Statistiek voor tandartsen (III)

Appels plus peren; meetniveau

A.H.B. Schuurs, tandarts*)
H.J. Duivenvoorden, methodoloog)**
M.A. van 't Hof, statisticus*)**

Uit *)de vakgroep Cariologie en Endodontologie, Academisch Centrum Tandheelkunde Amsterdam (ACTA), de **)vakgroep Medische Psychologie en Psychotherapie, Faculteit der Geneeskunde, Erasmus Universiteit te Rotterdam en ***)Medisch Statistische Afdeling, Faculteit der Geneeskunde en Tandheelkunde van de Katholieke Unviersiteit te Nijmegen.

Trefwoord: Statistiek

Datum van acceptatie: 28 september 1990.

Adres: Dr. A.H.B. Schuurs, ACTA, Louwesweg 1, 1066 EA Amsterdam.

Samenvatting
Een score is een getal waarmee een waarneming wordt gerepresenteerd, zoals lengte of de plaats van volgorde in een reeks waarnemingen. Scores kan men bewerken, maar appels kan men niet (zonder meer) bij peren optellen. Zo ook betekent het gemiddelde geloof = 1,5 (met 1 = katholiek, 2 = protestant en 3 = ongelovig) niets, maar het gemiddelde 8 van rapportcijfers geeft een indicatie van de leerprestaties (als de leerstof valide is gemeten en men niet heeft gefraudeerd). Welke rekenkundige bewerkingen mogen worden uitgevoerd, heeft te maken met het zogenoemde 'meetniveau'. Hier worden meetniveaus, 'centrale' maten (bijvoorbeeld het 'gemiddelde') en afwijkingen daarvan, onder andere de standaardafwijking, beschreven.

SCHUURS AHB, DUIVENVOORDEN HJ, VAN 'T HOF MA. Statistiek voor tandartsen (III). Appels plus peren; meetniveau. Ned Tijdschr Tandheelkd 1990; 97: 505-8.

1 INLEIDING

De gegevens uit onderzoek worden vaak rekenkundig bewerkt, onder andere wanneer men een gemiddelde berekent, de standaardafwijking bepaalt of bepaalde toetsen uitvoert.

De gemiddelde leeftijd van één 10-, 20- en 30-jarige is 20 jaar. Als aan een arts het cijfer 1 wordt toegekend, aan een tandarts 2 en aan een specialist 3, kunnen we het beroep, net als de leeftijd in het vorige voorbeeld 'middelen', met als uitkomst 1,5. Maar dit laatste gemiddelde kan niet worden geïnterpreteerd en bezit geen informatieve waarde. Berekeningen zijn dus niet altijd zinnig; de zin hangt af van het zogenoemde *meetniveau* van de variabelen. Er zijn enkele meetniveaus te onderscheiden die hierna, gerangschikt van 'laag' tot 'hoog', worden besproken en waarvan wordt vermeld welke rekenkundige bewerkingen zinnig zijn. De verstrekte informatie is in vele boeken te vinden.[1-4]

2 NOMINAAL

De toekenning van scores, de waarden 1, 2, 3 enzovoorts, gebeurt in dit geval uitsluitend om categorieën (klassen) te onderscheiden en aan te duiden, zoals hiervoor met de artsen, tandartsen en specialisten is gedaan. Echter, met evenveel recht kan aan de specialisten het getal 1 of 10 worden toegekend of kan men met de letter a, b en c werken. Er is sprake van een niet-geordende categorisering (klasse-indeling). De onderzoeker kan zelf bepalen welke scores (of codes) hij aan de categorieën toekent.

