

STATOR

periodiek van de VvS+OR jaargang 16, nummer 2, juli 2015

Ontwikkeling van jonge cannabisgebruikers vergeleken met leeftijdsgenoten; een Bayesiaans avontuur

Verdeelsleutel van CO₂-emissies bij gedeeld transport

Ronald Fisher: briljant statisticus en eugenicus

Simpsons paradox

Interbeoordelaar–betrouwbaarheid; het coderen van kenmerken van survey items voor de constructie van vragenlijstprofielen

Met Google op weg naar de automatische statisticus

STATOR is een uitgave van de Vereniging voor Statistiek en Operationele Research (VvS+OR). STATOR wil leden, bedrijven en overige geïnteresseerden op de hoogte houden van ontwikkelingen en nieuws over toepassingen van statistiek en operationele research. Verschijnt 3 keer per jaar.

Redactie

Joaquim Gromicho (hoofdredacteur), Annelieke Baller, Ana Isabel Barros, Kristiaan Glorie, Johan van Leeuwaarden, Guus Luijben (eindredacteur), Richard Starmans, Gerrit Stemerding (eindredacteur), Hilde Tobi en Vanessa Torres van Grinsven. Vaste medewerkers: Fred Steutel en Henk Tijms.

Kopij en reacties richten aan

Prof. dr. J.A.S. Gromicho (hoofdredacteur), Faculteit der Economische Wetenschappen en Bedrijfskunde, afdeling Econometrie, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, telefoon 020-5986010, mobiel 06-55886747, <j.a.dossantos.gromicho@vu.nl>.

Bestuur van de VvS+OR

Voorzitter: prof. dr. Jacqueline Meulman <president@vvs-or.nl>
 Secretaris: dr. Fetsje Bijma <bestuur@vvs-or.nl>
 Penningmeester: dr. Ad Ridder <penningmeester@vvs-or.nl>
 Overige bestuursleden: prof. dr. Eric Cator (SMS), prof. dr. Jeanine Houwing-Duistermaat (BMS), Maarten Kampert MSc., dr. Albert Wagelmans (NGB), dr. Michel van de Velden (ECS), dr. Jelte Wicherts (SWS), Nynke Krol (Young Statisticians)

Leden- en abonnementenadministratie van de VvS+OR

VvS+OR, Postbus 244, 6700 AE Wageningen, telefoon 0317-419572, e-mail <admin@vvs-or.nl>.
 Raadpleeg onze website over hoe u lid kunt worden van de VvS+OR of een abonnement kunt nemen op STATOR.

VvS+OR-website

www.vvs-or.nl

Sociale media

Wilt u uw vakgenoten ontmoeten en wilt u discussiëren over actuele thema's, volg dan de VvS+OR en de Young Statisticians via LinkedIn, Facebook, Twitter en Flickr.
 Sluit u aan bij de LinkedIn-groep van VvS+OR of Young Statisticians; bekijk foto's op <www.flickr.com/photos/vvs-or/sets>; like onze Facebook-pagina; volg de President van VvS+OR op <https://twitter.com/#!/dutchstat>.

Advertentieacquisitie

M. van Hootegem <hootegem@xs4all.nl>
 STATOR verschijnt in maart, juli en november.

Ontwerp en opmaak

Pharos, Nijmegen

Uitgever

© Vereniging voor Statistiek en Operationele Research
 ISSN 1567-3383

INHOUD

- 3** Over cannabis, Bayes, Fisher, Google en meer – redactioneel
- 4** Ontwikkeling van jonge cannabisgebruikers vergeleken met leeftijdsgenoten; een Bayesiaans avontuur
Mariëlle Zondervan-Zwijenburg
- 9** Steden – column
Fred Steutel
- 10** Verdeelsleutel van CO₂-emissies bij gedeeld transport
Steffie Naber, Wilco van den Heuvel, Diederik de Ree & Remy Spliet
- 14** Scholieren van Nederland, kom wiskunde studeren!
Remco van der Hofstad & Johan van Leeuwaarden
- 15** Ronald Fisher: briljant statisticus en eugeneticus
Joep Burger
- 16** Simpsons paradox
Edwin de Jonge
- 18** Interbeoordelaar–betrouwbaarheid; het coderen van kenmerken van survey items voor de constructie van vragenlijstprofielen
Frank Bais, Barry Schouten & Vera Toepoel
- 22** Toevalligheden en onmogelijkheden – column
Henk Tijms
- 24** Met Google op weg naar de automatische statisticus
Richard Starmans
- 28** Young Statisticians
- 28** IM Oud-voorzitter prof. dr. Kobus Oosterhoff
- 29** Persoonlijke herinneringen aan Kobus Oosterhoff
Piet Groeneboom
- 30** Gerrit Stemerding benoemd tot erelid van de VvS+OR
Jacqueline Meulman
- 31** Spinozapremie voor Aad van der Vaart



Over cannabis, Bayes, Fisher, Google en meer

Ook dit nummer van STATOR bevat een uiteenlopende verzameling voorbeelden van toepassingen van statistiek en operationele research. In zes artikelen kunt u zien hoe breed ons vak zich uitstrekt.

Mariëlle Zondervan-Zwijenburg laat zien hoe Bayes kan helpen bij het onderzoek naar de ontwikkeling de cognitieve functies van jonge cannabisgebruikers en Frank Bais, Barry Schouten en Vera Toepoel geven inzicht in de problemen met het coderen van survey items.

Steffie Naber, Wilco van den Heuvel, Diederik de Ree en Remy Spliet gebruiken de coöperatieve speltheorie om de emissie van kooldioxide bij distributieritten te verdelen over de klanten die bezocht worden en Edwin de Jonge beschrijft het verschijnsel dat het soms mogelijk is dat het weekloon voor een hele populatie stijgt, maar tegelijkertijd voor alle subgroepen daalt.

Joep Burger laat zien dat sir Ronald Fisher fantastische dingen heeft gedaan, maar toch verre van perfect was en Richard Starmans gaat na of de Googleaanse kijk op data-analyse kan leiden tot een automatische statisticus.

Onze vaste columnisten belichten weer verrassende aspecten: Fred Steutel vraagt zich af welke gemeente zich stad mag noemen en Henk Tijms geeft een tip hoe het schudden van een nieuw pak kaarten zó uitgevoerd kan worden dat de kans op vier perfecte handen fysisch onmogelijk wordt vanuit aards perspectief.

Zonder wiskunde geen statistiek of OR. STATOR-redacteur Johan van Leeuwaarden en zijn collega hoogleraar Remco van der Hofstad hebben een oproep geschreven waarmee zij bij scholieren de aandacht voor het vak wiskunde willen vergroten. De STATOR-redactie hoopt dat u deze tekst, die is aangeboden aan een van de landelijke dagbladen, goed kunt gebruiken.

Uiteraard ontbreken ook de Young Statisticians niet in dit nummer, zij hebben de Denen geïnspireerd om ook daar een YS te starten!

De VvS+OR is een vereniging, in dit nummer dan ook aandacht voor enkele van haar leden. Zoals voor Aad van der Vaart, oud-voorzitter van de VvS+OR, die de NWO-Spinozapremie voor zijn baanbrekend onderzoek in de statistiek heeft ontvangen. Hij krijgt de prijs voor zijn fundamenteel onderzoek naar modellen die bijvoorbeeld helpen bij het opsporen van genen die een rol spelen bij kanker.

We staan stil bij het overlijden van Kobus Oosterhoff in een bericht van het bestuur en een persoonlijke herinnering door Piet Groeneboom.

Tijdens de Statistische dag van 26 maart 2015 heeft de Algemene Ledenvergadering unaniem het voorstel van het bestuur aangenomen om Gerrit Stemerding te benoemen tot Erelid van de Vereniging. Gerrit is zeer belangrijk geweest voor de continuïteit van STATOR. Gerrit levert ieder jaar weer meer PK's aan de motor die STATOR gaande houdt. Zijn brede kennis en interesse, zijn bijzonder goed ontwikkelde taalgevoel en zijn perverse aanleg om fouten te signaleren maken hem tot de ideale eindredacteur.

Ten slotte een hartelijk welkom voor ons nieuwe redactielid Kristiaan Glorie, hij zal vanuit zijn achtergrond gaan bijdragen aan inspirerende, veelzijdige nummers.

Wij besluiten de inleiding op dit zeer gevarieerde nummer met de traditionele wens dat het u veel leesplezier moge bieden.

De STATOR-redactie



Cannabis sativa L., tekening van C. F. Millspaugh in *American medicinal plants*, vol. 2: t. 154 (1892)

Ontwikkeling van jonge cannabisgebruikers vergeleken met leeftijdsgenoten EEN BAYESIAANS AVONTUUR

Het onderzoeken en vergelijken van kleine groepen kan zo lastig lijken dat data uiteindelijk ongebruikt blijven. Het probleem speelt ook binnen het onderzoek naar het effect van cannabisgebruik op jonge leeftijd op het werkgeheugen. Dit artikel toont met een simulatiestudie aan wat de mogelijkheden voor deze datasets zijn binnen een latente-factor groeimodel. Aan de hand van de simulatieresultaten wordt er priorinformatie gezocht voor de vergelijking van werkgeheugenontwikkeling bij jongeren die cannabis gebruiken en jongeren die aangeven dat niet te doen. Het proces blijkt niet eenvoudig en de resultaten zijn verrassend. De potentie van de methode wordt geëvalueerd.

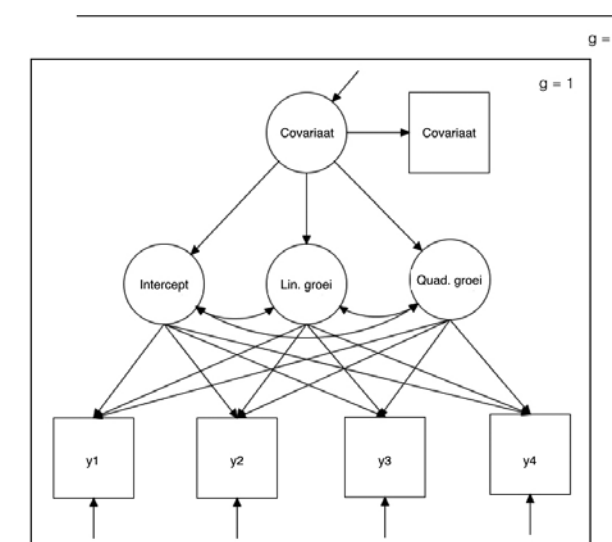
MARIËLLE ZONDERVAN-ZWIJNENBURG

Onderzoekers en beleidsmakers wereldwijd maken zich zorgen over de impact van cannabis op het ontwikkelende brein. Er zijn echter opvallend weinig longitudinale onderzoeken beschikbaar met betrekking tot het effect van cannabis op de cognitieve ontwikkeling. De onderzoeken die er zijn evalueren slechts twee meetmomenten: een voor het cannabisgebruik en een jaren later. Een verklaring voor dit gebrek aan onderzoek kan zijn dat de groep jongeren die frequent cannabis gebruikt relatief klein is en lastig op te nemen en behouden in een steekproef.

Op de Universiteit Utrecht bleek dat de onderzoekster Margot Peeters juist longitudinale data met betrekking tot cognitieve functies zoals werkgeheugen had verzameld van adolescenten op het speciaal onderwijs voor jongeren met gedragsproblemen, een risicogroep voor cannabisgebruik. Deze jongeren hadden ook informatie verstrekt over hun cannabisgebruik. Unieke data. Toch was er geen plan om de ontwikkeling van het werkgeheugen van de cannabisgebruikers te vergelijken met die van niet-gebruikende leeftijdsgenoten. Het aantal jongeren dat rapporteerde cannabis te gebruiken op het eerste meetmoment – de jongeren waren toen gemiddeld 13,3 jaar oud – was slechts 16, terwijl de groep die aangaf geen cannabis te gebruiken uit 250 jongeren bestond. Waar de onderzoekers normaal gesproken een latente-factor groeimodel zouden gebruiken met bijvoorbeeld alcoholgebruik als controle-variabele (figuur 1) om het verschil in ontwikkeling tussen de groepen te analyseren, hadden zij nu niet de zekerheid dat met deze beperkte data betrouwbare resultaten verkregen konden worden.

De grenzen en mogelijkheden binnen het latente-factor groeimodel

Om die reden werd een simulatiestudie gestart waarin de prestaties van twee schattingsmethoden binnen het latente-factor groeimodel werden onderzocht: maximum-likelihood-schatting met robuuste standaardfouten en Bayesiaanse schatting. De interesse binnen dit model lag op het verschil tussen de lineaire groeifactoren van de groepen, aangezien de kwadratische groeifactor gelijk gezet was over de groepen. In de simulatiestudie werden groepsgroottes voor de speciale groep van 5, 10, 25 en 50 onderzocht. De groeps grootte voor de



Figuur 1. Latente-factor groeimodel met covariaat voor twee groepen (g).

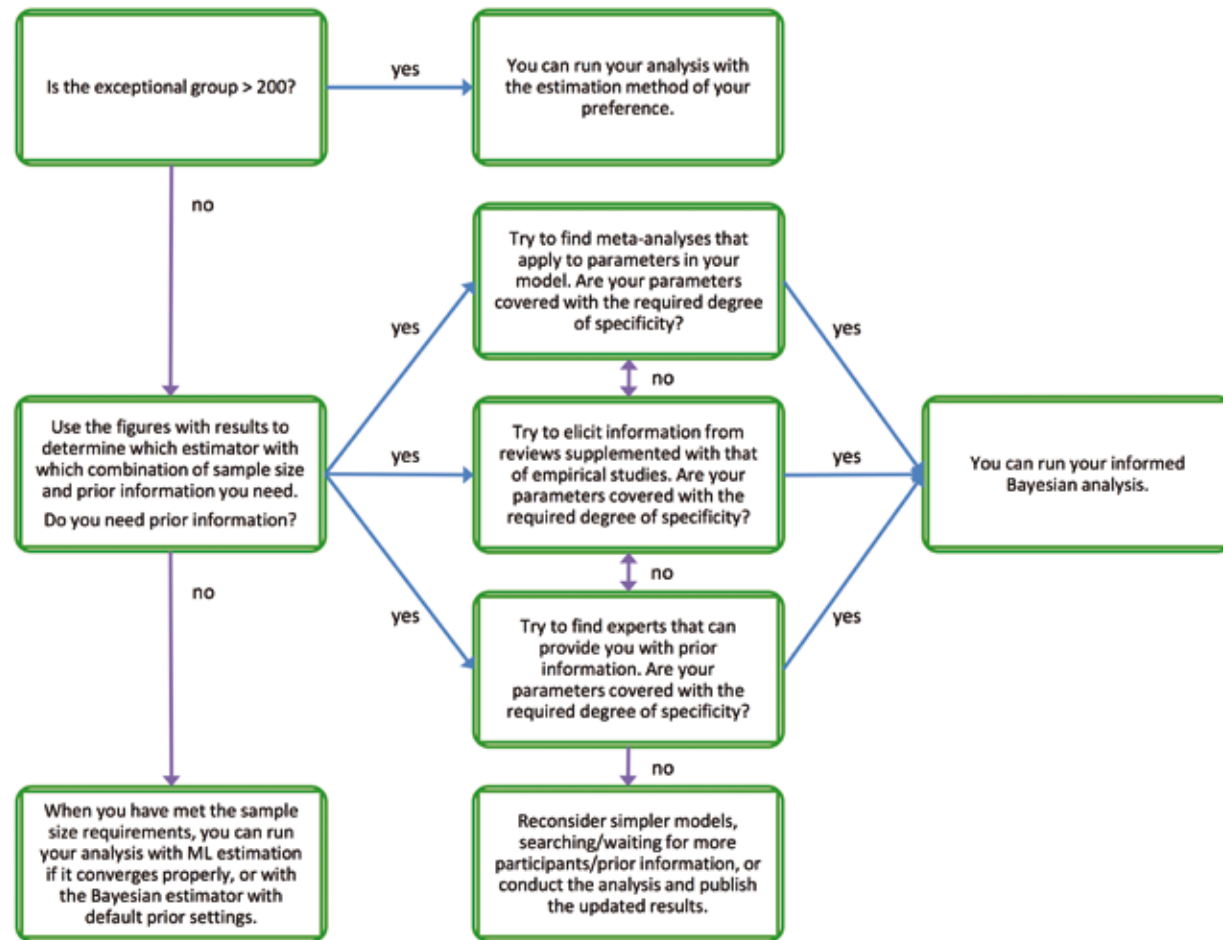
referentiegroep varieerde van 50 tot 10.000. Binnen de Bayesiaanse schatting werd ook de informativiteit van de zogenoemde priorinformatie gevarieerd: de beschikbare informatie van buiten de data die meegenomen wordt in de analyse.

