

*Casper Albers stapte over van de wiskundige naar de toegepaste statistiek omdat de waarheid, schoonheid en goedheid binnen de toegepaste statistiek zo veel moeilijker te bereiken zijn, betoogt hij in deze bijdrage. De toepassing in de psychologie is een extra uitdaging, aldus Albers, waarvoor hij graag zijn wiskundige safe zone verlaat. ‘Daarmee kan ik een bijdrage leveren aan het oplossen van grote vraagstukken waar de maatschappij voor staat.’*

VAN ONTWIKKELING TOT COMMUNICATIE

# TOEGEPASTE STATISTIEK

Mijn promotor Willem Schaafsma heeft mij met grote regelmaat meegegeven dat de wetenschapper zich moet laten leiden door drie platonische principes, te weten waarheid, schoonheid en goedheid. Met die waarheid, schoonheid en goedheid zit het wel goed binnen de wiskunde en de wiskundige statistiek. Dat wiskundige bewijzen 'waar' zijn, staat buiten kijf. Dat de wiskunde ons veel goeds heeft gebracht, moge ook duidelijk zijn. De schoonheid van de wiskunde en wiskundige statistiek is mijns inziens ook duidelijk.

Neem als voorbeeld de welbekende normale verdeling. Carl Friedrich Gauss bestudeerde astronomische data en stelde vier criteria op waaraan de verdeling van meetfouten van die data moest voldoen (Albers, 2018a):

- Kleine meetfouten zijn waarschijnlijker dan grote meetfouten;
- De verdeling van meetfouten is symmetrisch;
- De verdelingscurve moet een vloeiende vorm hebben;
- Als je een bepaalde kwantiteit meermalen gemeten hebt, is het gemiddelde van de metingen de meest voor de hand liggende schatter.

Het elegante van de normale verdeling is dat deze vier ogenschijnlijk eenvoudige criteria compleet voldoende zijn om de vorm van de verdeling exact te bepalen. Dus ook de schoonheid is aanwezig.

Maar dan de toegepaste statistiek. Waarheid, schoonheid en goedheid lijken overboord gekieperd te worden. Je weet als toegepast statisticus immers nooit zeker of je daadwerkelijk het juiste antwoord geeft. Je werkt doorgaans met aannames waarvan je weet dat ze eigenlijk niet waar zijn en helaas zijn *questionable research practices* aan de orde van de dag. Dit zijn onderzoekspraktijken waarin een loopje genomen wordt met statistische technieken, met het doel de resultaten mooier weer te geven dan ze zijn. Hierover later meer.

De reden waarom ik de overstap van de wiskundige naar de toegepaste statistiek gemaakt heb, is dat waarheid, schoonheid en goedheid binnen de toegepaste statistiek zo veel moeilijker te bereiken zijn. Dit maakt het naar mijn mening juist een uitdagender vakgebied.

De toepassing in de psychologie is een extra uitdaging. Voor mijn huidige baan werkte ik bij de Open University in Engeland aan modellen om de verkeersstromen op Britse snelwegen te voorspellen (Queen & Albers, 2009; Anacleto, Queen & Albers, 2013). Per rijstrook wisten we per minuut van de dag, en dat maandenlang, hoeveel voertuigen er overheen reden. Tellen hoeveel auto's ergens rijden is simpel. Zo simpel dat we computers het werk kunnen laten doen. Dat is binnen de psychologie een stuk uitdagender. Emoties, attitudes en zelfs gedrag laten zich een stuk lastiger langs de meetlat leggen. En net zoals een automobilist mogelijk de

snelheid aanpast als er een verkeerscamera op de auto gericht is, past een mens zijn gedragingen aan als hij of zij gemeten wordt (het bekende Hawthorne-effect). Die extra complexiteit is tevens een extra uitdaging.

Een tweede reden waarom ik de statistiek graag toepas, is dat ik daarmee een bijdrage kan leveren aan het oplossen van grote vraagstukken waar de maatschappij voor staat. Daar kom ik graag mijn wiskundige *safe zone* voor uit. Ik zal twee toepassingsgebieden toelichten.

## Het begrijpen van de dynamiek van een psychologisch proces is een essentiële voorwaarde om het proces zelf te begrijpen

### STATISTIEK TOEGEPAST IN DE PSYCHOPATHOLOGIE

Ik begin met de psychopathologie en ontwikkelingspsychologie. Een op de vijf volwassenen krijgt psychische problemen. Depressie en angststoornissen zijn hiervan de meest voorkomende problemen (Trimbos Instituut, 2016) en suïcide is doodsoorzaak nummer 1 onder Nederlandse jongeren. Empirisch onderzoek op dit gebied is enorm belangrijk maar ook enorm ingewikkeld. Er is namelijk een ingewikkeld samenspel van veel variabelen. Zo is het om goed te voorspellen hoe ernstig iemands winterdepressie is onvoldoende om alleen te kijken naar de hoeveelheid zonlicht. Een combinatie van maar liefst negen meteorologische variabelen is nodig om een goed beeld te krijgen van de situatie (Sarran et al., 2017).

