

STAtOR

periodiek van de VVSOR jaargang 23, nummer 3-4, december 2022

Oversterfte

Smart Surveys: apps en sensoren in
survey-onderzoek

Voetbalplaatjes, wie heeft ze niet verzameld?

Statistiek en filosofie: een voortschrijdende
samenspraak?

Politieke besliskunde

Optional stopping met Bayes factors

De Hoofdige Boer

Tourist trip planning; bezoekers spreiden
met behulp van slimme algoritmes

Hyperbolen 'tussn de duumn'

Vijfenveertig jaar statistiek

Wees voorzichtig met voxpops!



STAtOR

Jaargang 23, nummer 3-4, december 2022

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

Redactie

Joaquim Gromicho (hoofdredacteur), Annelieke Baller, Joep Burger, Caroline Jagtenberg, Guus Luijben (eindredacteur), Kerry Malone, Richard Starmans, Gerrit Stermerdink (eindredacteur), Vanessa Torres van Grinsven en Sanne Willems. Vaste medewerkers: Jelke Bethlehem, John Poppelaars, Gerard Sierksma en Henk Tijms.

Kopij en reacties richten aan

Prof. dr. J.A.S. Gromicho (hoofdredacteur), Universiteit van Amsterdam Faculteit Economie en Bedrijfskunde, Sectie Operations Management | Amsterdam Business School, Plantage Muidergracht 12, 1018 TV Amsterdam, j.a.s.gromicho@uva.nl

Bestuur van de VVSOR

Voorzitter: prof. dr. Casper Albers, db@vvsor.nl; Secretaris: Pieter Jongsma MSc, secretaris@vvsor.nl; Penningmeester: Judith ter Schure PhD, penningmeester@vvsor.nl; Algemeen bestuurslid: Marianne Jonker PhD, db@vvsor.nl; Webmaster: Eugenio Traini MSc, webmaster@vvsor.nl.

Voorzitters van de secties: prof. dr. ir. Mark van de Wiel (Biometrical Section); prof. dr. Albert Wagelmans (Section for Operations Research); dr. Eduard Belitser (Section Mathematical Statistics); dr. Rebecca Kuiper (Social Sciences Section); dr. Michel van de Velden (Economics Section); dr. Daniel Oberski (Section Data Science); Luise Gummi MSc (Young Statisticians) dr. Sanne Willems (Section Statistics Communication).

Leden- en abonnementenadministratie van de VVSOR

VVSOR, Maarsbergseweg 20, 3956 KW Leersum, admin@vvsor.nl. Raadpleeg onze website www.vvsor.nl over hoe u lid kunt worden van de VVSOR of een abonnement kunt nemen op STAtOR.

Voor advertenties

M. van Hootegem, hootegem@xs4all.nl
STAtOR verschijnt in maart, juni, september en december.

Ontwerp en opmaak

Pharos, Nijmegen

Uitgever

© Vereniging voor Statistiek en Operations Research
ISSN 1567-3383

Op naar het einde

Als het goed is wordt dit nummer van STAtOR aan het eind van 2022 bezorgd. Een roerig jaar komt dan aan zijn einde, de opluchting dat we COVID-19 grotendeels onder controle kregen werd overschaduwd door de oorlog in Oekraïne. Ook in ons blad komt iets aan zijn einde, dit is het laatste nummer dat werd verzorgd door Monique van Hootegem. Aan het eind van dit dubbelnummer leest u hoe de redactie haar als de feitelijke creator van STAtOR ziet. De volgende nummers zullen anders zijn.

COVID heeft duidelijk gemaakt hoeveel impact ons vakgebied heeft, in eerdere nummers hebt u over veel verschillende toepassingen kunnen lezen. Dit nummer start met een artikel van Lenny Stoeldraijer over oversterfte, een onderwerp dat nog niet aan de orde kwam.

Veel van ons dagelijks doen en laten is bekend zonder dat we ons dat realiseren. We geven onbewust van alles prijs via sociale media en smart phones maken het mogelijk onze verplaatsingen te volgen. In 'Smart Surveys: apps en sensoren in survey-onderzoek' vertellen Barry Schouten en Annemieke Luiten hoe zo'n gegevensverzameling ook bewust tot stand kan komen, ook gaan zij in op ethische aspecten.

Statistiek en filosofie hebben een grote wederzijdse invloed, Richard Starmans gaat in op een voortschrijdende samenspraak. Wie is er niet ooit in de verleiding gekomen eerder dan gepland te stoppen met het verzamelen van gegevens omdat de reeds verzamelde data al een duidelijk beeld opleveren? Rianne de Heide legt uit uit hoe Bayes Factors kunnen worden gebruikt bij Optional Stopping en zij gaat in op de mogelijkheden en beperkingen.

Joris Slootweg, Rob van der Mei en Frank Ottenhof vertellen hoe slimme algoritmes kunnen helpen bij het spreiden van grote aantallen bezoekers. Ronald Does ten slotte vertelt over zijn veelzijdige loopbaan: Vijfenvoertig jaar statistiek.

Columns zijn er ook. Henk Tijms verzamelt voetbalplaatjes, Gerard Sierksma bekent wel eens overdreven te hebben en hoe dat zijn geloofwaardigheid over een val met zijn racefiets in een ravijn aantastte, Gerrit Stermerdink geeft toe een koppige Achterhoeker te zijn, Bernard Zweers vervangt zijn collega John Poppelaars met een verhaal over politieke besliskunde, en Jelke Bethlehem maakt duidelijk dat voxpops geen vervanging van onderzoek kunnen zijn.

Samen met wat verenigingsnieuws is dit weer een gevarieerd nummer geworden, wij wensen u veel leesplezier!



4



8



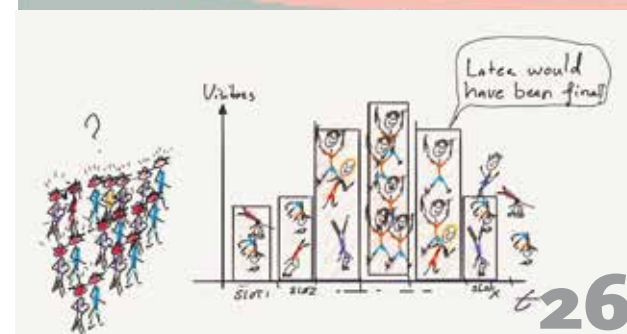
13



15



22



26



33

INHOUD

- 2 Op naar het einde
- 4 Oversterfte | LENNY STOELDRAIJER
- 8 Smart Surveys: apps en sensoren in survey-onderzoek | BARRY SCHOUTEN & ANNEMIEKE LUITEN
- 13 Voetbalplaatjes, wie heeft ze niet verzameld? – column | HENK TIJMS
- 15 Statistiek en filosofie: een voortschrijdende samenspraak? | RICHARD STARMANS
- 20 Politieke besliskunde – column | BERNHARD ZWEERS
- 21 Oproep om kandidaten te nomineren voor de Jan Hemelrijk en Willem R. van Zwet awards 2022
- 22 Optional stopping met Bayes factors; mogelijkheden en beperkingen | RIANNE DE HEIDE
- 25 De Hoofdige Boer – column | GERRIT STERMERDINK
- 26 Tourist trip planning; bezoekers spreiden met behulp van slimme algoritmes | JORIS SLOOTWEG, ROB VAN DER MEI & FRANK OTTENHOF
- 30 Hyperbolen 'tussn de duumn' – column | GERARD SIERKSMA
- 33 Vijfenvoertig jaar statistiek; Deel 2: periode 1990–2021 | RONALD DOES
- 38 Peilingpraktijken. Wees voorzichtig met voxpops! – column | JELKE BETHLEHEM
- 43 Monique van Hootegem, de feitelijke creator van STAtOR
- 44 LNMB/NGB conferentie op 16, 17 en 18 januari 2023
- 44 Young Statisticians
- 44 VVSOR Annual Meeting 2023 op 23 maart 2023



OVERSTERFTE

Sinds de start van de COVID-19-epidemie heeft Nederland een aantal perioden gekend waarin sprake was van oversterfte: het aantal overledenen lag in die perioden significant hoger dan wat verwacht mocht worden op basis van sterftcijfers in de jaren vóór de pandemie. In dit artikel wordt uitgelegd wat oversterfte inhoudt, wordt de methode van het CBS uitgelegd en worden de aantallen in 2020 en 2021 vergeleken met de sterfte aan COVID-19 (op basis van de CBS Doodsoorzakenstatistiek).

LENNY STOELDRAIJER

Een begrip dat (bijna) niemand kende, maar sinds het begin van de coronapandemie was dit bijna wekelijks in de media te volgen: 'oversterfte'. Dit betekent dat het aantal overledenen hoger is dan wat verwacht mocht worden. Volgens het Centraal Bureau voor de Statistiek (CBS) was de oversterfte in Nederland in 2020 gelijk aan 15 duizend en in 2021 bijna 16 duizend. Maar wat is oversterfte precies en waarom is dit cijfer zo belangrijk?

Wat is oversterfte?

Bij een epidemie of andere gebeurtenis is het onder andere belangrijk te weten wat de impact er van is op de sterfte. In het geval van de COVID-19-epidemie zou daarvoor het aantal overledenen aan COVID-19 zeer relevant zijn. In Nederland is er echter geen tijdig beeld van dit aantal,

omdat er geen meldplicht voor overlijden aan COVID-19 is aan het RIVM en de verwerking van de doodsoorzakenformulieren door het CBS, waarop voor iedere overledene de doodsoorzaak (waaronder COVID-19) staat vermeld, enkele maanden duurt.

Naast de sterfgevallen die direct toe te schrijven zijn aan COVID-19, is het mogelijk dat mensen overlijden als indirect gevolg van COVID-19, vanwege aandoeningen die bestonden vóór de COVID-19-epidemie. Bijvoorbeeld door overbezetting van de gezondheidszorg of zorgmijding. Ook kunnen de beperkende coronamaatregelen sterfgevallen door andere oorzaken dan COVID-19 juist hebben voorkomen, bijvoorbeeld door verminderde besmettingen door andere infectieziekten. Indirecte sterfgevallen als gevolg van COVID-19 worden niet meegenomen in de COVID-19-sterftcijfers.

Vanwege het gebrek aan tijdige aantallen en het ontbreken van indirecte sterftcijfers, wordt oversterfte beschouwd als een meer objectieve en vergelijkbare maatstaf die een tijdig beeld geeft van zowel de directe als de indirecte gevolgen van de COVID-19-epidemie.

Oversterfte is een tijdelijke, bijzondere stijging van het aantal overledenen die samenvalt met een bijzondere gebeurtenis, zoals een griep-epidemie, hittegolf of COVID-19-epidemie. De Wereldgezondheidsorganisatie (WHO) definieert oversterfte als 'The mortality above what would be expected based on the non-crisis mortality rate in the population of interest'. Ook het CBS spreekt van oversterfte wanneer het waargenomen aantal overledenen hoger is dan het verwachte aantal overledenen in dezelfde periode. Daarbij wordt uitgegaan van de sterfteontwikkelingen alsof er geen COVID-19-epidemie is geweest in Nederland.

Bepaling verwachte sterfte

Het verwachte aantal overledenen wanneer er geen COVID-19-epidemie was geweest, wordt door het CBS bepaald in twee stappen. Eerst wordt het verwachte aantal

overledenen in het hele jaar bepaald, daarna het verwachte aantal overledenen per week.

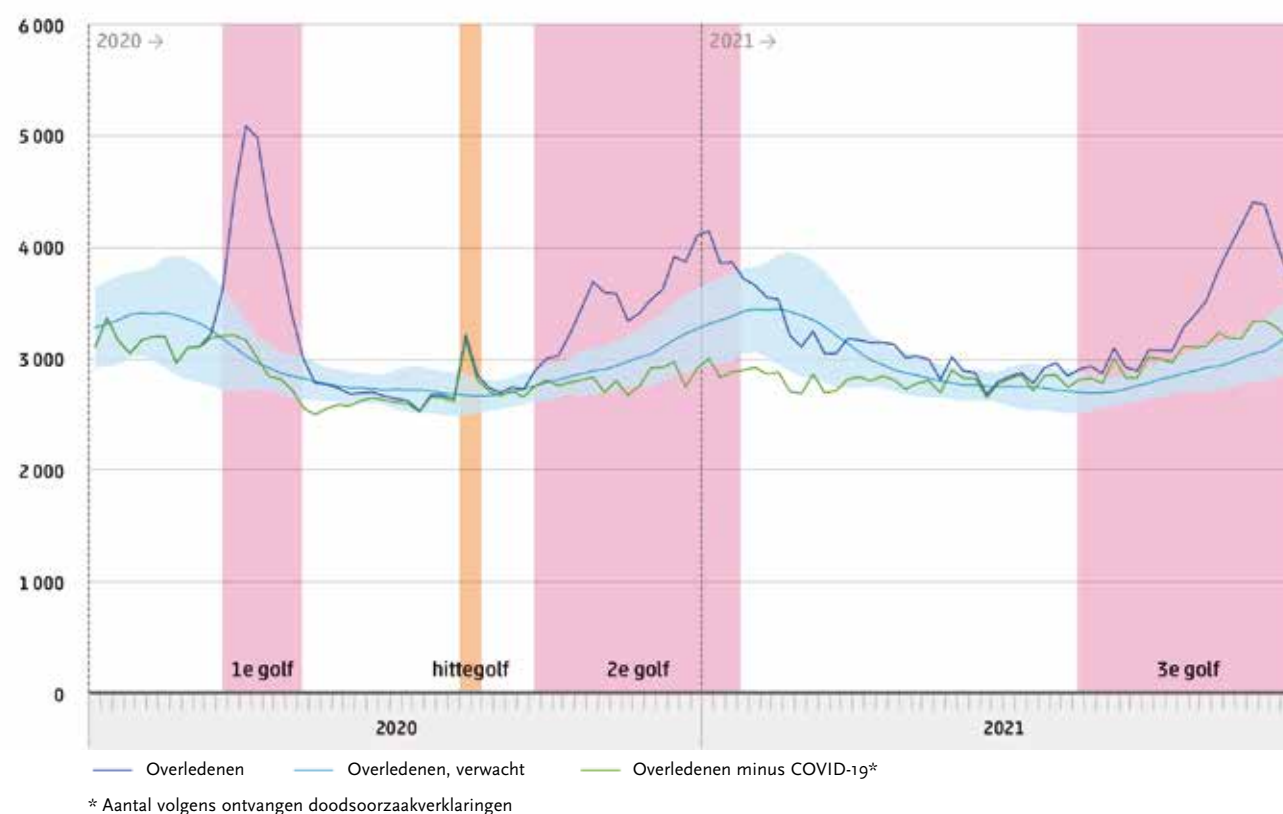
Het verwachte aantal overledenen in het hele jaar wordt bepaald op basis van de prognoses die het CBS jaarlijks maakt. Deze prognoses geven namelijk de meest waarschijnlijke toekomstige ontwikkelingen van de bevolking en de sterfte, uitgaande van de kennis die er op dat moment is. De prognoses houden rekening met het feit dat de bevolking continu verandert: er komen immigranten bij (die relatief jong zijn), we vergrijzen en we worden steeds ouder.

Het CBS gebruikt voor de prognose van de leeftijds- en geslachtsspecifieke sterftkansen een extrapolatiemodel: er wordt van uitgegaan dat de toekomstige trends een voortzetting zijn van de trends uit het verleden. In het model wordt niet alleen uitgegaan van de trends in Nederland, maar ook van de meer stabiele trends in andere West-Europese landen. Ook houdt het model rekening met het effect van rookgedrag op de sterfte. Er wordt gekeken naar trends over een lange periode, namelijk vanaf 1970.

Op basis van de geprognosticeerde leeftijds- en geslachtsspecifieke sterftkansen en de verwachte bevolkingsopbouw in dat jaar, wordt het verwachte aantal overledenen naar leeftijd en geslacht berekend voor een bepaald jaar. Voor 2020 is de verwachte sterfte 153.402 (ontleend aan de Kernprognose 2019-2060) en voor 2021 is deze 154.887 (ontleend aan de Bevolkingsprognose 2020-2070).

Het verwachte aantal overledenen per week is geschat op basis van de waargenomen sterfte in 2015 tot en met 2019. Eerst wordt voor elk jaar de sterfte per week bepaald. Vervolgens wordt per week een gemiddelde van de sterfte in die week en de zes omliggende weken bepaald. Toevallige fluctuaties krijgen op deze manier minder gewicht. Het seizoenspatroon wordt in zijn geheel gekalibreerd zodat de som van de weken optelt tot het jaartotaal uit de eerste stap.

Om er achter te komen of de hogere sterfte geen toeval is, wordt gekeken naar hoe de sterfte gewoonlijk



Figuur 1. Overledenen per week minus COVID-19 in 2020 en 2021; lichtblauw vlak: 95%-interval (voorlopige cijfers). Bron: CBS

schommelt, en wordt daarmee een boven- en een ondergrens van het verwachte sterftecijfer per week berekend. Alleen als de sterfte hoger is dan die bovengrens of lager dan de ondergrens van de verwachte sterfte spreken we van een bijzondere ontwikkeling. Het 95%-interval rond de verwachte sterfte is geschat op basis van de waargenomen spreiding in de sterfte per week in de periode 2015 tot en met 2019.

Internationaal perspectief

Er bestaat geen standaardmethode om de verwachte sterfte te schatten. De methode kan afwijken vanwege het doel, de data die ervoor beschikbaar zijn, of keuzes die gemaakt worden. Verschillende methoden kunnen tot andere bevindingen leiden voor wat betreft de oversterfte. Vergelijkbare uitkomsten met verschillende methodes versterken de bevindingen echter.

Het CBS heeft aan het begin van de COVID-19-epidemie een methode gebruikt waarbij gecorrigeerd werd voor seizoensgebonden factoren. Omdat die methode gebaseerd was op een zeer korte waarneemperiode (tien

weken voorafgaand aan de COVID-19-epidemie), is het CBS in de loop van 2020 overgestapt op een meer robuuste methode voor de langere termijn. In juli 2020 heeft het CBS in samenwerking met het Planbureau voor de Leefomgeving (PBL) een artikel gepubliceerd waarin een aantal verschillende methoden vergeleken zijn (Husby et al., 2020). De uitkomsten voor de eerste golf van de COVID-19-epidemie waren voor deze methoden vergelijkbaar.

Naast het CBS hebben andere instanties schattingen van de oversterfte gepubliceerd. In Nederland houdt het Rijksinstituut voor Volksgezondheid en Milieu (RIVM) sinds de griep-epidemie van 2009 wekelijks het aantal overledenen in de gaten met gegevens van het CBS. Het doel hiervan is de impact van een epidemie of incident in beeld te brengen. De sterftegegevens worden daarom gecorrigeerd voor griep-epidemieën en andere incidenten in het verleden.

Een Europese sterfitemonitor is EuroMoMo, gericht op het opsporen en meten van oversterfte als gevolg van seizoensgriep, pandemieën en andere bedreigingen voor de volksgezondheid. Ook Eurostat, het bureau voor de statistiek van de Europese Unie, heeft verschillende

rapporten over de oversterfte in de Europese Unie gepubliceerd met de sterftegegevens die in de loop van de pandemie meer gedetailleerd zijn verzameld (Eurostat, 2022). Verder is er een project van de non-profit organisatie Global Change Data Lab, Our World in Data, die uitgebreid over de oversterfte publiceert. Daarnaast zijn er nog verscheidene wetenschappelijke publicaties geweest waarin verschillende methoden gebruikt en naast elkaar gezet zijn (Nepomuceno, 2021).

Over het algemeen komen de schattingen van de oversterfte van andere instanties en publicaties overeen met de oversterfte die het CBS publiceert of is uit te leggen waar de verschillen door komen. De verschillende methodes vullen elkaar ook aan. Het is immers onbekend hoe de sterfte zich daadwerkelijk had ontwikkeld als COVID-19 er niet was geweest.

Oversterfte in Nederland

In de sterftecijfers zijn gedurende de COVID-19-epidemie drie golven met oversterfte te onderscheiden op basis van de methode die het CBS hanteert: de eerste golf in het voorjaar van 2020 (week 13 tot met 18 van 2020), de tweede golf die startte in het najaar van 2020 en die duurde tot het begin van 2021 (week 39 van 2020 tot en met week 3 van 2021) en de derde golf die startte rond het einde van de zomer van 2021 en rond de jaarwisseling van 2021/2022 eindigde (week 33 tot en met week 52 van 2021). In deze perioden lag het aantal overledenen hoger dan de verwachte sterfte en buiten het interval van verwachte fluctuaties in de sterfte. In het midden van augustus 2020 was ook een korte periode van oversterfte te zien, maar die viel samen met een hittegolf (week 33 en 34 van 2020). Zie ook figuur 1.

Oversterfte in vergelijking met COVID-19-sterfte

In 2020 en 2021 overleden in totaal 30 duizend meer mensen dan verwacht als er geen COVID-19-epidemie was geweest. De totale sterfte aan COVID-19 op basis van de Doodsoorzakenstatistiek, informatie die het CBS op een later moment ontvangt, in deze jaren betrof 40 duizend. De drie golven van oversterfte die zichtbaar waren in 2020 en 2021, vallen samen met de perioden waarin veel mensen aan COVID-19 overleden. Met name de stijging, de pieken en daarna de daling in de oversterfte en

de COVID-19-sterfte vinden bij iedere golf op ongeveer hetzelfde moment plaats.

Gedurende de eerste oversterftegolf van de COVID-19-epidemie kwam de oversterfte volledig overeen met het aantal door het CBS geregistreerde overledenen aan COVID-19. In de tweede oversterftegolf van de COVID-19-epidemie overleden meer mensen aan COVID-19 dan er oversterfte was. Mogelijk dat een deel van de overledenen aan COVID-19 al kwetsbaar was en nog een beperkte levensverwachting had door andere aanwezige aandoeningen. Tijdens de derde oversterftegolf was de oversterfte juist hoger dan het aantal mensen dat aan COVID-19 overleed. Mogelijk dragen andere doodsoorzaken bij aan de oversterfte tijdens die periode. Na de eerste en tweede golf overleden er nog steeds mensen aan COVID-19, maar was de sterfte aan andere doodsoorzaken lager. Hierdoor was er in die periode geen sprake meer van oversterfte.

Recente cijfers over de oversterfte en COVID-19-sterfte zijn te vinden op de website van het CBS.

LITERATUUR

- CBS. (2022). <https://www.cbs.nl/nl-nl/reeksen/sterfte-per-week>
- EUROMO. (2022). <https://www.euromomo.eu/graphs-and-maps>
- Eurostat (2022). https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Excess_mortality_-_statistics
- Husby, T., Stoeldraijer, L., & Visser, H. (2020). Oversterfte corona-epidemie: toepassing van een dynamisch regressiemodel. Statistische Trends, 29 juli 2020. <https://www.cbs.nl/nl-nl/achtergrond/2020/31/oversterfte-corona-epidemie-toepassing-van-een-dynamisch-regressiemodel>
- Nepomuceno, M.R., Klimkin, I., Jdanov, D.A., Galarza, A.A., & Shkolnikov, V. (2021). Sensitivity of excess mortality due to the COVID-19 pandemic to the choice of the mortality index, method, reference period, and the time unit of the death series. medRxiv 2021.07.20.21260869; doi: <https://doi.org/10.1101/2021.07.20.21260869>
- RIVM. (2022). <https://www.rivm.nl/monitoring-sterftecijfers-nederland>
- Stoeldraijer, L., Schürmann, R., & Van der Toorn, J. (2022). Sterfte, oversterfte en COVID-19-sterfte in 2020 en 2021. Statistische Trends, 23 juni 2022. <https://www.cbs.nl/nl-nl/longread/statistische-trends/2022/sterfte-oversterfte-en-covid-19-sterfte-in-2020-en-2021>

LENNY STOELDRAIJER is statistisch onderzoeker bij het Centraal Bureau voor de Statistiek en gepromoveerd op het onderwerp sterfte.
E-mail: l.stoeldraijer@cbs.nl



Wat kopen Nederlanders, hoeveel bewegen ze en hoe verplaatsen ze zich?

