

Casper Albers

Psychometrie & Statistiek
Rijksuniversiteit Groningen
c.j.albers@rug.nl



Column Casper grijpt een kans

Lies, damn lies en tijdreeksen

Caspers Albers schrijft op regelmatige basis in dit blad een column over alledaagse statistische onderwerpen.

Stel dat je lobbyt voor een of andere maatregel. Om je argumenten kracht bij te zetten, wil je laten zien dat niet alleen jij dit een goed idee vindt, maar dat het volk dit ook wil. Je besluit wat mensen op straat te ondervragen, in een ochtend-, middag- en avondsessie. Elke keer ondervraag je 100 personen. 's Ochtends sta je bij een basisschool en krijg je steun van 62 van de 100 personen. 's Middags en 's avonds, op andere locaties in de stad, heb je minder succes: beide keren zijn maar 38 mensen het met je eens. In het persbericht van het lobbybedrijf schrijf je vervolgens "Bijna tweederde van de personen die we op maandagmorgen ondervraagd hebben, steunt ons voorstel. Nieuwe regering: luister naar het volk en pak die kans!"

Vanzelfsprekend is dit valsspelen. Het is een schoolvoorbeeld van de uitspraak 'Lies, damn lies and statistics' (die doorgaans onterecht wordt toegekend aan de Britse oud-premier Benjamin Disraeli [2]). Vanzelfsprekend had de uitspraak moeten zijn dat $m = 138$ van de $n = 300$ ondervraagden, dus $\hat{p} = m/n = 46\%$, het voorstel steunden. Bij voorkeur vul je dit nog aan met een uitspraak die de onzekerheid kwantificeert. Dit kan bijvoorbeeld met een 95% betrouwbaarheidsinterval via de normale benadering van de binomiale verdeling,

$$\hat{p} \pm 1,96 \times \sqrt{\frac{1}{n} \hat{p}(1 - \hat{p})} = (40,3\%; 51,6\%)$$

(waarbij 1,96 het bijbehorende percentiel van de standaardnormale verdeling is). Bovenstaande is niet de beste methode om zo'n interval te berekenen, maar wel een makkelijke. Het kan net iets beter via de *plus-four*-methode [1] waarbij $\hat{p} = m/n$ vervangen wordt door $\hat{p} = (m + 1,96) / (n + 2 \times 1,96)$, dat levert (40,4%; 51,7%).

Zo zijn er nog een aantal methoden om een puntschatter van een proportie te voorzien van een (betrouwbaarheids)interval. Het is makkelijk om je te richten op de wiskundige aspecten van al deze methoden. Ongeacht de wiskundige details geven alle aanpakken ongeveer hetzelfde beeld: vier à vijf van de tien ondervraagden zijn het met de stelling eens. Door alleen naar de wiskundige kant van de statistiek te kijken mis je echter een belangrijk aspect: de grote verschillen in de ochtend-, middag- en avondmeting moeten wel een alarmbelletje laten rinkelen. Misschien is de populatie waaruit gepeild werd niet zo homogeen: de ouders op het schoolplein staan blijkbaar veel positiever tegenover het voorstel dan andere personen. Waarom is dit? Wellicht gaat de maatregel over extra investeringen in de kwaliteit van basisscholen: de ouders van de kinderen op die school hebben daar natuurlijk direct belang bij en zullen dat eerder de kosten waard vinden dan een willekeurige andere Nederlander. Een goed statistisch model houdt met dit soort aspecten rekening.

Met gewoon het rapporteren van de 62% steun van de ochtendploeg kom je tegenwoordig niet meer makkelijk weg. Wetenschappers, journalisten, politici en andere belanghebbenden zijn tegenwoordig (redelijk) goed getraind in het herkennen van dit soort valsspelen (al kan het altijd beter). Er is echter één soort data waarbij steevast — bewust of onbewust — verkeerde conclusies getrokken worden: longitudinale of tijdreeksdata.

Tijdreeksdata

Tijdreeksen bestaan uit metingen die elkaar opvolgen in de tijd. Het kan de dagelijkse middagtemperatuur in De Bilt zijn, de hartslag die je hartslagmeter meet tijdens het hardlopen, of welke andere reeks gegevens ook. Veel tijdreeksen zijn fundamenteel anders dan gewone steekproeven, in de zin dat ze nooit 'klaar' zijn. Bij een gewone steekproef bepaal je doorgaans vooraf hoeveel

mensen je wilt ondervragen, en dan doe je dat en rapporteer je de resultaten. Een tijdreeks kan eindig zijn — bijvoorbeeld wanneer je wilt weten wanneer op de dag voor kerstmis de meeste omzettingen in de supermarkt gedraaid worden: zodra die dag voorbij is, is je meting voorbij. Vaak houden tijdreeksen echter nooit op. Denk aan de peilingen van Maurice de Hond. Elke week komt er weer een nieuw antwoord op de vraag “Waar zou u op stemmen als er vandaag verkiezingen zouden zijn geweest?”, ondanks dat we weten dat er vandaag helemaal geen verkiezingen zijn. Elke keer komt er weer een piepklein stukje extra informatie. Bij een gewone steekproef zou geen krant er ook maar over denken om een nieuwsbericht te wijten aan hoe de resultaten gewijzigd zijn omdat de steekproefgrootte van $n = 1292$ naar $n = 1293$ is gegaan. Maar bij de politieke peilingen van De Hond is het wekelijks raak: ook als alle verschuivingen binnen de foutenmarge zijn, worden deze uitgebreid uitgemeten in de media en komen de experts om te duiden waar de verschuivingen door komen.