Veel variabelen tegelijk meten is voor dit soort modellen niet eens voldoende. Je moet ze ook nog eens met grote regelmaat meten. Πάντα ρεῖ (panta rhei), een uitspraak toegeschreven aan de Griekse filosoof Heraclitus, betekent 'alles stroomt'. Alles verandert continu, zo ook de mens. Om de processen die ten grondslag aan menselijk gedrag liggen goed te begrijpen, is het dus van vitaal belang om regelmatig dezelfde variabelen bij dezelfde personen te meten. Dit is nodig om inzicht te verkrijgen in de gecompliceerde aspecten van menselijk gedrag. Die complexiteit uit zich

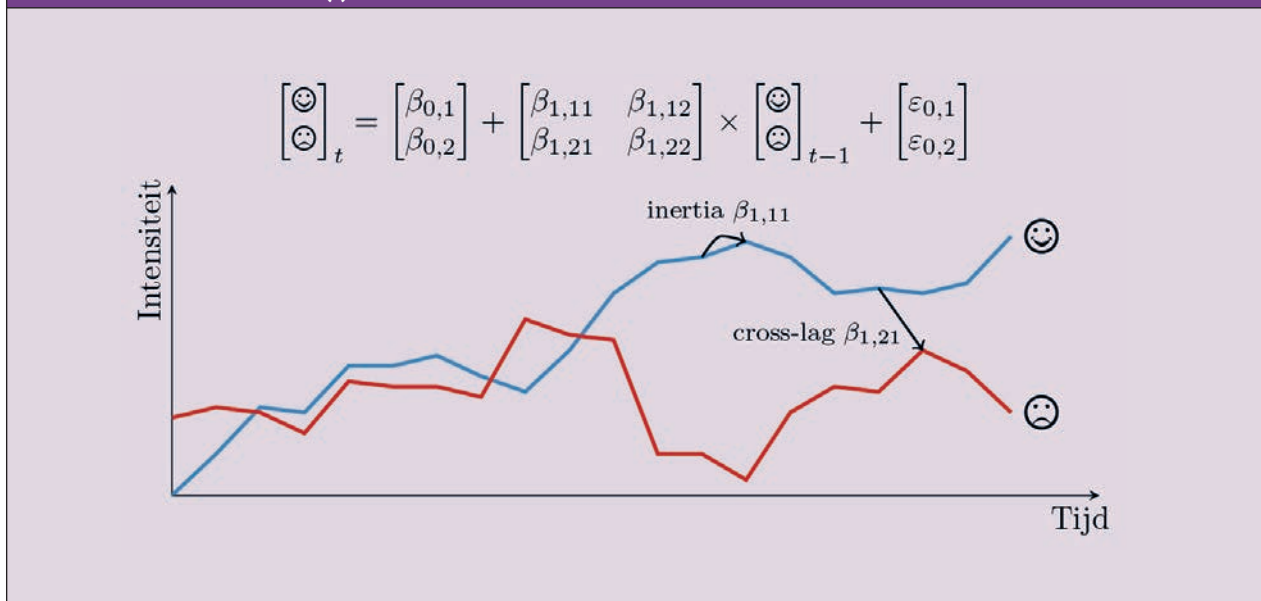
door fluctuaties in gedrag en emotionele gesteldheid over de tijd. Deze fluctuaties hangen af van de context, van interindividuele verschillen, en van toevallige verstoringen. Het begrijpen van de dynamiek van een psychologisch proces is een essentiële voorwaarde om het proces zelf te begrijpen.

Tot voor kort was het om praktische redenen onmogelijk of, op z'n minst, enorm moeilijk om bijvoorbeeld bij depressie-onderzoek met voldoende regelmaat de mentale toestand van proefpersonen te meten. Dankzij technologische vooruitgang is dit nu wel mogelijk. Sinds een jaar of tien worden *experience sampling method* en *ecological momentary assessment* steeds meer gebruikt. Via deze methodes vullen deelnemers aan een onderzoek meerdere keren per dag via hun smartphone een korte vragenlijst in. Er wordt bijvoorbeeld vijf keer per dag gevraagd om een viertal emoties te kwantificeren. Op die manier worden intensieve longitudinale data verkregen. Vervolgens wordt via een regressiemodel bekeken hoe de emotie op tijdstip  $t$  samenhangt met die op het vorige tijdstip,  $t - 1$ . Bij een enkele emotie verkrijgt je het zogenaamde AR(1)-model en bij het modelleren van meerdere variabelen tegelijk het VAR(1)-model (zie Figuur 1).

Voor gebruik in de praktijk zijn (V)AR modellen doorgaans te elementair (Krone, Albers & Timmerman, 2015; 2016) en wordt er gebruik gemaakt van ingewikkeldere modellen, maar het AR- en VAR-model liggen wel aan de basis van deze methoden. Deze modellen voor intensieve longitudinale data laten zien dat niet alleen de vraag hoe iemand zich gemiddeld gesproken voelt van belang is om iets over iemands welzijn te kunnen zeggen; de temporele dynamiek geeft ons inzichten die we niet uit de gemiddeldes kunnen halen (Cramer et al., 2016; Hamaker et al., 2015; Koval & Kuppens, 2012). Tevens geeft het bestuderen van trajecten over de tijd de onderzoeker de mogelijkheid om bijvoorbeeld te voorspellen of en wanneer een depressieve patiënt een terugval kan verwachten (Van de Leemput et al., 2014; Slofstra et al., 2018), zodat de therapeut proactief kan optreden.

De kracht van deze modellen zit hem in de mogelijkheid om tegelijkertijd interindividuele en intra-individuele uitspraken te doen, oftewel zowel iets te zeggen over de vergelijking tussen personen als de dynamiek van een individu. Als het aantal deelnemers aan een studie te groot wordt, wordt de hoeveelheid individuele modellen onoverzichtelijk en biedt het meerwaarde om personen met gelijksoortige dynamiek te clusteren (Krone et al., 2018; Ernst et al., 2019).

FIGUUR 1. HET BIVARIATE VAR(1) MODEL



De meeste tijdsreeksmodellen maken een onrealistische aanname van stationariteit, wat impliceert dat de zogenaamde emotionele inertia niet kan wijzigen over de tijd. Zeker wanneer de metingen tot doel hebben een patiënt te volgen die onder behandeling staat, is dit een groot probleem. Immers, het hele doel van de behandeling is om de interne dynamiek van de patiënt te wijzigen. Die dynamiek kan zowel geleidelijk (Bringmann et al., 2017) als abrupt (Hamilton, 1989) wijzigen. Een nieuw model (Albers & Bringmann, 2019) dat beide types verandering tegelijk aan kan, dient de komende jaren uitgebreid te worden om ook multivariate veranderingen aan te kunnen, alsmede om bruikbaar te zijn in de context van psychologische netwerken en clusteringsmethoden.

