellen vormen thans het standaard-repertoire van de psychometrie voor de analyse van tests en examens. In Nederland worden ze bijvoorbeeld door het Cito gebruikt om de eind-examens in het voortgezet onderwijs te analyseren. Figuur 6 laat de responsfuncties zien voor een eindexamen Engels uit het VO. Het is goed om te beklemtonen dat de horizontale as in deze figuur een latente variabele representeert waarvoor de posities van de kandidaten samen met de parameters voor de opgaven uit de responsen moeten worden geschat. Dit is een belangrijk

verschil tussen IRT modellen en *quantal response* modellen, waarmee de regressie van responsen op een gemeten, experimentele variabele worden bestudeerd.

Adapteren

Een belangrijk kenmerk van alle IRT modellen is dat ze de effecten van personen en de opgaven op de responsen door middel van aparte parameters scheiden. Hiermee kwam Binet's oorspronkelijke idee, om zijn opgaven adaptief te kiezen bij het intelligentieniveau van zijn Parijse leerlingen, beschikbaar voor iedere test of examen dat met behulp van een IRT model op één of meer dimensies schaalbaar blijkt. Het enige wat gedaan moet worden, is zorgen dat er tijdens de test een goede match ontstaat tussen de schattingen van de persoonsparameter en de (vooraf geschatte) item-parameters.

Frederic Lord experimenteerde al met adaptieve testen in de periode 1960-1970. Zijn grote probleem was echter het gebrek aan de rekenkracht die bij adaptieve testafname nodig is. Itemparameters kunnen vooraf zonder enige tijdsrestrictie wor-



Figuur 6. Responsfuncties voor een eindexamen Engels uit het VO

den geschat, maar de schatting van de persoon-parameter, θ , en de selectie van de items moet in *real time* gebeuren. Eén van zijn pogingen om een potlood-en-papier versie van een adaptieve test te maken was de *flexilevel test*. Hierbij kreeg de kandidaat een boekje met opgaven, een vast item om te beginnen en een antwoordformulier in de vorm van een schrapkaart. Als de kandidaat een antwoordveld wegkraste, werd het nummer van de volgende opgave zichtbaar. Iets later werd op beperkte schaal met adaptieve testen op main-frame computers geëxperimenteerd. Voorlopers waren de examens in de faculteit psychologie aan de *University of Minnesota* door David Weiss, personeelselectie bij de *Civil Service* in de VS en de selectie van beroepsmilitairen in de VS met de *Armed Services Vocational Aptitude Test Battery* (ASVAB).

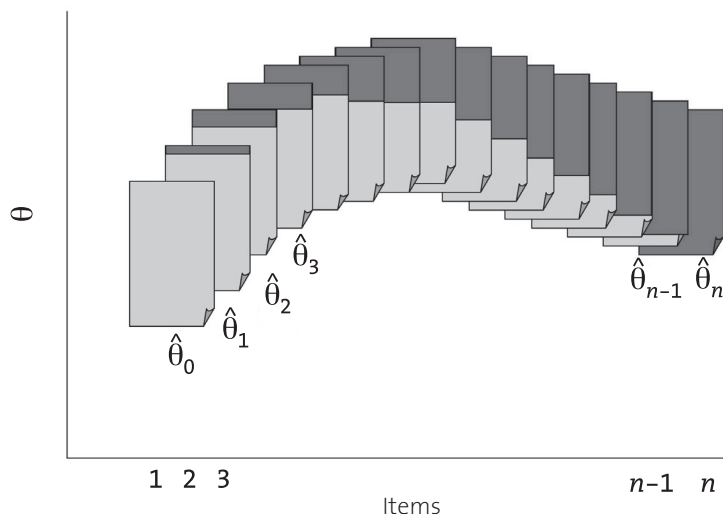
Nieuwe Problemen

De grote doorbraak vond echter plaats toen in de 90-er jaren adaptief toetsen met een personal computer mogelijk werd. De eerste grootschalige examens die adaptief werden afgenomen,

waren de eerder genoemde ASVAB, een *teacher certification* examen (PRAXIS) en de *Graduate Record Examination* (GRE), de toelatingstest voor de *Graduate Schools* in de VS. Daarna volgden wereldwijd andere examens.

Maar tegelijkertijd werd duidelijk dat er nieuwe problemen opgelost moesten worden. Een eerste voorbeeld is de noodzaak om adaptieve toetsen voor iedere kandidaat aan dezelfde inhoudelijke specificaties te laten voldoen. Examenorganisaties leggen aan de selectie van de opgaven voor hun toetsen vaak honderden randvoorwaarden op met betrekking tot een inhoudelijke classificatie, verdeling van de goede antwoorden over de alternatieven, leesbaarheidsindices, e.d.. Als ze adaptief worden, moeten dezelfde randvoorwaarden aan de itemselectie voor iedere kandidaat worden opgelegd.

Het tweede voorbeeld was pijnlijker. Toen de GRE in 1994 adaptief werd, voelden de *coaching schools* in de VS, die studenten tegen groot geld op deze test voorbereiden, zich in de hoek gedrongen. Eén van hen, *Kaplan Educational Centers*, zond zijn werknemers naar een aantal testcentra om de test meerdere keren af te leggen en de items



Figuur 7. Adaptief toetsen met schaduwtests

uit het hoofd te leren. Omdat het item-selectie-algoritme in feite kapitaliseerde op een beperkt aantal zeer goede items in de itembank, lag het hele examen op straat.

Een derde voorbeeld is de differentiële tijdsdruk waaronder kandidaten kunnen komen te werken. Items verschillen in de tijd die ze vragen (een factor 5-7 is niet ongebruikelijk). Als de selectie voor een kandidaat ongunstig uit zou vallen, kan hij onder grote tijdsdruk komen te staan om de test af te maken, wat uiteraard oneerlijk is.

Een algemene benadering waarmee aan de Universiteit Twente wordt gewerkt om dergelijke problemen op te lossen is door middel van schaduwtests. Het principe kan worden uitgelegd met behulp van figuur 7. Horizontaal staan daar de opeenvolgende items in een adaptieve test; de verticale as representeert de te schatten vaardigheid θ . In plaats van items selecteert het algoritme telkens een gehele test uit de bank. Deze schaduwtests worden niet als zodanig afgenomen, maar vormen een tussenstap. Ze worden zo samengesteld dat ze (i) alle eerdere items die de kandidaat heeft beantwoord bevatten (de donkere gedeelten in figuur 7), (ii) optimaal zijn bij de huidige $\hat{\theta}$ en (iii) voldoen aan alle aan de adaptieve toets op te leggen randvoorwaarden. Het item dat afgenomen wordt, is het beste vrije item uit de schaduwtest. Omdat alle randvoorwaarden

aan iedere schaduwtest worden opgelegd, voorkomen we dat we buiten de oplossingsverzameling moeten selecteren en voldoet de adaptieve test dus automatisch aan deze randvoorwaarden. Bovendien kan met dit schema iedere bestaande test onmiddellijk in een adaptieve versie worden omgezet.

Schaduwtests worden in real time geselecteerd met een geoptimaliseerde solver voor geheeltallig programmeren. In het optimaliseringsmodel worden tijdens de test de benodigde randvoorwaarden bijgehouden. Het probleem van de itemveiligheid kan worden opgelost door bij de selectie van de schaduwtests een kansmechanisme in te bouwen. Toepassingen bij diverse praktische problemen zijn te vinden in van der Linden (2005, hfd. 9) en van der Linden en Glas (in druk).

LITERATUUR

- Binet, A., & Simon, Th. A. (1905). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'Année Psychologie*, 11, 191-336.
- Fischer, G. H. & Molenaar, I. W. (1995), *Rasch models: Foundations, recent developments, and applications*. New York: Springer.
- Linden, W.J. van der (2005). *Linear models for optimal test design*. New York: Springer.
- Linden, W. J. van der & Hambleton, R. K. (Eds.) (1997). *Handbook of modern item response theory*. New York: Springer.
- Linden, W. J. van der & Glas, C. A. W. (in druk). Statistical aspects of adaptive testing. In C. R. Rao & S. Sinharay (Eds.) *Handbook of statistics* (Volume 27: Psychometrics). Amsterdam: Elsevier.
- Stigler, S. M. (1986). *The history of statistics: The Measurement of uncertainty before 1900*. Cambridge, MA: Harvard University Press.
- Wolf, T. H. (1973). *Alfred Binet*. Chicago: The University of Chicago Press.

WIM J. VAN DER LINDEN is psychometricus en hoogleraar Meetmethoden en Data-analyse bij de Faculteit der Gedragwetenschappen van de Universiteit Twente. Zijn onderzoek richt zich op de testtheorie, in het bijzonder responsmodellen met latente variabelen.
E-mail: w.j.vanderlinden@utwente.nl

SIX SIGMA IN DE ZORG



De gezondheidszorg is een boeiende bedrijfstak. Er wordt marktwerking geïntroduceerd, alle systemen zijn op de schop genomen en iedereen heeft er een mening over. De kwaliteit van zorg is aanhoudend onderwerp van discussie en de patiëntveiligheid zou, ook volgens internationale studies, onder de maat zijn (Institute of Medicine 1999). *Captains of industry* worden uitgenodigd om na een rondleiding van enkele weken hun mening te geven over wat er allemaal sneller en beter kan in de zorg. Wat doet de zorg zelf in deze woelige tijden? In dit artikel wordt uit de doeken gedaan wat twee ziekenhuizen hebben bereikt met het invoeren van Six Sigma, een uit de Verenigde Staten afkomstige integrale methode voor kwaliteitsverbetering.

JAAP VAN DEN HEUVEL

Six Sigma is ontwikkeld door Motorola en bekend geworden door General Electric. Veel grote Amerikaanse multinationals gebruiken Six Sigma en ook een aantal Nederlandse bedrijven is er inmiddels mee aan de slag gegaan. Six Sigma is een integrale aanpak voor kwaliteitsverbetering die uit vier elementen bestaat.