2.1 Bewerkingen

Men kan tellen hoeveel personen in de verschillende categorieën aanwezig zijn; deze frequenties zijn ook in percentages (= aantal per klasse gedeeld door het totale aantal, maal honderd) of fracties (aantal per klasse gedeeld door totale aantal) uit te drukken en visueel overzichtelijk weer te geven in balkdiagrammen (afb. 1) en lijngrafieken. Ook is de **modus, de klasse met de grootste frequentie** (= klasse met het grootste aantal waarnemingen), te bepalen. Een en ander is in tabel I geïllustreerd; de modus aldaar is vet gedrukt.

Op nominale gegevens kunnen statistische toetsen worden uitgevoerd, die gebruik maken van de frequenties in de klassen, de zogenoemde tabelanalyse, die later zal worden besproken.

2.2 Dichotomie

De meest simpele vorm van een nominale variabele is de zogenaamde *dichotomie*. Een dichotome variabele kent slechts twee klassen, bijvoorbeeld carieus en niet-carieus of mannen en vrouwen. Variabelen met meer dan twee klassen worden gedichotomiseerd door alle klassen tot twee samen te voegen. Men kan variabelen van elk meetniveau dichotomiseren. Ter illustratie: men kan cariësvrije elementen onderscheiden van elementen met respectievelijk klasse I-, II-, III-, IV- en V-cariës maar men kan deze vijf (nominale) klassen van Black ook samenvoegen tot één klasse, namelijk 'carieuze elementen', waardoor een gedichotomiseerde variabele ontstaat met de twee klassen 'cariësvrij' en 'carieus'. Dichotome en gedichotomiseerde variabelen kunnen vaak als 'metrisch' (par. 4) worden behandeld, dat wil zeggen men kan het gemiddelde berekenen en onder bepaalde omstandigheden kan de t-toets of variantieanalyse worden toegepast. Bij dichotome variabelen is het gemiddelde gelijk aan de fractie (proportie). Ter illustratie: kenden we aan mannen de code 0 en aan vrouwen de code 1 toe en vinden we als het gemiddelde voor de variabele 'geslacht' = 0,65, dan betekent dat, dat 65% vrouw is.

3 ORDINAAL

Bij dit meetniveau worden met de getallen 1, 2, 3 enzovoorts – of de letters a, b, c – niet alleen de klassen aangeduid, maar ook een rangorde. Zo kunnen een aantal goudlegeringen naar hardheid worden ingedeeld. Of daarbij nu cijfers of letters worden toegekend maakt niet uit, mits het laagste cijfer of de eerste letter de zachtste legering en het hoogste cijfer of laatste letter de hardste legering representeert – of omgekeerd. De indeling van een populatie in sociale klassen is een ander voorbeeld van het ordinale meetniveau.

3.1 Bewerkingen

Als maat voor het centrum voor ordinale gegevens kan de mediaan worden gebruikt. De **mediaan is de middelste waarde** van de waarnemingen: erboven en eronder liggen een gelijk aantal waarnemingen, namelijk 50%. De mediaan is het vijftigste percentiel (P50).Om de mediaan te bepalen moeten de waarnemingen eerst naar grootte worden gerangschikt. Bij een oneven aantal is de middelste waarde de mediaan. In de getallenreeks 0, 5, 20, 30, 70 is de mediaan 20. Bij een even aantal worden de twee middelste waarnemingen opgeteld en gedeeld oor twee: van de reeks 5, 20, 30, 90 is de mediaan (20 + 30)/2 = 25. Beschikt men over twee reeksen getallen dan is de mediaan van beide reeksen samen echter niet gelijk aan hun optelsom gedeeld door twee.

Op ordinale metingen kunnen later te bespreken rangtoetsen worden toegepast.

4 METRISCH (ratio)

Het is nuttig onderscheid te maken tussen continue en discrete variabelen.

