

STAtOR

Hoe veilig zijn pincodes?

Ras, BOAS, CFR

**Clio's stiefkind en de geschiedenis van
het probabilistische denken**

De ene peiling is de andere niet

**Optimaliseren van de inpak-strategie
voor bezorgbussen**

Rare jongens die Britten!

Nog vijf maanden te leven

Modelmatig overheidsbeleid



STAToR

Jaargang 24, nummer 2, juli 2023

STAToR is een uitgave van de Vereniging voor Statistiek en Operations Research (VWSOR). STAToR wil leden, bedrijven en overige geïnteresseerden op de hoogte houden van ontwikkelingen en nieuws over toepassingen van statistiek en operations research. Verschijnt 3 of 4 keer per jaar.

Redactie

Joaquim Gromicho (hoofdredacteur), Annelieke Baller, Caroline Jagtenberg, Guus Luijben (eindredacteur), Kerry Malone, Richard Starmans, Gerrit Stermerdink (eindredacteur), Vanessa Torres van Grinsven en Laura Zwep. 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, stator@wsor.nl

Bestuur van de VWSOR

Voorzitter: prof. dr. Casper Albers, db@wsor.nl; Secretaris: Pieter Jongma MSc, secretaris@wsor.nl; Penningmeester: dr. Rebecca Kuiper, penningmeester@wsor.nl; Algemeen bestuurslid: dr. Marianne Jonker, db@wsor.nl; Webmaster: Eugenio Traini MSc, webmaster@wsor.nl.

Voorzitters van de secties: prof. dr. ir. Mark van de Wiel (Biometrical Section); prof. dr. Albert Wagelmans (Section for Operations Research); prof. dr. ir. Frank van der Meulen (Section Mathematical Statistics); dr. Rebecca Kuiper (Social Sciences Section); dr. Michel van de Velden (Economics Section); dr. Iris Yocarini (Section Data Science); Luise Gummi, BSc (Young Statisticians); dr. Sanne Willems (Section Statistics Communication).

Leden- en abonnementenadministratie van de VWSOR

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

Voor advertenties

Prof. dr. J.A.S. Gromicho, stator@wsor.nl
STAToR verschijnt in maart, juli en november.

Uitgever

© Vereniging voor Statistiek en Operations Research
ISSN 1567-3383

Vakantieprobleem?

Dit nummer van STAToR verschijnt aan het begin van de vakantietijd, daarom bieden wij enkele verstrooiende artikelen.

Vakantie is een periode die voor veel mensen begint met een zekere stress: het optimaal inpakken van koffers, rugzakken of bagageruimtes. Steevast zal blijken dat iets dat men dringend nodig heeft helemaal onderin ligt. Als dit al een probleem kan zijn bij zoiets overzichtelijks als een koffer, dan moet het voor pakketbezorgers een veel complexere puzzel zijn. Voor hen telt namelijk iedere minuut, te lang zoeken naar een pakket heeft financiële gevolgen. Louise Zwep heeft uitgebreid naar dit probleem gekeken.

Kattenliefhebber Richard Gill heeft zich op een honds probleem gestort. Er zijn mensen die van honden houden met een bijzonder uiterlijk. Zo zijn honden met een korte snuit, zoals mopshonden, veel in trek. Vaak wordt op zo'n eigenschap doorgefokt, maar dat kan tot gezondheidscomplicaties leiden. Al te korte snuiten geven namelijk problemen met de ademhaling. Daarom zijn in het kader van het dierenwelzijn sommige rassen verboden. Die beslissing is onder meer gebaseerd op meetgegevens. En Richard heeft, zoals we dat van hem gewend zijn, kritisch naar die gegevens gekeken: hoe zijn ze verzameld, zit er misschien een bias in etc. Zijn conclusie is duidelijk, op deze gegevens kan men geen beleid baseren.

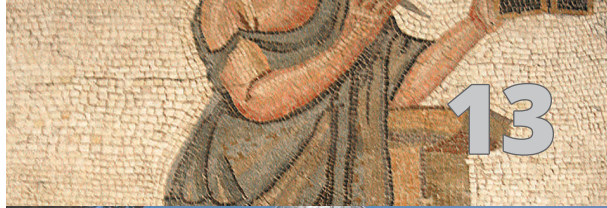
Denken over kansen is waarschijnlijk al zo oud als de mensheid. De ontwikkeling daarvan loopt parallel met ontwikkelingen in de wetenschap. Richard Starmans heeft een gedegen overzicht over deze ontwikkeling geschreven, met veel interessante details.

En over kansen gesproken: wat moeten we denken bij een uitspraak als "u hebt nog vijf maanden te leven"? Sanne Willems gaat in op de statistiek achter zo'n uitspraak.

Verder telt dit nummer een viertal zeer verschillende columns, daar zal dus zeker iets naar uw smaak bij zitten. En met nieuws van de Young Statisticians en wat verenigingszaken is het blad weer vol. Het is het tweede nummer Post Monique, we beginnen het hele proces langzamerhand steeds beter onder de knie te krijgen en onze bewondering voor haar werk is alleen maar gegroeid.

Wij wensen u een goede vakantie, met veel leesplezier!

De STAToR-redactie



INHOUD

2 Vakantieprobleem?

4 Hoe veilig zijn pincodes? – column
| Henk Tijms

6 Ras, BOAS, CFR | Richard D. Gill

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

13 Clio's stiefkind en de geschiedenis van het probabilistische denken | Richard Starmans

20 De ene peiling is de andere niet
– column | Jelke Bethlehem

24 Young statisticians

25 Optimaliseren van de inpak-strategie voor bezorgbussen | Louise J. Zwep

28 Rare jongens die Britten! – column
| Gerrit Stemerding

29 Save the date: Annual Meeting 21 maart 2024

30 Nog vijf maanden te leven | Sanne Willems

33 Statistica Neerlandica in een nieuw jasje
| Ernst Wit, Casper Albers

34 Modelmatig overheidsbeleid – column
| John Poppelaars



Hoe veilig zijn pincodes?

Je mag alles van me weten, behalve mijn pincode. Een historische slogan uit de jaren negentig net als de onvergetelijke reclamespot *Foutje, bedankt*. De tv en radio spots over de pincode gaven de kijkers en luisteraars adviezen hoe ze hun pincode moesten beschermen tegen kwaadwillenden. Een pincode bestaat uit vier cijfers. De betaalpas wordt geblokkeerd als je driemaal een verkeerde code intikt op dezelfde dag. Hoe veilig is deze bescherming als een kwaadwillende door verlies of diefstal een pasje van een ander in handen krijgt? Zijn er niet bepaalde cijfercombinaties die door heel veel mensen gebruikt worden? Het fenomeen van populaire getallen is welbekend uit loterijen. Voorbeelden te over. Jaren geleden bestond de Nederlandse lotto 25 jaar en

toen was het rijtje 1-2-3-4-5-6 verreweg het meest ingevulde rijtje. Het zal tegenwoordig niet anders zijn. Recentelijk publiceerde een Engelse krant dat in de UK National Lottery ongeveer 10 000 personen wekelijks de getallen 1-2-3-4-5-6 kiezen. Uiteraard heeft dit rijtje dezelfde minuscule kans als elk ander rijtje dat de jackpot er op valt, maar als dit zou gebeuren dan moet de jackpot met talloze andere deelnemers gedeeld worden. In de Zuid-Afrikaanse lotto werden op 1 december 2020 de zes opeenvolgende getallen 5-6-7-8-9-10 getrokken en moest de jackpot door vele deelnemers gedeeld worden. In de Filipijnse 6/55 lotto werden op 1 oktober 2022 de zes getallen 9-18-27-36-45-54 getrokken en waren er 433 winnaars van de jackpot. De psychologie

nr	pin	freq	nr	pin	freq
1	1234	10.71%	11	9999	0.451%
2	1111	6.02%	12	3333	0.419%
3	0000	1.88%	13	5555	0.395%
4	1212	1.20%	14	6666	0.391%
5	7777	0.745%	15	1122	0.366%
6	1000	0.616%	16	1313	0.304%
7	2000	0.613%	17	8888	0.303%
8	4444	0.526%	18	4321	0.293%
9	2222	0.516%	19	2001	0.290%
10	6969	0.512%	20	1010	0.285%

Tabel 1: Top twintig paswoorden

achter het kiezen van getallen is fascinerend. Zo is in Japan er een voorkeur voor oneven getallen: bij het geven van een cadeau is het aan te bevelen de gift uit drie, vijf of zeven pakjes te laten bestaan. Het getal acht is een geluksgetal in Azië: een analyse van krantenadvertenties in China, Taiwan en Hongkong liet zien dat 8 verreweg het meest populaire niet-nul getal is in prijzen van producten, een product wordt verleidelijker door een 8 in de prijs op te nemen.

Terug naar de vraag hoe veilig pincodes zijn. Voor eerstejaars studenten is mijn favoriete kansopgave over de complement regel: wat is de kans dat onder 500 gestolen pinpassen er tenminste één pinpas is waarmee de dief succes heeft binnen drie pogingen op dezelfde dag? Het antwoord is $1 - \left(1 - \frac{3}{10\,000}\right)^{500} = 0.1393$, een kans van bijna 14%. De kans op een geslaagde poging tot misbruik van een pinpas zou nog groter zijn wanneer het dievengilde op de hoogte zou zijn van de resultaten van een studie van 3,4 miljoen vier-cijferige digitale paswoorden uit vrijelijk beschikbare databestanden. De analyse werd uitgevoerd door de bekende data analyst Nick Berry en gepubliceerd op zijn blog DataGenetics. Berry had uiteraard geen toegang tot bestanden met pincodes voor betaalpassen, maar aangenomen mag worden dat de data voor de gebruikte paswoorden een goed beeld geven van gebruikte pincodes. In de tabel worden de twintig meest

gebruikte vier-cijferige paswoorden gegeven met daarbij de relatieve frequenties. Bijna 11% van de 3,4 miljoen onderzochte paswoorden is 1234. Uiteraard bevat de tabel ook de 'usual suspects' 1111, 2222, ..., 9999. De top twintig vertegenwoordigt 26,8% van de gebruikte paswoorden. Een vreemde eend in de bijt van meest populaire paswoorden lijkt 2580 op plaats 22. Bij nader inzien is de populariteit van dit paswoord eenvoudig te verklaren: 2580 staat in een rechte lijn van boven naar beneden in het midden van het toetsenbord op een mobiele telefoon. Wat zijn de minst gebruikte paswoorden? Uit Nick Berry's studie bleek dat 8068, 8093 en 9629 de drie minst gebruikte paswoorden zijn met relatieve frequenties 0,00074%, 0,00089% en 0,00095%.

Hoewel het dievengilde zich niet onder de lezers van STATOR bevindt, wordt een ieder ten zeerste afgeraden zijn pincode te veranderen en 8068 als pincode te nemen. Deze column is alleen bedoeld om een inkijkje te geven in de fascinerende wereld van hoe mensen hun getallen kiezen bij loterijen en pincodes. De studie van Nick Berry is voer voor psychologen!

Henk Tijms is emeritus hoogleraar operations research aan de Vrije Universiteit en auteur van diverse leerboeken over operations research en kansrekening. Zijn nieuwste boek is "A First Course in Probability for Computer and Data Science", World Scientific Press, 2023.
e-mail: h.c.tijms@xs4all.nl



Ras, BOAS, CFR

De statistiek achter een nieuwe wet op fokken met kortsnuitige honden

Richard D. Gill

De auteur werd geraadpleegd door verenigingen van fokkers van twee zeldzame oude hondenrassen: de Griffon Bruxellois en de Affenpinscher. Ze waren van mening dat recente wetten, bedoeld om gezondheidsproblemen bij kortsnuitige honden te bestrijden, gebaseerd waren op een foutieve interpretatie van statistisch

onderzoek. De auteur houdt helemaal niet van honden, laat staan het fokken van rashonden, maar nog minder van foute statistiek. Dit is een kort relaas van een nog niet beslechte strijd rondom de kwestie correlation vs. causation.



De Griffon Bruxellois in het Arnolfini portret van Jan van Eyck

Inleiding

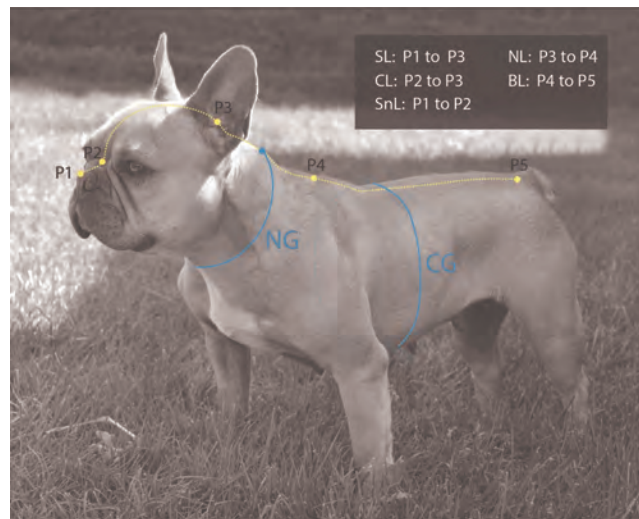
De laatste jaren kwamen de mopshond (engels: *pug*), en zowel de engelse als de franse bulldog, erg in de mode. Er kon dus veel verdiend worden met het fokken van de deze honden. Dit leidde tot een explosieve groei van hun aantal. In enkele decennia kwam een grote populatie voort uit een heel klein aantal voorouders. Er werd ook nog sterk geselecteerd op het bekende externe kenmerk: een platte snuit. Zulke honden worden 'brachycefiaal' genoemd. Eén van de ziektes waar deze honden vaak chronisch aan lijden heet BOAS, oftewel *brachephalic obstructive airways syndrome*. De hond heeft grote problemen bij het ademen, veroorzaakt door vervormingen van de interne luchtwegen en bijbehorende zachte weefsels.

Maar niet alle honden met korte snuiten hebben ernstig last van BOAS. Een belangrijk voorbeeld hiervan is de *Griffon Bruxellois*, bij kunstliefhebbers bekend van het beroemde "Arnolfini portret", geschilderd omstreeks 1434 door Jan van Eyck, waarschijnlijk in de stad Brugge. De huidige mopshond stamt waarschijnlijk van de *Griffon Bruxellois* af. Men denkt dat dit schilderij het huwelijksportret is van de voorname Italiaans bankier Arnolfini en zijn duidelijk hoogzwangere bruid. (Op een spiegel aan de muur achter de geportretteerden ziet men ook de schilder zelf en een tweede bezoeker. Het schilderij is dus eigenlijk een huwelijkscertificaat; de schilder en de andere persoon zijn de getuigen.)

Nu is een syndroom niets anders dan het simultaan aanwezig zijn van een aantal medische klachten samen met bepaalde lichaamskenmerken of andere externe factoren. Een hond kan per definitie alleen BOAS hebben als het een korte snuit heeft. Maar is

die korte snuit de oorzaak van de ademhalingsproblemen, die leiden tot kortademigheid en moeheid na fysieke inspanning? BOAS wordt door diergeneeskundigen gediagnosticeerd door het luisteren naar het ademen met een stethoscoop, en door het vergelijken van dat geluid in rust en na inspanning. De diagnose wordt zelfs voorzien van een gradatie, van 1 tot 4. Honden met ernstige BOAS leven kort, en hun leven is een lijdensweg. Per definitie kan BOAS alleen bestaan in kortsnuitige honden. De mens is de oorzaak van het bestaan van kortsnuitige honden en dus voor het bestaan van de ziekte BOAS. Geassocieerde interne vervormingen zijn de directe oorzaak van chronische moeite met ademen, leidend tot een algemeen slechte gezondheid en een korte levensduur.