De afgelopen jaren zijn de modellen voor intensieve longitudinale psychologische data als paddenstoelen uit de grond geschoten. Doorgaans zijn deze modellen gebaseerd op enkele verstandige aannames waarna via uitgebreide computersimulaties is aangetoond dat die modellen, binnen een bepaalde context, goed werken. Dat is uiterst nuttig. Maar uiteindelijk gaat het erom dat de modellen in de praktijk ook bruikbare antwoorden opleveren. Benchmarking-onderzoeken zijn een standaardmethode binnen onder meer de informatica, maar nog niet binnen de psychologie. In een recente studie (Bastiaansen, et al., 2019) analyseerden twaalf geselecteerde teams, bestaande uit zowel methodolo-

gische als klinische experts, elk exact dezelfde dataset. Momenteel liggen de antwoorden van de verschillende aanpakken nog zodanig ver uit elkaar dat ze slechts van beperkt praktisch nut zijn. Tot nu toe wint de statistische predictie het dus niet van de klinische predictie (Meehl, 1954) en het is nog maar de vraag of dit in de toekomst anders kan zijn.

Het is ook de vraag of dit überhaupt wenselijk is. Mijns inziens kunnen dit soort statistische modellen het beste gebruikt worden om het klinische oordeel te verbeteren en niet om deze te vervangen. Dit verbeterde oordeel zal leiden tot verbeterde klinische diagnostiek en dus tot verbeterde behandeling van de meest kwetsbare personen in de samenleving.

#### STATISTIEK TOEGEPAST IN DE OMGEVINGSPSYCHOLOGIE

Een ander toepassingsgebied is de energietransitie en dan met name de omgevingspsychologische kant ervan. Rapport na rapport (IPCC, 2018) laat zien dat er enorme veranderingen moeten plaatsvinden, willen we de opwarming van de aarde beperken tot 1,5°C. Met alleen technologische innovaties en extra windmolens redden we het niet; een gedragsverandering is nodig. Voordat op gedragsverandering aangestuurd kan worden, is het van belang dat gedrag te begrijpen. Mijn interesse gaat met name uit naar het gedrag

van huishoudens rond hun energieverbruik en het vinden van de determinanten van opvattingen rond klimaatverandering.

Hierbij spelen meerdere sociaal-psychologische en economische theorieën een rol (Van der Linden, 2017; Poortinga et al., 2019; van der Werff & Steg, 2015). Elke theorie afzonderlijk bestudeert een onderdeel van het geheel, bijvoorbeeld hoe attitudes en zelfeffectiviteit invloed hebben op intenties en hoe dit weer invloed heeft op gedrag. Theorieën kunnen afzonderlijk worden bestudeerd, in een energiecontext, maar het is veel interessanter ze gelijktijdig te bestuderen (zie figuur 2). Dan kan er zowel confirmatief naar de reeds bestaande getheoretiseerde verbanden worden gekeken als exploratief naar verbanden die individuele theorieën overstijgen. Zo'n geïntegreerde aanpak leidt dus tot inzichten die zonder deze grootschalige aanpak niet verkregen kunnen worden.

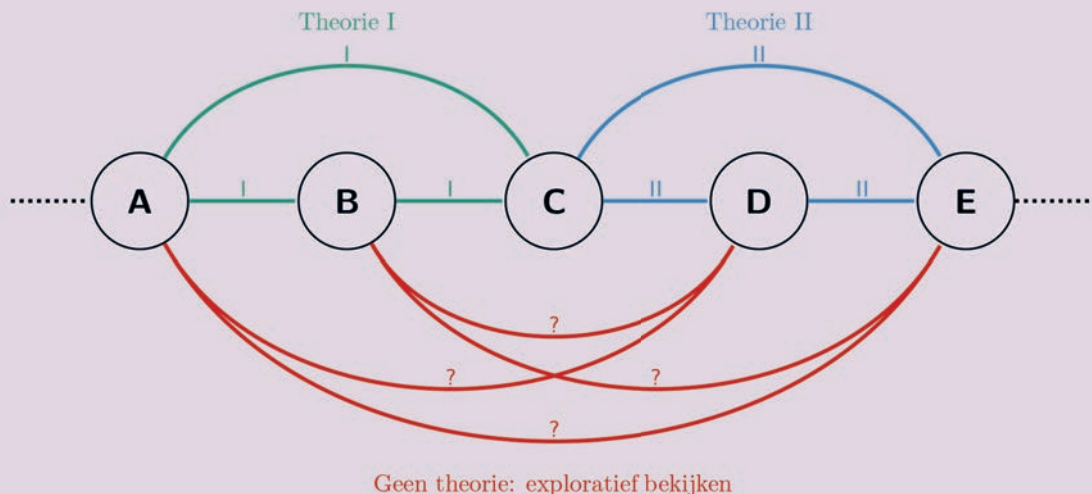
Uiteindelijk is het doel om de psychologische variabelen te koppelen aan het energieverbruik. Omdat energieverbruik vanzelfsprekend ook afhangt van tal van andere variabelen, zoals bouwtechnische aspecten van de woning, sociaaleconomische variabelen en het weer (Namazkhan, Albers & Steg, 2019) is dit geen gemakkelijke klus. Het is als het op elkaar

proberen te zetten van twee LEGO-blokjes van verschillend formaat en hiertoe dienen op maat gemaakte statistische modellen ontwikkeld te worden. Omdat met deze aanpak het aantal gemeten variabelen snel stijgt, raakt het model snel onoverzichtelijk. Het psychologische netwerkmodel (Borsboom & Cramer, 2013), tot nu toe met name binnen de psychopathologie gebruikt, biedt een visuele uitkomst om door de bomen toch een bos te zien.