Smart Surveys: apps en sensoren in survey-onderzoek

Smart surveys zijn enquêtes waarbij sensoren worden ingezet voor het beantwoorden van vragen die respondenten veel tijd kosten en/of waarvoor gedetailleerde kennis en herinnering nodig is. In dit artikel bespreken we het onderzoek dat onder andere is gedaan door een Europees consortium van statistiebureaus naar de mogelijkheden en uitdagingen. We illustreren het onderzoek met toepassingen op het meten van consumptie, mobiliteit en lichaamsbeweging.

BARRY SCHOUTEN EN ANNEMIEKE LUITEN

Sommige vragen die mensen moeten beantwoorden in surveys zijn niet eenvoudig. Wie weet bijvoorbeeld precies hoeveel kilometer een verplaatsing was, of wat precies de aankomsttijd was (zoals in het CBS-verplaatsingsonderzoek gevraagd wordt), of hoeveel minuten gefietst wordt in een gemiddelde week (zoals in

het CBS-gezondheidsonderzoek). Sensoren zoals locatiesensoren en bewegingssensoren kunnen assisteren bij het beantwoorden van deze moeilijke vragen. Surveys waar sensoren worden ingezet noemen we *smart surveys*. Smart surveys kunnen het beantwoorden makkelijker en interessanter maken voor respon-

denten, en daarbij de kwaliteit van de gegevens verbeteren.

Aan de andere kant, smart surveys kennen hun eigen problemen. Er kunnen nieuwe fouten worden geïntroduceerd doordat niet iedereen in staat of bereid is om sensormetingen uit te voeren. Niet iedereen heeft een smartphone, en menig respondent maakt zich zorgen over de privacy van dit soort metingen. Ook als een respondent wel bereid en in staat is tot sensormetingen, kunnen de sensoren zelf fouten maken. In veel gevallen is het dan ook nodig om de respondent te betrekken bij het interpreteren of corrigeren van de metingen. Ook is het nodig om te investeren in nieuwe logistiek, IT-infrastructuur en de ontwikkeling van (machine learning) algoritmes om de sensordata te duiden.

Sinds enkele jaren doet het CBS onderzoek naar smart surveys, waarbij we duidelijk proberen te krijgen of de potentie van smart surveys opweegt tegen de uitdagingen. De afgelopen twee jaar is dit onderzoek, met subsidie van Eurostat, opgepakt door een Europees consortium van statistiebureaus (Eurostat, 2022). De belangrijkste vragen die we moeten beantwoorden zijn:

- Waar ligt de balans tussen actieve en passieve dataverzameling; hoeveel moet de respondent doen in termen van correctie of duiding van de metingen, en wanneer wordt die inspanning voor de respondent (te) belastend.
- Welke nieuwe methodologie moet ontwikkeld worden op het gebied van machine learning en respondent benadering; wat voor strategieën moeten worden gebruikt zodat respondenten ons vertrouwen, begrijpen wat de bedoeling is, en expliciet toestemming geven om hun sensormetingen met ons te delen.
- Wat zijn de ethische en wettelijke implicaties van het verzamelen van dit soort data. Voldoet het meten met sensoren aan het principe van data minimalisatie, zoals beschreven in de wet op de privacy en gegevensbescherming (AVG/GDPR).

In de volgende paragraaf illustreren we de mogelijkheden en methodologische vragen met drie voorbeelden: het meten van consumptie, het meten van mobiliteit en het meten van fysieke bewegingen. In elk voorbeeld beschrijven we waarom het thema zich leent voor smart metingen, welke smart metingen worden gebruikt, welke nieuwe methodologie wordt ingezet op het gebied van machine learning, welke respondent-interactie gevraagd wordt, en hoe de smart survey eruit ziet.

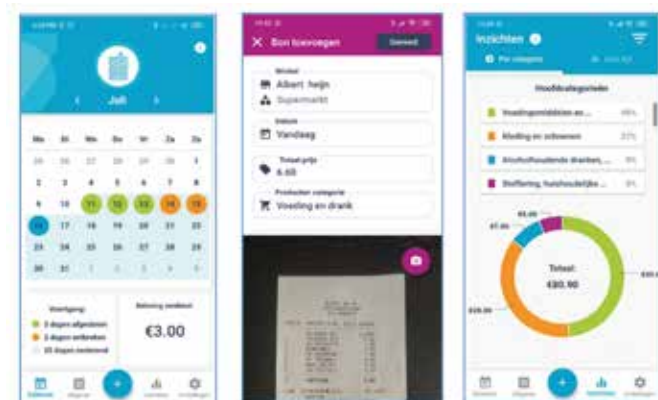
Toepassingen

1. Het meten van consumptie met de HuishoudBudget App

Een van de meest belastende onderzoeken van het CBS is het Budget Onderzoek. Huishoudens houden een maand lang hun uitgaven van meer dan 20 euro bij, en een week lang alle uitgaven. Alle huishoudleden houden gedurende de dag een papieren dagboekje bij. Die dagboekjes worden aan het eind van de dag overgenomen in een webvragenlijst. Mede door die belasting is de respons op dit onderzoek laag. Ook is er sprake van meetfouten, doordat uitgaven vergeten worden of verkeerd worden ingevoerd.

Om de belasting voor de respondenten te verlagen en de kwaliteit van de metingen te verbeteren is een smartphone-app gebouwd waarbinnen respondenten aankopen meteen kunnen invoeren (zie figuur 1). Dat kan door aankoopbonnen in hun geheel te scannen, zodat niet langer individuele aankopen stuk voor stuk moeten worden ingevoerd. De gekochte artikelen van de scan worden via Optical Character Reading herkend. Machine learning algoritmes classificeren deze vervolgens in categorieën van COICOP (Classification of Individual Consumption according to Purpose). De COICOP is een classificatie van consumptieve uitgaven die wordt beheerd door de Verenigde Naties. De categorieën worden getoond als cumulatief overzicht in de app, zodat deelnemers een overzicht krijgen van het totaal en de aard van hun uitgaven.

In de pilots met deze app is gevarieerd met de mate van respondent-interactie door hen al dan niet de gelegenheid te geven om de resultaten van de scan te laten corrigeren.



Figuur 1. De HuishoudBudget App

Het laten corrigeren van de scans heeft een positieve invloed op de kwaliteit van de data en lijkt ook noodzakelijk. Het percentage aankoopbedragen dat minder dan 1 euro afwijkt van het ware bedrag stijgt van 11% in de oorspronkelijke scans tot 64% in de door deelnemers gecorrigeerde scans. Ook het correcte aantal producten stijgt van 16% naar 28%. Actieve respondent-interactie lijkt voornamelijk nodig om tot correcte metingen te komen in deze app. Deze voorbeelden tonen aan dat de scan zeker (nog) niet in alle gevallen leidt tot correcte identificatie en classificatie.

Details van deze pilots zijn te vinden in De Groot e.a. (2022) en Akkermans e.a. (2022).

2. Het meten van verplaatsingen door middel van een app

Het onderzoek 'Onderweg in Nederland' dat het CBS in opdracht van Rijkswaterstaat uitvoert, vraagt deelnemers om gedurende één dag al hun verplaatsingen van locatie naar locatie bij te houden en in te voeren in een webvragenlijst. Het onderzoek is belastend voor deelnemers: precieze tijden van verplaatsingen worden gevraagd, postcodes van begin- en eindlocatie moeten worden gezocht en ingevoerd, en van elke verplaatsing moet worden opgegeven met welk vervoermiddel deze werd uitgevoerd. Dit leidt tot meetfouten als gevolg van onzorgvuldig invullen, omdat deelnemers vertrouwen op hun geheugen, verplaatsingen vergeten of de vragenlijst invullen als de dag nog niet is afgelopen.

Smartphones bevatten locatie-sensoren die een groot deel van deze lastige taak kunnen overnemen. Door middel van GPS kan de positie van een smartphone tot 5

meter nauwkeurig worden bepaald. Dezelfde sensor kan tonen met welke snelheid een telefoon zich verplaatst. Machine learning algoritmes bepalen wanneer iemand onderweg is en wanneer een locatie is aangedaan. Met deze informatie kan, gekoppeld aan informatie over wegenet en openbaar vervoer infrastructuur, worden afgeleid met welk vervoermiddel een verplaatsing is gedaan (figuur 2). Het is in principe mogelijk dat de deelnemer, behalve het downloaden van de app en het invullen van wat achtergrondkenmerken nog nauwelijks naar de app hoeft om te kijken. Zo ver zijn we nog niet.

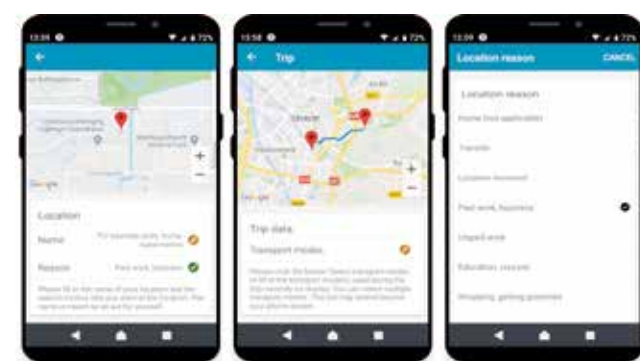
Een actieve rol voor respondenten is nog nodig om context te geven bij een verplaatsing (waarom was de respondent onderweg) en om het vervoermiddel in te voeren. Ook is de app nog niet in staat om onderscheid te maken tussen een 'echte' locatie en, bijvoorbeeld, stilstaan op de snelweg in een file. Het onderzoek naar hoeveel correcties die respondent moet geven en hoeveel het algoritme kan corrigeren is nog lopend.

Meer details over dit onderzoek zijn beschreven in McCool e.a. (2021).

3. Het meten van lichaamsbeweging met een beweegmeter

Waar in de vorige twee voorbeelden gebruik werd gemaakt van een smartphone app, wordt in dit voorbeeld gebruik gemaakt van externe sensors (beweegmeters) die deelnemers van het CBS krijgen (figuur 3).

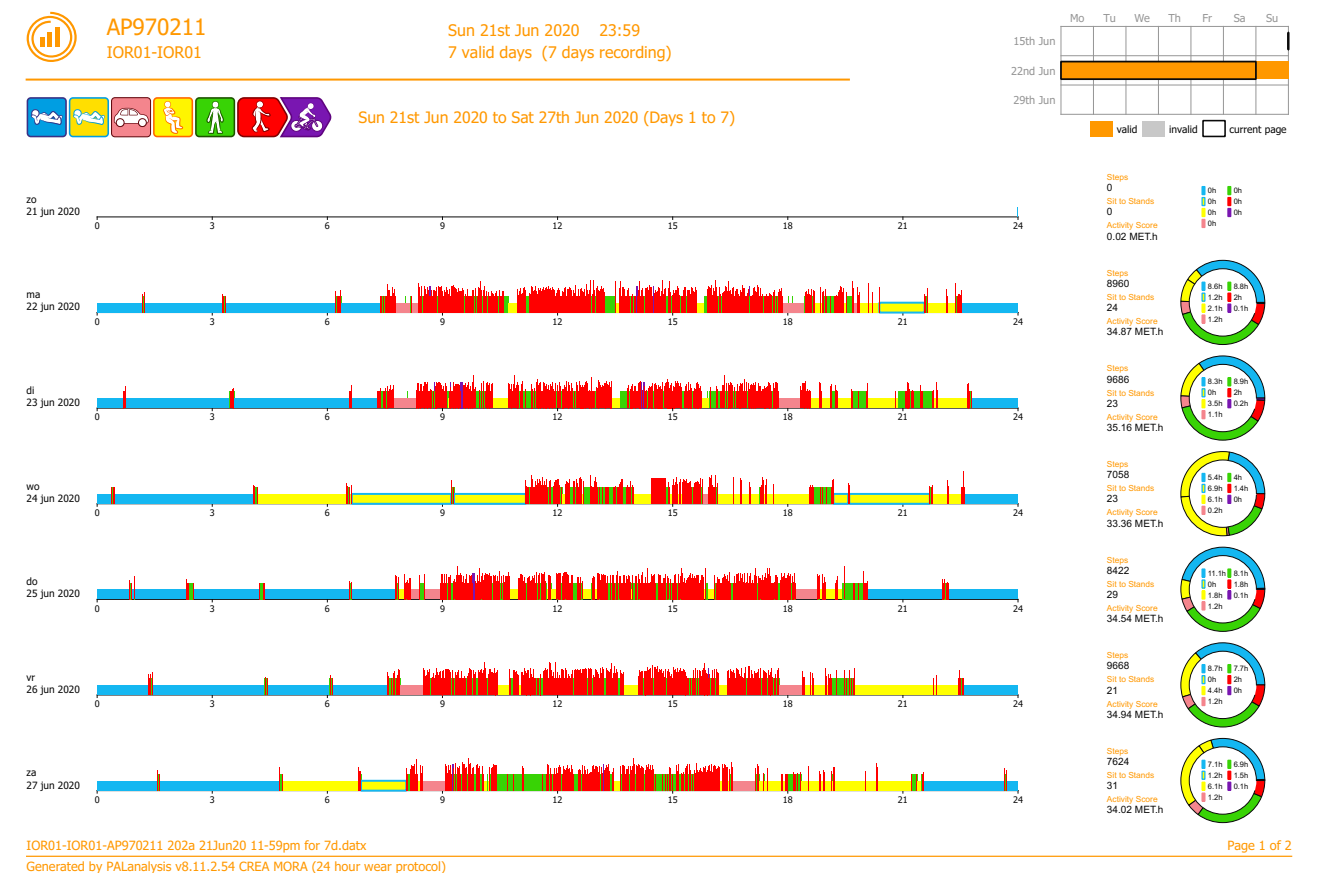
In CBS-surveys over gezondheid en leefstijl wordt lichaamsbeweging gemeten door middel van een vragenlijst. Deze vraagt respondenten om te schatten hoeveel tijd zij in een 'gemiddelde week' besteden aan sport,



Figuur 2. App om verplaatsingen te meten



Figuur 3. Beweegmeters



Figuur 4. Data van de beweegmeter

fietsen, huishoudelijke taken, tuinieren, etc. Dit is geen gemakkelijke taak, en onderzoek toont aan dat mensen de mate van lichamelijke activiteit schromelijk overschatten (bijvoorbeeld Kapteyn e.a. 2018). Het meten van lichaamsbeweging lijkt dan ook een goede kandidaat voor meting via sensoren.

In vier pilots is respondenten die een vragenlijst invulden gevraagd of zij aanvullend wilden meedoen met een onderzoek met beweegmeters. Voor het onderzoek is gebruik gemaakt van de activPAL™ beweegmeter die op het bovenbeen wordt gedragen. De meter meet de positie van het been en of er wel of niet activiteit is. Op basis daarvan berekent een algoritme of iemand ligt, zit, staat, loopt of fietst, en met welke intensiteit. De meters vergen geen interactie met de deelnemer: de data worden opgeslagen op de meter en – na terugsturen – uitgelezen en verder verwerkt (figuur 4).

In deze pilots was een belangrijke onderzoeksvraag of mensen mee willen doen met dit soort onderzoek, en wat voor soort mensen dat zijn: alleen topsporters, of ook mensen die (te) weinig bewegen. In elk van de pilots was het aandeel respondenten op de vragenlijst dat

bereid was om mee te doen met aanvullend onderzoek ongeveer 50%. Afhankelijk van de initiële respons op de vragenlijst komt dat neer op 11 tot 38% van de steekproef. De deelnemers zijn vooral mensen die voldoen aan de beweegnorm van 150 minuten minimaal matig intensief bewegen per week (RIVM, 2020). Toekomstig onderzoek richt zich op het vinden van manieren om ook minder actieve deelnemers te vinden. Zie Luiten e.a. (2022) voor een gedetailleerde beschrijving van de opzet en resultaten van de pilots.

Discussie

In drie voorbeelden hebben we laten zien dat smart surveys kansen bieden. Mensen zijn bereid om mee te doen met dit soort metingen, hoewel die bereidheid (nog) niet op het zelfde niveau is als bij vragenlijsten. Aanvullend onderzoek is nodig hoe wij de representativiteit van de respons kunnen verbeteren.

In de inleiding is beschreven dat er drie belangrijke vragen zijn die beantwoord moeten worden om de

haalbaarheid van smart surveys goed te kunnen beoordelen. In de voorbeelden die beschreven zijn, zijn eerste stappen gezet, maar het eindoordeel is nog verre van duidelijk en kan verschillend zijn voor verschillende toepassingen.

Methodologische vragen die er nog leven betreffen de te ontwikkelen machine learning algoritmes. Ook deze zijn feilbaar en hebben een foutmarge op de voorspellingen. Een beweging kan bijvoorbeeld met 80% kans lopen zijn en met 20% kans rennen. Bovendien zullen de algoritmes herhaaldelijk of zelfs doorlopend moeten worden getraind, bijvoorbeeld door de introductie van nieuwe producten in het Budget onderzoek. Respondenten zijn onmisbaar bij het verbeteren van nauwkeurigheid. Welke respondent-interactie noodzakelijk en niet te belastend is, hoe deze interactie op een begrijpelijke manier in te bouwen in een user interface, en hoe om te gaan met de onzekerheden in machine learning voorspellingen als respondenten het ook niet kunnen weten zijn aspecten waar nader onderzoek nodig is.

Ook met betrekking tot logistiek en IT zijn er nog legio vragen. We gebruiken privé apparaten van onze respondenten en hebben complexe IT nodig om sensoren aan te sturen. Hoe deze afdoende te beveiligen is nog onvoldoende bekend.

We verzamelen nieuwe vormen van data die respondenten zelf ook niet volledig kennen. Dit roept de vraag op of respondenten wel weten waarvoor ze toestemming geven. Hoeveel en welke informatie moeten we respondenten geven opdat zij werkelijk 'informed consent' kunnen geven? De wet op de informatiebescherming roept de vraag op hoe en waar data voorbewerkt kunnen worden vanuit het gezichtspunt van data minimalisatie. Classificatie van producten in de Budget app zou bijvoorbeeld aanzienlijk nauwkeuriger en sneller kunnen als de data van het apparaat van de respondent naar het CBS gezonden konden worden, geclassificeerd, en daarna weer teruggestuurd naar de respondent. Vanwege de AVG is dit echter problematisch.

Tot slot is er nog de open vraag wat te doen als de nieuwe vormen van data niet perfect zijn, wat meestal het geval zal zijn, en hoe we moeten omgaan met meta-informatie over kwaliteit. Geldt de AVG ook voor metadata?

Zo ja, hoe kunnen we dan nog weten of data wel goed genoeg zijn, bijvoorbeeld als locatiemetingen uitbijters bevatten of gaten.

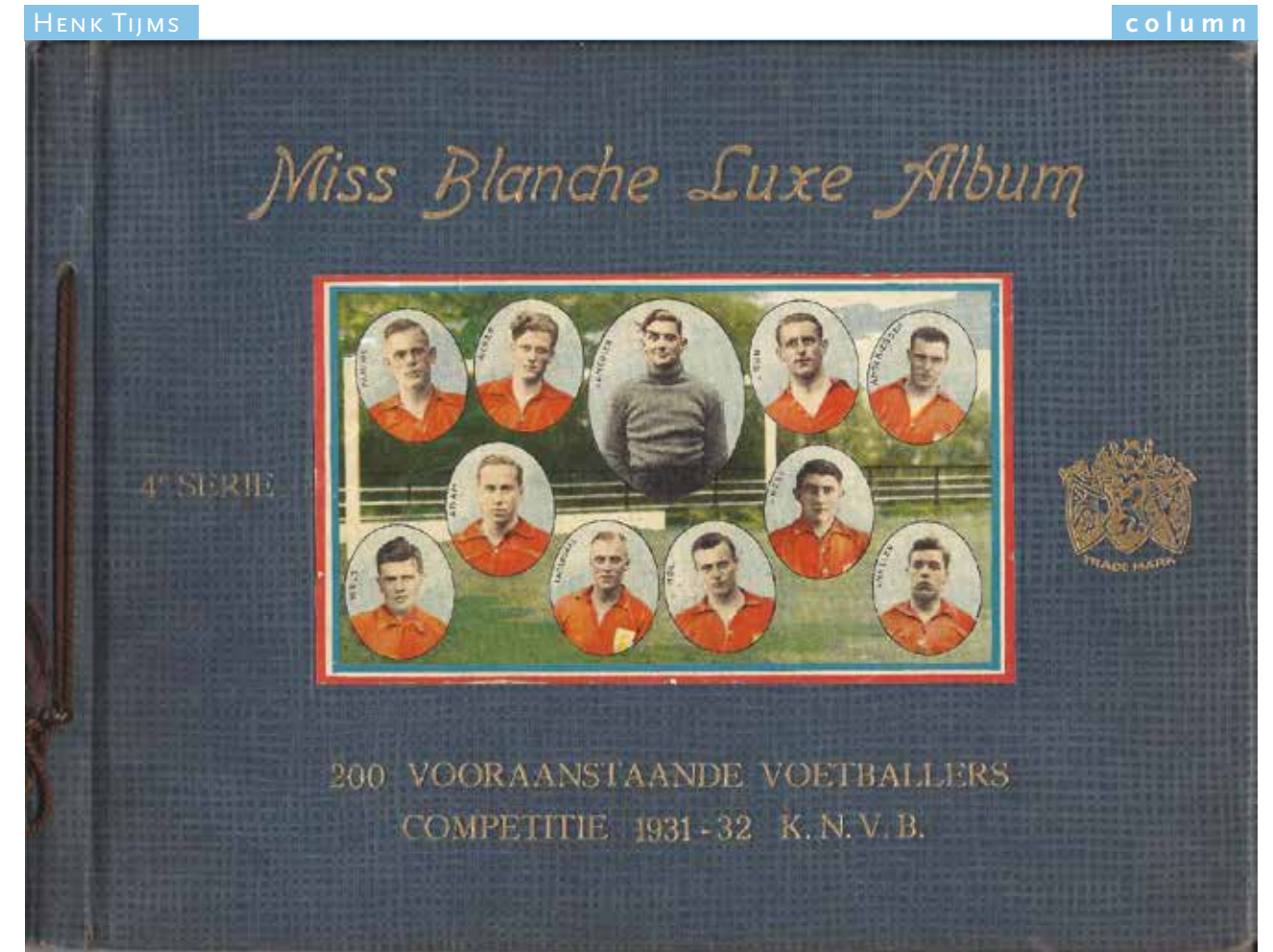
In de komende jaren is aanvullend onderzoek voorzien dat ons inzicht in smart surveys verder zal verdiepen.

LITERATUUR

- Akkermans, J., Rodenburg, E., Groot, J. de, Schouten, B., Martin Bernia, E., Balsa Criado, V., Horcajo, T., Gauche, C., & Osier, G. (2022). *The role of interviewers*. Geraadpleegd van CROS website ESSnet Smart Surveys.
- De Groot, J., Oerlemans, T., Rodenburg, E., Schouten, B., Lope Mariscal, A.C., Bernia, M., Poch, J., Horcajo, T., Blasé Criado, V., Gauche, C., & Osier, G. (2022). *Thematic paper on the Consumption pilot*. Geraadpleegd van https://ec.europa.eu/eurostat/cros/system/files/220413_wp2_-_deliverable_2.1_-_consumption.pdf.
- Eurostat. (2022). Projectdocumentatie ESSnet Smart Surveys, [https://ec.europa.eu/eurostat/cros/content/ess\[1\]net-smart-surveys_en](https://ec.europa.eu/eurostat/cros/content/ess[1]net-smart-surveys_en).
- Kapteyn, A., Banks, J., Hamer, M., Smith, J. P., Steptoe, A., Van Soest, A., & Wah, S.H. (2018). What they say and what they do: comparing physical activity across the USA, England and the Netherlands. *J Epidemiol Community Health*, 72(6), 471-476.
- Luiten, A., Toepoel, V., Schouten, B., Cierpiat-Wolan, M., Kapica, K., Szlachta, P., Lusyne, P., & Van der Beken, H. (2022). *Health – measuring physical activity*. Deliverable 2.1. Essnet Smart Statistics. Geraadpleegd van Thematic papers per pilot - Physical activity (europa.eu).
- McCool, D., Lugtig, P., Mussmann, O., & Schouten, B. (2021). An app-assisted travel survey in official statistics: Possibilities and challenges. *Journal of Official Statistics*, 37, 149-170. <http://dx.doi.org/10.2478/JOS-2021-0007>.
- RIVM. (2020). Beweegrichtlijnen. Geraadpleegd van <https://www.rijksoverheid.nl/documenten/publicaties/2020/07/07/beweegrichtlijnen>.

