2001 is een geschikt jaar om enkele wetenswaardigheden over de mediaan onder de aandacht te brengen, vindt **Arthur Bakker**. Daarbij gaat hij vooral in op de etymologie, de geschiedenis en de rol van Fermat, Cournot en Fechner.

De mediaan in het middelpunt

De mediaan van een oneven aantal geordende getallen is simpelweg het middelste getal. Bij een even aantal ligt het iets ingewikkelder. In principe voldoet ieder getal tussen de middelste twee getallen, inclusief deze middelste getallen zelf, maar de conventie is om hiervan het gemiddelde te nemen. Deze twee middelste getallen noemt Stigler in het Engels de *comedians*, een woordspeling die in het Nederlands helaas niet werkt. Waarom heet de mediaan zo en waarom is ze bedacht? Op deze vragen geeft dit artikel een antwoord door met zevenmijlslaarzen door de geschiedenis van de mediaan te gaan en daarbij vooral te letten op mooie jaartallen zoals 1601 en 1801.

Etymologie

In het Latijn is *medius* het meest gangbare woord voor ‘in het midden’ of ‘middelste’, maar er zijn ook andere vormen. Zoals ook *mons* en *montanus* allebei ‘berg’ betekenen, worden *medius* en *medianus* allebei gebruikt voor ‘in het midden’ of ‘middelste’. Bij de Romeinse architect Vitruvius uit de eerste eeuw voor Christus zijn formuleringen te vinden als *contra quaternas columnas medianas*, ‘tegenover de middelste vier zuilen’. In dezelfde eeuw schreef de redenaar Cicero bijvoorbeeld *a mediana linea*, ‘vanaf de middelste lijn’. Rond 1400 werd het Latijnse *medianus* ook gebruikt in een medische context, namelijk in *vena mediana,* de middelste ader in de onderarm. Onder deze naam raakte deze ader bekend in verschillende talen. In het Middelnederlands werd bijvoorbeeld geadviseerd ‘zo doe hem laten die mediane’. Vergeleken bij andere woorden voor het midden of de middelste, bleef het gebruik van het woord ‘mediaan’ echter zeer beperkt.

Op zoek naar de werkelijke waarde

Voor het ontstaan van de statistische mediaan moeten we terug naar de tijd waarin gezocht werd naar methodes om de werkelijke waarde te bepalen, vooral binnen de astronomie, navigatie en muntenmakerij. Anders dan nu ging men ervan uit dat er altijd een werkelijke waarde was. Soms werden alleen die waarden gebruikt die goed bij de theorie pasten, soms koos men een waarde uit een cluster in het midden. Vaak was dit dan een meting die onder gunstige omstandigheden totstandgekomen was. De Arabieren rond de tiende eeuw gebruikten ook wel het gemiddelde van de twee extreme waarden: het bereikmidden. Dit idee is niet eens zo gek als we bedenken dat veel verschijnselen redelijk symmetrische data opleveren. Het duurde tot in de zestiende eeuw voordat het rekenkundig gemiddelde van meer dan twee waarden gedefinieerd werd. De astronoom Tycho Brahe, gestorven in 1601, paste dit gegeneraliseerde gemiddelde vermoedelijk als eerste toe om observaties te combineren en zo tot een preciezer resultaat te komen.

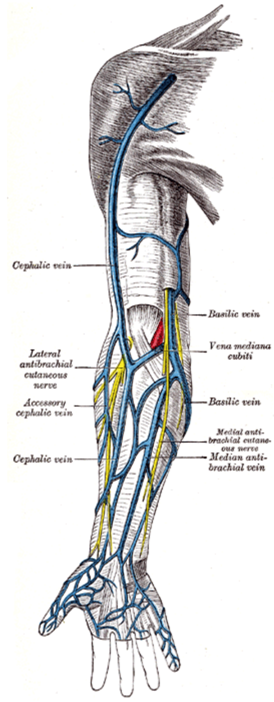

fig. 1 Onderarm met de vena mediana cubiti

De mediaan was toen nog lang niet in beeld. Wel is er één tekstfragment gevonden dat misschien op de mediaan duidt. Het stamt uit 1599, twee jaar voor Brahes dood, en het staat in een navigatieboek van Edward Wright.

