

Schoolkenmerken en leerlinggedrag

Enkele methodologische merkwaardigheden in het onderzoeksbedrijf

P. G. Swanborn

Dat men bij enquêtes een vrij groot aantal eenheden, aselekt getrokken nodig heeft om met succes multivariate analyse te kunnen doen, bijv. om de relatieve invloed van enkele onafhankelijke variabelen ten opzichte van een afhankelijke variabele vast te stellen, is genoegzaam bekend. Vaak ziet men echter een merkwaardige anomalie in het onderzoeksbedrijf als het gaat om de invloed van contextuele variabelen, dat zijn kenmerken van groepen waar de individuen, die het object van analyse vormen, deel van uitmaken. Niet zelden doet een onderzoeker uitspraken over de invloed van 'de cohesie van de groep' of van 'de gemiddelde motivatie van de groep' op het gedrag van de leden als zijn onderzoek niet meer dan 10 of 12 groepen omvatte die toevallig aanwezig waren (Blau 1960). Vaak wordt een conclusie getrokken over de invloed van de status van de buurt op bijv. de criminaliteit van de bewoners of het aspiratieniveau van de kinderen als slechts enkele tientallen buurten worden onderscheiden (bijv. Sewell en Armer 1966); frequent ontmoet men onderzoeksontwerpen van onderwijsresearchers waarin gepoogd wordt om bijv. de invloed van het curriculum op de schoolprestaties van kinderen te meten bij een totaal van niet meer dan bijv. 25 scholen die in de onderzochte regio aanwezig zijn.¹

Het lijkt wel alsof bij contextuele kenmerken andere methodologische richtlijnen ten aanzien van aselekte trekking en benodigd aantal eenheden gelden dan bij individuele kenmerken; nochtans lijkt hiervoor geen enkele grond te zijn.

Het is de bedoeling van deze notitie om dit punt nader te analyseren. De multiniveau-analyse (andere termen als contextuele analyse, structuur-analyse e.d. liggen in dezelfde sfeer) verheugt zich momenteel weer in een toenemende belangstelling sinds de eerste algemeen bekend geworden publicaties van Blau (1960) en Davis (1961). Vooral in de onderwijsresearch

bestaat er aandacht voor, misschien mede dankzij de omstandigheid dat hier zo keurig verschillende hiërarchische niveaus (leerling; vriendengroepje; klas; school) onderscheiden kunnen worden. We zullen daarom onze opmerkingen concretiseren met een voorbeeld uit de onderwijsresearch.

Stel dat een onderzoeker de invloed wil nagaan van een aantal schoolkenmerken op de vorderingen van de leerlingen. Een mogelijkheid daartoe doet zich uiteraard alleen voor als hij verschillende scholen in zijn onderzoek betreft. Veronderstel verder dat hij leerlingen van een 20-tal scholen onderzoekt, bij een leerlingenaantal per school van gemiddeld 150. Als onafhankelijke variabelen (schoolkenmerken) fungeren bijv. het aantal uren besteed aan sociale en creatieve vorming, gedichotomiseerd (variabele A) en het materieel budget van de scholen, gedichotomiseerd (variabele B). Als afhankelijke variabele (C) figureert een of andere vorderings- of prestatiemeting.

Men kan nu het design voor multivariate analyse op minstens twee manieren vormgeven.

Design I: de scholen vormen de analyse-eenheden ($N = 20$). De afhankelijke variabele C is de per school berekende mediaan van de individuele prestaties, of een andere centrummaat. Gewoonlijk wordt aan dit design de benaming 'analyse op groepsniveau' verbonden.

Design II: de leerlingen vormen de analyse-eenheden ($N = 3000$). Aan elke leerling i van school j wordt een score A_j en een score B_j toegekend, alsmede een score C_i voor de eigen prestatie. We noemen dit 'contextuele analyse'.