De resultaten toonden aan dat er, ondanks kleine steekproeven en grote verschillen tussen de groeps-groottes, nauwelijks sprake was van afwijkende schattingen ten opzichte van de ingegeven populatiewaarde. Ook werd de populatiewaarde vaak genoeg gevonden in de 95% betrouwbaarheidsintervallen. Wel resulteerden maximum-likelihood-berekeningen regelmatig in onmogelijke parameterschattingen in de variantie-covariantie-matrix waardoor geen van de resultaten van de analyse nog zinvol geïnterpreteerd kon worden. Zeer problematisch was echter de statistische power om de significantie van een klein effect aan te tonen. Alleen zeer specifieke priorinformatie bood de mogelijkheid om met voldoende power te toetsen of het verschil tussen de latente groei-factoren van de groepen statistisch significant was. Op basis van deze resultaten kwam een beslisboom tot stand (figuur 2).

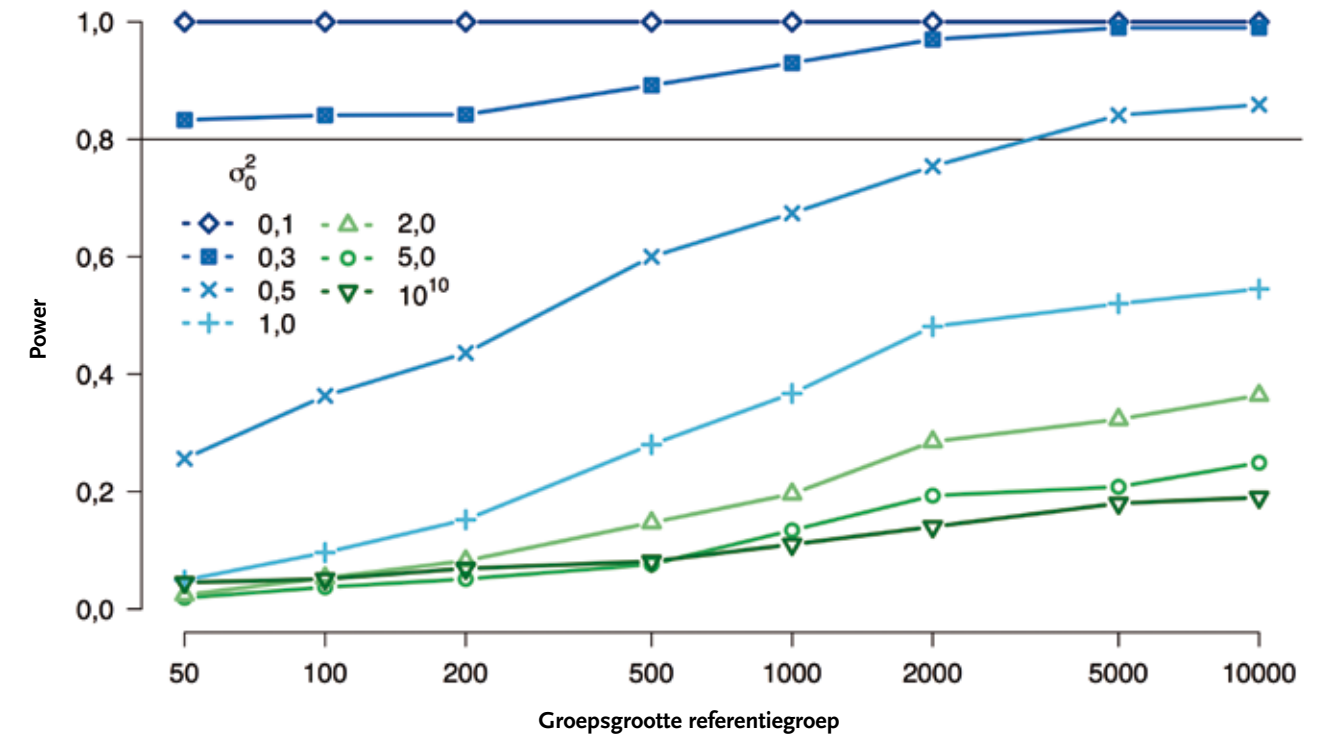
Het volgen van de beslisboom voor de beschikbare werkgeheugen-data leidde tot de evaluatie van figuur 3. Deze figuur toont hoe specifiek de priorinformatie voor de groeifactoren moet zijn in combinatie met de steekproefgrootte van de referentiegroep (dit zijn jongeren die geen cannabis gebruiken) om voldoende statistische power te behalen om een klein effect in de data aan te tonen wanneer de groep van frequente cannabisgebruikers slechts 10 (figuur 3a) of 25 (figuur 3b) deelnemers omvat. De conclusie is dat met de huidige dataset de variantie rond de schatting voor de priors 0,3 mag zijn.

Op zoek naar priorinformatie

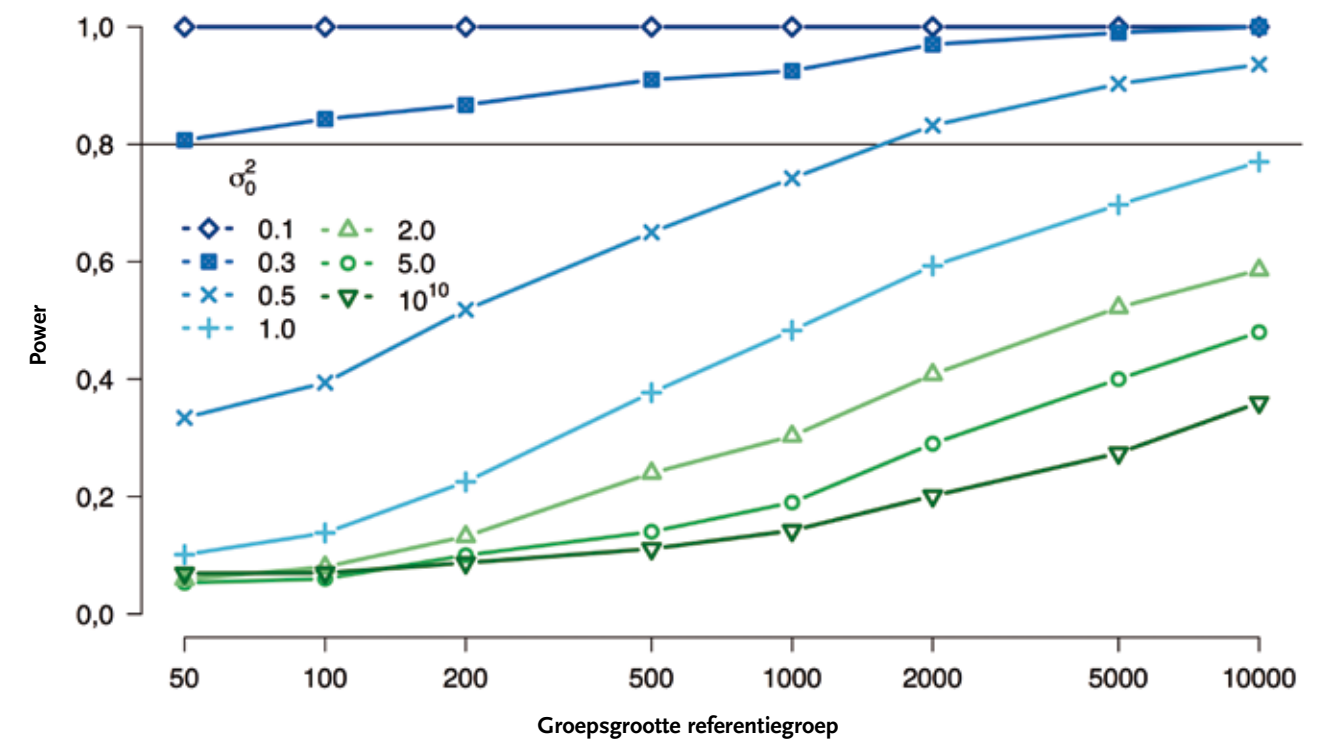
Gestuurd door de beslisboom (figuur 2), begon het zoeken naar meta-analyses voor priorinformatie. Helaas werden deze niet gevonden. Vervolgens werd, op instructie van diezelfde beslisboom, gezocht naar reviews. Negen reviews werden gevonden die in totaal 36 unieke relevante studies bespraken. Dit resultaat leverde een nieuwe uitdaging op: het vertalen van tekst uit reviews



Figuur 2. Beslisboom gebaseerd op de resultaten van de simulatiestudie



3a. Steekproefgrootte speciale groep is 10



3b. Steekproefgrootte speciale groep is 25

Figuur 3. Statistische power (y-as) in relatie tot steekproefgrootte in de referentiegroep (x-as) en priorvariantie op de latente groeifactoren (lijnen) als de steekproefgrootte van de speciale groep 10 is (a) en als die 25 is (b)

Steden

Volgens Jurian Meijering – zie zijn bijdrage in de vorige editie van *STATOR* 16(2015)¹ – woont driekwart van de Europese bevolking in steden. Hierbij wordt niet verteld wat onder een stad wordt verstaan, maar min of meer ‘uit het ongerijmde’ volgt uit de aanname dat ook in Nederland driekwart van de mensen in steden woont, welke Nederlandse gemeenten steden zijn. Heel wat meer dan de grootste vijf: Amsterdam, Rotterdam, Utrecht, Groningen en Eindhoven, waarbij ik wat Eindhoven betreft mijn twijfels heb; aan elkaar geplakte dorpen: Woensel, Gestel, Strijp, Tongelre en nog een paar.

Op internet vind ik een lijst van de grootste gemeenten in Nederland. Als ik voor Nederland tot een totaal van 75 procent van de bijna 17 miljoen Nederlanders wil komen – zeg 12,5 miljoen – moet ik op zijn minst alle gemeenten met 35.000 of meer inwoners meetellen. Dit betekent dat ook plaatsen als Nissewaard, Hoogeveen en Woerden tot de steden moeten worden gerekend. En ook Katwijk, waar bijna alle mensen Haasnoot heten en ook Almelo, waar regelmatig het enige stoplicht van rood op groen springt. Het dorp Haaksbergen, waar ik opgroeide, mag gelukkig een dorp blijven, al is het aantal inwoners sinds mijn jeugd verdubbeld.

In het boek *Triumph of the city* legt Edward Glaeser uit waarom steden beter zijn dan dorpen. In de steden gebeurt het allemaal – ‘daar is het te doen’, zoals dat tegenwoordig heet. Daar worden dingen uitgevonden, daar vindt innovatie plaats. Bovendien, de musea, schouwburgen en concertzalen zijn in de steden; en ook de beste scholen en universiteiten. Maar, Katwijk, Almelo? In Katwijk is inderdaad een museum: het Katwijks museum, en Almelo heeft het Stadsmuseum. Mijn dorp Haaksbergen heeft een spoorwegmuseum, en je kunt in de vakanties met de stoomtrein naar Boekelo en terug.

Afgelopen week waren mijn vrouw en ik in New York, en inderdaad we zijn naar verschillende musea geweest en naar de schouwburg. Nu weer thuis in de vijfde stad van Nederland. Toch weer even wennen.

FRED STEUTEL is emeritus hoogleraar kansrekening aan de TU Eindhoven.

E-mail: <fsteutel@xs4all.nl>

naar informatie waarop priorverdelingen gebaseerd kunnen worden. Kernzinnen uit de reviews werden onderzocht op aanduidingen van richting en sterkte van het mogelijke effect van cannabisgebruik en op indicaties van de zekerheid die de reviews met betrekking tot dit effect uitten. De informatie-synthese leidde tot de conclusie dat het effect negatief en klein geschat kon worden. De variantie van de prior moest dusdanig zijn dat enige onzekerheid over het bestaan van een effect tot uitdrukking werd gebracht.

Zoals ook aangegeven in de beslisboom, moet informatie van reviews aangevuld worden met dat van empirische studies. Reviews leiden voornamelijk tot schattingen van effectgroottes, maar deze zijn alleen bruikbaar voor de prior wanneer deze aangevuld worden met ongestandaardiseerde informatie over de afhankelijke variabele. Om die reden werd het zoeken naar priorinformatie vervolgd met het zoeken naar empirische studies waarin de werkgeheugentaak van de beschikbare data gebruikt werd. Ruim 600 studies waarin de naam van de taak opgenomen was werden geëvalueerd. Dertien studies bleven over waarin de taak daadwerkelijk uitgevoerd werd door gezonde jongeren of jongeren met een gedragsstoornis als ADHD. Op basis van deze studies kon een ontwikkelingslijn gecreëerd worden door gewogen gemiddelde scores per leeftijdsgroep uit te rekenen. Om de studies te wegen en om te bepalen voor welke groep zij relevant waren, werd de hulp ingeroepen van twee experts: gedragspsychologe Wenneke van de Schoot-Hubeek en Reinout Wiers, hoogleraar ontwikkelingspsychopathologie.* De experts gaven aan dat vier studies relevant waren voor jongeren in het speciaal onderwijs die geen cannabis gebruiken. Aangezien al deze steekproeven kinderen van gemiddeld 10 jaar betroffen, resulteerde dit in slechts één gewogen gemiddelde. Ook voor de frequente cannabisgebruikers kon één datapunt berekend worden. Alle andere steekproeven werden door de experts niet direct relevant bevonden: deze studies betroffen normaal ontwikkelende kinderen. Zelfs wanneer de data van deze laatste groep alsnog in overweging genomen werd, kon geen duidelijke ontwikkeling in het werkgeheugen uit de empirische informatie afgeleid worden. Toch was de informatie niet nutteloos: het was bijvoorbeeld duidelijk dat een waarde lager dan 60 voor de intercept groeifactor hoogst onwaarschijnlijk zou zijn. Zodoende konden informatieve priors voor de

groeifactoren gespecificeerd worden. Op basis van de beschikbare informatie kon echter geen onderscheid gemaakt worden tussen de priors voor de twee groepen. Ook kon de beoogde maximale variantie van 0,3 niet gespecificeerd worden: de onzekerheid over de waarden van de groeifactoren was daarvoor niet specifiek genoeg.

Analyse

De beslisboom (figuur 2) leidde nu naar het punt waarop gesuggereerd wordt om simpeler modellen te heroverwegen, net als het zoeken naar of wachten op meer participanten en priorinformatie. De andere optie was om de analyse alsnog uit te voeren en de resultaten te publiceren ter bevordering van toekomstig onderzoek dat met behulp van de nieuwe resultaten eventueel wel genoeg informatie kan verzamelen. Aangezien de toepassing van simpelere modellen geen interessante optie was, er geen zicht was op nieuwe informatie en de onderzoekers graag wilden bijdragen aan de opbouw van kennis, werd gekozen voor de laatste optie.

De resultaten van de analyse waren verrassend: er werd een statistisch significant verschil tussen de beide groepen gevonden. Een sensitiviteitsanalyse met minimaal-informatieve prior-verdelingen toonde echter aan dat de prior en de data dusdanig verschilden dat niet alleen de grootte van de betrouwbaarheidsintervallen, maar ook de parameterschattingen veranderden. Vooral de resultaten voor de speciale groep waren beïnvloed door de priors. Dit komt doordat de data voor de speciale groep minder informatie bevatten. Desalniettemin kan de ongelijke beïnvloeding als alarmerend ervaren worden omdat het verschil tussen de groepen er ook door beïnvloed wordt.

En nu? Wat betekent dit?