Allereerst bestaat de filosofie van Six Sigma eruit dat beslissingen gebaseerd dienen te zijn op

gegevens: de data, en niet op emoties of veronderstellingen. Er wordt dus veel gemeten binnen Six Sigma. Zeker in de zorg is dit niet erg gebruikelijk gelet op de grote weerstand tegen de prestatie-indicatoren. In de praktijk blijkt meten wel erg heilzaam voor het oplossen van knelpunten.

Een tweede kenmerk van Six Sigma is de projectmatige aanpak. Projecten worden onderverdeeld in vijf fases (*Define, Measure, Analyse, Improve* en

Control), afgekort DMAIC, die onderverdeeld zijn in twaalf projectstappen. Iedere stap kent vaste criteria waaraan de uitkomst wordt getoetst. Een volgende stap mag pas uitgevoerd worden als het resultaat van de voorgaande stap bekend is. Zo moet in de Define fase helder worden welke kwaliteitsindicator, de zogenaamde CTQ (Critical to Quality), verbeterd zal worden. Tevens dient de meetprocedure gevalideerd te zijn. In de Analyse fase wordt de procesprestatie gemeten, worden de doelstellingen van het project bepaald en worden potentiële invloedsfactoren op de CTQ geïdentificeerd. Er ontstaat daardoor een vergelijking:

$$Y = f(X_1) + f(X_2) + f(X_3) \dots$$

Hierbij is Y de waarde van de CTQ en alle X'en zijn de potentiële invloedsfactoren. In de Improve fase worden de belangrijkste invloedfactoren geïdentificeerd en wordt de relatie met de CTQ vastgesteld. Tot slot worden verbeteracties ontworpen. In de Control fase wordt het kwaliteitssysteem aangepast aan de nieuwe situatie, wordt de nieuwe procesprestatie bepaald en na een periode van een half jaar opnieuw gemeten. Vervolgens wordt het project afgesloten. Hieronder staan alle fases en stappen genoemd.

Een derde kenmerk van Six Sigma is het gebruik van een groot aantal kwaliteit-verbeter-instru-

menten en statistische technieken. In de tool box van Six Sigma zitten algemeen bekende verbeter-instrumenten zoals het Ishikawadiagram, de Root Cause Analyse en de Failure Mode & Effect Analyse. Deze instrumenten worden ingezet om vast te stellen of er een relatie bestaat tussen de CTQ en de mogelijke invloedsfactoren. Van de belangrijkste invloedsfactoren wordt met behulp van statistische technieken de exacte relatie met de CTQ vastgesteld. Het gebruik van (uitgebreide) statistiek is één van de kenmerken van Six Sigma. Sigma betekent zoals bekend standaardafwijking en zes sigma staat voor een foutenmarge van 3,4 op de miljoen gebeurtenissen. In de praktijk is dit natuurlijk een fouten-marge die niet wordt gehaald. Wanneer we in ons ziekenhuis 5,6 procent postoperatieve infecties (zie www.cwz.nl) realiseren is dat landelijk vergeleken een hele goede score, maar nog geen zes sigma kwaliteit. Dit is zeker in dit geval ook niet haalbaar en de naam Six Sigma symboliseert dan ook meer de ambitie om foutloos te werken en de nadruk op statistische onderbouwing van beweringen.

De laatste belangrijke eigenschap van Six Sigma is de degelijke inbedding in de organisatie. Medewerkers worden opgeleid tot een zogenaamde Green Belt of Black Belt. Een Green Belt

Define	Project identification; project proposal; project selection
Measure	1. Selecteer de <i>Critical To Quality</i> (CTQ) 2. Operationaliseer de CTQ 3. Valideer de meetprocedure
Analyse	4. Bepaal de proces-prestatie 5. Bepaal de project-doelstellingen 6. Identificeer mogelijke invloedsfactoren
Improve	7. Selecteer de belangrijkste invloedsfactoren 8. Bepaal de relatie tussen de CTQ en de invloedsfactoren 9. Ontwerp verbeteracties
Control	10. Pas het kwaliteit-beheerssysteem aan 11. Bepaal de nieuwe proces-prestatie 12. Sluit het project af

krijgt een opleiding van zes dagen gespreid over een half jaar en is voor maximaal twee dagen per week vrijgesteld voor het project. In die periode doet hij of zij ook zijn eerste project. Projecten dienen naast het verbeteren van de kwaliteit ook een concrete besparing op te leveren. Voor Green Belt projecten bedraagt dat een bedrag van minimaal 20.000 euro. Black Belt projecten dienen minimaal 100.000 euro op te brengen. De opleiding van een Black Belt duurt vier weken en is eveneens gespreid over een half jaar. De Black Belt is volledig vrijgesteld voor projecten. Het is gangbaar om ook een zogenaamde Master Black Belt aan te stellen die als programma manager van alle Six Sigma projecten fungeert en die ook Green Belts begeleidt met hun projecten. Een laatste functie die de Six Sigma organisatie kenmerkt is die van Champion. Dit is de manager die verantwoordelijk is voor het project en die bewaakt dat het project volledig wordt uitgevoerd. In de ideale situatie heeft de Champion geen omkijken naar het project, maar als er vertragingen optreden kan hij vanuit zijn hiërarchische positie belemmeringen wegnemen. De organisatie ziet er in eerste instantie wellicht wat merkwaardig uit, zeker als het gaat om de naamgeving, maar in de praktijk blijkt het erg goed te werken.

Six Sigma in het ziekenhuis

In 2002 is het Rode Kruis Ziekenhuis (RKZ) in Beverwijk als eerste ziekenhuis in Nederland begonnen met het invoeren van Six Sigma. Het traject begon met het opleiden van een eerste groep Green Belts in 2002. Daarna zijn jaarlijks twee groepen van ongeveer vijftien personen opgeleid. De terminologie, de strakke projectmatige aanpak en de nadruk op statistiek deden in eerste instantie wel wat stof opwaaien. Ook hadden enkele verpleegkundigen tijdens de opleiding moeite met de statistiek. Maar een goede statistische onderbouwing bleek hen te helpen bij het

doorvoeren van veranderingen. Bij één project kon de angel uit de discussie worden gehaald door statistisch aan te tonen dat verschillen in ligduur niet afhankelijk waren van de artsen, terwijl dit bij oppervlakkige beoordeling van de cijfers wel zo leek. Verder kon aangetoond worden dat de ligduur wel significant afhankelijk was van het behandelprotocol. Zonder al teveel discussie is het protocol daarop aangepast. Al snel bleek dus dat er met Six Sigma grote stappen gemaakt konden worden met verbeteracties die in het verleden steeds, om emotionele redenen, stagneerden (Van den Heuvel 2004). Het RKZ beschikte al over een verbetermethodiek die onderdeel uitmaakte van het in 2000 gecertificeerde ISO 9001 kwaliteitssysteem (Van den Heuvel 2005a). Toen Six Sigma in 2002 werd toegevoegd aan het kwaliteitssysteem nam het verbetermanagement een grote vlucht. In totaal zijn er 63 Green Belts opgeleid en zijn er 44 projecten gestart. In 2004 kon hiermee een besparing van meer dan een miljoen euro worden gerealiseerd. Projecten konden op nagenoeg alle afdelingen worden gestart. Begin 2004 voorzag het RKZ financiële problemen. Naast de lopende Six Sigma projecten is toen een extra slag gemaakt in de vorm van een groot aantal doelmatigheidsprojecten. Deze projecten werden uitgevoerd binnen de Six Sigma structuur, hadden een beperkte organisatorische impact maar resulteerden wel aan het einde van het jaar in een extra besparing van 1 miljoen euro. Het RKZ kon het jaar 2004 dan ook afsluiten met een positief resultaat vóór reserveringen van bijna 2 miljoen euro (Van den Heuvel 2005b).

In 2005 startte het Canisius Wilhelmina Ziekenhuis (CWZ) in Nijmegen eveneens met Six Sigma. Inmiddels zijn twee groepen met in totaal 33 Green Belts opgeleid. Zij werken aan 24 projecten waarvan de totale opbrengst op dit moment op ruim 985.000 euro wordt geraamd. In de praktijk zien we overigens vaak dat Green Belts hun besparingen te laag inschatten. Met behulp van

de ervaringen die in het RKZ waren opgedaan konden veel sneller kansrijke projecten worden geïdentificeerd en gestart. De grootste winst kan geboekt worden met projecten die de beschikbare capaciteiten (bedden, OK's) beter benutten. Hierdoor is het na het uitvoeren van een project mogelijk met dezelfde capaciteit meer patiënten te behandelen en dus meer inkomsten te genereren. Een mooi voorbeeld dat in beide ziekenhuizen is uitgevoerd gaat over het op tijd beginnen op de operatiekamers. Iedereen weet dat er te laat wordt begonnen en iedereen heeft het idee dat het komt doordat de ander steeds te laat is. Het zoeken naar oorzaken ontaardt dan al snel in wederzijdse beschuldigingen en niet tot een oplossing. Met behulp van Six Sigma wordt de CTQ (op tijd beginnen) bepaald en gemeten (hoe laat beginnen we nu eigenlijk?). Vervolgens worden alle oorzaken van verstoringen opgespoord en geanalyseerd. Al snel bleek dat er niet één groep verantwoordelijk was voor het te laat starten, maar een complex van factoren, vooral van organisatorische aard. Hierdoor werd enerzijds medewerking van alle betrokkenen verkregen en anderzijds konden maatregelen getroffen worden op de meest relevante invloedsfactoren. Het project leverde in beide ziekenhuizen besparingen op van meer dan 200.000 euro en voorkwam een hoop ergernis, niet in de laatste plaats bij de patiënt.