Is de onderzoeker alleen in tussen-school-verschillen geïnteresseerd, dan zal hij design I kiezen. Als de onderzoeker daarentegen tegelijk met de invloed van schoolkenmerken ook nog de invloed van individuele onafhankelijke variabelen wil vaststellen, dan zal hij in het algemeen design II kiezen.

De beide designs zijn in termen van doelstelling en te verklaren variantie niet onderling vergelijkbaar. Ten aanzien van de methodologische problemen die wij aan de orde stellen leiden zij echter niet tot een uiteenlopende benadering, zoals wij hieronder zullen laten zien.

Wij stellen nu eerst het probleem van het kleine aantal contexten aan de orde; vervolgens wordt de selectieproblematiek behandeld. Tenslotte gaan we in op enkele voor contextuele variabelen specifieke 'bezwarende omstandigheden'.

Het aantal te onderzoeken contexten

Sociologen zijn gewend aan de gedachte dat bij enquêtes enkele honderden individuen ondervraagd moeten worden om redelijk betrouwbare conclusies te kunnen trekken over bijv. de relatieve invloed van enkele onafhankelijke variabelen.

Hoe groter de steekproef, hoe groter de kans op een adequate representatie van de uni- en multivariate verdelingen in de populatie. Hoe groter de steekproef, des te groter ook de kans dat inconstante meethouten tegen elkaar gaan wegvallen. Hoe groter de steekproef tenslotte, des te groter zijn de mogelijkheden voor multivariate analyse. Er is geen reden te bedenken waarom niet eenzelfde maatstaf zou moeten worden gehanteerd als het gaat om het aantal contexten wanneer men onderzoek wil doen naar de relatieve invloed van diverse kenmerken van deze contexten op het gedrag van individuen. Dit betekent, in ons voorbeeld, dat men een steekproef van enkele honderden scholen zou moeten trekken.

Toch zijn er legio voorbeelden van contextuele analyses waarbij met 10 of 20 contexten wordt gewerkt.

Zelfs als de contexten — hier, concreet, de scholen — aselekt zouden zijn getrokken is bij een steekproef van een dergelijke geringe omvang de kans op vertekening erg groot. Bovendien, waar de meting van veel variabelen onderhavig is aan onder meer inconstante fouten, kan dit bij een klein aantal eenheden tot een grote invloed van foutieve classificering leiden.

Maar, zo kan men zich afvragen, wordt de situatie niet verbeterd als men van design I overgaat naar design II?

De onderzoeker houdt zichzelf voor de gek als hij denkt het gestelde probleem hiermee op te lossen. Weliswaar wordt het aantal eenheden in ons voorbeeld nu 150-maal vergroot, maar een afwijkende steekproef-samenstelling, al dan niet op grond van foutieve metingen, wordt ook 150-maal zo zwaar gewogen. Een foutief gemeten school komt op 150 foutief gemeten leerlingen te staan. Het 'opblazen' van de steekproef op deze wijze is misschien aantrekkelijk voor de onderzoeker die graag multivariate analyse bedrijft; vanuit een oogpunt van levering van betrouwbare onderzoeksresultaten is het een zinloze zaak.

Het is wellicht goed om nog wat verder te concretiseren voor welke moeilijkheden men bij multivariate analyse bij dergelijke kleine aantallen komt te staan.

We gaan daarbij uit van het hierboven geïntroduceerde geval met twee onafhankelijke variabelen bij een aantal van 20 scholen (design I).

De multivariate verdeling van een 2×2 -tabel wordt bepaald door de univariate verdelingen en de onderlinge correlatie. In het eenvoudigste geval is er geen correlatie tussen de onafhankelijken, en zijn de verdelingen niet scheef. Bij 20 scholen betekent dit dat er 5 scholen zijn van het AB-type; 5 van het $\bar{A}\bar{B}$ -type; 5 $A\bar{B}$ -scholen en nog eens 5 $\bar{A}B$ -scholen. Een antwoord op de vraag of de invloed van A op C groter of kleiner is dan de invloed van B op C berust nu op niet meer dan 10 eenheden, te weten die in de $\bar{A}B$ - en de $A\bar{B}$ -groep. Indien in de $\bar{A}B$ -groep de C-score hoger is dan in de $A\bar{B}$ -groep, constateert men dat de invloed van B op C groter is dan die van A op C, bijvoorbeeld.