In verschillende landen wordt momenteel vanuit de politiek actie ondernomen, deels onder druk van maatschappelijke "pro natuur en tegen dierenleed" bewegingen. In Nederland is het ministerie van landbouw, natuur en voedselkwaliteit (LNV) verantwoordelijk voor het welzijn van gezelschapsdieren. Enige jaren geleden is opdracht gegeven aan de Faculteit Diergeneeskunde van Universiteit Utrecht (de enige veterinaire school van Nederland) om te komen tot criteria waaronder fokken van honden verboden zou moeten zijn. Deze criteria dienden wetenschappelijk onderbouwd te zijn, en gebaseerd op makkelijk te bepalen externe kenmerken waarvan bovendien wetenschappelijk bewezen is dat zij oorzaak zijn van gezondheidsklachten en dus het "onwelzijn" van dieren. In eerste instantie heeft Utrecht problemen hiermee gezien. Uit WOB verzoeken blijkt dat de opdracht in eerste instantie niet aangenomen werd omdat men problemen zag bij BOAS. Maar uiteinde-



Figuur 1: CFR = afstand P1–P2 / afstand P5–P3, gemeten met een meetlint uit oma's naaidoos

lijk is de opdracht geaccepteerd en het resultaat is een publicatie (ook in Engelse vertaling beschikbaar) die een “stoplichten” systeem voorstelt.

Verschillende kenmerken worden genoemd die gemeenschappelijk hebben dat ze inderdaad eenvoudig te bepalen zijn door simpele externe metingen of waarnemingen. Sommige worden voorzien van gradaties. Tenslotte is een tabel ontstaan van groen, oranje en rood gekleurde kenmerken. Uiteraard is rood het ergste. Rood betekent: verbod op fokken met deze honden, ook verbod op eigenaar zijn van zulke honden, en ook verbod of vertoning (in media, advertenties, et cetera) van zulke honden.

Hier zal ik het hebben over het rode stoplicht gebaseerd op een eenvoudig te bepalen maat voor kortsnuitigheid: de *cranio-facial ratio* of CFR, zie figuur 1. Het wordt vaak als een percentage uitgedrukt. Utrecht komt tot een totaalverbod op fokken, houden of tonen van honden met een CFR onder de 30%.

Hun rapport Hagen (2019) aan het ministerie van LNV is een literatuurstudie. Naast anatomische, biologische en genetische werken wordt zwaar geleund op een tweetal epidemiologische studies uit Engeland: de artikelen Packer, Hendricks e.a. (2015) en Liu e.a. (2017), allebei verschenen in het tijdschrift *PLoS ONE*. Hoofdauteur van het eerste artikel is Rowena Packer van het “Royal Veterinary College” in Hatfield; een onderdeel van de University of London. De laatste en senior auteur van het tweede artikel is Jane Ladlow, van de University of Cambridge. Deze twee dames zijn duidelijk de wereldleiders op dit gebied. Een derde, heel belangrijke, literatuurverwijzing is de encyclopedie over brachycefale honden Packer en O’Neill (2022) (Hatfield groep), met arti-

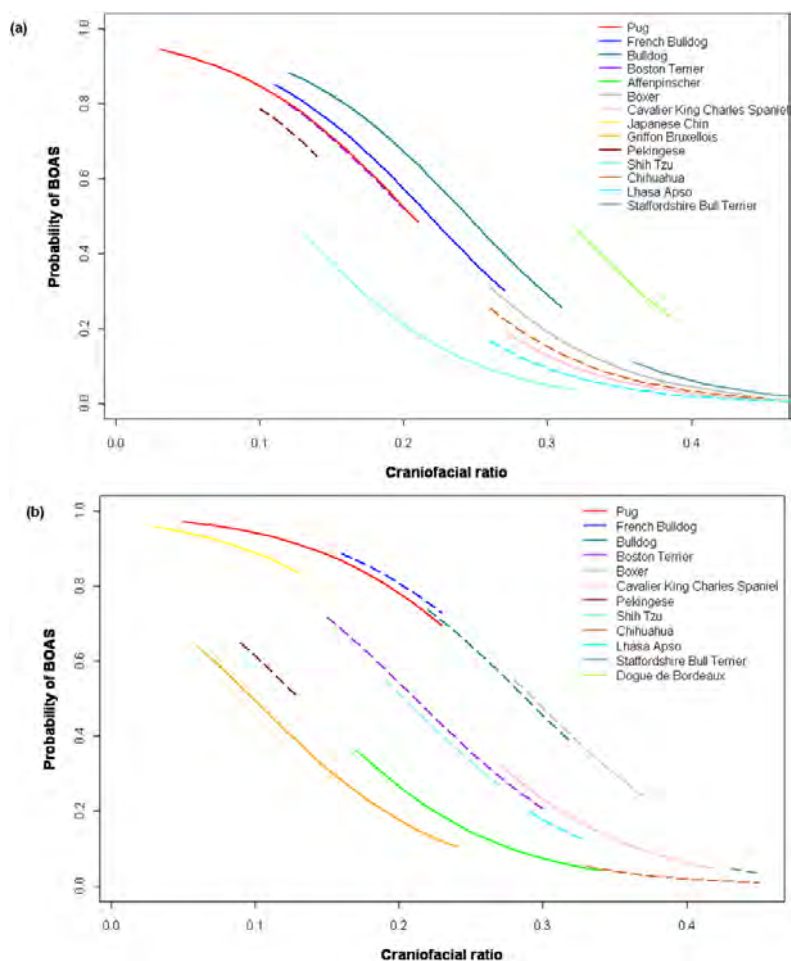
kelen ook van Ladlow en anderen uit Cambridge, alsook van nog meer deskundigen.

Data en statistiek

Het artikel van Packer, Hendricks e.a. (2015) is een verslag van twee studies, allebei exploratief van aard. Bij rashonden van diverse brachycefale soorten zijn verschillende metingen gedaan, waaronder CFR; en is de aan- of afwezigheid van BOAS genoteerd. Vervolgens is logistische regressie uitgevoerd met verschillende covariaten in een poging goede predictoren van BOAS te ontdekken. Om redenen die straks duidelijk zullen zijn is ervoor gekozen om “ras” als een “random effect” mee te nemen. De resultaten werden samengevat in een tweetal grafieken die de volle aandacht van de Utrechtse veterinairs kregen, een voor studie 1 en de andere voor studie 2, zie figuur 2. Het onderschrift is het oorspronkelijk gepubliceerde.

Laat ik eerst iets zeggen over de aard van de twee data-sets. “Studie 1” omvat honden die onder diergeneeskundige behandeling waren bij Packer en haar collega’s bij het Royal Veterinary College (of doorverwezen waren naar deze instelling). Dit zijn dus zieke honden. Naar aanleiding van de resultaten van Studie 1 werd een nieuwe studie opgezet, waarin honden van rassen zijn opgenomen waar BOAS voorkwam in Studie 1, en waar een poging is gedaan honden van rassen die in Studie 1 zeer weinig voorkwamen extra te rekruteren. Het zijn dus niet honden die bekend waren bij de dierenarts omdat ze ziek waren, maar voornamelijk honden in het bezit van fokkers.

Laten we een blik werpen op een aantal “sum-



Figuur 2: Predicted probability of brachycephalic dog breeds being affected by brachycephalic obstructive airway syndrome (BOAS) across relevant craniofacial ratio (CFR) and neck girth ranges. The risks across the CFR spectrum are calculated by breed using GLMM equations based on (a) Study 1 referral population data and (b) Study 2 non-referral population data. For each breed, the estimates are only plotted within the CFR ranges observed in the study populations. Dotted lines show breeds represented by <10 individuals. The breed mean neck girth is used for each breed (as stated in Table 2). In (b), the body condition score (BCS) = 5 (ideal bodyweight) and neuter status = neutered.

mary statistics” van deze twee studies, zie tabel 1 en tabel 2. In studie 1 (honden doorverwezen naar de diergeneeskundigen in Hatfield) waren maar twee Brusselse Griffons (waarvan één BOAS had) en maar één Affenpinscher (die het niet had). Voor studie 2 zijn dus veel meer honden van deze twee rassen gerekruteerd, namelijk 20 en respectievelijk 31. BOAS kwam voor bij maar 10% van deze honden. De slechte toestand van de mopshonden en de bulldogs (Engels en Frans) is duidelijk, alsook van Japanse Chin, Shi Tzu, de Boston Terrier en de Boxer.

Wat in ieder geval duidelijk is, is dat de mate van BOAS sterk afhankelijk is van het ras; en het lijkt er ook op dat het verband met CFR niet hetzelfde is voor de verschillende rassen. Het artikel Liu et al. (2017) was een vervolg op Packer, Hendricks e.a. (2015). Men had een steekproef alleen van mopshonden en van bulldogs (beide soorten), als zijnde

de rassen die het meeste last hebben van BOAS. Een belangrijke conclusie was dat het effect van BOAS verschilde per ras; dit was statistisch significant ondanks de kleine mate van spreiding van CFR binnen elk van de drie rassen. Zoals gezegd, de laatste jaren is erg intensief gefokt (met veel inteelt als gevolg) op het uiterlijk kenmerk van een platte snuit. Men heeft door fokken CFR heel klein en bijna constant gekregen.

Nu een terugblik op de mooie curves van Packer, Hendricks e.a. (2015). Volgens de Utrechtse veterinairs hebben Packer et al. *bewezen* dat het verband tussen CFR en BOAS er uitziet zoals in deze plaatjes. Maar dat hebben ze niet bewezen; dat hebben ze verondersteld! Bovendien, de analyse had de veronderstelling van random “ras”-effecten nodig om een prachtige curve voor elk verschillend ras te krijgen: veel rassen zijn in beide studies in erbarmelijk kleine

Study 1	n	Affected %	Median craniofacial ratio (IQR)	Median neck girth (IQR)	Min-Max predicted BOAS risk
Pug	32	88	0.08 (0.06)	32.2 (4.82)	0.69–0.97
French Bulldog	13	70	0.19 (0.06)	33.0 (6.25)	0.73–0.89
Bulldog	16	63	0.22 (0.11)	42.2 (7.58)	0.38–0.74
Boston Terrier	6	83	0.14 (0.04)	30.2 (2.55)	0.21–0.72
Japanese Chin	0	–	–	–	–
Pekingese	3	67	0.12	31.3	0.50–0.65
Dogue de Bordeaux	6	67	0.36 (0.03)	55.5 (9.15)	0.22–0.47
Griffon Bruxellois	2	50	0.13	22.6	0.24–0.55
Boxer	13	18	0.31 (0.06)	41.0 (5.85)	0.27–0.55
Shih Tzu	13	8	0.20 (0.07)	28.5 (3.25)	–
Chihuahua	5	40	0.34 (0.17)	20.0 (1.20)	0.01–0.05
Cavalier King Charles Spaniel	26	4	0.39 (0.07)	31.2 (4.85)	0.05–0.32
Affenpinscher	1	0	0.20	21.1	–
Staffordshire Bull Terrier	16	6	0.50 (0.07)	39.2 (4.92)	0.04–0.05

Tabel 1: Packer 0-1

Study 2	n	Affected (%)	Median craniofacial ratio (IQR)	Median neck girth (IQR)	Min-Max predicted BOAS risk
Pug	32	91	0.12 (0.06)	31.8 (3.10)	0.48–0.95
French Bulldog	4	75	0.18 (0.05)	35.3 (3.70)	0.30–0.85
Bulldog	6	33	0.25 (0.08)	43.8 (9.75)	0.26–0.88
Boston Terrier	2	50	0.23	28.2	0.84–0.96
Japanese Chin	10	80	0.04 (0.06)	23.8 (3.38)	–
Pekingese	3	0	0.11	28.0	0.66–0.79
Dogue de Bordeaux	1	0	0.36	55.1	–
Griffon Bruxellois	20	10	0.15 (0.06)	24.2 (3.23)	0.11–0.64
Boxer	4	50	0.30 (0.07)	38.2 (8.42)	0.02–0.31
Shih Tzu	7	43	0.22 (0.05)	29.2 (5.00)	–
Chihuahua	3	0	0.41	19.2	0.04–0.45
Cavalier King Charles Spaniel	11	18	0.36 (0.05)	29.2 (4.10)	0.02–0.26
Affenpinscher	31	10	0.23 (0.08)	23.6 (4.30)	0.01–0.19
Staffordshire Bull Terrier	2	0	0.45 (0.02)	38.8 (6.72)	0.04–0.36

Tabel 2: Packer 0-2

aantallen aanwezig.

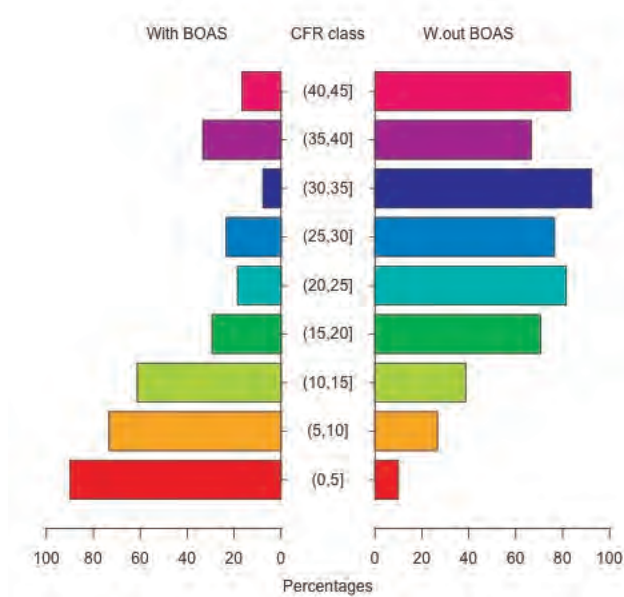
Ik heb een rapport geschreven voor mijn opdrachtgevers (fokkers van Brussels Griffon en van Affenpinscher), Gill (2022), waarin ik hun standpunt volledig onderbouw. De Utrechtse onderzoeksgroep heeft geen statistici. Ze hebben beweringen uit de twee Engelse studies niet goed weergegeven en ze zagen veel meer in Packer's resultaten dan verantwoord is. Ik geef hier een voorbeeld: er zijn genoeg data om het effect van CFR niet-parametrisch te modelleren, zie figuren 3 en 4.