BARRY SCHOUTEN werkt als senior wetenschappelijk medewerker bij het Centraal Bureau voor de Statistiek in Den Haag en als bijzonder hoogleraar bij de Universiteit Utrecht. Hij is coördinator van een gezamenlijk dataverzameling-innovatienetwerk tussen het CBS en de Universiteit Utrecht. E-mail: jg.schouten@cbs.nl

ANNEMIEKE LUITEN werkt als wetenschappelijk medewerker bij het Centraal Bureau voor de Statistiek in Heerlen waar zij onderzoek doet naar de rol van sensormetingen in enquêtes. E-mail: a.luiten@cbs.nl



VOETBALPLAATJES wie heeft ze niet verzameld?

Voor het wereldkampioenschap voetbal in Qatar hebben Albert Heijn, Jumbo en andere supermarktketens klaarblijkelijk afgezien van acties met plaatjes van de voetballers die acte de présence geven op dit besmette evenement in Qatar. Het verzamelen van voetbalplaatjes is wereldwijd een rage onder de jeugd. Menig lezer van STA^tOR zal zich vroeger ook aan deze verslavende activiteit overgegeven hebben. Wereldberoemd zijn de voetbalalbums van de gebroeders Guiseppe en Benito Panini. In 1961 brachten ze vanuit hun simpele krantenkiosk in de Italiaanse stad Modena hun eerste album uit met plaatjes van alle voetballers die in dat jaar actief waren in de Italiaanse voetbalcompetitie Serie A. Dit was het begin van een wereldwijd imperium dat de gebroeders Panini opbouwden en dat hen schatrijk maakte. Overigens waren de Panini-albums niet de eerste in hun soort. In Ne-

derland werden jaren dertig en veertig van de vorige eeuw al diverse albums met plaatjes van sporthelden uitgebracht door het legendarische sigarettenmerk Miss Blanche van de Vittoria Egyptian Cigarette Company uit Rotterdam, een sigarettenmerk dat ook de inspiratie vormde voor de psychologische roman *Miss Blanche* uit 2004 van Nelleke Noordervliet over een Rotterdamse tabakswinkelier die zijn leven ziet veranderen door een toevallige ontmoeting met een kettingrokende Turkse schone. De Miss-Blanche-albums zijn heden ten dage ware *collectors items*, zeker het album dat in 1932 werd uitgebracht met plaatjes van tweehonderd vooraanstaande voetballers uit de eerste en tweede klas van de Nederlandse voetbalcompetitie 1931-1932, waaronder de befaamde internationals Leo Halle, Puck van Heel en Bas Paauwe. Fabrikanten van margarinemerken als Leeuwenzegel en

Planta ('pukeltjes-affaire') deden in de jaren vijftig ook al pogingen om met plaatjes van bekende Nederlandse sporters de verkoop van hun producten te stimuleren. Nog steeds ben ik de trotse bezitter van het complete Leeuwenzegelalbum *Nederlandse Sport Successen* uit 1953 met onder meer plaatjes van heroïsche sporthelden als Kees Broekman, Wim van Est en Piet Kraak.

Voor het Panini-album voor het Europees kampioenschap voetbal 2020 moesten 678 stickers worden verzameld. De stickers konden worden gekocht in pakketjes van zes verschillende stickers. Voor het geval dat je niet kunt ruilen met andere verzamelaars zijn interessante vragen: hoeveel pakketjes heb je nodig om met een kans van 50% of meer een complete collectie stickers te verkrijgen en hoeveel is het benodigde aantal stickers als je een succeskans van 90% of meer wilt hebben? Onder de aanname dat de stickers uniform verdeeld zijn over de pakketjes, geeft Monte Carlo simulatie met 100.000 runs als antwoord dat 780 respectievelijk 993 pakketjes nodig zijn. Een analytische aanpak is ook mogelijk met behulp van een absorberende Markovketen. De Markovketen heeft de 679 toestanden $i = 0, 1, \dots, 678$, waarbij toestand i betekent dat je tot nu toe i verschillende stickers verzameld hebt. Toestand 0 is de begintoestand en toestand 678 is een absorberende toestand. De 1-staps overgangskansen in de Markovmatrix P zijn simpel en worden gegeven door hypergeometrische kansen. De kansverdeling van het aantal benodigde pakketjes voor een complete verzameling van stickers vind je door de matrix P herhaald met zichzelf te vermenigvuldigen: de kans dat je meer dan n pakketjes nodig hebt is 1 minus het $(0, 678)$ de element van het n -voudig matrix product van P met zichzelf.

De berekeningen voor het *coupon collector's* probleem met stickers in dezelfde frequenties zijn veel simpeler als je bij aankoop maar één sticker krijgt. Dan kun je de kansverdeling van het aantal benodigde aankopen voor een complete verzameling van stickers berekenen met de bekende inclusie-exclusie formule uit de kansrekening (of met een absorberende Markovketen) en kunnen voor de verwachtingswaarde m_s en de variantie v_s van het aantal benodigde aankopen bij s verschillende stickers de expliciete formules

$$m_s = s \sum_{k=1}^s \frac{1}{k} \text{ en } v_s = s^2 \sum_{k=1}^s \frac{1}{k^2} - s \sum_{k=1}^s \frac{1}{k}$$

gegeven worden. Deze expliciete formules berusten op het feit dat het aantal aankopen nodig om van verschil-

lende stickers naar $i+1$ verschillende stickers te gaan een geometrische kansverdeling met succeskans $(s-i)/s$ heeft. Voor s voldoende groot kunnen $\sum_{k=1}^s 1/k$ en $\sum_{k=1}^s 1/k^2$ worden benaderd door $\ln(s) + \delta$ en $\pi^2/6$, waarbij $\delta = 0,5772\dots$ de Euler-Mascheroni constante is.

De aanname dat de stickers random verdeeld zijn over de pakketjes en dus met dezelfde frekwenties voorkomen, is cruciaal in bovenstaande berekeningen. Een empirische studie van Tim Paulden, 'Soccer fans in shiny sticker shock' in het Britse statistiek tijdschrift *Significance* Vol.18(2021)3, leidt tot grote twijfels of deze aanname vervuld is geweest bij het Panini-album voor het wereldkampioenschap voetbal in 2018. Verbazingwekkend is dit voor mij niet. Ook bij het album *Voetbalsterren* uit 1969 van de Utrechtse uitgever Vanderhout gaan hardnekkige geruchten dat de uitgever van sommige voetballers (waaronder Johan Cruyff) minder plaatjes heeft laten drukken om de verzamelaars hongerig te houden.

Als de frequentie waarmee een sticker voorkomt niet voor elke sticker hetzelfde is, wordt de situatie veel ingewikkelder. Voor het geval dat de stickers alleen per stuk verkocht worden, dan kunnen de verwachtingswaarde en de standaardafwijking van het aantal benodigde aankopen voor een complete verzameling nog wel analytisch berekend worden. Dit gaat met numerieke integratie van twee oneigenlijke integralen die met de 'Poissonization' truc van A.N. Kolmogorov verkregen worden. Zo vind je bijvoorbeeld de waarden 61,217 en 35,976 voor de verwachtingswaarde en de standaardafwijking van het aantal benodigde worpen met twee zuivere dobbelstenen totdat elk van de elf mogelijke uitkomsten 2, ..., 12, voor de som verschenen is. Een approximatie voor de kans dat meer dan n worpen nodig zijn om deze elf uitkomsten te verkrijgen is 1 minus het product van de kansen $1 - e^{-np_k}$ over $k=2, \dots, 12$, waarbij p_k de kans is op de som k in één worp. Deze simpele approximatie is algemeen toepasbaar voor het *coupon collector's* probleem met ongelijke kansen, ook voor het probleem met meerdere coupons per aankoop zoals bij het Panini-album.

HENK TIJMS is emeritus-hoogleraar operations research aan de Vrije Universiteit en auteur van diverse leerboeken over operations research en kansrekening, waaronder het boek *Basic probability; What every math student should know* (World Scientific Press, 2021, 2e druk).
E-mail: h.c.tijms@xs4all.nl



STATISTIEK EN FILOSOFIE: EEN VOORTSCHRIJDENDE SAMENSPRAAK?

RICHARD STARMANS

Precies honderd jaar geleden ontving Albert Einstein de Nobelprijs voor Natuurkunde en de wijze waarop dit geschiedde was in menig opzicht curieus. Pas op 10 November 1922 stuurde het Nobelcomité Einstein een telegram met de mededeling dat hem alsnog de Nobelprijs voor – nota bene – het jaar 1921 was toegekend. Deze merkwaardige toewijzing 'met terugwerkende kracht' was in nog een ander opzicht verlaat. Zij vond pas plaats zeventien jaar na Einsteins *annus mirabilis* 1905, waarin hij een viertal artikelen had gepubliceerd over respectievelijk het foto-elektrisch effect, de Brownse beweging, de speciale relativiteitstheorie en de massa-energie vergelijking, en ruim zeven jaar nadat de algemene relativiteitstheorie in 1915 het licht zag.

Als klap op de vuurpijl besloot het Nobelcomité uiteindelijk Einstein de prijs toe te kennen 'for his services to theoretical physics and especially for the law of the photoelectric effect'. De relativiteitstheorie werd niet expliciet bekroond! Comitévoorzitter Svente Arrhenius beseftte dat hij de wetenschappelijke wereld een verklaring schuldig was en tijdens de *Presentation Speech* op 10 December 1922 verdedigde hij manmoedig zijn commissie door het controversiële karakter van de relativiteitstheorie te benadrukken. 'Most discussion centres on his theory of relativity. This pertains essentially to epistemology and has therefore been the subject of lively debate in philosophical circles. It will be no secret that the famous philosopher Henri Bergson has challenged his theory, while other philosophers have acclaimed it wholeheartedly.' Dat klinkt uiteraard als een nogal krampachtige apologie voor de werkwijze van de commissie, waarbij de relativiteitstheorie lijkt te worden verbannen uit het domein van de fysica naar de 'speeltuinen' der filosofen.

Einstein en Bergson

Toch sloeg Arrhenius met zijn verwijzing naar de Franse filosoof Henri Bergson (1859–1941) de plank niet geheel mis. Eerder dat jaar, op 6 april 1922, had de wijsgerig zeer onderlegde Einstein tijdens een lezing aan de Société Française de Philosophie te Parijs Bergson ontmoet. Diens status als filosoof, intellectueel en diplomaat was enorm en hij deed in internationale bekendheid niet voor Einstein onder. Sedert 1914 maakte Bergson deel uit van de Académie Française, de Franse regering zond hem in 1917 als speciale gezant naar de Amerikaanse president Wilson in een poging de VS actiever bij WOI te betrekken en hij zou in 1927 de Nobelprijs voor Literatuur ontvangen. De ontmoeting leidde tot een jarenlange controversie tussen Einstein en Bergson, waarmee vele vooraanstaande wetenschappers en filosofen zich zouden bemoeien en die tot op heden onderwerp vormt van debat (Campo, 2021; Canales, 2015). Tijdens de bijeenkomst in Parijs nam Bergson het woord en kritiseerde de representatie van de tijd in de relativiteitstheorie, mede tegen de achtergrond van het fameuze tweeling-experiment, dat toen ook al buiten de natuurkunde opzien baarde. Bergson feliciteerde Einstein met de experimentele resultaten, maar stelde dat achter de relativiteitstheorie een foutieve metafysica schuilging. Tijd werd volgens hem als het ware 'geometriseerd', langs een afstandsmaat gelegd, gefixeerd en in stukken verdeeld, die geheel te overzien waren. Deze fysische tijd was volgens Bergson 'passief', veronderstelde een universum dat paradoxaal genoeg juist een terugkeer naar het determinisme impliceerde, waarbij alles vastlag, wording en creatie niet mogelijk wa-

Wachtenden bij de ingang van de Académie Française tot Henri Bergson naar buiten komt. Parijs, 24 januari 1918. Foto: Excelsior, journal illustré quotidiens

ren; een werkelijkheid die zich bovendien onttrok aan de menselijke ervaring en het bewustzijn min of meer obsoleet verklaarde. Het was feitelijk louter een gemeten tijd of kloktijd, een (wiskundige) abstractie die voorbijging aan de werkelijke tijd. Deze laatste behelsde volgens Bergson de zuivere tijd (*temps pur*) of reële duur (*durée réelle*) die 'actief' was, niet ontdaan van de ervaring en 'gegeven aan het bewustzijn'. Deze reële duur maakte verandering, wording en creatie mogelijk en was noodzakelijk voor het begrijpen van het universum (Starmans, 2022).

Einstein nam de zaak aanvankelijk luchtig op en riposteerde met de inmiddels beroemde opmerking dat er naast fysische tijd weliswaar psychologische tijd kan bestaan, maar beslist geen filosofische tijd. Voor Bergson stond er evenwel veel op het spel en hij zette op twee fronten de aanval in. Allereerst zocht hij steun voor zijn positie in de academische wereld, ook bij vooraanstaande fysici. Zo benaderde hij Hendrik Lorentz, zonder dat Einstein daarvan op de hoogte was en ontmoette hij diverse malen Albert Michelson met wie hij de implicaties van het beroemde Michelson-Morley experiment voor zijn 'zaak' besprak. Daarnaast verscheen nog in hetzelfde jaar zijn boek *Durée et simultanéité. À propos de la théorie d' Einstein* (1922). Het maakte duidelijk dat de aanval op Einstein geen losse flodder was, maar een volgende stap in een groot project, dat verder ging dan louter de representatie van de tijd of zelfs maar de relativiteitstheorie en waarvan de uitgangspunten reeds in zijn vroege werk waarneembaar waren. Met behulp hiervan tracht Bergson de wetenschappen, die na de fragmentatie van de kennis in de 19e eeuw waren ontstaan, te verrijken of aan te vullen met een eigen metafysica.

De controverse zou verstrekkende gevolgen hebben voor de verhouding tussen natuurkunde en filosofie, die op dat moment nog betrekkelijk harmonieus was (Starmans, 2022). Zij maakte van Bergson, de man van *l'évolution creative* en *élan vital* een tragische held, die vanuit een hooggestemd utopisme precies het tegenovergestelde bereikte van wat hij beoogde. In een tijd waarin de Franse filosofie zich steeds meer afwendde van de natuurwetenschappen en in Europa een schisma tussen analytische en continentale filosofie ontstond, geloofde hij oprecht in een wisselwerking en kruisbestuiving tussen natuurwetenschap en wijsbegeerte en probeerde hij de eendracht te herstellen. Tevergeefs; Bergson vond in de wetenschap weinig bijval en is thans vooral een boegbeeld in de filosofische traditie die hij verfoeide. De relatie tussen natuurkunde en filosofie (zoals Bergson die voorstond) verslechterde in rap tempo en lijkt heden ten dage een dieptepunt te hebben bereikt. Nobelprijswinnaar Richard

Feynmann verklaarde ooit dat filosofie voor de fysica even nuttig is als de ornitholoog voor de grasparkiet, Steven Hawking proclameerde in 2011 de dood van de wijsbegeerte en de vorig jaar overleden fysicus en eveneens Nobelprijslaureaat Steven Weinberg sprak schamper over de *unreasonable ineffectiveness of philosophy* (Starmans, 2022). Dat suggereert een onvervalste conflicthypothese betreffende fysica en filosofie. Daartegen is wel het een en ander in te brengen, maar we moeten hier volstaan met de constatering dat de huidige 'dialoog' tussen beide vooral mogelijk werd nadat filosofen hun conceptie van wijsbegeerte fors hadden bijgesteld.

Geldt dit nu voor alle wetenschappen? Hoe is bij voorbeeld de verhouding tussen statistiek en filosofie? Doemt ook hier een conflicthypothese op, zijn het NOMA's (*non overlapping magisteria*) of tekent zich een samenspraak af, wederzijds begrip of wellicht zelfs toenadering? We beperken ons tot een bescheiden bloemlezing van saillante voorbeelden, die hopelijk de contouren van de problematiek zichtbaar maken.

De genealogie van de onzekerheid

Anders dan de natuurkunde kan de statistiek niet bogen op een lange, tot de klassieke oudheid teruggaande relatie met de (natuur)filosofie. Toch is er een evidente inhoudelijke verbondenheid. De Griekse (natuur)filosofen worstelden al met variatie en verandering in de natuur, onzekerheid en onvolledigheid van de kennis, en toeval en onbepaaldheid. Toen medio 17e eeuw kansrekening en statistiek opkwamen en trachtten greep te krijgen op deze weerbarstige fenomenen, bleef dit dan ook niet onopgemerkt. Toch was de ontvangst onder filosofen aanvankelijk niet erg hartelijk en ook niet onverdeeld gunstig. Sceptis kwam allereerst vanuit de ethiek, die de probabilistische redeneerwijze wantrouwde, waardoor statistiek dikwijls in het beklaagdenbankje terecht kwam (Starmans, 2018). Daarnaast vanuit de epistemologie: ware kennis is immers axiomatisch-deductief, het denken over onzekerheid en verandering behelst hooguit een vorm van 'doxa', een geïnformeerde of prudente mening. Dat veranderde eind 19e eeuw toen de probabilistische revolutie zich aandienende, vele empirische disciplines daarin meegingen en de wetenschapsfilosofie niet kon achterblijven. Een voorzichtige samenspraak kondigde zich aan. Een klassiek voorbeeld hiervan betreft de wijsgerige belangstelling voor de diverse interpretaties van het kansbegrip (klassiek, frequentistisch, logisch, *propensity*, subjectief/objectief Bayesiaans). Zo verankert het zeer toegankelijke

Philosophical introduction to probability (2005) van de Italiaanse wetenschapsfilosoof Maria Carla Galavotti ontwikkelingen binnen de kansrekening in de wijsgerige traditie. Pascal, Bernoulli, Bayes en Laplace worden geduid vanuit het empirisme van onder meer Francis Bacon, Hume, Stuart Mill en Whewell. Het logicisme van De Morgan, Boole, Jevons en Keynes krijgt reliëf tegen de achtergrond van filosofen als Wittgenstein, Waismann en Carnap. C.S. Peirce's propensity benadering wordt door de filosoof Popper 'gegeneraliseerd', maar ook diverse vormen van (hedendaags) Bayesianisme passeren de revue. Dit alles vanuit enkele minimale formele beginselen, waaronder Bayes' rule en Kolmogorovs axiomatisering van de kansrekening.

Een tweede en nog saillant voorbeeld betreft de wijze waarop statistische ideeën vorm hebben gekregen in de Bayesiaanse confirmatietheorie, die nog steeds toonaangevend is onder wetenschapsfilosofen die een rationele reconstructie van wetenschap nastreven. Fameus zijn *Scientific Reasoning; the Bayesian approach* uit 2006 van Colin Howson en Peter Urbach en *Bayes or Bust* uit 2005 van John Earman. Allerm minst traditioneel Bayesiaans, maar evenzeer relevant is het werk van Ronald N. Giere, die sedert eind jaren 60 zijn wetenschapsfilosofie opbouwde vanuit theoretisch-statistische kwesties, zoals controversen tussen Pearson en Fisher, maar ook het verschil tussen Neymans intervalschattingen en Bayesiaanse benadering van *credible intervals*. Zoekt men een 'actievare' wijsgerige bijdrage in de ontwikkeling van het probabilistische denken, dan kan a fortiori het werk van de filosofen Hans Reichenbach, Irving John Good en Patrick Suppes worden genoemd, die niet alleen binnen de probabilistische benadering van causaliteit baanbrekend werk verrichtten, maar ook in ruimere zin bijdroegen aan de kennisleer.

Uiteraard kan men bij het duiden van de relatie tussen statistiek en filosofie ook een andere invalshoek kiezen en op zoek gaan naar denkers voor wie beide disciplines twee kanten van dezelfde medaille vormden en die zo bijdroegen aan de 'genealogie' van het denken over onzekerheid. Men komt dan al gauw uit bij de vermaarde statisticus, maar vergeten filosoof Karl Pearson, wiens *The Grammar of Science* uit 1905 grote invloed uitoefende op Einstein en één van de belangrijkste boeken uit de vroege wetenschapsfilosofie zou vormen. We besluiten deze summiere opsomming met filosofen die, wellicht behept met enige vorm van bergsoniaanse hubris, het vigerende formele paradigma van de kansrekening uitdaagden. Bekend is in dit opzicht het boek *The Probable and the Provable*, geschreven door de filosoof L.J. Cohen, die een

alternatieve probabilistische methode ontwikkelde, welke niet gebukt gaat onder de tekortkomingen van wat hij benoemt als de 'Pascaliaanse' traditie. Het boek is onder rechtsfilosofen nog steeds populair en sluit aan op psychologisch onderzoek naar probabilistische paradoxen in alledaags redeneren, zoals dat in de jaren 70 opkwam en vooral door psychologen als Daniel Kahneman en Amos Tversky tot bloei is gebracht. Hoe dan ook, deze onvolledige opsomming suggereert dat van een harde conflict-hypothese geen sprake hoeft te zijn en dat enige vorm van wederzijds begrip of interactie tussen statistiek en filosofie niet op voorhand dient te worden uitgesloten.

Andrew Gelman en Deborah Mayo

Is dit alles nu echt overtuigend? Nee, volgens de Amerikaanse Bayesiaanse statisticus Andrew Gelman (1965), die al meer dan tien jaar in artikelen, lezingen en blogs pleit voor meer wisselwerking en kruisbestuiving tussen statistiek en filosofie. Al in zijn artikel 'Philosophy and the practice of Bayesian statistics' (2011) bepleit hij het eminente belang van filosofie voor zowel het statistisch onderzoek, als de statistische praktijk. De auteur stelt dat in dit opzicht te veel sprake is van *freewheelen* onder zijn collega's. Hij benadrukt zelf primair statisticus te zijn en slechts een deel van de relevante filosofische literatuur in zijn beschouwing te hebben betrokken. Zo beperkt hij zich tot iconen als Karl Popper en Thomas Kuhn en spoort zijn collega's aan om direct aansluiting te zoeken bij meer recent wetenschapsfilosofisch onderzoek.