4.1 Continuum

Aan veel metingen ligt een continuum ten grondslag. Ter verduidelijking van het begrip continuum: we kunnen de afstand tussen twee voorwerpen meten in meters, maar tussen twee willekeurige schaalpunten, zeg tussen 3 en 4 m, kan altijd nog een extra waarde (= meetpunt) een plaats vinden, bijvoorbeeld één in cm en daartussen weer één in nog kleinere eenheden, in mm, enzovoorts; in wezen kan de afstand elke waarde aannemen uit het continuum. Een continue variabele kent dus een oneindig aantal waarden. Continue grootheden kunnen, na ingedeeld te zijn in *klassen*, in een histogram met aan elkaar grenzende balken of in een lijn met knikken (veelhoek, polygoon) worden uitgebeeld.

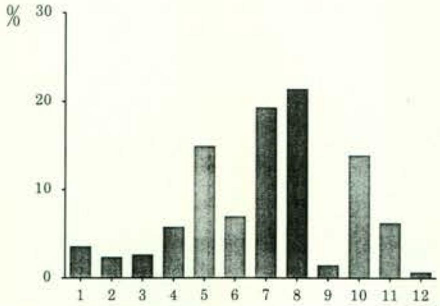
4.1.1. Discreet en tellingen

Een andere verschijningsvorm van het metrische meetniveau is de *discrete* variabele, die praktisch gezien alleen gehele waarden kan aannemen. Een discrete variabele kent een beperkt (eindig) aantal klassen (vergelijk een dobbelsteen met zes en een verkeerslicht met drie klassen). Voorbeelden van tellingen zijn het aantal elementen in de mond, de toename van het aantal carieuze vlakken en het aantal tandartsbezoeken in 1989.

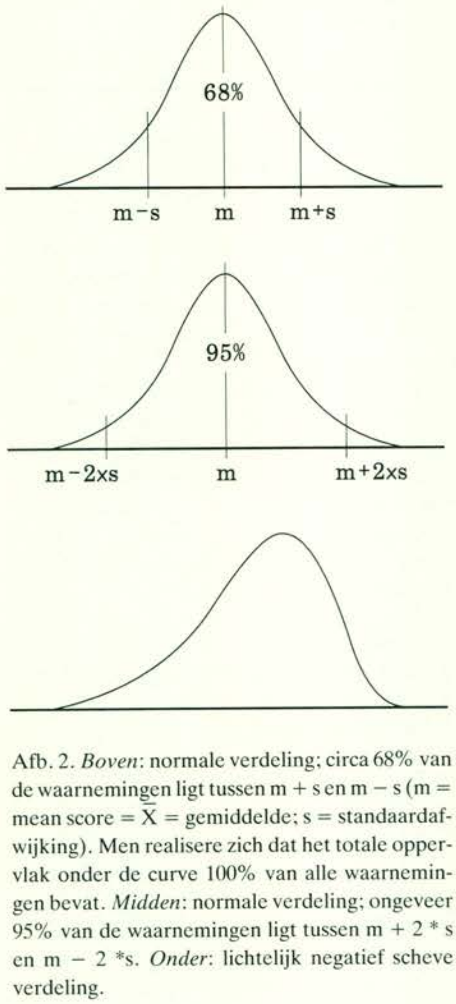
4.2 Bewerkingen

Iemand van 180 cm lengte is tweemaal zo lang als iemand van 90 cm. Dit soort rekenkundige bewerkingen zijn zinvol voor gegevens van metrisch niveau.

De maat voor het centrum bij metrische metingen is het **rekenkundige gemiddelde**, kortweg gemiddelde genoemd (arithmetic mean). Dit gemiddelde en de standaardafwijking (variatie; zie par. 4.4) worden *parameters* genoemd. Op bepaalde voorwaarden mogen klassieke toetsen, zoals t-toetsen en variantie-analyse (ook wel parametrische toetsen genoemd) worden gebruikt, mits aan de eisen van 'normaliteit' is voldaan. Toetsen waarbij deze eis niet



Afb. 1. Balkdiagram: frequenties van het aantal ondervraagde tandartsen in procenten, per regio, zoals vermeld in tabel I.