Voor het geval A correleert met B, wordt het aantal $\bar{A}B$ en $A\bar{B}$ nog kleiner.² Naarmate de correlatie sterker wordt, neemt het multicollineariteitsprobleem in betekenis toe. Dit wil zeggen dat de schatting van de relatieve invloeden van A en B op C bemoeilijkt wordt doordat deze schatting op een zeer klein aantal eenheden komt te berusten (nl. de eenheden die ver van de regressielijnen van A op B en B op A liggen, als we, hoewel zeer gekunsteld in het geval van een 2×2 -tabel, van regressielijnen mogen spreken.³ Op zich is dat niet zo'n bezwaar als zonder meetfouten is gemeten of wanneer de steekproef niet van de populatie zou afwijken; maar dat zijn in de onderzoekspraktijk vrijwel uitgesloten gevallen.

[- Bij *scheefheid* van beide variabelen wordt het aantal 'uitbijters' waarop de schatting van de relatieve invloed wordt gebaseerd eveneens kleiner. Het blijkt dat voor deze eenvoudige 2×2 -situatie geldt dat uitspraken als bovenbedoeld slechts op maximaal de helft van het aantal eenheden, dat toch al erg beperkt was, worden gebaseerd. Zoals hierboven toegelicht, verandert er niets wezenlijks aan de situatie wat betreft steekproef- en meetfouten door overgang op design II.

In de praktijk zijn de moeilijkheden tengevolge van het gebruikelijke geringe aantal contexten nog groter dan in de simpele illustratie hierboven, doordat niet met twee, maar met een veel groter aantal contextuele kenmerken wordt gewerkt. Soms ontmoet men zelfs de bekende situatie dat er meer variabelen dan eenheden zijn. De lezer zal inzien dat mede door de dan onvermijdelijk optredende correlaties de invloeden van de diverse kenmerken niet uit elkaar zijn te pluizen.

Vuistregels voor de relatie tussen de betrouwbaarheid van uitspraken en het benodigde aantal eenheden,² en daarmee voor de zinvolheid van bepaalde multivariate operaties, worden door de inductieve statistiek gegeven; althans hier vindt men een zekere basis. Dit brengt ons echter op de vraag in hoeverre de inductieve statistiek van toepassing kan zijn, en daarmee op het probleem van de selectie.

De selectie van de te meten contexten

De scholen (algemeen: de contexten) worden vaak niet aselekt gekozen. Meestal neemt een onderzoeker die scholen welke zich in zijn regio aanbieden. Soms ook wordt opzettelijk geselecteerd door bijv. contrasterende contexten (contrasterend op een of enkele van de onafhankelijke variabelen) te kiezen. Tegelijkertijd blijft het meestal de bedoeling om parameters voor relatieve invloed voor een bepaald universum van scholen te schatten. Deze doelstelling is echter moeilijk te realiseren bij een selectie trekking. Deze kan immers zodanig zijn dat juist die scholen worden uitgekozen waarbij het verband tussen de onafhankelijke en de afhankelijke variabelen sterk is, of juist die deelverzameling van scholen waar dit ontbreekt of erg zwak is.

In ons voorbeeld werden contextuele variabelen gebruikt. De lezer zal echter constateren dat de redenering invariant is met betrekking tot het gebruik van wat voor variabelen dan ook. Het is de gebruikelijke redenering waarmee aselechte trekking en een behoorlijk aantal eenheden in het gangbare enquête-onderzoek worden beargumenteerd. De conclusie luidt voorlopig, dat wil men de invloed van schoolvariabelen met een acceptabele mate van betrouwbaarheid (en generaliseerbaarheid) vaststellen, dat men dan minstens vergelijkbare criteria voor trekking en aantal moet hanteren als bij analyse op individueel niveau; dan wel, dat men moet grijpen naar designs van geheel andere aard (experimenten, simulatie).