Six Sigma heeft ook een positief effect op de kwaliteit van de zorg. Wanneer het productieproces in bijvoorbeeld een autofabriek niet goed is georganiseerd dan zal de klant die de auto koopt daar weinig last van hebben. Zelfs als het proces veel fouten en verspillingen voortbrengt zal de klant, mits er een goede eindcontrole is, ook daarvan weinig merken. In de zorg ligt dat geheel anders. Een slecht georganiseerd zorgproces resulteert direct in een slechte beleving van de kwaliteit door de patiënt. Verspillingen en fouten heten in de zorg complicaties en dienen in het belang van de patiënt en diens veiligheid altijd

vermeden te worden. Doordat de patiënt onderdeel uitmaakt van het zorgproces snijdt het mes met Six Sigma dus aan twee kanten. De kosten dalen en de kwaliteit stijgt.

Het geld voor het oprapen

Met Six Sigma kunnen, zo blijkt ook uit de industrie, grote resultaten worden geboekt op het gebied van kostenbesparing en kwaliteitsverbetering. Wanneer we Six Sigma vergelijken met Total Quality Management (TQM) dan zijn er op het eerste gezicht heel veel overeenkomsten. TQM is echter vooral een filosofie die van bovenaf door de leiding over de organisatie wordt uitgestort. Six Sigma begint concreet met het opleiden van medewerkers die de problemen binnen de organisatie kennen en die er last van hebben. Het blijkt steeds weer heel motiverend als medewerkers in staat gesteld worden om zelf problemen op te pakken en op te lossen. Slechts zelden weet een Green Belt **niet** of hij moet beginnen, bijna altijd weet hij niet **waar** hij moet beginnen. De verbetermogelijkheden liggen namelijk, ook in ziekenhuizen, voor het oprapen.

LITERATUUR

- Institute of Medicine. (1999) *To Err is Human: Building a Safer Health System*. National Academy Press. Washington.
- Van den Heuvel, J.; Does, R.J.M.M.; and Vermaat, M.B. (2004) Six Sigma in a Dutch hospital: Does it work in the Nursing Department? *Quality and Reliability Engineering Int.* 20, pp. 419-426.
- Van den Heuvel, J.; Koning, L.; Bogers, A.J.J.C.; Berg, M.A.; Van Deijen, M.E.M. (2005a) An ISO Quality Management System in a Hospital: Bureaucracy or just Benefits. *International Journal of Health Care Quality Assurance* 18 (5) pp. 361-369.
- Van den Heuvel, J.; Does, R.J.M.M.; Verver, J.P.S. (2005b) Six Sigma in Health Care: Lessons learned from a Hospital. *International Journal of Six Sigma and Competitive Advance* 1 (4).

JAAP VAN DEN HEUVEL is voorzitter Raad van Bestuur Canisius Wilhelmina Ziekenhuis in Nijmegen.
E-mail: <j.v.d.heuvel@cwz.nl>.



SCHIJN BEDRIEGT SOMS, EN SOMS NIET

Sommige verschijnselen die bijzonder lijken, zijn onvermijdelijk. Zo is het is niet zinvol om te klagen over het feit dat (ongeveer) de helft van alle ziekenhuizen onder het gemiddelde presteert: dit zit praktisch opgesloten in het begrip gemiddelde.

Onlangs was het Nijmeegse hartcentrum in het nieuws: het aantal sterfgevallen bij hartoperaties zou daar twee maal zo hoog zijn als het landelijk gemiddelde. Hoe bijzonder zou dat eigenlijk zijn? Hier kan de statistiek uitkomst bieden

IVO ADAN EN FRED STEUTEL

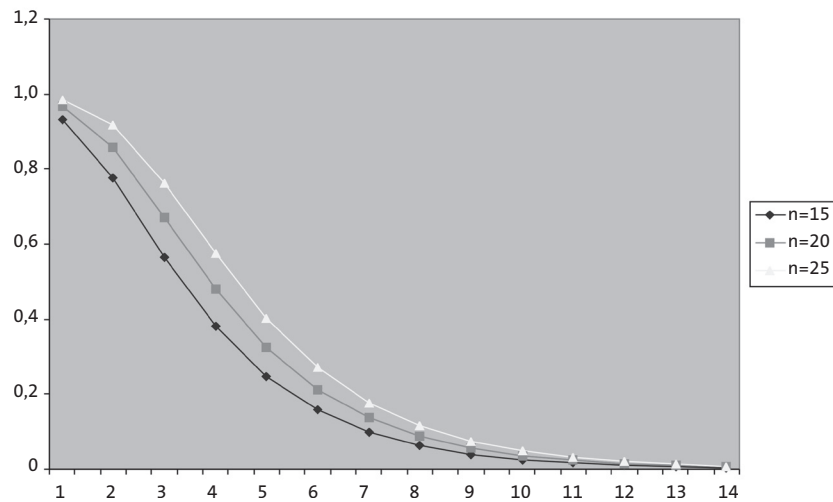
Om bovenstaande vraag te kunnen beantwoorden heb je een model nodig en een aantal gegevens. Een heel eenvoudig model gaat ervan uit dat alle hartcentra evenveel operaties doen, waarbij een door het toeval bepaald aantal sterfgevallen optreedt. Het ligt voor de hand om dit aantal binomiaal verdeeld te veronderstellen: onafhankelijke experimenten met een zekere kans op een sterfgeval, 'succes'. Omdat de succeskans klein is en het aantal operaties groot, zullen de aantallen sterfgevallen dan bij goede benadering Poisson-verdeeld zijn met telkens dezelfde verwachting. We willen nu de volgende kans uitrekenen: $P(\text{Max} > 2 \text{ Gem})$, waarbij Max het maximum is van een aantal onafhankelijke, Poisson verdeelde aantallen en Gem het gemiddelde daarvan.

We formuleren de oorspronkelijke vraag nog een keer: Wat is de kans dat in het meest onfortuinlijke van de (in aanleg gelijkwaardige) hartcentra het aantal sterfgevallen meer dan twee

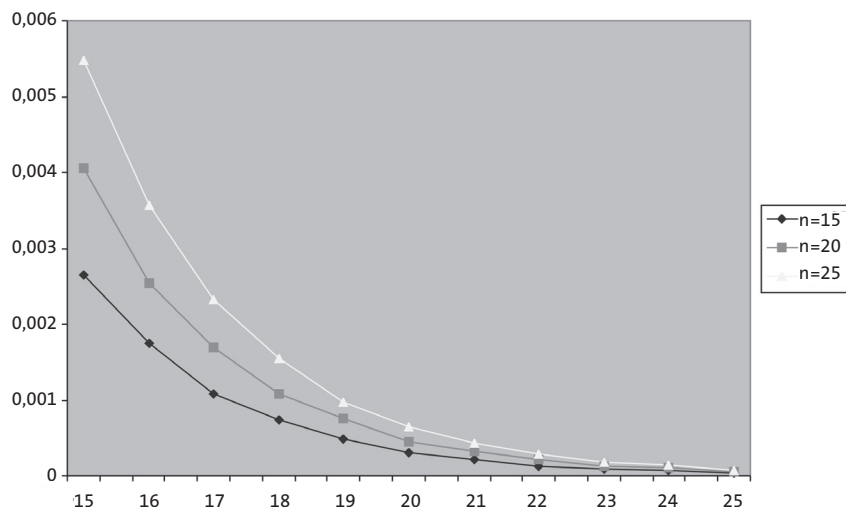
keer zo hoog is als het landelijk gemiddelde? Als we er van uitgaan dat de aantallen sterfgevallen onafhankelijk zijn en Poisson verdeeld, dan moeten we nog twee parameters kiezen: n het aantal Nederlandse hartcentra en μ het verwachte aantal sterfgevallen per hartcentrum per jaar. We zullen straks waarden voor deze parameters invullen.

In de onderstaande figuren hebben we voor een aantal waarden van n de kans $P(\text{Max} > 2 \text{ Gem})$ uitgezet als functie van μ . Deze kans is niet eenvoudig exact uit te rekenen, maar met een simpel simulatieprogrammaatje gaat dat heel snel en heel nauwkeurig. Voor dit soort simulaties verwijzen we naar het boekje *Spelen met kansen* van Henk Tijms.*

Het resultaat laat zien dat deze kans lang niet voor alle waarden van n en μ klein is. Omdat de sterftcijfers niet openbaar gemaakt worden, moeten we naar μ een slag slaan. Dat doen we als volgt: we lezen op internet dat er in Nederland



Figuur 1: De kans $P(\text{Max} > 2 \text{ Gem})$ uitgezet als functie van μ voor $1 \leq \mu < 15$.



Figuur 2: De kans $P(\text{Max} > 2 \text{ Gem})$ uitgezet als functie van μ voor $15 \leq \mu \leq 25$.

ongeveer 15.000 hartoperaties per jaar worden uitgevoerd in 15 hartcentra. De sterftekans bij zo'n operaties is volgens een andere site (natuurlijk afhankelijk van de conditie van de patient en de kwaliteit van het hartcentrum) ongeveer 2%. Dat geeft gemiddeld per jaar 300 sterfgevallen en per hartcentrum gemiddeld 20; we kunnen dus $\mu = 20$ kiezen.

Bij de waarden $n=15$ en $\mu=20$, wordt de kans heel erg klein: 0.0003; er zou in Nijmegen best iets

aan de hand kunnen zijn, maar we weten natuurlijk niet wat; het kan aan het ziekenhuis liggen of aan de patienten.

* Henk Tijms, *Spelen met Kansen*. Epsilon, Utrecht, 1999.

Ivo ADAN is universitair hoofddocent stochastische beslis-kunde aan de TU Eindhoven.

E-mail: <iadan@win.tue.nl>.

FRED STEUTEL is emeritus hoogleraar kansrekening aan de TU Eindhoven; hij is redacteur van STATOR.

E-mail: <f.w.steutel@tue.nl>.