> 'Exacte waarheid tussen de onconstante golven van de zee moet niet nagestreefd worden, hoewel goede nieuwe instrumenten goed gebruikt kunnen worden. Toch kunnen we met aandachtige toewijding slechts zo dicht bij de waarheid komen als de aard van de zee, ons gezichtsvermogen en de instrumenten ons dulden. Zelfs als er onenigheid is tussen de observaties hoeven ze niet allemaal stuk voor stuk verworpen te worden; maar zoals wanneer veel pijlen op een doelwit geschoten zijn dat later weggehaald is, zo kan hij die zijn verstand gebruikt en uit wil vinden waar het doelwit stond, toch het midden tussen [the middle place amongst] alle pijlen zoeken: zo komt temidden van vele observaties de middelste waarschijnlijk het dichtste bij de waarheid.'
> (uit Eisenhart 1971)

Helaas geeft Wright geen getalvoorbeelden, zodat niet helemaal zeker is of hij nu de middelste bedoelt of het midden van het bereik. Maar omdat hij in de derde zin schrijft over observaties die niet stuk voor stuk verworpen hoeven te worden, lijkt hij het toch over de middelste *observatie* te hebben en niet het midden van het *bereik*. Bij het bereikmidden zou men namelijk wel degelijk alle waarden behalve de extremen verwerpen, aldus Eisenhart (1971).
Aangezien dit een eenzaam en discutabel voorbeeld is, kunnen we niet spreken van de geboorte van de mediaan. Er was ook lang geen behoefte aan een andere centrummaat dan het gemiddelde, omdat men tot in de negentiende eeuw ervan uitging dat foutenverdelingen symmetrisch waren. Wel werd in enkele andere contexten duidelijk dat er verschillende soorten middens zijn. Achteraf kunnen we hier de mediaan en het gemiddelde in herkennen. Er volgen een paar voorbeelden.

Voorbeelden van verschillende soorten middens

Voorbeeld 1
Fermat vroeg in 1629 naar het punt in een driehoek dat een minimale som van de afstanden tot de hoekpunten op zou leveren. Dit was een ander punt dan het zwaartepunt. Fermats punt blijkt achteraf equivalent te zijn met de mediaan, en het zwaartepunt met het gemiddelde (Monjardet 1991).

Voorbeeld 2
Huygens maakte in 1669 bij het berekenen van levensverwachting voor het eerst onderscheid tussen waarschijnlijke en verwachte levensduur. Tegenwoordig zouden we zeggen: mediane en gemiddelde levensduur (Stamhuis 1996).

Voorbeeld 3
Boscovich ontwikkelde rond 1755 een methode om in observatiegegevens een lineair verband te vinden; daarbij stelde hij als criterium dat de som van de afwijkingen tot deze lijn minimaal moesten zijn (Eisenhart 1961). Hoewel dit achteraf equivalent blijkt te zijn met de mediaan, leidde ook dit incident niet tot de geboorte van de statistische mediaan. Hiervoor moeten we in de kansrekening zijn.
Aan de wieg van de kansrekening stonden onder andere Cardano (1501-1576), Pascal (1623-1662) en Fermat (1601-1665). Het duurde echter tot 1843 voordat Cournot (1801-1877) als eerste de term *valeur médiane* gebruikte voor de mediaan van continue kansverdelingen. Fechner (1801-1887) bedacht later de Duitse naam *Centralwerth*. De mediaan van een rijtje getallen, zoals die nu in de schoolboeken staat, werd toen echter nog steeds niet gebruikt.
Liefhebbers van zoveel-honderdste geboortejaren kunnen hun hart ophalen met dit rijtje geboortejaren. Omdat Cardano niet direct iets met de mediaan te maken heeft, worden hier alleen Fermat, Cournot en Fechner besproken.

