

Interview

We leven in een wereld van getallen en cijfers. Maar wat zegt de brij van statistieken die dagelijks over ons wordt uitgestort nu écht? We vroegen een hoogleraar, een lerares en een cabaretier om een rondleiding door het cijferwoud.

Door **Maarten Keulemans** Foto's **Els Zweerink**

1.1 Pas op, cijfers! Lees eerst deze instructies

Cijfers zijn vaak wazig

Daar, achter die boom! Kijk uit, cijfers! 'Probeer rustig te blijven', adviseert lerares wiskunde Hester Vogels. 'Een deel van de bevolking raakt meteen een soort van in paniek als er cijfers bij komen kijken.'

'Mensen hebben ook die kwantificatiedrang', zegt Jan Beuing, wiskundige en cabaretier. 'Als er iemand ziek wordt, is het meteen: wat zijn mijn kansen? We willen getallen zien. Getallen zijn op de een of andere manier harder dan woorden.'

Terwijl ze dat lang niet altijd zijn, benadrukt hoogleraar wiskunde Johan van Leeuwen van de TU Eindhoven. 'Ik zie veel stellige uitspraken over getallen die in feite onzeker zijn. Maar die onzekerheden halen de krant niet. Zo ga je heel snel naar: dit is waar.'

We zitten bijeen aan een tafel vol krantenknipsels, printjes en bierviltjes. Want leuk, zo'n statistiekspecial, maar laten we bij het begin beginnen. Een spoedcursus statistiek. Voor dummies.

Want hoe erg is het eigenlijk als het RIVM bij een steekproef onder 380 omwonenden van een fabriek iemand vindt met een bloedwaarde van 147 ppb, terwijl 89 ppb de norm is? Wat zegt het als de misdaad in Dronten vijfmaal hoger ligt dan in Lelystad, en een in de Waal gevonden lichaam 'zo goed als 100 procent zeker' van een vermiste Eritreeër is?

De eerste waarschuwing is dit: getallen zijn vaak eerder wazige vlekjes dan duidelijke piketpaaltjes op een getallenlijn. Natuurlijk, dat Manchester 2 keer scoorde tegen Ajax en dat er in Nederland 12 provincies zijn, ligt vast. Maar anders wordt het bij afgeleide cijfers.

De professor: 'We hebben steeds meer data, en de krant staat dagelijks vol samenvattende statistiekjes, die die data op een hoop gooien. De charme van statistiek is nu juist dat je daarbij niet uitkomt op een getal, maar het oprekt: hier is de ondergrens en hier de bovengrens, en je kunt er vrij zeker van zijn dat het getal ertussenin ligt.'

En die keer toen *de Volkskrant*, op last van Zwitsers onderzoek, kopte: 'Arafat voor 83 procent zeker vergiftigd'?

De wiskundigen proberen rustig te blijven.

De lerares: 'Zo'n getal klinkt opeens heel feitelijk en precies. Je denkt: er zit een berekening achter.'

De professor: 'Het is té precies. En het klinkt alsof het heel wat is. Terwijl je dit in de kansrekening niet eens 'zeer waarschijnlijk' zou noemen. Een gangbare vuistregel is bijvoorbeeld: pas bij 95 procent zekerheid is iets overtuigend.'

De cabaretier: 'Ik probeer zoiets altijd maar te vertalen. Stel je voor dat je een pistool hebt met zes kamers en één kogel. Je kunt dan wel zeggen: ik heb 83 procent kans om te overleven. Maar de juiste reactie is natuurlijk: ik heb



een kans van eenzesde dat ik dood ben!

Vergeet het gemiddelde, check de spreiding

Ach, het gemiddelde. Dat keurige poldercompromis, het brave midden tussen de cijfers. Wat kan daar nou mis mee zijn?

'Wantrouw het gemiddelde', zegt Jan Beuving, de cabaretier.

'Je moet beseffen dat het gemiddelde sterk wordt beïnvloed door

uitschieters', zegt Hester Vogels, de lerares. 'Als ze bij de dokter zeggen dat de gemiddelde levensverwachting nog twee jaar is, denken mensen gevoelsmatig: ik zal daar wel heel dichtbij zitten. Maar dat hoeft niet zo te zijn. Het kan ook betekenen dat je twee weken hebt, of juist tien jaar.'