OPERATIONS RESEARCH VOOR RANGEERPLANNING

Treinen variëren in lengte, in de spits zijn ze langer dan daarbuiten. Dat betekent dat gedurende een dag treinstellen moeten worden toegevoegd of afgekoppeld om tot de optimale lengte van een trein te komen. Afgekoppelde treinstellen moeten naar een rangeerterrein worden gebracht, ze kunnen niet langs de perrons blijven staan tot ze weer nodig zijn. Ook worden treinstellen die 's nachts 'geparkeerd' zijn inwendig schoongemaakt waarvoor ze langs een speciaal perron geplaatst dienen te worden.

Dat alles resulteert in een groot aantal rangeerbewegingen. Dit artikel gaat over wiskundige methoden om deze rangeerbewegingen deels geautomatiseerd in te plannen. Het artikel is gebaseerd op de dissertatie van de auteur (Lentink 2006).

RAMON M. LENTINK

Dagelijks vervoert NS Reizigers 1.000.000 reizigers over 2.800 kilometer spoor en via 385 stations. NS Reizigers gebruikt hiervoor 2.800 treinstellen met 228.000 zitplaatsen. Dit vervoersproces vereist

veel coördinatie en een complex planningsproces. Dit planningsproces omvat onder andere het ontwikkelen van de dienstregeling, het maken van diensten voor personeel en materieel, en

het maken van operationele rangeerplannen. Huisman et al (2005) geven een overzicht van het gebruik van Operations Research voor de diverse onderdelen van dit planningsproces.

Zoals bij alle vormen van personenvervoer kent het vervoeren van reizigers per trein de karakteristieke pieken in de vraag naar vervoer, te weten de ochtend- en de avondspits. In deze perioden zet NS Reizigers bijna al het beschikbare materieel in voor het reizigersvervoer. Echter, tussen deze pieken wordt er minder materieel ingezet in verband met de efficiëntie van de treinloop. Het materieel dat tijdelijk niet nodig is voor het uitvoeren van de dienstregeling wordt opgesteld op een rangeerterrein. Op deze wijze is de hoofdinfrastructuur maximaal beschikbaar voor het vervoer van reizigers.

Ongeveer 130 planners bij NS Reizigers houden zich bezig met het maken van rangeerplannen voor ongeveer 30 rangeerterreinen. De rangeerplanning wordt voor elke dag en voor ieder rangeerterrein opnieuw gemaakt. Deze rangeerplanning is momenteel een belangrijke bottleneck in het planningsproces bij NS Reizigers. De reden hiervoor is dat bijna elke wijziging in een ander onderdeel van het planningsproces wijzigingen van één of meer rangeerplannen tot gevolg heeft.

Aangezien NS Reizigers bijna geen nachttreinen rijdt, wordt in het begin van de ochtend vrijwel al het materieel van de rangeerterreinen gehaald en toegewezen aan de dienstregeling. Daarom is het doel van de rangeerplanning om het logistieke proces in de volgende ochtend zo soepel mogelijk op te starten. Andere karakteristieken waarmee rekening wordt gehouden zijn robuustheid van de planning (het minimaliseren van het effect van kleine verstoringen op de planning) en efficiënt gebruik van de productiemiddelen (onder andere energie en infrastructuur).

Naast het opstellen van treinstellen zijn er diverse andere processen die een belangrijk onderdeel van de rangeerplanning vormen. Dit

zijn het routeren van de treinstellen over de infrastructuur van het station, het plannen van het personeel dat de rangeeractiviteiten gaat uitvoeren, en het schoonmaken van en klein onderhoud aan treinstellen.

Voor realistische instanties is het gehele probleem dusdanig complex dat er in mijn onderzoek voor is gekozen om het probleem op te splitsen in verschillende deelproblemen. De deelproblemen die worden beschouwd zijn:

- * Het toewijzen van aankomende treinstellen aan vertrekkende treinen.
- * Het opstellen van treinstellen op de opstelsporen van het rangeerterrein.
- * Het plannen van het inwendig reinigen van treinstellen.
- * Het routeren van treinstellen over de infrastructuur van het station en het rangeerterrein.

Met personeelsplanning, uitwendig reinigen en onderhoud van treinstellen wordt niet expliciet rekening gehouden. In dit artikel worden de eerder genoemde deelproblemen geïntroduceerd en wordt er een indicatie gegeven van de achterliggende wiskundige modellen en algoritmen, welke hiervoor worden gebruikt.

Het toewijzen van treinstellen

Het toewijzen van aankomende treinstellen aan vertrekkende treinen is het probleem dat in de eerder genoemde opsplitsing als eerste wordt opgelost. Voor iedere trein die aankomt of vertrekt beschrijft de dienstregeling het tijdstip en het perron van aankomst of vertrek. Daarnaast beschrijft de dienstregeling ook de exacte volgorde van typen treinstellen in iedere trein. Deze configuraties mogen niet worden gewijzigd.

Het grootste deel van de treinstellen in aankomende treinen is op voorhand al toegewezen aan de vertrekkende treinen. Dit geldt met name voor doorgaande treinen, die een korte periode langs een per-

ron halteren om reizigers in en uit te laten stappen, waarna de treinen hun weg vervolgen naar andere stations. Voor de resterende treinstellen moet de rangeerplanner een toewijzing maken.

De treinstellen worden onderverdeeld in families en binnen een familie bestaan vaak twee typen. De typen binnen een familie onderscheiden zich door het aantal wagons in dit type.

De belangrijkste doelstellingen zijn om aankomende (resp. vertrekkende) treinstellen uit dezelfde trein zoveel mogelijk toe te wijzen aan dezelfde vertrekkende (resp. aankomende) trein. Simpel geformuleerd: een vertrekkende trein bestaat zoveel mogelijk uit treinstellen uit eenzelfde aankomende trein. Daarnaast is het wenselijk om zo veel mogelijk toewijzingen te maken met een minimaal verschil in tijd tussen aankomst en vertrek. Deze doelstellingen resulteren in een minimale hoeveelheid rangeerwerk.

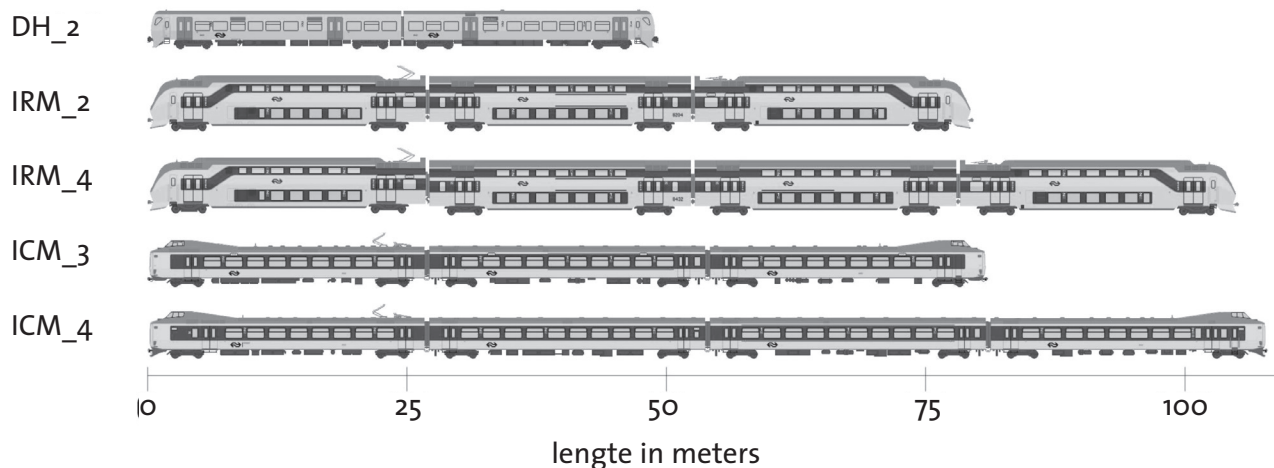
Dit probleem is gemodelleerd als een geheel-tallig lineair programmeringsprobleem, waarin voor elke trein een kortste pad probleem wordt opgelost en waarin de resultaten van deze kortste paden simultaan aan elkaar worden toegewezen.

Het opstellen van treinstellen

Gegeven een toewijzing van treinstellen, moeten de treinstellen worden opgesteld op opstelsporen van het rangeerterrein. De toewijzing geeft informatie over de aankomst- en vertrektijden van de treinstellen die opgesteld moeten worden. Daarnaast zijn er sporen die van één of twee kanten te benaderen zijn. De sporen die van twee kanten te benaderen zijn resulteren in extra flexibiliteit maar ook in extra complexiteit voor het maken van een opstelling.

Bij het opstellen spelen twee restricties een belangrijke rol. De eerste is dat de lengte van het opstelspoor ten alle tijden voldoende moet zijn om de treinstellen te kunnen opstellen. De tweede is dat een treinstel een ander treinstel niet in de weg mag staan. Zulke blokkerende treinstellen zullen eerst weg gerangeerd moeten worden, voor de andere treinstellen kunnen aankomen of vertrekken. Dit levert extra complexiteit en is dus ongewenst.

Naast het opstellen van zoveel mogelijk treinstellen is het van belang om (geschatte) routingskosten te minimaliseren en robuuste plan-



Figuur 1: Enkele voorbeelden van typen treinstellen.

nen te maken. Onderdelen van robuustheid zijn sporen met één type treinstel (hier is de volgorde van treinstellen op een opstelspoor niet van belang) en het groeperen van treinstellen voor dezelfde vertrekkende trein naast elkaar op hetzelfde spoor in de juiste volgorde.

Het probleem kan worden gemodelleerd als een Set Partitioning Probleem. Het nadeel van deze formulering is dat het leidt tot een exponentieel toenemend aantal beslissingvariabelen. Dit kan worden omzeild door het toepassen van kolomgeneratie, waarin slechts een deel van alle beslissingsvariabelen expliciet wordt beschouwd.

