

Enige kritische kanttekeningen bij het artikel van Van Doorn, Saris & Lodge "Discrete or continuous measurement: What difference does it make?"

KM 10(1983) pag 104-120

Ab Mooijaart *

Jaap van der Pol *

In bovenstaand artikel vergelijken de auteurs twee manieren om attitudes te meten: 1) door gebruik te maken van categorische variabelen en 2) door gebruik te maken van continue variabelen (lijnstukken en getallen). In deze notitie willen wij voornamelijk ingaan op de methodologische aspecten van hun onderzoek. Ook zullen enkele data-analytische zaken aan de orde komen omdat deze van belang zijn in verband met de door hun verkregen resultaten.

Een eerste opmerking is dat de onderzoekers in feite geen categorische metingen hebben verricht, maar de continue variabelen hebben gediscrètiseerd, hetgeen iets anders is dan de titel doet vermoeden. Bovendien geldt dat - indien variabelen categorisch (discreet) zijn - lineaire verbanden tussen de variabelen niet voor de hand liggen en in dat geval is Pearson's product-moment-correlatie niet zo'n goede maat om een verband tussen variabelen weer te geven. Analyse met behulp van niet-lineaire modellen (zoals recentelijk uitvoerig beschreven door Gifi (1981)) ligt daarom meer in de rede. Ook is de keuze van associatiemaat niet bindend: vanuit een meer sociologische invalshoek, waar discrete variabelen als geslacht, godsdienst, status vooral van belang zijn, heeft Yule (1912) andere maten voorgesteld die meer geschikt zijn voor categorische gegevens.

Ook de op blz. 105 genoemde nadelen van de categorische schalingsmethode zijn vatbaar voor kritiek.

- Discrete variabelen kunnen wel degelijk op intervalniveau gemeten zijn. Bijvoorbeeld bij de methode van "equal appearing intervals" kunnen rating scales schattingen geven van schaaltypes van een hoger niveau dan ordinaal niveau (BirnbauM (1974), Orth (1982)). Dit is in tegenstelling met wat Van Doorn, Saris & Lodge beweren in punt 1.
- In het algemeen geldt, en dat is reeds sinds het begin van deze eeuw herhaaldelijk aangetoond (zie ook b.v. de controversE Yule en Pearson zoals beschreven door MacKenzie (1978)), dat wanneer de product-moment-correlatie als associatiemaat voor discrete of voor

*Vakgroep Methoden en Technieken

gediscretiseerde variabelen gebruikt wordt te lage correlaties zullen worden gevonden (een mooi voorbeeld hiervan wordt gegeven door Champney en Marshall (1939), waar correlaties op basis van metingen in centimeters en millimeters met elkaar vergeleken worden. Eén van de gevallen geeft een stijging in correlatie te zien van .602 naar .762. Bovendien geldt in hun onderzoek dat de variabele met het minste aantal categorieën nog altijd 9 categorieën heeft tegenover drie categorieën bij Van Doorn, Saris & Lodge. Het effect op de correlatie zal daardoor in het laatste geval nog veel groter zijn).

- Dergelijke lage correlaties hebben tot gevolg dat schattingen van betrouwbaarheden en validiteiten ook laag zullen uitvallen, althans als we de klassieke testtheorie als leidraad nemen. Immers, deze betrouwbaarheidscoëfficiënten zijn uitgedrukt in termen van correlaties. Door dit argument zal punt twee van Van Doorn, Saris & Lodge in een ander daglicht moeten worden gezien.
- Wat betreft punt 3: Dat bij het gebruik van categorische variabelen in sociaal-wetenschappelijk onderzoek sprake is van verlies aan informatie is bij het gebruik van categorische variabelen geen evidente zaak. Immers, het dwingen van respondenten om "genuanceerde" antwoorden te geven betekent geenszins dat die personen ook genuanceerd denken (een genuanceerd oordeel hebben). Dit probleem zou enigszins ondervangen zijn als bij de respondenten ook daadwerkelijk categorische metingen waren verricht (en niet zoals in dit onderzoek de continue metingen gediscretiseerd waren). De vraag dringt zich dan ook op of we hier niet te maken hebben met *schijnnaauwkeurigheid*.
- In verband daarmee valt het op dat de auteurs niet vermelden hoe inter-individuele verschillen t.a.v. het schaalgebruik beschouwd moeten worden (betekent de score 68 bij persoon A hetzelfde als de score 68 bij persoon B?). Bij grofmazige metingen lijkt o.i. dit probleem van minder belang, hetgeen zou pleiten voor categorische metingen. (Coombs (1964) gaat in feite uit van al dit soort methodologische problemen en probeert ze door niet-metrische oplossingen te zoeken zoveel mogelijk te omzeilen.)