Al met al kan een onderzoeker die zowel de informatieve priors als de data valide en betrouwbaar acht concluderen dat er een statistisch significant, middelgroot (Cohens $d = 0,54$) negatief effect is van frequent cannabisgebruik op jonge leeftijd op de ontwikkeling van het werkgeheugen. Onze conclusie was echter dat er alsnog meer informatie nodig is om tot betekenisvolle

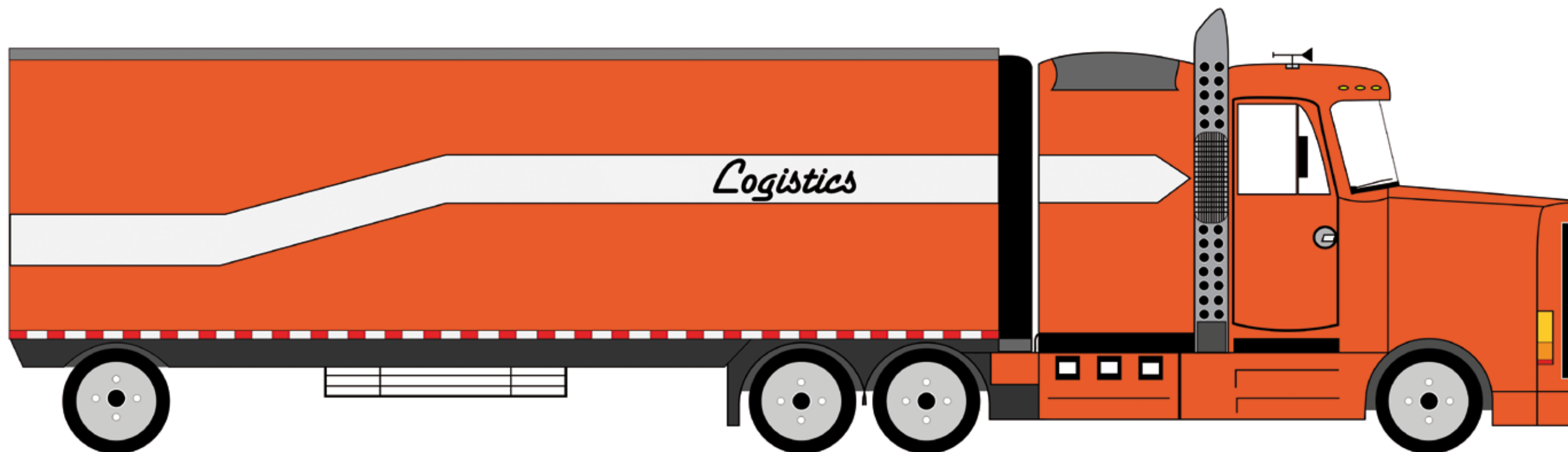
uitspraken over het effect te komen, omdat de sensitiviteitsanalyse belangrijke verschillen aantoonde. Ondanks deze uitkomst, zouden we het onderzoek niet als zinloos willen beschouwen. Door het zoeken naar priorinformatie leerden we meer over het effect, de afhankelijke variabele en de populatie dan dat we hadden geleerd van een gebruikelijke literatuurstudie voor een introductie sectie. Ook maakt de prior het mogelijk om kritischer op de data te reflecteren. In algemene zin maakt dit onderzoek ook duidelijk dat het includeren van priorinformatie uit de literatuur niet per se een makkelijke uitweg is voor het probleem van kleine steekproeven. Een onderdeel van het probleem is ook een vicieuze cirkel: data zijn lastig te verzamelen voor de speciale groep, maar om diezelfde reden is er ook weinig informatie beschikbaar in de literatuur. Deze cirkel zou doorbroken kunnen worden door meer data te delen en studies te publiceren ongeacht significantie. Daarnaast kan voor eenvoudig te beoordelen parameters de hulp ingeschakeld worden van experts: hun expertise en huidige kennis kan omgezet worden in priorverdelingen. In ieder geval biedt Bayesiaanse analyse met informatieve priors een unieke kans om grote vragen over kleine populaties te evalueren.

* Wenneke van de Schoot-Hubeek is gedragspsychologe in het speciaal onderwijs voor Horizon Jeugdzorg en Onderwijs en lid van het NIP-Crisis Interventie Netwerk Schoolpsychologen. Reinout Wiers is hoogleraar ontwikkelingspsychopathologie aan de Universiteit van Amsterdam met als onderzoeksgebied cognitieve processen betrokken bij verslavingsgedrag.

LITERATUUR

Zondervan-Zwijenburg, M., Peeters, M., Depaoli, S., & Van de Schoot, R. (2014). *Pushing the limits: Sample size requirements with unbalanced subgroups in latent growth models*. Manuscript submitted for publication.

MARIËLLE ZONDERVAN-ZWIJENBURG is promovendus aan de Universiteit Utrecht bij de afdeling Methoden en Statistiek voor Sociale Wetenschappen. Haar onderzoek richt zich op het gebruik van informatieve priors ten behoeve van het Consortium Individual Development (CID). Het CID is gefinancierd door het Zwaartekracht-programma van het ministerie van Onderwijs, Cultuur en Wetenschap en het NWO (NWO Zwaartekracht 024.001.003). De beschreven masterthesis, begeleid door Rens van de Schoot en Margot Peeters, werd bekroond met de Peter G. Swanbornprijs voor hoogwaardig empirisch-theoretisch onderzoek.
E-mail: <M.A.J.Zwijenburg@uu.nl>



VERDEELSLEUTEL VAN CO₂-EMISSIONS BIJ GEDEELD TRANSPORT

STEFFIE NABER, WILCO VAN DEN HEUVEL, DIEDERIK DE REE & REMY SPLIET

Het merendeel van de klimaatwetenschappers lijkt er niet meer aan te twijfelen: de uitstoot van koolstofdioxide (CO₂) draagt bij aan de opwarming van de aarde. In de afgelopen decennia zijn steeds meer mensen zich hier bewust van geworden, en steeds meer van hen proberen daarom hun eigen CO₂-uitstoot terug te dringen. Bedrijven spelen hier slim op in door 'milieuvriendelijke' producten aan te bieden. Dit wordt bijvoorbeeld gedaan door de *carbon footprint*, de totale hoeveelheid CO₂ die gepaard gaat met de vervaardiging van een product, te vermelden. Naast het productieproces beïnvloedt ook het transport van de producten in belangrijke mate de carbon footprint. De totale CO₂-uitstoot van een rit kan vrij nauwkeurig berekend worden. Maar hoe bepaal je, voor een distributierit waarin meerdere klanten bevoorrad worden, welk deel van die CO₂-uitstoot moet worden toebedeeld aan welke klant?

Omdat hier momenteel geen gouden standaard voor is, verdelen veel bedrijven de emissies naar eigen inzicht. Meestal wordt dan de zogeheten ster-methode gebruikt,

waarin de totale emissie verdeeld wordt naar rato van wat de emissie zou zijn geweest als de klant afzonderlijk bevoorrad zou worden, de zogeheten stand-alone emissie. Aan de hand van figuur 1 illustreren we de toepassing van de ster-methode. Hierin zijn de afstanden in kilometers van de klanten (A t/m E) tot het depot en die van de route die de logistieke dienstverlener aflegt weergegeven. Stel dat er één voertuig gebruikt wordt met een uitstoot van 0,8 kg CO₂/km. In dat geval bedraagt de totale CO₂-uitstoot van de route 12 kg (de lengte is 15 km). Gezien klant A op 2 km van het depot ligt en de som over de afstanden van alle klanten tot aan het depot 15 km is, wordt 2/15 deel, dus 1,6 kg CO₂, toebedeeld aan klant A.

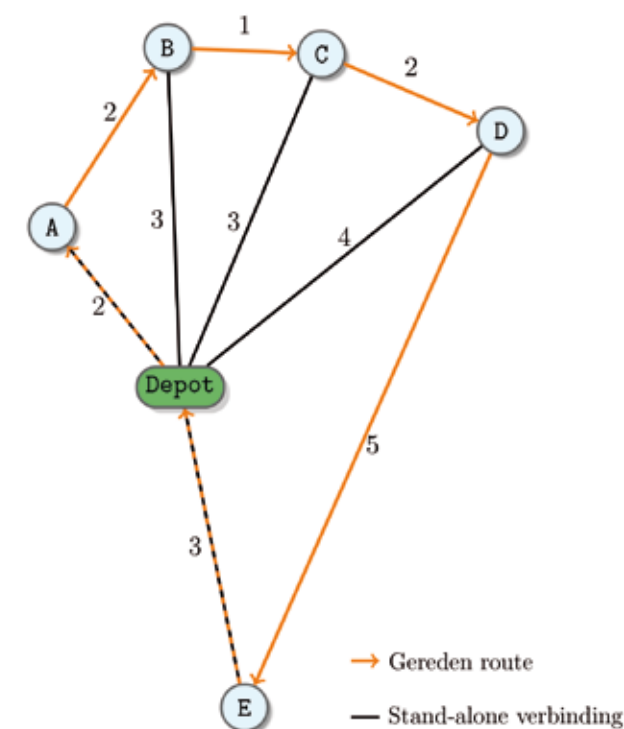
Gezien de eenvoud en inzichtelijkheid van de ster-methode, wordt een dergelijke methode ook voorgesteld in verschillende initiatieven om tot een gouden standaard te komen (European committee for Standardization, 2012; COFRET deliverable 2.1, 2013; DEFRA, 2013; DSLV, 2013). Een van de nadelen van deze ster-methode is echter dat de toebedeelde emissie niet direct afhan-

kelijk is van de locatie van een klant ten opzichte van de andere klanten op de route. Ligt een klant bijvoorbeeld vrij geïsoleerd ten opzichte van de overige klanten, dan lijkt het logischer om een relatief groter gedeelte van de CO₂-emissies toe te delen aan deze klant. Klant E ligt bijvoorbeeld veel geïsoleerder dan klant B en beide krijgen nu toch dezelfde hoeveelheid emissie toebedeeld. In dit artikel wordt daarom een aantal alternatieve allocatiemethoden voorgesteld. In een casus worden ze op verschillende criteria zowel onderling als met de ster-methode vergeleken.

Coöperatieve speltheorie

Het toedelingsvraagstuk benaderen we vanuit een speltheoretische invalshoek, om precies te zijn vanuit de coöperatieve speltheorie. Het spel bestaat uit het verdelen van de totale emissie van een distributierit over de klanten (de spelers) van wie de producten vervoerd worden. In standaard coöperatieve speltheorie is de basis van de samenwerking tussen de spelers vaak een besparing van kosten (of toename in winst) en de centrale vraag is hoe deze besparing verdeeld moet worden. In ons spel voegt de logistieke dienstverlener de klanten samen om kosten of tijd te besparen, en niet zozeer emissies (al zal er in het algemeen wel een reductie in emissies plaatsvinden). Alhoewel klanten een zo laag mogelijke emissie

nastreven, zullen zij, in tegenstelling tot standaard speltheorie, het spel niet verlaten wanneer hun toebedeelde emissie hoger is dan hun stand-alone emissie. Enerzijds hebben zij vaak geen inzicht in de CO₂-uitstoot van de gereden route en van de mogelijke alternatieven,



Figuur 1. De toepassing van de ster-methode

anderzijds hechten zij vaak ook meer waarde aan tijd en geld dan aan milieuvriendelijkheid. Hetzelfde geldt voor coalities van klanten, die in een gezamenlijke route een kleinere hoeveelheid emissies zouden verdelen. In het voorbeeld zouden bijvoorbeeld klanten A, B, C en D een coalitie kunnen vormen. Wanneer deze coalitie apart bediend zou worden, dan is de totale uitgestoten emissie 8,8 kg (de lengte van de route is 11 km), terwijl de huidige toebedeelde emissie aan deze coalitie meer is, namelijk 9,6 kg. Deze allocatie zal daarom niet als eerlijk en wenselijk worden ervaren en een dergelijke allocatie noemen we instabiel. De verzameling van alle stabiele allocaties wordt de zogeheten *core* van het spel genoemd.

Allocatiemethoden

De alternatieve allocatiemethoden die we vergelijken zijn de Shapley-waarde (Shapley, 1953), de Nucleolus (Schmeidler, 1969), en aangepaste vormen van de Lorenz-allocatie (Arin, 2003) en Equal Profit Method, EPM, (Frisk et al, 2010). We zullen de methodes nu kort toelichten.

De Shapley-waarde van een klant is gedefinieerd als de gemiddelde marginale emissie over alle coalities van andere klanten, waarbij de marginale emissie gelijk is aan de emissie die erbij komt als je een klant toevoegt aan een coalitie van andere klanten. Hoewel dit eerlijk lijkt, garandeert dit niet dat de uiteindelijke toedeling in de core van het spel ligt en dus stabiel is. De Nucleolus is een andere toewijzing, die kan worden geïnterpreteerd als het 'middelpunt' van de core. Theoretisch gezien kan de core echter leeg zijn. In dit geval is de Nucleolus een allocatie die het minst instabiel is, oftewel, het is een allocatie die de restricties opgelegd door de core zo min mogelijk overschrijdt. Een nadeel van de Nucleolus is dat het weinig inzicht biedt in hoe een eventuele veran-

dering van een route door zal werken in de allocatie. De laatste twee methoden die wij toegepast hebben, leveren een andere allocatie in de core. De Lorenz-allocatie is een allocatie in de core die elke klant in absolute zin een zo gelijk mogelijke emissie toebedeelt. Dit wordt gedaan door een allocatie in de core te zoeken die de absolute verschillen in toewijzing tussen de klanten minimaliseert. De EPM is gebaseerd op hetzelfde mechanisme, maar minimaliseert niet de absolute verschillen tussen klanten, maar de verschillen tussen de ratio's van hun toewijzing en hun stand-alone emissie. Omdat er bij een lege core geen allocatie bestaat voor de Lorenz-allocatie en de EPM, wordt in dat geval de Nucleolus gebruikt.

Toetsingscriteria

De voorgestelde methoden vergelijken we op vier criteria die belangrijk zijn voor de acceptatie van een allocatiemethode in de praktijk: stabiliteit, consistentie, robuustheid en rekentijd. De stabiliteit meten we als percentage allocaties dat zich in de core bevindt. Met consistentie bedoelen wij dat de allocatie op een voorspelbare manier verandert, wanneer de karakteristieken van een probleeminstantie veranderen. Dit betekent bijvoorbeeld dat als de vraag van een klant toeneemt terwijl de vraag van de anderen gelijk blijft, dat dan de emissie die wordt toebedeeld aan die klant ook toeneemt. Om een score voor consistentie te bepalen voeren we een lineaire regressie uit. Hierin meten we in hoeverre de toebedeelde emissie gerelateerd is aan de grootte van de bestelling, de afstand van de klant tot het depot, en de gemiddelde afstand van de klant tot de andere klanten. Als uiteindelijke score van consistentie gebruiken we de R^2 van de regressie. De robuustheid van een methode is de mate van constantheid in de toebedeelde emissie aan een klant die herhaaldelijk dezelfde bestelling plaatst, terwijl de ei-

genschappen van andere klanten veranderen. Als laatste is de rekentijd vergeleken, omdat dit een belangrijk aspect is voor het gebruik van de methoden in de praktijk.

Casus

Om de vijf voorgestelde methoden te testen op de criteria bekijken we een modeketen die elke week haar winkels bevoorraadt vanuit het magazijn. Hiertoe worden altijd dezelfde 13 routes gebruikt, waarin 9 tot 11 klanten worden bevoorradt. Voor een bepaald jaar is van elke week bekend hoe hoog de vraag was van elke klant. Er zijn dus $52 \times 13 = 676$ instanties waarvoor de emissies verdeeld dienen te worden. In deze casus wordt dit gedaan met behulp van de vijf eerder beschreven methoden.

Na het toepassen van de methoden op deze instanties blijkt dat de core voor geen van de instanties leeg is. Per definitie vinden de Nucleolus, de Lorenz-allocatie en de EPM daarom in alle gevallen een stabiele allocatie. De Shapley-waarde en de ster-methode vinden in respectievelijk 99% en 35% van de gevallen een core-allocatie (tabel 1). De Shapley-waarde en Nucleolus hebben de laagste R^2 en scores daarmee het slechtst op consistentie. De ster-methode scoort met een R^2 van 0,79 het hoogst. Voor deze instanties is de gemiddelde rekentijd slechts enkele seconden voor de Nucleolus, Lorenz-allocatie en EPM, en zelfs minder dan één seconde voor de ster-methode en de Shapley-waarde.

Om de robuustheid van de methoden goed te kunnen meten hebben we zelf probleeminstanties gegeneerd. Voor deze instanties bekijken we in hoeverre de emissie die wordt toebedeeld aan een bepaalde klant varieert naarmate de locatie en vraag van andere klanten veranderen. Op basis van de variatiecoëfficiënt concluderen we dat de ster-methode het minst robuust is, en de Nucleolus het meest.

Conclusie

Ons onderzoek laat zien dat er niet één beste allocatiemethode lijkt te zijn, en dat het cruciaal is om prioriteiten te stellen onder de verschillende criteria. Iedere allocatiemethode heeft zo zijn eigen voor- en nadelen. Als de rekentijd en eenvoud van een methode het belangrijkste aspect is, dan is de ster-methode een geschikte keuze. Echter, als stabiliteit van belang is valt deze methode direct af. De Lorenz-allocatie en EPM zijn stabiele

allocaties die het meest consistent zijn. De Nucleolus is aan te raden als de meeste waarde wordt gehecht aan robuustheid van een methode.

Dit artikel is gebaseerd op het onderzoek beschreven in: Naber, S.K., De Ree, D.A., Spliet, R., & Van den Heuvel, W. (2015). Allocating CO2 emission to customers on a distribution route. *Omega*, 54, 191-199.