Schrijver (2003) heeft een model ontwikkeld waarin het toewijzen van aankomende aan vertrekkende treinstellen en het opstellen geïntegreerd worden opgelost. Hij is in staat om betere oplossingen te vinden ten koste van extra reken-tijd.

Het routeren van treinstellen

Gegeven een opstelling van treinstellen moeten in dit probleem routes voor treinstellen van en naar de juiste opstelsporen worden gezocht. Hierbij moet rekening worden gehouden met diverse typen reserveringen van de infrastructuur. In tegenstelling tot de dienstgeregelde treinen zijn de aankomst- en vertrektijden hierbij enigszins flexibel.

Infrastructuur kan op één moment slechts door één trein worden gereserveerd om conflicten in de uitvoering van het plan te voorkomen. Hierbij geldt dat de routes van de doorgaande treinen al eerder zijn bepaald en een hogere prioriteit hebben dan de routes die volgen uit het rangeerplan. Daarnaast gelden er plantijden voor verschillende activiteiten welke moeten worden gerespecteerd.

We zijn op zoek naar korte routes, waarin de

starttijden zo min mogelijk afwijken van de voorkeurstijdstippen.

Het probleem wordt trein-voor-trein opgelost met een uitbreiding van A* Search. A* Search op zichzelf is een uitbreiding van Dijkstra's algoritme voor het kortste pad probleem, waarbij rekening wordt gehouden met schattingen van resterende kosten. A* Search geeft richting aan Dijkstra's algoritme. Het trein-voor-trein oplossen leidt tot sub-optimale oplossingen. Om dit effect te beperken wordt een verwisselingsheuristiek toegepast.

Het reinigen van treinstellen

Alle treinstellen die overnachten op het rangeerterrein moeten inwendig gereinigd worden. Dit reinigen vindt plaats langs een speciaal perron. De plantijden voor het reinigen van typen treinstellen worden gegeven in man-minuten. Aangezien het aantal schoonmakers varieert verschillen de doorlooptijden voor het reinigen van een treinstel gedurende de nacht.

We nemen aan dat er langs één zijde van het reinigingsperron wordt gereinigd. Het spoor langs de andere zijde wordt gebruikt voor het routeren van treinstellen die moeten worden gereinigd of die net gereinigd zijn.

De doelstelling van het reinigingsproces is om zoveel mogelijk treinstellen kort na aankomst te reinigen. Het reinigen vlak voor vertrek conflicteert met de globale doelstelling om de volgende ochtend zo soepel mogelijk te beginnen. Het reinigen halverwege tussen aankomst en vertrek leidt ertoe dat treinstellen zowel voor als na het reinigen moeten worden opgesteld. Dit leidt tot extra complexiteit, die eventueel vermeden zou kunnen worden.

Het probleem kan worden geformuleerd als een geheeltallig lineair programmeringsprobleem. Dit model kan efficiënt worden opgelost met commercieel beschikbare software.

Slotwoorden

In dit artikel wordt het probleem van de operationele rangeerplanning besproken. Ondanks dat dit probleem slechts een klein onderdeel van het gehele planningsproces is van NS Reizigers, is dit een complex probleem.

Om dit proces te stroomlijnen kan gebruik gemaakt worden van wiskundige modellen en algoritmen. In dit artikel wordt slechts verwezen naar deze modellen en algoritmen. In Lentink (2006) worden de modellen en algoritmen in detail beschreven en worden rekenresultaten gepresenteerd.

LITERATUUR

Huisman, D., Kroon, L.G., Lentink, R.M., Vromans, M.J.C.M. (2005). Operations Research in passenger railway transportation, *Statistica Neerlandica*, 59(4), 467-498.

Lentink, R.M. (2006). *Algorithmic decision support for shunt planning*. Dissertatie. Rotterdam: Erasmus Universiteit Rotterdam.

Schrijver, A. (2003). *Planning van opstelsporen*. Research rapport. Amsterdam: Centrum voor Wiskunde en Informatica.

RAMON LENTINK studeerde van 1994 tot en met 1999 Bedrijfseconometrie aan de Vrije Universiteit te Amsterdam. Sindsdien is hij werkzaam bij ORTEC, waar hij zich in eerste instantie heeft gericht op modellen en algoritmen voor personeelsplanning in de luchtvaart en op het spoor. In de periode 2000 tot en met 2005 heeft hij naast zijn deeltijdwerk voor ORTEC een promotieonderzoek aan de Erasmus Universiteit verricht. Dit onderzoek was geïnitieerd door dr. Richard Freling en is afgerond onder begeleiding van promotoren prof. dr. Leo Kroon en prof. dr. ir. Jo van Nunen. Samen met Dennis Huisman won hij in 2002 de eerste prijs in de Management Science in Railroad Applications Student Competition georganiseerd door de Rail Applications Special Interest Group, een onderdeel van het Institute for Operations Research and the Management Sciences. Sinds 2006 is hij voltijds in dienst van ORTEC, waar zijn aandachtsgebied is verbreed naar algoritmiek binnen de gehele logistieke tak van ORTEC. E-mail: <rlentink@ortec.nl>.

AGENDA

6-16 juni 2006

Het Econometrisch Instituut in Rotterdam viert haar 50-jarig bestaan. Van 6 tot 16 juni 2006 zijn er vijftien prominente buitenlandse toponderzoekers te gast in Rotterdam. Nadere informatie via <info@ei50.org> of op <www.ei50.org>.

10-14 juli 2006

Ter ere van de 65ste verjaardag van Piet Groeneboom organiseert het Lorenz Center in Leiden een meerdaagse bijeenkomst. Tijdens een 3-daagse workshop op 10-12 juli geeft een aantal experts op het gebied van de kansrekening en statistiek een voordracht. Verder presenteren specialisten hun werk aan jonge onderzoekers. Voor informatie zie www.lorentzcenter.nl/lc/web/2006/20060710/info.php3?wsid=189.

2-7 juli 2006

Het belangrijkste doel van ICOTS-7 is uitwisseling van de nieuwste inzichten omtrent statistiekonderwijs. Informatie: <www.maths.otago.ac.nz/icots7>.

27-31 augustus 2006

De 27ste ISCB conferentie 2006 vindt plaats in Geneve en is een internationaal forum voor het presenteren van methoden, toepassingen en nieuwe inzichten op het terrein van biostatistiek bij medisch onderzoek en de medische praktijk. Informatie: <www.iscb2006.info>.

18-20 september 2006

Doel van de Sixth Annual ENBIS meeting in Wroclaw (Polen) is om statistici uit de praktijk samen te brengen met academische statistici, consultants, Six Sigma black belts, en andere professionals die zich bezig houden met bedrijfs- en industriële statistiek. Zie ook: <www.enbis.org>..



Eugene Lukacs (links) en Béla Székfalvi-Nagy

KANSREKENING OP REIS

FRED STEUTEL

Wie verre reizen doet kan veel verhalen. Veel van de reizen die ik maakte, betroffen bezoeken aan congressen of hadden anderzijds een kanstheoretisch karakter. Soms kreeg een dergelijke reis toch een heel persoonlijk cachet. Een paar voorbeelden.

Gevlekte pingpongballen

Jaren geleden werkte ik aan het volgende probleem: kies willekeurig ('random') n punten op een cirkel en plaats in elk van die punten een cirkelboogje met een vaste lengte. De bedoeling is dat is de hele cirkel met zulke boogjes overdekt raakt. Je kunt nu vragen stellen als: Wat is bij gegeven n de kansverdeling van de lengte van de grootste nog-niet-overdekte cirkelboog? Of: Hoe groot moet n gekozen worden om de kans op volledige overdekking minstens 95 procent te maken? Aan dit soort vragen is door veel mensen met plezier gerekend. Leuke wiskunde, maar het echte probleem, dat te maken had met de over-

dekking van virussen door anti-lichamen, was een dimensie hoger: kies willekeurig n punten op een bol, en breid elk van die n punten uit tot een bolkapje van gegeven omvang. Hoe groot moet n zijn om er (bijna) zeker van te zijn dat de hele bol overdekt is, d.w.z. het virus uitgeschakeld? Dat was een veel moeilijker probleem, waaraan maar weinig echt gerekend kon worden. Simulatie lag dus voor de hand, maar dat was toen, veertig jaar geleden, ook niet eenvoudig. Tijdens een congres-uitstapje in Hongarije kwam ik in de bus naast een dame te zitten – vriendin van de beroemde Zolotarev, die twee stoelen verder wiskundig in gesprek was. Haar naam kwam mij bekend voor, en zij bleek de dochter te zijn van een van de mensen die aan het virusprobleem hadden gewerkt. Zij was geen wiskundige, maar ze wist alles van de manier waarop haar vader simulaties had uitgevoerd: een groot aantal pingpongballen werd bedekt met een soort eierrek, een plaat met vrij kleine ronde openingen, waar die pingpongbal-

len een klein stukje doorheen staken. Vervolgens werd die plaat met ballen bespoten met zwarte verf. Na drogen werden de pingpongballen goed geschud en werd de operatie herhaald. Na iedere spuitbeurt werd gemeten welk deel van de ballen zwart geworden was. Zij was te klein om te begrijpen waar het allemaal toe diende, maar ze vertelde dat ze jarenlang met zwartgekleurde pingpongballen had gespeeld.