Eerste analyses zijn veelbelovend. Zowel de analyse van een grootschalig Nederlands buurtinitiatief (Bhushan et al., 2019) als die van Ronde 8 van de *European Social Survey* (Verschoor et al., 2019) laten zien dat dergelijk netwerkmodellen van toegevoegde waarde zijn binnen de omgevingspsychologie. Verbanden tussen variabelen zijn in de meeste Europese landen zeer overeenkomstig en dit impliceert dat beleidswijzigingen in alle landen een vergelijkbaar effect zullen hebben (Bouman et al., 2019).

De vervolgstap in dit onderzoek is om de dynamiek binnen huishoudens te onderzoeken. Wie zijn de *agents of change*: wie en wat binnen een huishouden kan er voor zorgen dat het energiegedrag groener wordt? Momenteel verzamelen we data in verschillende Europese landen bij verschillende typen huishoudens om antwoord op deze vragen te krijgen.

FIGUUR 2. SCHEMATISCH OVERZICHT VAN DE KOPPELING TUSSEN MEERDERE THEORIEËN. THEORIE I POSTULEERT DE ONDERLINGE VERBANDEN TUSSEN A, B EN C; THEORIE II DIE TUSSEN C, D EN E. OVER DE IN ROOD WEERGEGEVEN VERBANDEN, BIJVOORBEELD DIE TUSSEN A EN D, IS (NOG) GEEN THEORIE.



## COMMUNICATIE

Het ontwikkelen van een statistisch model voor een toepassing is maar het halve verhaal. De andere helft is er voor zorgen dat het model en haar uitkomsten door de belanghebbenden goed begrepen worden. Deze belanghebbenden zijn doorgaans zelf geen statisticus, maar sociaal-wetenschapper, beleidsmaker, therapeut of patiënt. Zonder goede statistische communicatie geen toegepaste statistiek. Juist in deze tijd van *fake news* en *framing* (Albers, 2019b) kan een goede communicatie van wetenschappelijke resultaten niet meer gezien worden als een fijne bonus, maar dient het een essentieel ingrediënt van de wetenschappelijke methode te zijn.

En statistiek is niet altijd even makkelijk uit te leggen. Nobellaureaat Daniel Kahneman onderscheidt twee denksystemen (Kahneman, 2011): *thinking fast*, waarbij moeiteloos en intuïtief wordt gedacht, en *thinking slow*, waarbij analytisch wordt gedacht. Bij statistische vraagstukken is uit de context niet altijd duidelijk dat er analytisch gedacht moet worden, waardoor het brein in de verkeerde modus staat. En dan is statistiek inderdaad moeilijk.

Deze moeite met statistiek, die zelfs door hoogopgeleide wetenschappers wordt ervaren, is een van de hoofdoorzaken van de replicatiecrisis (Open Science Foundation, 2015) die sinds een aantal jaar de psychologie, en ook de andere empirische wetenschappen, in haar greep heeft. Kort samengevat behelst de crisis dat onderzoekers vaak methodologisch onverantwoorde stappen zetten in de wetenschapscyclus, waardoor de gevonden resultaten vertekend en daarmee onjuist zijn. Deze onverantwoorde stappen, of *questionable research practices* (John, Loewenstein & Prelec, 2012), worden in uitzonderlijke gevallen – gevallen van fraude – opzettelijk gemaakt, maar doorgaans komen zij voort uit een onvoldoende methodologisch inzicht van de wetenschapper.

Hoewel methodologen als Paul Meehl (Meehl, 1978, 1990), Adriaan de Groot (De Groot, 1961) en Jacob Cohen (Cohen, 1962; 1994) een halve eeuw geleden al waarschuwden voor weeffouten in het wetenschappelijk proces, lijkt er het afgelopen decennium pas echt momentum te zijn om de crisis aan te pakken. Het is de taak van de statisticus om hierbij een actieve rol te spelen. Dit kan op een meerdere manieren, waaronder:

- Het onder de aandacht brengen van minder bekende methodologische kennis (Ernst & Albers, 2017; Lakens et al., 2018; Albers, 2019c);
- Het wijzen op methodologische fouten in gepubliceerde high-impact artikelen (Brown, Albers & Ritchie, 2017; Albers, 2015);

## Tot nu toe wint statistische predictie het niet van klinische predictie en het is de vraag of dit in de toekomst anders kan zijn

- Het actief meewerken aan de introductie van technieken die de weeffouten in de wetenschap kunnen verkleinen, zoals pre-registratie (Williams & Albers, 2019) en open science (Universiteitsbibliotheek Groningen, 2018);
- Het ontwikkelen en aanbieden van open access methodologisch onderwijs aan wetenschappers en het algemeen publiek.