Overigens wil ik niet op grond van deze plaatjes suggereren dat een verantwoord "rood stoplicht" zou kunnen zijn: CFR onder de 20%. Mijn conclusie, op basis van eigen analyses en de geciteerde literatuur, is dat de relatie tussen CFR en BOAS rasafhankelijk is en dat we voorlopig voor veel rassen

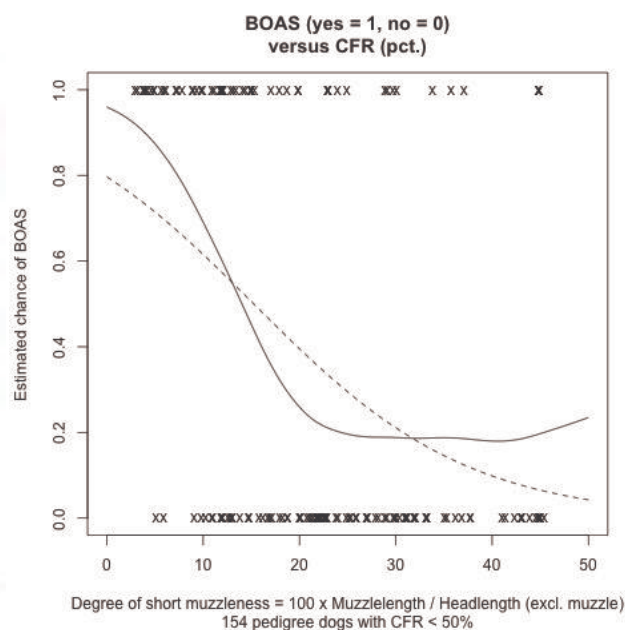
veel te weinig data hebben om met behulp van alleen epidemiologisch onderzoek te komen tot een eenvoudig ras-onafhankelijk stoplichten-systeem. Hondenfokkers en hun beroepsverenigingen en wetenschappers kunnen wel samen tot verstandige regels komen. Het is al bekend, voornamelijk dankzij engels onderzoek, hoe met verstandige fokprogramma's het BOAS probleem in een decennium of twee sterk verminderd kan worden.

Het vervolg

Mijn rapport heb ik omgebouwd tot een preprint, Gill (2022), en aangeboden aan *PLoS ONE*, het tijdschrift waarin Packer en Ladlow publiceerden. Tot mijn verassing werd mijn bijdrage onmiddellijk door een "desk-editor" geweigerd op formele gronden:



Figuur 3: 154 rashonden met CFR < 50%. 12 verschillende rassen bij elkaar



Figuur 4: BOAS versus CFR (logistisch effect versus gladde effect CFR)

het was geen origineel onderzoek. Ik had alleen data geanalyseerd uit andere onderzoeken (gepubliceerd in hun tijdschrift) en beredeneerd dat de getrokken conclusies misleidend, zo niet onjuist waren, en aangetoond dat dit al andere onderzoekers daadwerkelijk sterk misleid had. Uiteraard heb ik bij de redactie geklaagd. Ik mocht een nieuwe poging doen; mijn artikel moest nu heten "Commentary on ...". Opnieuw werd het artikel geweigerd, nu op grond van de mening dat ik niets nieuws en positiefs te melden had. Ik klaagde opnieuw. Tot nu toe had geen enkele persoon met relevante vakkennis naar het artikel gekeken. Het resultaat was: volstrekte stilte. Als lezers van dit stukje geïnteresseerd zijn om mijn preprint te lezen en te "reviewen" op de site pubpeer.com, graag.

Overigens heb ik het artikel ook opgestuurd aan Rowena Packer en Jane Ladlow. Ze reageerden allebei zeer goedkeurend maar wilden mij niet publiekelijk steunen: de wereld van diergeneeskundigen gespecialiseerd in kortsnuitige honden is heel klein en men wil zich niet mengen in politieke discussies in andere landen, en ook niet collega's afvallen. (Momenteel is een rechtszaak gaande in Oslo, en ook in andere Europese landen wordt juridische strijd geleverd.) Uiteraard heb ik ook contact proberen

te leggen met de Utrechtse veterinairs, maar die onthouden zich van commentaar.

Dit is natuurlijk allemaal jammer want zonder een instemmend peer review kan mijn werk makkelijk genegeerd worden door Nederlandse politici. (Als mijn werk Gill (2022) afgekraakt kan worden hoor ik dat ook graag, dan trek ik het ook onmiddellijk in!) Toch heb ik een weg gevonden om hun aandacht te krijgen. De Tweede Kamer heeft een speciale commissie die over deze zaken gaat. Eens per jaar ontmoet de minister de leden van de commissie, de verslagen van hun jaarlijkse bijeenkomsten zijn op internet te vinden. Het is fascinerend leesmateriaal. Bijvoorbeeld, uit het verslag van de commissie begin 2023:

Tweede Kamer, Dierenwelzijn, Commissiedebat 26 januari 2023

De heer Haverkort (VVD): "We hebben tijdens het vorige debat over dierenwelzijn een hele discussie gehad over aan de ene kant de inzichten die langs de lijn van de Cambridge methode gehanteerd worden en aan de andere kant de lijn van het ministerie, dat met de Universiteit Utrecht iets heeft afgesproken over de verhouding tussen de grootte van de kop

en de lengte van de neus. Ik meen dat die 30% moet zijn, uit mijn hoofd gezegd. Dat lijkt me ook vrij objectief. Ik kan mij nog niet aan de indruk onttrekken dat het dus vrij eenvoudig zou moeten zijn om die objectieve criteria voor kortsnuiten op korte termijn vast te stellen. Die kunnen dan heel snel op het lijstje van dieren komen die we niet meer willen houden. Kan de minister daar nog wat helderheid over verschaffen?"

Minister Adema: "Ik denk dat dat een wat te brede kwalificatie is. Het gaat echt om dieren die lijden aan het feit dat ze kortsnuitig zijn. Niet elke kortsnuitige hond komt er dus op. Het gaat om eigenschappen die maken dat dieren lijden aan hun doorgefokte eigenschappen."

De heer Haverkort (VVD): "Om het even heel scherp te hebben: zegt de minister hierbij dan dat de inzichten van de Universiteit Utrecht in het kader van de kortsnuitige honden en de verhouding tussen kop en neus, niet voldoende zijn om lijden van kortsnuitige honden te toetsen? Komt de minister daarmee terug op de beleidslijn van zijn voorganger? Dan hebben we namelijk wellicht nog meer te bespreken met elkaar. Het zou mij vervelend verrassen als dat zo is."

Hier wordt "de Cambridge methode" (dus: Ladlow c.s., luisteren naar ademhaling voor en na fysieke inspanning) vergeleken met de veel eenvoudiger "Utrecht methode". De commissie dacht dat de minister al besloten had om "Utrecht" aan te houden. Zo te zien, heeft de minister wel door dat "Utrecht" een botte bijl is (ook al uitgevoerd met een meetlint uit oma's naaidoos).

Ik heb geprobeerd met de vijf of zo parlementaire leden van de commissie in e-mail contact te komen, maar kreeg alleen reactie van een partij waar ik liever geen zaken mee zou doen. Ik vond het in het bijzonder jammer dat de Partij Voor de Dieren geen belangstelling toonde voor zich ontwikkelende wetenschappelijke inzichten over dierenleed. Toen zag ik dat men kan e-mailen aan de Griffier van de Commissie. Dat heb ik gedaan en kreeg het antwoord dat mijn bericht doorgegeven is aan alle commissieleden. Toen kreeg ik ook eindelijk reactie dat men best met mij wilde praten als de huidige drukke periode in de Tweede Kamer (na afloop van de verkiezingen Provinciale Staten) enigszins voorbij is. Leve onze democratie!

Richard Gill is emeritus hoogleraar statistiek en erelid van de VVSOR.
E-mail: gill@math.leidenuniv.nl



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

**OPROEP OM KANDIDATEN TE
NOMINEREN VOOR DE JAN
HEMELRIJK EN WILLEM R. VAN ZWET
AWARDS 2023**

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. De winnaar van afgelopen jaar was Fleur Theulen. Sinds 2012 is er ook een prijs voor dissertaties: de Willem R. van Zwet Award. Deze werd vorig jaar gewonnen door Collin Drenth.

De afgelopen jaren kwam het overgrote deel van de nominaties uit de hoek van de operations research en de mathematische statistiek. Zonder iets af te willen doen aan de goede kwaliteit van deze nominaties, is de jury ervan overtuigd dat ook in de andere secties uitstekend promotie- en afstudeeronderzoek gedaan wordt. Ook statistisch onderzoek met een toegepaste karakter kan in aanmerking komen voor deze prijs.

Wij willen dan ook supervisors (begeleiders) uit alle vakgebieden van harte uitnodigen om een uitmuntende afstudeerscriptie (Master) of dissertatie (Ph.D.) te nomineren. 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 21 januari 2024 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
Prof. dr. Casper Albers, voorzitter VVSOR



foto: Jean-Pol GRANDMONT

Clio's stiefkind en de geschiedenis van het probabilistische denken

Richard Starmans

Veerman tussen twee culturen

De Nederlandse wiskundige en wetenschapshistoricus E.J. Dijksterhuis (1892-1965) is vooral bekend vanwege zijn inmiddels klassieke *De Mechanisering van het Wereldbeeld* uit 1950, waarin hij de ontwikkeling van de natuurwetenschappen beschrijft vanaf de Pre-socratici, ongeveer 600 voor Christus tot en met de natuurkunde van Isaac Newton in de vroege 18^e eeuw. Voor dit werk ontving hij in 1951, amper één jaar na het verschijnen ervan, de P.C. Hooftprijs

der Letteren, die in die tijd afwisselend werd toegekend aan bijdragen op het gebied van fictie -met name proza en poëzie- en non-fictie, vooral historische studies of wijsgerige essays. In weerwil van dit onmiskenbare eerbetoon zou de laureaat, wederom één jaar later een kritisch artikel het licht doen zien, getiteld *Clio's stiefkind* (1952). Hierin beklaagt Dijksterhuis zich dat historici, doorgaans geschoold in de humaniora of letteren, weinig of geen kennis van de wetenschapsgeschiedenis bezitten en evenmin waardering ervoor tentoonspreiden. Het vak

gold toen nog als een relatief pril deelgebied van de historiografie. Dat alles was voldoende reden voor Dijksterhuis om Clio, de Muze van de geschiedschrijving ter verantwoording te roepen en te afficheren als de in sprookjes optredende, spreekwoordelijke boze stiefmoeder. De onvrede van Dijksterhuis had alles te maken met zijn hooggestemde ambitie om de wetenschapsgeschiedenis om te smeden tot de verbindende schakel tussen de natuurwetenschappen en de humaniora, tussen alfa en bèta. De dikwijls gepercipieerde kloof tussen beide werelden zou kunnen worden overbrugd door resultaten uit de exacte wetenschappen in een historische en culturele context te bestuderen en toegankelijk te maken; ideeëngeschiedenis in optima forma. Dat dit anno 1952 als een vorm van hooggestemd utopisme kon worden beschouwd bleek wel enkele jaren later toen de Britse chemicus en romanschrijver C.P. Snow (1905-1980) zijn *Two Cultures and the scientific revolution* deed verschijnen.

Volgens Snow was het intellectuele leven in de Westerse samenleving verscheurd door een diepge wortelde tegenstelling tussen de “sciences” en de “literary intellectuals”. Die tegenstelling kent uiteraard een lange traditie; zij gaat terug tot Aristoteles, krijgt vervolgens gestalte in het trivium en quadrivium tijdens de scholastiek en wordt vooral manifest in de 19^e eeuw toen een fragmentatie van de kennis optrad en bijvoorbeeld William Dilthey het “Erklären” in de natuurwetenschappen stelde tegenover het “Verstehen” in de humaniora.

Snow ging een stap verder en benadrukte de implicaties voor de moderne tijd. De kloof was verdiept en met name het feit dat beleidsmakers en politici doorgaans in de niet-exacte traditie waren geschoold zag hij als een belangrijk obstakel om grote problemen in de samenleving op te lossen. Ofschoon Snow zijn standpunt later iets reviseerde, veranderde de sombere strekking ervan niet en ook vandaag de dag lijkt er weinig reden tot optimisme; Dijksterhuis’ verwachting is niet ingelost en zijn project allerminst voltooid. Wel heeft de wetenschapsgeschiedenis inmiddels een emancipatieproces doorlopen en een volwaardige en geïnstitutionaliseerde plaats ingenomen in het palet van historische aandachtsvelden, met eigen leerstoelen, conferenties en tijdschriften. Toch achtte de Groningse wetenschapshistoricus Klaas van Berkel (1953) het essay *Clio’s stiefkind* in 1990 nog voldoende saillant en karakteristiek om een door hemzelf geredigeerde bloemlezing met opstellen van de hand van Dijksterhuis -door hem getypeerd als “veerman tussen twee culturen”- onder dezelfde titel te doen

verschijnen (Dijksterhuis, 1990).

Bij dit alles dringt zich een nieuwe vraag op: welke rol speelt de geschiedenis van het probabilistische denken (ruwweg kansrekening en statistiek) eigenlijk binnen de nu gevestigde algemene wetenschapsgeschiedenis? Herhaalt de geschiedenis zich en is ook hier sprake van een ondergeschoven positie, waardoor historici van de statistiek zich met enig recht te buiten kunnen gaan aan zelfbeklag en op hun beurt Clio verwijten maken? Dat zou zo maar kunnen, bijvoorbeeld omdat het vak moet concurreren met de historiografie van meer gevestigde disciplines, zoals wiskunde, astronomie, natuurkunde, scheikunde en biologie, waarop traditionele wetenschapshistorici doorgaans hun aandacht vestigen. Of komt het probabilistische denken er juist allerminst bekaaid van af en is er volop sprake van waardering en kruisbestuiving? Ook dat zou zomaar kunnen, gelet op het feit dat de laatste 100 jaar vele wetenschappen een probabilistische wending hebben ondergaan. Hun concepten en methoden en zelfs het achterliggende wereldbeeld worden primair begrepen en benaderd vanuit kansrekening en statistiek. Vele van de in de 19^e eeuw ontstane nieuwe “variatie-en-verandering rijke” wetenschappen konden bovendien vooral in wisselwerking met de statistiek tot bloei komen; psychologie, sociologie, economie, biologie, genetica, landbouwwetenschap, epidemiologie. Een volledig antwoord op deze vraag overstijgt de reikwijdte van dit korte essay. We beperken ons tot een viertal sail-lante momenten of episodes uit de historiografie van de statistiek: de “dageraad”, de “gouden era”, de “statistische stijl als historische bouwsteen” en tot slot de “grenserving”.