fig. 2 Pierre de Fermat

Pierre de Fermat (1601-1665)

Tegenwoordig is het ondenkbaar dat een amateurwiskundige belangrijke bijdragen levert aan een wiskundig vakgebied, maar Fermat deed dit op verschillende terreinen. Hij was een rustige, aimabele man, die een hoge functie had in de gemeente Toulouse. Het bekendst is hij als grondlegger van de getaltheorie. Eeuwenlang heeft zijn laatste stelling wiskundigen hoofdbrekens bezorgd. Pas in 1994 is de stelling door Andrew Wiles bewezen na acht jaar geconcentreerde arbeid. Verder legde Fermat zoals gezegd samen met Pascal de wiskundige basis voor de kansrekening. Minder bekend is dat Fermat onafhan-

kelijk van Descartes de analytische meetkunde uitvond. En dan telt zijn verzameld werk nog vele delen over andere wetenschappen. Door middel van brieven communiceerde hij met wiskundigen over zijn vondsten, maar het liefst gaf hij ze problemen op.
Een van die problemen was een meetkundig probleem dat hij aan Toricelli schreef:

> ‘Gegeven drie punten in het vlak bepaal het punt waarin de som van de afstanden tot deze drie punten minimaal is.’
> (Fermat verzameld werk band IV, p. 141)

Dit punt wordt nog steeds het punt van Fermat of Toricelli genoemd. Een constructie van Fermats punt is te vinden in figuur 3. Op de zijden van de originele driehoek zijn gelijkzijdige driehoeken geplaatst (er mag geen hoek groter dan $120^{\times\circ}$ bij zitten). Door middel van een handige rotatie van $60^{\times\circ}$ om een van de hoekpunten van de originele driehoek kan bewezen worden dat het punt F inderdaad een minimale som.

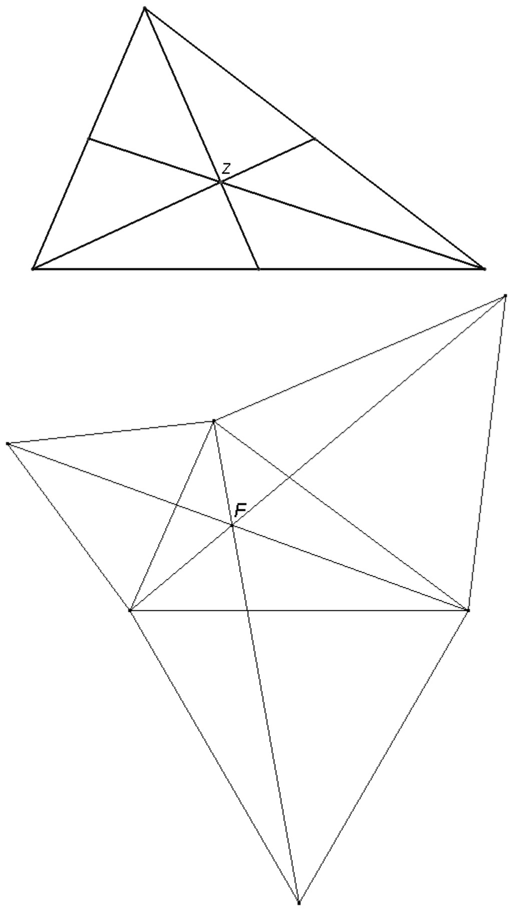

fig. 3 Het zwaartepunt (Z) en het punt van Fermat (F)

Als we dit probleem eendimensionaal opvatten, zoeken we het getal M dat minimale som van de afwijkingen heeft tot M. Binnen de zoektocht naar de beste waarde is dit wel eens als criterium genomen, bijvoorbeeld door Boscovich rond 1755, en de mediaan blijkt aan dit criterium te voldoen. De lezer wordt uitgedaagd zelf een elegant bewijs te vinden voor de equivalentie van de mediaan en Fermats punt, en voor de equivalentie van het gemiddelde en het zwaartepunt.