Specifieke 'bezwarende omstandigheden'

De 20 scholen in ons voorbeeld vormen evenzovele specifieke, complexe, subculturen, die op een groot aantal kenmerken van elkaar verschillen. Deze kenmerken zijn over het algemeen veel minder *bekend*, veel frequenter *gecorreleerd* (door een geringere logische afstand); veel minder op hun invloed *getoetst*; en, voorzover onderzoeksresultaten bekend zijn, van veel geringer *invloed* dan kenmerken van individuen.

De meting van contextuele kenmerken staat in het algemeen nog in de kinderschoenen, terwijl de operationalisering van vele individuele kenmerken langzamerhand een relatief vaste vorm heeft verkregen. De correlaties tussen individuele (meestal als onafhankelijk optredende) kenmerken als beroep vader, kerkelijke gezindte, opleiding e.d. zijn redelijk bekend; we weten wat we in de analyse constant moeten houden, en we

weten wat we met een bepaalde operationalisatie meten. Als we daarentegen op een bepaalde schoolvariabele het etiket 'variabele A' plakken, dan kan de verkregen indeling tegelijkertijd mogelijk een indeling volgens de onbekende, althans ongemeten variabelen 'B' of 'C' weerspiegelen. De onderzoeker die aan variabele A een bepaalde invloed toekent, heeft nauwelijks verweer tegen critici die gemakkelijk kunnen beweren dat niet A, maar in feite B of C de bepalende factor is, en dat het etiket dus moet worden vervangen. Men vergelijkte voor een instructief en komisch verhaal hierover Hauser 1970. In 13 pagina's doet deze onderzoeker verslag van een uiterst serieuze analyse rondom de contextuele variabele 'sex-ratio' (kwantitatieve verhouding jongens/meisjes) op een aantal (11!) scholen en geeft hij een theoretische interpretatie van het gevonden verband met aspiratieniveau; in de daarop volgende drie pagina's breekt hij het voorgaande tot op de grond toe af. Zo is het, in ons voorbeeld, zeer wel denkbaar dat de scholen in de 'plus'-groep voor sociale en creatieve vorming juist de scholen zijn in intellectuele en hoge SES-buurtten, terwijl de 'min'-scholen juist de scholen zijn in lage SES-buurtten. In dit voorbeeld zijn beide variabelen nog redelijk gemakkelijk te meten; vaststelling van de relatieve invloed van elk stuit echter op de genoemde bezwaren bij de gebruikelijke contextuele analyse-onderzoekingen. Heel wat moeilijker wordt het wanneer A, B, C etc. variabelen zijn die 'gesyntetiseerde' opinies, attitudes, onderlinge relaties betreffen. Terwijl in een puur individueel niveau-analyse de opstelling van een causaal model waarin opinies, attitudes e.d. onderling worden geordend al een toer is, lijkt iets dergelijks vooralsnog praktisch onmogelijk bij contextuele analyse.

Hauser (Hauser 1970) geeft twee alternatieve wegen aan om uit deze moeilijkheden te raken. In de eerste plaats adviseert hij om niet zo maar allerlei contextuele variabelen 'te bedenken' en de contexten hierop in te delen, maar pas dan een contextuele variabele te gebruiken wanneer de theoretische interpretatie van het verwachte verband gegeven kan worden. Meestal brengt dit met zich mee dat ook andere variabelen, tussenschakels in het causale proces, gemeten moeten worden; en dit laat weer de mogelijkheid van toetsing van een bepaalde interpretatie toe. In de tweede plaats oppert hij de mogelijkheid om de 'complexe subculturen' waarover wij hierboven spraken ongegroepeerd te laten maar 'ieder voor zich te laten spreken' via covariantie-analyse (vgl. Schuessler 1969). Weliswaar heeft men dan nog geen interpretatie voorhanden, maar men vermijdt irrelevante groepering en geeft elke school de kans om in de analyse het eigen karakter te laten uitkomen. Overigens levert het beschouwen van de afzonderlijke contexten als evenzovele categorieën van een nominale varia-