De dageraad: Whewell en Todhunter

In 1837 publiceerde de Engelse wetenschapper, filosoof en anglicaanse geestelijke William Whewell (1794-1866) zijn befaamde studie *History of the inductive sciences, from the Earliest to the Present Times*, die vaak als beginpunt van de wetenschapsgeschiedenis wordt beschouwd. In die tijd was de beoefening ervan vrijwel uitsluitend het domein van praktiserende, dikwijls vermaarde natuurwetenschappers, die zich in hun vrije tijd wijdden aan de bestudering van de geschiedenis van hun vak en hun bevindingen optekenden in inleidende hoofdstukken van leerboeken of handboeken. Al deze studies hadden een hoog technisch gehalte, waren geschreven “door ingewijden voor ingewijden” en bleken voor doorsnee historici ontoegankelijk. Daarbij waren zij meestal uitsluitend op interne ontwikkelingen

binnen het vakgebied gericht, de resultaten, theorieën en experimenten, zonder verankering in een algemene culturele of filosofische traditie. Deze internalistische benadering zou tot grote successen leiden (Snelders, 1979), maar kende ook kritiek. Tegenstanders waarschuwden voor vormen van presentisme, “whig-history” en finalisme (Jonker, 2011). Natuurwetenschappelijke ontwikkelingen worden dan vanuit een hedendaags standpunt gezien, waarbij de superioriteit van de huidige kennis een vertrekpunt vormt, de geschiedenis een rationele reconstructie ondergaat, en het wetenschappelijk bedrijf als een objectieve en waardenvrije queeste naar waarheid wordt voorgesteld, een ware zegetocht, waarbij de successen en hoogtepunten worden uitvergroot en minder succesvolle bijdragen of zijpaden van onderzoek stelselmatig onderbelicht. Door hyperspecialisatie van de exacte wetenschappen dreigt de zuiver internalistische historische invalshoek zich bovendien te verliezen in steeds specifiekere, alleen nog door ingewijden te begrijpen detailstudies. Als antidotum ontstond daarom al spoedig een externalistische traditie, waarbij wetenschap volledig wordt geanalyseerd, begrepen en verklaard vanuit externe factoren. Deze betreffen allereerst de culturele, politieke, institutionele en maatschappelijke con-

text, maar ook privé-opvattingen, drijfveren en oogmerken van onderzoekers zijn, waarbij niet zelden een persoonlijke, b.v. marxistische ideologie van de historicus vooropstaat en als algemeen interpretatiekader fungeert (Starmans, 2023). Tegenstanders van deze aanpak benadrukken op hun beurt dat de externalistische invalshoek gepaard kan gaan met een ideologische vooringenomenheid, die afbreuk doet aan de historische werkelijkheid, elke vorm van objectiviteit en rationaliteit van wetenschap a priori uitsluit of ten minste ondermijnt. Of zij stellen dat de wetenschapsgeschiedenis niet als een “blackbox” mag worden benaderd door historici of sociologen die natuurwetenschappelijk ongeschoold zijn en elke kennis van de specialistische deelgebieden ontberen. Sommigen zochten de oplossing van dit dilemma in een ideeënhistorisch internalisme, waarbij wel degelijk inhoudelijke (wiskundige, experimentele en technische) ontwikkelingen worden geanalyseerd in diepgravende historische case-studies, maar niet worden ontdaan van hun intellectuele, wijsgerige context en andere culturele determinanten. In dit licht was het vroege ideeënhistorische werk van Whewell baanbrekend en hij mag dan ook worden beschouwd als heraut van deze stroming, die pas echt gestalte kreeg met de historisch-wijsgerige bij-

National Portrait Gallery, London



Francis Galton en zijn biograaf Karl Pearson (links)

dragen van de natuurkundigen en filosofen Ernst Mach en Pierre Duhem. Ook de twee beroemdste Nederlandse wetenschapshistorici, de voornoemde E.J. Dijksterhuis en de chemicus Reyer Hooykaas passen in deze traditie. Hoe dan ook, in het werk van Whewell en in andere studies uit de beginperiode van de wetenschapsgeschiedenis was inderdaad weinig aandacht voor de opkomst van kansrekening en statistiek. Liever dan bij de pakken neer te zitten of te wachten op Clio's aandacht, besloten vele generaties statistici de geschiedschrijving dan maar zelf ter hand te nemen. De wegbereider van deze traditie was zonder meer de Britse wiskundige Isaac Todhunter (1820-1884) die zelf belangrijke bijdragen leverde aan calculus, geometrie, mechanica en statistiek, maar ook de geschiedschrijving ervan beoefende. Zo publiceerde hij al in 1865 *A Mathematical Theory of Probability from the Time of Pascal to that of Laplace*. In een buitengewoon systematische opgebouwde studie van ruim 650 bladzijden bespreekt hij eerst de wegbereiders in het probabilistische denken, waaronder Cardano, Kepler en Galilei om dan de wieg van de kansrekening te situeren medio 17^e eeuw met het werk

van Pascal, Fermat, Huygens, Bernoulli en anderen. Via onder meer Montmort, De Moivre, Euler, d'Alembert, Bayes, Lagrange en Condorcet komt hij dan uit bij Laplace, wiens bijdragen hij als een culminatiepunt beschouwt. Naast Christiaan Huygens -die merkwaardigerwijs slechts vijf bladzijden krijgt toebedeeld- komen ook internationaal minder beroemde landgenoten aan bod, waaronder Van Schooten, De Wit, Van Hudde en 's Gravesande. De studie is ook na bijna 170 jaar nog steeds zeer lezenswaardig. Maar ook latere generaties statistici kozen een pragmatische opstelling om te zorgen dat hun bijdragen direct in de annalen der wetenschapsgeschiedenis konden worden opgetekend, onder meer door het laten vervaardigen van (intellectuele) biografieën door erudiete familieleden, geestverwanten of leerlingen. We beperken ons hier tot de founding fathers van de statistiek: Francis Galton (1822-1911), Karl Pearson (1857-1936) en Ronald Fisher (1890-1962). Zo kon Francis Galton erop rekenen dat zijn protegé en opvolger Karl Pearson deze taak op zich nam, hetgeen resulteerde in *The life, letters and labours of Francis Galton* (1914, 1924, 1930). Zelf zou Karl Pearson de zorg voor zijn intel-



Designed by starline / Freepik

lectuele nalatenschap toevertrouwen aan zijn zoon Egon Pearson, die dit ter hand nam in onder meer Karl Pearson: an appreciation of some aspects of his life and work (1938). Egon ging als historicus echter veel verder blijkens zijn omvangrijke *The history of statistics in the 17th and 18th centuries* uit 1929 en het latere *Studies in the history of statistics and Probability* uit 1969. Bij Fisher ging het weer iets anders: zijn dochter, Joan Fisher Box, die getrouwd was met de statisticus George Box, schreef in 1978 een biografie van haar vader R.A. Fisher, *life of a scientist*. Natuurlijk kan door deze summiere opsomming een gechargeerd beeld van de vroege geschiedschrijving van kansrekening en statistiek ontstaan, maar vast staat dat door de dikwijls internalistische aanpak veel historisch materiaal aan de vergetelheid werd onttrokken, statistische bijdragen relatief snel een bescheiden plaats in de ideeëngeschiedenis konden innemen en op deze manier bovendien canonvorming werd versterkt. Traditionele historici daarentegen zetten soms vraagtekens aangaande de vereiste academische distantie en objectiviteit, die nodig zijn voor bijvoorbeeld een kritische, intellectuele biografie. Zij zijn dan ook geneigd deze dageraad van de geschiedenis van de statistiek, die met Todhunter begon als de “amateurperiode” te typeren, een noodzakelijke aanloop tot een “professionele” historiografie. We laten dit hier verder rusten, maar besluiten met de constatering dat kansrekening en statistiek ook in de geschiedschrijving van de wiskunde lange tijd een uiterst bescheiden plaats innamen. Een saillant voorbeeld is de bekende studie van Dirk Struik, *Geschiedenis van de Wiskunde* uit 1948.

De gouden era

Die gezochte professionalisering van de historiografie van de statistiek diende zich trouwens al spoedig aan en luidde een gouden era in, die we hier kort schetsen en die ruwweg het laatste kwart van de 20^e eeuw bestreek. Deze “periode” nam een aanvang in 1975 toen de recentelijk overleden Ian Hacking *The Emergence of Probability* deed verschijnen, waarin de focus ligt op de 17^e eeuw, een studie die 15 jaar later een vervolg kreeg met *The Taming of Chance*, dat de revolutie in het probabilistische denken in de 19^e eeuw centraal stelt. Beide niet-technische, ideeënhistorische werken toonden een historisch-wijsgerig perspectief op het probabilistische denken, dat bijna paradigmatisch is geworden en waarin Dijksterhuis’ hooggestemde ambities resoneren. In de tussenliggende jaren werden vele projecten op het gebied van de historiografie van de statistiek

gestart. Eén van deze betraf de vraag in welke mate in de 19^e eeuw een probabilistische revolutie heeft plaatsgevonden. Het jarenlange project, uitgevoerd in Duitsland aan de universiteit van Bielefeld, leidde tot een ruim 1000 pagina’s tellend tweedelig standaardwerk *The Probabilistic Revolution I, II* (Krüger, 1987a, 1987b) waaraan tal van toonaangevende denkers bijdroegen, waaronder Thomas Kuhn, Hacking en vele anderen. Daarnaast zagen in de jaren 80 en 90 van de vorige eeuw vele studies het licht, waarin de toenemende professionalisering zichtbaar werd; het historische “gehalte” nam toe onder meer door gedetailleerd bronnenonderzoek, internalistische en ideeënhistorische bijdragen wisselden elkaar af in hoog tempo.

Eén van de belangrijkste bijdragen betraf het werk van de Deense statisticus Anders Hald (1913-2007), die zich niet alleen toelegde op strikt internalistische bijdragen over bijvoorbeeld de parametrische statistiek vanaf Bernoulli tot en met Fisher (Hald, 2007), maar traditionele (ideeën-)historische benaderingen niet uit de weg ging (Hald, 1998, 2003). Dat laatste gold a fortiori voor het werk van bijvoorbeeld de Amerikaanse historicus Steven Stigler, die op basis van diepgaand bronnenonderzoek vele belangwekkende bijdragen leverde variërend van de wortels van het Bayesiaanse denken tot de bekende Wet van Stigler over de Eponymie (Starmans, 2022a). Belangrijk is dat naast vakstatistici intellectuelen van velerlei pluimage bijdroegen aan deze bloeiperiode. We noemen hier slechts bijdragen van de Amerikaanse historicus Theodore Porter, allereerst *The rise of Statistical Thinking, 1820-1900* (1990) en later *In numbers we trust; the pursuit of objectivity in science and public life* (1995), vervolgens *Classical Probability in the Enlightenment* (1988) van de hand van de Amerikaanse wetenschapshistorica Lorraine Daston en tot slot de door de Duitse psycholoog Gerd Gigerenzer geredigeerde bundel *The Empire of Chance: how probability changed science and everyday life* (1989). Ook in Nederland was deze bloeiperiode waarneembaar. Zo verscheen in 1992 in *Genewa*, het toenmalige tijdschrift van de Nederlandse Vereniging voor de Geschiedenis van de Natuurwetenschappen en Geneeskunde een themanummer, getiteld *De met cijfers bedekte negentiende eeuw; toepassing van statistiek en waarschijnlijkheidsrekening in Nederland en Vlaanderen tussen 1840 en 1920*. De bundel werd geredigeerd door Ida Stamhuis en Annemarie de Knecht-van Eekelen en gaf een goed overzicht van de probabilistische wending in de wetenschappen met name vroege toepassingen in de astronomie, geneeskunde, plantkunde

en genetica. Het tijdschrift beoogde daarmee een omissie te corrigeren betreffende het werk van Krüger, waarin de Nederlandse en Vlaamse bijdragen onderbelicht waren gebleven. Tot slot moet worden opgemerkt dat deze “gouden era” mede kon ontstaan door de historisering van de wetenschapsfilosofie, die reeds eerder in gang werd gezet door filosofen als Karl Popper, Thomas Kuhn en Paul Feyerabend, maar in dit tijdvak min of meer werd voltooid. Zowel historici, als wetenschapsfilosofen onderkennen in toenemende mate het belang van het probabilistische denken en de daar uit voortkomende kruisbestuiving betaalde zich ook op dit terrein uit. Bijdragen op het gebied van de filosofie van de statistiek van met name de Amerikaanse filosofe Deborah Mayo, de Italiaanse historica Maria Carla Galavotti en later van de statisticus Andrew Gelman moeten in dit verband worden genoemd (Starmans, 2022b).

De statistische stijl van Crombie

Dit brengt ons bij een derde saillante ontwikkeling in de historiografie van de statistiek. Deze kwam voort uit de geschiedschrijving zelf en betrof de identificatie van de statistische stijl als historische bouwsteen. Waar historici doorgaans structuur aanbrengen in de stroom en volgtijdelijkheid van gebeurtenissen via traditionele periodisering met identificatie van tijdperken, scholen en stromingen, koos met name de Australische wetenschapshistoricus Alistair C. Crombie (1915-1996) aan het eind van zijn carrière voor een radicaal andere invalshoek. In 1995 verscheen zijn driedelige magnum opus *Styles of Scientific Thinking in the European Tradition; the history of argument and explanation especially in the Mathematical and Biomedical Sciences and Arts*. In dit boek, waaraan hij sedert de vroege jaren 80 had gewerkt identificeert de auteur een zestal stijlen van redeneren, die de loop van de wetenschapsgeschiedenis hebben bepaald en ook meer recht doen aan de dynamiek van het wetenschappelijk bedrijf zelf. Het gaat om:

- de deductieve/postulationele stijl
- de experimentele stijl
- de hypothetisch-analoge stijl
- de classificierend/taxonomische stijl
- de statistisch/probabilistische stijl
- genetisch/evolutionair/historische stijl.

De stijlen zijn niet strikt gedefinieerd; het zijn geen methoden, paradigma's (Kuhn), *epistèmès*

(Foucault), onderzoeksprogramma's (Lakatos) of onderzoekstradities (Laudan) in strikte zin, al is er in sommige opzichten verwantschap. Ze volgen elkaar niet op, sluiten elkaar niet uit, overlappen elkaar soms, of spelen simultaan een rol bij een nieuwe theorie of ontdekking of zelfs het ontstaan van nieuwe disciplines. Dat neemt niet weg dat zij voldoende onderscheidend zijn; niet alleen als structurerend principe voor historici en filosofen, maar ook bij het bestuderen van menselijke en kunstmatige intelligentie. Het gaat om stijlen van redeneren, die uiteenlopende cognitieve vaardigheden betreffen, waarover mensen of “artificial agents” moeten beschikken (stellingen bewijzen, inductief en analoog redeneren, hypothetiseren, classificeren, processen identificeren, beslissen, etc.) en die bijgevolg ook in de AI en in machine learning een rol spelen. Met enige goede wil kan een zesdimensionale ruimte worden opgespannen, waarin elk probleem of elke taak kan worden gerepresenteerd.

Uiteraard spelen niet alle combinaties in elke discipline of in elke historische periode een even grote rol. Belangrijk is ook dat ze op verschillende tijdstippen “ontstonden”, niet logisch uit elkaar zijn af te leiden en eigen concepties van kennis en criteria voor waarheid met zich kunnen meebrengen. Hoe dan ook, aan het stijlenconcept en de interpretatie ervan werd ook door anderen gesleuteld. Zo besefte ook Hacking in het kielzog van Crombie dat met dit kader als het ware de bouwstenen van de wetenschapsgeschiedenis kunnen worden blootgelegd. In Nederland is het vooral de Amsterdamse wetenschapshistoricus Chunlin Kwa, die verantwoordelijk is voor de popularisering en uitwerking van het stijlenconcept, allereerst in zijn *De ontdekking van het weten, een alternatieve geschiedenis van de wetenschap* (2006) maar ook in zijn latere werk. Daarin wordt niet alleen de relevantie voor de humaniora, de sociale wetenschappen en de techniek beargumenteerd, maar fungeert het stijlenconcept ook als basis vormt voor een inleiding in de wetenschapsfilosofie, bijvoorbeeld in *Kernthema's in de wetenschapsfilosofie* (2014) en in *Wat is waarheid? Basisboek wetenschapsfilosofie* (2017). Het gevolg van dit alles is dat met het identificeren van de “statistische stijl” het probabilistische denken een van de bouwstenen van de wetenschapsgeschiedenis -om niet te zeggen- de ideeëngeschiedenis is geworden, daarmee als vanzelf deel uitmaakt van (de geschiedenis van) de afzonderlijke vakwetenschappen, bij uitstek geschikt is om de eenheid of verscheidenheid van de wetenschappen te duiden of bijvoorbeeld de rationaliteit van de wetenschappelijke kennis te

bestuderen. Met de identificatie van de statistische stijl is feitelijk na het “stiefmoederlijke” begin en na de eerder geschetste gouden era duidelijk de voltooiing van een emancipatieproces zichtbaar.