Antoine-Augustin Cournot (1801-1877)

Bij gebrek aan wiskundigen die in 1701 geboren zijn, maken we een sprong naar 1801, toen Cournot en Fechner geboren werden. De mediaan werd tot deze tijd niet gebruikt in de huidige statistische betekenis. Wel waren grote wiskundigen als Laplace, Legendre en Gauss op de hoogte van allerlei eigenschappen van de mediaan bij continue verdelingen. Zij gebruikten alleen andere namen, zoals *valeur probable* en *milieu de probabilité* voor de mediaan. De termen illustreren dat de kansrekencontext sterk aanwezig was, en dat men nog steeds uitging van symmetrische foutenverdelingen. Voor Cournot waren de gebruikte termen voor de mediaan niet acceptabel. Hij koos daarom in 1843 de term *valeur médiane* in de context van kansverdelingen, en uit de eerste paragraaf blijkt dat dit een geschikte naam is. Het woord *médiane* werd buiten de medische context nog erg weinig gebruikt en betekende ‘middelste’ of ‘in het midden’, net als *milieu* en *moyen*, die veel meer gebruikt werden.
Cournot is verder bekend geworden door zijn spot met de *l’homme moyen*, ‘de gemiddelde mens’ van Quetelet. Cournot vond dat een gemiddelde van allerlei maten ‘simpelweg een onmogelijk persoon’ opleverde (Cournot 1843, p. 141). Hij vond dit net zo onzinnig als het nemen van het ‘gemiddelde’ van een verzameling rechthoekige driehoeken. De gemiddeldes van de zijden hoeven geen rechthoekige driehoek op te leveren, dus de essentiële eigenschap gaat verloren.

Gustav Theodor Fechner (1801-1887)

Zo’n drie decennia later, in 1874, gaf Fechner de mediaan een Duitse naam: *Centralwerth*. Lang voordat de psychologie en sociologie als disciplines bestonden, probeerde hij psychologische en sociologische verschijnselen te beschrijven met middelen die in de astronomie hun diensten bewezen hadden. Hij vond daarbij de mediaan makkelijker te berekenen dan het gemiddelde, maar hij had ook theoretische motieven om andere centrummaten dan het gemiddelde te gebruiken. In de paragraaf over Fermat stond dat de mediaan de som van de absolute afwijkingen minimaliseert, zoals het gemiddelde de som van de afwijkingen tot de tweede macht minimaliseert. Mediaan en gemiddelde zijn dus speciale gevallen van de zogenaamde *Potenz-mittelwerthen*, de waarden waarbij de som van de afwijkingen tot de *n*-de macht minimaal is: $\Sigma|x_i - x|^n$ is minimaal.
In tegenstelling tot wat de meeste wetenschappers in die tijd veronderstelden, ging Fechner ervan uit dat verdelingen normaal gesproken asymmetrisch zijn. Dit staat haaks op de naam die Galton in 1889 gaf aan de beroemde symmetrische foutenwet: de normale verdeling (David 1995).

De aanname dat verdelingen symmetrisch zijn is een mogelijke reden waarom de mediaan zo laat ontstaan is ten opzichte van het gemiddelde. Bij scheve verdelingen komen gemiddelde en mediaan namelijk uit elkaar te liggen. Fechner gebruikte nu zijn *Potenz-mittelwerthen* om de scheve verdelingen te karakteriseren. Het verschil tussen gemiddelde en mediaan is een maat voor de scheefheid. In figuur 4 is de mediaan lager dan het gemiddelde omdat deze verdeling scheef naar rechts is. Een dergelijk fenomeen treedt ook op bij salarisverdelingen.