Weense elegantie in Washington

Door mijn belangstelling voor karakteristieke functies (Fourier-getransformeerden van kansverdelingen) was ik in contact gekomen met Eugene Lukacs, auteur van een aantal boeken op dat terrein. Ik ontmoette hem in het wiskunde-dorado Oberwolfach, waar hij een bijeenkomst organiseerde. Hij raakte geïnteresseerd in mijn werk, waaraan hij uitgebreid aandacht gaf in de tweede druk van zijn bekendste boek. Lukacs (1906), geboren in Szombathely, Hongarije, en opgeleid in Wenen verhuisde in 1938 met zijn Weense vrouw Elizabeth naar de VS. Toen ik hem leerde kennen werkte hij aan de Catholic University of America in Washington. Wat later had ik een sabbatical van een half jaar aan de Johns Hopkins Universiteit en woonde ik met vrouw en kinderen in Baltimore, een uur rijden van Washington. Op een zondag werden we door Eugene en Elizabeth Lukacs uitgenodigd voor de lunch. Zij woonden in een buitenwijk van Washington, maar als je hun huis binnenkwam waande je je in Wenen: Europese meubelen, Europese wandversiering en vooral Weens porcelein, koffie uit zeer breekbare kopjes, alleen voor de volwassenen, en zelfgemaakte versnaperingen op dito schoteltjes. De niet meer zo jonge gastvrouw was Europees gekleed en had een elegantie, waar wij niet echt aan mee konden doen, en die je nu in Europa misschien ook alleen nog in Wenen ziet. In alle opzichten gesterkt reden wij terug naar Baltimore.

Jaren later bezocht ik een herdenkingsbijeenkomst voor Lukacs – 1987 in Helsinki overleden – in Bowling Green, waar Eugene na zijn pensionering in Washington jarenlang een plek had. In maart was dat een koude en desolate plek. De campus werd bijna helemaal omgeven door een begraafplaats, die tot mijn verbazing vele graven herbergde van gesneuvelden in de Burgeroorlog van de jaren 1860. Elizabeth Lukacs was te oud om nog *acte de présence* te geven. We hebben nog een aantal jaren kerstkaarten uitgewisseld. Met haar overlijden in 1993 is een stukje Wenen uit Washington verdwenen.

Tossen in Hue

Vijftien jaar geleden was ik op bezoek in Hanoi, uitgenodigd als tegenprestatie voor verleende diensten aan Vietnamese wiskundigen die de Universiteit van Amsterdam bezochten. Alle *invited speakers* kregen 200 Amerikaanse dollars. Enerzijds kon je daar makkelijk een week van slapen en eten, maar anderzijds waren er in Hanoi ook hotels die 200 dollar per dag kostten. Weer anderzijds werd mij verteld dat het jaarsalaris van een hoogleraar ook ongeveer 200 dollar bedroeg. Eén van mijn verplichtingen was het geven van een eenvoudige voordracht over kansrekening voor een groep leraren in Hue, de oude keizerstad, berucht van het Tet-offensief van 1968. Ik had een aardig verhaal over gooien met muntstukken: doe één worp met n munten, laat liggen wat 'kop' boven heeft, en herhaal dit met de overblijvende munten, totdat alle munten 'kop' laten zien. De wat onverwachte vraag luidde: Wat is de kansverdeling van het aantal munten in de laatste worp? Afgezien van het taalprobleem – de leraren kenden maar weinig Engels – was er een materiële handicap: het Vietnamese geld bestond geheel uit bankbiljetten, er waren geen munten.

*FRED STEUTEL is emeritus hoogleraar kansrekening aan de TU Eindhoven; hij is redacteur van STATOR.
E-mail: <f.w.steutel@tue.nl>.*



POLITIEK, CULTUUR, TAAL EN RELIGIE

Een analyse van het stemgedrag tijdens het Eurovisie Songfestival

De Fransen houden niet van liedjes in het Engels. Nederland baseert de puntenverdeling vooral op de eigenschappen van de liedjes en heeft een sterke voorkeur voor liedjes gezongen door vrouwen. De Baltische staten spelen elkaar systematisch punten toe. Cyprus en Ierland hebben een sterke voorkeur voor liedjes uit landen met dezelfde religie. Turken woonachtig in Europa stemmen massaal op de Turkse bijdrage aan het Eurovisie Songfestival...

Een econometrische analyse van dertig jaar Eurovisie Songfestival biedt een fraai inkijkje in het Europese stemgedrag. Zo heeft een aantal landen een uitgesproken voorkeur voor de liedjes afkomstig van buurlanden en andere landen in hun nabijheid, zelfs na correctie voor overeenkomsten in taal, cultuur en religie. Het ligt voor de hand dit te interpreteren als bewijs voor ‘vriendjespolitiek’. Echter, de veelgehoorde beschuldigingen van politiek gekleurd stemgedrag aan het adres van Scandinavië, Oost-Europa en het voormalige Joegoslavië worden niet ondersteund door de data. Alleen tegen de Baltische staten is er overtuigend bewijs voor het bevoordelen van elkaars liedjes.

LAURA SPIERDIJK EN MICHEL VELLEKOOP

Het Eurovisie Songfestival. Volgens sommigen een leuk en spannend evenement en volgens anderen juist een afschuwelijk *camp* fenomeen dat het beste zo snel mogelijk van de televisie kan verdwijnen... Over smaak valt uiteraard niet te twisten. Echter, welke kant men ook kiest, het songfestival is zonder enige twijfel voer voor econometristen en statistici. De historische gegevens van de afgelopen vijftig jaar – inmiddels op eenvoudige wijze via internet te bemachtigen¹ – bevatten een rijke hoeveelheid informatie over het stemgedrag van de deelnemende landen en bieden een fraai inkijkje in het Europese stemgedrag tijdens het Eurovisie Songfestival.

De regels van het songfestival zijn simpel. De deelnemende landen sturen een liedje in. Tijdens het festival worden deze ten gehore gebracht en daarna worden de liedjes beoordeeld door de deelnemende landen zelf, die de punten 1,2,3,4,5,6,7,8,10,12 mogen toekennen. Uiteraard mogen de landen geen punten aan zichzelf toekennen. Omdat er de laatste decennia altijd meer dan 20 landen meededen, kent elke jury aan een groot aantal liedjes geen enkel punt toe. Vóór 1997/1998 werden de punten door professionele jury's toegekend. In 1997/1998 is men overgestapt op *televoting*, waarbij iedereen kan stemmen via telefoon of *sms*.²

land	# jaren	# stemmen	land	# jaren	# stemmen	land	# jaren	# stemmen
België	26	526	Italië	17	335	Polen	8	184
Bosnië	9	207	Joegoslavië	13	253	Portugal	27	550
Cyprus	21	448	Kroatië	11	254	Roemenië	5	118
Denemarken	22	450	Letland	4	92	Rusland	7	162
Duitsland	28	574	Litouwen	4	91	Slovenië	9	207
Estland	9	208	Luxemburg	19	366	Slowakije	3	70
Finland	24	482	Macedonië	3	70	Spanje	29	596
Frankrijk	28	579	Malta	14	315	Turkije	25	520
Griekenland	23	479	Monaco	5	89	Verenigd Koninkrijk	29	596
Hongarije	4	94	Nederland	25	512	Zweden	28	579
Ierland	27	554	Noorwegen	28	573	Zwitserland	25	506
Israel	24	491	Oostenrijk	26	532			

Tabel 1: Het aantal deelgenomen jaren per land in de periode 1975-2003 en het aantal geanalyseerde stemmen per land.

Met het Eurovisie Songfestival 2006 in het vizier (18 en 20 mei in Athene) steken beschuldigingen van vriendjespolitiek tussen o.a. Oost-Europese landen en de Baltische staten opnieuw de kop op in de media. Zulke aantijgingen zijn overigens zo oud als het festival zelf en behoren inmiddels tot de jaarlijks terugkerende folklore. Wel lijken de beschuldigingen elk jaar extremer te worden. Zo beweerde *runner-up* Malta dat de winnaar van vorig jaar (Griekenland) gewonnen heeft dankzij de hoge scores toegekend door andere landen met een substantiële religieus *orthodoxe* bevolking zoals Roemenië, Macedonië

en Bosnië. Genoeg redenen dus om het Eurovisie Songfestival aan een grondige econometrische analyse te onderwerpen.

Aanpak

Ons onderzoek richt zich op de periode 1975-2003. We beperken ons tot deze jaren omdat het qua spelregels een relatief homogene periode is. We nemen alleen de landen in beschouwing die in deze periode ten minste drie keer meededen; dit zijn er totaal 36. Tabel 1 geeft een overzicht van de geanalyseerde landen en het aantal jaren dat

landen	gemiddelde bias (in punten)
Griekenland → Cyprus	7.7
Cyprus → Griekenland	9.2
België → Nederland	0.4
Nederland → België	1.5
Denemarken → Zweden	3.5
Zweden → Denemarken	2.2
Letland → Estland	4.0
Estland → Letland	5.9
Letland → Litouwen	4.6
Litouwen → Letland	6.1
Estland → Litouwen	0.3
Litouwen → Estland	6.1
Macedonië → Roemenië	6.8
Roemenië → Macedonië	7.7

Tabel 2: Gemiddelde bias over de periode 1975-2003 tussen verschillende landen.

deze landen participeerden gedurende de periode 1975-2003. Om erachter te komen of vriendjespolitiek inderdaad een rol speelt tijdens het Eurovisie Songfestival, richten we ons op een aantal factoren. Ten eerste kijken we of geografische aspecten invloed hebben op de puntenverdeling. Uiteraard corrigeren we eerst voor overeenkomsten in taal en cultuur tussen landen. Immers, voorkeuren voor naburige landen zouden ook veroorzaakt kunnen worden door linguïstische en culture banden tussen de betreffende landen. Als een land na correctie voor laatstgenoemde factoren nog steeds een significante voorkeur voor naburige landen heeft, kan dit op politiek gekleurd stemgedrag duiden. Tevens onderzoeken we of beschuldigingen zoals die van Malta inzake mogelijk ‘religieus’ stemgedrag door de data worden ondersteund.