Grensvervaging

De aangekondigde vierde episode is enigszins complex en wordt hier slechts aangestipt. Het zou onjuist zijn om in de geschetste historiografie van een triptiek van opkomst, bloei en neergang te spreken. Dat laatste is zeker niet het geval, al is er wel sprake van enige windstille na de stroom van succesvolle publicaties in de periode 1975-2000. Die windstille betreft een zekere terughoudendheid inzake de bestudering van de geschiedenis van het probabilistische denken na WO II en vooral die van de vroege 21^e eeuw. Deels komt dit doordat het beschrijven van de recente geschiedenis, waar de auteurs dikwijls zelf deel van uitmaken, altijd een hachelijke zaak is. Velen willen meer distantie, al toonden de hierboven geschetste voorbeelden uit het verleden aan dat dit dikwijls helemaal niet nodig is. Deels komt het ook omdat het probabilistische denken vele gezichten heeft. De statistiek kent zelf traditiegetrouw vele benaderingen en invalshoeken, en er is bovendien al lange tijd wisselwerking met bijvoorbeeld operations research, epidemiologie en econometrie. Belangrijker nog is het gegeven dat heden ten dage het probabilistische denken ook wordt gevoed vanuit “verwante” vaak nieuwe disciplines waaronder data science, machine learning, data mining, knowledge discovery in databases, pattern recognition, causal inference, de subsymbolische AI (neurale netwerken en het daaruit voortkomende deep learning). De voornoemde aandachtsgebieden zijn niet “mutually exclusive” of “totally exhaustive”, sommige lijken zelfs in onbruik te geraken of zijn geannexeerd door een groter onderzoeksdomein; er is in deze ontwikkelingsgang onmiskenbaar sprake van grensvervaging, maar alle stromingen komen voort uit of bouwen voort op de probabilistische revolutie. Ook die heterogeniteit maakt de geschiedschrijving problematischer dan die van de gecodificeerde en “voltooide” bijdragen die hierboven werden besproken. Met name de spectaculaire ontwikkelingen in data science en machine learning leiden soms tot historische beschouwingen in blogs, onderzoeksartikelen of handboeken, vaak wederom door vakspecialisten, internalistisch, “door ingewijden voor ingewijden” en fragmentarisch, vergelijkbaar met de situatie 170 jaar geleden in de beginperiode van de historiografie van het probabilistische denken.

Soms gaat het daarbij ook hier om de pioniers of protagonisten van het eerste uur (Sejnowski, 2018). Al met al lijkt de geschiedenis zich enigszins te herhalen. Nu vele wetenschappen sterk afhankelijk van deze data-intensieve en op de “learning” metafoor gebaseerde disciplines is het zaak een volgende stap te zetten; wederom om canonvorming te bevorderen, nieuwe bijdragen relatief snel een bescheiden plaats in de ideeëngeschiedenis te laten innemen, en bovenal middels nauwkeurig bronnenonderzoek van het overvloedig beschikbare “historische” materiaal, de met enige moeite verkregen aandacht van Clio niet te verliezen. Mede in het licht van de geschetste grensvervaging is daarmee een voor de hand liggende, maar belangrijke uitdaging voor de huidige historiografie van het probabilistische denken benoemd.

Literatuur

- Dijksterhuis, E.J. (1990) *Clio's stiefkind*, Bert Bakker.
- Hald, A. (1998) *A History of Mathematical Statistics from 1750 to 1930*. New York: Wiley.
- Hald, A. (2003) *A History of Probability and Statistics and Their Applications before 1750*. Hoboken, NJ: Wiley.
- Hald, A. (2007) *A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713–1935*. New York: Springer
- Jonker, E. (2011). “Van relativisme naar oordeelsvorming. Recente tendensen in de wetenschapsgeschiedschrijving”. In: *Studium*, 4(1).
- Krüger, Lorenz; Daston, Lorraine and Heidelberger, Michael. (Eds.). (1987a) *The Probabilistic Revolution, Volume I: Ideas in History*, MIT Press.
- Krüger, Lorenz, L. Daston, M. Heidelberger, G. Gigerenzer & M. S. Morgan (1987b): *The Probabilistic Revolution, Volume II*, Cambridge/MA, 1987.
- Sejnowski, T.J. (2018) *The Deep Learning Revolution; artificial intelligence meets human intelligence*, MIT Press.
- Snelders, H.A.M. (1979) “Terugzien naar het verleden, een les voor het heden”, Wageningen.
- Starmans, R.J.C.M. (2022a) Starmans, R. J. C. M. (2022). Stiglers Wet, het Mattheüseffect of Selectiebias? In: *STAtOR*, 23(2).
- Starmans, R.J.C.M. (2022b) *Statistiek en Filosofie; een voortschrijdende samenspraak*. In: *STAtOR*, 23(3).
- Starmans, R.J.C.M. (2023) “Het verlate gelijk van Robert Merton: over instituties, normen, crises en de genealogie van serendipiteit”. In: *Filosofie-Tijdschrift*, 33(3).

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



De ene peiling is de andere niet

In de gemeenteraad van Alphen a/d Rijn was er in 2014 en 2015 discussie over het instellen van koopzondagen. De meningen liepen ver uiteen. Zo wilde de VVD de winkeliers zelf laten beslissen over hun openingstijden, terwijl de SGP en ChristenUnie de winkels op zondag gesloten wilden houden. De lokale coalitiepartijen (CDA, Nieuw Elan, VVD, D66 en ChristenUnie) kwamen er niet uit. Daarom besloten ze om met een peiling onder de inwoners van Alphen a/d Rijn te vragen wat zij ervan vonden. Zo'n peiling kun je op diverse manieren doen, maar welke manier is de beste? In Alphen a/d Rijn bleek het mogelijk om verschillende manieren van peilen met elkaar te vergelijken.

Het oorspronkelijk plan van de politici van Alphen a/d Rijn was om inwoners op straat in de winkelcentra te interviewen over het instellen van koopzondagen. Toen duidelijk werd dat dit geen representatieve steekproef zou opleveren, besloten ze om ook een peiling te doen met het AlphenPanel, het online burgerpanel van de gemeente Alphen a/d Rijn.

Om alle inwoners van Alphen a/d Rijn de gelegenheid te geven hun mening kenbaar te maken, werd ook nog eens een online peiling met zelfselectie georganiseerd. En zo waren er drie peilingen tegelijkertijd. Ze gebruikten dezelfde vragenlijst en werden op hetzelfde tijdstip uitgevoerd. Dat biedt een mooie gelegenheid om deze drie verschillende manieren van peilen met elkaar te vergelijken.

Peilen in de winkelcentra

De eerste manier van peilen was het interviewen van de bezoekers in de winkelcentra in de gemeente.

	Burgerpanel-peiling	Online peiling	Doelpopulatie
Aarlanderveen	2	1	1
Alphena/dRijn	70	55	66
Benthuizen	3	13	3
Boskoop	13	18	14
Hazerswoude-Dorp	4	8	5
Hazerswoude-Rijndijk	4	2	5
Koudekerka/dRijn	3	2	4
Zwammerdam	1	1	2
Totaal	100	100	100

Tabel 1: De verdeling van de respons en de doelpopulatie over de kernen (percentages)

Op één dag, zaterdag 17 januari 2015, tussen 11.00 en 15.00 uur, kon het winkelende publiek een vragenlijst invullen. Ze werden daarvoor op straat aangesproken door vrijwilligers van de deelnemende politieke partijen. Die gedroegen zich niet als neutrale enquêteurs, maar waren duidelijk herkenbaar als vertegenwoordiger van een politieke partij. Je kunt je afvragen of dit geen negatieve invloed heeft gehad op de objectiviteit en representativiteit van de enquête. Je kon hooguit concluderen dat deze peiling representatief was voor het winkelende publiek op zaterdagmiddag. Daarmee werden groepen genegeerd die om wat voor reden dan ook niet op zaterdag winkelen of niet door de politici wensten te worden aangesproken.

Peilen met een burgerpanel

De tweede manier van peilen was het inschakelen van het AlphenPanel. Dit online burgerpanel bestond uit een steekproef van inwoners van de gemeente Alphen a/d Rijn die toegezegd hadden om op gezette tijden een vragenlijst in te vullen. Het panel bestond aanvankelijk uit 450 personen die zichzelf hadden aangemeld. Het is echter later uitgebreid tot 1600 personen die via loting uit het bevolkingsregister waren geselecteerd. Dit panel was redelijk representatief. Je kon het dus gebruiken om uitspraken te doen over de gehele bevolking van Alphen a/d Rijn.

Online peilen met zelfselectie

De derde manier van peilen was een online peiling. De steekproef hiervoor was gebaseerd op zelfselectie en niet op een aselect trekken. De vragenlijst werd aangeboden via een link op de website van de gemeente Alphen a/d Rijn. Ook werd de link vermeld in berichtgeving in de media. Iedereen kon

de vragenlijst invullen. Dus ook mensen die niet in Alphen a/d Rijn woonden.

Specifieke groepen deden oproepen aan hun leden om toch vooral de vragenlijst in te vullen. Zo schreef de kerkenraad van de Gereformeerde Kerk in Boskoop (een van de dorpen in de gemeente) op haar website:

“Graag roepen wij u op aan deze enquête mee te doen en uw mening kenbaar te maken. De tendens in de samenleving is dat de zondag steeds meer gezien wordt als een gewone dag in plaats van een geheiligde dag, een dag van afzondering en rust. Laten we daarom met elkaar onze verantwoordelijkheid nemen, onder biddend opzien tot de Heere die alles regeert en bestuurt.”

De Hervormde gemeente in Benthuizen (ook een dorp in de gemeente) riep op 3 februari 2015 haar leden op om te gaan stemmen:

“Op de website van de Gemeente Alphen aan de Rijn kunt u en kan jij een enquête invullen over het onderwerp koopzondagen. Het moderamen beveelt invulling van deze enquête bij u en jou van harte aan.”

En dan was er ook nog de oproep van de gemeente Alphen a/d Rijn op haar website om de enquête meer dan één keer in te vullen:

“Ook u bent van harte uitgenodigd om aan deze enquête deel te nemen, ook al heeft u ook al eerder bij een van de winkelcentra in de gemeente meegedaan.”

Dit soort oproepen droeg zeker niet bij tot de representativiteit van de steekproef. In tegendeel, de representativiteit werd er juist door aangetast.

Representatieve steekproeven?

De drie manieren van peilen overziend, kwam de peiling via het burgerpanel nog het dichtst in de buurt van een representatieve steekproef. Daarom kun je deze gegevens het beste gebruiken om een indruk te krijgen van de mening van de Alphenaren. Het is verstandig om de twee andere aanpakken



(a)



(b)

Figuur 1: (a) De stedelijke kern Alphen a/d Rijn en (b) het dorp Benthuisen. Foto's: Jelke Bethlehem.

te negeren. Er is een hardnekkig misverstand dat problemen met peilingen vanzelf verdwijnen als de steekproef maar groot genoeg is. Helaas is dit niet zo. Vertekeningen ten gevolge van mankementen in de opzet en uitvoering peiling blijven bestaan, hoeveel gegevens je ook verzameld. Daarom is het niet verstandig om de gegevens van de winkelcentrum-peiling en de online peiling te voegen bij die van de burgerpanel-peiling. Dit zou een redelijk goed gegevensbestand alleen maar vervuilen. De politieke partijen namen uiteindelijk ook het verstandige besluit de uitkomsten van de burgerpanel-peiling leidend te laten zijn bij de politieke besluitvorming.

De uitkomsten

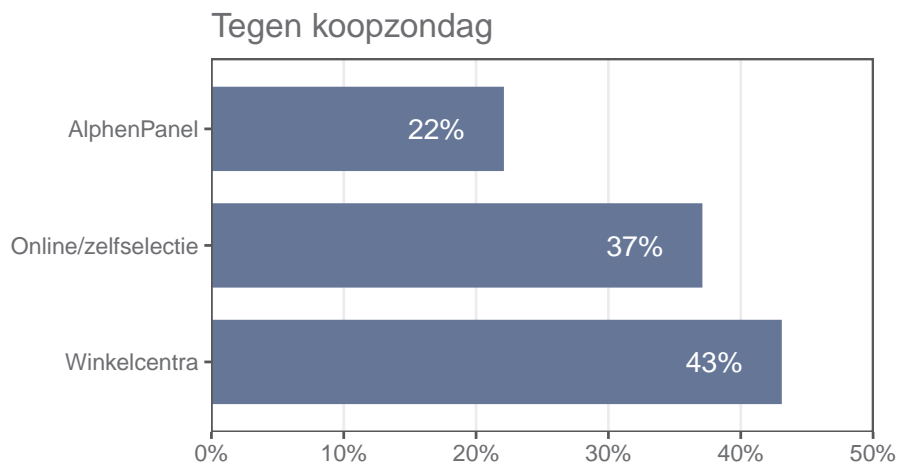
Op 3 maart 2015 maakten de politieke partijen de uitkomsten van de drie enquêtes bekend. Voor de peiling in de winkelcentra werden 754 vragenlijsten ingevuld. De online peiling leverde 1550 ingevulde formulieren op. Voor de burgerpanel-peiling werden 1600 leden van het panel uitgenodigd om de vragenlijst in te vullen. Uiteindelijk deden 857 leden dat. Dit komt neer op een respons van 54%. Dit lijkt een redelijk responspercentage. De respons had hoger kunnen zijn omdat de leden van het panel ermee hadden ingestemd om af en toe een vragenlijst in te vullen. Bovendien was het een aansprekend onderwerp. De gemeente Alphen a/d Rijn had op dat moment 107.000 inwoners. Zo'n 66% woonde in de stedelijke kern Alphen a/d Rijn. De overige 34% woont in de dorpen daaromheen. Bij de burgerpanel-peiling en de online peilingen is geregistreerd waar de respondenten woonden. Daardoor kon je de verdeling van de respons over de kernen in deze peilingen vergelijken met de verdeling in de

doelpopulatie van alle inwoners van de gemeente. Zie hiervoor tabel 1.

Hoe zat de respons in elkaar?

De responsverdeling van de burgerpanel-peiling lijkt veel op de verdeling in de doelpopulatie. Er zitten wat meer respondenten uit de stedelijke kern Alphen a/d Rijn in de peiling dan verwacht zou kunnen worden op grond van de verdeling van de bevolking over de kernen: 70% in plaats van 66%. Voor de overige kernen bedraagt het verschil hooguit 1%. De conclusie kan zijn dat de burgerpanel-peiling in hoge mate representatief is met betrekking tot de verdeling over de kernen. Bij de online peiling is de representativiteit behoorlijk aangetast. Er zijn flinke verschillen tussen de verdeling in de respons en de verdeling in de doelpopulatie. Zo woont slechts 3% van de doelpopulatie in Benthuisen, maar in de online peiling is dat 13%. De Benthuisenaren zijn duidelijk oververtegenwoordigd in de online peiling. Hetzelfde, zij het in wat mindere mate, is te zien bij Boskoop (18% in plaats van 14%) en Hazerswoude-Dorp (8% in plaats van 5%). De oververtegenwoordiging van bepaalde dorpen moet er toe leiden dat een of meer andere dorpen of kernen ondervertegenwoordigd is. Dat is in de online peiling vooral de stedelijke kern Alphen a/d Rijn. In de peiling is slechts 55% afkomstig uit Alphen a/d Rijn, terwijl het 66% had moeten zijn.

Waarom zijn de dorpen Benthuisen, Hazerswoude-Dorp en Boskoop oververtegenwoordigd in de online peiling? Een aannemelijke verklaring is dat in deze dorpen veel aanhangers van de Christelijke partijen wonen. Benthuisen is een dorp waarin de SGP en ChristenUnie altijd goed scoren.