LITERATUUR

- Arin, J. (2003). *Egalitarian distributions in coalitional models: The Lorenz criterion*. Technical report. Leioa: University of the Basque Country.
- COFRET deliverable 2.1. Existing methods and tools for calculation of carbon footprint of transport and logistics, May 2013. Gevonden op www.cofret-project.eu.
- DEFRA. Guidance on measuring and reporting greenhouse gas (ghg) emissions from freight transport operations, May 2013. Gevonden op <http://archive.defra.gov.uk>.
- European committee for Standardization. En16258:2012, 2012. Gevonden op www.duurzamelogistiek.nl.
- DSL. Berechnung von treibhausgasemissionen in spedition und logistik, 2013. Gevonden op www.spediteure.de.
- Frisk, M., Göthe-Lundgren, M., Jörnsten, K., & Rönnqvist, M. (2010). Cost allocation in collaborative forest transportation. *European Journal of Operational Research*, 205(2), 448-458.
- Schmeidler, D. (1969). The nucleolus of a characteristic function game. *SIAM Journal of Applied Mathematics*, 17(6), 1163-1170.
- Shapley, L. (1953). A value for n-person games. *Annals of Mathematics Studies*, 28, 307-317.

STEFFIE NABER is promovenda bij de afdeling Maatschappelijke Gezondheidszorg van het Erasmus Medisch Centrum Rotterdam. Ten tijde van het onderzoek was Steffie werkzaam als stagiaire bij TNO op de afdeling Sustainable Transport and Logistics. E-mail: <s.naber@erasmusmc.nl>

WILCO VAN DEN HEUVEL is universitair hoofddocent aan het Econometrisch Instituut, Erasmus Universiteit Rotterdam, en doet onderzoek naar optimalisatie in productieplanning. E-mail: <wvandenheuvel@ese.eur.nl>

DIEDERIK DE REE is business consultant bij Quintiq, waar hij grote ondernemingen en multinationals ondersteunt in het oplossen van complexe bedrijfspuzzels en het verwezenlijken van hun strategische visie. Tijdens het onderzoek was Diederik werkzaam als consultant bij TNO op de afdeling Sustainable Transport and Logistics en betrokken bij het onderzoeksproject COFRET. E-mail: <diederik.de.ree@gmail.com>

REMY SPLIET is universitair docent aan het Econometrisch Instituut van de Erasmus Universiteit Rotterdam en doet onderzoek naar voertuigrouting. E-mail: <rspliet@ese.eur.nl>

METHODE	STABILITEIT (% ALLOCATIES IN CORE)	CONSISTENTIE (R^2)	GEMIDDELDE REKENTIJD (SECONDEN)
Ster-methode	35	0,79	0,2
Shapley-waarde	99	0,59	0,5
Nucleolus	100	0,60	3,3
Lorenz	100	0,73	3,3
EPM	100	0,74	3,3

Tabel 1. Scores van de allocatiemethoden op verschillende toetsingscriteria voor de casus



SCHOLIEREN VAN NEDERLAND: KOM WISKUNDE STUDEREN!

REMCO VAN DER HOFSTAD & JOHAN VAN LEEUWAARDEN

Sta jij op dit moment voor je studiekeuze? Houd je van een pittige puzzel, weet je wel raad met wiskunde, lach je om de rekentoets en zoek je juist meer inhoud? Kies dan voor wiskunde! Je weet vast al dat wiskunde prachtig is, met voorbeelden als de stelling van Pythagoras, priemgetallen en de mysterieuze tekeningen van Escher. Maar er is nog veel meer. Wiskunde is overal aanwezig en van groot belang voor onze maatschappij. Tijd dus voor een eigentijds beeld van wiskunde aan de hand van drie (niet wiskundige) stellingen. Daarna is de keuze aan jou.

Beste baan ter wereld!

CareerCast maakt ieder jaar een ranking van de beste banen, gebaseerd op salaris, baankansen en het plezier van de werknemers. In 2014 voerde het beroep van wiskundige deze lijst aan. Dit jaar een derde plaats voor wiskundige, naast een eerste plaats voor actuaire, een vierde plaats voor statisticus en de zesde plaats voor data scientist: vier wiskundige beroepen in de top 6! De banen zijn aansprekend, de salarissen hoog en de vraag naar wiskundigen stijgt. Steeds meer banen vragen om wiskundige vaardigheden. De beroepsmogelijkheden voor wiskundigen zijn eindeloos. Zo kunnen ze bijvoorbeeld aan de slag bij grote bedrijven, overheidsinstellingen, zakelijke dienstverlening en het onderwijs. Het analytische denken dat je als wiskundige leert, gecombineerd met de vaardigheden om te rekenen en data te begrijpen, geeft dus een fantastische basis voor allerlei carrières. Al overtuigd? Blijf in ieder geval doorlezen!

Zonder wiskunde geen welvaart

Onze kenniseconomie drijft op wiskunde. Jaarlijks verdienen we in Nederland 160 miljard met wiskunde. Dat blijkt uit een uitgebreide studie uitgevoerd door Deloitte. De wiskunde in Nederland zorgt voor 1,4 miljoen banen en daarmee voor 25% van de Nederlandse werkgelegen-

heid en circa 30% van het bruto nationaal product (die 160 miljard dus). De vraag naar wiskundigen is inmiddels véél groter dan het aanbod. Onze wereld bestaat uit duizelingwekkend veel getallen. Geld, digitale communicatie, informatie, alles wordt uitgedrukt in getallen. Een wiskundige heeft inzicht in getallen en kan er patronen in ontdekken. Bedrijven staan te springen om mensen die met data overweg kunnen. En nee, de verplichte rekentoets is daarvoor niet genoeg. Kortom: krijg je al zin om wiskunde te gaan studeren?

Wiskunde is uitdagend

We horen je al denken: 'Klinkt goed, maar wiskunde is toch heel moeilijk?' Dat klopt. Maar is dat niet juist een reden om het te gaan studeren? Bij wiskunde leer je logisch en rationeel denken. Dat is hard nodig in onze complexe wereld. Je kunt wiskunde dus zien als het oefenen van je probleemoplossend vermogen. Als je wiskundige sommen kunt kraken, dan zul je ook wel een goede kijk hebben op lastige economische, bestuurlijke of strategische vraagstukken. Kortom, je scherpt je blik, zowel binnen als buiten de wiskunde. Wiskunde boekt successen op terreinen als economie, gezondheidszorg, logistiek, landbouw, sport, energie, verkeer, hightech en nog veel meer. Heb je aanleg voor wiskunde? Kijk dan eens goed in de spiegel en vraag je af of je deze uitdaging aan wilt gaan. We zien je graag in onze collegebanken verschijnen!

Bovenstaande oproep werd eerder aangeboden aan de landelijke dagbladen. REMCO VAN DER HOFSTAD en JOHAN VAN LEEUWAARDEN zijn beiden hoogleraar wiskunde aan de Technische Universiteit Eindhoven. Remco is woordvoerder van Platform Wiskunde Nederland (PWN) en Johan is lid van De Jonge Akademie (KNAW) en STATOR-redacteur. Zij zetten zich in om de wiskunde in Nederland te promoten. E-mail: <r.w.v.d.hofstad@tue.nl>, <j.s.h.v.leeuwaarden@tue.nl>

RONALD FISHER

briljant statisticus en eugenicus

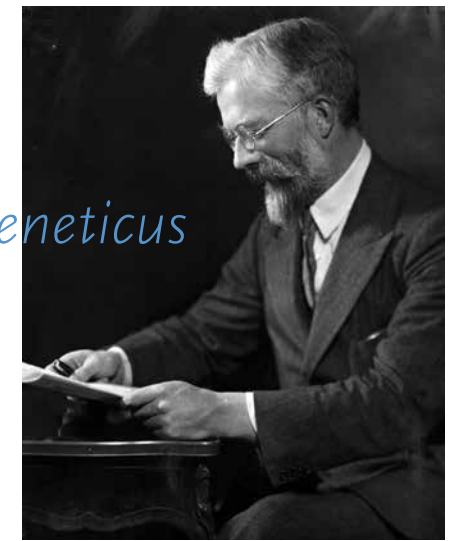
JOEP BURGER

Helden heb ik niet. Ze doen me te veel denken aan de histerie onder de eenzijdig geïnformeerde bevolking toen de Noord-Koreaanse dictator Kim Jong-il overleed. Jezus van Nazareth de zelloot zoals Reza Aslan die heeft beschreven, vind ik geloofwaardiger dan Jezus Christus de messias waarover de pastoor vertelde. Volgens filmmaker Quentin Tarantino verkleedt Superman zich als mens – bril en maatpak van Clark Kent zijn het kostuum – om onze tekortkomingen te laten zien. In evolutionaire termen zijn helden *Darwinian demons*: superorganismen met eigenschappen die natuurlijke selectie bevoordeelt. Ze bestaan niet, omdat evolutie door natuurlijke selectie slechts voortbouwt op wat er al is, en omdat elk voordeel zijn nadeel heeft. In Dawkins' *The Greatest Show on Earth* staan daarvan fascinerende voorbeelden.

Dat neemt niet weg dat er mensen zijn die fantastische dingen hebben gedaan. Neem Ronald Fisher (1890-1962). Staat niet bekend als methodoloog maar is voor methodologen zeker relevant als auteur van onder andere *The Design of Experiments*. Hierin beschrijft hij een dame die beweert dat ze kan proeven in welke volgorde thee en melk zijn ingeschonken. Fisher introduceert de twee fundamentele kenmerken van een experiment waarmee een dergelijke bewering getoetst kan worden: replicatie en randomisatie. (De resultaten van het experiment geeft hij niet; David Salsburg wel.) Welbekende concepten als *confounding factor* en p-waarde zijn van Fishers hand. Hij stelde een kans van een op twintig (0,05) voor als drempel om een effect significant te noemen, waarschijnlijk niet vermoedend dat deze suggestie zo dogmatisch gebruikt gaat worden dat publicatie van een onderzoek vaak daarvan afhangt. Verder: uitvinder van ANOVA, Maximum Likelihood en permutatietest. En naar wie zouden de F-test en F-verdeling vernoemd zijn?

Fisher verzette zich hevig tegen zogenoemde observationele studies die een correlatie vonden tussen roken en longkanker. Uiteenlopende studies met elk hun eigen tekortkomingen lieten echter een zeer consistent beeld zien waardoor weinigen nog twijfelden aan een causaal verband. Maar Fisher zou waarschijnlijk alleen overtuigd kunnen worden door een experiment waarvan een ethische commissie stekelhaar krijgt.

Old school? Hypothese-gedreven onderzoek is immers uit, data-gedreven onderzoek is (weer) in. Maar wie zich zorgen maakt om de selectiviteit van big data



bronnen, kan Fishers Linear Discriminant Analysis gebruiken. Toegepast op bijvoorbeeld Twitter reduceren profielfoto's hiermee tot *Fisherfaces*, als input voor *machine-learning-algoritmes* waarmee achtergrondkenmerken kunnen worden afgeleid. Wie durft te beweren dat zijn werk over tachtig jaar nog relevant is?

Behalve statisticus was Fisher ook nog geneticus. Ook daarin blonk hij uit als een van de grondleggers van de populatiegenetica, de studie van de verdeling en dynamiek van genfrequenties in populaties. Daarmee legde hij de basis van de synthese van Darwins evolutietheorie en Mendels erfelijkheidsleer. Zonder Mendel te beschuldigen van fraude ontdekte Fisher dat Mendels data te mooi waren om waar te zijn. Gewoon een voorbeeld van onbewust zien wat je wil zien.

Was Sir R.A. Fisher perfect? Natuurlijk niet. Hij kon humeurig zijn, zijn teksten staan bekend als moeilijk te volgen, en ook zijn colleges trokken geen volle zalen. Meest opmerkelijke feit is dat hij actief lid was van the British Eugenics Society, een club die, weliswaar voor de Holocaust, (vrijwillige) sterilisatie wilde legaliseren om de genetische kwaliteit van de populatie verbeteren. Met zijn negen kinderen zal hij zijn eigen genen en die van zijn vrouw hoog hebben ingeschat.

Ronald Fisher: geen held, wel een reus op wiens schouders wij allen mogen staan.

LITERATUUR

- Aslan, R. (2013). *Zealot. The life and times of Jesus of Nazareth*. New York: Random House.
- Dawkins, R. (2009). *The greatest show on earth. The evidence for evolution*. New York: Free Press.
- Fisher, R. A. (1935). *The design of experiments*. Edinburgh: Oliver and Boyd.
- Salsburg, D. (2002). *The lady tasting tea. How statistics revolutionized science in the twentieth century*. New York: Freeman/Owl Book.

JOEP BURGER is statistisch onderzoeker/data-analist en werkzaam bij het Centraal Bureau voor de Statistiek als methodoloog. E-mail: <jbur@cbs.nl>



SIMPSONS PARADOX

EDWIN DE JONGE

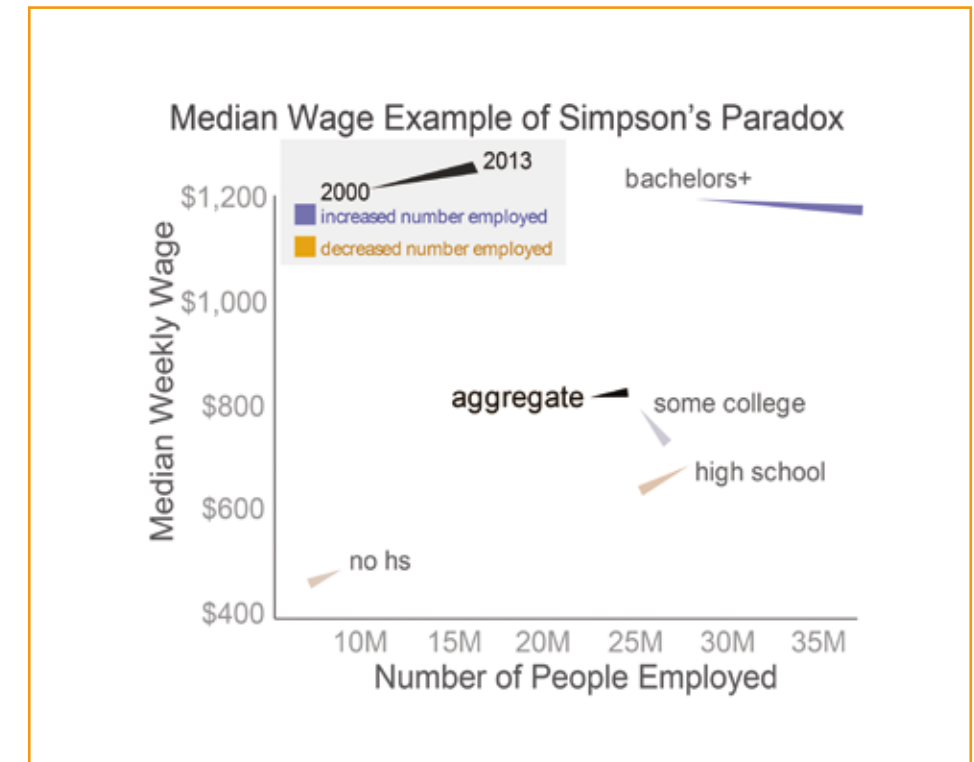
Tussen 2000 en 2013 steeg het mediane weekloon gecorrigeerd voor inflatie in de VS met 0,9% (Bureau of Labor Statistics, 2014). Het is interessant om te weten hoe deze stijging naar opleidingsniveau verdeeld is. Ongetwijfeld zullen hier verschillen in optreden. Tabel 1 laat zien hoeveel het weekloon tussen 2000 en 2013 gestegen of gedaald is, gesplitst naar (Amerikaans) opleidingsniveau.

Je verwacht dat sommige groepen het beter dan 0,9% doen en sommige slechter, maar voor alle groepen daalt het mediane inkomen terwijl het totaal stijgt. Hoe kan dat? Is er sprake van een datafout? Nee, dit is een reëel, maar tegenintuïtief effect en heet Simpsons paradox (Simpson, 1951). Als je onbekend bent met deze paradox, stop even met lezen en bedenk hoe dit kan.

De verklaring dat het totaal stijgt en het loon van alle opsplitsingen daalt zit in het feit dat tussen 2000 en 2013 de grootte van deze groepen ook veranderd is.