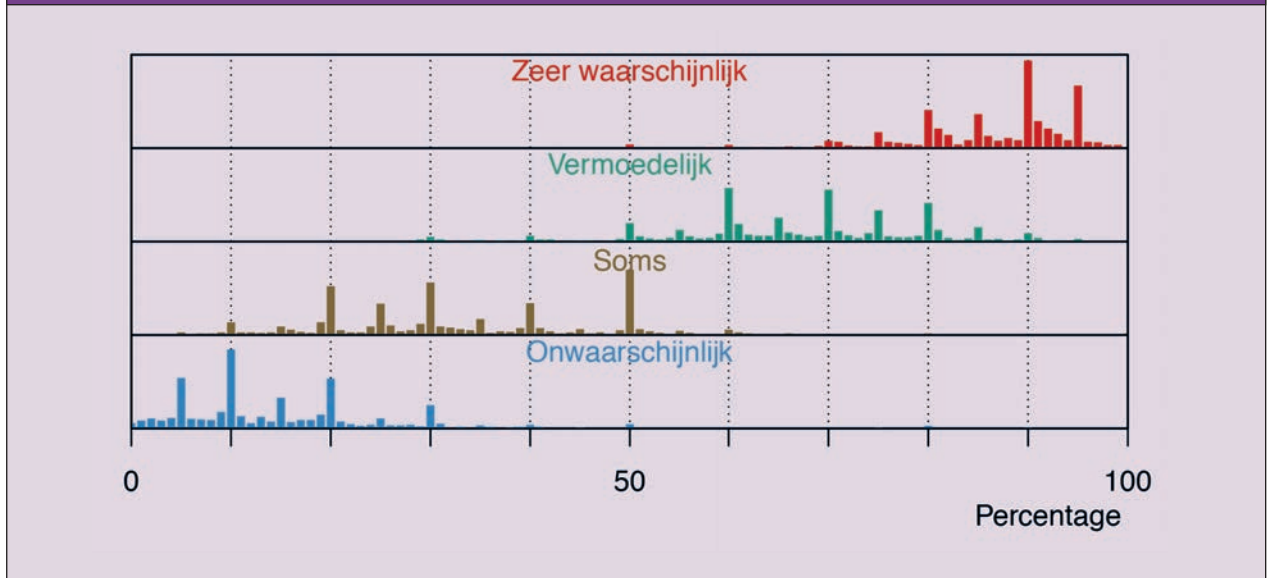
Deze taken leiden niet direct tot nieuwe wetenschappelijke kennis en zouden door de purist dus niet als taak van de onderzoeker bestempeld worden. De wetenschapper heeft echter ook een onderwijstaak en die dient zich zeker niet tot alleen het klaslokaal te beperken.

Terug naar de moeilijke statistiek. Er is veel onderzoek gedaan naar welke statistische vraagstukken mensen moeilijk vinden. Het driedeurenprobleem is een overbekend voorbeeld.<sup>1</sup> Het blijkt dat zelfs duiven beter te trainen zijn in het oplossen van dit probleem dan mensen (Albers, 2019a; Herbranson & Schroeder, 2010). Er is echter nog verrassend weinig onderzoek gedaan naar *waarom* mensen dit soort vragen zo moeilijk vinden, noch naar welk type mens er extra moeite mee heeft of wat men kan doen om de moeilijkheden weg te halen. Het is bijvoorbeeld erg interessant om te kijken vanaf welke leeftijd intuïtie mensen de verkeerde kant op stuurt bij vragen als het driedeurenprobleem. Als

1 Noot van redactie: In een quiz wordt een deelnemer geconfronteerd met drie gesloten deuren. Achter een van de deuren staat een auto (of een ander waardevol voorwerp), achter de andere twee een geit (of iets anders van 'weinig' waarde). De deelnemer mag een deur aanwijzen en krijgt als prijs datgene wat zich achter die deur bevindt. Als de deelnemer een deur heeft aangewezen, opent de presentator een van de andere deuren waarachter een geit staat. De presentator geeft de deelnemer daarna de mogelijkheid om te wisselen van gesloten deur, dus om in plaats van de eerst gekozen deur te kiezen voor de andere nog gesloten deur. Wat moet de deelnemer doen? Kan hij beter wisselen van deur, of maakt het niets uit? Is de kans op het winnen van de auto groter als de deelnemer van deur wisselt? Bron: <https://nl.wikipedia.org/wiki/Driedeurenprobleem>



FIGUUR 3. VERDELING VAN DE INTERPRETATIE VAN VIER KANSWOORDEN (WILLEMS, ALBERS &amp; SMEETS, 2019).



FIGUUR 4. DRIE VERSCHILLENDE VISUELE REPRESENTATIES VAN DEZELFDE VERHOUDINGEN.



duiven het beter kunnen dan mensen, omdat duiven niet gehinderd worden door menselijke intuïtie, dan kunnen kinderen het mogelijk beter dan volwassenen; maar tot welke leeftijd?

Onlangs is uitgezocht hoe het zit met de interpretatie van Nederlandstalige kanswoorden (Willems, Albers & Smeets, 2019; zie figuur 3). Net als bij soortgelijk Engelstalig onderzoek vonden wij dat de ene persoon een woord als ‘meestal’ totaal anders interpreteert dan de ander. Dit heeft vergaande gevolgen voor risicocommunicatie, bijvoorbeeld in flyers met informatie over ziekten. Wat wij ook onderzocht hebben is of statistici, die beroepsmatig met kanswoorden bezig zijn, homogener zijn in hun interpretaties van kanswoorden dan leken. Dit bleek niet het geval. Uitgebreide kanstraining zorgt dus niet voor een eenduidiger kansbegrip, althans niet

wanneer het kansbegrip omgezet wordt in woorden. Welke variabelen dan wel een rol spelen bij deze heterogeniteit is iets dat ik de komende jaren zal onderzoeken.

#### VISUALISATIE

Een belangrijk onderdeel van statistische communicatie is datavisualisatie. Het gebruik van visualisaties om kwantitatieve informatie inzichtelijk te maken, gaat eeuwen terug (Friendly, 2006). Door technologische vooruitgang, de beschikbaarheid van softwarepakketten zoals R (R Core Team, 2018) en ggplot2 (Wickham, 2016) en de mogelijkheid tot animaties en interactieve visualisaties, maar ook door de opkomst van ingewikkeldere statistische datasets en modellen, zijn datavisualisaties aan een opmars bezig in de sociaalwetenschappelijke literatuur.