Gelman gaat vervolgens een stap verder en stelt dat het in de hedendaagse statistiek gebruikelijk is om statistische scholen of uitgangspunten te relateren aan of zelfs te verankeren in filosofische paradigma's of posities. Hij verwijst daarbij onder meer naar het inductivisme en het hypothetisch-deductieve-model als concepties van wetenschap. Vervolgens typeert hij diverse verworvenheden van de 'klassieke' statistiek als manifestaties van hypothetisch-deductief redeneren; de significantie-testen en MLE van Ronald Fisher, het hypothesetoetsen zoals ontwikkeld door Egon Pearson en Jerzy Neyman, en tot slot de theorie van betrouwbaarheidsintervallen, waarvan Neyman eveneens de grondlegger was. De Bayesiaanse statistiek daarentegen wordt volgens Gelman in de statistiek algemeen beschouwd als een proeve van inductivisme. Beide stromingen worden door de auteur als tegenovergestelde visies gepresenteerd. Gelman laat het niet bij dit alles en stelt zich tot doel een nieuwe filosofie voor de Bayesiaanse statistiek te ontwikkelen, die

empirisch adequaat is en de bestaande statistische praktijk beter verdisconteert. Hiertoe kritiseert de auteur de eerder geschetste vigerende typering van het veld en stelt dat de Bayesiaanse statistiek juist als een succesvolle instantiatie van het hypothetico-deductieve model moet worden beschouwd en niet meer of minder inductief is dan de hier genoemde vormen en uitingen van de ‘klassieke’ statistiek. De meest succesvolle toepassingen van Bayesiaans redeneren overstijgen de standaardopvatting waarbij redeneren wordt opgevat als het updaten van een prior-verdeling tot een posteriori-verdeling met behulp van een likelihood-functie op basis van nieuwe data, waarna vervolgens de ‘graad van geloof’ hieraan dient te worden aangepast, aldus Gelman. Het gaat er veeleer om diverse modellen te ontwikkelen (*model fitting*), die vervolgens worden geëvalueerd (*model checking*) en aan een opvolging van strenge tests worden onderworpen, waarna het sterkste model overblijft. Omdat volgens Gelman (vrijwel) alle modellen toch ‘fout’ zijn, dat wil zeggen, niet de kansverdeling bevatten door welke de beschikbare data zijn gegenereerd, is er veeleer sprake van een *trial-and-error* aanpak, die historisch gesproken met Karl Popper kan worden verbonden, maar in de recente literatuur in Deborah Mayo de meest prominente pleitbezorger vindt.

Daarmee is dan gelijk een tweede hoofdrolspeler genoemd. De Amerikaanse filosoof Deborah Mayo publiceerde in 1996 *Error and the growth of experimental knowledge*, waarvoor zij twee jaar later de vijfjaarlijkse Lakatosprijs voor wetenschapsfilosofie ontving. Vanuit een historisch-filosofische invalshoek en met behulp van probabilistische methoden ontwikkelt de auteur een methodiek om hardnekkige problemen in de wetenschapsfilosofie (inductie, rationaliteit, vooruitgang, objectiviteit, etc.) aan te pakken. Het gaat daarbij niet zozeer over de ondersteuning van hypothesen met behulp van waarschijnlijkheidsargumenten, maar om een *error-statistical approach*, waarbij het analyseren, manipuleren, simuleren en beheersen van (patronen van) fouten gekoppeld wordt aan een methode van *severe testing* en *learning from error*. Het werk biedt een nieuwe kijk op de experimentele praktijk, maar tevens een fundamentele kritiek op het Bayesianisme. In 2018 gaat zij nog een stap verder met haar boek *Statistical inference as severe testing: How to get beyond the statistics wars*. Daarin beargumenteert zij dat haar methodology van severe testing een cruciale rol kan spelen in actuele theoretische en praktische statistische ‘kwesties’, zoals het conflict tussen frequentisten en bayesianen, p-hacking, de replicatiecrisis en allerlei andere ongemakken die sedert Ioannides’ essay ‘Why most

published research findings are false’ uit 2005 hoog op de onderzoeksagenda’s staan.

Dit alles was uiteraard koren op de molen van Gelman, die ondanks de anti-Bayesiaanse positie van Mayo mogelijkheden zag tot inpassing in zijn grote project. Zo ging hij in 2019 de uitdaging aan door Mayo’s visie voor te leggen aan een internationaal forum van statistici, werkzaam binnen en buiten de academie, wat tot verrassende stellingnames en positiebepalingen heeft geleid (Gelman, 2019). Het debat is nog verre van voltooid.

Causaliteit

Toch kunnen deze hooggestemde ambities niet probleemloos worden gegeneraliseerd en de studie naar causaliteit is daarvan een treffend voorbeeld. In 2018 verscheen het boek van Rani Lill Anjum en Stephen Mumford, *Causation in science and the methods of scientific discovery*. Beide filosofen formuleren daarin ex cathedra onder meer negen ‘normen’ voor *scientific discovery* en doen dit vanuit een analyse van causaliteit; een benadering die deels een voortschrijdende kritiek op Hume behelst, sterk metafysisch is georiënteerd, de *sciences* als een blackbox benadert en nauwelijks aansluit bij de (experimentele) methodologie en praktijk van natuur- of sociaalwetenschappelijk onderzoek. Op geen enkele wijze wordt getracht de kloof tussen filosofie en methodologie/wetenschap wat betreft causaliteit te overbruggen. Zelfs traditionele filosofische bijdragen -zoals hierboven beschreven- of succesvolle designs in de sociale wetenschappen worden genegeerd en dat geldt a fortiori voor moderne wiskundige (probabilistische en computationele) benaderingen van causal inference. Dat laatste is vooral teleurstellend, omdat daar nu juist sedert het begin van deze eeuw spectaculaire vooruitgang is geboekt en causaliteit, een voor velen obsoleet en achterhaald begrip, een opmerkelijke terugkeer beleefde in het wetenschappelijk discours. Men denke aan bijdragen zoals die van Clark Glymour, Donald Rubin, Judea Pearl, James Robins, Mark van der Laan, Guido Imbens en anderen. Hoe dan ook, het kwam beide auteurs te staan op een ziedende reactie van voornoemde Clark Glymour, die het boek in de *Notre Dame Philosophical Reviews* in 2019 vernietigend recenseerde. Het was diezelfde Glymour die in *Causation, prediction and search* (1993) als een van de eersten *causal discovery* computationeel ging benaderen, wat onder meer leidde tot het TETRAD-project en een opmaat vormde tot een thans florerend onderzoeksgebied. In zijn recensie verwijt de getergde Glymour de auteurs het be-

drijven van armchair-philosophy, gebrek aan kennis van en belangstelling voor (causaliteit in de) wetenschappen: ‘*they continue the tradition in recent philosophy of science of avoiding the real scientific problems or even reporting on new frameworks -once radical- to solve them.*’ Hun doel is ‘*to provide and justify norms for science. The justification is supposed to be metaphysical.*’ Om vervolgens te besluiten met: ‘*Yes, I read the rest of the book, but you shouldn’t. The whole thing is a mixture of banality, light criticism of other philosophers about the disunity of science and plurality of causes...*’

Ontdoet men Glymours reactie van het (soms onnodige) polemische venijn, dan blijkt hij opmerkelijk genoeg allerm minst een aanhanger van de conflicthypothese. De sleutel ligt zijns inziens in het door Anjum en Mumford genegeerde werk van de kantiaanse filosoof Michael Friedman, die in *The dynamics of reason* (2001) evenals Gelman laat zien hoe kruisbestuiving wel mogelijk is. Dat dit allemaal geen sinecure is, bewijst ook *Causality: philosophical theory meets scientific practice* (Illary, 2014) dat enerzijds een doorwrochte studie vormt over de vele gezichten van de causaliteit en wel degelijk recht doet aan de rol ervan in de wetenschappen, maar anderzijds eveneens de succesvolle contemporaine statistische benaderingen van causaliteit buiten beschouwing laat. Helemaal lijkt de kilte van twee kanten te komen. In formele benaderingen is doorgaans weinig belangstelling voor de filosofische literatuur en met name voor de vele gezichten van causaliteit. Judea Pearl zet met *The book of why* (2018) weliswaar enige schreden op dit terrein, maar gaat voorbij aan het feit dat causaliteit dikwijls juist kwalitatief en deterministisch is, de *logic of justification* betreft en een pragmatisch, discourse- of taalhandelingsperspectief vraagt (Starmans, 2020). Hoe dan ook, indien er naast het dualisme van ‘fysische’ en ‘filosofische’ tijd, ook een schisma van ‘filosofische’ en ‘statistisch/wetenschappelijke’ causaliteit manifest wordt, is dat uiteraard moeilijk te rijmen met Gelmans hooggestemde ambities.

Epiloog

De fysisch en godsdienstfilosoof Ian Barbour publiceerde in 2000 het invloedrijke boek *When science meets religion; Enemies, strangers or partners*. De auteur was van mening dat de vele opvattingen over de relatie tussen wetenschap en geloof een ordenend, structurerend principe behoeften. Hij koos daartoe een viertal invalshoeken of ‘modellen’: achtereenvolgens het model van Conflict, van Integratie / Harmonie, van Onafhankelijkheid en van

Dialog, van waaruit hij alle thema’s en onderzoeksvragen van het boek stelselmatig benaderde. Dit eenvoudige kader is goed toepasbaar wanneer de wisselwerking tussen filosofie en afzonderlijke disciplines moeten worden geduid. We laten de relatie tussen natuurkunde en filosofie voor wat zij is en evenmin committeren we ons op voorhand aan één van Barbours modellen. Gelmans filosofische preoccupaties suggereren een mogelijkheid van Integratie/ Harmonie, de bloemlezing uit de tweede paragraaf toont toch tenminste een voortschrijdende samenspraak of dialoog; daartegenover lijkt het causaliteitsonderzoek de metafoor van conflict / schisma te voeden. Hoe dan ook, de relatie tussen statistiek en filosofie kent nog tal van uitdagingen of het nu gaat om big data, foutief gespecificeerde parametrische modellen, de erosie van het modelbegrip, de verzoening van statistical learning en machine learning, causale interpretaties van inferentiële statistiek, de oude vete tussen bayesianen en frequentisten, de replicatiecrisis en p-hacking of het vanuit de ethiek gevoede wantrouwen jegens statistiek en opake deep learning algoritmiek en de daarmee gepaard gaande roep om explainable en responsible datascience.

LITERATUUR

- Campo, A., & Gozzano, S. (2021). *Einstein versus Bergson. An enduring Quarrel on Time*. De Gruyter.
- Canales, J. (2015). The physicist and the philosopher: Einstein, Bergson and the debate that changed our understanding of time. Princeton University Press.
- Gelman, A. (2019). Many perspectives on Deborah Mayo’s ‘Statistical inference as severe testing: How to get beyond the statistics wars’. https://www.researchgate.net/publication/333309216_Many_perspectives_on_Deborah_Mayo%27s_Statistical_Inference_as_Severe_Testing_How_to_Get_Beyond_the_Statistics_Wars
- Starmans, R.J.C.M. (2018). The Predicament of Truth: on Statistics, Causality, Physics and the Philosophy of Science. In Mark J. Van der Laan & Sherri Rose (Eds.), *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*. Springer Series in Statistics, Springer.
- Starmans, R.J.C.M. (2020). Prometheus unbound or Paradise regained: the concept of causality in the contemporary AI-data science debate. *Journal of the French Statistical Society*, 161(1), 4–41.
- Starmans, R.J.C.M. (2022) Een Griekse Tragedie: Einstein, Bergson en de Tijd. *Filosofie-Tijdschrift*, 32(3), 40–45.

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

Politieke besliskunde

Toen een vriend van mij, afgestudeerd in de Internationale Betrekkingen, hoorde dat 'besliskunde' de Nederlandse vertaling van Operational Research is, was hij behoorlijk van slag. 'Hebben jullie de kunst van het maken van beslissingen nu ook al geclaimd?' In een wereld waar algoritmes een steeds prominentere rol spelen, was hij van mening dat de kunst van het maken van politieke besluiten toch nog wel echt aan de alfa's toebehoorde.

In theorie is politiek prima in een wiskundig model te gieten. Een politicus neemt beslissingen die aan bepaalde budgettaire of juridische voorwaarden moeten voldoen, met een zeker doel voor ogen. Het bepalen van het doel is waar voornamelijk de politieke voorkeur naar voren komt: een liberaal kiest voor het maximaliseren van het nut van de totale samenleving, terwijl een socialist het minimale nut zo hoog mogelijk wilt krijgen.

Maar gebruiken politici in de praktijk ook de theorie van de besliskunde? Laten we daarvoor als casus een politiek voorstel van afgelopen zomer nemen. Op 20 juli, terwijl heel Europa gebukt ging onder een hittegolf, presenteerde de Europese Commissie het plan Save Gas for a Safe Winter. Hierin wordt landen opgedragen om voor 31 maart 2023 15% van hun totale gasverbruik te vermindern. Als je tussen de regels van de nieuwsberichten leest, is het opvallend hoeveel verwijzingen er naar Operational Research zijn.

Om te beginnen het percentage van 15 procent, hoe

is de commissie daarop gekomen? Frans Timmermans, vicevoorzitter van de Europese Commissie, legt dat uit in de uitzending van *Nieuwsuur* van 20 juli: 'We hebben gekeken naar het gemiddelde gasverbruik over de laatste vijf jaar en dan hebben we gekeken naar wat we missen als de Russen helemaal geen gas meer leveren. Dan kom je uit op 15% en die 15% kunnen we dus maar het beste besparen, dan weten we zeker dat niemand in de kou hoeft te zitten en we onze industrie aan de gang houden.'

Dit verhaal klinkt logisch, maar als je er met een wiskundige blik naar kijkt is er nog wel het een en ander op aan te merken. De hoeveelheid beschikbaar gas als Rusland stopt met leveren, wordt door Timmermans gezien als een vast gegeven. Laten we voor het wiskundige perspectief s gebruiken om deze hoeveelheid aan te duiden. Verder is de vraag (d) is stochastisch en is de beslissing die genomen moet worden met welke factor (α) de vraag afgeschaald moet worden. We moeten er dus voor zorgen dat $(1-\alpha)d \leq s$. De doelfunctie wordt nergens expliciet gegeven, maar het beperken van de economische schade lijkt een logisch streven.

Het eerste zwakke punt in Timmermans' verhaal, is dat hij claimt dat we met een 15% reductie van het gasverbruik sowieso genoeg gas hebben om aan alle vraag te voldoen. Met een stochastische vraag is zulke zekerheid natuurlijk nooit te geven. Ten tweede, is het behoorlijk beperkt om slechts naar de afgelopen vijf jaar te kijken

om de kansverdeling van de vraag te bepalen. Tenslotte, resulteert het vervangen van de stochastische variabele d door zijn verwachting, zoals in het huidige plan is gedaan, over het algemeen niet tot de optimale oplossing. Een betere oplossing zouden we kunnen krijgen als we dit probleem zouden modelleren als een two-stage recourse problem, iets dat Timmermans zelf eigenlijk ook al suggereert: 'De keuze die we vandaag hebben is om nu al solidair te zijn, of te wachten tot het ons wordt opgedrongen – met veel hogere kosten.'

Hoe de landen 15% van hun gasverbruik gaan reduceren mogen ze zelf bepalen. Als gast in de podcast *Cafe Europa* bracht Diederik Samson, stafchef van Frans Timmermans, een mogelijke oplossing ter tafel. Sectors die veel gas verbruiken maar weinig economische waarde toevoegen zouden als eerste van het gas moeten worden afgesloten. Als we dit vraagstuk zien als een probleem waarin we zoveel mogelijk geld willen verdienen gegeven de beperkte gascapaciteit, is de analogie met een knapsack probleem snel gemaakt. Samsons oplossing zou dan ook de optimale oplossing zijn, die door het greedy algoritme voor dit probleem zou worden gegeven.

In het plan van de EU worden echter ook nog extra richtlijnen voor landen gegeven waar ze op zouden kunnen sturen. Zo zouden sectoren die cruciaal zijn voor EU-brede supply chains of installaties die ernstig beschadigd worden als ze niet worden gebruikt, gespaard moeten blijven. Dit zijn uiteraard simpele uitbreidingen voor het knapsack probleem. Ook zou men ervoor kunnen kiezen om elke sector nog wel een minimale hoeveelheid gas te geven zodat er nog wel een beperkte hoeveelheid economische activiteit kan plaatsvinden.

Het voorbeeld van Save Gas for a Safe Winter laat zien dat er zeker mogelijkheden zijn om besliskunde in de politieke besluitvormen toe te passen. Dat wordt in de praktijk echter slechts beperkt gedaan, dus daar is voor de wiskundigen nog veel te winnen. Waarschijnlijk is de eerste stap het onder de aandacht brengen van onze methoden bij politici. Een dialoog met beleidsmakers levert waarschijnlijk ook veel nieuwe interessante wiskundige problemen op, waarmee we besliskunde veel breder kunnen toepassen en samen met de alfa's politici kunnen helpen om optimale besluiten te nemen.

NOOT

1. Europese Commissie komt met controversieel plan om gas te besparen. NRC, 20 juli 2022

BERNARD ZWEERS, Doing The Math
E-mail: bernard@doingthemath.nl



UITMUNTENDE MASTER'S OF PH.D. THESIS BEGELEID?

OPROEP OM KANDIDATEN TE NOMINEREN
VOOR DE JAN HEMELRIJK
EN WILLEM R. VAN ZWET AWARDS 2022

Ter bekroning van een uitzonderlijke afstudeerprestatie aan een Nederlandse instelling voor wetenschappelijk onderwijs/hoger beroepsonderwijs looft de VVSOR al vanaf 1989 een scriptieprijs uit. In 2014 kreeg deze de naam *Jan Hemelrijk Award*. Sinds 2012 is er ook een prijs voor dissertaties: de *Willem R. van Zwet Award*.

De VVSOR roept op tot nominaties voor deze prijzen. Beide prijzen bestaan uit een oorkonde en een geldbedrag van 1.000 euro. De prijzen zullen worden uitgereikt tijdens de Annual Meeting van de VVSOR, op donderdag 23 maart 2023. Genomineerd kunnen worden personen die van september 2021 tot en met augustus 2022 zijn afgestudeerd respectievelijk gepromoveerd, dat wil zeggen gedurende het meest recente afgesloten academische jaar.

Hierbij worden supervisors (begeleiders) opgeroepen om een uitmuntende afstudeerscriptie (Master) of dissertatie (Ph.D.) te nomineren voor de Jan Hemelrijk dan wel Willem R. van Zwet Award 2022. In aanmerking voor de prijzen komen zowel originele theoretische bijdragen aan als inventieve toepassingen van bestaande theoretische concepten uit de statistiek en/of operations research.

De indiening van een nominatie dient vergezeld te gaan van een aanbevelingsbrief van de supervisor van de genomineerde. De precieze procedure voor beide prijzen, alsmede de reglementen en het nominatieformulier zijn te downloaden via de website van de VVSOR, www.vvsor.nl. De nominatie dient uiterlijk 22 januari 2023 binnen te zijn.

Namens de VVSOR,
Dr. Ad Ridder, juryvoorzitter Jan Hemelrijk Award
Prof. dr. Jelle Goeman, juryvoorzitter Willem R. van Zwet Award
Dr. Sander Scholtus, Secretaris der beide jury's

OPTIONAL STOPPING MET BAYES FACTORS mogelijkheden en beperkingen

Optional stopping betekent ‘naar de resultaten kijken om te beslissen of je nog wat data aan het experiment toe wilt voegen’. Vaak wordt er beweerd dat de Bayes factor voor hypothesesetsen geldig blijft tijdens optional stopping. Echter, dat hangt af van de specifieke situatie.

RIANNE DE HEIDE

Er zijn drie verschillende wiskundige definities te geven voor het fenomeen *optional stopping*. Of we in de praktijk kunnen zeggen ‘de Bayes factor methode voor hypothesesetsen blijft geldig tijdens optional stopping’ is een subtiele kwestie, die afhangt van de specifieke kenmerken van de gegeven situatie: welke modellen en welke priors worden gebruikt, en wat is het doel van de analyse.

De afgelopen jaren bleken de conclusies van papers verrassend vaak onjuist wanneer de experimenten werden gerepliceerd. Een deel van deze *replicability crisis* wordt veroorzaakt doordat in praktijk noodzakelijke voorwaarden voor het gebruik van de klassieke (frequentistische) methodes worden geschonden (John, Loewenstein & Prelec, 2012). Een van die voorwaarden is dat het experimentele protocol van te voren volledig moet zijn vastgelegd. In de praktijk passen onderzoekers het juist vaak aan, door onvoorziene omstandigheden, of ze verzamelen gegevens tot ze een bevredigend resultaat zien. Dit laatste wordt *optional stopping* genoemd, en kan er voor zorgen dat hypotheses veel vaker ten onrechte worden verworpen dan de foutgaranties van de statistische methodes beloven. Hypothesetoetsen met *Bayes factors* wordt al lang bepleit als een alternatief voor traditionele hypothesesetsmethodes dat verschillende problemen kan oplossen, en in het bijzonder werd al vroeg beweerd dat Bayesiaanse methodes geldig blijven tijdens optional stopping (Lindley, 1957; Raiffa & Schlaifer, 1961; Edwards, Lindman, & Savage, 1963). In het licht van de replicability crisis stonden deze claims kortgeleden opnieuw in de belangstelling (Wagenmakers, 2007; Rouder, 2014; Schönbrodt et al., 2017; Yu et al., 2014; Sanborn & Hills, 2014). Maar wat betekenen ze wiskundig? Het blijkt dat verschillende auteurs verschillende dingen bedoelen met ‘Bayesiaanse methodes kunnen omgaan met optional stopping’; en

bovendien blijken dergelijke claims vaak alleen geldig te zijn in informele zin of in beperkte situaties. In het paper (Hendriksen, De Heide & Grünwald, 2021) geven we een systematisch overzicht en een formalisering van dergelijke claims, en leggen hun relevantie voor de praktijk uit: kunnen we er op vertrouwen dat Bayes factor hypothesesetsen geldig blijven tijdens optional stopping, of niet? We zullen zien dat het antwoord genuanceerd ligt. Ten tweede breiden we het bereik van die claims uit tot meer algemene situaties, waarin ze nooit formeel zijn geverifieerd en waarvoor dat niet triviaal is. In het paper (De Heide & Grünwald, 2021) lichten we die claims vervolgens toe voor een publiek van methodologen en toegepaste statistici met behulp van computersimulaties.

Bayesiaans leren

Bayesianisme gaat over een bepaalde interpretatie van het begrip *waarschijnlijkheid*: als *geloofsgraden*. Een Bayesiaan begint met het uitdrukken van dit geloof als een waarschijnlijkheidsfunctie. We noemen dit de *prior*, en we duiden het aan met $\mathbb{P}(\theta)$, waarbij θ de parameter (of meerdere parameters) van het model is. Na specificatie van de prior, verkrijgen we de data D en likelihood $\mathbb{P}(D|\theta)$. Nu kunnen we de posterior $\mathbb{P}(\theta|D)$ berekenen met behulp van de stelling van Bayes:

$$\mathbb{P}(\theta|D) = \frac{\mathbb{P}(D|\theta)\mathbb{P}(\theta)}{\mathbb{P}(D)}.$$

Stel dat we een nulhypothese $\mathcal{H}_0$ willen toetsen tegenover een alternatieve hypothese $\mathcal{H}_1$. We kunnen dit op een Bayesiaanse manier doen met *Bayes factors*: we beginnen met de *prior odds* $\mathbb{P}(\mathcal{H}_1)/\mathbb{P}(\mathcal{H}_0)$, ons geloof voordat we de data zien. Vaak geloven we dat beide hypotheses even waar-

schijnlijk zijn, en dan zijn onze prior odds 1 staat tot 1. Daarna verzamelen we de data D , en we werken onze odds bij met onze nieuwe kennis, met behulp van de stelling van Bayes:

$$\text{posterior odds } (\mathcal{H}_1 \text{ vs. } \mathcal{H}_0) = \frac{\mathbb{P}(\mathcal{H}_1|D)}{\mathbb{P}(\mathcal{H}_0|D)} = \frac{\mathbb{P}(\mathcal{H}_1)}{\mathbb{P}(\mathcal{H}_0)} \frac{\mathbb{P}(D|\mathcal{H}_1)}{\mathbb{P}(D|\mathcal{H}_0)}.$$

De posterior odds is onze bijgewerkte overtuiging over welke hypothese waarschijnlijker is.

Optional stopping: drie begrippen

Een wenselijke eigenschap van hypothesesetsen is geldigheid tijdens optional stopping: we verzamelen wat gegevens, we kijken naar de tussenresultaten, en besluiten op basis daarvan of we stoppen of nog even doorgaan en nog wat meer data verzamelen. Informeel noemen we het ‘gluren naar de data tot dusver om te beslissen of we nog meer data gaan verzamelen of niet’ *optional stopping*, maar als we het wat preciezer willen maken wat het nou echt betekent als een toets geldig blijft tijdens optional stopping, blijkt het dat verschillende benaderingen (frequentistisch, subjectief en objectief Bayesiaans) leiden tot verschillende interpretaties en definities. Het blijkt dat we drie verschillende wiskundige concepten kunnen onderscheiden, die we identificeren en formeel definiëren in het paper (Hendriksen, De Heide & Grünwald, 2021).