bele in een normale regressie-analyse al voordelen op boven groepering van contexten op een meer of minder fictieve variabele; Hauser vond op deze wijze aanzienlijk meer variantie verklaard dan via de sex-ratio groepering.

Bovenstaande, noodzakelijkerwijs vluchtige impressies over de specifieke aard van contextuele kenmerken, lijken de eerder geformuleerde conclusie over de noodzaak van de gebruikelijke methodologische criteria bij contextuele analyse in hoge mate te versterken.

In deze korte bijdrage hebben wij slechts één aspect van de multiniveau-analyse aan de orde willen stellen, te weten de wijze van selectie en het aantal te kiezen contexten in een natuurlijke situatie. Veel andere problemen die aan deze vorm van analyse verbonden zijn (bijv. de multicollineariteit van individuele en contextuele variabele als deze inhoudelijk identiek zijn; de invloed van de gebruikelijke vergroving, i.c. dichotomisering; de interpretatie van contextuele invloeden via 'tussengelegen' relationele kenmerken e.d.) zijn buiten beschouwing gelaten of slechts terloops genoemd.

De conclusie luidde, dat wanneer men in een natuurlijke situatie de invloed van een aantal contextuele kenmerken op individueel gedrag wil meten (al dan niet samen met de meting van de invloed van individuele variabelen), men dan de contexten aselekt moet trekken en in voldoende groot getal om multivariate analyse zinvol te maken. Over de feitelijke grootte van N kan uiteraard, gezien de afhankelijkheid van probleemstelling en object, in het algemeen weinig gezegd worden; het noemen van een minimale N van 100 is echter wellicht nuttig. Gezien de bezwarende praktische implicaties van een dergelijke eis doet zich de vraag voor, of, althans in het concrete veld van de onderwijsresearch, het kiezen van een experimenteel design, simulatie of case-studies niet meer voor de hand liggen dan voortzetting van de enquête-praktijk.

Noten

1. Deze notitie werd geïnspireerd door het ontwerp voor een vervolgonderzoek van het Sociologisch Instituut te Groningen naar de invloed van schoolvariabelen (Dr. G. W. Meijnen).
2. Bij een 'negatieve' correlatie nemen AB en $\bar{A}\bar{B}$ relatief af. De regressielijnen — als we daarvan mogen spreken in een 2×2 -tabel — komen dan door $\bar{A}\bar{B}$ en $A\bar{B}$ te lopen, en AB en $\bar{A}B$ moeten als 'uitbijters' beschouwd worden, waarop

- de schatting van de relatieve invloed van A en B op C komt te berusten.
3. Introductie van het begrip 'regressielijnen' biedt het voordeel dat ons voorbeeld hierdoor gekoppeld kan worden aan de gebruikelijke interpretatie van het multi-collineariteitsprobleem.

Literatuur

- Blau, P. M., Structural Effects, *American Sociological Review* 25, 1960, 178-193.
- Davis, J. A. e.a., A Technique for Analyzing the Effects of Group Composition, *American Sociological Review* 26, 1961, 215-225.
- Hauser, R. M., Context and Consex: A Cautionary Tale, *American Journal of Sociology* 75, 1970, 645-664.
- Schuessler, K., Covariance Analysis in Sociological Research, in E. F. Borgatta (ed.), *Sociological Methodology 1969*, Jossey-Bass 1969, 219-244.
- Sewell, W. H. en J. Michael Armer, Neighbourhood Context and College Plans, *American Sociological Review* 31, 1966, 159-168.