## HET ACADEMISCH ONDERZOEKSKLIJMAAT

Om goed onderzoek te kunnen doen is een goede onderzoeksomgeving onontbeerlijk. Hoewel er ontzettend veel goed gaat aan deze universiteit, wil ik er toch op wijzen dat de academische omgeving, zowel voor studenten als medewerkers, beter kan en beter moet.

De focus ligt nu te veel op rendement en dat heeft een nadelig effect. Het klopt dat de karige bijdrage van de overheid aan wetenschap en hoger onderwijs niet helpt, maar dit is zeker niet de enige schuldige. Om tegenwoordig promovendus te kunnen worden, dient eigenlijk wel een research masteropleiding gevolgd te zijn. En om zo'n opleiding te kunnen volgen, is het zeker een voordeel om tijdens de bacheloropleiding deelgenomen te hebben aan een *honours college*. Om daar binnen te komen, zijn uitstekende cijfers en extra-curriculaire activiteiten gedurende de middelbare school van groot belang.

Scholieren die een academische carrière ambiëren, moeten hier dus al rond hun zestiende mee bezig zijn. Laatbloeiërs, zij-instromers en twijfelaars wordt het moeilijk gemaakt. Dit systeem sorteert erop voor dat er te veel homogeniteit is aan de universiteit, zowel qua type persoonlijkheden als achtergrond. Dit is het duidelijkst terug te zien in de vertekende genderbalans. Los van de evidente wenselijkheid van diversiteit van ideeën, is het voor het draagvlak van het hoger onderwijs en de wetenschap van belang dat de maatschappij zich vertegenwoordigd ziet binnen de academie. Hoewel er qua diversiteit aan de universiteit voortgang geboekt wordt, gaat het te traag.

Het zogenaamde Tenure Track-systeem heeft zeker goede punten maar ook minder goede kanten. De lijst met Tenure Track-eisen staat vol met targets die behaald moeten worden en dan geldt natuurlijk de wet van Goodhart (Strathern, 1997): 'Als een maatstaf een doel wordt, is het geen goede maatstaf meer.' De succeskans bij beursaanvragen van nwo is erg laag. Dit komt omdat het aantal aanvragen het afgelopen decennium flink gestegen is, terwijl de financiële middelen niet meebewogen.

Die vele extra aanvragen komen niet uit een sterk gegroeide intrinsieke motivatie maar vooral omdat men beurzen moet binnenhalen van het systeem. Het niet binnenhalen van beurzen wordt als een mislukking gezien, terwijl het gezien de succeskans juist de *default* is dat dit niet lukt (Albers, 2018b). Bijkomend probleem is dat ook dit systeem de homogenisering in de hand werkt: iedereen moet exact dezelfde targets halen en moet dus dezelfde kwaliteiten in huis hebben. De targets zijn ook nog grotendeels solo-targets, samenwerking wordt onvoldoende beloond. Dit leidt tot een toxisch competitief systeem.

Een alternatief systeem, waarbij de medewerker punten kan verdienen met bijvoorbeeld publicaties en media-optredens, en vervolgens vanaf een bepaald aantal punten op kan gaan voor bevordering, zal al helpen die homogenisering tegen te gaan. Immers, de ene academicus zal punten verzamelen omdat zij succesvolle beursvoorstellen schrijft, de ander omdat hij een brug slaat tussen academie en maatschappij.

Het Tenure Track-systeem dwingt medewerkers om continu nèt harder te bewegen dan wenselijk. De medewerker moet immers niet alleen aantonen dat hij of zij goed werk levert, het moet ook in een korte tijdsperiode aangetoond. Deze dwang zorgt niet alleen voor onnodige werkstress onder de medewerkers, maar remt ook de creativiteit af. Zoals de Jedi-meester Yoda zei: *'The greatest teacher, failure is'* (Johnson, 2017). De ruimte om fouten te kunnen maken is essentieel voor succes, zowel voor studenten als voor medewerkers. Die ruimte moet dan ook worden geboden.

Een goede werkomgeving maakt goede wetenschap mogelijk. De statistiek kan worden ingezet in tal van toepassingsgebieden om maatschappelijke problemen aan te pakken. Zoals John Tukey ooit zei: *'The best thing about being a statistician is that you get to play in everyone's backyard.'* Hij sloeg de spijker op z'n kop. Een succesvolle aanpak van maatschappelijke problemen vereist echter niet alleen goede wiskunde en goede methodologie, maar ook goede communicatie. (CA)

Om te weten of de boodschap die via een statistische grafiek wordt uitgedragen goed overkomt, moeten we weten hoe mensen grafieken interpreteren. Er zijn uitgebreide richtlijnen omtrent goede datavisualisatie (Wilkinson, 2005). Er is ook enig onderzoek gedaan naar hoe grafieken geïnterpreteerd worden (Simkin & Hastie, 1987; Cleveland & McGill, 1984; Spence, 1990). Er is echter verrassend weinig onderzoek gedaan naar waarom het op

die manier geïnterpreteerd wordt. Zo schatten mensen kansen anders in wanneer deze via horizontale staafdiagrammen worden gepresenteerd dan wanneer deze via verticale staafdiagrammen worden gepresenteerd, ook al is de informatie wiskundig gezien equivalent (zie figuur 4). We weten niet goed waarom. Er is literatuur over de psychologische aspecten van het verwerken van grafische informatie (Wagemans et al., 2012), maar hoe dit in een



## Het maken van de visualisatie is de ene helft van het verhaal, zorgen dat de visualisatie goed begrepen wordt de andere helft

statistische context in zijn werk gaat, is onvoldoende bekend.