Subjectief Bayesiaans optional stopping

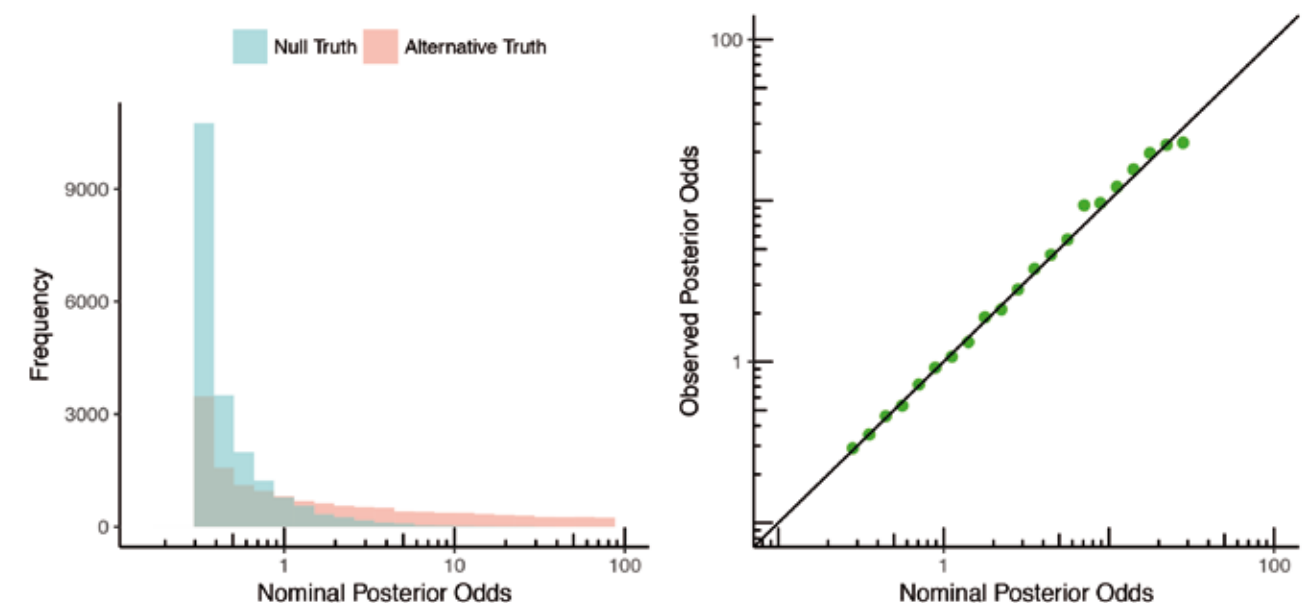
Het eerste begrip noemen we *subjectief Bayesiaans optional stopping* oftewel *onafhankelijkheid*. Als je het puur subjectief Bayesiaans beschouwt — alleen als je werkelijk in je prior gelooft — dan wordt Bayesiaans updaten van prior naar posterior niet beïnvloed door de gebruikte stopregel: je eindigt met dezelfde posterior als je de steekproefgrootte n van te voren had vastgelegd, of als hij bijvoorbeeld was bepaald omdat je het resultaat na n datapunten goed genoeg vond. In die zin is een subjectief Bayesiaanse procedure niet afhankelijk van de stopregel.

Kalibratie

De tweede vorm van optional stopping noemen we *kalibratie*. Zoals (Rouder, 2014) schrijft: ‘Als een herhalingsexperiment een posteriorkans van 3,5 staat tot 1 oplevert in het voordeel van de nulhypothese, dan verwachten we dat de data 3,5 keer zo waarschijnlijk door de nulhypothese zijn geproduceerd, als door de alternatieve hypothese.’ In meer wiskundige taal kan dit worden uitgedrukt als:

$$\text{post-odds } (\mathcal{H}_1 \text{ vs. } \mathcal{H}_0 | \text{“post-odds } (\mathcal{H}_1 \text{ vs. } \mathcal{H}_0 | D) = a\text{”}) = a.$$

We zeggen dat deze vergelijking *kalibratie van de posterior odds* uitdrukt. Het blijkt dat deze kalibratie niet standhoudt als je er geen puur subjectieve visie van Bayesianisme op nahoudt, en in het bijzonder geldt dit niet voor



Figuur 1. Posterior odds in een experiment waarin we toetsen of de verwachtingswaarde van een normale verdeling 0 is ($\mathcal{H}_0$), versus niet-nul ($\mathcal{H}_1$), van 20:000 experimenten. (a) de empirische steekproevenverdeling als histogram onder $\mathcal{H}_0$ (blauw) en $\mathcal{H}_1$ (roze). (b) Kalibratieplot: de geobserveerde posterior odds als een functie van de nominal posterior odds

de *default* priors die enkele Bayesianen in de psychologie bepleiten (Wagenmakers, 2007; Rouder et al., 2012). Om een eerste idee van de problemen hieromtrent te krijgen: default priors zijn soms afhankelijk van de data. In dat geval is het onduidelijk wat *optional stopping* nou eigenlijk betekent, omdat als je een prior $P_1(\theta)$ gebruikt die gebaseerd is op een steekproefgrootte van n , maar je bent gestopt bij $n' < n$, dan had je eigenlijk prior $P'_1(\theta)$ gebaseerd op n' moeten gebruiken... maar dan zou je kunnen stoppen bij $n'' < n'$, etc. Zie ons paper (De Heide & Grünwald, 2021) voor een uitgebreide discussie en veel voorbeelden.

Frequentistisch optional stopping

De derde vorm is een frequentistische interpretatie van het omgaan met optional stopping, die gaat over het controleren van de Type-I-fout van een experiment. Een Type-I-fout treedt op als we de nulhypothese verwerpen wanneer deze waar is, ook wel een *vals positief* genoemd. De frequentistische interpretatie van geldigheid tijdens optional stopping is dat de Type-I-foutgarantie nog steeds geldt als we de onderzoeksopzet — en dus de stopregel — niet van tevoren vastleggen, maar we mogen stoppen als we een significant resultaat zien. In het geval dat $\mathcal{H}_0$ enkelvoudig is (bestaat uit slechts één hypothese), is er een bekend verband tussen Bayes factors en Type-I-foutwaarschijnlijkheden: als we $\mathcal{H}_0$ verwerpen dan en slechts dan als de posterior odds in het voordeel van $\mathcal{H}_0$ kleiner zijn dan een vast niveau α , dan hebben we een garantie dat onze Type-I-fout ten hoogste α is. En interessant genoeg geldt dit niet alleen voor een vaste steekproefgrootte, maar ook tijdens optional stopping. Echter, voor *meervoudige* $\mathcal{H}_0$, geldt dit verband niet. Behalve in het speciale geval waar *alle* vrije parameters in $\mathcal{H}_0$ zogenoemde *nuisance parameters* zijn, die een groepsstructuur bezitten en die zijn uitgerust met de bijbehorende right-Haar prior, en worden gedeeld met $\mathcal{H}_1$, zoals we bewijzen in (Hendriksen, De Heide & Grünwald, 2021). Maar voor algemene priors en voor meervoudige $\mathcal{H}_0$ is dit meestal niet het geval.

Conclusie

Er zijn drie verschillende wiskundige definities te geven aan het fenomeen *optional stopping*. Of we in de praktijk kunnen zeggen ‘de Bayes factor methode voor hypothesetoetsen blijft geldig tijdens optional stopping’ is een

subtiele kwestie, die afhangt van de specifieke kenmerken van de gegeven situatie: welke modellen en welke priors worden gebruikt, en wat is het doel van de analyse.

LITERATUUR

- Edwards, W., Lindman, H., & Savage, L.J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70(3), 193–242. <https://doi.org/10.1037/h0044139>.
- De Heide, R., & Grünwald, P.D. 2021. Why optional stopping can be a problem for Bayesians. *Psychonomic Bulletin & Review*, 28(3), 795–812.
- Hendriksen, A., de Heide, R., & Grünwald, P. 2021. Optional stopping with Bayes factors: a categorization and extension of folklore results, with an application to invariant situations. *Bayesian Analysis*, 16(3), 961–989.
- John, L.K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*.
- Lindley, D.V. (1957). A statistical paradox. *Biometrika*, 44(1/2), 187–92. <https://doi.org/10.2307/2333251>.
- Raiffa, H., & Schlaifer, R. (1961). *Applied statistical decision theory*. Harvard University Press.
- Rouder, J.N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, 21(2), 301–308. <https://doi.org/10.3758/s13423-014-0595-4>.
- Rouder, J.N., Morey, R.D., Speckman, P.L., & Province, J.M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56(5), 356–374.
- Sanborn, A.N., & Hills, T.T. (2014). The frequentist implications of optional stopping on Bayesian hypothesis tests. *Psychonomic Bulletin & Review*, 21(2), 283–300. <https://doi.org/10.3758/s13423-013-0518-9>.
- Schönbrodt, F.D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, 22(2), 322–339. <https://doi.org/10.1037/met0000061>.
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of P values. *Psychonomic Bulletin & Review*, 14(5), 779–804. <https://doi.org/10.3758/bf03194105>.
- Yu, E.C., Sprenger, A.M., Thomas, R.P., & Dougherty, M.R. (2014). When decision heuristics and science collide. *Psychonomic Bulletin & Review*, 21(2), 268–282. <https://doi.org/10.3758/s13423-013-0495-z>.

RIANNE DE HEIDE werkt als universitair docent bij de afdeling wiskunde van de Vrije Universiteit Amsterdam. Ze is geïnteresseerd in de wiskundige en filosofische grondslagen van de statistiek en machine learning. Haar onderzoek betreft in het bijzonder (sequentieel) hypothesetoetsen, Bayesiaans leren, banditproblemen, en sinds kort ook explainable AI. Ze won in 2022 de Willem R. Van Zwet Award voor haar proefschrift *Bayesian learning: challenges, limitations and pragmatics*. Website: <https://riannedeheide.github.io>
E-mail: r.de.heide@vu.nl

De Hoofdige Boer

‘Elk weet, waar ‘t Almensch Kerkje staat, en kent de laan, die derwaart gaat.’ Zo begint het gedicht van Staring over de Hoofdige Boer. De weg naar dat kerkje werd doorsneden door een modderige ondiepe voorde die men moest oversteken om ter kerke te kunnen gaan. Dat gaf vaak schade aan kleding en schoeisel, daarom werd op aandringen van de predikant een brug gebouwd. Iedereen blij, behalve Scholte Stuggink die het niet op die nieuwlichterij had voorzien. Hij bleef, voorzien van lieslaarzen, de oude doorsteek gebruiken. Hij was koppig, hoofdfig in het Nederlands van de 19e eeuw, en bleef doen zoals zijn voorouders ook altijd hadden gedaan.

Aan die Hoofdige Boer moest ik denken toen ik problemen had met een softwareleverancier en koppig voet bij stuk hield. Tenslotte ben ik ook een Achterhoeker, net als Starings Scholte Stuggink.

Ik loop al heel lang mee in dit vak, en dan met name in de gegevensverwerking. Daar heeft zich in de loop der jaren een heilvolle standaardisatie voltrokken. Niet alleen heilvol omdat daarmee het gemak voor de gebruiker toeneemt, maar vooral omdat er fouten worden vermeden.

De rekencentra van Nederlandse universiteiten hadden begin jaren tachtig vorige eeuw verschillende merken en types computer in gebruik. Toen ik in 1982 bij het Rekencentrum van de toenmalige Landbouwhogeschool ging werken was daar een DEC-10 systeem in gebruik. Dat was in zoverre een afwijkend concept omdat men niet de gebruikelijke 32-bits woordlengte hanteerde, maar 36-bits. Dat had voordelen voor de rekennauwkeurigheid, van een getal konden er iets meer decimalen worden opgeslagen. Onder andere Tilburg, Twente en Rotterdam hadden dit ook, of de iets grotere versie DEC-20. Verder waren er 32-bits IBM-systemen in Delft, Nijmegen en Leiden en 60-bits CDC-systemen in Amsterdam, Utrecht en Groningen en nog wat minder gangbare merken. PC's waren er nog niet, die kwamen pas enkele jaren later.

Een programma dat ik graag in Wageningen wilde hebben was LISREL, voor Linear Structural Relations, geschreven door Karl Jöreskog. De distributie daarvan werd geregeld via een Amerikaans bedrijf. Kleinere organisaties hadden niet zelf de mogelijkheid een product over te zetten naar andere systemen, dat besteedden ze uit aan bijvoorbeeld een bevriende universiteit die over het gewenste computersysteem beschikte. Zo kwam de DEC-10

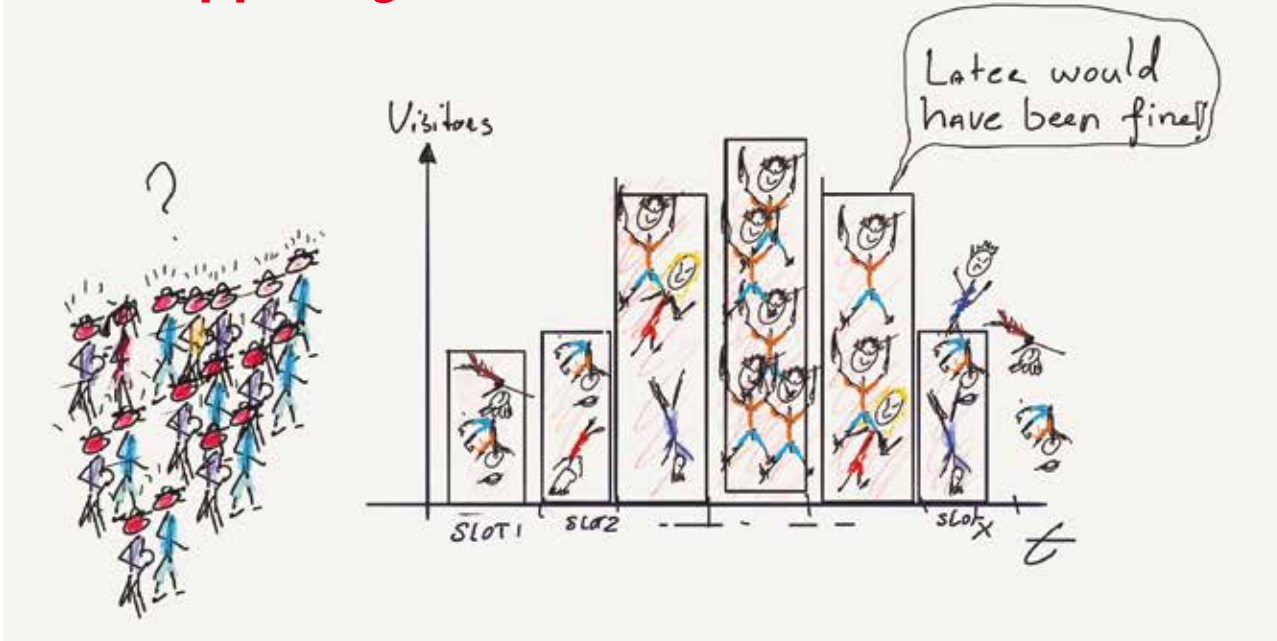
versie van LISREL via een Australische universiteit. Wat er precies is misgegaan met die conversie is me nooit duidelijk geworden, maar ik kreeg in Wageningen afwijkende uitkomsten van de bijgeleverde set testen. Op een dag probeerde ik het nog eens en plotseling was het wél goed. Na enige tijd van alles te hebben gecontroleerd en vergeleken ontdekte ik dat de juiste uitkomsten alleen werden geproduceerd als ik de geheugencapaciteit voor mijn job klein hield. Om sneller te kunnen werken met zo'n gecompliceerd programma had ik de eerste keer gekozen voor meer geheugen. Ik vond dat ik het programma op deze manier niet aan de gebruikers beschikbaar mocht stellen, het moest gewoon onder alle geheugenpartities de juiste uitkomsten geven. De Amerikaanse distributeur wist geen raad met het probleem, en de Australische universiteit was van mening dat het correct was, ik moest de gebruikers maar verplichten met weinig geheugen te werken. Ik heb het probleem toen voorgelegd aan Jöreskog die mij steunde in mijn besluit het programma in deze staat niet te accepteren. Gelukkig hadden we de licentie nog niet betaald, ik heb het contract *on hold* gezet totdat er een oplossing kwam.

Helaas bleef de distributeur me bestoken met herinneringsnota's en dreigde met advocaten. Op een gegeven ogenblik werd ik gebeld door juristen van het Ministerie van Landbouw, zij hadden van de Nederlandse ambassade in de VS het verzoek gekregen die vervelende man bij de Agricultural University tot enige spoed te manen. Ik voelde me in mijn koppigheid zo langzamerhand een echte Hoofdige Boer en heb ze uitgelegd dat ik niet van plan was ook maar iets te betalen voor een niet correct werkend product. Dat de oorspronkelijke maker van dat product me hierin steunde gaf de doorslag. Ik heb alle ontvangen materiaal zoals tapes en manuals naar het ministerie gestuurd die het in een *diplomatic pouch* naar Amerika verzond alwaar de Landbouwwattaché het overhandigde aan de distributeur.

Weken commotie, ergernis en veel vergeefs werk, alleen omdat wij op een ander computersysteem werkten dan de oorspronkelijke maker. Gelukkig liggen die ruige pioniersjaren inmiddels achter ons.

GERRIT STEMERDINK is eindredacteur van *STATOR*.
E-mail: gjstemerding@hotmail.com

Tourist trip planning



BEZOEKERS SPREIDEN MET BEHULP VAN SLIMME ALGORITMES

Op drukke bestemmingen kan het plezier van een dagje uit belemmerd worden door lange wachtrijen en grote drukte. Hierdoor kunnen mensen veel minder doen dan ze oorspronkelijk van plan waren. Het coördineren van bezoekers op basis van hun persoonlijke voorkeuren kan helpen dit effect te verminderen of weg te nemen. In dit artikel schetsen we een methode om deze coördinatie slim te bewerkstelligen zodat op drukke bestemmingen de bezoekerstevredenheid aanzienlijk verbeterd kan worden.

JORIS SLOOTWEG, ROB VAN DER MEI & FRANK OTTENHOF

Een dagje Amsterdam of pretpark kan snel verpest worden doordat je continu in de rij staat of veel activiteiten volgeboekt zijn. Daardoor kunnen mensen veel minder activiteiten ondernemen dan ze oorspronkelijk van plan waren. Buiten een klein groepje vroege boekers, moet bijna iedereen ofwel lang in de rij staan, of 's ochtends vroeg bij het opengaan van de laatste boekingsloten nog proberen een plekje te bemachtigen. Met slimme coördinatie en spreiding kan zowel de gemiddelde bezoekerstevredenheid als het gemak verbeterd worden. Hierbij is het wel noodzakelijk om specifiek rekening te houden met de diversiteit in wensen en eisen van de bezoekers. The

Driving Force ontwikkelt hiervoor het SuperLinda platform waarop bezoekers geholpen worden een dagje uit te plannen. Een ander doel van SuperLinda is om bezoekers goed te spreiden.

Laten we als voorbeeld een dagje Amsterdam nemen. Ruim voor vertrek gaat iemand zich oriënteren. Op dit punt kan het platform al helpen met spreiden door leuke alternatieve activiteiten voor drukke highlights en relatief rustige dagen voor te stellen. Uiteindelijk maakt de bezoeker een keuze, maar zelfs dan is het altijd een hoop werk om een leuk programma voor een dag of weekendje samen te stellen. Daarnaast is het voor sommige activi-

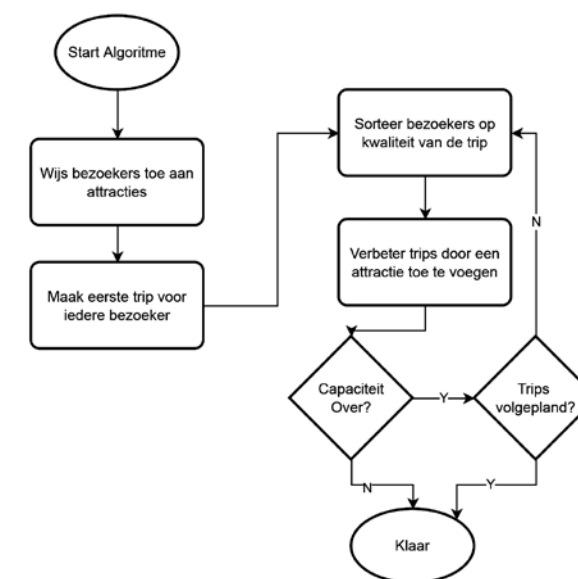
teiten alleen mogelijk ze te boeken als je meteen bij het openen van het slot reserveert. Een algoritme dat dit voor je doet of je ermee helpt op basis van jouw voorkeuren kan dan prettig zijn. Dit probleem wordt onderzocht in veel varianten als het Tourist Trip Design problem.

Deze algoritmes kunnen bezoekers helpen een dagje uit te plannen. Helaas houden ze in het beste geval hoogstens rekening met de verwachte drukte op een locatie. Dit terwijl het op drukke bestemmingen juist noodzaak is om tijdens het maken van routes deze coördinatie en spreiding meteen mee te nemen. Daarnaast zie je dat als men al probeert meerdere routes gecoördineerd te creëren, de algoritmes vaak maar werken voor enkele honderden bezoekers. Dit terwijl juist in de locaties waar coördinatie relevant is er vaak miljoenen bezoekers per jaar zijn. Algoritmes moeten dus in staat zijn om snel oplossingen uit te rekenen voor tienduizenden bezoekers op dagelijkse basis.

Gecoördineerde reisplannen aanmaken

Wij stellen een algoritme voor dat op basis van opgegeven voorkeuren gecentraliseerd een reisplan maakt. Dus iedere bezoeker geeft van tevoren aan hoe hij de activiteiten waardeert, in het geval van SuperLinda via een gesorteerd wensenlijstje. Deze wensenlijst wordt voor het algoritme aangepast naar een score per activiteit. Daarnaast geeft een bezoeker aan hoe groot de groep is waar ze mee reizen, de gewenste vorm van transport en de gewenste aankomst- en vertrektijd. Op basis van deze informatie rekent het algoritme een schema uit voor alle bezoekers.

Zoals veel problemen met routing is ook dit probleem optimaal op te lossen met Integer Linear Programming (ILP). Je voegt routing constraints toe en tijdsgebonden capaciteit constraints op de activiteiten en optimaliseert de totale som van de beloningen die bezoekers opstrijken door activiteiten te bezoeken. Dit leidt gegarandeerd tot een optimale oplossing, helaas zorgen met name de routing constraints ervoor dat het probleem al snel praktisch onoplosbaar wordt door de benodigde rekentijd. De grootste instantie die wij op deze manier opgelost hebben bestaat uit 80 toeristen met slechts 5 locaties. Dit is natuurlijk een verre van realistische probleemgrootte. Een andere manier om het probleem op te lossen is door iedere bezoeker één voor één een optimale tour te bieden volgens het First Come First Served principe. De rekentijd neemt aanzienlijk af doordat je de routes één voor één uit kan rekenen. Helaas gaat met deze methode de gemiddelde tevredenheid fors achteruit, doordat de vroeg-



Figuur 1. Een stroomdiagram van ons voorgestelde algoritme

komers schaarse capaciteit bezet houden die veel beter aan anderen gegeven had kunnen worden.