HOOGSTE OPLEIDING	MEDIANE LOON (2000-2013)
Totaal	+ 0,9%
Lager dan high school	- 7,9%
High school	- 4,7%
College	- 7,6%
Bachelor of hoger	- 1,2%

Tabel 1. Stijging van het mediane loon per opleidingsniveau tussen 2000 en 2013



Figuur 1. Comet chart. Bron: Armstrong & Wattenberg (2014)

Er zijn meer werknemers die hoger opgeleid zijn, dus het mediane loon voor alle werknemers is toegenomen, maar per opleidingsniveau afgenomen. Simpsons paradox is genoemd naar de statisticus E.H. Simpson, die in 1951 daarover voor het eerst publiceerde. De paradox is een bijzonder geval van mix-effecten, waarbij de ontwikkeling van aggregaat-cijfers opgebouwd is uit echte ontwikkeling en uit ontwikkelingen van populatiegroottes.

Een bekend geval van Simpsons paradox trad op in 1973 toen de University of California-Berkeley aangeklaagd werd voor discriminatie (Freedman, Pisani & Purves, 2007; Bickel, Hammel & O'Connell, 1975). De universiteit accepteerde 44% van de mannelijke kandidaten en maar 35% van de vrouwelijke. Maar toen onderzoekers keken naar de cijfers ontdekten ze iets verrassends: opgesplitst naar faculteiten werden vrouwelijke kandidaten licht bevoordeeld (statistisch significant).

Het effect treedt vaak op in data-analyses maar wordt als zodanig dus niet altijd onderkend, ondanks (of dankzij?) het feit dat het meer dan 50 jaar geleden beschreven is.

Een interessante manier om je bewust te worden of het effect in jouw data optreedt, is het maken van een zogenaamde comet chart (Armstrong & Wattenberg, 2014). In deze grafiek (zie figuur 1) wordt populatie-

grootte op de x-as gezet en het aggregaat-cijfer op de y-as. De verandering tussen 2000 en 2013 wordt aangegeven met behulp van een pijl: de komeet. Uit een dergelijke grafiek wordt duidelijk hoe het aggregaat zich verhoudt tot haar onderdelen: het wordt ook duidelijker hoe verschillen in ontwikkeling opgebouwd zijn.

Blijft de vraag over voor de lezer: hoe bewust ben je van dit effect? Hoe vaak treedt het bij jou op?

LITERATUUR

Bureau of Labor Statistics (2014). *Current population survey labor force statistics – table a-4. employment status of the civilian population 25 years and over by educational attainment*. Gevonden op <http://www.bls.gov/webapps/legacy/cpsa-tab4.htm>.

Simpson, E.H. (1951). The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society, Ser. B*, 13, 238-241.

Freedman, D., Pisani, R., & Purves, R. (2007). *Statistics* (4th edition). New York: W.W. Norton & Company.

Bickel, P. J., Hammel, E. A., & O'Connell, J. W. (1975). Sex bias in graduate admissions: Data from Berkeley. *Science* 187(4175), 398-404.

Armstrong, Z., & Wattenberg, M. (2014). Visualizing statistical mix effects and Simpson's paradox. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12).

EDWIN DE JONGE is methodoloog bij het Centraal Bureau voor de Statistiek.

E-mail: <ejne@cbs.nl>



INTERBEOORDELAAR–BETROUWBAARHEID

Het coderen van kenmerken van survey items voor de constructie van vragenlijstprofielen

Tegenwoordig worden dezelfde surveys vaak afgenomen via meerdere modes, zoals in persoon, via de telefoon en het web. Afhankelijk van de wijze van afname kunnen de antwoorden van respondenten echter verschillen. Om inzicht te krijgen in voor wat voor soort vragen deze antwoorden verschillen, hebben we de items van enkele grote surveys, waaronder de Enquête Beroepsbevolking van het Centraal Bureau voor de Statistiek, door meerdere beoordelaars op hun kenmerken laten coderen. Uit het onderzoek blijkt dat de interbeoordelaar–betrouwbaarheid relatief laag is voor een groot aantal van de kenmerken. Dit suggereert dat onderzoekers voorzichtig moeten zijn in het beoordelen en het gebruik van kenmerken om inzicht te krijgen in mode-specifieke meetfouten.

FRANK BAIS, BARRY SCHOUTEN & VERA TOEPOEL

Tegenwoordig worden steeds meer surveys afgenomen via meerdere survey modes. Een gebruikelijke vorm hiervan is in eerste instantie afname via het web en aansluitend via telefoon en/of in persoon. Deze vorm is relatief gunstig, omdat de aanvankelijke afname via het web kosteneffectief is en de follow up afname via telefoon en/of in persoon zorgt voor een verhoging van het responspercentage. Afname via het web is immers relatief goedkoop en door in meerdere modes de survey af te nemen wordt een groter gedeelte van de groep beoogde respondenten aangesproken voor potentiële deelname. Een gevolg van afname via meerdere modes is dat antwoorden van respondenten op dezelfde vragen in verschillende modes niet per definitie vergelijkbaar zijn (Klausch, Hox & Schouten, 2013). De verschillende modes kunnen tijdens het antwoordproces leiden tot ver-

schillende meetfouten, welke mode-specifieke meetfouten worden genoemd. Deze mode-specifieke meetfouten maken het vergelijken van antwoorden in verschillende modes moeilijk en onbetrouwbaar. Rekening houden met mode-specifieke meetfouten wordt dikwijls onvoldoende gedaan, omdat dit moeilijk is, veel tijd kost, of omdat men zich hier niet van bewust is. In ons onderzoek proberen we inzicht te krijgen in het optreden en de omvang van deze mode-specifieke meetfouten. We gebruiken hier onder andere de Enquête Beroepsbevolking (EBB) van het Centraal Bureau voor de Statistiek (CBS) voor, welke één van de surveys is die dient voor de productie van de officiële CBS-statistieken. Het doel is uiteindelijk om te proberen om specifieke typen surveys en specifieke survey modes op elkaar toe te spitsen om zo mode-specifieke meetfouten te minimaliseren.

Het optreden en de omvang van mode-specifieke meetfouten worden onder andere bepaald door de kenmerken van de vragenlijst (Beukenhorst, Engelen, Van der Laan, Meertens en Schouten, 2013). We willen deze kenmerken samenvatten in wat we vragenlijstprofielen zullen noemen en onderzoeken in hoeverre deze vragenlijstprofielen variatie in antwoordgedrag kunnen verklaren. Indien die variatie substantieel blijkt, zou het mogelijk moeten zijn om op basis van deze profielen uiteindelijk respondent, vragenlijst en survey mode op elkaar toe te spitsen. Als een eerste stap willen we vragenlijstprofielen construeren op basis van kenmerken van een groot aantal items van enkele omvangrijke nationale surveys. Hiervoor zal elk item in de vragenlijst op de aanwezigheid van elk kenmerk afzonderlijk beoordeeld moeten worden. Voordat dergelijke vragenlijstprofielen geconstrueerd kunnen worden, zal onderzocht moeten worden in hoeverre betrouwbaar kan worden vastgesteld dat er sprake is van een bepaald kenmerk voor een bepaald item. In het huidige onderzoek zijn daarom alle items op elk geselecteerd kenmerk gecodeerd door meerdere beoordelaars om te onderzoeken in hoeverre items betrouwbaar gecodeerd kunnen worden. Oftewel, voor welke proportie items zijn meerdere beoordelaars het erover eens dat een bepaald kenmerk zich op bepaalde wijze al dan niet manifesteert?

Opzet coderingsonderzoek

In het huidige onderzoek zijn alle items van de Enquête Beroepsbevolking van het CBS en de tien kernstudies Bezit, Wonen, Inkomen, Gezondheid, Persoonlijkheid, Religie & Etniciteit, Politiek & Waarden, Familie & Huishouden, Werk & Scholing en Sociale Integratie & Vrije Tijd van het LISS Panel (geadministreerd door CentERdata) door twee of drie beoordelaars gecodeerd op 16 item-kenmerken. Aan de hand van typologieën van Saris en Gallhofer (2007) en Gallhofer, Scherpenzeel en Saris (2007), een aantal focusgroepen en een proefcodering is een lijst met uiteindelijk 16 item-kenmerken geconstrueerd. De beoordelaars van zowel deze focusgroepen en proefcodering als de daadwerkelijke codering zijn werknemers van Universiteit Utrecht (UU), het CBS en CentERdata en hebben relatief veel kennis van en ervaring met survey onderzoek. Tijdens de coderingen werden de beoordelaars geacht zich strikt te houden aan de omschrijving en categorieën van de kenmerken. Zeven kenmerken hebben volgens de literatuur relatief veel invloed in het opwekken van mode-specifieke meetfouten

en zullen we derhalve de lastige kenmerken noemen: Inhoud van de vraag, lastig taalgebruik van de vraag, emotionele lading, potentieel vermoeden van een filtervraag, gevoelige informatie, centraliteit en complexiteit van de antwoordopties. Deze kenmerken zijn gecodeerd door drie beoordelaars. De overige negen kenmerken tijd, condities, herinnering, hypothetische situatie, berekeningen, dubbelloops, mismatch, formulering en toelichting noemen we de makkelijke kenmerken. Deze kenmerken zijn gecodeerd door twee beoordelaars. Zie tabel 1.

Uitkomsten coderingsonderzoek

Voor elk kenmerk is de interbeoordelaar–betrouwbaarheid over alle items van alle surveys berekend. Zie tabel 1. Deze bestaat uit de totale proportie items waarvoor door beide of alle drie de beoordelaars dezelfde categorie gecodeerd is. Als vuistregel is aangehouden dat een interbeoordelaar–betrouwbaarheid van 0,80 of hoger acceptabel is. Voor wat betreft de makkelijke kenmerken bleken de kenmerken formulering en toelichting niet betrouwbaar gecodeerd te kunnen worden. Voor wat betreft de lastige kenmerken bleken inhoud van de vraag, lastig taalgebruik van de vraag, emotionele lading, potentieel vermoeden van een filtervraag, gevoelige informatie en centraliteit niet betrouwbaar gecodeerd te kunnen worden. Slechts het aangepaste kenmerk inhoud van de vraag met één objectieve en één subjectieve categorie en het kenmerk complexiteit van de antwoordopties van de lastige kenmerken bleken betrouwbaar gecodeerd te kunnen worden.

Verklaringen voor lage interbeoordelaar–betrouwbaarheid

Een eerste mogelijke verklaring voor de gevonden lage interbeoordelaar–betrouwbaarheid is het idee dat het relatief moeilijk is om de kenmerken sluitend te definiëren. Ook al zijn de kenmerken gebaseerd op de literatuur en extensieve discussies tussen de beoordelaars, het is voor bepaalde kenmerken vermoedelijk relatief moeilijk voor de beoordelaars om te bepalen welke categorie van toepassing is doordat de definitie niet doorslaggevend blijkt voor relatief veel items. Veel items blijken niet eenduidig te coderen op basis van de afgesproken definitie. Daarom verwachten wij dat er voor de meeste surveys relatief veel items zullen zijn die niet eenvoudig beoordeeld kunnen worden op bepaalde kenmerken, ongeacht hoe strak deze kenmerken conceptueel ook gedefinieerd mogen

zijn. De interpretatie van items kan niet alleen van beoordelaar tot beoordelaar verschillen, maar is ook afhankelijk van zaken zoals persoonlijkheid, voorgeschiedenis en stemming van de dag, welke allemaal hun invloed kunnen hebben op de wijze waarop een kenmerk geïnterpreteerd wordt en vervolgens hoe bepaalde items worden gecodeerd voor het desbetreffende kenmerk. Vanuit dit perspectief bezien zal de interbeoordelaar–betrouwbaarheid gedeeltelijk afhangen van welke beoordelaars de items coderen. De verwachting is dat zowel andere be-

oordelaars als dezelfde beoordelaars die dezelfde items nog een keer opnieuw coderen, zullen resulteren in verschillende interbeoordelaar–betrouwbaarheden.

Omgaan met lage interbeoordelaar–betrouwbaarheid

Hoe nu om te gaan met lage interbeoordelaar–betrouwbaarheid? Een eerste optie is het simpelweg negeren van

alle items waarvoor geen consensus gevonden is in het coderen, en om de vragenlijstprofielen puur te baseren op de items waarvoor alle beoordelaars het met elkaar eens waren voor een bepaald kenmerk. Dit heeft niet de voorkeur, omdat er op deze wijze informatie ongebruikt blijft en vragenlijstprofielen vooral voor wat betreft de lastige kenmerken gebaseerd zullen zijn op een beperkt aantal items. Een tweede optie is om de kenmerken te verfijnen en herdefiniëren op basis van alle items waarvoor geen consensus bereikt is. De nieuwe definitie wordt dan aangepast en aangevuld op basis van de inhoud van deze items met behulp van de literatuur. Op deze wijze bestrijken de definities alle beoordeelde items en kunnen vragenlijstprofielen voor elk kenmerk geconstrueerd worden op basis van alle items.

Om het subjectieve karakter van het coderen te voorkomen, is een derde optie het computeriseren van de definities van de kenmerken en het eigenlijke codeerproces. Door de definities en de beslissingen voor een categorie van alle kenmerken te formaliseren, worden subjectieve interpretaties van de beoordelaars voor items die binnen het eerder genoemde grijze gebied vallen vermeden. Het valt echter te bezien in hoeverre dit computeriseren realistisch blijkt, wat met name voor de lastige kenmerken vermoedelijk moeilijk zal zijn. Bovendien zijn ook hier beoordelaars en programmeurs nodig en ontkomt men niet aan het subjectieve karakter van het opstellen en programmeren van regels en beslissingen. Om zowel het interpretatieve karakter van het herdefiniëren als het computeriseren van de kenmerken te vermijden, is een vierde optie het construeren van schalen met meerdere categorieën in plaats van enkel de categorieën van toepassing en niet van toepassing. Dit betekent dat bijvoorbeeld geen, één, twee of drie beoordelaars voor een bepaald item hebben aangegeven dat een bepaald kenmerk van toepassing is. Hier kan een vragenlijstprofiel op gebaseerd worden dat de aanwezigheid van een bepaald kenmerk gradueel weergeeft, afhankelijk van hoeveel beoordeelaars hebben aangegeven dat sprake van het kenmerk zou zijn.

Conclusie

Survey items kunnen niet betrouwbaar gecodeerd worden. Dit geldt voor de meeste lastige kenmerken, maar ook voor de makkelijke kenmerken formulering en toelichting. De lage interbeoordelaar–betrouwbaarheid kan mogelijk verklaard worden doordat kenmerken moeilijk sluitend te definiëren zijn en dat enige subjectieve interpretatie tijdens het beoordelen onvermijdelijk lijkt en

bovendien afhankelijk is van de beoordelaar. Dit is van belang voor elke onderzoeker die met item-kenmerken werkt. Dat wat de één bestempelt als een vraag over bijvoorbeeld kennis, gedrag of attitude, kan door iemand anders op andere wijze geclassificeerd worden. Dit kan een belangrijke oorzaak zijn van tegenstrijdige conclusies over item-kenmerken in de literatuur. Doordat survey items niet betrouwbaar gecodeerd kunnen worden op hun kenmerken, is het moeilijk om deze te gebruiken voor onderzoek naar mode-specifieke meetfouten. Toekomstig onderzoek zal zich moeten richten op het omgaan met lage interbeoordelaar–betrouwbaarheid.

LITERATUURLIJST

- Beukenhorst, D., Buelens, B., Engelen, F., Van der Laan, J., Meertens, V., & Schouten, B. (2013). *The impact of survey item characteristics on mode-specific measurement bias in the Crime Victimization Survey*. Discussion paper 201416. Den Haag: Statistics Netherlands.
- Gallhofer, I., Scherpenzeel, A., & Saris, W. E. (2007). *The codebook for the SQP program*. Gevonden op <www.sqp.nl>.
- Klausch, L. T., Hox, J., & Schouten, B. (2013). Measurement effects of survey mode on the equivalence of attitudinal rating scale questions. *Sociological Methods and Research*, 42, 227-263.
- Saris, W. E., & Gallhofer, I. (2007). Estimation of the effects of measurement characteristics on the quality of survey questions. *Survey Research Methods*, 1, 29-43.