Welke visualisatietechniek optimaal is, hangt vanzelfsprekend ook van de context af: moet snel een oppervlakkig inzicht verkregen worden, of moet het juist een boodschap zijn die goed blijft hangen? Is de boodschap vrij eenvoudig of juist gecompliceerd en genuanceerd? De keuze hangt ook af van de doelgroep die men wil bereiken, bijvoorbeeld medewetenschappers of patiënten (Smeets, 2014).

Neem het toepassingsgebied van de psychopathologie, waarin netwerkvisualisaties, en zelfs netwerkanimaties, een prominente rol innemen. Er worden technisch gezien vernuftige visualisaties gemaakt. Maar of de boodschap uit de grafiek ook goed overkomt bij de therapeut en de patiënt, is nog maar de vraag (Bos et al., 2019). Indien er structureel verkeerde interpretaties gemaakt worden, kan dit leiden tot suboptimale behandelingen of zelfs verkeerde klinische diagnoses. Kortom, het maken van de visualisatie is de ene helft van het verhaal, zorgen dat de visualisatie goed begrepen wordt de andere helft.

De komende jaren wil ik me richten op het slaan van een brug tussen het statistische model en het gebruik ervan. Dit houdt in dat ik wil bijdragen aan het verbeteren van de wetenschappelijke cultuur om *questionable research practices* tegen te gaan. Tevens wil ik me storten op de vragen rond de interpretatie en communicatie van kansen, onzekerheden en risico's. Ten slotte wil ik middels nieuw onderzoek meewerken aan een cultuurverschuiving richting het gebruik van visualisaties die bewezen effectief zijn. Dit zal eraan bijdragen dat de toegepaste statistiek van waarde blijft voor de toepassing.

### OVER DE AUTEUR

Casper Albers is hoogleraar in de Toegepaste Statistiek en Datavisualisatie aan de Rijksuniversiteit Groningen. Dit is een licht bewerkte versie van de rede die hij uitsprak op dinsdag 19 maart 2019 bij de aanvaarding van zijn leerstoel. E-mail: [cj.albers@rug.nl](mailto:cj.albers@rug.nl), [www.casperalbers.nl](http://www.casperalbers.nl).

## Literatuur

- Albers, C.J. (2015). Dutch research funding, gender bias, and Simpson's paradox. *Proceedings of the National Academy of Sciences*, 112(50), E6828–E6829.
- Albers, C.J. (2018a). De Moivre–Gauss–Laplace: extraordinarily normal. *Nieuw Archief voor Wiskunde, Vijfde serie, deel 19(1)*, 37–38.
- Albers, C.J. (2018b). Mislukking. *UKrant Groningen*, 20 november.
- Albers, C.J. (2019a). Aan intuïtie heb je weinig bij het oplossen van statistische puzzels. *De Volkskrant*, 22 januari.
- Albers, C.J. (2019b). Klimaatseptica geven antwoord op niet-relevante vragen. *De Volkskrant*, 20 februari.
- Albers, C.J. (2019c). The problem with unadjusted multiple and sequential testing. *Nature Communications*, 10: 1921..
- Albers, C.J. & Bringmann, L. F. (2019). Inspecting gradual and abrupt changes in emotion dynamics with the time-varying change point autoregressive model. Ingediend voor publicatie.
- Anacleto, O., Queen, C. M. & Albers, C. J. (2013). Multivariate forecasting of road traffic flows in the presence of heteroscedasticity and measurement errors. *Journal of the Royal Statistical Society*, 62(2), 251–270.
- Bastiaansen, J. A., Kunkels, Y. K., ..., Albers, C. J. & Bringmann, L. F. (2019). Time to get personal? The impact of researchers' choices on the selection of treatment targets using the experience sampling methodology. Ingediend voor publicatie.
- Bhushan, N., Mohnert, F., Sloot, D., Jans, L., Albers, C. J. & Steg, E. M. (2019). Using a Gaussian graphical model to explore relationships between items and variables in environmental psychology research. *Frontiers in Psychology*, 10: 1050.
- Borsboom, D. & Cramer, A. O. J. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annual review of clinical psychology*, 9, 91–121.
- Bos, F. M., Snippe, E., Bruggeman, R., Wichers, M. E. & van der Krieke, L. (2018). Insights of Patients and Clinicians on the Promise of the Experience Sampling Method for Psychiatric Care. *Psychiatric Services*, geaccepteerd voor publicatie.
- Bouman, T., Verschoor, M., Steg, E. M., Böhm, G., Fisher, S., Poortinga, W., ... Albers, C. J. (2019). When worry about climate change leads to climate action and policy support. Ingediend voor publicatie.
- Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D. & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods*, 22(3), 409–425.
- Brown, N. J. L., Albers, C. J. & Ritchie, S. J. (2017). Contesting the evidence for limited human lifespan. *Nature*, 546, E6–E7.
- Cleveland, W. S. & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387), 531–554.
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: a review. *Journal of Abnormal Social Psychology*, 65, 145–153.
- Cohen, J. (1994). The earth is round ( $p < .05$ ). *American Psychologist*, 49, 997–1003.
- Cramer, A. O. J., van Borkulo, C., Giltay, E. J., van der Maas, H. L. J., Kendler, K. S. et al. (2016). Major depression as a complex dynamical system. *PLoS ONE*, 11(12), e0167490.
- de Groot, A. D. (1961). *Methodologie*. Den Haag: Mouton & Co.