Hoewel dit aspect wellicht vergezocht lijkt, is al vaker in de economie aangetoond dat religie de beslissingen van individuen kan beïnvloeden.³ Derhalve is het interessant om de rol van religie te onderzoeken. Ten derde kijken we of dertig jaar Eurovisie de beschuldigingen aan het adres van Turkse immigranten (en andere inwoners van de Europese Unie met een Turkse afkomst) bevestigt. Sinds de introductie van televoting in 1997/1998 wordt beweerd dat migratielanden zoals Nederland en Duitsland relatief veel punten aan de Turkse inzending geven (‘patriottisch’ stemgedrag). Hoewel dergelijke effecten ook een rol kunnen spelen bij andere etnische groepen, zijn betrouwbare migratiedata voor deze groepen vooralsnog niet beschikbaar. Daarom beperken we ons tot de in Europa wonende Turken. Tot slot zullen we de resultaten betrekken op de veelgehoorde beschuldigingen van vriendjespolitiek in de media aan het adres van een aantal landen-groepen, zoals de Baltische staten, het voormalige Joegoslavië en bepaalde Oost-Europese landen. Hoewel er al redelijk wat econometrisch onderzoek naar het Eurovisie Songfestival is gedaan⁴, richten de bestaande studies zich op de *gemiddelde* invloed van bepaalde factoren op het stemgedrag tijdens het songfestival, waarbij gemiddeld wordt over alle deelnemende landen. Wij kiezen voor een gedetailleerdere aanpak en onderzoeken het effect van bepaalde factoren op het stemgedrag *per individueel land*, zodat we ook per land kunnen kijken naar het waarheidsgehalte van beschuldigingen van vriendjespolitiek.

Model

De historische Eurovisie gegevens vormen een data set in drie dimensies: het aantal punten van een juryland toegekend aan een liedland in een bepaald jaar. Als te verklaren variabelen kiezen we de zogenaamde *voting bias* b_{tij} . Dit is het aantal punten s_{tij} dat land i in jaar t toekent aan een

liedje j , minus het gemiddelde aantal punten s_{tj} dat er in dat jaar door de overige landen aan het lied van land j wordt toegekend.

De reden dat we de bias als verklarende variabele kiezen en niet het aantal toegekende punten, is dat de bias veel informatiever is over de mate van over- of onderwaardering van een lied door een bepaald land. Immers, een hoog (laag) aantal punten aan een lied is pas exceptioneel als het lied van de andere landen gemiddeld genomen juist weinig (veel) punten krijgt. Zo eindigde Griekenland in 1992 als vijfde op een totaal van 23 landen. In dit jaar kreeg het 12 punten toebedeeld door Cyprus. In 2003 werd Griekenland slechts 17de van de 26 deelnemende landen, maar in dat jaar kreeg het land wederom 12 punten van Cyprus. Gegeven de uitslag zijn de 12 punten in 2003 een veel sterkere indicatie voor een bias dan in 1992.

Tabel 2 toont een overzicht van de gemiddelde bias tussen enkele landen die regelmatig beschuldigd worden van vriendjespolitiek, berekend over de periode 1975-2003. Deze tabel laat bijvoorbeeld zien dat Cyprus gemiddeld 9.2 punten meer aan Griekenland geeft dan de overige deelnemers aan het songfestival.

Voor elk land schatten we een regressiemodel om de bias van het betreffende land ten opzichte van de deelnemende liedjes te verklaren uit een groot aantal verklarende variabelen. De factoren die we beschouwen zijn:

- *Geografische factoren*: (1) een indicator variabele die aangeeft of het juryland en het liedland buuren zijn en (2) de afstanden tussen de hoofdsteden van jury- en liedland.
- *Religie*: een indicator variabele die weergeeft of landen een gemeenschappelijke religie hebben.⁵
- *Linguïstische en culturele factoren*: de lexico-statistische⁶ afstand tussen de talen van het juryland en het liedland, de afstanden tussen vier culturele dimensies⁷ van het juryland en het liedland.

- *Performance factoren*: (1) lied gezongen door man, vrouw of groep, (2) een indicatorvariabele die aangeeft of het lied in het Engels is gezongen, (3) een indicatorvariabele die aangeeft of het lied in het Frans is gezongen, (4) de positie van een lied in de volgorde van optreden en (5) een indicatorvariabele die aangeeft of het songfestival plaatsvindt in het liedland.
- *Etnische factoren*: een indicatorvariabele die aangeeft of een juryland behoort tot de groep van landen die sinds de jaren zestig een relatief grote populatie van Turkse afkomst heeft.

Omdat we voor elk individueel land de invloed van de verschillende factoren op de voting bias schatten, is het mogelijk dat de invloed van deze factoren verschilt per land. Deze aanpak biedt daardoor inzicht in de verschillen in het stemgedrag tussen landen. Het regressiemodel dat we schatten voor elk land i schatten heeft dus de volgende vorm:

$$(1) \quad b_{tij} = \alpha_i + \sum_{i=1}^k X_{tij} \beta_i + \varepsilon_{tij}$$

$$E(\varepsilon_{tij} \mid X_{tij}) = 0.$$

Hier is b_{tij} de voting bias van land i naar land j , α_i een intercept, β_i een K -dimensionale kolomvector van coëfficiënten, X_{tij} een K dimensionale rijvector van verklarende variabelen (hierboven beschreven) en ε_{tij} een storingsterm. Voor elk land schatten we model (1) met behulp van de kleinste kwadraten methode.⁸

Resultaten

Omdat we voor maar liefst 36 landen model (1) schatten met 19 verklarende variabelen (d.w.z. $K=19$), richten we ons alleen op de variabelen die significant zijn. Omdat dit er nog steeds veel zijn, geven we geen tabel met schattingsresultaten maar beschrijven we in woorden de belangrijkste

land	significante effecten
Cyprus	hoge waardering voor landen die zingen in een verwante taal lage waardering voor landen die later op de avond optreden hoge waardering voor landen met een gemeenschappelijke religie
Griekenland	hoge waardering voor liedjes gezongen door mannen hoge waardering voor landen die zingen in een verwante taal
Estland	hoge waardering voor landen die later op de avond optreden hoge waardering voor dichtbij gelegen landen hoge waardering voor buurlanden
Letland	hoge waardering voor liedjes gezongen door groepen hoge waardering voor liedjes in het Frans hoge waardering voor buurlanden hoge waardering voor landen met een gemeenschappelijke religie
Litouwen	hoge waardering voor duetten hoge waardering voor liedjes in het Engels hoge waardering voor liedjes in het Frans hoge waardering voor buurlanden
Macedonië	hoge waardering voor gastland hoge waardering voor nabije landen, tenzij een buurland lage waardering voor buurlanden
Nederland	lage waardering voor liedjes door mannen lage waardering voor liedjes door groepen lage waardering voor duetten hoge waardering voor liedjes in het Frans hoge waardering voor landen met vergelijkbare cultuur
België	lage waardering voor liedjes uit het gastland lage waardering voor landen met vergelijkbare cultuur
Frankrijk	lage waardering voor liedjes in het Engels
Monaco, Roemenië & Rusland	geen enkele factor beïnvloedt de bias significant

Tabel 3: Overzicht van landen en de factoren die hun voting bias significant beïnvloeden.

resultaten. Voor een uiteenzetting van de volledige resultaten verwijzen we naar Spierdijk en Vellekoop (2006). De schattingsresultaten maken duidelijk dat taal de belangrijkste factor is in het verklaren van de voting bias. Negen landen prefereren liedjes gezongen in een taal die dicht bij hun eigen taal ligt, terwijl twee landen (Zwitserland en Malta) juist van exotische talen houden. Geografische factoren spelen na taal de belangrijkste rol. Zes landen hebben een voorkeur voor de liedjes van buurlanden, zelfs na correctie voor overeenkomsten in taal en cultuur (Estland, Letland, Litouwen, Slowakije, Slovenië en Zwitserland). Echter, Macedonië heeft juist een afkeer van de bijdragen van buurlanden. Bij veertien landen speelt de afstand tot het liedland een significante rol in het verklaren van de voting bias, waarbij de meeste landen de liedjes prefereren van landen die dicht bij hen in de buurt liggen. Religie beïnvloedt het stemgedrag van zeven van de 36 landen. Een opvallend resultaat is dat het stemgedrag van twee landen die bekend staan als sterk religieus (het katholieke Ierland en het orthodoxe Cyprus) inderdaad significant beïnvloed wordt door godsdienst. Tevens tonen de resultaten aan dat landen met een relatief grote Turkse populatie een significant positieve bias hebben naar de Turkse inzending. Tabel 3 geeft een overzicht van de meest opvallende resultaten.

De schattingsresultaten laten zien dat er grote verschillen zijn in het stemgedrag van de deelnemende landen. De voting bias van sommige landen is niet te verklaren uit de genoemde factoren (wat de suggestie wekt dat deze landen 'objectief' stemmen), terwijl die van andere landen juist sterk afhangt van deze variabelen (suggererend dat die landen 'subjectief' stemmen). De 'braafste' jongetjes van de klas zijn Rusland, Roemenië en Monaco (hun stemgedrag is onverklaarbaar uit de door ons beschouwde factoren), terwijl de Baltische staten, Cyprus en Griekenland

en landen van het voormalige Joegoslavië het 'stoutst' zijn (hun stemgedrag hangt heel sterk af van de door ons bekeken factoren). Het is opvallend dat de gemiddelde bias van Roemenië aan Macedonië wel heel hoog is (zie Tabel 1), maar dat het stemgedrag van Roemenië tegelijkertijd niet uit de beschouwde factoren verklaard kan worden. Ook is het opvallend dat het stemgedrag van Nederland als enige van de deelnemende landen grotendeels afhangt van de eigenschappen van het liedje (d.w.z. taal en samenstelling performers) en niet van andere factoren (op cultuur na).

Door de periode voor en na de introductie van televoting te beschouwen, wordt duidelijk dat zowel patriottisch als religieus stemgedrag sterk zijn toegenomen sinds televoting de professionele jury's verving. Dit mag gerust opvallend genoemd worden. Onlangs heeft de Eurovisie organisatie namelijk overwogen om terug te gaan naar het systeem van professionele jury's vanwege de enorme kritiek op televoting. Uiteindelijk heeft men besloten om televoting toch te handhaven. Onze resultaten tonen aan dat de invloed van niet aan liedjeskwaliteit gerelateerde factoren sinds de introductie van televoting aanzienlijk gegroeid is.