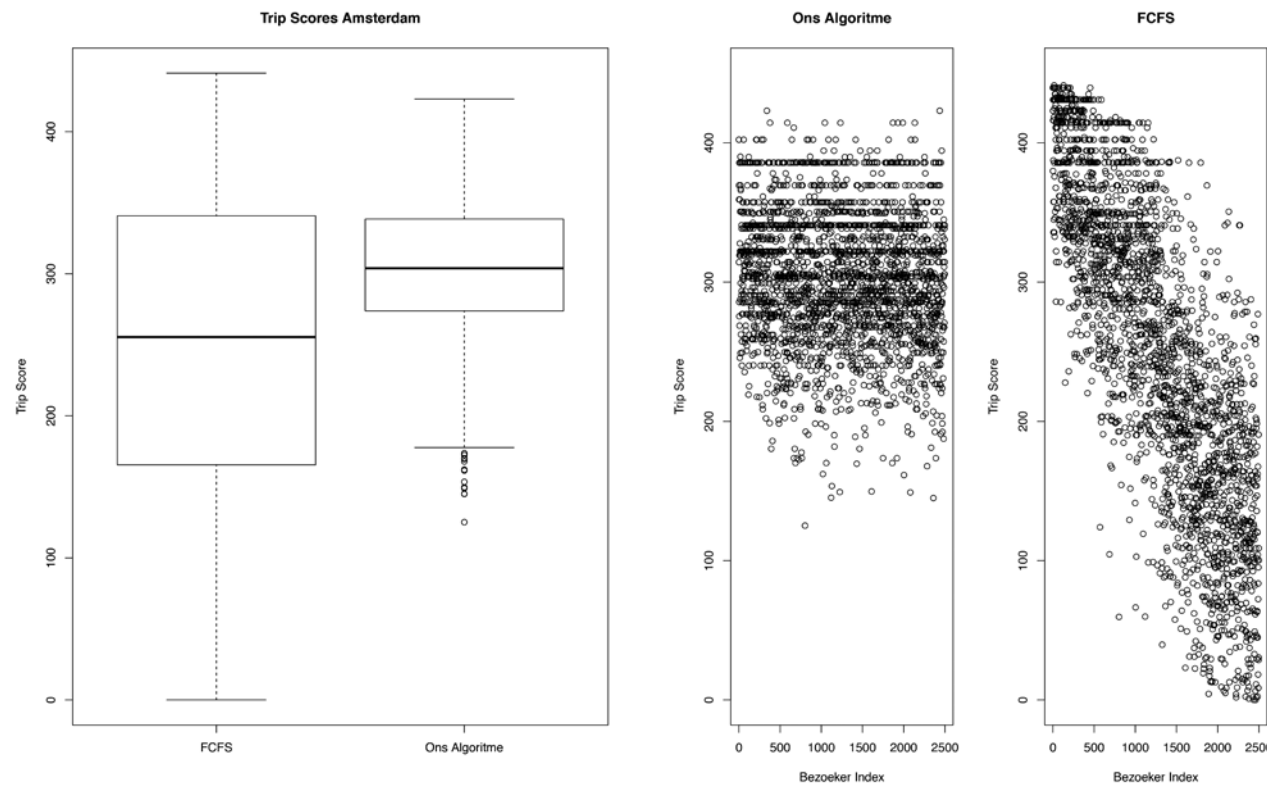
Naar een schaalbaar algoritme

Vanuit dit inzicht stellen wij een algoritme voor dat snel en schaalbaar is. Het algoritme bestaat uit 5 stappen en in figuur 1 tonen wij het als stroomdiagram.

Stap 1 is bezoekers toewijzen aan activiteiten op basis van de dagelijkse capaciteit. Doordat problemen als slotkeuze en routing hierbij wegvallen, is dit Integer Linear Program prima op te lossen voor realistische probleeminstaties. Doordat er geen rekening gehouden wordt met spreiding op de dag of de reisafstanden kan het gebeuren dat sommige toeristen niet alles kunnen doen wat aan hen toegewezen wordt.

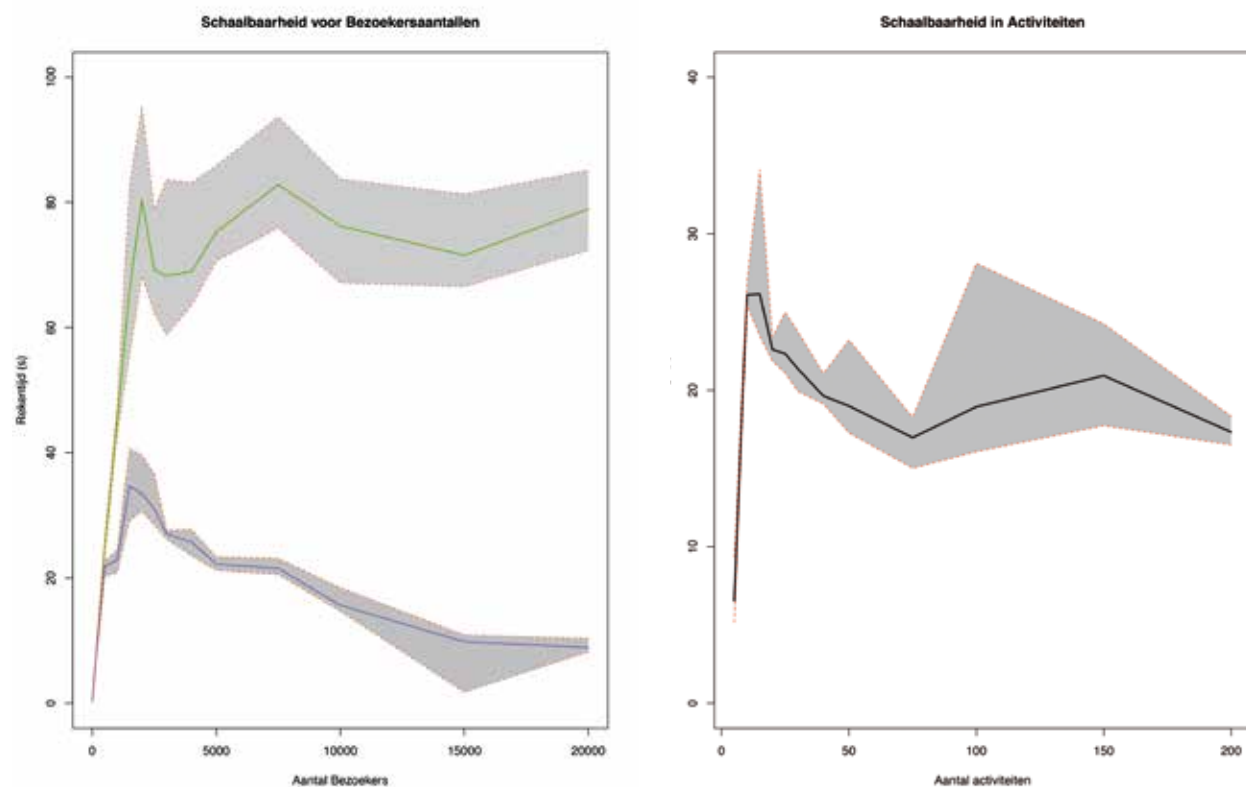
Stap 2 maak voor iedere toerist een route waarbij de beschikbare activiteiten beperkt zijn tot de set van activiteiten die hem of haar is toegewezen. Aangezien deze set van activiteiten beperkt is kan dit routing probleem per bezoeker heel snel opgelost worden. Indien sommige toegewezen activiteiten niet in het programma van de bezoeker passen worden ze op dit moment niet ingepland. De vrijgekomen capaciteit kan in latere stappen van het algoritme door andere bezoekers gebruikt worden.

Stap 3 sorteer de gebruikers oplopend op basis van de trip score. Hierdoor wordt in de volgende verbeterstap prioriteit gegeven aan bezoekers die de minste activiteiten van hun wensenlijst kunnen doen.



Figuur 2. Een boxplot die de verdeling van de tevredenheid uitgedrukt in trip score van bezoekers voor FCFS en ons algoritme weergeeft

Figuur 3. Een plot voor de tevredenheid van iedere bezoeker waarbij de index de volgorde van aankomst aangeeft



Figuur 4. De rekentijd als het aantal beschikbare activiteiten of het aantal gebruikers toeneemt. Opvallend is dat de rekentijd bij extreme bezoekersaantallen zelfs afneemt door de zware versimpeling van het routeringsvraagstuk

Stap 4 probeer de trips van de gebruikers uit te breiden met een extra activiteit die in hun top 3 activiteiten zit. Bepaal voor die gebruiker ook meteen of zijn trip volgeboekt is, zodat indien er realistisch gezien geen activiteiten bijgeboekt kunnen worden deze gebruiker niet meer meegenomen wordt in de vervolgstappen. Stap 5 kijk of er nog gebruikers zijn wier trip uitgebreid kan worden en of er nog capaciteit op de activiteiten is. Indien beide het geval is herhalen we alles vanaf stap 3.

De resultaten

De gemiddelde tevredenheid gaat aanzienlijk omhoog ten opzichte van FCFS. Daarnaast wordt ook de spreiding van de tevredenheid veel minder omdat het algoritme de capaciteit eerlijk over de bezoekers verdeelt zoals getoond in figuur 2. Dit komt vooral omdat voorkomen wordt dat de eerste gebruikers de rest blokkeren, zoals goed te zien is in figuur 3. Daarnaast blijkt dat op kleine instanties het gat met de optimale oplossing beperkt blijft tot ongeveer 5%. Doordat het voorgestelde algoritme naast goede oplossingen ook snel werkt is het ook in staat om in geval van noodzakelijke aanpassingen gedurende dag, snel een nieuwe oplossing te berekenen. Dit stelt bijvoorbeeld pretparken ook in staat om in geval van incidenten snel te kunnen schakelen en de negatieve impact op de bezoekerstevredenheid van bijvoorbeeld ongepland onderhoud zoveel mogelijk te beperken.

Deze methode leidt tot goede resultaten en is daarnaast goed schaalbaar. Hij profiteert optimaal van de diversiteit in voorkeuren die bezoekers in deze setting hebben. Daarnaast wordt het complexe routing probleem in de stappen van dit algoritme bijna triviaal gemaakt doordat de routes continu extreem kort zijn en daardoor simpel optimaal lokaal op te lossen. Verder blijkt dat toename in het aantal activiteiten of bezoekers slechts beperkt tot een toename van de rekentijd leidt zoals getoond wordt in figuur 4. Dit doordat er in het probleem voor de rekentijd twee natuurlijke limieten zijn: het aantal activiteiten dat een bezoeker kan ondernemen en de capaciteit van de activiteiten. Sterker nog: bij extreme drukte wordt de rekentijd zelfs lager omdat

de schema's door de beperkte capaciteit extreem kort worden.

Discussie

Het spreiden van bezoekers met behulp van coördinatie op basis van persoonlijke voorkeuren lijkt tot aanzienlijke verbeteringen van bezoekerstevredenheid te kunnen leiden. Praktijkttests zijn nodig om de algoritmes en de waardering van trips goed te evalueren. Daarnaast zijn er nog meerdere interessante varianten te bestuderen, zoals bijvoorbeeld online varianten van ons voorgesteld algoritme waarbij gespreid wordt op basis van verwachte klantenstromen.

JORIS SLOOTWEG werkt sinds 2020 bij the Driving Force en is PhD student bij het CWI/VU. Zijn onderzoek is deel van het COCONUT project dat zich tot doel stelt knelpunten in de infrastructuur op te lossen of te verbeteren door slimme algoritmes op basis van coördinatie en personalisatie. Infrastructuur wordt in dit onderzoek zeer breed opgevat en bevat bijvoorbeeld terreinen als energie, toerisme, luchthavens, wegen en openbaar vervoer.
E-mail: slootweg@cw.nl

ROB VAN DER MEI is als hoogleraar Toegepaste Wiskunde en Manager Research and Strategy verbonden aan het Centrum Wiskunde & Informatica en de Vrije Universiteit Amsterdam. Hij richt zich met name op het modelleren en optimaliseren van logistieke processen in de praktijk, en slaat daarmee een belangrijke brug tussen theorie en toepassing. In 2021 ontving hij voor zijn werk samen met collega Sandjai Bhulai de prestigieuze Huibregtsprijs. De laatste jaren is hij vooral werkzaam op het gebied van ambulance- en brandweerplanning, het reduceren van wachtlijsten in de ouderenzorg, en het optimaliseren van personenmobiliteit en vrachtlogistiek. Hij is tevens projectleider van het COCONUT project.
E-mail: mei@cw.nl

FRANK OTTENHOF is directeur van the Driving Force. Van origine werd bij the Driving Force verkeersmanagementsoftware ontwikkeld. Recent is dit uitgebreid met software om toerisme en boerderijwinkels te managen. Hij is betrokken bij het COCONUT project vanwege zijn domeinkennis vanuit de infrastructuur. Ook focust hij zich op de implementeerbaarheid van het onderzoek.
E-mail: frank.ottenhof@thedrivingforce.nl



HYPERBOLEN 'TUSSN DE DUUMN'

De tuimeling in een vijf meter diep Zuid-Frans ravijn, fiets en lijf ongedeerd, brachten mijn gedachten deze zomer terug naar een jeugdig trauma van te vaak overdrijven en daardoor te vaak niet te worden geloofd als het wel echt was gebeurd. Tamelijk vernederend was dan het gebaar met de beide handen van de toehoorder met een mompelend 'tussn de duumn zeker'.

In tijden van dikke duimen en nepnieuws werd ik deze zomer dus gewelddadig herinnerd aan een traumatisch, maar toch ook erg grappig gebaar met de duim en de wijsvinger. In het gezin waarin ik opgroeide werd de uitdrukking 'tussn de duumn' veelvuldig non-verbaal gehanteerd bij vermeende overdrijvingen van voorgedane gebeurtenissen, vooral als het getallen betrof. 'Tussn de duumn', Nedersaksisch voor 'tussen de duimen' is een gebaar van beide handen, waarbij de afstand tussen de twee duimen, naar elkaar gericht, beduidend kleiner is dan die tussen de beide wijsvingers die elk een hoek van 90 graden maken met hun aanpalende duim. Het verhaal met de overschatte aantallen lag dan 'tussen de beide wijsvingers', het werkelijke aantal lag 'tussn de duumn'. Overdrijven en chargeren was mijn probate pubermiddel om aandacht te trekken en te krijgen. Maar het veelvuldige gebruik van ongepaste hyperbolen leidde te vaak ook tot het tussn-de-duumngebaar van de toehoorder, terwijl ik de waarheid en niets anders dan ...

Hoe dat zat. Eerst iets meer achtergrond. Ik kom uit een groot, naoorlogs, gereformeerd gezin met tien kinderen.

Op twee na zitten of zaten ze in het onderwijs. Vader, moeder, pake (inderdaad, Friese roots) en ooms, allemaal schoolmeesters. Mijn vader was 'directeur der Gereformeerde School voor MULO' (later MAVO) en part-time leraar aan het Gereformeerd Lyceum, beide te Kampen. In ons huis aan de stadsgracht die als open riool de passende naam Burgel heeft, had hij een rokerig studeerkamertje voor zichzelf. In de kasten, tot aan het plafond, overheersten de donkerrode ruggen van de boeken van de wiskundigen P. Wijdenes en F. Schuh en alle edities – tot dan – van het *Nieuw Tijdschrift voor Wiskunde*. De reeds rugloze *Groote Tafel in Vijf Decimalen* van J. Versluys en P. Wijdenes (Noordhoff N.V., 1936) heb ik in mijn studententijd tussen de schone sokken en onderbroeken meegesmokkeld naar Groningen, alwaar ik inmiddels wiskunde moest (sic!) gaan studeren. Dat laatste was inderdaad met tamelijk veel tegenzin. Als puber lagen mijn ambities elders: bruggen en torens bouwen, architect, water- en wegenbouw in Delft. Zoiets toen.

Om je overeind te houden in een groot gezin zijn vanzelfsprekend daadkracht en manifestatiedrang basisvereisten, hoewel ik hier niet moet overdrijven over de grootte van het gezin. Toen ik op 17-jarige leeftijd 'de deur uitging' was het aantal competitieve broers slechts twee. De nummer vier in de rij van tien was vijf jaar jonger dan ik en telde dus niet mee. Ik heb slechts twee keer meegeemaakt dat de oudste (ik scheel –13 maanden met hem) de derde (met wie ik +13 maanden scheel) een tand door de lip sloeg en hem ook een keer van een lichte bloedneus heeft voorzien. Mij heeft hij alleen een keer met een

hamer bedreigd omdat ik het vertikte de afwas te doen terwijl hijzelf aan de beurt was. Die poging tot doodslag (een hyperbole volgens de Engelsen, voor de kegelsnede hebben ze het woord *hyperbola*) heb ik ruimschoots overleefd. Opgroeien als de nummer twee in een door het mannelijk geslacht gedomineerd gezin, was vanaf mijn wieg al geen sinecure. Alle cruciale stappen in mijn ontwikkeling waren suffe herhalingen van de dertien maanden eerdere oohs en aahs voor de nummer één. Meer en meer heb ik me teruggetrokken en me alleen in onweersaanbare situaties opzichtig gemanifesteerd. Op het Gereformeerd Lyceum te Kampen koos ik mijn momenten. We hadden daar les van een, voor mij, inspirerende leraar wiskunde met, voor de andere klasgenoten, de bijpassende naam dr. Dorleijn. Orde houden was niet zijn sterkste punt; 's ochtends met het scheerschuim om de mond de klas binnenkomen hielp niet. Maar inspirerend was hij wel, zeker voor mij. Vooral de kegelsneden wist hij boeiend tevoorschijn te toveren. In de pauze vertelde hij me meer over hyperbolen en assenstelsels: De beide assen, tezamen met de hyperbool, zijn te vangen in de compacte binaire formulering: $xy \in \{0,1\}$, waarbij $xy = 0$ de x- en de y-as representeren. De laatste x en y moesten wel met een hoofdletter worden geschreven, maar dat vond ik toen al onzin. Ik vond dat allemaal super interessant. De hyperbool met de coördinaatassen als asymptoten werd daarna mijn contemplatieve reddingsboei in onbegrepen puberuren. Urenlang kon ik de hyperboolkromme in gedachten volgen richting zijn asymptoten en indenken dat in het oneindige 'hiernamaals' de beide losse delen van de hyperbool zich verenigden.

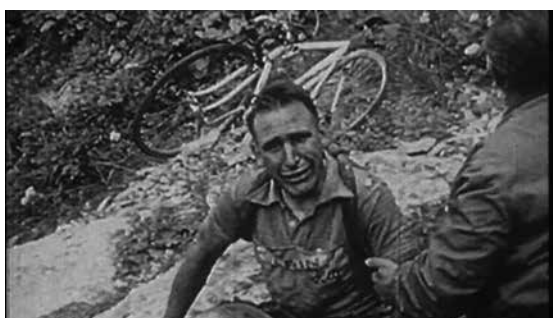
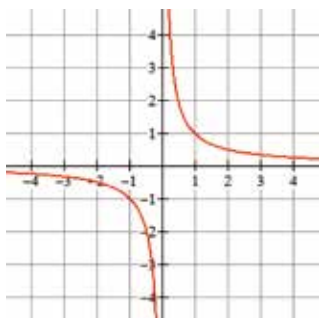
Van mijn virtuele reizen naar het hyperbolische hiernamaals snapte mijn gelovige vader natuurlijk geen biet. Voor mij was belangrijk dat ik met zulke, achteraf toch wel absurde, verhalen een tijdje zijn aandacht had. Daarbij heeft hij nooit het tussn-de-duumngebaar gemaakt, dat dan zeker een negatieve connotatie voor mij zou hebben gehad. Mijn hyperbolische gepassioneerdeheid moet hem wel hebben doen besluiten mij naar Groningen 'te verbannen' om zo mijn hoogdravende Delftse ambities de grond in te boren. Het moet geweest zijn op een zomerse augustusavond, zo tegen halfnegen tijdens zijn avondkoffie met biscuitje, dat hij me op besliste toon ontbood in zijn studiehok. Aan de muur boven zijn bureau hing een meer dan levensgroot portret van de kennelijk voor hem grote Abraham Kuyper. 'Jij gaat niet naar Delft. Mamma en ik hebben een kamer geregeld in Groningen en je bent ingeschreven bij wis- en natuurkunde. Verder kun je je melden bij de gereformeerde studentenvereni-

ging Hendrik de Cock en je bent overgeschreven van de kerk in Kampen naar die in Groningen-Zuid met dominee Stam als toezichhouder.' In de deuropening ving ik nog iets op over 'alle verleidingen van de satan in het zondige westen'.

Het werd dus Groningen en niet Delft. En zo vond ik mezelf dan als amper 17-jarige studentje terug in de collegebanken van de grote Offerhauszaal van het Groningse Academieggebouw, luisterend naar de op enigszins cynische toon gegeven colleges van dr. Verdenius, wiskunde lector: 'Voor elke $\varepsilon > 0$ – hoe klein ook – bestaat er $\delta > 0$ zodanig dat...' en dan volgde er iets over continuïteit en limieten. Het contrast met mijn Delftse wolkenkrabbers kon niet groter zijn. Bij Verdenius' limieten droomde ik weg op de glijbaan van Dorleijns hyperbool. En dat ging als volgt.

Les extrêmes ne se touchent pas, mais se transforment l'un en l'autre. Het startpunt van de hyperboolreis was het punt (1,1) (want, $1 \times 1 = 1$) op de hyperbool in het eerste kwadrant van het assenstelsel ' $xy=0$ '. De virtuele reis ging eerst rechtsaf naar beneden richting de positieve tak van de x-as. Steeds sneller ging het, steeds dichterbij die positieve x-as, zonder die ooit te bereiken. In het oneindige van dit deel van de reis geschiedde een wonder. Daar transformeerde de hyperbool van positief 'boven' de x-as naar negatief er 'onder', om daarbij aan de overzijde tevoorschijn te komen in het derde kwadrant. In dat kwadrant ging de val onstuitbaar bergafwaarts, via het punt (-1,-1), diep naar beneden naar de oneindige hel van de negatieve y-as. (Zo idioot verbeeldde ik mij dit!) En in dat diepe negatieve oneindige geschiedde het tweede wonder, de tweede transitie: van de hel naar de hemel in een ondeelbaar moment. De hyperboollijn transformeerde daar van negatief naar positief en van oneindig diep daaronder tot oneindig hoog daarboven. Terug dus in het eerste kwadrant. Terug naar waar de reis begon bij $1 \times 1 = 1$. Best grappig toch zo'n virtuele puberreis.

Bijna-doodervaringen overkomen je. Het gebeurde dit jaar in juli, op vakantie in Zuid-Frankrijk met mijn nieuwe racefiets: donkerrood, compleet van carbon en met schijfremmen en met wielen die een ruim een kwart van de totale fietsprijs kostten. Voor de kenner, een Trek Domane met 32mm banden. Niks niet tussn-de-duumn deze keer. Gedurende vier weken in de Franse Alpen maak ik een dikke 15.000 hoogtemeters. Vanuit de camping gaat de weg eerst Keutebergsteil naar beneden. Daarna zo'n zes kilometer vals plat naar het begin van de beklimming van



(links) Hyperbool tusen de assen:
 $xy \in \{0,1\}$

(rechts) De iconische ravijnval van
70 meter van Wim van Est in 1951

de col met een gemiddelde stijging van zo'n 6%. Heel wat anders dan een fietspad in de Noord-Drentse Onlanden. Na elke kilometer verschaft een bordje informatie met: de huidige hoogte, het stijgingspercentage van de komende kilometer en het aantal kilometers naar de top. Na maximaal 10 kilometer klimmen, de variantie van die 6% heeft als bovengrens zeker 10%, schieten de eerste hyperbolen als fata morgana's door m'n kop. Van opgeven heb ik nooit gehoord, hoogstens onderbreekt een hoognodige pisset (in welertermen) het ritme van de pedaalslag. Vergezichten gaan over in diepe ravijnen. Voor me flitst de kalkwitte bergtop met radartoren wenkend tussen de bergsparren door. In de buurt van de ruïne van het 12^{de}-eeuwse klooster is het komende gemiddelde stijgingspercentage 8% en is het nog een dikke 8 kilometer naar de top. Bij elke passage van een info-bordje geeft mijn fietscomputertje al enige tijd 'nog 54 minuten naar de top' aan. Het fietstempo loopt dus, langzaam weliswaar maar wel zeker, terug. Het is 31 graden en de honderden vliegen kunnen mijn tempo bijbenen. Natuurlijk kom ik boven. Het laatste deel van de klim heb ik de wind achter en geniet ik van de slechts 4% stijging. Boven is het suizen van de wind om de radartoren het enige geluid. Twee wielrenners passeren, met een *bonne route* als groet. Mijn bidon is bijna leeg en het laatste gelletje gaat erin. Dalen, heerlijk. Diep in het stuur. Suizend windje. Niets houdt me tegen. Wel altijd doortrappen, ook al is dat niet nodig voor de snelheid, maar van belang voor de betere controle over de fiets bij snelheden van zo'n 70 per uur op het rechte stuk. Op het vals plat is mijn bidon leeg en besluit die te gaan vullen bij de eerstvolgende *fromage de chevreboerderij*.