Opvallend is overigens dat de onderzoekers een log-transformatie op de continue metingen uitvoeren. Zij doen dat op basis van theoretisch-psychofische argumenten. Een meer platvloerse argumentatie is dat daardoor grote verschillen worden "plat geslagen", vooral aan de uiteinden van de schaal.

- Naast de keuze van de onderzoekers voor de product-moment-correlaties is er nog een belangrijke, andere, aanpak mogelijk, welke door de auteurs niet genoemd wordt. Het ontbreken hiervan leidt volgens ons tot een onverantwoord groot hiaat in het artikel. Deze aanpak, die aansluit bij punt 4 van Van Doorn, Saris & Lodge, is die van Pearson. Pearson (1900) postuleert een onderliggende, continue, variabele - de 'latent response variable' - waarbij elke categorie van de geobserveerde variabele overeenkomt met een *interval* van de response variabele. De bijbehorende correlaties van dit soort latente variabelen worden wel de polychorische correlaties genoemd. In het hier besproken onderzoek zou een dergelijke aanname ook zeer voor de hand hebben gelegen, ook omdat de plaatjes bij het artikel aangeven dat in dit onderzoek heel wel aan de assumptie van normaliteit, welke een vereiste is, voldaan kan zijn. Gezien het feit dat deze polychorische correlaties, of eventueel een andere correlatie, gecorrigeerd voor continuïteit (zie b.v. Don, 1981), meestal groter zijn dan de 'normale' product-moment-correlaties van de categoriescores is het zeer de vraag of de beweringen van de auteurs in dat geval zouden stand houden. (Een dergelijke analyse is tegenwoordig mogelijk met behulp van LISREL V, zoals ontwikkeld door Jöreskog en Sörbom (1982)). Wij betwijfelen de juistheid van de resultaten van Van Doorn, Saris & Lodge dan ook en vinden het jammer dat de auteurs nagelaten hebben ook deze, methodologisch zo belangrijke, variant in hun onderzoek te betrekken.

Literatuur

- Birnbaum, M.H. Using contextual effects to derive psychophysical scales. *Perception & Psychophysics*, 1974, 16, 89-96.
- Champney, H. & Marshall, H. Optimal refinement of the rating scale. *Journal of Applied Psychology*, 1939, 23, 323-331.
- Coombs, C.H. *A Theory of Data*. New York: John Wiley & Sons, 1964.
- Don, F.J.H. A Note on Sheppard's Corrections for Grouping and Maximum Likelihood Estimation. *Journal of Multivariate Analysis*, 1981, 11, 452-458.
- Gifi, A. Non-linear multivariate analysis. Leiden: Department of Datatheory, 1981.
- Jöreskog, K.G. & Sörbom, D. *LISREL, version V*. Chicago: International Educational Resources, 1982.
- MacKenzie, D. Statistical Theory and Social Interests: A Case-Study. *Social Studies of Science*, 1978, 8, 35-83.
- Orth, B. A Theoretical and Empirical Study of Scale Properties of Magnitude-Estimation and Category-Rating Scales. In: Bernd Wegener (ed.) *Social Attitudes and Psychophysical Measurement*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1982.
- Pearson, K. Mathematical contributions to the theory of evolution, VII. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London*, 1900, A, 195, 1-47.
- Yule, G.U. On the Methods of Measuring Association between Two Attributes. *Journal of the Royal Statistical Society*, 1912, 75, 579-652.