'Vandaar dat ik het woord gemiddelde altijd een beetje probeer te vermijden', zegt de cabaretier. 'Ik heb het liever over de verwachting. Bij een verwachting heb je ook het idee dat het níét uit

Jan Beuving, Hester Vogels en Johan van Leeuwarden.

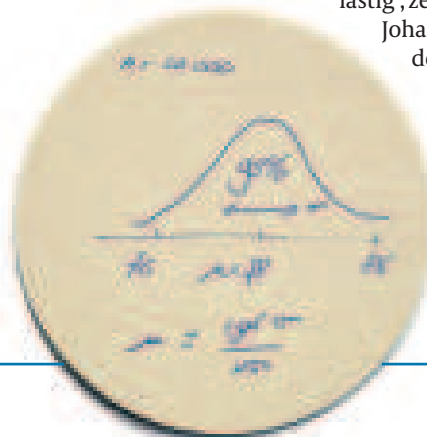
kan komen. Terwijl een gemiddelde je het idee geeft dat dit is wat er hoort te gebeuren.'

'Uitrekenen hoe het zit, is best lastig', zegt de professor,

Johan van Leeuwarden.

'Maar het visualiseren, dat helpt.'

En zowel de professor als de lerares pakken een bierviltje en tekenen een 'normaalverdeling'. ➔



► De lerares: 'Zo'n normaalverdeling zegt je waar de grote bups zit. Die μ hier, dat is het gemiddelde. En σ , dat is de zogeheten standaardafwijking. Binnen één standaardafwijking zit altijd ongeveer 68 procent van je data. En binnen twee standaardafwijkingen ongeveer 95 procent.'

De standaardafwijking: die vind je door te berekenen hoe ver de waarden in een reeks af zitten van het gemiddelde - en dáárvan weer het gemiddelde te nemen. 'Maar met één keer klikken in Excel heb je hem ook', zegt de professor.

Dus toen het Nibud in het nieuws bracht dat ouders hun kind gemiddeld 5 euro voor het rapport geven? Moeilijke blikken aan tafel. De lerares: 'Ligt het nou tussen de 4,90 euro en de 5,10 euro? Of tussen 2 en 8 euro? Dat moet je weten.'

'Dit is wel een ding hoor', zegt de professor. 'Het onderzoek van het Nibud bevat meer interessante feiten die in de nieuwsberichten niet terugkwamen. En het is een kleine moeite om er even bij te zetten: let op, de standaardafwijking is drie euro. Of: het gemiddelde is 5 euro, maar de meeste mensen geven 1 euro. Dat plaatst het in perspectief.'

'Er gaat zó'n sterke werking uit van het woord gemiddelde', vindt de cabaretier. 'Wat blijft er hangen? Die 5 euro. En bij het volgende rapport denk je: o ja, 5 euro geven, dat is kennelijk zoals het hoort.'

Kijk uit met de winnende oliebol

In Ede, dat u maar weet, wonen de gelukkigste mensen van Nederland. Van de inwoners geeft 88,7 procent aan gelukkig te zijn; de nummer twee, Alphen aan den Rijn, blijft steken op 88,5 procent. Aldus de Atlas voor Gemeenten, in een statistiek die vorige maand breed de media haalde.

'Ik moet denken aan de oliebollentest', zegt de cabaretier. 'Daar komt ook altijd een winnaar uit. Mensen willen graag een winnaar, dat is ook helemaal niet erg. Maar stel: de winnende oliebol heeft een gemiddeld jurycijfer van 8,5 en de nummer twee een 8,4. Misschien heeft de winnende oliebol wel een paar tienden en wat zesjes, en de nummer twee vijf keer een

1.2

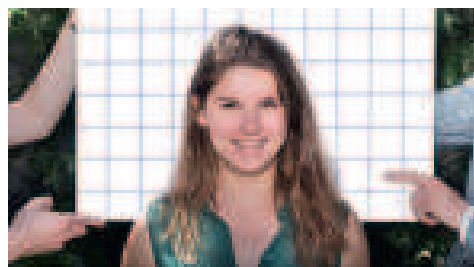
DE PROFESSOR

Johan van Leeuwen (1978) is hoogleraar wiskunde aan de TU Eindhoven en gespecialiseerd in kansrekening, data en sociale netwerken.