Afb. 2. *Boven*: normale verdeling; circa 68% van de waarnemingen ligt tussen m + s en m − s (m = mean score = $\overline{X}$ = gemiddelde; s = standaardafwijking). Men realisere zich dat het totale oppervlak onder de curve 100% van alle waarnemingen bevat. *Midden*: normale verdeling; ongeveer 95% van de waarnemingen ligt tussen m + 2 * s en m − 2 *s. *Onder*: lichtelijk negatief scheve verdeling.

wordt gesteld of die voor lagere meetniveau's kunnen dienen, heten non-parametrisch. Met 'normaliteit' wordt bedoeld dat de gegevens normaal verdeeld zijn, hetgeen grafisch uitgebeeld in de welbekende Gausse-verdeling (klokvormige curve) resulteert (afb. 2). Deze verdeling laat zien waar het gemiddelde ligt en vertelt iets over de spreiding; aan beide zijkanten (staarten) liggen weinig waarnemingen, naar het midden toe meer en meer.

Tellingen lenen zich ook voor bewerkingen als het berekenen van het gemiddelde, t-testen en variantie-analyse. In de praktijk blijkt echter vaak dat tellingen 'scheef' (Eng.: skewed) verdeeld zijn, dat wil zeggen dat de top van de verdeling naar links of naar rechts (afb. 2) ligt. Door in die gevallen de gegevens te transformeren (door kwadrateren in geval van een negatieve scheve verdeling of worteltrekken bij een positief scheve, waardoor 'uitschieters' verdwijnen) kunnen toch klassieke toetsen worden toegepast.

4.3 Rekenkundig gemiddelde

Om het aantal carieuze caviteiten van een groep uit te drukken is het rekenkundige gemiddelde (aangeduid als X, met een streepje erboven), $\overline{X}$ een goede maat voor het centrum. Een nadeel van het rekenkundig gemiddelde is dat extreme waarden, zogenaamde 'uitbijters', een vertekend beeld voor ogen kunnen spiegelen. Stel dat het gemiddelde bruto jaarlijkse inkomen van vijf tandartsen *f* 100.000 is. Met dit gemiddelde is het mogelijk dat ieder van de vijf tandartsen ongeveer evenveel verdient, bijvoorbeeld: *f* 95.000, *f* 98.000, *f* 100.000, *f* 102.000 en *f* 105.000. Maar het gemiddelde *f* 100.000 kan ook het resultaat zijn van de volgende inkomensreeks: *f* 40.000, *f* 50.000, *f* 60.000, *f* 70.000 en *f* 280.000. De 'uitbijter' *f* 280.000 trekt het gemiddelde fors omhoog en wekt de indruk dat alle tandartsen veel verdienen.

Een andere illustratie van het mogelijk op een dwaalspoor worden gezet door een gemiddelde, is te vinden in de rapportage over cariës. Het DMF-getal voor een groep kan relatief laag zijn, omdat velen geen en weinigen veel cariës hebben. De conclusie dat cariës overal is verdwenen geldt dan niet. Als uitschieters aanwezig zijn, is het informatief naast het gemiddelde tevens de mediaan te vermelden. Bijvoorbeeld, in de eerste serie inkomens van de tandartsen is de mediaan = *f* 100.000 en in het tweede geval *f* 60.000.

Het voorbeeld over het gemiddelde inkomen van de vijf tandartsen betreft een 'steekproef'. Het gemiddelde inkomen van de populatie, in dit geval dat van alle tandartsen, kennen we niet. In een latere bijdrage wordt duidelijk dat aan de hand van een steekproef toch binnen zekere marges uitspraken over het gemiddelde van de populatie kunnen worden gedaan.

4.4 Spreidingsmaten

Het gemiddelde geeft geen aanwijzing voor de variatie, voor de spreiding van gegevens. De standaardafwijking (zie 4.4.1) doet dat wel. Een cruciale rol bij de berekening van de staandaardafwijking speelt de variantie.