Figuur 2: De percentages tegenstanders van de koopzondag in de verschillende peilingen

Zo kreeg de combinatie SGP-ChristenUnie bij de gemeenteraadsverkiezingen van 2010 zelfs 46% van de stemmen. Als dan een kerk in Benthuizen haar leden oproept om toch vooral mee te doen aan de online peiling, dan is die oververtegenwoordiging verklaarbaar. Ook in Boskoop en Hazerswoude-Dorp trekken de christelijke partijen (CDA, SGP, ChristenUnie) traditioneel veel kiezers. En ook in Boskoop deed een kerk een oproep om mee te doen met de online peiling.

Voor of tegen koopzondagen?

Doel van de het onderzoek van de politieke partijen was het krijgen van meer inzicht in de mening van de inwoners van de gemeente over koopzondagen. Dus werd de vraag gesteld of men voor of tegen koopzondagen was. Figuur 2 toont de uitkomsten voor de drie manieren van peilen. Merk op dat de uitkomsten van de burgerpanel-peiling en de online peiling zijn gewogen om te corrigeren voor een mogelijk gebrek aan representativiteit.

Het percentage tegenstanders verschilt nogal van peiling tot peiling. Het varieert tussen de 22% en de 43%. De uitkomst van de burgerpanel-peiling zal het dichtst liggen bij het werkelijke percentage in de doelpopulatie. Daarom lijkt 22% een redelijke goede schatting van het percentage tegenstanders. Je moet er wel rekening mee houden dat schattingen op basis van een steekproef altijd een onzekerheidsmarge hebben. Voor de 22% tegenstanders is die marge gelijk aan bijna 3%. Dit betekent dat het exacte percentage tegenstanders in de gehele bevolking met grote waarschijnlijk zal liggen tussen de 19% en 25%.

Bij de online peiling is de schatting van het per-

centage tegenstanders (na weging) aanzienlijk hoger: maar liefst 37%. Dat is dus 15 procentpunten meer dan de 22% van de burgerpanel-peiling. Hiermee zit de online peiling er behoorlijk naast. De voor de hand liggen verklaring van deze afwijkende schatting is een oververtegenwoordiging van personen die vanuit hun geloof tegen de koopzondag zijn.

Bij de peiling in de winkelcentra is de schatting van het aantal tegenstanders van de koopzondag nog groter: 43%. Dat is bijna dubbel zoveel als bij de burgerpanel-peiling. Een verklaring hiervoor valt lastig te geven zonder verder onderzoek. Misschien is het wel zo dat juist mensen die op zaterdag winkelen geen behoefte hebben aan winkelen op zondag. En mensen die graag op zondag willen winkelen, omdat ze op zaterdag niet kunnen, zitten niet in de peiling.

De verschillen tussen de drie peilingen zijn te groot om aan toevallige afwijkingen in de steekproef te kunnen toeschrijven. Er is hier sprake van grote systematische verschillen. De conclusie moet zijn dat de uitkomsten van de online peiling en de peiling in de winkelcentra verkeerde uitkomsten opleveren.

Het lijkt erop dat de peiling in de winkelcentra de grootste afwijkingen had. Dat die peiling het zo slecht deed, ligt niet aan het feit dat het een mondelinge (face-to-face) peiling was. Dat de peiling in de winkelcentra het zo slecht deed had vooral te maken met het selectiemechanisme van de steekproef. Dat leverde een steekproef op die verre van representatief was.

Conclusie

Het onderzoek naar de koopzondagen in Alphen a/d Rijn bood de gelegenheid om drie verschillende

manieren van peilen met elkaar te vergelijken. De verschillen bleken behoorlijk groot te zijn. Een burgerpanel is goed te gebruiken voor een peiling. Dan moet dit panel zijn gevuld door het trekken van een aselechte steekproef van inwoners die is geloot uit de hele bevolking. Elke inwoner van de gemeente moet dezelfde kans hebben gehad om in de steekproef te komen.

Een online peiling op basis van zelfselectie is een slecht meetinstrument. De representativiteit kan op allerlei manieren zijn aangetast. Zo kunnen ook personen van buiten de doelgroep de vragenlijst invullen, en daarmee de uitkomsten 'vervuilen'. Verder komt het vaak voor dat een vragenlijst meer dan één keer kan worden ingevuld. Ook is het mogelijk de uitkomsten te manipuleren door aandrang op bepaalde groepen om toch maar de vragenlijst in te vullen.

Ook een face-to-face peiling op straat is een slechte peiling. Immers, je kunt alleen mensen interviewen die op dat moment op straat zijn. Iedereen die, om wat voor reden dan ook, niet op straat is, blijft buiten het onderzoek. Daarmee is de doelgroep van het onderzoek in feite ingeperkt tot de mensen op straat op dat tijdstip. En die groep hoeft niet representatief te zijn voor alle inwoners.

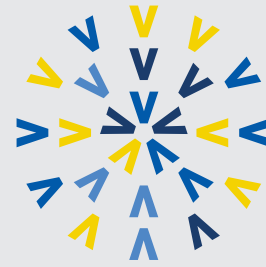
Voor wat betreft de koopzondagen was het een lastige situatie in Alphen a/d Rijn: de stedelijke kern was voor en de dorpen er omheen waren tegen. Welke besluit ook werd genomen, het zou altijd teleurstellend zijn voor een deel van de inwoners. Uiteindelijk besloot de gemeenteraad in maart 2015 met een grote meerderheid (28 tegen 10 stemmen) om koopzondagen ongelimiteerd toe te laten. Wel werd een motie van het CDA en de lokale partij Nieuw Elan aanvaard waarin werd vastgelegd dat de winkels op zondag pas om 13.00 uur open mochten.

Er is een uitgebreide analyse gedaan van de drie peilingen in Alphen a/d Rijn. Het verslag hiervan is terug te vinden in (Bethlehem, 2015).

Bronnen

Bethlehem, J.G. (2015), Over goed en slecht peilen: de casus van de koopzondag. Bestuurswetenschappen 69-4, blz. 23-39.

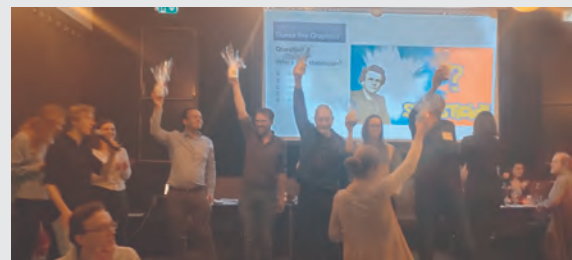
Jelke Bethlehem is expert op het gebied van steekproeven, vragenlijsten en weergave van onderzoeksresultaten.
e-mail: mail@jelkebethlehem.nl



Young statisticians

The Young Statisticians had a lot of fun organizing a pub quiz during the dinner of the VVSOR Annual Meeting on the 23rd of March in Utrecht. After the conference on the topic "Statistics and operations research for robust decision making" the attendees had the option to have dinner together at the Restaurant de Rechtbank. In between the courses of delicious food and conversations about the talks we listened to during the conference, we tested the annual meeting attendees on questions about statistics theory, funny graphs and the VVSOR.

Congratulations to the winning team, who brought home our signature Young Statisticians prize. Want to know what it is? Join our next pub quiz to find out!



Next event

To keep informed on our future events, sign up for our newsletter, follow us on LinkedIn or Instagram or keep an eye on our website: wsor.nl/young-statisticians.



Optimaliseren van de inpak-strategie voor bezorgbussen

Louise J. Zwep

Wanneer heb jij voor het laatst een pakket bij je thuis te laten bezorgen? Vast niet zo lang geleden, gezien de toenemende populariteit van e-commerce in de afgelopen decennia. Nu steeds meer mensen hun toevlucht nemen tot online winkelen, groeit de behoefte om de efficiëntie van pakketbezorging te verbeteren. Een cruciale factor daarbij is het optimaliseren van de inpak-strategie voor bezorgbussen. De manier waarop pakketten worden ingepakt heeft een directe invloed op de totale levertijd, omdat deze bepaalt hoe gemakkelijk elk pakket bij aankomst kan worden gepakt. De optimalisatie van de inpak-strategie leidt tot een minimalisatie van het aantal keren dat de bus moet worden herpakt.

In mijn thesis (Zwep, 2023) was het doel een methode te ontwikkelen voor het creëren van een verpakingsoplossing die zich houdt aan het Last In First Out (LIFO) principe. Dit zorgt ervoor dat een pakket

door de bezorger kan worden bereikt zodra deze de bestemming van het pakket bereikt. Deze specifieke inpakmethode wordt het 3D-Bin Packing Problem with Loading Constraints (3L-BPP) genoemd. Om dit probleem aan te pakken werd een formulering van een Mixed Integer Linear Programming (MILP) model opgesteld. Daarnaast werd een heuristische oplossingsmethode bedacht om snel een initiële oplossing te vinden.

Mixed Integer Linear Program (MILP)

Het MILP is gebaseerd op het 3D-Bin Packing Problem (3D-BPP), waarvoor al een MILP-model bestond. Er waren echter enkele problemen met dit model, zoals:

1. De pakketten konden zweven
2. Er was een gebrek aan zekerheid over de stabiliteit van de pakketten
3. De oplossingen voldeden niet aan het LIFO-principe.

Deze problemen werden een voor een aangepakt door voor elk probleem specifieke constraints

in te voeren, wat resulteerde in een model met in totaal 49 regels aan constraints. Ik zal niet ingaan op al deze constraints, maar ik zal ter illustratie laten zien hoe het eerste punt, de vliegende pakketjes, werd aangepakt. Ten eerste werden de vloer en de planken van de bezorgbus gedefinieerd als pakketten, zodat elk pakket direct aan een ander pakket kan worden gekoppeld langs de z-as.

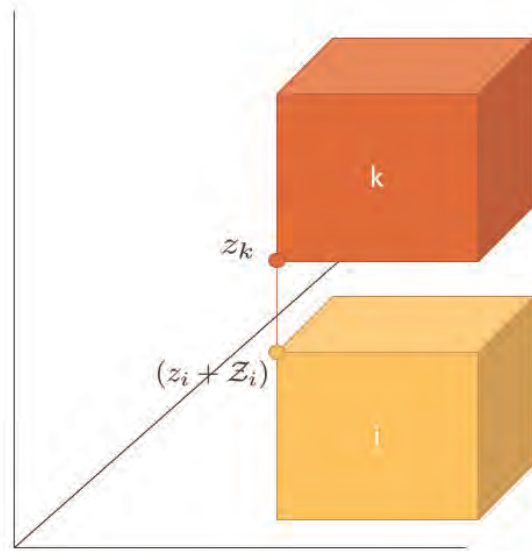
Zeg dat P de verzameling pakketten is die in het busje moet worden geplaatst. Variabele z_i geeft de z-coördinaat voor pakket i , en Z_i is de lengte van dit pakket in de richting van de z-as. De binaire variabele e_{ik} wordt op 1 gezet wanneer pakket i onder pakket k staat. Voor deze specifieke constraints worden binaire variabelen n_{ik} ingevoerd. Deze variabelen worden op 0 gezet als pakket k direct bovenop pakket i ligt, en op 1 in alle andere gevallen. Daarnaast worden de parameters M en BM , waarbij $BM > M$, ingevoerd voor modelleringsdoeleinden. Vervolgens worden de volgende constraints toegevoegd:

$$z_k - (z_i + Z_i) + (1 - e_{ik})M \leq n_{ik}BM \quad \forall i, k \in I : i \neq k \quad (1)$$

$$\sum_{i \in P: i \neq k} n_{ik} \leq |P| - 1 \quad \forall k \in P \quad (2)$$

$$n_{ik} \in \{0, 1\} \quad \forall i, k \in I \quad (3)$$

De constructie van de constraints is zodanig dat n_{ik} ongelijk aan 1 moet worden gekozen voor ten minste één pakket i voor elk pakket k . Dit wordt gemodelleerd door de constraints (2). Het is te zien dat de constraints (1) vereisen dat als twee pakketten i en k niet op elkaar worden geplaatst, n_{ik} gelijk moet zijn aan 1. Dit komt omdat als $e_{ik} = 0$, de rechterkant van de ongelijkheid groter moet zijn dan M , en de enige manier om dit te bereiken is door n_{ik} op 1 te stellen. Als $e_{ik} = 1$, is dit deel van de ongelijkheid gelijk aan 0. Het eerste deel van de ongelijkheid staat voor de onderkant van het bovenste pakket - de bovenkant van het onderste pakket, die minstens eenmaal gelijk moet zijn aan 0, wil de rechterkant ook nul zijn. Deze afstand is ook weergegeven in figuur 1. Door deze constructie worden de pakketten gedwongen direct op minstens één ander pakket te staan, en kunnen ze dus niet in de lucht zweven. Merk op dat planken wel mogen zweven (omdat ze aan de muur vastzitten), en dat de constraints (2) dus niet geldt voor planken.

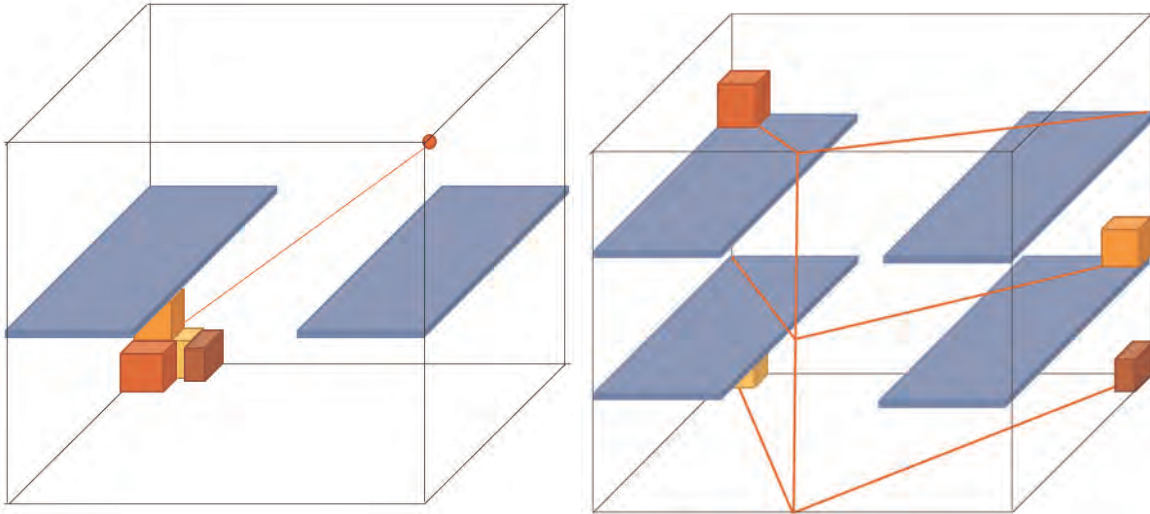


Figuur 1: Afstand tussen z_k en $z_i + Z_i$ van pakketten i en k in een 3D-ruimte

Heuristische oplossingsmethode

Het MILP leverde een analytisch correcte oplossing op, maar die was tijdrovend. Daarom werd een heuristische oplossingsmethode ingevoerd om snel een (weliswaar niet noodzakelijk perfecte) oplossing te vinden. Deze methode was gebaseerd op de "afstand tot de hoek vooraan-boven-rechts" (Distance to the Top-Right-Corner (DTRC) in het Engels), die de verste afstand meet tussen alle mogelijke plaatsingsposities van het volgende pakket en de, zoals de naam al doet vermoeden, hoek vooraan-rechts van de bestelwagen. Dit garandeert een compacte indeling van alle pakketten in de bezorgbus.