DE LERARES

Hester Vogels (1993) is docent wiskunde bij scholengemeenschap UniC voor havo en vwo in Utrecht. Daarnaast heeft ze een populaire site met uitlegvideo's over wiskunde, Wiskunjelers.nl.



DE CABARETIER

Jan Beuving (1982) studeerde wiskunde en is cabaretier, tekstschrijver, sport- en taalcolumnist. Aan zijn theatershows kunnen, zoals hij zelf zegt, 'bèta's hun hart ophalen, en de alfa's hun cijfers.'



8,4. Dan vind ik die 8,4 toch betrouwbaarder.'

De lerares maakt een ruwe berekening op een bierviltje: ze moeten daar

in Ede ruim 100 duizend mensen hebben bevraagd om de gemeente met 95 procent betrouwbaarheid tot winnaar uit te roepen - haast alle inwoners. 'Natuurlijk is het leuk om een winnaar aan te wijzen, maar als je het onderzoek nog een keer zou uitvoeren, was misschien Alphen aan den Rijn bovenaan gekomen. En iedereen onthoudt nu Ede. Ik heb helemaal niet onthouden dat Alphen ook heel hoog zat.'

De cabaretier: 'En kijk eens naar het geheel. Wat mij opvalt is dat de zogenaamd ongelukkigste gemeente Rotterdam ook boven de 80 procent zit. Een van de conclusies is: we zijn in Nederland gewoon hartstikke rete-gelukkig.'

Een pluim rondom getallen, zodat je ziet hoe wijd de onzekerheidsmarge is en welke getallen overlappen; dat mag eigenlijk niet ontbreken, vinden de drie. Zo breng je op het oog harde getallen terug tot kansboogjes, en wordt de nummer één gewoon een van de winnaars: vandaag heeft Ede gewonnen, maar Alphen, Amstelveen, Amersfoort, Gouda, Deventer, Velsen en Apeldoorn hadden net zo goed gekund.

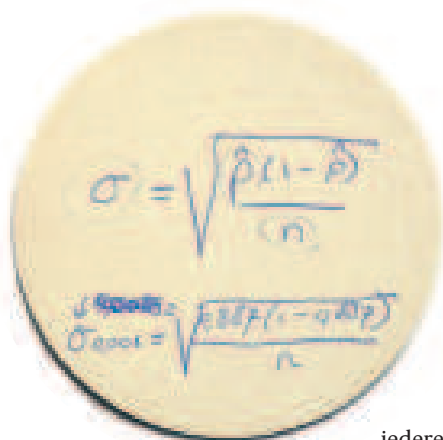
De cabaretier: 'Ik vind dat er een klein disclaimer bij zou moeten. Pas op, dit is statistisch onderzoek, de kans dat het de volgende keer anders uitvalt is vrij groot.'

De lerares: 'Of presenteer het zoals de Peilingwijzer de verkiezingspeilingen presenteert, met een foutmarge eromheen. Zodat je kunt zien dat het toeval erin zit.'

Pas op met stijgingen en dalingen

Ah, het stijgt, het daalt. Het groeit. Het stagneert. Pas op!

Neem het aantal verkeersdoden. In 2015 waren het er 621, het jaar erop 629. Een stijging, voor het tweede opeenvolgende jaar!, constateerden de nieuwsmedia vorige maand unaniem. Waarna experts en lobbyisten er hun lievelingsinterpretatie aan gaven: het kwam on-



getwijfeld door het sms'en, het 130 kilometer rijden, of nee, het elektrisch fietsen.

De cabaretier: 'Ten eerste moet je constateren dat het een toename is. We hebben geteld en er zijn gewoon meer verkeersdoden dan vorig jaar.'

De professor: 'Maar om nou op basis hiervan te zeggen dat er een opwaartse trend is, gaat me te ver. Dat zie je al door gewoon te kijken naar die lijn, puur op het oog.'