De spreiding in de waarnemingen kan op verschillende manieren worden aangegeven. Zo ziet men soms naast het gemiddelde de laagste en hoogste waarden vermeld: $\bar{X}$ = 27 jaar, *range* (= spreidingsgebied) 18–39 jaar. Als alternatief geeft men wel de *spreidingsbreedte* op, het verschil tussen de grootste en kleinste waarneming: $\bar{X}$ = 27 jaar, spreidingsbreedte is (39–18 =) 21 jaar. De *gemiddelde afwijking* is te vinden door alle waarnemingen (aangeduid als X) te schrijven als deviatie (afwijking) in *positieve* zin van het rekenkundig gemiddelde, dus $\bar{X} - X$. Hierna worden de deviaties gesommeerd en de uitkomst daarvan gedeeld door het aantal waarnemingen. De gemiddelde afwijking wordt weinig gebruikt in de statistiek. Meer populair is de standaardafwijking, die ook op het principe 'afwijking van het gemiddelde' is gebaseerd.

Tabel I. Verdeling van tandartsen naar provincie (n = 413) uit de totale tandartspopulatie (N = 5036) die ziekenfondsverzekerden behandelen.*)

Regio	Absolute aantal (n) (= frequentie)	% (= relatieve frequentie)
1. Groningen	15	3,6
2. Friesland	10	2,4
3. Drenthe	11	2,7
4. Overijssel	24	5,8
5. Gelderland	62	15,0
6. Utrecht	29	7,0
7. Noord-Holland	80	19,4
8. Zuid-Holland	**89**	21,5
9. Zeeland	6	1,5
10. Noord-Brabant	58	14,0
11. Limburg	26	6,3
12. Flevoland	3	0,7
Totaal	413	100,0

*) Ontleend aan Statistisch Overzicht 1987, Commissie Tandheelkundige Statistiek, Zeist.

4.4.1 Variantie en standaardafwijking als spreidingsmaten

De berekening van de variantie van een steekproef is eenvoudig, geïllustreerd aan de hand van het aantal vullingen van tien studenten (tabel II). Als eerste worden alle gevonden aantallen (aangeduid als X) geschreven als deviatie van het gemiddelde cijfer $\bar{X}$. Voor student A is de deviatiescore: 4 – 6 = –2, voor B: 7 – 6, enzovoorts. Door nu alle deviatie-scores te kwadrateren en op te tellen (de Griekse sigma Σ betekent som, dus $\Sigma x^2 = 44$) en vervolgens te delen door het aantal n, of liever door n min 1 (in dit geval: 10 – 1 = 9) vinden we de *variantie* $s^2 = 4,9$.

De variantie zegt iets over de spreiding, maar is niet goed interpreteerbaar. De vierkantswortel uit de variantie levert de beter interpreteerbare *standaardafwijking* op: $s = \sqrt{4,9} = 2,2$ vullingen.

De gewoonte te delen door n –1 in plaats van door n, is niet belangrijk voor grote waarden van n, want het verschil tussen beide uitkomsten is verwaarloosbaar. Bij kleine waarden van n is het wel van belang, hetgeen het beste wordt geïllustreerd voor n = 1. Bij deling door n zou dan steeds een spreiding = 0 worden gevonden, hetgeen onzin is. Bij deling door n – 1 komt de spreiding uit op 0/0, dat wil zeggen een onberekenbare grootheid, hetgeen wel reëel is.

Tabel II. Aantal restauraties bij een groep van tien studenten, de aantallen genoteerd als deviatiescores en hun kwadraten, gevolgd door de berekening van de standaardafwijking s.