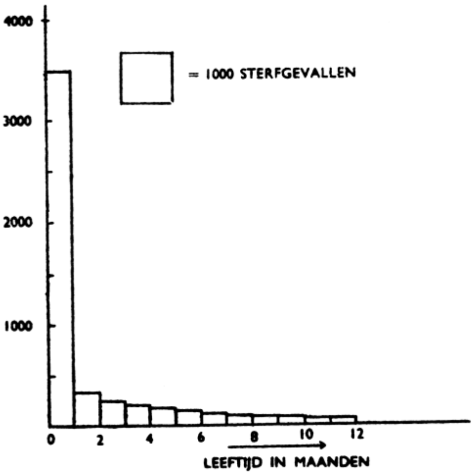

fig. 4 Verdeling van babysterfte in 1953 (uit Freudenthal 1966)

Galton gebruikte de term *median value* in 1882 als eerste in het Engels. Ook zorgde hij voor de doorbraak van de mediaan als middelste waarde van een rij geordende getallen. Het was zelfs zo dat hij de mediaan en de kwartielen intuïtief veel duidelijker vond dan het gemiddelde en de voorloper van de standaarddeviatie.

Samenvatting

Hoewel het woord *medianus* al zeker 2000 jaar oud is, wordt het pas sinds anderhalve eeuw gebruikt voor de middelste waarde van een rij geordende getallen. Vergeleken met het gemiddelde dat al in Pythagoras' tijd bekend was, is de mediaan laat ontstaan. Er zijn wel enkele contexten vanaf de zeventiende eeuw waarin we met moderne ogen de mediaan kunnen herkennen, maar het duurde tot diep in de negentiende eeuw voordat de ons bekende mediaan gebruikt werd. Een verklaring hiervoor is de aanname dat fenomenen, zoals fouten rond de 'werkelijke' waarde, symmetrisch verdeeld waren. Binnen de natuurwetenschappen was daar een goede reden voor, maar binnen de sociale en economische wetenschappen ontstond uiteindelijk aandacht voor asymmetrische verdelingen. Daarmee ontstond ook de behoefte aan een alternatieve centrummaat die minder gevoelig is voor extreme waarden en dichter bij de meerderheid van de data ligt.
Voor het onderwijs zou dit kunnen betekenen dat de mediaan minder simpel is dan ze lijkt. Om haar zinvol te gebruiken moeten leerlingen namelijk iets weten van uitschieters en scheve verdelingen.

Arthur Bakker, Freudenthal Instituut, Utrecht

Literatuur

Cournot, A.A. (1843). *Exposition de la Théorie des chances et de Probabilités*. Paris: L. Hachette. (herdruk 1984, ed. Bernard Bru. Paris: J. Vrin.)

David, H.A. (1995). First (?) Occurences of Common Terms in Mathematical Statistics. *The American Statistician*, *49*(2), 121-133.

David, H.A. (1998). First (?) Occurences of Common Terms in Probability and Statistics—A Second List, with Corrections. *The American Statistician*, *52*(1), 36-40.

Eisenhart, C. (1961). Boscovich and the Combination of Observations. In M. Kendall & R.L. Plackett (1977). *Studies in the History of Statistics and Probability. Volume II.* London: Charles Griffin & Company Limited.

Eisenhart, C. (1971). *The development of the concept of the best mean of a set of measurements from antiquity to the present day. 1971 A.S.A. Presidential Address.* (Unpublished).

Fermat, P. de (1891). *Oeuvres de Fermat*. *Tome I.* Paris: Gauthier-Villars.

Freudenthal, H. (1966). *Waarschijnlijkheid en statistiek*. Haarlem: De Erven F. Bohn. Derde druk.

Monjardet, B. (1991). Eléments pour une histoire de la médiane métrique. In J. Feldman, G. Lagneau & B. Matalon (eds.). *Moyenne, Milieu, Centre*. Paris: Éditions de l'École des Hautes Études en Sciences Sociales.

Stamhuis, I.H. (1996). Christiaan Huygens correspondeert met zijn broer over levensduur. *De Zeventiende Eeuw, 12*, 161-170.

Stigler, S.M. (1977). Fractional Order Statistics, with Applications. *Journal of the American Statistical Association*, *72*, 544-550.

Een overzichtelijke website met biografieën van wiskundigen is http://www-history.mcs.st-and.ac.uk/history/ .