

Casper Albers

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



Column Casper grijpt een kans

De Amerikaanse verkiezingen gepeild

De verkiezingen in de Verenigde Staten waren amper voorbij, of de media rolden over elkaar heen met de ‘De peilers zaten er weer enorm naast!’-berichten. Klopt dat wel?

Ik schrijf deze column enkele dagen nadat de Amerikaanse media Joe Biden en Kamala Harris tot winnaar uitgeroepen hebben. Van twee staten (Georgia en North Carolina) hebben de media de winnaar nog niet aangewezen, maar Biden zit boven de 270 en heeft dus vooralsnog gewonnen. Wel met een slag om de arm: nog geen enkele staat heeft formeel de resultaten vastgelegd en team-Trump is nog druk bezig om via rechtszaken de uitslag te veranderen.

Trump had van tevoren al aangegeven dat zijn strategie zou zijn om *fraude* te roepen als de resultaten hem niet bevielen. Sterker nog, hij had in een debat met Clinton in 2016 al aangegeven de uitslag van de verkiezing alleen te accepteren als hij won. Die hoeven we dus niet serieus te nemen. Maar het verloop van de ‘tussenstand’ bij het tellen van de stemmen was natuurlijk wel opvallend. In veel *swing states* ging Trump steeds meer aan kop, totdat er een keerpunt kwam en Biden in ging lopen. Bij een aantal

staten was het inlopen onvoldoende om de staat blauw te laten kleuren, maar bij andere staten niet. Toen Biden genoeg punten had om het Commonwealth of Pennsylvania te winnen, werd hij tot winnaar uitgeroepen.

Maar hoe ongewoon is het, dat een kandidaat flink voorligt maar uiteindelijk toch verliest? Laten we die vraag beantwoorden door naar Pennsylvania te kijken. De einduitslag is nog niet bekend, maar wel de stand na het tellen van 99% van de stemmen – en laten we die voor nu dan maar als einduitslag interpreteren. Voor het gemak laten we de circa 90 duizend stemmen die op een andere kandidaat zijn uitgebracht of ongeldig zijn buiten beschouwing: het gaat om Biden versus Trump, de rest is bijzaak.

Kandidaat	Tussenstand	Stand na 99% tellen
Joe Biden	2 798 695	3 365 846
Donald Trump	3 118 571	3 320 510
Totaal	5 917 226	6 686 586



Van de eerste 5,9 miljoen stemmen kreeg Trump er 3,1 miljoen: met 52,7% een duidelijke meerderheid. Als alle stemmen in totaal willekeurige volgorde geteld worden, is de statistiek relatief eenvoudig.

Op de middelbare school komen regelmatig statistiekvragen langs als: “Uit een urn met in totaal N ballen, waarvan R rode en $B = N - R$ blauwe ballen worden willekeurig en zonder terugleggen n ballen getrokken. Wat is de kans dat X , het aantal blauwe ballen, gelijk is aan b ($\leq n$)?” Deze kans is te berekenen met de hypergeometrische verdeling:

$$P(X = b) = \frac{\binom{B}{b} \binom{R}{n-b}}{\binom{R+B}{n}}.$$

Met miljoenen stemmen is de kans op een heel specifiek aantal stemmen natuurlijk nagenoeg nul, interessanter zijn cumulatieve kansen: wat is de kans dat, bij $R = 3320510$ Trumpstemmen en $B = 3365846$ Bidenstemmen, er na het tellen van $n = 5917226$ stemmen er $b = 2798695$ of minder op Biden gegeven worden? Die kans is

$$P(X \leq b) = \sum_{k=0}^b P(k).$$

Deze kans is lastig te berekenen, omdat er astronomisch grote getallen in alle tellers en noemers staan. Wassili Hoeffding [1] heeft een bovengrens voor deze kans ontwikkeld:

$$P(X \leq b) = e^{-2t^2n},$$

waarbij t het verschil is in de proportie Bidenstemmen bij de eindstand (0,5034) versus die proportie bij de tussenstand (0,4730): $t = 0,0304$. Hiermee is de bovengrens

$$P(X \leq b) = e^{-10936},$$

oftewel een microscopische kans: er komen zo'n 25 nullen na de komma voordat er een ander getal komt. Soms vinden gebeurtenissen met microscopisch kleine kansen gewoon plaats, maar in dit geval is het veel logischer dat de verklaring is dat het model – dat impliceert dat het tellen van kansen in willekeurige volgorde gebeurt – niet klopt. En dat dan de kansen opeens totaal anders worden, spreekt voor zich. Stel je neemt je voor om eerst alle rode ballen te tellen en vervolgens pas alle blauwe. Dan heb je gegarandeerd een moment (mits $R > 0$) dat een tussenstand laat zien



Op 7 november, vier dagen na de verkiezingen, roept CNN Biden tot winnaar uit

dat 100% van de ballen rood zijn, zelfs als de blauwe ballen enorm in de meerderheid zijn.

Volgens team-Trump klopt het model niet omdat de Democraten massaal gefraudeerd hebben. Hier is, tot op het moment van schrijven van deze column, geen bewijs voor geleverd. Volgens team-Biden is de verklaring anders: Democraten stemmen (vanwege corona) veel meer per post en de poststemmen worden als laatste geteld. Dat dit een vertekening zou opleveren in de ‘tussenstanden’ werd van tevoren door de peilers al genoemd. Dat Pennsylvania superspannend zou worden ook: volgens de laatste pre-election-voorspelling van Nate Silvers FiveThirtyEight had Biden een 50,2% kans om deze staat te winnen.

En die andere staten? Dat ging eigenlijk grotendeels ook zoals voorspeld. Er waren wat uitzonderingen, zo kreeg Trump onverwachts veel stemmen in Florida, maar dat het spannend zou worden in staten als Arizona en Georgia was al bekend. Van tevoren gaf Nate Silver Trump een 10% kans op de overwinning, 1% kans op ‘gelijkspel’, en Biden dus 89% kans op winst. Als Biden ook nog Georgia in de wacht sleept, staat hij op 306 kiesmannen. In 26 van de 100 simulaties van Silver kwam Biden op minder uit, in 1 op exact 306 en dus in 73 hoger. De peilingen hadden Biden dus iets te hoog ingeschat, maar niks wat buiten de onzekerheidsmarges viel. Bij andere peilers vallen gelijksoortige conclusies te trekken. De hijgerige berichten die stelden dat de peilingen er meters naast zaten, waren niet meer dan dat: hijgerig en voorbarig. Wat dat betreft zijn de Democraten niet de enige winnaar van deze verkiezing: de peilingbureaus kunnen tevreden terugkijken. ❖

Referentie

- 1 W. Hoeffding, Probability inequalities for sums of bounded random variables, *Journal of the American Statistical Association* 58(301) (1963), 13–30.