Aan deze heuristische methode is een aanpassing gedaan om haar meer geschikt te maken voor een bezorgbus. In plaats van te meten vanaf de bovenhoek, wordt nu gemeten vanaf de middenlijn van de bezorgbus, die kan worden gezien als de achterdeur van de bus. Door de pakketten in omgekeerde aflevervolgorde te rangschikken, worden ze gestapeld op een manier die het LIFO-beleid volgt. De verandering in de afstandsmeting wordt ook getoond in figuur 2.



Figuur 2: Verandering van de heuristiek van het meten van de langste afstand tot (L, W, H) naar het meten van de langste afstand tot de lijn $(\frac{L}{2}, W, z)$ met $z \in [0, H]$

Resultaten

Om de werkzaamheid van de MILP en de bijbehorende heuristiek te evalueren is een dataset van 789 verschillende ritten geanalyseerd. Deze dataset, beschikbaar gesteld door PostNL, bevat historische data van gereden ritten, met daarbij de afmetingen van alle pakketten en de aflevervolgorde. De resultaten toonden zeer goede prestaties van de heuristiek, want deze genereerde een initiële oplossing voor alle ritten in de dataset, waarbij 98,9% van de oplossingen binnen 3 seconden werd gevonden. Voor de resterende 1,1% van de gevallen leidde de integratie van een genetisch algoritme (GA) tot een oplossing binnen 90 seconden.

De initiële oplossing die door de heuristiek werd gegenereerd, werd verder verbeterd door optimalisatie technieken van de MILP. Dit resulteerde in de meest gunstige resultaten wat betreft het minimaliseren van het aantal verkeerd geplaatste pakketten. Om de toepasbaarheid van het model in de praktijk te testen, werd een test uitgevoerd op een van de pakketbezorgingsdepots van PostNL. Een door het model gegenereerde oplossing werd geïmplementeerd voor één rit, waarbij de bestelwagen werd gecompartmenteerd en elk pakket werd toegewezen aan een specifiek compartiment met behulp van handscanners en barcodes, zoals te zien is in figuur 3. Deze praktische toepassing toonde het potentieel van toegepaste wiskunde in real-world situaties!



Figuur 3: Foto's van de proef waarin het model werd uitgetest in een hub voor pakketbezorg. De door het model berekende locatie van de pakketten werd via een scanhandschoen aangegeven.

Literatuur

L. J. Zwep. *Optimizing the packing strategy for parcel delivery vans*. 2023. URL: <https://repository.tudelft.nl/islandora/object/uuid:f127e334-ae93-44dc-93ba-301688a62427?collection=education>.

Louise Zwep behaalde in mei 2023 haar master in Applied Mathematics aan de TU Delft. Haar interesse ligt bij operations research, waar zij in haar bachelor ook haar thesis over optimalisatie van ambulance standplekken heeft gedaan.



Rare jongens die Britten!

Een goed verhaal heeft een 'running gag', een steeds terugkerende grap. Zo ook de strips van Asterix, waar in vrijwel ieder deel Obelix wel een keer opmerkt dat het 'rare jongens' zijn, en dat zijn dan Romeinen maar ook Gothen, Belgen, Noormannen of Britten.

Wat de Britten betreft ben ik geneigd Obelix gelijk te geven. We weten allemaal dat ze daar afwijkende maten hebben voor zo ongeveer alles. En dan niet alleen zaken als miles, shillings en gallons. Zo is er een Guinea, een hoeveelheid geld die eigenlijk alleen gebruikt wordt voor de tarieven van artsen en advocaten, of de Stone die slaat op lichaamsgewicht. Maar ik ben ook ooit op ons vakgebied iets bijzonders tegengekomen, zoiets als links rijden in de kansverdeling!

Omstreeks 1983 werd het mogelijk te communiceren tussen de computersystemen van de universiteiten. Een rudimentaire vorm van e-mail was al langer mogelijk binnen hetzelfde systeem, maar nu

ging het verder. IBM financierde het EARN-project, het European Academic Research Network, waar in principe alle universiteiten aan konden deelnemen. Ook was het gekoppeld aan vergelijkbare netwerken in o.a. de VS en Canada. Ik was hoofd van de groep Toepassingen bij het rekencentrum van de Landbouwhogeschool en moedigde mijn mensen aan vrijuit te experimenteren met deze nieuwe technieken. Dat heb ik geweten, de creativiteit vierde hoogtij. Zo vroeg een van mijn programmeurs gelijktijdig de interne klok op van alle Nederlandse universitaire mainframes om te constateren dat er geen dezelfde tijd aangaf. En mijn beste kracht zocht alle Canadese universiteiten af op zoek naar Nederlands klinkende gebruikersnamen. Met een van de gevonden personen kreeg hij een intensieve mailwisseling, drie maand later ging hij op bezoek en binnen het jaar vertrok hij definitief, hij is nog steeds gelukkig getrouwd.

Door die communicatie tussen universiteiten

ontstonden ook honderden informele netwerken van mensen met een gedeelde interesse, een verre voorloper van chatgroepen. Als je een vraag of opmerking had stuurde je die naar het adres van die groep en die werd dan automatisch gemaild naar alle deelnemers. De onderwerpen waren legio, van genealogie tot beurskoersen, StarTrek of TeX.

Een van die groepen bestond uit SPSS-gebruikers en het duurde niet lang of ik werd gezien als een van de autoriteiten op dit gebied. Op een dag kwam er een vraag in de groep van iemand uit het VK die een parametervrije toets had gebruikt en wilde weten hoe hij de gerapporteerde p-waarde moest interpreteren. Ik had wel even tijd voor een antwoord en begon een uitleg te typen, maar halverwege merkte ik dat mijn uitleg niet klopte. Uit de gegevens bleek duidelijk een hoge samenhang, maar de p-waarde was juist laag. Ik heb die gegevens overgebracht op mijn Wageningse computer en dezelfde toets gedraaid. Nu klopte het wel! We zijn bijna twee weken aan het zoeken geweest en uiteindelijk ontdekte ik dat de afwijking zat in de Error-functie van de En-

gelse computer. In Fortran kan men allerlei functies aanroepen die men niet zelf hoeft te programmeren. Die functies kunnen computer-afhankelijk zijn en zitten daarom niet in de kern van het programma maar worden tijdens de compilatie 'gelinked'. En het functiepakket van die specifieke Engelse computer gaf niet p maar 1-p als uitkomst. Zogezegd links rijden in de statistiek

Hoe kwam het nu dat dit niet was gezien bij het testen van de conversie naar dit systeem? Heel simpel, de betreffende toets werd tot $N=40$ exact berekend, slechts daarboven werd de benadering met de normaalverdeling gebruikt. De door SPSS bijgeleverde testset voor dit probleem had dacht ik 35 waarnemingen, werd dus exact berekend en gaf de juiste uitkomst. Niemand had ooit een test gedraaid met meer dan 40 waarnemingen en dus bleef het statistisch links rijden op dit computersysteem onopgemerkt. Om nogmaals Obelix te citeren: "rare jongens die Britten".

Gerrit Stemerding is eindredacteur van STATOR.
e-mail: gjstemerding@hotmail.com



Save the date: Annual Meeting 21 maart 2024

De Annual Meeting (AM) van 2024 zal weer plaatsvinden bij In de Driehoek in Utrecht. De voertaal zal Engels zijn. Meer informatie over het programma en over het bijwonen van de Annual Meeting volgt in het volgende nummer van STATOR en op de VVSOR-website.

Wanted: new member of the Annual Meeting Committee



Als je wilt helpen bij de organisatie van de Annual Meeting, of als je een goed idee hebt voor het thema van dit jaar, stuur ons dan een e-mail.

annualmeeting@vvsor.nl



Nog vijf maanden te leven

Sanne Willems

Op basis van gegevens van eerdere patiënten schatten statistici de levensduur voor nieuwe patiënten. Hoe doen we dat en wat is het resultaat van zo'n berekening?

Onderstaande blogpost over overlevingsduur is overgenomen van blog.vvsor.nl met Sanne's toestemming.

Hoe lang heb ik nog? Het is een vraag die onherroepelijk op tafel komt tijdens een slechtnieuwsgesprek waarin een patiënt te horen krijgt dat zij een ernstige ziekte heeft. De arts kan dan vaak een inschatting geven op basis van bijvoorbeeld ervaring. Maar ook statistiek speelt hier een belangrijke rol. Met behulp van overlevingsanalyse op data van eerdere patiënten worden schattingen steeds beter. Hoe moet je zo'n getal interpreteren? In deze post neem ik je mee in de totstandkoming van een sinister getal.

Gecensureerde data

De eerste stap om overlevingsduur te schatten is informatie verzamelen over patiënten. Een groep patiënten wordt bijvoorbeeld geobserveerd vanaf hun diagnose en hun gegevens, zoals de datum van de diagnose en de sterfdag, worden genoteerd. Het

lijkt misschien makkelijk om met die diagnose- en sterfdata een gemiddelde levensduur vanaf diagnose te berekenen, maar in werkelijkheid is de analyse een stuk gecompliceerder. Namelijk, niet van elke patiënt wordt de sterfdag geobserveerd.

Stel bijvoorbeeld dat een aantal patiënten uit de groep verhuist naar een ander ziekenhuis, dan is niet bekend hoe het na de verhuizing met hen gaat. Het kan ook voorkomen dat een aantal patiënten zich nog goed voelt wanneer een studie wordt stopgezet. Dan blijft het de vraag hoelang zij nog leven na het stoppen van de studie. In beide gevallen blijft de exacte sterfdag onbekend en kan daardoor dus niet worden meegenomen in de berekening van de gemiddelde levensduur.

Deze patiënten simpelweg weglaten uit de berekening is zonde, want er is wel iets cruciaals bekend over de overleving van deze patiënten; zij leefden namelijk nog minimaal tot het moment dat ze verhuisden of de studie stopte. Dit is nuttige informatie, want het negeren van deze patiënten zou een onderschatting van de levensduur geven.

Dit soort data waarbij alleen minimale overleving bekend is in plaats van de exacte sterfdag noemen we gecensureerde data. Deze data bevatten nuttige informatie die belangrijk zijn om een goede schatting te maken, maar ze maken de statistische

analyse wel een stuk lastiger. Simpelweg een gemiddelde nemen over de levensduur sinds diagnose kan niet meer.

Overlevingsanalyse

Voor een correcte analyse is een aparte tak binnen de statistiek ontstaan: de overlevingsanalyse. Deze techniek wordt vooral veel toegepast in de medische statistiek om de levensduur, of vrolijker, de tijd tot genezing, te berekenen. Maar in principe kan het gebruikt worden voor allerlei tijdsschattingen. Denk bijvoorbeeld aan de levensduur van een apparaat of batterij, hoelang het duurt voordat een klant zijn lidmaatschap opzegt, hoelang een klant aan de telefoon moet wachten op een klantenservicemedewerker, hoelang een busreis naar Parijs duurt, etc.

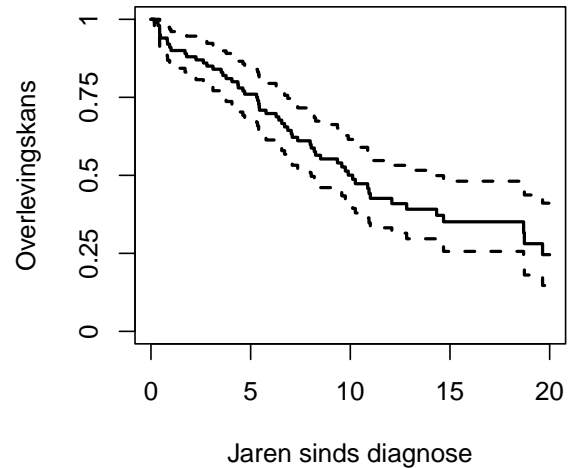
Door de gecensureerde data is het berekenen van een gemiddelde levensduur niet mogelijk. Daarom wordt in de overlevingsanalyse niet het gemiddelde berekend, maar de kans dat je nog leeft tot en met een bepaald tijdstip. Voor meerdere tijdstippen wordt bekeken hoeveel patiënten er nog onder observatie waren, en hoeveel van hen overleden sinds het vorige tijdstip. Op die manier worden de kansen geschat dat iemand bijvoorbeeld na twee weken, een maand, zes maanden en een jaar nog leeft.

Overlevingscurve

Deze schattingen worden weergegeven in een grafiek, de overlevingscurve. In figuur 1 zie je een voorbeeld met verzonden data. Op de horizontale as zie je de tijd sinds diagnose en op de verticale as staat de kans dat een patiënt dan nog leeft. We zien bijvoorbeeld dat de overlevingskans na vijf jaar ongeveer 0,75 is, ofwel 75 procent.

Een overlevingscurve begint altijd bij één, want op het moment van diagnose leven alle patiënten nog. Daarna daalt de grafiek, omdat er door de tijd heen helaas steeds meer patiënten overlijden.

De kansen worden berekend op basis van een steekproef, en daarom is er altijd wat onzekerheid in de schattingen. De twee stippellijnen in figuur 1 geven die onzekerheid aan. De echte curve ligt waarschijnlijk tussen de twee stippellijnen, en hoogstwaarschijnlijk dicht bij de middelste lijn.



Figuur 1: Voorbeeld van een overlevingscurve (ononderbroken lijn) voor verzonden data met indicatie van de onzekerheid over de curve (stippellijnen)

Mediane overleving

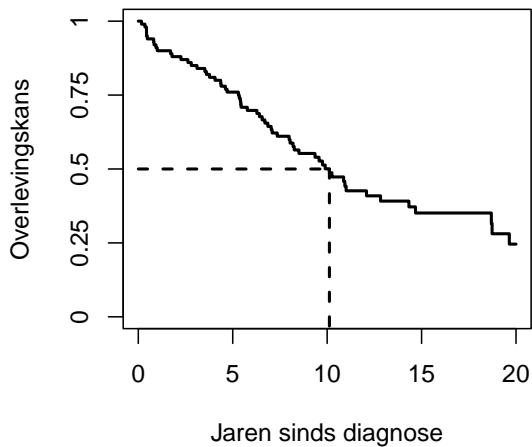
Patiënten krijgen niet altijd de volledige overlevingscurve te zien omdat die lastig te interpreteren zijn. Vaak wordt alleen de kans van één tijdstip gegeven, zoals de kans om na vijf jaar nog te leven.

Een andere optie is de mediane overleving. Dit is het moment waarop naar schatting de helft van de patiënten nog leeft. Ofwel, het moment waarop de overlijdenskans precies 50 procent is.

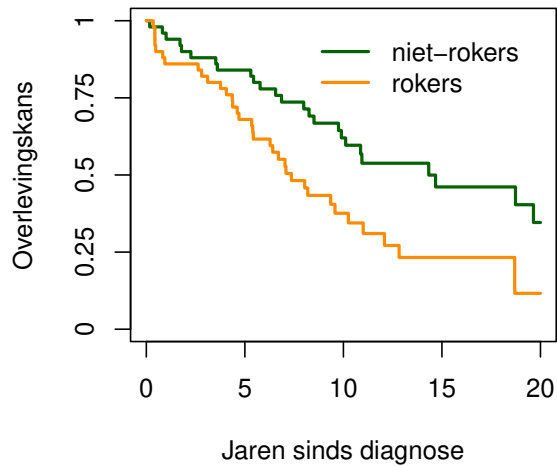
In figuur 2 is de overlevingscurve uit figuur 1 nog een keer weergegeven, nu met een horizontale lijn bij een kans van 0,5. Als je de lijn volgt tot deze de grafiek raakt, dan kun je op die plek de mediane overlevingsduur aflezen op de horizontale as, in dit voorbeeld ongeveer tien jaar. Dat betekent dat 50 procent van de patiënten na tien jaar nog in leven zijn.