Dan nog zes kilometer naar de camping die links van de weg ligt. Tien meter voor de brug over de Lauzon begint links de klim naar de camping. Er rest nog zo'n 300 meter van gemiddeld 16%. Smerig steil dus. Opnieuw passeer ik de vrijwel droogstaande meanderende Lauzon. Op de brug neem ik een laatste slok water en giet de rest over de roestige reling. Vervolgens rechtdoor langs de *poubelles* naar boven. In het laatste rechte deel zit een scherpe onoverzichtelijke, steil omhoog buigende, bocht naar links. Het asfalt van de weg is daar naar rechts weg-

gezakt met in het midden een vervelende kloof. Ik ga hier staan in de pedalen. Flarden van overdrijvingen schieten door mijn hoofd: 'duwen en trekken, verzuring in de bovenbenen, heerlijk, excelsior, hyperbola, hyperbole, die engelsen zijn niet alleen brexitdom'. Dan doemt een rode auto op. Vanuit de linkse hoek met de bocht te ruim nemend. De berm rechts naast me is smal. Ik voel doornige braamscheuten.

En dan gaat het dus mis. Mijn voorwiel schiet in een braamstruik ... het derde kwadrant ... ik tuimel omlaag: drie keer hals over kop achterover. Dan de dreun, een fikse, tegen het houten beschot van de stacaravan aldaar. De bewoner komt aangesneld. Ik sta op. Drie braamkrassen op mijn rechterarm. Verder niks. Ongedeerd. Het binnenwerk van de helm ligt er wel scheef bij. Ik schat, dit alles na een val van een meter of zes, precies tussen twee platanboomstammen met een tussenruimte van een meter, door. Ongedeerd dus. Ook de fiets is ongeschonden. De hyperbool van het eerste kwadrant schiet geruststellend door mijn hoofd. De dame van de rode auto is uitgestapt en roept, laten we het houden op enigszins in shock: '*Tu aurais dû rester à gauche*'. Ik had dus links moeten houden, volgens dat wijf (*excusé le mot*). De goede mobil-homeman helpt me weer op de fiets en duwt me tot mijn schoenen weer vast in de pedaalclips zitten. Op eigen kracht overbrug ik de laatste tweehonderd gravelmeters van 12%. De drie schrammetjes en mijn scheefstaande helm maken mijn verhaal aan Rita volstrekt ongeloofwaardig. Het oude trauma borrelt weer op: dit is de waarheid en niets anders dan... Ze schenkt me een groot glas koele cola in.

Diezelfde dag 's avonds ben ik met Rita teruggegaan naar de plaats 'des tuimelings'. Ze is geen statisticus en zeker geen wiskundige, maar haar gezicht verraadt: 'geen zes meter'. Ik stamel iets als: gevoelsafstand ... inschattingsfoutje ... tolerantie marge. Laten we het houden op 'tussn de duumn'.

GERARD SIERKSMA is emeritus hoogleraar Operations Research aan de Rijksuniversiteit Groningen.
E-mail: g.sierksma@rug.nl

Deel 2: periode 1990–2021

Fabricage van scheerkoppen bij Philips in Drachten. Foto: Philips | Robert Smallegange

Vijfenveertig jaar statistiek

In 2021 ben ik met emeritaat gegaan na 45 jaar werkzaam te zijn geweest in de statistiek. In deel 1 beschreef ik mijn verleden als mathematisch statisticus bij het CWI en als medisch statisticus aan de Universiteit van Maastricht. In deel 2 zal ik ingaan op mijn werkzaamheden als statistisch adviseur bij het CQM (Philips) en als industrieel statisticus aan de Universiteit van Amsterdam.

RONALD J. M. M. DOES

Na achtereenhalf jaar gewerkt te hebben bij de Universiteit van Maastricht vertrok ik als universitair hoofddocent voor een nieuw avontuur in Brabant. Ik ging werken als industrieel statisticus bij het CQM. Het CQM was in 1979 door Jos de Kroon opgericht om als onafhankelijk adviesbureau binnen Philips diensten op het terrein van de kwantitatieve methoden op commerciële wijze aan te

bieden. Mijn voornaamste taak richtte zich op het implementeren van statistische procesbeheersing binnen de diverse onderdelen van Philips. Het beheersen van variatie in processen, zodanig dat de uitkomsten van de kwaliteitskenmerken voorspelbaar zijn, is essentieel voor een goede bedrijfsvoering. Een aantal klanten van Philips uit de automobiel- en vliegtuigindustrie eiste dat

ook van hun toeleveranciers. In mijn CQM tijd heb ik verschillende bedrijfsonderdelen van Philips mogen helpen bij deze implementatie. Met name bij Philips Stadskanaal (dioden), Philips Drachten (scheerapparaten), Philips Nijmegen (halfgeleiders), Philips Roermond (keramische geleiders) en Philips Santa Ana, VS (echografie) ben ik maanden actief geweest. Alhoewel het CQM geen onderzoeksinstituut was, was er wel ruimte voor wetenschappelijk onderzoek. Met name op het gebied van de operations research heeft het CQM een flink aantal hoogleraren voortgebracht (bijvoorbeeld de volgende collega's uit mijn CQM tijd: Dick den Hertog, Hein Fleuren, Ton de Kok, Peter van Laarhoven en Henk Zijm).

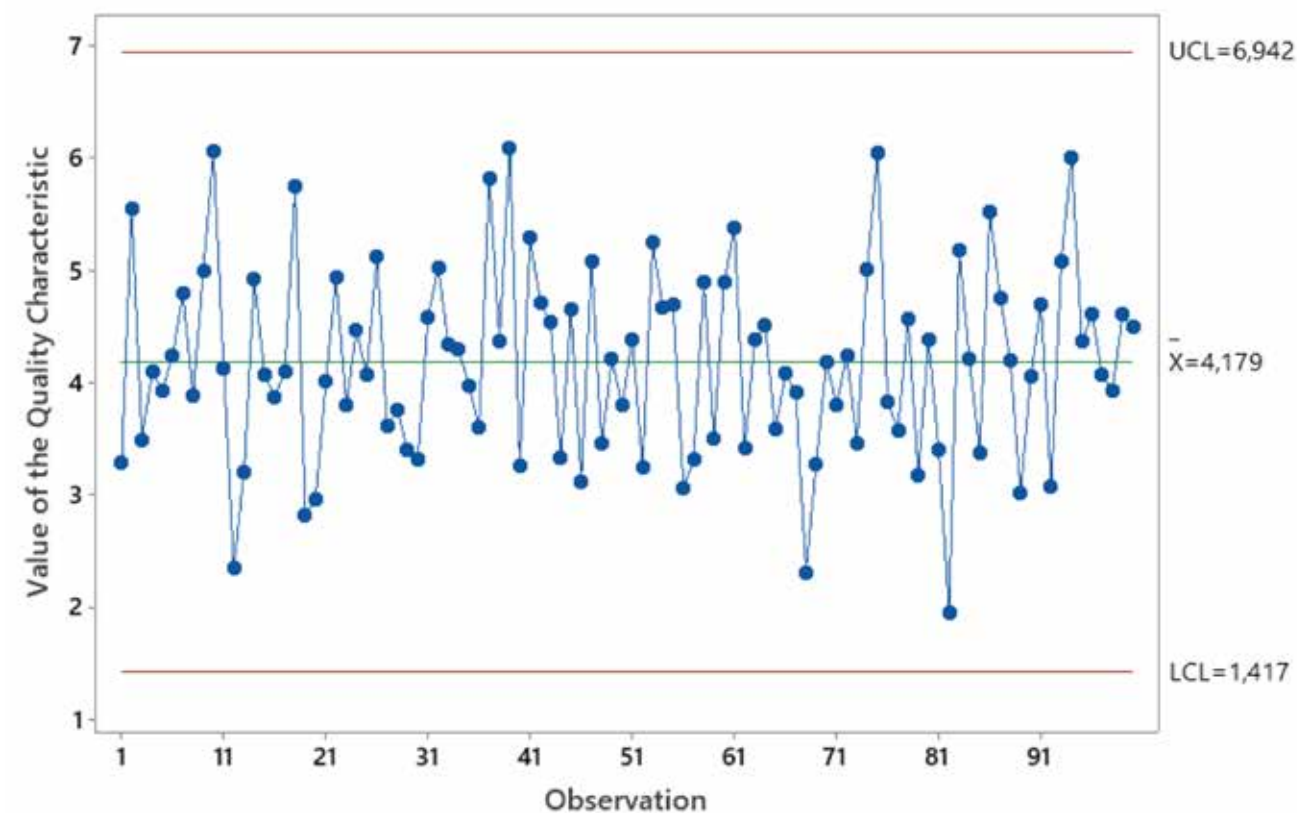
Regelkaart

Mijn onderzoek concentreerde zich al snel op de regelkaart: een grafiek met op de ene as de tijdstippen waarop een kwaliteitskenmerk gemeten wordt en op de andere as de waarde van het kwaliteitskenmerk. Of het onderliggende proces al dan niet statistisch beheerst is, wordt beoordeeld door het toevoegen van regelgrenzen (UCL

en LCL) die statistisch bepaald zijn (zie voorbeeld).

Met name het vaststellen van de regelgrenzen, de steekproeven, de onderliggende variatiebronnen en corresponderende verdelingsfunctie aannamen zijn onderwerpen voor nadere studie. Mijn eerste artikel over dit onderwerp verscheen in *Statistica Neerlandica* (zie Does en Schriever, 1992). Met de statistici van het CQM werd ook nog een handboek geschreven over industriële statistiek en kwaliteit (zie Banens e.a., 1994). Verschillende hogescholen hebben dit boek gebruikt als lesmateriaal.

Aan de Universiteit van Amsterdam was begin negentiger jaren een opleiding Statistiek gestart met 3 afstudeervarianten waaronder bedrijfs- en industriële statistiek. Voor een hoogleraar Industriële Statistiek werd het CQM gepolst om geschikte kandidaten naar voren te schuiven. Op 1 oktober 1991 werd ik benoemd voor 1 dag per week binnen de vakgroep Wiskunde van de toenmalige Faculteit voor Wiskunde en Informatica. De samenwerking met Chris Klaassen, hoogleraar Mathematische Statistiek, kon ook weer voortgezet worden. De combinatie van werkzaamheden duurde tot 1 mei 1994. Vanaf die datum aanvaardde ik het voltijds hoogleraarschap en verhuisden we weer naar het westen van het land.



Voorbeeld van een Shewhart regelkaart voor individuele metingen

Industriële statistiek

Het vakgebied industriële statistiek kan alleen maar bloeien als er een wisselwerking bestaat tussen het wetenschappelijk onderzoek en de toepassing ervan in het bedrijfsleven. Om dit te creëren richtte ik in 1994 het Instituut voor Bedrijfs- en Industriële Statistiek (IBIS UvA) op: een adviesbureau binnen de Universiteit van Amsterdam, dat vanuit zijn deskundigheid op het gebied van methoden en technieken de verbetering van de bedrijfsprocessen van zijn klanten ondersteunt. Meteen hadden we twee grote klanten (Hollandse Signaalapparaten (het tegenwoordige Thales) en ASML), die we hielpen met het implementeren van statistische procesbeheersing. Gelukkig werd het team van adviseurs snel uitgebreid met onder andere de komst van twee oud-collega's van het CQM, Kit Roes en Albert Trip. Zij gingen ook aan de slag met hun proefschrift met als onderwerp de regelkaart. De praktijkvoorbeelden uit Philips Stadskanaal, Philips Drachten en ASML vormden belangrijke ingrediënten voor het onderzoek. Het onderzoek had ook impact, hetgeen bijvoorbeeld bleek uit het artikel in *Technometrics* (zie Roes en Does, 1995). Samen met Kit en Albert schreef ik ook nog een standaardwerk op het gebied van statistische procesbeheersing dat zowel in het Nederlands als in het Engels verscheen (zie Does, Roes en Trip, 1996 en 1999).

Voor de nieuwe studierichting Statistiek ontwikkelde ik het vak industriële statistiek en zorgden we ervoor dat de studenten konden afstuderen bij bedrijven zoals Organon, General Electric, Philips, Thales en ASML. Het was een inspirerende tijd binnen de vakgroep Wiskunde waarvan ik ook nog een periode voorzitter was. In 1992 mochten we Robbert Dijkgraaf begroeten als hoogleraar Mathematische Fysica en het was een voorrecht zo'n collega in de vakgroep te hebben.

Lean Six Sigma

In de afgelopen decennia is er een enorme ontwikkeling waar te nemen hoe we mede met behulp van de industriële statistiek een organisatie effectief kunnen mobiliseren om kwaliteitsproblemen op te lossen. Volgens de evolutietheorie worden de beste ideeën, methoden en benaderingen behouden en de niet zo zinvolle weggegooid. De integrale kwaliteitszorg (TQM) beweging eind vorige eeuw (met de theorieën van Deming, Juran, Ishikawa, Feigenbaum en Taguchi) onderscheidde zich door een focus op kwaliteitsverbetering. Dit stond lijnrecht tegenover in-

spectie en kwaliteitsbeheersing, de dominante stroming in de jaren daarvoor. Maar ook al bestond TQM uit vele nuttige innovaties in management, de vele mislukkingen uit de praktijk gaven aan dat het gebrek aan programmabetrokkenheid van het management een probleem bleek. Dat veranderde met de komst van Six Sigma: een kwantitatief kwaliteitsverbetering programma gebaseerd op Jurans Trilogie (plannen, beheersen en verbeteren), projectmanagement methoden en een focus op resultaat, vertaald in financiële en strategische termen. Voor de industriële statistici was Six Sigma een reïncarnatie van hun vakgebied. En deze ontwikkeling ging hand in hand met de digitale revolutie. Denk hierbij aan computertechnologie, statistische software, de veel betere toegankelijkheid en makkelijke verplaatsing van data, communicatie via internet, en het gebruik van PowerPoint voor educatie en projectrapportages. Six Sigma behelst systematische innovatie, in plaats van enkel het streven naar foutloze producten of diensten.

Externe factoren waren ons in de jaren negentig gunstig gezind. Toen General Electric onder de bezielende leiding van Jack Welch het Six Sigma programma omarmde, kwam het in een stroomversnelling. In Nederland werden de drie technische universiteiten gevraagd om het programma met statistische kennis te ondersteunen. Oud collega Wim Albers van de TU Twente tipte ons en zodoende hadden we in 1996 ons eerste project bij General Electric Plastics in Bergen op Zoom te pakken. Snel volgden meer Amerikaanse gelieerde bedrijven zoals DAF Trucks als onderdeel van PACCAR en Douwe Egberts als onderdeel van Sara Lee. Een extra impuls voor het verbeterprogramma Six Sigma werd ontdekt door een project bij Nationale Nederlanden. Daar was het adviesbureau McKinsey ook actief en zij leidden in 3 dagen Lean experts op. IBIS UvA pakte dit snel op en integreerde beide aanpakken tot Lean Six Sigma. Lean Six Sigma is meer dan kwaliteitsverbetering in de engere zin van het woord. Lean Six Sigma wordt veel ingezet om doorloop- en bewerkingstijden te reduceren, verspillingen te elimineren, kostenreducties te realiseren, omzetten en marktaandeel te verhogen. IBIS UvA werd vanaf 2000 marktleider op het gebied van Black en Green Belt opleidingen. In de loop van tijd hebben we duizenden professionals in het bedrijfsleven opgeleid. Ook het certificeringsprogramma voor Black en Green Belts vond gretig aftrek. Meer dan 3000 personen deden examen bij de Universiteit van Amsterdam. IBIS UvA was commercieel een groot succes, hetgeen bleek uit de ruim 12 miljoen euro dividend die uitgekeerd werd aan de Universiteit in de periode 1998–2015 toen IBIS UvA een BV was binnen de Holding. Het

aantal medewerkers is echter nooit hoger dan 8 geweest, waaronder altijd 3 à 4 jonge promovendi.

Een echt keerpunt voor het onderzoek van IBIS UvA was de komst van Søren Bisgaard als deeltijd hoogleraar in 2000. Bisgaard was op het terrein van de bedrijfs- en industriële statistiek een persoon van wereldfaam. Zijn onderzoekinteresses waren zeer divers. In het begin van zijn carrière hield hij zich vooral bezig met design of experiments. Later verschoof zijn aandacht meer naar operations management, Lean Six Sigma, healthcare engineering en tijdreeksanalyse. Op alle gebieden was hij een autoriteit. Zijn inspiratie, grote kennis en netwerk hebben IBIS UvA internationaal gepositioneerd. Door onze focus op het verbeteren van processen in het bedrijfsleven, werd het in 2009 tijd om te wisselen van faculteit. Op 1 april 2009 werden we onderdeel van de Faculteit Economie en Bedrijfskunde, eerst binnen de Amsterdam School of Economics, en in 2014 als aparte sectie Operations Management binnen de Amsterdam Business School.

Operations Management

In totaal zijn 14 promovendi onder mijn leiding gepromoveerd op het terrein van de industriële statistiek. Veelal was de regelkaart het onderzoeksonderwerp. Dat trok ook internationaal aandacht en er melden zich 3 Pakistani en een Zuid-Afrikaan die ook aan de Universiteit van Amsterdam wilden promoveren. Die samenwerking leidde tot mooie publicaties (zie bijvoorbeeld Schoonhoven e.a., 2011 en het overzichtsartikel van Does e.a., 2020). Op het terrein van de regelkaart kwamen we uiteindelijk met een verdelingsvrije oplossing om de regelgrenzen te bepalen, waarbij de asymptotische statistiek uit mijn begin carrière goed van pas kwam (zie Goedhart e.a., 2020).

Mijn onderzoek werd ook breder. Er hebben 8 promovendi gewerkt aan onderwerpen uit de operations management, Lean Six Sigma en healthcare engineering. Bijzonder was dat een voorzitter van de Raad van Bestuur van het Rode Kruis Ziekenhuis (Jaap van den Heuvel) en twee master black belts (Gerard Niemeijer van het UMCG en Bart Lameijer van ABN AMRO) in staat bleken om proefschriften te schrijven over Lean Six Sigma onderwerpen die vanuit de gezondheidszorg en de financiële dienstverlening waren gevoed. De voorzitter werd later hoogleraar Healthcare Management en de twee master black belts vervolgden hun carrière op een universiteit.

Bestuurlijk was ik direct betrokken bij de oprichting van de volgende internationale verenigingen: Internatio-

nal Society for Business and Industrial Statistics (ISBIS), European Network for Business and Industrial Statistics (ENBIS) en International Statistical Engineering Association (ISEA). In die verenigingen was ik jarenlang directeur, secretaris, penningmeester of lid van het bestuur. Door onze bijdragen aan kwaliteitsverbetering en de industriële statistiek volgden de internationale erkenningen in de loop der jaren elkaar snel op. Zo werd ik Fellow van de American Society for Quality (ASQ) in 2007 en Fellow van de American Statistical Association in 2014. Ik ontving de Hunter Award in 2008 van de Statistics Division van de ASQ vanwege mijn consultancy activiteiten; de Shewhart Medal van de ASQ en de Box Medal van ENBIS in 2019 vanwege mijn statistische bijdragen en in 2021 de Lancaster Medal van de ASQ vanwege mijn internationale bijdragen aan operational excellence. Ook IBIS UvA viel in de prijzen met de Excellence in Practice Gold Award van de European Foundation for Management Development vanwege de implementatie van Lean Six Sigma in het Universitair Medisch Centrum Groningen (zie Does en De Jong, 2014). De uitreiking was een bijzonder moment omdat de Silver Award winnaar de University of Oxford was en die vertegenwoordigers waren echt jaloers op ons.

Toekomst van de industriële statistiek

Wat is de toekomst van de industriële statistiek? Uiteraard kan ik dit niet echt voorspellen. Echter, in de afgelopen dertig jaar ben ik wel opgetrokken met de internationale collega's om aan het front mee te denken wat er gebeurt met het vakgebied. Allereerst moeten we vooral de scope op wat we doen verbreden. In de toekomst moeten we ons meer gaan bezighouden met de grote, complexe en ongestructureerde problemen. Typische voorbeelden daarvan zijn de duurzame ontwikkelingsdoelstellingen die in 2015 door de Verenigde Naties zijn vastgesteld als de nieuwe mondiale duurzame ontwikkelingsagenda voor 2030. Maar ook in organisaties zijn vele grote problemen te vinden die een integrale aanpak nodig hebben. Professionals begrijpen dat dergelijke projecten uiteindelijk gericht zijn op het voldoen aan behoeften van de klant en het verbeteren van de concurrentiepositie van een organisatie. Om het proces van het oplossen van de grote, complexe en ongestructureerde problemen te systematiseren kan instrumentarium worden ingezet om op systematische wijze, de behoeften en verwachtingen van de klant in kaart te brengen, onderzoek en diagno-

ses uit te voeren om het probleem te identificeren, een oplossing te ontwikkelen voor het probleem en tevens de oplossing te borgen met geschikte methoden en technieken. Sinds de dagen van Shewhart vormen vooral de data gerelateerde-instrumenten en stappenplannen de basis voor ons beroep, die gebaseerd zijn op een wetenschappelijke methode. Denk bijvoorbeeld aan het Lean Six Sigma DMAIC stappenplan en het gebruik van diverse kwaliteitsinstrumenten. Toch denk ik dat er een extra stap nodig is en daarom ben ik de afgelopen jaren heel actief geweest binnen ISEA. Het heeft geleid tot een Handbook of Statistical Engineering (zie Hare e.a., 2021). Statistical Engineering is een verdere evolutie van ons vakgebied die we met Lean Six Sigma hebben toegepast. Economische, bedrijfskundige, analytische, computer en statistische kennis komen zo mooi samen. Innovatie is een zeer belangrijk strategisch onderwerp. Een organisatie kan het zich niet veroorloven dit te negeren. Daarom heb je een organisatie nodig die toegerust is om deze innovatieve taken systematisch uit te voeren.

Tot besluit. Terugkijkend op mijn 45-jarige carrière als statisticus kan ik heel tevreden zijn. Na een gedegen mathematisch statistische scholing onder leiding van Willem van Zwet, heb ik mij ontwikkeld als toegepast statisticus. De eerste jaren in de medische statistiek en psychometrie, de laatste 30 jaar in de industriële statistiek. Die laatste switch rond mijn 35-ste bleek een gouden greep. De



Boeken op het terrein van de industriële statistiek

combinatie van een voet in de praktijk en de andere voet in de academie was voor mij ideaal. Vele bedrijven heb ik mogen begeleiden met de implementatie van statistische procesbeheersing en Lean Six Sigma. Van de 9 boeken op deze terreinen, waar ik co-auteur van was, zijn meer dan 25.000 exemplaren verkocht.

Mijn wetenschappelijke publicaties zijn ruim 6.000 keer geciteerd en leverde eind 2022 een *h-index* op van 44. Met name de artikelen die vanuit de praktijk werden geïnitieerd worden veel geciteerd (zie bijvoorbeeld De Koning e.a., 2006 met bijna 650 citaties). Ook mocht ik 25 promovendi begeleiden en aan vele duizenden professionals en studenten lesgeven. Dat ik bij mijn afscheid door Zijne Majesteit de Koning benoemd werd tot Ridder in de Orde van Nederlandse Leeuw was een bekroning op mijn carrière.