De lerares: 'Het zijn maar twee datapunten he? Heel riskant om op basis daarvan meteen een stijgende lijn te trekken.'

De professor: 'Ik vind dit een prachtig voorbeeld van hoe metingen inspiratie kunnen geven. Want hier begint het pas. De volgende stap is dat je op zoek gaat naar de verklarende variabelen. Is er een onderliggende factor die dit patroon kan verklaren? Waardoor kwam die daling? En is het toeval dat het nu lijkt te stagneren?'

De Groningse statisticus Casper Albers rekende het voor zijn column in *Nieuw Archief voor Wiskunde* uit: het flupje omhoog wijkt maar eentiende standaardafwijking af van de trendlijn omlaag. 'Niet significant, niet eens een beetje', aldus Albers.

'Vergeet niet dat het de afgelopen twintig jaar drie keer eerder is voorgekomen dat er even een stijging was van het aantal verkeersdoden', zegt de cabaretier. 'Waarna de daling weer doorging.'

Maak relatief weer absoluut

En wees extra voorzichtig met cijfers waar 'procent' achter staat, waarschuwt de lerares. Een beruchte valkuil: procent van wát eigenlijk? 'Percentages voelen heel prettig. Iedereen weet wat het is. Maar in werkelijkheid is het gewoon hartstikke lastig.'

Ze denkt terug aan die keer toen er trammelant was over het eten van rood en bewerkt vlees: 'Dat zou de kans op kanker met 20 procent verhogen. Ja, maar om wat voor aantallen gaat het écht? Als je van 0,00001 procent risico naar 0,00003 procent risico gaat, is dat ook een toename van 200 procent, terwijl de kans nog steeds... nou ja, nihil is.'

Bij het vlees eten kwam het erop neer dat van

1.3

KRIJG EEN AFWIJKING

De onzekerheid is met een simpele formule snel te berekenen. Stel, in een gemeente zegen 82 van 100 onder-vraagde mensen gelukkig te zijn. Met een 95%-betrouwbaarheidsinterval geldt dan:

$$p \pm 1,96 \times \sqrt{(p(1-p))/n}$$

Waarbij p de kans is en n de steekproefgrootte. In dit geval dus:

$$82\% \pm 1,96 \times \sqrt{(82\%(1-82\%)/100)} \\ = (74,5\%; 89,5\%)$$

In woorden: bij een andere steekproef zal de uitkomst 19 van de 20 keer ergens liggen tussen de 74,5 procent en de 89,5 procent gelukkige mensen.

iedere 1.000 mensen die weinig bewerkt vlees eten er 56 ooit darmkanker krijgen, tegenover 66 onder de veel-etters. Van 56 naar 66, dat is inderdaad 18 procent méér. Maar in absolute aantallen zijn het ook 'slechts' 10 extra gevallen op iedere 1.000 mensen – 1 procent kans dat je er als bewerkt-vleeseter mee te maken krijgt.

'In het *NOS Journal* werd het op een gegeven moment zelfs: als vleeseter heb je 20 keer zoveel kans op kanker', herinnert lerares Vogels zich. 'Je moet als lezer of kijker kritisch opletten. Als er statistieken worden gepresenteerd, maken de media gewoon veel fouten.'

De cabaretier: 'Je moet ook niet te bang zijn. Kijk even reëel wat er aan de hand is. Je kunt op gezond verstand meestal een heel end komen.'

De professor: 'Je moet jezelf denk ik dwingen het te vertalen in woorden. Het is ook gewoon interessant om er meer over te weten.'

Ah, zonlicht. Daar is de uitgang uit het cijferwoud, met al zijn hobbels en valkuilen, zijn data-moerassen, kansbomen en mistige statistieken. We nemen afscheid, een tafel vol volgekrabbelde viltjes blijft achter.

Maar... wat staat daar nou? Een van de wiskundigen moet het hebben opgeschreven, toen er even niemand keek, achterop een viltje van het biermerk Liefmans.

'Statistiek is ook léúk', staat daar in speelse, blauwe hanepoten. ●