Het nadeel van het samenvatten van een hele curve in één enkel getal is natuurlijk dat je veel informatie verliest. Met een mediane overleving heb je wel een goede eerste indruk van de levensduur, maar je weet niets over het verloop van de kansen door de tijd heen. Je weet bijvoorbeeld van de eerste helft patiënten niet of zij vlak na diagnose al sterven of pas na een jaar of acht.

Je hoort wel eens "artsen zeiden: 'u heeft waarschijnlijk nog maar drie maanden te leven,' en kijk, na twee jaar leef ik nog steeds!". Dat een voorspelling ernaast zit komt vaak door de grote verschillen tussen overleving van patiënten. In figuur 2 zien we dat ook: de mediane overleving is tien jaar, maar 25



Figuur 2: Voorbeeld van een overlevingscurve (ononderbroken lijn) voor verzonnen data met de berekening van de mediane overleving



Figuur 3: Overlevingscurve per subgroep patiënten (rokers en niet-rokers) voor verzonnen data

procent van de patiënten leeft zelfs nog na twintig jaar.

Soms wordt een overlevingskans na verloop van tijd aangepast. Als een patiënt zich bijvoorbeeld bovengemiddeld lang goed voelt, dan is dit reden om de verwachte overleving naar boven bij te stellen.

Gepersonaliseerde kansen

Bij de tot nu toe getoonde overlevingscurves zijn de data van alle patiënten op een hoop gegooid. Maar je kunt je voorstellen dat er veel factoren zijn die de overlevingskansen beïnvloeden waardoor deze uiteenlopen, bijvoorbeeld de levensstijl van een patiënt of bepaalde fysieke eigenschappen. Als patiënt wil je natuurlijk graag dat daar rekening mee wordt gehouden bij de schatting; je hebt immers het meest aan een gepersonaliseerde overlevingscurve die helemaal op jou is afgestemd.

Een simpele manier om verschillen tussen patiënten mee te nemen is om een aparte curve te maken per subgroep patiënten. Je kunt bijvoorbeeld een splitsing maken tussen patiënten die wel en niet roken en voor beide groepen een overlevingscurve maken, zoals in figuur 3. Op deze manier kun je de groepen makkelijk vergelijken. We zien in figuur 3 bijvoorbeeld dat de niet-rokers veel hogere overlevingskansen hebben, en dus ook een hogere mediane overleving.

Het vergelijken van groepen op deze manier kan alleen als je maar een paar factoren tegelijk bekijkt. Stel dat bijvoorbeeld ook gezonde en ongezonde eetgewoontes worden meegenomen. Je krijgt dan

vier curves om te vergelijken: voor niet-rokers met gezonde eetgewoontes, voor niet-rokers met ongezonde eetgewoontes, en dezelfde splitsing voor de rokers. Hoe meer groepen, hoe lastiger om alle curves te vergelijken.

Voor veel factoren, en ook voor continue data zoals leeftijd, bedachten statistici andere oplossingen. De extra informatie wordt meegenomen in een model dat alles tegelijk vergelijkt. Uit het model wordt per persoon een overlevingscurve gemaakt, gebaseerd op gegevens van eerdere patiënten die het meest op die persoon lijken.

Het blijven schattingen

Helaas kunnen de modellen nooit een perfecte schatting geven. Dat is altijd afhankelijk van de data die beschikbaar zijn; in hoeverre lijken eerdere patiënten op een nieuwe patiënt en zijn er genoeg om een betrouwbare schatting te maken? Bovendien is het nog de vraag of alle factoren die invloed hebben op overleving wel in het model zijn opgenomen. Is er niks belangrijks over het hoofd gezien?

Al met al geven overlevingsanalyses goede indicaties van de tijd die een patiënt nog rest. Niet meer dan dat. Het blijven schattingen. Op de vraag 'hoe lang heb ik nog?' volgt dan ook een minder precies antwoord dan het tijdens het slechtnieuwsgesprek misschien lijkt.

Sanne Willems is Assistant Professor in Applied Statistics - Methodology & Statistics, Institute of Psychology at Leiden University.

e-mail: s.j.w.willems@fsw.leidenuniv.nl



Statistica Neerlandica in een nieuw jasje

Ernst Wit, Casper Albers

De VVSOR is opgericht in de zomer van 1945, toen nog als Vereniging voor Statistiek. Een van de eerste acties die de vereniging ondernam, begin 1946, was de oprichting van een eigen *peer-reviewed* tijdschrift, Statistica Neerlandica.

Het doel van Statistica Neerlandica was om wetenschappelijke inzichten op het gebied van de statistiek, en dan met name de wiskundige statistiek, biostatistiek en psychometrie te delen. Het was opgezet als een blad van en voor de vereniging, met artikelen en mededelingen over verenigingsactiviteiten in vooral het Nederlands, en sporadisch een bijdrage in het Engels, Frans en Duits.

In de bijna acht decennia dat het tijdschrift bestaat, zijn er de nodige zaken veranderd. Voor verenigingsmededelingen richtte de VVSOR in 1976 de 'Methoden en Data Nieuwsbrief' op, die later overging in Kwantitatieve Methoden en die ging op haar beurt weer over in STAtOR. De lingua franca van de wetenschap is inmiddels enkel nog Engels. We sturen geen artikelen en we zoeken geen reviewers meer per fax of post. Mede dankzij het internet is de wetenschap, en dus ook Statistica Neerlandica, dynamischer en internationaler geworden.

De afgelopen jaren lag Statistica Neerlandica in goede handen van de editors Miroslav Ristic, Marijtje van Duijn en Nan van Geloven. Zij hebben aangegeven toe te zijn aan nieuwe uitdagingen en de vereniging is hen erg dankbaar voor hun inzet. Het bestuur van de vereniging heeft het afgelopen jaar naar nieuwe editors gezocht en in de tussentijd heeft Casper Albers op de winkel gepast. Met ingang van mei 2023 is er een nieuw editorial team voor Statistica Neerlandica: Ernst Wit (Lugano, Zwitserland), Edwin van den Heuvel (Eindhoven) en Veronica Vinciotti (Trento, Italië) nemen het stokje over.

Behalve een aanpassing in zowel de editors als associate editors, vindt er ook een inhoudelijke wijziging plaats. Om beter aan te sluiten bij de moderne, dynamische wetenschap is Statistica Neerlandica vanaf heden een *fastturn-around* journal. Het doel is om de duur tussen het indienen van het artikel en de publicatie online terug te brengen naar maximaal twee maanden. Hierdoor blijven artikelen niet

meer maandenlang achter de schermen liggen tot het peer-review en herhalende revisieproces zijn afgerond, maar worden nieuwe ideeën sneller met de academische gemeenschap gedeeld.

Vanzelfsprekend kan deze versnelling alleen door ook wijzigingen te maken in het publicatieproces. We vragen aan auteurs expliciet om artikelen te submitten waar op inhoudelijk en tekstueel vlak geen grote veranderingen bij nodig zijn. De papers moeten een duidelijk narratief hebben, duidelijk maken wat ze toevoegen aan de literatuur en geen verdere revisie vereisen. Het oordeel "major revision" zal komen te vervallen: nieuwe submissions worden ofwel afgewezen, ofwel geaccepteerd na enkele "minor revisions".

Naast reguliere onderzoeksartikelen, variërend van wiskundig-statistische ontwikkelingen tot interessante statistische en beslistkundige toepassingen, staat Statistica Neerlandica nu ook open voor short notes, waarin nieuwe ideeën gepresenteerd kunnen worden zonder een complete theoretische onderbouwing. Daarnaast zijn bijdragen waarin nieuwe en belangrijke ontwikkelingen op het gebied van statistiek kort en krachtig worden uitgelegd welkom. Dergelijke artikelen zijn vaak een nuttig startpunt voor promovendi en andere (jonge) onderzoekers die een nieuwe methode onder de knie willen krijgen.

Doordat auteurs nu veel sneller weten waar ze aan toe zijn, hopen we dat deze nieuwe werkwijze ook leidt tot een verhoogde kwantiteit en kwaliteit van de artikelen die worden aangeboden. Omdat Statistica Neerlandica namens onze vereniging wordt uitgegeven, hopen we vooral op meer bijdragen door leden van onze mooie vereniging. Voor medewerkers van Nederlandse universiteiten geldt een deal die met Wiley gemaakt is, waardoor gratis open access publiceren automatisch is. We kijken uit naar jullie werk!

Meer informatie

Voor meer informatie, en voor het indienen van een nieuw artikel, kunt u terecht op onlinelibrary.wiley.com/journal/14679574.



Modelmatig overheidsbeleid

Uitdagingen met wiskundige modellen in wet- en regelgeving

De overheid maakt steeds meer gebruik van wiskundige modellen om beleidsbeslissingen te onderbouwen en wetgeving te ontwikkelen wat efficiëntieverbeteringen oplevert. Het gebruik van modellen in wetgeving brengt echter ook uitdagingen en risico's met zich mee met mogelijk nadelige gevolgen voor de betrokken burgers of bedrijven.

Introductie

In toenemende mate maakt de overheid gebruik van wiskundige modellen om beleidsbeslissingen te onderbouwen en wetgeving te ontwikkelen, bijvoorbeeld in fraudebestrijding of bij de ontwikkeling van klimaatbeleid. Modellen kunnen helpen om complexe vraagstukken snel en efficiënt op te lossen en beleidsbeslissingen op een objectieve en rationele manier te nemen. Echter, het gebruik van modellen in wetgeving brengt ook uitdagingen en risico's met zich mee met mogelijk nadelige gevolgen voor de betrokken burgers of bedrijven. Ondanks deze risico's blijft het gebruik van modellen in wetgeving groeien en is de verwachting dat deze trend zich zal voortzetten. Het is daarom van groot belang dat de overheid zich afvraagt of het gebruik van een model überhaupt nodig is om een beleidsbeslissing te nemen en, als dat het geval is, hoe het de betrouwbaarheid en toepasselijkheid van modellen waarborgt.

Betrouwbaar en toepasselijk

Dat de inzet van modellen in het ontwikkelen en uitvoeren van overheidsbeleid voordelen heeft is evident. Modellen stellen beleidsmakers in staat om verschillende scenario's te verkennen en te analyseren hoe een beleidsmaatregel zich in de toekomst kan ontwikkelen. Dit kan helpen bij het identificeren van de meest effectieve beleidsinstrumenten en het optimaliseren van de uitkomsten ervan. Beleidsmakers zijn, zoals vele beslissers in het bedrijfsleven, geneigd om beslissingen te nemen op basis van intuïtie en ervaring, wat kan leiden tot beslissingen die niet altijd gebaseerd zijn op de feiten. Modellen kunnen helpen bij het verminderen van deze subjectiviteit en ervoor zorgen dat beleidsbeslissingen worden genomen op basis van objectieve analyses. Door de gebruikte modellen en data te publiceren kan de transparantie en de verantwoording van beleid worden verbeterd. Dit kan bijdragen aan het vergroten van het vertrouwen van burgers in de overheid en het versterken van de democratische legitimiteit van beleidsbeslissingen.

Er zitten ook risico's aan het gebruik van modellen in beleidsontwikkeling. Een van de grootste zorgen is de betrouwbaarheid van de modellen. Modellen zijn gebaseerd op aannames en data die mogelijk niet volledig, accuraat en waardevrij zijn. Als gevolg hiervan kunnen de uitkomsten van de modellen foutief zijn en beleidsbeslissingen die op basis ervan zijn genomen verkeerd uitpakken. Een bekend voorbeeld hiervan is de toeslagenaffaire, waarbij duizenden ouders onterecht werden beschuldigd van fraude door een foutief voorspelmodel. De betrouwbaarheid van de Corona modellen van het RIVM is ook twijfelachtig aangezien het effect van maatregelen zoals afstand houden en schoolsluitingen als aanname in de modellen zijn gestopt en niet het resultaat waren van de toepassing van de modellen.

Een ander probleem is de toepasselijkheid van de modellen. Kan het model de vraag waarop het antwoord wordt gezocht wel beantwoorden? Een voorbeeld daarvan is het model dat de NAM gebruikt om de schade aan woningen als gevolg van aardbevingen vast te stellen. Daarvoor gebruikt de NAM een model dat de kans berekent dat een aardbeving schade aan een woning toebrengt. Echter, de vraag die beantwoord moet worden is of de waargenomen schade aan een woning veroorzaakt is door een aardbeving. Vergelijkbaar is de inzet van het stikstofdispositie voorspelmodel AERIUS om piekbelastende boerenbedrijven te vinden terwijl uitspraken op

dat detailniveau met geen enkel model mogelijk zijn vanwege de hoge mate van complexiteit en onzekerheid over de processen die tot stikstofdispositie leiden.

Het hele kader is het halve antwoord

Veel van de problemen met de inzet van wiskundige modellen in beleidsontwikkeling en uitvoering zijn te voorkomen. Het bedrijfsleven past al jaren, met succes, wiskundige modellen in hun besluitvorming toe. Daar kan de overheid van leren. Sleutel in kwalitatief goede besluitvorming, al of niet ondersteund met wiskundige modellen, is een goede afbakening van het vraagstuk of doel.

De *afbakening* is de eerste fase van ieder beslustraject. In deze fase worden de situatie en complicaties beschreven en wordt een beeld geschetst van het te bereiken doel (*What does good look like*). Door alle stakeholders te betrekken, inclusief (een vertegenwoordiging van) degene op wie het beleid betrekking heeft of invloed erop kan uitoefenen, wordt draagvlak gecreëerd. Zo ontstaat een gedragen, gemeenschappelijk beeld van het te bereiken doel inclusief zicht op acceptabele oplossingsrichtingen. In deze fase wordt ook duidelijk of überhaupt een wiskundig model nodig of mogelijk is. Immers niet ieder vraagstuk leent zich voor de inzet van wiskundige modellen, in veel gevallen is gezond verstand voldoende om passend beleid te formuleren. Als al besloten wordt een wiskundig model te gaan gebruiken, wordt in deze fase ook duidelijk aan welke randvoorwaarden en condities het model en de modeluitkomsten zullen voldoen (*Meta model condities*) zodat achteraf door iedere stakeholder getoetst kan worden of aan die condities is voldaan.

Wiskundige modellen kunnen van grote waarde zijn in het analyseren en oplossen van complexe vraagstukken en kunnen daarmee een welkom aanvullend instrument zijn in het ontwikkelen en uitvoeren van regeringsbeleid. Belangrijk daarbij is dat de overheid zorgdraagt voor, onder andere, een goed afbakeningsproces waarin met alle stakeholders doel en scope worden vastgesteld, waarbij ook de noodzakelijkheid en wenselijkheid van het gebruik van een wiskundig model wordt getoetst. Alleen op die manier kan modelgebaseerde wetgeving worden ontwikkeld die rechtvaardig, effectief en ethisch verantwoord is.

John Poppelaars is oprichter en CEO van Doing the Math e-mail: john@doingthemath.nl