FRANK BAIJS heeft een onderzoeksmaster psychologie gevolgd aan de Universiteit van Amsterdam met als major Methoden & Statistiek. Sinds januari 2014 is hij promovendus op de afdeling Methoden & Statistiek van de Universiteit Utrecht. Zijn positie betreft een samenwerkingsproject tussen de UU en het CBS waarin met behulp van kenmerken van respondenten en vragenlijsten geprobeerd wordt om inzicht te verkrijgen in mode-specifieke meetfouten.
E-mail: <f.bais@uu.nl>

BARRY SCHOUTEN is als senior methodoloog werkzaam bij het Centraal Bureau voor de Statistiek, na een PhD in wiskundige statistiek aan de Vrije Universiteit van Amsterdam. Hij doet onderzoek naar methoden voor reductie en correctie van nonrespons en meetfouten in surveys. Hij heeft verschillende publicaties op dit terrein.
E-mail: <bstn@cbs.nl>

VERA TOEPOEL is universitair docent bij Methoden & Statistiek aan de Universiteit Utrecht. Zij is gepromoveerd aan de Universiteit van Tilburg met de thesis 'A closer look at web questionnaire design'. Haar onderzoeksinteresse ligt op: survey design, selectie-effecten, meet-effecten, modus van afname etc. Ze is voorzitter van het Nederlandstalig Platform voor Survey Onderzoek en heeft verscheidene publicaties op het gebied van survey-onderzoek, o.a. het boek *Doing surveys online* (2015).
E-mail: <v.toepoel@uu.nl>

KENMERK	OMSCHRIJVING	CODENUMMER EN CODECATEGORIE	α
Tijd	Naar welke tijd refereert het item?	1 verleden 2 heden 3 toekomst	0,85
Conditie	Bevat het item condities of voorwaarden?	0 nee 1 ja	0,89
Herinnering	Is er sprake van een bepaalde herinnering?	0 geen herinnering 1 herinnering non-specifiek 2 herinnering < 1 maand 3 herinnering > 1 maand	0,85
Hypothetische situatie	Is er sprake van een concrete, specifieke hypothetische situatie?	0 nee 1 ja	0,98
Berekeningen	Dient respondent een bepaalde berekening uit te voeren?	0 nee 1 ja	0,94
Dubbelvragen	Bevat het item meerdere vragen of is het potentieel verwarrend?	0 nee 1 ja	0,96
Mismatch	Vraag en antwoordcategorieën sluiten niet op elkaar aan.	0 nee 1 ja	0,98
Formulering	Is de vraag geformuleerd als een stelling?	0 nee 1 ja	0,57
Toelichting	Is er toelichting aanwezig bij een item?	0 nee 1 ja	0,71
Inhoud van de vraag 1	Naar wat voor inhoudelijk aspect wordt gevraagd?	1 feitelijk gedrag 2 overig feitelijk 3 mening 4 tevredenheid 5 overig subjectief	0,56
Inhoud van de vraag 2	Naar wat voor inhoudelijk aspect wordt gevraagd?	1 objectief 2 subjectief	0,90
Lastig taalgebruik vraag	Bevat vraag moeilijke woorden of zinsconstructies?	0 nee 1 ja	0,61
Emotionele lading	Bevat het item emotionele woorden of lading?	0 nee 1 ja	0,75
Vermoeden van filtervraag	Kan respondent vermoeden dat vraag filtervraag zou kunnen zijn?	0 nee 1 ja	0,62
Gevoelige informatie	Bevat het item gevoelige informatie van maatschappelijke of huishoudelijke aard?	0 nee 1 ja	0,53
Centraliteit	Valt de vraag buiten de kennis- of belevingswereld van respondent?	0 nee 1 ja	0,59
Complexiteit antwoordopties	Bevatten antwoordopties moeilijke woorden of zinsconstructies?	0 nee 1 ja	0,91

Tabel 1. Omschrijving, categorieën en interbeoordelaar–betrouwbaarheden α van de kenmerken



Toevalligheden en onmogelijkheden

De Engelse krant *The Sun* had onlangs een opmerkelijk nieuwsbericht. De krant bracht het bericht dat het derde kind van een Engels echtpaar was geboren om 7.43 uur 's ochtends en wel precies op hetzelfde tijdstip als waarop elk van de eerdere twee kinderen van het echtpaar was geboren. Uitgaande van de wat wonderlijke veronderstelling van een tijdsvenster van 12 uur per dag voor geboortes, rekende de krant haar lezers voor dat de kans dat drie opeenvolgende kinderen uit eenzelfde gezin op precies hetzelfde tijdstip geboren zullen worden gelijk is aan 1 op de $720 \times 720 \times 720$, oftewel een kans van ongeveer 1 op de 370 miljoen. Attente lezers van *STATOR* zullen ongetwijfeld direct opmerken dat de krant dezelfde klassieke fout maakte die ook menigeeen maakt wanneer het antwoord 1/49 wordt gegeven op de vraag wat de kans is dat twee willekeurig gekozen personen op eenzelfde dag van de week jarig zijn. De krant had natuurlijk moeten stellen dat de kans 1 op de $720 \times 720 = 518.400$ is. Nog steeds een heel kleine kans, maar in het licht van ongeveer 3 miljoen gezinnen in de UK met drie of meer kinderen is het in *The Sun* met chocoladeletters gebrachte nieuws minder opmerkelijk dan in eerste instantie gedacht zou kunnen worden.

Zo zijn er in het dagelijks leven vele zogenaamde toevallige gebeurtenissen die in het juiste perspectief geplaatst in het geheel niet toevallig zijn. In 1986 won mevrouw Evelyn Mary Adams in een tijdsbestek van vier maanden tweemaal de jackpot in de New Jersey State

Lottery, waar de kans ongeveer 1 op de 17 biljoen is dat een gegeven persoon twee keer de jackpot wint bij twee keer deelname met elke keer één ingevuld rijtje. Hoe verast moet je zijn over een dergelijke gebeurtenis? Een gebeurtenis zoals mevrouw Adams overkwam, wordt minder verrassend wanneer je bedenkt dat er vele loterijen zijn zoals de New Jersey State Lottery en talloze lotospelers die elke week weer hun geluk beproeven met vrijwel altijd meerdere ingevulde rijtjes. Onder realistische aannames kan met eenvoudige kansrekening aangetoond worden dat met een kans dicht bij 1 er ergens onder die miljoenen en miljoenen lottospelers iemand zal zijn is die in een kort tijdsbestek twee of meer keer de jackpot zal winnen. Een ander opmerkelijk gebeuren vond op woensdag 21 juni 1995 plaats in de Duitse lotto. Bij de trekking op die dag werden dezelfde zes getallen 15-25-27-30-42-48 getrokken als op zaterdag 20 december 1986. Een herhaling van eerder getrokken getallen gebeurde toen voor het eerst in het bestaan van de Duitse lotto nadat er 3016 trekkingen waren geweest. Deze gebeurtenis is bij nadere beschouwing minder verrassend dan het lijkt, net zo min als het gebeuren om in een schoolklas van 23 leerlingen twee kinderen te treffen die op eenzelfde dag jarig zijn. Wel is uitzonderlijk toevallig een gebeuren in de Bulgaarse lotto waarin op 6 en 10 september 2009 in twee opeenvolgende trekkingen dezelfde zes getallen getrokken werden uit de getallen 1 tot en met 42. Deze gebeurtenis is weliswaar verbazingwek-

kend, maar in het licht van de vele lotto's in de wereld die ieder jaar weer vele trekkingen kennen, is deze gebeurtenis niet zodanig ongeloofwaardig dat meteen aan Bulgarenfraude gedacht moet worden.

Er zijn vele andere voorbeelden te geven van zogenaamde toevalligheden die bij nadere analyse toch niet zo toevallig zijn. Je zou kunnen spreken over het loterij-principe (de befaamde kansrekenaars Persi Diaconis en Frederick Mosteller noemen het de wet van de heel grote getallen): hoe klein ook de kans op een bepaalde gebeurtenis is, bij voldoende gelegenheden tot optreden zal de gebeurtenis altijd een keer plaatsvinden op grond van toeval. Is dit wel zo? Zijn er niet gebeurtenissen die zo onwaarschijnlijk zijn dat ze nooit zullen optreden? Waar trek je een grens? Iemand die zich over deze vragen boog, was de wiskundige Émile Borel (1871–1956). Deze beroemde Franse wiskundige – die met de Borel verzameling, het Borel-Cantelli lemma en de sterke wet van de grote aantallen fundamentele bijdragen aan de kansrekening heeft geleverd – publiceerde voor een breed publiek het boek *Les Probabilités et La Vie*, een boek dat in 1962 bij Dover in een Engelse vertaling verscheen als *Probability and Life*. In dit boek onderscheidde Borel drie gevallen voor kansen die zo klein zijn dat ze praktisch gesproken verwaarloosbaar zijn: een gebeurtenis met een kans van 10^{-6} is fysisch onmogelijk vanuit menselijk perspectief, een gebeurtenis met een kans van 10^{-15} is fysisch onmogelijk vanuit aardse perspectief en een gebeurtenis met een kans van 10^{-50} is fysisch onmogelijk vanuit kosmisch perspectief. Dit principe staat bekend als Borels wet. Uiteraard is deze wet niet een wiskundige stelling, maar een illustratieve vuistregel ontworpen door een wetenschapper die lang had nagedacht over extreem onwaarschijnlijke gebeurtenissen. In weerwil van de suggestieve titel *Probability and Life* van zijn boek, ging Borel in dit boek niet in op evolutionaire of abiogenese gerelateerde zaken (in zijn latere boek *Probabilité et Certitude* zou Borel dit wel doen). Echter in discussies tussen creationisten en evolutionisten wordt Borels wet veelvuldig als argument gebruikt.

Laat ik me niet begeven in discussies hoe waarschijnlijk het is dat een Schepper wel of niet bestaat, maar me richten op een down-to-earth-probleem. Wat is de kans op vier perfecte handen in bridge? Preciezer geformuleerd, stel dat de 52 kaarten in een volledig random volgorde liggen, wat is dan de kans dat bij het delen van de kaarten elk van de vier spelers 13 kaarten krijgt van dezelfde soort (klaveren, ruiten, harten, schoppen)? Deze kans is onvoorstelbaar klein en wordt gegeven door $4,47 \times 10^{-28}$. Om een voorstelling te krijgen van hoe klein

deze kans wel is, het is ongeveer de kans om 91 keer achter-elkaar kop te gooien bij het werpen met een zuivere munt. Dit valt onder de noemer onmogelijk op een aardse schaal. Om deze uitspraak te onderstrepen, stel dat vier miljard mensen bridge zouden spelen en elke dag 25 keer de kaarten zouden delen, dan zou gemiddeld ongeveer 1 keer in de 245 biljoen jaar sprake zijn van vier perfecte handen. Niettemin bracht de *Daily Mail* op 25 november 2011 groot het nieuws dat tijdens een bridgepartij in een dorpje in graafschap Warwickshire elk van de vier bridgespelers bij het delen 13 kaarten van dezelfde soort had gekregen. Het nieuws in de krant ging vergezeld met een foto van vier betrouwbaar uitzijnde pensionado's, ook nog trouwe kerkgangers, om duidelijk te maken dat hier geen sprake was van een misplaatste grap. De krant stelde in het begeleidende commentaar dat ook veilig aangenomen mocht worden dat dit de eerste keer in de geschiedenis van bridge was dat sprake was van vier perfecte handen in het bridgespel. De redacteur van het artikel had blijkbaar niet de moeite genomen enig zoekwerk met Google te doen, want dan was het hem snel duidelijk geworden dat al ettelijke malen in een krant bericht was over het optreden van vier perfecte handen in bridge. Altijd gebeurde dit bij een lokale bridgepartij en nimmer bij een professioneel bridgetoernooi.

De wereld is vergeven van de grappenmakers. Op 1 april 1959 zou bij het bridgespel in de eerbiedwaardige St. James' club in Londen sprake zijn geweest van vier perfecte handen bij de bridgespelende Lords in deze club. Het kan natuurlijk ook zijn dat bridgespelers naar eer en geweten de krant op de hoogte stellen van het ongelooflijke gebeuren dat hen overkomen is. De verklaring kan dan alleen maar zijn dat de kaarten niet goed geschud zijn. Stel dat een nieuw pak kaarten waarin de kaarten op soort geordend zijn, worden geschud door tweemaal een *perfect riffle* (de stapel halveren en dan de twee helften links-rechts afwisselend in elkaar ritsen) toe te passen, dan zijn na het schudden de kaarten weer in de oorspronkelijke volgorde en treedt de situatie van vier perfecte handen op wanneer elke speler 13 kaarten tegelijk van bovenaf gedeeld krijgt. Het schudden van de kaarten om te bereiken dat de kaarten in een voldoende random volgorde komen te liggen, zonder de kaarten een kwartier in de wasdroger te laten ronddraaien, is handmatig veel lastiger dan vaak beseft wordt.

HENK TIJMS is emeritus hoogleraar operations research aan de Vrije Universiteit en auteur van diverse leerboeken over operations research en kansrekening.
E-mail: <tijms@quicknet.nl>



MET GOOGLE OP WEG NAAR DE AUTOMATISCHE STATISTICUS

RICHARD STARMANS

Big data is een *moving target* met een aanhoudend verschuivende en dynamisch aan te passen betekenis. Het is functioneel vaag en context-afhankelijk gedefinieerd als data die de gebruikelijke technieken van *storage* en *retrieval* overstijgen, niet alleen wat betreft veelheid, maar ook inzake snelheid, verscheidenheid in type, vorm en bron, daarnaast variatie in kwaliteit en status van data en niet in de laatste plaats betreffende het gedistribueerde karakter ervan. Zo beschouwd is het verschijnsel allerminst nieuw en kent in feite een lange genealogie (Starmans, 2015a). Tegen deze achtergrond vormt *data science* de technisch-wetenschappelijke discipline die zich toelegt op het managen van deze veelheid aan data: het verkrijgen of produceren, opslaan, bewerken, toegankelijk maken, beveiligen, analyseren en interpreteren ervan. Informatica en statistiek kunnen daarbij als de toeleverende, meer canonieke disciplines worden beschouwd: informatica van oudsher gericht op datastruc-

turen, algoritmieken en schaalbaarheid, statistiek gericht op het analyseren en interpreteren van gegevens.

Ofschoon de liaison tussen beide in de praktijk dikwijls nog broos, onwennig en soms zelfs moeizaam blijkt, kan met enige goede wil toch een drietal pijlers of 'zuilen' worden onderkend, waardoor deze samenwerking wordt geschraagd (Starmans, 2015b). De eerste betreft database en *information retrieval* technologie. De laatste jaren is door met name Google en Yahoo gewerkt aan technieken om vele computers in data-centra te clusteren en terabytes aan data toegankelijk te maken en te bewerken: denk bij voorbeeld aan BigTable, Google File Systems, aan een programmeertaalparadigma als Map Reduce, en de *open source* variant Hadoop. De tweede zuil betreft intelligente algoritmen uit de hoek van *computational intelligence*, dat *machine learning* (ML) en *data mining* omvat en een veelheid aan uiteenlopende technieken behelst, zoals neurale netwerken,

genetische algoritmen, *decision trees*, *association rules*, support vector-machines, ensemble-technieken, en Bayesian learning. Ruwweg passen deze in de traditie van de subsymbolische *artificial intelligence* (AI). De derde pijler tot slot wordt gevormd door de wiskundige, inferentiële statistiek, waarbij op basis van schattingstheorie en het dualisme van steekproef en populatie kan worden geregeneerd met onzekere informatie: parameterschatting, hypothesetoetsen, *p-values* en betrouwbaarheidsintervallen. De successen van bedrijven als Google en Yahoo hebben vooral bijgedragen aan de versteviging van de eerste pijler, door zowel productontwikkeling als vernieuwend informatica-onderzoek.

Daarnaast financieren voornoemde bedrijven ook regelmatig universitair of privaat gefinancierd onderzoek, hetgeen dikwijls al borg staat voor forse media-aandacht. Zo maakte Google vorig jaar bekend een bedrag van 750.000 euro te investeren in het project 'The Automatic Statistician; an artificial intelligence for data science', dat plaatsvindt aan de Universiteit van Cambridge onder leiding van Zoubin Ghahramani. Het initiatief is relatief kleinschalig en kent nog allerminst wereldschokkende bevindingen, maar trekt wel volop aandacht en niet alleen door de betrokkenheid van Google of door de eloquente en uitermate doordachte titel ervan. Evenmin wordt de aandacht gevoed door zwaarmoedige overpeinzingen over de dreigende teloorgang van het beroep van statisticus, dat enige jaren geleden door Google's topeconoom Hal Varian nog werd getypeerd als de meest sexy professie van de 21e eeuw. Integendeel, *The Automatic Statistician* kent belangwekkende filosofische implicaties, die ook de overige zuilen van data science betreffen. Enige aspecten van de allerminst onproblematiese relaties tussen AI, Statistiek en de 'Googleaanse' kijk op data-analyse zullen hier kort voor het voetlicht worden gebracht.