Vriendjespolitiek: mythe of realiteit?

De eerste hamvraag is uiteraard of de veel gehoorde beschuldigingen van vriendjespolitiek aan het adres van verschillende landengroepen zoals het voormalige Joegoslavië, Scandinavië en Oost-Europa ondersteund worden door onze analyse. Het antwoord op deze vraag is verrassend. Hoewel veel landen een duidelijke voorkeur of afkeer voor hun burens hebben en er dus wel degelijk geografisch gestemd lijkt te worden, is er alleen voor de Baltische staten sterk bewijs voor vriendjespolitiek. Deze landen hebben een sterke voorkeur voor elkaar, zelfs na correctie voor overeenkomsten in taal, cultuur en religie. Voor de andere landen die regelmatig beschuldigd worden, toont

het model aan dat niet-politieke factoren (zoals taal, cultuur en religie) het stemgedrag voor het grootste gedeelte kunnen verklaren. Zo blijkt de enorme bias van Griekenland en Cyprus in grote mate verklaard te kunnen worden uit de gemeenschappelijke taal en religie.

Wanneer wint Nederland het Eurovisie Songfestival?

Chauvinistisch als we zijn, is de tweede hamvraag natuurlijk wanneer ons eigen land weer de Eurovisieoverwinning in de wacht gaat slepen. De laatste was in 1975, toen de Enschedese groep Teach-in won met het lied *Ding-a-dong*. Gaan de dames van Treble ons dit jaar naar de victorie leiden? Nederland heeft in elk geval de pech slechts twee burens te hebben. Hoewel wij België rijkelijk met punten bedelen (de gemiddelde bias in de periode 1975-2003 was +1.53 punten), zijn onze zuiderburen een stuk minder hartelijk naar ons toe (gemiddelde bias slechts +0.38 punten). Duitsland is iets deller op ons dan België, maar zal ons met een gemiddelde bias van +0.58 punten zeker niet aan de overwinning helpen. Kortom, als Nederland voor het zoveelste jaar wederom laag eindigt, dan hoeven we dat niet langer alleen te wijten aan onze slechte liedjes. Nederland heeft gewoon niet genoeg burens en de weinige burens die we hebben laten ons behoorlijk in de steek. Wat dat betreft kan Nederland eerder een verre vriend dan een niet zo'n goede buur gebruiken. Ofwel, het is de hoogste tijd om de Nederlands-Cypriotische betrekkingen te versterken...

LITERATUURVERWIJZINGEN

Dyen, I., Kruskal, J. & Black, P. (1992), 'An indo-european classification: A lexostatistical experiment', *Transactions of the American Philosophical Society* 82 (5).
Fenn, D., Suleman, O., Efstathiou, J. & Johnson, N.F. (2005), 'How does Europe make its mind up? Connections, cliques and compatibility between countries in the Eurovision Song Contest', te verschijnen in *Physica A*.
Ginsburgh, V. & Noury, A. (2004), *Cultural voting: The*

Eurovision Song Contest. Working Paper.

Haan, M., Dijkstra, G. & Dijkstra, P. (2005), 'Expert judgement versus public opinion - Evidence from the Eurovision Song Contest', *Journal of Cultural Economics* 29, 59-78.

Hofstede, G. (1980), *Culture's Consequences*, Sage, Beverly Hills, California.

Hofstede, G. (1996), *Cultures and Organizations*, McGraw-Hill, London.

Iannaccone, L. (1998), 'Introduction to the economics of religion', *Journal of Economic Literature* 36, 1465-1496.

Kuran, T. (1994), 'Religious economics and the economics of religion', *Journal of Institutional and Theoretical Economics* 150, 769-775.

Spierdijk, L. & Vellekoop, M. (2006), 'Geography, culture, and religion: Explaining the bias in Eurovision Song Contest voting'. Zie www.home.math.utwente.nl/~spierdijk/.

White, H. (1980), 'Heteroskedasticity consistent covariance matrix estimator and a direct test for heteroskedasticity', *Econometrica* 48, 817-838.

1. Zie bijvoorbeeld <www.eurovisioncontest.co.uk> en <<http://members.fortunecity.com/mcdeil69/1980.htm>>
2. Sinds 2004 is er bovendien een halve finale om te bepalen welke landen door mogen naar de finale.
3. Zie bijvoorbeeld Iannaccone (1998) en Kuran (1994).
4. Zie bijvoorbeeld Fenn et al (2005), Haan et al. (2005) en Ginsburgh en Noury (2004).
5. We hebben hier alleen gekeken naar religies die in substantiële mate vertegenwoordigd zijn in een land, dat wil zeggen dat zij onder ten minste 20% van de gelovigen in een land beleden moeten worden.
6. De lexicostatistische afstand tussen twee talen is gebaseerd op het percentage aan elkaar verwante woorden in de talen. Voor een precieze uiteenzetting van dit begrip, zie Dyen et al. (1992).
7. We hebben de vier door Hofstede (1980, 1996) gedefiniëerde *cultural dimensions* gebruikt, namelijk *masculinity*, *individualism*, *uncertainty avoidance* en *power distance*.
8. We berekenen de heteroskedasticiteit-robuste standaardfouten met behulp van de procedure van White (1980)

LAURA SPIERDIJK wordt per 1 juni 2006 benoemd tot Rosalind Franklin Fellow bij de Rijksuniversiteit Groningen. Tot haar aanstelling in Groningen op 1 juni a.s. is Spierdijk universitair docent bij de afdeling Toegepaste Wiskunde en het Financial Engineering Lab van de Universiteit Twente. E-mail: <L.Spierdijk@utwente.nl>.

MICHEL VELLEKOOP is als universitair hoofddocent werkzaam bij de afdeling Toegepaste Wiskunde en het Financial Engineering Lab van de Universiteit Twente. E-mail: <M.H.Vellekoop@utwente.nl>.

Methodoloog - Analist gezocht!



INTERVIEW • NSS

INTERVIEW • NSS BV is een gerenommeerd Nederlands marktonderzoeksbureau met een wereldwijd werkterrein. Centraal staat voor ons de relatie tussen organisaties en hun klanten, potentiële klanten, medewerkers en andere doelgroepen.

INTERVIEW • NSS BV kenmerkt zich als een professionele en uitdagende organisatie om voor te werken. De bedrijfscultuur is open, collegiaal en innovatief. Wij zijn geïnteresseerd in mensen die ons helpen bij het bereiken van onze ambitieuze doelen, zowel young professionals als mensen die een volgende stap in hun loopbaan willen maken.

Wat ga je doen:

Als **Methodoloog - Analist** maak je deel uit van een van de productieafdelingen (Branding & Media, of Contact & Channel Management). Je ondersteunt projectleiders bij het onderzoeksdesign (steekproef / vragenlijst), bij complexe wegingen en allerlei vormen van toegepaste analyses. Bij grote of complexe onderzoeken ga je soms mee naar de klant voor methodologisch advies of de presentatie van bevindingen. Samen met je collega methodologen blijf je op de hoogte van de ontwikkelingen in het vakgebied en draag je bij aan de ontwikkeling van nieuwe technieken en oplossingen.

Wat vragen wij van jou:

Je hebt als achtergrond toegepaste wiskunde (bv. econometrist of bedrijfswiskundige) en/of enkele jaren werkervaring in een soortgelijke functie. Je hebt voldoende methodologische bagage en kennis van toegepaste statistiek, steekproeftheorie en weging en enige programmeerervaring, bij voorkeur in een database-omgeving. Daarnaast heb je ervaring met statistische software (zoals SPSS) of ben je in staat dat snel op te pakken. Je werkt graag met cijfers en data en kan moeilijke zaken helder uit leggen aan klanten of collega's. Je bent handig met computers en in het automatiseren van standaardoplossingen met scripts en macro's. Je werkt systematisch en secuur. Documenteren van je werk is dan ook geen probleem. Bij voorkeur heb je kennis van databases en database theorieën. Je vindt het leuk van elkaars 'eigen' wijsheden te leren en houdt daarbij het gemeenschappelijk doel voor ogen.

Ben je geïnteresseerd?

Dan zijn wij benieuwd naar jou! Stuur je sollicitatiebrief en curriculum vitae voor 29 mei as. naar Anneke Avis, e-mail: sollicitaties@interview-nss.com. Meer informatie over ons bedrijf kun je vinden op de website: www.interview-nss.com



woensdag 14 juni

Dag voor Statistiek en Besliskunde 2006

Hoofdsprekers: Jan de Leeuw & Trevor Hastie

Ter gelegenheid van het vijftigjarig jubileum van het Econometrisch Instituut wordt de Dag voor Statistiek en Besliskunde dit jaar op **woensdag 14 juni** gehouden in het Erasmus Expo en Congrescentrum op het Woudestein complex van de **Erasmus Universiteit Rotterdam**. Dit complex bevindt zich bij de Brienoordbrug, en bij het metrostation Kralingse zoom.

Dag voor Statistiek en Besliskunde is de nieuwe naam voor wat vroeger de Statistische Dag heette. Het bestuur beoogt hiermee het brede werkveld van de VVS ook in de naam van de jaarlijkse ontmoeting zichtbaar te maken.

Het programma begint om 10.35 uur. Vanaf 10.00 uur wordt u ontvangen met koffie en thee. De beide hoofdsprekers zijn **Jan de Leeuw** en **Trevor Hastie**. Daarnaast hebben de secties ieder een eigen programma samengesteld. Om 16.30 uur biedt het Econometrisch Instituut een borrel aan ter afsluiting.

Zie voor meer informatie het programmaboekje dat bij deze STATOR is gevoegd.