- Ernst, A.F. & Albers, C.J. (2017). Regression assumptions in clinical psychology research practice — a systematic review of common misconceptions. *PeerJ*, 5, e3323.
- Ernst, A.F., Timmerman, M.E., Jeronimus, B.F. & Albers, C.J. (2019). Insight Into Individual Differences in Emotion Dynamics With Clustering. *Assessment*, geaccepteerd voor publicatie.
- Friendly, M. (2006). A Brief History of Data Visualization. In C. Chen, W. Härdle & A. Unwin (Red.), *Handbook of Computational Statistics: Data Visualization* (Deel 3). Heidelberg: Springer-Verlag.
- Hamaker, E.L., Ceulemans, E., Grasman, R. & Tuerlinckx, F. (2015). Modeling affect dynamics: State of the art and future challenges. *Emotion Review*, 7(4), 316–322.
- Hamilton, J.D. (1989). A new approach to the economic analysis of non-stationary time series and the business cycle. *Econometrica*, 57, 357–384.
- Herbranson, W.T. & Schroeder, J. (2010). Are birds smarter than mathematicians? Pigeons (*Columba livia*) perform optimally on a version of the Monty Hall dilemma. *Journal of Comparative Psychology*, 124, 1–13.
- IPCC. (2018). Global warming of 1.5°C. Geneva, Switzerland: IPCC.
- John, L.K., Loewenstein, G. & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23, 524–532.
- Johnson, R. (regie/scenario). (2017). *Star Wars: Episode VIII – The Last Jedi*. San Francisco: Lucasfilm.
- Kahneman, D. (2011). *Thinking Fast and Slow*. New York: Farrar, Straus en Giroux.
- Koval, P. & Kuppens, P. (2012). Changing emotion dynamics: Individual differences in the effect of anticipatory social stress on emotional inertia. *Emotion*, 12(2), 256–267.
- Krone, T., Albers, C.J., Kuppens, P. & Timmerman, M.E. (2018). A multivariate statistical model for emotion dynamics. *Emotion*, 18, 739–754.
- Krone, T., Albers, C.J. & Timmerman, M.E. (2015). A comparative simulation study of AR(1) estimators in short time series. *Quality & Quantity*, 1–21.
- Krone, T., Albers, C.J. & Timmerman, M.E. (2016). Comparison of estimation procedures for multilevel AR(1) models. *Frontiers in Psychology: Quantitative Psychology and Measurement*, 7.
- Lakens, D., Adolfs, F.G., Albers, C.J., ... & Zwaan, R. (2018). Justify your alpha. *Nature Human Behaviour*, 2, 168–171.
- Meehl, P.E. (1954). *Clinical versus statistical prediction: A theoretical analysis and review of the evidence*. University of Minnesota Press.
- Meehl, P.E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald and the slow progress of soft psychology. *Journal of consulting and clinical psychology*, 46, 806–834.
- Meehl, P.E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, 66, 195–244.
- Namazkhan, M., Albers, C.J. & Steg, E.M. (2019). The role of environmental values, socio-demographics and building characteristics in setting room temperatures in winter. *Energy*, 171, 1183–1192.
- Open Science Foundation. (2015). Estimating the reproducibility of psychological science. *Science*, 349, aac4716.
- Poortinga, W., Whitmarsh, L., Steg, E.M., Böhm, G. & Fisher, S. (2019). Climate change perceptions and their individual-level determinants: a cross-European analysis. *Global environmental change*, 55, 25–35.
- Queen, C.M. & Albers, C.J. (2009). Intervention and causality: Forecasting traffic flows using a dynamic Bayesian network. *Journal of the American Statistical Association*, 104(486), 669–681.
- R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.
- Sarran, C., Albers, C.J., Sachon, P. & Meesters, Y. (2017). Meteorological analysis of symptom data for people with seasonal affective disorder. *Psychiatry Research*, 257, 501–505.
- Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perceptions and Performance*, 16, 683–692.
- Simkin, D. & Hastie, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association*, 82, 454–465.
- Slofstra, C., Nauta, M.H., Bringmann, L.F., Klein, N.S., Albers, C.J. et al. (2018). Individual negative affective trajectories can be detected during different depressive relapse prevention strategies. *Psychotherapy and Psychosomatics*, 87, 243–245.
- Smeets, I. (2014). *Het exacte verhaal. Wetenschapscommunicatie voor bèta's*. Amsterdam: Nieuwezijds.
- Strathern, M. (1997). Improving ratings: audits in the British university system. *European Review*, 5, 305–321.
- Trimbos Instituut. (2016). *Depressie. Enkele cijfers*. Factsheet, INFO16.
- Universiteitsbibliotheek Groningen. (2018). Open peer review. An interview with Casper Albers. *Open Science Newsletter*, 28 maart.
- Van de Leemput, I., Wichers, M., Cramer, A.O.J., Borsboom, D., Tuerlinckx, F. et al. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences*, 111(1), 87–92.
- Van der Linden, S.L. (2017). Determinants and measurement of climate change risk perception, worry, and concern. In: *The Oxford encyclopedia of climate change communication*. Oxford, UK: Oxford University Press.
- Van der Werff, E. & Steg, E.M. (2015). One model to predict them all: Predicting energy behaviours with the norm activation model. *Energy Research & Social Science*, 6, 8–14.
- Verschoor, M., Albers, C.J., Poortinga, W., Böhm, G. & Steg, E.M. (2019). Attributes to climate change and energy: A network analysis of the European Social Survey. Ingediend voor publicatie.
- Wagemans, J., Elder, J.H., Kubovy, M., Palmer, S.E., Peterson, M.A. et al. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological bulletin*, 138, 1172–1217.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.
- Wilkinson, L. (2005). *The Grammar of graphics* (2de ed.). New York: Springer.
- Willems, S.W., Albers, C.J. & Smeets, I. (2019). Variability in the interpretation of Dutch probability phrases – a risk for miscommunication. *arXiv preprints*, 1901.09686.
- Williams, M. & Albers, C.J. (2019). Dealing with distributional assumptions in preregistered research. *Meta-Psychology*, 3:1592.