Het homo mensurae-principe

Kevin Murphy, Senior Research Scientist van Google licht op de website van Cambridge de samenwerking toe: '*In recent years, machine learning has made tremendous progress in developing models that can accurately predict future data. However, there are still several obstacles in the way of its more widespread use in the data sciences. The first problem is that current Machine Learning (ML) methods still require considerable human expertise in devising appropriate features and models. The second problem is that the output of current methods, while accurate,*

is often hard to understand, which makes it hard to trust.' Volgens Murphy is het zonneklaar dat '*the automatic statistician project from Cambridge aims to address both problems, by using Bayesian model selection strategies to automatically choose good models / features, and to interpret the resulting fit in easy-to-understand ways, in terms of human readable, automatically generated reports. This is a very promising direction for ML research, which is likely to find many applications at Google and beyond.*' De initiatiefnemers gaan nog een stap verder: '*making sense of data is one of the great challenges of the information age we live in. While it is becoming easier to collect and store all kinds of data (...), there are relatively few people trained in the statistical and machine learning methods required to test hypotheses, make predictions, and otherwise create interpretable knowledge from this data. The Automatic Statistician project aims to build an artificial intelligence for data science, helping people make sense of their data.*'

Bovenstaande citaten bevatten volop verwijzingen naar klassieke AI-debatten, zwaar verankerd in het aloude homo mensurae-principe, waarbij uiteenlopende, mogelijk tegenstrijdige aspecten van 'de menselijke maat' gestalte krijgen: de behoefte aan begrip en verklaring (*intelligibility*), ondersteuning bij saaie, lastige taken, opheffing van de willekeur van menselijke keuzes, het ideaal van mechanische productie van kennis, dreiging de mens te vervangen of te overheersen, maar ook de evolutie van data via informatie / betekenis naar kennis. Hoe krijgen deze AI-issues nu gestalte in statistiek en big data? Daarover is het project nog vaag. Uiteraard is ook de AI geëvolueerd, AI (kennisrepresentatie) en *brute force* rekenkracht stonden lange tijd tegenover elkaar. Belangrijker zijn echter enkele hedendaagse ontwikkelingen binnen AI, statistiek en data-analyse, die de reikwijdte van dit project van Ghahramani cum suis weliswaar overstijgen, maar tegen de achtergrond waarvan elk streven een *AI for data science* te ontwikkelen moet worden geduid. Een viertal passeert hier kort de revue.

De naderende singulariteit

De eerste ontwikkeling is moeilijk te missen en betreft de recente herleving van de dystopie van de AI, mede doordat beroemdheden als de ict-pionier Bill Gates, de fysicus Stephen Hawking en de linguïst en politiek agitator Noam Chomsky de risico's van AI voor de mensheid benadrukten. Zij grepen daarbij onder meer terug op de vaak in apocalyptische termen geduide notie van de singulariteit. Precies tien jaar geleden publiceerde de

uitvinder en futuroloog Ray Kurzweil zijn boek *The singularity is near; when humans transcend biology* (2005), dat een bestseller werd. Al vele jaren daarvoor had Kurzweil de toon gezet in het AI-debat, onder meer met studies als *The Age of Intelligent Machines* (1990), waarin hij samen met beroemdheden als zijn leermeester Marvin Minsky, Daniel Dennett en Douglas Hofstadter, de genealogie van de AI schetst en de haalbaarheid van haar 'klassieke' ambities. Het overtreffen van de mens vindt zijn bekroning in voornoemde studie over de singulariteit, een grillig begrip met vele connotaties. In de wiskunde is het een merkwaardig punt, dat niet is gedefinieerd, oneindig groot wordt, of asymptotisch gedrag veroorzaakt. In de kosmologie is het een punt met oneindig klein volume en oneindig grote dichtheid. In de singulariteit werken systemen elkaar tegen, worden opgeblazen, evenwichten verdwijnen of ontstaan. Bij Kurzweil is de singulariteit het moment dat de technologisch vooruitgang door exponentiele groei 'oneindig' wordt of althans door de mens zo 'begrepen' wordt en de mens als zelfstandige levensvorm overgaat in deze totale intelligentie, die het heelal doordringt, met de menselijke geest niet te bevatten, laat staan te beheersen. Ofschoon weinigen zullen stellen dat de voorspelling reeds is uitgekomen, zijn sommige voorwaarden voor deze totale intelligentie wel al manifest. Een verwant aspect betreft het ideaal van het transhumanisme, dat volgens sommige de uiterste consequenties trekt uit de ambities van het humanisme en stelt dat we technologie stelselmatig moeten gebruiken om de beperkingen van het menselijk lichaam te overwinnen en *la condition humaine* te verbeteren: hogere intelligentie, meer geluk en hogere moraliteit door *cognitive enhancement*, nanorobots en computers die in het lichaam worden geïmplanteerd en een verlengstuk van de mens worden. Door deze mens-machine-integratie zal de grens tussen deze 'verbeterde' werkelijkheid en daadwerkelijke virtuele werelden vervagen, een concrete stap op weg naar Kurzweils singulariteit. Sommigen beschouwen het transhumanisme als een nieuwe cesuur, een post-darwiniaans era, anderen stellen dat het transhumanisme juist deel uitmaakt van de evolutie en de evolutietheorie de mens nooit als voltooid heeft beschouwd. In het recente *Superintelligence* (2014) van Nick Bostrom worden de grenzen opnieuw verlegd. Hoe dan ook, nadat AI jaren lang symbool stond voor niet uitgekomen verwachtingen en utopisch leek, overheerst in het hedendaagse debat vooral de dystopie. Bezorgdheid is er niet alleen bij futurologen en doemdenkers,

maar ook in meer gematigde kringen. De open brief die Stephen Hawking en raketgeleerde Elon Musk recentelijk opstelden als een manifest voor *Robust and Beneficial AI* werd door vele vooraanstaande, ook Nederlandse wetenschappers ondertekend. Al met al zal elke poging een AI voor data science te ontwikkelen tegen deze achtergrond gestalte krijgen. Als aspect van de menselijke maat komt bij de automatische statisticus zeker ook de morele dimensie om de hoek kijken, gelet op de langdurige, moeizame relatie die ethiek en statistiek onderhouden. Wellicht dat Disraeli's notoire verzuchting dat '*there are lies, damned lies and statistics*' middels een AI voor data science beslissend kan worden ontkracht.

De datarevolutie

Tegenover de dystopie van de AI staat paradoxaal genoeg de utopie van de datarevolutie. Of het nu gaat om wetenschap, economie en welvaart, gezondheidszorg, milieu- en klimaatproblematiek, rechtspraak of de nationale veiligheid, het belang van big data en data science voor het huidige tijdsgewicht wordt in alle toonaarden benadrukt. Soms tekent zich een heuse mythetvorming af, en wordt zelfs een panacee voor problemen van velerlei aard in het vooruitzicht gesteld. In hun recente boek *Big Data: A Revolution that will Transform how we Live, Work and Think* (2013) scharen Viktor Mayer-Schönberger en Kenneth Cukier zich nadrukkelijk in deze traditie. Op soms aanstekelijke wijze schetsen zij de contouren van een gedatificeerde wereld, waarin de werkelijkheid in zijn volle omvang langzamerhand gecodeerd wordt in data en op die manier toegankelijk en kenbaar wordt. Volgens Davenport (2012) zou de data scientist het meest sexy beroep van de 21e eeuw worden en de mogelijkheden van het vak worden nog dagelijks in tal van publicaties bezongen. De eudaimonische voltooiing van de menselijke zoektocht naar zelfrealisatie kan volgens deze protagonisten immers alleen dankzij de data scientist worden begeleid en gerealiseerd. Uiteraard wordt ook de keerzijde enigszins belicht, maar het optimisme overheerst, zoals indertijd in de vroege dagen van de AI. Omdat de al dan niet vermeende dreiging van AI pas via big data gestalte kan krijgen is de tegenstelling op zijn minst paradoxaal. We zullen hier niet trachten een en ander te ontwarren, maar duidelijk mag zijn dat ook dit aspect de achtergrond bepaalt waartegen het debat over een AI voor data science wordt gevoerd.

Google versus Wigner

Een derde ontwikkeling speelt zich af op methodologisch niveau en betreft het spanningsveld tussen enerzijds de computationele, op veel data en ML algoritmen gebaseerde, data-analyse en anderzijds de wiskundige, inferentiële statistiek, ruwweg de tweede en derde pijler van data science (Starmans, 2015a). Het waren de Google-onderzoekers Alon Halevy, Peter Norvig en Fernando Pereira die onder verwijzing naar (Wigner, 1960) in een lucide maar door simplificaties ook enigszins omstreden artikel 'The Unreasonable Effectiveness of Data' claimden dat wanneer het om mensen gaat en niet om elementaire deeltjes, eenvoudige wiskundige formules en elegante theorieën van beperkt nut. Om gedrag van mensen te begrijpen en te voorspellen kan men beter vertrouwen op vele petabytes van ruwe, ongeannoteerde, ongestructureerde, desnoods vervormde data. Veel data en algoritmen versus het statistisch-wiskundig modellen, *computation versus estimation* (Halevy, 2009). Tegenover het monisme van de data-analyse met een in data gecodeerde waarneembare werkelijkheid, staat het dualisme van de statistiek met een waarneembare steekproef en de achterliggende 'theoretische' populatie. De 'data die voor zich spreken' versus de statistische theorie, inclusief aannamen en veronderstellingen, waarbij het model de (benadering van de) ware datagenererende kansverdeling moet bevatten. Uiteraard zijn er pogingen tot verzoening ondernomen (Starmans & Van der Laan, 2013), maar vanzelfsprekend bepaalt deze visie op data science mede welke conceptie van AI nodig en wenselijk is, zoals ook blijkt uit de volgende paragraaf.

Selectie en verklaring

Het laatste aspect van de onderhavige problematiek, dat hier wordt belicht is ook statistisch-methodologisch van aard, maar zeer concreet en raakt de ambitie van The Automatic Statistician 'to automatically select and explain models'. Murphy's conclusie dat de afhankelijkheid van menselijke interventie bij ML het voornaamste obstakel voor grootschalige toepassing ervan vormt, kent een cruciaal en dikwijls onderbelicht aspect: de keuze voor een bepaalde data-analytische of statistische techniek geschiedt vaak ad hoc of op basis van subjectieve, persoonlijke gronden. Dit, terwijl de resultaten sterk kunnen afhangen van karakteristieken van de gekozen

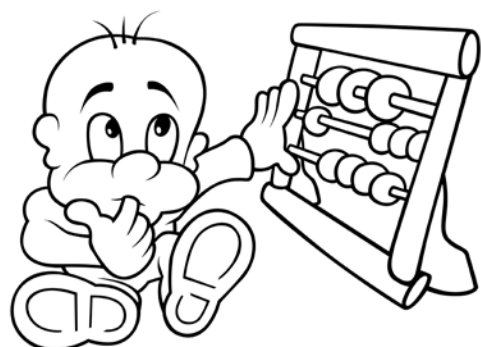
algoritmen of methoden. Een AI van data science kan bij de oplossing hiervan een belangrijke rol spelen. De verklaring of rechtvaardiging van de conclusies is zo mogelijk nog gewichtiger. Dat bleek niet alleen in de vroege dagen van expertsystemen, later beslissingsondersteunende systemen, maar ook in dit project, ja zelfs vanuit een hedendaagse Googleaanse kijk op data-analyse of vanuit een naderende singulariteit, waarin mensen de inferentiële mechanismen sowieso niet langer begrijpen. De vele verwijzingen naar de menselijke maat in het streven naar een AI voor data science, kan niet langs een dimensie van maximaliseren of minimaliseren van menselijke worden geschetst, maar heeft vele aspecten: interpreteerbare resultaten, vertrouwen, de mens die uiteindelijk betekenis moet toekennen, ook in het licht van een naderende singulariteit. Zo beschouwd is de keuze van Google voor een AI ook minder verrassend dan men op basis van de Wigner-Google controversie zou verwachten. Hoe dan ook, juist door de opmars van data science vormt de recente liaison van AI met data-analyse en statistiek ook voor de Filosofie van de AI nieuwe uitdagingen.

LITERATUUR

- Davenport, T.H., & Patil, D. J. (2012). Data scientist: The sexiest job of the 21st century. *Harvard Business Review*, 10.
- Halevy, A., Norvig, P., & Pereira, F. (2009). The Unreasonable effectiveness of data. *IEEE Intelligent Systems*, 24(2), 8-12.
- Mayer-Schönberger, V., & Cukier K. (2013). *Big data: A revolution that will transform how we live, work and think*. London: John Murray Publishers.
- Starmans, R. J. C. M., & Van der Laan, M. J. (2013). Inferential statistics versus machine learning; a prelude to reconciliation (in Dutch). *STATOR*, 24(2).
- Starmans, R. J. C. M. (2015a). The advent of data science: some considerations on the unreasonable effectiveness of data. In P. Buhlmann, P. Drineas, M. Kane & M. Van der Laan (Eds.), *Handbook of big data; Handbooks of modern statistical methods*. New York: Chapman & Hall/CRC Press.
- Starmans, R. J. C. M. (2015b). The ant, the spider and the bee; Bacon's unintended preamble to the data science debate (in Dutch). *Filosofie*, 25(3).
- Wigner, E. P. (1960). The unreasonable effectiveness of mathematics in the natural sciences. *Communications in Pure and Applied Mathematics*, 13(1).

RICHARD STARMANS is verbonden aan de Faculteit Bèta-wetenschappen (Department of Information and Computing Sciences) van de Universiteit Utrecht. Hij doet onderzoek op het snijvlak van filosofie, statistiek en informatica. E-mail: <starmans@cs.uu.nl>

Young Statisticians



ISCB 2015

The city of Utrecht will host the 36th International Society for Clinical Biostatistics annual conference in 2015, August 23–27. Before the ISCB 2015 conference starts, there will be an informal young researchers gathering in a pub in Utrecht. The plan is simple: we gather at 18.30h if you want to join for a simple dinner. From 20.30h, there will be drinks. There will be no science slam or formal agenda. It is all about meeting other PhD students and young researchers and having a cold drink together. Two drinks are included. Location is to be announced, <www.iscb2015.info>.



YOUNG STATISTICIANS DENMARK

We would like to introduce you to Young Statisticians Denmark, founded this year and partly inspired by the Dutch Young Statisticians as two of our members co-founded the organization. So if you are thinking of a position in Denmark, make sure to sign up with them! <www.youngstatisticians.dk>

NEW BOARD

Let us introduce our newest member of the board: Kees Mulder. Kees is a researcher focusing on Bayesian analysis of circular data, in which he is currently undertaking a PhD at Utrecht University. He is an enthusiastic organizer, with a broad sense of statistical humour. Welcome, Kees! We would also like to take the opportunity to once again thank Iris Eekhout and Nadia Vendrig for their contribution to the Young Statisticians.

YS-board: Nynke Krol, Sanne Willems, Kees Mulder
E-mail: <chair@youngstatisticians.nl>

UPCOMING ACTIVITIES

Keep updated by subscribing to our mailing list, send an e-mail to <chair@youngstatisticians.nl> and fill out our members survey. And are you interested in organizing an activity or have an idea, for example for our next workshop? Please send an email!

Oud-voorzitter prof. dr. Kobus Oosterhoff overleden

Jacobus Oosterhoff is 7 mei 1933 in Leeuwarden geboren en studeerde aanvankelijk een jaar geologie en daarna wiskunde aan de Universiteit van Amsterdam, waar hij in 1963 cum laude het doctoraal examen aflegde, met hoofdrichting Mathematische Statistiek en Waarschijnlijkheidsrekening. Hij promoveerde in 1969, wederom cum laude, tot doctor in de Wis- en Natuurkunde op het proefschrift *Combination of one-sided statistical tests*, met als promotor professor W. R. van Zwet.

Hij was van 1961 tot 1970 verbonden aan de afdeling Mathematische Statistiek van het Mathematisch Centrum in Amsterdam, eerst als assistent, maar vanaf 1963 als wetenschappelijk medewerker, in de jaren 1967-1970 als sous-chef van de deze afdeling. Van 1967 tot 1969 had hij de leeropdracht 'Biometrika' aan de Universiteit van Amsterdam. Van 1971 tot 1974 was hij lector in de Mathematische Statistiek en Waarschijnlijkheidsrekening aan de Katholieke Universiteit te Nijmegen en vanaf 1974 tot zijn emeritaat hoogleraar Mathematische Statistiek aan de Vrije Universiteit in Amsterdam. Daar stond hij aan de basis van de huidige onderzoeksgroep in dit vak en heeft hij als decaan van de voormalige faculteit der Wiskunde en Informatica op bestuurlijk gebied een belangrijke bijdrage geleverd. Daarnaast was hij vanaf 1971 lange tijd adviseur van de afdeling Mathematische Statistiek van het Mathematisch Centrum.