LITERATUUR

- Banens, P.J.A., Does, R.J.M.M., Van Dongen, G.B.W., Engel, J., Hasselaar, M.M.A., Van Lieshout, R.A.J.M., Praagman, J., Schriever, B.F., Trip, A., & Van der Veen, H. (1994). *Industriële statistiek en kwaliteit*. Kluwer Bedrijfswetenschappen.
- De Koning, H., Verver, J.P.S., Van den Heuvel, J., Bisgaard, S., & Does, R.J.M.M. (2006). Lean Six Sigma in healthcare. *Journal for Healthcare Quality*, 28, 4–11.
- Does, R.J.M.M. & De Jong, L.J. (2014). An industry approach transforms healthcare: a seven year journey. *EFMD Global Focus*, 8(3), 15–18.
- Does, R.J.M.M., Goedhart, R., & Woodall, W.H. (2020). On the design of control charts with guaranteed conditional performance under estimated parameters. *Quality & Reliability Engineering International*, 36, 2610–2620.
- Does, R.J.M.M., Roes, K.C.B., & Trip, A. (1996). *Statistische procesbeheersing in bedrijf*. Kluwer Bedrijfswetenschappen.
- Does, R.J.M.M., Roes, K.C.B., & Trip, A. (1999). *Statistical process control in industry*. Kluwer Academic.
- Does, R.J.M.M., & Schriever, B.F. (1992). Variables control chart limits and tests for special causes. *Statistica Neerlandica*, 46, 229–245.
- Goedhart, R., Schoonhoven, M., & Does, R.J.M.M. (2020). Non-parametric control of the conditional performance in Statistical Process Monitoring. *Journal of Quality Technology*, 52, 355–369.
- Hare, L.B. (Ed.-in-chief), Does, R.J.M.M., Hoerl, R.W., Schall, S.O., & Snee, R.D. (Eds). (2021). *Handbook of statistical engineering*. eBook via www.isea-change.org.
- Roes, K.C.B., & Does, R.J.M.M. (1995). Shewhart-type charts in nonstandard situations (with discussion). *Technometrics*, 37, 15–40.
- Schoonhoven, M., Nazir, H.Z., Riaz, M., & Does, R.J.M.M. (2011). Robust location estimators for the X-bar control chart. *Journal of Quality Technology*, 43, 363–379.

RONALD DOES is emeritus hoogleraar Industriële Statistiek aan de Universiteit van Amsterdam.
E-mail: r.j.m.m.does@uva.nl

PEILINGPRAKTIJEN

Wees voorzichtig met voxpops!

Steeds vaker maken media gebruik van *voxpops*. Een voxpop is een kort videofragment waarin een onderzoeker de ‘gewone’ man of vrouw vraagt over een meestal actueel onderwerp. Voxpops worden soms ook wel *straatpeilingen* genoemd. Dat wekt de indruk dat ze een vorm van opinieonderzoek zijn. Een soort peilingen dus. Dit is echter meestal niet het geval. Daarom is voorzichtigheid geboden met het gebruik en de interpretatie van voxpops. Besef wat het wel is (een journalistieke impressie) en wat niet (een vorm van opinieonderzoek)!

Voxpop komt van *vox populi*. Dat is Latijn voor ‘de stem van het volk’ of ‘de stem des volks’. Het is de mening van de spreekwoordelijke gewone man (of vrouw), zoals tv-, radio- of krantenmakers die vastleggen in korte straatinterviews. Ze noemen dat wel *voxpoppen*. Daar is niks mis mee, als je het maar niet verward met publieksonderzoek. Voxpops hebben namelijk ernstige tekortkomingen als je ze wilt gebruiken of wilt zien als een representatieve peiling.

Meestal spreekt de *voxpopper* mensen op straat aan om ze hun mening te laten geven over een bepaald nieuwsfeit. De antwoorden zijn gevarieerd en onvoorspelbaar. Daardoor kan de onderzoeker een mix van meningen aan elkaar plakken, bijvoorbeeld een van een voorstander van een voorstel en een van een tegenstander. Dit gebeurt volstrekt subjectief en dus hoeft het resultaat niet representatief te zijn. De voorkeur lijkt eerder uit te gaan naar een ‘leuke’ reactie die het goed doet ‘op camera’ dan naar een ‘representatieve reactie’. Dat is misschien wel begrijpelijk vanuit journalistieke overwegingen, maar het heeft niets te maken met enig wetenschappelijk streven. De selectie van ondervraagden zal nooit zelfs maar in de buurt komen van een aselechte steekproef.

Voor leuke voxpops moet je geschikte locaties vinden waar je allerlei mensen kunt aanspreken. Het moet druk genoeg zijn en voldoende mensen moeten meewerken. Dus op naar drukke plaatsen als markten (de Albert

Cuypmarkt in Amsterdam), winkelcentra (de koopgoot in Rotterdam) of stations (het Centraal Station in Utrecht). Door deze locaties op te zoeken, maken de onderzoekers natuurlijk een scherpe voorselectie uit de doelpopulatie van ‘alle gewone mensen’. Dit leidt tot een selectief, en dus vertekend beeld.

Je krijgt een representatief onderzoek als je de mensen volstrekt willekeurige selecteert. Selectie mag niet ingegeven zijn door wat die mensen te melden hebben. Door je niet te richten op het trekken van een aselechte steekproef maar uit te gaan van interessante reacties, lopen voxpops het grote risico dat meningen van bepaalde groepen oververtegenwoordigd raken en het beeld bepalen. Ook kan het gebeuren dat bepaalde minderheden helemaal niet aan bod komen, zoals ouderen die niet meer hun eigen boodschappen doen of mensen die om andere redenen de straat niet op gaan. Het tijdstip van het opnemen van voxpops kan eveneens een rol spelen. Als je dat bijvoorbeeld ’s morgens doet, kun je scholieren en studenten mislopen, of mensen die dan al lang op kantoor zitten.

Hier zijn drie concrete voorbeelden van het gebruik van voxpops. Ze zijn afkomstig uit de NPO1-uitzendingen van Eenvandaag, Nieuwsuur en de NOS op drie op een volgende dagen: 27, 28 en 29 november 2019.

Forum Groningen

Na meer dan vijftien jaar van plannen maken en bouwen werd op 27 november 2019 het Forum Groningen geopend. Dit markante gebouw van tien verdiepingen en een oppervlakte van 17.000 vierkante meter, moet een ontmoetingsplek worden voor alle Groningers. In het gebouw zijn onder meer een filmhuis, het Stripmuseum, een restaurant, studieplekken en een bibliotheek ondergebracht. Er was echter veel discussie over het nieuwe gebouw. De vraag dringt zich dan ook op wat de Groningers er eigenlijk zelf van vinden? EenVandaag maakte een reportage met daarin de vier voxpops die in figuur 1 zijn weergegeven. Zie ook EenVandaag (2019).

Op het eerste gezicht lijken de geïnterviewden behoorlijk divers. Keurig de helft man en de helft vrouw.



‘Het is een beetje groot gebouw. Een groot gedrocht.’



‘Het is een heel mooi gebouw geworden, maar je ziet er helemaal niets meer van.’



‘Ik vind het niet passen bij de omgeving.’



‘Ik vind het niet verkeerd. Er komt hier meer leven in de brouwerij.’

Figuur 1. Voxpops in EenVandaag (op 27 november 2019)

Sommigen waren jonger en anderen ouder. Maar wel allemaal personen met een Nederlandse achtergrond. Als je vervolgens naar hun meningen kijkt, dan is de helft positief over het nieuwe gebouw en de andere helft negatief. Eerlijk verdeeld weliswaar, maar dit beeld klopt niet met de werkelijkheid. Uit degelijker peilingen van Dagblad van het Noorden, RTV Noord en OOG TV vlak na de opening bleek dat 70% van de Groningers het gebouw geen aanwinst vindt voor de stad. Zie Groninger Krant (2019). De Voxpops geven geen goed beeld van de verhouding voor-

en tegenstanders. Eigenlijk zouden drie geïnterviewden kritisch moeten zijn, en slechts een enkele lovend.

Nutri-score

Vanaf 2021 komt op alle verpakkingen in de supermarkt een nieuw logo: de Nutri-Score. Met vijf letters en vijf kleuren moet het ons consumenten helpen om verschillende producten te vergelijken en makkelijker gezondere



‘Nee, nooit gezien. Nooit op gelet.’



‘Het is nog geen gewoonte om dit te doen. Het staat niet op elk product.’



‘Ik heb niet naar de Nutri-Score gekeken.’



‘Dat kan ze stimuleren om op te letten bij het kopen van producten.’

Figuur 2. Voxpops in Nieuwsuur (op 28 november 2019)

keuzes te maken. In Frankrijk gebruiken ze de Nutri-Score al langer. Daarom ging correspondent Frank Renout van het programma Nieuwsuur in Frankrijk naar een grote supermarkt en vroeg de klanten wat ze ervan vonden. Zie figuur 2. Zie ook Nieuwsuur (2019). Voxpops dus! De antwoorden zijn uit het Frans vertaald in het Nederlands.

Je moet je afvragen hoe informatief voxpops zijn als ze in een ander land op een parkeerplaats van één supermarkt zijn gemaakt. En als deze meningen ergens voor staan, kun je deze reacties dan ook nog eens naar de Nederlandse situatie vertalen? Verder valt het op dat de geïnterviewden overwegend vrouw zijn. Was het een opname doordeweeks overdag, als meer vrouwen dan mannen boodschappen doen? Kortom, de vier voxpops zeggen heel weinig over de Franse en nog minder over de Nederlandse bevolking.

Black Friday

Black Friday is in de Verenigde Staten de dag na Thanksgiving Day. Die vrijdag hebben de meeste werknemers vrij. Black Friday is het begin van het seizoen voor de kerstaankopen. Omdat dan veel mensen gaan winkelen, adverteren veel winkels met aanbiedingen.

Deze rage is ondertussen overgeplant of overgewaaid naar Nederland. Maar er is discussie: Black Friday zet aan tot onnodige aankopen en stimuleert koopgekte. En dat vlak voor Sinterklaas! Vooral kleine winkeliers zien er niks in. Ze worden echter welhaast gedwongen om mee te doen. Maar hun marges worden kleiner, ze verdienen nauwelijks meer.

Op 29 november 2019 maakte de NOS een reportage over Black Friday in Zoetermeer. Volgens de NOS was de dubbelheid van Black Friday nergens beter te zien als in Zoetermeer. Alsof dat feitelijk was vastgesteld. In de oude Dorpsstraat deden echter nauwelijks mensen mee. Maar

in de Promenade even verderop was het heel druk en knal-den de mega-kortingen je tegemoet, zoals dat dan hoort op een echte Zwarte Vrijdag. Zie de voxpops in figuur 3.

De drie voxpops wekken de indruk dat de meerderheid Black Friday eigenlijk wel ziet zitten. Slechts één doet er niet aan mee. De geïnterviewden zijn trouwens wel vrij divers: een vrouw en twee mannen, en iemand met een migratieachtergrond. Toch kloppen de verhoudingen niet. In dat jaar peilde onderzoeksbureau Motivaction de mening in Nederland. Zie Motivaction (2019). Daaruit bleek dat slechts een derde van de Nederlanders van plan was op 29 november, of op de dagen erna, op koopjesjacht te gaan. Maar ja, de mensen die er niets in zien, die lopen niet in de winkelstraat en krijg je niet voor een leuk voxpopje voor de camera natuurlijk ...

De populariteit van voxpops

Uit bovenstaande voorbeelden blijkt dat je voxpops absoluut niet kunt gebruiken om een representatief beeld te schetsen van wat er aan de hand is. Het gepresenteerde beeld is standaard verre van representatief, en betreft een hoogst arbitraire, heel kleine verzameling individuele impressies.

Toch zijn voxpops populair bij journalisten. Waarom? Kathleen Beckers (2017) deed onderzoek onder 253 Belgische journalisten. Daaruit bleek dat het voornamelijk audiovisuele journalisten zijn die er graag en veel mee werken. Ze denken hun publiek zo beter bij het onderwerp te kunnen betrekken. Kijkers/luisteraars/lezers breng je op deze manier dicht bij het onderwerp, en zij kunnen zich identificeren met de geïnterviewde personen. En in de hevige concurrentiestrijd tussen de verschillende media is wat het publiek wil steeds vaker leidend. En het publiek wil op personen gericht nieuws, 'human interest'. Voxpops zijn een eenvoudige manier om dit menselijke

aspect aan het nieuws toe te voegen. Mensen als u en ik mogen ook hun zegje doen!

Volgens Beckers (2017) zien sommige journalisten voxpops zelfs als een vorm van democratisering van het nieuws. Voxpops vertegenwoordigen immers de bevolking in het nieuws en zouden daarom de publieke opinie weergeven. Een mooie ambitie allicht, maar een streven dat via voxpops niet wordt waargemaakt: een beperkt aantal van die voxpops (drie of vier), in feite achteraf geselecteerd, levert geen representatief beeld op.

Dat de voxpop bijna een verplicht nummer is in tal van journalistieke reportages, stemt overigens de journalisten zelf niet per se vrolijk. Zo waren de journalisten in het onderzoek van Beckers behoorlijk negatief over het gebruik ervan: 63% van de journalisten was van mening dat er in nieuwsberichten te veel voxpops worden gebruikt.

Mark Easton (2019), de BBC News home editor, gaat op de website van de BBC dieper in op het wel en wee van voxpops. Hij constateert dat er forse kritiek is op het gebruik van voxpops. Die kritiek is zeker terecht voor voxpops waarin je een simplistische vraag stelt aan bezoekers van bijvoorbeeld een winkelcentrum en de voxpop bestaat uit een ja-antwoord, een nee-antwoord en een hilarisch antwoord van iemand die de vraag niet heeft begrepen. Zulke voxpops noemt Easton *patronising* en *disrespectful*.

Toch wil hij voxpops niet afwijzen als slechts zinloze opvulling van nieuwsberichten. Het moet mogelijk zijn om goede vragen te stellen en te luisteren naar wat geïnterviewden zeggen, hoe scherp geformuleerd of slecht geïnformeerd die meningen misschien ook zijn. Ze leren kunnen ons iets over wat er borrelt in en onder de oppervlakte van de samenleving.

De Duitse mediaorganisatie Deutsche Welle waarschuwt dat het eenvoudig is om voxpops te manipuleren. Als je als journalist of onderzoeker een specifieke mening hebt over het onderwerp van de voxpop, dan is het simpel

om, bewust of onbewust, een draai te geven aan het gesprek. Dan kan bijvoorbeeld met een suggestieve vraagstelling of door antwoorden op een bepaalde manier te monteren. Onethisch natuurlijk, want in het hele productieproces van een voxpop dient de journalist neutraal en objectief te zijn, maar zal werkelijk elke journalist altijd de verleiding te sturen en spelen met voxpops kunnen weerstaan?

De BBC (2019a) raadt aan om niet meer dan spaarzaam gebruik te maken van voxpops. Want, nogmaals volgens die gerespecteerde omroeporganisatie (BBC 2019b), voxpops 'are a tool of illustration, not a tool of research'. En zo is het maar net!

BRONNEN

- BBC. (2019a), *BBC Election Guidelines*. https://www.bbc.com/editorialguidelines/general_election_guidelines_221119.pdf
- BBC. (2019b), *Guidance: Opinion polls, surveys, questionnaires, votes and 'straw polls'*. <https://www.bbc.com/editorialguidelines/guidance/surveys>.
- Beckers, K. (2017). Vox pops in the news: The journalists' perspective. *Communications*, 43(1). https://www.researchgate.net/publication/320248026_Vox_pops_in_the_news_The_journalists'_perspective.
- Deutsche Welle, *The Vox Pop*. DW-AKADEMIE Fortbildungszentrum Hörfunk. <https://www.dw.com/download/26027346/the-vox-pop.pdf>.
- Easton, M. (2019). *Why vox pops are important*. <https://www.bbc.com/news/uk-46946442>.
- EenVandaag. (2019). *Van aardbeving tot referendum: het mega bouwproject Forum Groningen gaat na jaren eindelijk open voor publiek*. <https://eenvandaag.avrotros.nl/item/het-omstreden-groninger-forum-na-8-jaar-bouwen-275-miljoen-euro-en-fel-debat-nu-eindelijk-open-voor/>.
- Groninger Krant. (2019). *Groningers vinden Forum Groningen geen gezicht*. <https://groningerkrant.nl/groningen/groningers-vinden-forum-groningen-geen-gezicht/>.
- Motivaction. (2019). *Black Friday: 40% meer deelnemers dan in 2018*. www.motivaction.nl/kennisplatform/nieuws-en-persberichten/black-friday-40-meer-deelnemers-dan-in-2018.
- Nieuwsuur. (2019). *Nutri-Score 'goede stap' naar gezonder leven maar 'meer maatregelen nodig'*. <https://nos.nl/nieuwsuur/artikel/2312478-nutri-score-goede-stap-naar-gezonder-leven-maar-meer-maatregelen-nodig.html>
- Wikipedia. (2020). *Stem van het volk*. https://nl.wikipedia.org/wiki/Stem_van_het_volk: (2 april 2020).

Aan deze column over voxpops is ook meegewerkt door Joop van Holsteijn, Instituut Politieke Wetenschap van de Universiteit Leiden.

JELKE BETHLEHEM werkte bij het CBS en is emeritus hoogleraar aan de Universiteit Leiden. Hij is een expert op het gebied van steekproeven, vragenlijsten en weergave van onderzoeksresultaten. Deze onderwerpen behandelt hij regelmatig in zijn blog. E-mail: mail@jelkebethlehem.nl



'Ziet er goed uit. Hoe meer kortingen, hoe meer ik kan kopen.'



'Ik doe er niet aan mee. Ik ga geen dingen kopen die ik niet nodig heb.'



'Al die kortingen neem je mee.'

Figuur 3. Voxpops bij de NOS (op 29 november 2019)

Monique van Hootegem, de feitelijke creator van STAtOR

De redactie krijgt eigenlijk alleen maar lovende commentaren op dit blad, al 23 jaar lang. Natuurlijk is dat in de eerste plaats omdat we een boeiende inhoud weten te brengen, met veel afwisseling. Maar die inhoud zou niet worden opgemerkt zonder de meer dan voortreffelijke presentatie waar al die jaren Monique van Hootegem verantwoordelijk voor is geweest.

Voor iedere publicatie geldt dat de presentatie essentieel is voor de waardering. Een slordig en onrustig uiterlijk leidt namelijk af van de inhoud. Toen de ad-hoc redactie 23 jaar geleden als proef het nulnummer van STAtOR samenstelde zocht men dan ook een professionele vormgever. Redactielid Han Oud wist wel zo iemand, zo ongeveer bij hem om de hoek woonde Monique van Hootegem, *the rest is history*.

Wat is nu zo uniek aan de opmaak van STAtOR? Wij denken dat het wezenlijke van die opmaak is dat hij niet opvalt. De tot uiterste doorgevoerde consistentie daarvan zorgt ervoor dat alle aandacht naar de inhoud gaat. De opmaak is dienend en speelt geen eigen hoofdrol. Misschien heeft de gemiddelde lezer het nooit gemerkt, en zo hoort het ook, maar simpele zaken als een altijd even grote witruimte rond kopjes en subkopjes dragen enorm bij aan een rustgevend uiterlijk. En zo'n dienend uiterlijk is dan het ideale vehikel om de inhoud te laten schitteren. Bij jou ging het altijd om de presentatie van de inhoud, niet om de presentatie zelf. Dát is het waardoor ons blad opvalt: de vanzelfsprekende onopvallendheid.

Toen we jaren geleden uit kostenoverwegingen van het oorspronkelijke formaat en een echte drukker overgingen naar A4 en een 'drukfabriek' was dat een uitdaging voor je. Vormgevers zijn niet zo dol op A4, dit formaat leent zich minder goed voor een boeiende opmaak. Maar ook hier wist je de dienende opmaak van het oorspronkelijke formaat schijnbaar zonder moeite over te zetten naar het daarvoor minder geschikte A4.

Als taalliefhebber had je grote aandacht voor een correct taalgebruik. Daarbij lette je niet alleen op de juiste

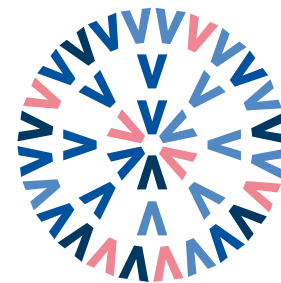


Foto: Pieter Bosch

spelling, maar ook op het plaatsen of niet plaatsen van komma's, het consequent gebruiken van de Nederlandse decimale komma in plaats van de Angelsaksische decimale punt, en op nog veel meer details. Menige nodeloos ingewikkelde zin werd door jou gesignaleerd en met een suggestie tot verbetering aan de auteur voorgelegd.

Na 23 jaar neem je afscheid van ons. Dank voor alles wat je voor ons blad gedaan hebt. We weten dat dit véél meer is geweest dan alleen teksten vormgeven. Je hebt bijvoorbeeld ook gezorgd voor de verzending en je hield de administratie bij van auteurs die geen VVSOR-lid waren en daarom een jaar lang ons blad ontvingen. Je maakte de jaarverslagen en nog veel meer. Je was de ziel, de inspirator en de feitelijke creator van STAtOR zoals we dit blad al 23 jaar kennen. Maar bovenal heb je ons met ieder nummer steeds weer geïnspireerd om door te gaan en nóg zo'n mooi nummer te maken. Je zult ontzettend gemist worden.

Met heel veel dank, de STAtOR-redactie



LNMB/NGB conferentie

16, 17 en 18 januari 2023

Het jaarlijkse LNMB/NGB congres zal plaatsvinden op 16, 17 en 18 januari 2023 in Hotel Kontakt der Kontinenten in Soesterberg en niet zoals jarenlang traditie was in Lunteren.

De eerste twee dagen worden door het LNMB georganiseerd en is bedoeld voor academici; de laatste dag wordt door het NGB georganiseerd en is met name gericht op operations research experts in het bedrijfsleven.

Houd voor de aankondiging van het thema en de registratiemogelijkheden de NGB website in de gaten: www.vvsor.nl/operations-research/.

Met vriendelijke groet,
Het bestuur van het Nederlands Genootschap voor Besliskunde (NGB)



YOUNG STATISTICIANS

has a new board: We have six board members from different backgrounds, and we're all enthusiastic about statistics; get to know us on the Young Statisticians section of the VVSOR webpage!

We are kicking off the academic year with a statistics pub quiz in September, where we will challenge participants with questions on R, statistical theory, funny graphs, and more! We're excited to see existing members and to meet incoming students and others who are enthusiastic about statistics. More exciting events such as Statistics Cafés and company visits will follow later. To keep informed on our future events, or to subscribe to our newsletter, make sure to visit our website: vvsor.nl/young-statisticians.



VVSOR ANNUAL MEETING 2023

thema

Statistics and Operations Research for robust decision making

donderdag 23 maart 2023

in Utrecht

Evenals vorig jaar zal de Annual Meeting ook dit jaar een hybride meeting zijn. We zullen bij elkaar komen bij In de Driehoek in Utrecht en zullen de bijeenkomst ook virtueel toegankelijk maken voor sprekers en toehoorders voor wie het niet mogelijk is om fysiek aanwezig te zijn.

Als een fysieke bijeenkomst vanwege COVID in maart 2023 niet mogelijk is, zal de AM volledig als online streaming event plaatsvinden met een iets aangepast programma.

De ALV en de uitreiking van de VVSOR Awards zijn onderdeel van de meeting.

De AM 2023 zal Engelstalig zijn. Het definitieve programma en informatie over hoe u de annual meeting kunt bijwonen kunt u enkele weken van tevoren vinden in de volgende editie van STAtOR en op de website van de VVSOR: <https://www.vvsor.nl/articles/vvsor-annual-meeting-2023/>.

E-mail: annualmeeting@vvsor.nl