Een veelgemaakte denkfout bij tijdreeksdata is dat men inzoomt op een klein deel van de reeks metingen en daar conclusies aan verbindt zonder het grotere plaatje te bekijken. Bij de afgelopen Tweede Kamerverkiezingen behaalde het CDA 19 zetels, een stijging van 6 ten opzichte van de verkiezingen van 2012. Dit werd door CDA'ers uitbundig als iets moois gevierd. Het is echter ook een daling van 2 zetels ten opzichte van de uitslag van 2010, of een gigantische aderlating ten opzichte van de uitslagen van enkele decennia geleden.

In een tijdreeks zit zo nu en dan een opvallende patroon — een daling of stijging of juist een periode van stabiliteit — en dat kan gewoon toeval zijn. Het kan natuurlijk ook een ‘echte’ oorzaak hebben. Of dit zo is, zie je niet als je alleen inzoomt op die opvallende stijging en daling. Stel dat ik vandaag elke minuut met een dobbelsteen gooi. Uiteindelijk heb ik dan een tijdreeks van 1440 worpen. Wellicht gooi ik wel ergens vandaag vijf keer op rij een zes. Als je daarop inzoomt, klinkt dat extreem onwaarschijnlijk: de kans om vijf keer op rij een zes te gooien is erg klein ($(\frac{1}{6})^5 = \frac{1}{7776}$). Die kleine kans zou je kunnen aangrijpen als bewijs dat de dobbelsteen niet zuiver is. Maar, in de context van 1440 opeenvolgende worpen, is het gooien van vijf zessen niet heel raar. Waarschijnlijk had je het evengoed raar gevonden als het niet vijf opeenvolgende zessen waren, maar vijf opeenvolgende enen, tweeën, ..., of vijven. De kans dat zoiets gebeurt, ergens op de dag, is al meer dan 50%.

Voorbeeld: verkeersdoden

Een treffend voorbeeld van het inzoomen op een (te) klein deel van de data haalde vorige maand de media. Op 2 mei jl. kwam het CBS met nieuwe cijfers over het aantal dodelijke verkeersslachtof-

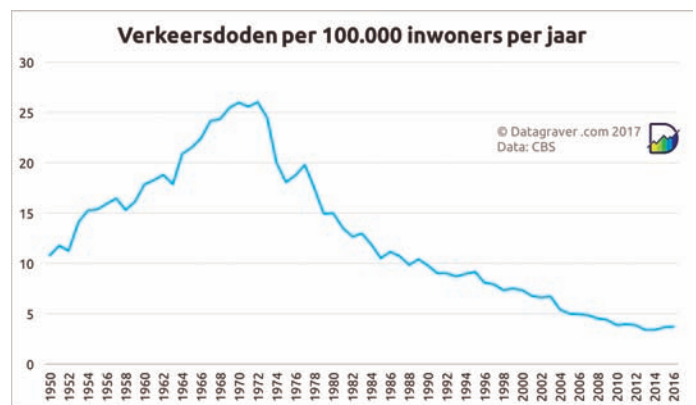
fers: in 2016 waren dit er 629, in 2015 waren het er 621, in 2014 waren het er 570. Reden voor de media om met afschrikwekkende krantenkoppen te komen: ‘Aantal verkeersdoden blijft stijgen’ (*Hart van Nederland*), ‘Aantal verkeersslachtoffers stijgt opnieuw’ (*RTL nieuws*). Feitelijk gezien correcte koppen natuurlijk, maar wel misleidend. Verschillende lobby-instituten gebruikten dit nieuws om — voor hun gunstige — voorstellen te doen. Zo adviseerde de politie om meer geld naar de politie te sturen voor verkeerscontroles en wou Veilig Verkeer Nederland meer geld voor voorlichting over veilig verkeer in Nederland. Nu is er natuurlijk weinig mis in het investeren in betere verkeersveiligheid, maar kom dan wel met de juiste onderbouwing.

Gemiddeld vielen er de afgelopen tien jaar jaarlijks 660,2 doden op de weg, met een standaarddeviatie van 69,3. De score van 2016 zat dus wederom onder dat gemiddelde en de stijging van 8 doden, hoe tragisch ook, was maar een tiende van de standaarddeviatie.

De grafiek [3] zoomt nog verder uit en laat zien wat het verloop is van de verkeersslachtoffers vanaf 1950, gecorrigeerd voor de groei in de bevolking. Uit die grafiek is op geen enkele manier reden tot zorg voor nu te halen.

De veranderingen in de afgelopen twee jaar blijven *ruim* binnen de waardes die de afgelopen tien jaar gemeten zijn. Desondanks wist *de Volkskrant* (krantenbericht ‘Waarom het aantal verkeersdoden weer stijgt’) precies uit te leggen waar die stijging toch aan lag: te weinig controle, te veel smartphones en steeds meer bejaarden op e-bikes — verklaringen die gevonden werden door onder meer te praten met belanghebbenden. Dat is geen *statistiek*, dat zijn *lies* en *damn lies*.

De juiste conclusie had moeten zijn dat het aantal verkeersdoden al een jaar of tien stabiel is op een niveau wat zo'n factor 7 lager ligt dan in begin jaren zeventig. ☹️



Referenties

- 1 A. Agresti en B.A. Coull, Approximate is better than exact for interval estimation of binomial proportions, *The American Statistician* 52 (1988), 119–126.
- 2 P.F. Velleman, Truth, damn truth, and statistics, *Journal of Statistics Education* 16(2) (2008).
- 3 Figuur gemaakt door Datagraver.com, op basis van openbare data van het Centraal Bureau voor de Statistiek