Hij was Elected Member of the ISI, IMS fellow en zeer actief in de Bernoulli Society. Ook van de VvS+OR was hij een betrokken lid, zo was hij voorzitter van 1976 tot 1979 en publiceerde hij vier artikelen in *Statistica Neerlandica*.

Wij gedenken hem met grote eerbied en respect en wensen zijn familie sterkte toe bij dit verlies.

BESTUUR VvS+OR

IN MEMORIAM

Prof. dr. Kobus Oosterhoff (1933–2015)



Persoonlijke herinneringen

Het onverwachte bericht van het overlijden van Kobus Oosterhoff was voor mij een grote schok. Eind januari was hij nog bij mij thuis geweest, omdat ik hem graag mijn nieuwe, samen met Geurt Jongbloed geschreven boek wilde geven met een opdracht erin. Hij vertelde dat hij net weer was goedgekeurd om auto te rijden. Hij liep wat gebogen, maar was verder nog helemaal de Kobus Oosterhoff zoals ik hem had gekend.

Iedereen heeft wel een aantal mensen in zijn leven die een belangrijke rol hebben gespeeld en Kobus Oosterhoff was zeker zo iemand voor mij. Ik was zijn tweede promovendus, de eerste was Wilbert Kallenberg. Kobus kwam altijd op woensdag naar het Mathematisch Centrum om te praten met mij en Richard Gill, zijn andere promovendus. In het boek dat ik hem in januari gaf, had ik geschreven: 'Voor Kobus, in dank voor de voortreffelijke begeleiding van mijn eerste onderzoekstapjes' en zo voelde ik dat ook. Kobus had echt plezier in zijn taak als promotor en zoiets is heel belangrijk om de moed erin te houden tijdens de vier jaar dat je met zo'n proefschrift bezig bent. In tegenstelling met hoe het tegenwoordig meestal gaat liet hij mij geheel vrij in wat ik ging doen en vroeg toen het eerste hoofdstuk af was: 'En, wat ga je nu doen?' Hij las alles altijd heel nauwkeurig, iets wat zijn eerste promovendus ook nog heeft gememoreerd, en zei vaak tegen mij: 'Je laat de mensen wel werken!', daarmee bedoelend dat ik niet voldoende uitleg verschaftte bij de dingen die ik opschreef. We hebben samen drie artikelen geschreven, waaronder één artikel samen met Frits Ruymgaart over large deviations in de *Annals of Probability* van 1979. Ik hoorde van Richard Gill dat dit artikel Kobus' meest geciteerde artikel is.

Volgens zijn oudste zoon was het moeilijk hem van mening te doen veranderen, maar dat is toch één keer gebeurd in mijn herinnering. Ik had aan een organisatie in Nederland beloofd na het eerste wereldcongres van de Bernoulli Society in Tashkent in 1986 een bezoek aan de

refuseniks in Moskou te brengen. Hoewel geenszins illegaal was dit geheel tegen de Russische 'partijlijn', maar ook tegen de toenmalige Nederlandse partijlijn. Vanuit Moskou was zelfs gedreigd het congres te cancelen als mensen de refuseniks zouden opzoeken. Onzin natuurlijk, want het congres was een grote propagandagebeurtenis. Ook had ik samen met een aantal andere kansrekenaars/statistici een soort dreigbrief gekregen die liep via kanalen in Moskou en Nederland. In het vliegtuig van Moskou naar Tashkent was zelfs nog tegen mij gezegd: 'Als je maar weet Piet dat het doorzetten van dat bezoek aan de refuseniks ernstige gevolgen zal hebben voor je positie in de Bernoulli Society!' Een betere aanmoediging om wel te gaan had ik niet kunnen krijgen! Kobus zei hierover tijdens zijn laatste bezoek aan mij: 'Nou dat klinkt wel dreigend, zeg!', waarbij ik eigenlijk niet zeker weet of hij een grapje maakte of ernstig was (iets waarvan ik wel vaker niet zeker was bij hem, omdat hij geneigd was grappen met een heel ernstig gezicht te maken). Kobus was aanvankelijk tegen dat bezoek aan de refuseniks. Maar toen Peter Bickel en Fred Steutel hadden besloten met mij mee te gaan – Laurens de Haan was al van begin af aan bij de actie betrokken – besloot ook Kobus met ons mee te gaan. Laurens de Haan schreef mij hierover nog op 1 maart: 'Hij heeft toen van harte meegedaan. Hij was/is een aardige man die zich ook altijd verantwoordelijk gedragen heeft.' Zo is het.

Een uitvoeriger stuk over zijn wetenschappelijk werk zal elders verschijnen. Hierboven een zeer karakteristieke foto van Kobus, waarbij hij niet door maar over zijn brillette heen kijkt. De foto is door Chris Klaassen genomen op de tachtigste verjaardag van Kobus' eigen promotor, professor W. R. van Zwet.

PIET GROENEBOOM

PIET GROENEBOOM is emeritus hoogleraar Statistiek aan de TU Delft.
E-mail: <p.groeneboom@tudelft.nl>



GERRIT STEMERDINK BENOEMD TOT ERELID VAN DE VVS+OR

Tijdens de Statistische Dag op 26 maart 2015 heeft de Algemene Ledenvergadering unaniem het voorstel van het bestuur van de VVS+OR aangenomen om Gerrit Stemerding te benoemen tot erelid van de Vereniging voor Statistiek en Operationele Research.

Spil binnen de Sectie Statistische Programmatuur

Gerrit Stemerding is geboren op 3 augustus 1941. Hij begon aan het begin van de 70'er jaren als actief lid van de VVS in de contactgroep Statistiek en Computer, die later uitgroeide tot de Sectie Statistische Programmatuur. Al snel nam Gerrit een belangrijke rol in de organisatie van de vereniging door vele jaren plaats te nemen in het bestuur van de Sectie Statistische Programmatuur en de Coördinatiecommissie van de vereniging. Gerrit is tot en met 2008 op zijn minst 25 jaren een spil geweest binnen de Sectie Statistische Programmatuur. Verder is hij niet alleen een trouwe bezoeker van de Statistische Dag, maar was hij betrokken bij de organisatie van vrijwel alle tweejaarlijkse Symposia Statistische Software (SSS).

De man van STATOR

De meeste VVS+OR leden kennen Gerrit Stemerding nu vooral als de man van STATOR. Vanaf het derde nummer van de eerste jaargang staat Gerrit genoteerd als redactielid. Oftewel hij begint nu al aan zijn 16e jaar in de redactie. Zijn oog voor detail binnen deze redactie is niet ongemerkt gebleven, al snel werd Gerrit de eindredacteur van STATOR.

Elected Member van het International Statistical Institute

De eerste 25 jaar van zijn loopbaan was Gerrit Stemerding werkzaam bij universiteiten en andere onderzoeksinstituten als statisticus met statistische software als specialisme. Het was toen de tijd van de vroege pensionering en zo ging hij in 1994 als 53-jarige, niet geheel vrijwillig, met pensioen als gevolg van een reorganisatie. Hij is er niet de man naar om de extra vrije tijd niet nut-

tig te besteden en een mogelijkheid hiervoor diende zich meteen aan. Zoltan Kenessey de directeur van het International Statistical Institute (ISI) had bij zijn komst naar Nederland met verbazing het grote aantal vroeg gepensioneerd gade geslagen en het ISI Senior Executive Corps opgericht waar vrijwilligers met kennis van statistiek zich bezig konden houden met nuttige zaken voor ontwikkeling van de statistiek. Gerrit Stemerding werd ook lid van dit Corps waar hij zich aanvankelijk bezig hield met computer problemen in het ISI-kantoor. Daarnaast werkte hij mee het opbouwen van de ISI Glossary of Statistical Terms. Dit project, gebaseerd op de Engelstalige uitgave *A Dictionary of Statistical Terms* van Kendall en Buckland (1957), is thans een elektronisch woordenboek waarin 3.500 statistische begrippen in 31 talen worden weergegeven. Het is van groot belang voor de verbreiding van de statistiek in alle delen van de wereld. Gerrit Stemerding is nu ruim 20 jaar lid van het ISI Senior Executive Corps en inmiddels Elected Member van ISI. Naast allerlei organisatorisch en redactioneel werk heeft hij zich ook bezig gehouden met de organisatie van een ISI-workshop voor statistici uit ontwikkelingslanden. En op de World Statistics Congresses is hij altijd te vinden in de ISI booth waar hij voorlichting geeft over het werk van het ISI. Hij speelt hier een rol als ambassadeur van de statistiek in de gehele wereld.

Ambassadeur van de statistiek

Hij speelde de rol als ambassadeur van de statistiek ook op kleinere schaal als hij aan de studievoorzichtingsavonden op de middelbare school van zijn dochters meewerkt om leerlingen uit te leggen wat een statisticus wel niet allemaal doet. Wij menen dat een erelidmaatschap van Gerrit Stemerding een passend blijk van waardering van de zijde van de vereniging is voor een lange loopbaan gewijd aan het uitdragen van statistiek binnen de vereniging en ook ver buiten de vereniging.

JACQUELINE MEULMAN, VOORZITTER VVS+OR



SPINOZAPREMIE VOOR AAD VAN DER VAART

Voor zijn baanbrekend onderzoek in de statistiek ontvangt hoogleraar Stochastiek Aad van der Vaart de NWO-Spinozapremie. Hij ontvangt de prijs voor zijn fundamenteel onderzoek naar modellen die bijvoorbeeld helpen bij het opsporen van genen die een rol spelen bij kanker.

Foto: NWO/Ivar Pel

Vier Nederlandse wetenschappers hebben de Spinozapremie gewonnen. De premie staat ook wel bekend als de Nederlandse Nobelprijs en is de hoogste wetenschappelijke prijs van Nederland: 2,5 miljoen euro. Een van de winnaars, de Leidse hoogleraar statistiek Aad van der Vaart. Aad van der Vaart was van 2003 tot 2007 voorzitter van de VVS+OR.

Van der Vaart (1959) verdient deze prestigieuze wetenschapsprijs omdat hij met zijn onderzoek naar wiskundige statistiek tot de internationale top van zijn vakgebied behoort, meldt NWO in het laudatio. Zijn boeken over schattingstheorieën zijn invloedrijke werken. Met zijn toegepaste statistiek levert Van der Vaart waardevolle bijdragen aan onder andere medische beeldvorming en statistische genetica. Vijftien jaar geleden zette hij in Nederland de Bayesiaanse niet-parametrische statistiek op de kaart. Mede dankzij Van der Vaarts verdiensten is deze richting uitgegroeid tot een toonaangevend onderzoeksonderwerp, aldus het laudatio. Op maandag

14 september 2015 vindt de feestelijke uitreiking van de NWO-Spinozapremies plaats in de Nieuwe Kerk in Den Haag.

Wat is uw reactie op deze prijs?

'Het is natuurlijk mooi om deze prijs te krijgen. Ik zie het als een carrièreprijs voor de afgelopen vijftientig jaar, niet voor één specifiek onderdeel. Maar een van de belangrijkste gebieden is inderdaad mijn onderzoek naar Bayesiaanse statistiek in oneindig veel dimensies. De prijs zal ertoe bijdragen om de statistiek in Nederland een steviger plaats te geven.'

Wat houdt de Bayesiaanse statistiek in?

'De Bayesiaanse statistiek is een van de twee paradigma's van de statistiek. Het uitgangspunt is een zogenaamde a-priori verdeling die in kansen uitdrukt wat je al weet over een bepaalde situatie. Bijvoorbeeld dat maar een klein aantal genen actief zijn, en misschien de genen

uit een bepaald netwerk eerder dan anderen. Nieuwe informatie en metingen verwerk je door deze a-priori verdeling bij te stellen. Dat lijkt heel logisch, maar nieuw is dat we dit principe, de regel van Bayes, ook toepassen bij hele complexe modellen en met enorm veel data. De a-priori verdeling is dan een complex object. Dat geeft de mogelijkheid om veel relevante informatie aan de analyse toe te voegen, maar dit zou er toe kunnen leiden dat het eindresultaat van de analyse wordt bepaald door de aannamen en niet de data. Dat is het traditionele bezwaar tegen de Bayesiaanse methode'

Wat is uw bijdrage geweest aan de vernieuwing binnen dit vakgebied?

'Voor het jaar 2000 dacht men dat die negatieve kant overheerst. Samen met anderen heb ik laten zien dat dit niet het geval is, mits de methoden op de juiste manier worden ingezet. Onze wiskundige analyses leggen bloot hoe de eigenschappen van de a-priori verdelingen doorwerken in het eindresultaat.'

Op welke manieren past u deze vorm van statistiek toe?

'Een belangrijke toepassing is in de moderne genetica. Medisch onderzoekers zijn nu bijvoorbeeld in staat om een volledig genoom af te lezen, of van alle genen te meten of zij actief zijn bij een bepaalde vorm van kanker. De analyse van zulke grote datasets vereist nieuwe statistische technieken om te achterhalen hoe al die genen samenwerken in netwerken. De Bayesiaanse denkwijze helpt hierbij, bijvoorbeeld door als a-priori informatie toe te voegen dat van alle genen in een bepaalde situatie er slechts een klein aantal belangrijk is. Ik heb daarover veel geleerd tijdens mijn samenwerking met medici aan de Vrije Universiteit, waar ik tot drie jaar geleden werkte.'

Beeldanalyse PET-scans

'Een ander voorbeeld is de beeldanalyse van PET-scans waarmee bijvoorbeeld gekeken kan worden welke delen van het brein actief zijn. Van tevoren krijgt de patiënt een kleine hoeveelheid radioactieve stof in het bloed gespoet. Met complexe algoritmes wordt de gemeten straling omgezet in een 3D-beeld dat informatie oplevert over wat er in het hoofd gebeurt. Door het radioactieve verval en afbuiging van de straling met kansen te modelleren krijg je betere reconstructies en kun je de verschillen tussen proefpersonen makkelijker ontdekken.'

Fundamentele en toegepaste wiskunde

'Ik wil benadrukken dat ik me zeker niet alleen met toegepaste wiskunde bezig houd. De kern is puur wiskun-

Wie is Aad van der Vaart?

- 1959 wordt op 12 juli geboren in Vlaardingen
- 1983 studeert cum laude af in de wiskunde aan de Universiteit Leiden
- 1987 promoveert in Leiden op statistische schattingen in hoogdimensionale ruimtes
- 1987 docent aan de Vrije Universiteit Amsterdam
- 1988 wint de C.J. Kok-prijs voor uitzonderlijke begaafdheid op het gebied van onderzoek
- 1990 docent aan de Texas A & M University (VS)
- 1996 wordt hoogleraar stochastiek aan de Vrije Universiteit Amsterdam
- 1995 hoogleraar aan de Université Paris XI (Frankrijk)
- 2000 Miller Fellow aan de University of California in Berkeley (VS)
- 2000 wint de 5-jaarlijkse Van Dantzigprijs voor de beste Nederlandse onderzoeker onder de 40 op het gebied van de statistiek en operations research
- 2003–2007 President Netherlands Society for Statistics and Operations Research
- 2003–2011 gastonderzoeker aan Harvard School of Public Health (VS)
- 2009 wordt lid van de KNAW
- 2012 wordt hoogleraar stochastiek aan de Universiteit Leiden
- 2013 krijgt een Advanced Grant van € 2,5 miljoen van de EU voor grensverleggend onderzoek

dig. Ik formuleer noodzakelijke begrippen en ontwikkel theorieën. Die kunnen worden toegepast op allerlei situaties. Maar het doel is een theorie te formuleren die universeel is.'

Wat gaat u doen met de premie van 2,5 miljoen euro?

'Daar moet ik nog over nadenken. Misschien ga ik meer inzetten op medische toepassingen. Een andere mogelijkheid is sterrenkunde. Hier in Leiden hebben we fantastische sterrenkundigen die met enorm grote datasets werken. Ze meten zoveel dat ze het niet eens allemaal kunnen opslaan. Met statistische modellen kun je informatie trekken uit die ingewikkelde data. Met het geld zal ik in elk geval een aantal nieuwe mensen aanstellen. In de wiskunde zijn de onderzoekskosten voornamelijk personeelskosten. Misschien dat zij zich ook op de Bayesiaanse statistiek zullen richten. Maar ik wil niet alles inperken van tevoren. Ik wil graag goede mensen aantrekken met hele goede ideeën die ik zelf nog niet bedacht had.'

Het interview verscheen eerder op de website van de Faculty of science van de Universiteit Leiden en is met toestemming overgenomen. Bron: <<http://iturl.nl/sn86S>>