Student	Aantal X	Deviatiescore $(X-\bar{X})$	Deviatiescore² $(X-\bar{X})^2$
A.	4	–2	4
B.	7	1	1
C.	10	4	16
D.	2	–4	16
E.	6	0	0
F.	5	–1	1
G.	6	0	0
H.	7	1	1
I.	8	2	4
J.	5	–1	1
	— +	— +	— +
Totaal	60	0*)	44

gemiddelde: $\bar{X} = 60/10 = 6$
variantie: $s^2 = 44/(10\text{-}1) = 4,9$
standaardafwijking: $sd = \sqrt{s^2} = \sqrt{4,9} = 2,2$ (ook sd genoemd)

*) De deviatiescores leveren na optelling altijd 0 op.

5 NUT VAN GEMIDDELDE EN STANDAARDAFWIJKING

Met behulp van het gemiddelde $\bar{X}$ en de standaardafwijking s wordt het aantal vullingen van de studenten compact beschreven. Presentatie van het gemiddelde en de standaardafwijking is vooral van belang, omdat niet voor iedere variabele een histogram kan worden gepubliceerd en men toch een groot aantal gegevens compact wil weergeven. De interpretatie van het gemiddelde en standaardafwijking is dan dat het aantal vullingen voor circa 68% van de studenten ligt tussen $\bar{X} - s$ en $\bar{X} + s$ (afb. 2 boven), natuurlijk op voorwaarde dat de verdeling normaal is, dat wil zeggen een Gausse-curve; ruim twee derde deel van het aantal vullingen in ons voorbeeld ligt tussen (6 – 2,2 =) 3,8 en (6 + 2,2 =) 8,2. Nemen we $\bar{X} \pm 2 * s$ (resp. 10,4 en 1,6), dan ligt ongeveer 95% van alle vullingen tussen die twee waarden.

6 SLOT

Afhankelijk van de aard van het meetniveau van de meetgegevens en hun verdeling (normaal of scheef) moeten de statistische parameters worden gekozen voor beschrijvingen in publikaties. In het algemeen wil men beschrijven wat het meeste voorkomt (dit wordt de maat voor centrale tendentie genoemd). Afhankelijk van de situatie (meetniveau, scheefheid) kan gekozen worden voor modus (nominale variabelen), mediaan (ordende variabelen en scheef verdeelde metrische variabelen) of voor gemiddelde (redelijk normaal verdeelde metrische variabelen). Naast deze maten is het nuttig de afwijkingen van de centrale tendentie (samenvattend) te vermelden. Voor nominale variabelen kan dat alleen in een histogram, voor ordinale en scheef verdeelde metrische variabelen zijn de percentielpunten (of range) het meeste geschikt en voor normaal verdeelde metrische variabelen wordt in het algemeen de standaardafwijking gepresenteerd.

In de rest van deze serie artikelen speelt bij de bespreking van statistische grootheden (toetsen) het meetniveau of de scheefheid van de verdeling dan ook steeds een belangrijke rol.

SUMMARY

STATISTICS (III)

Apples plus pears; level of measurement

Key word: Statistics

A score is a numerical expression of an observation, such as length or number in rank ordered scores. Scores may be manipulated. A mean score 8 for arithmetics represents an indication of the achievement, but the mean score 1,5 for 'belief' (from 1 Roman Catholic, 2 Anglican and 3 = not religious) is meaningless, due to the so called 'level of measurement'. In relation to the measurement scales, measures of central tendency and deviations thereof, among which the standard deviation, are presented.

LITERATUUR

[1]SIEGEL S, CASTELLAN NJ. Nonparametric statistics. New York: McGeaw-Hill Book Company, 1988.
[2]WILLEMSEN EW. Understanding statistical reasoning. San Francisco: W.H. Freeman and company, 1974.
[3]VAN DEN BRINK WP, KOELE P. Statistiek. Deel I. Datareductie. Meppel: Boom, 1985.
[4]NIE NH, HULL CH, JENKINGS JG, et al. Statistical package for the social sciences. New York: McGraw-Hill Book Company, 1975.