

Multivariate analyse

Een inleiding voor criminologen en andere sociale wetenschappers

LIEVEN PAUWELS*

C.C.J.H. Bijleveld, J.J.F. Commandeur (2008). Den Haag: Boom Juridische Uitgevers.

Dit boek bevat een introductie tot enkele grote klassieke multivariate analysetechnieken die volgens de auteurs tot de standaard toolkit van iedere criminoloog zouden moeten behoren en trekt precies daarom mijn aandacht: is dit boek werkelijk bedoeld is voor een bijzonder publiek, met name criminologen? Als lesgever in het vakgebied criminologie vind ik het belangrijk dat het methodeonderwijs een specifieke vakgebonden invulling krijgt via het geven van domeinspecifieke voorbeelden, terwijl de multivariate methoden eerder neutraal zijn en door diverse sociale wetenschappers gebruikt worden. Na lectuur wordt duidelijk dat dit boek ongetwijfeld in hoofdzaak door criminologen (in spe) en misschien minder door andere sociale wetenschappers zal gebruikt worden. Statistiek in de praktijk aanleren kan best vanuit een vakspecifieke invalshoek, al was het alleen maar om studenten de expliciete link tussen statistiek en hun inhoudelijke discipline te tonen. Niet alleen is dit boek doorspekt met voorbeelden uit de empirische onderzoekspraktijk van de criminoloog-vorser, ook is dit boek gericht op het minimaliseren van de mathematische achtergrond van statistische analyseprocedures. Hier herken ik spontaan de modale criminologiestudent, bij wie de schrik voor mathematische formules er toch relatief sterk in zit, zeker in vergelijking met studenten uit onze zusterdisciplines zoals psychologie en sociologie. Ik ben geleid door de nobele doelstelling van de auteurs dat zij met hun boek studenten uit de criminologie dichter bij de toegepaste multivariate statistiek willen brengen. In de eerste plaats slagen de auteurs erin de mathematische achtergrond tot een minimum te beperken, maar tevens slagen zij er in abstracte begrippen iets meer levendig voor te stellen dan statistische handboeken die de mathematische achtergrond beklemtonen. Sommige wetenschappers vinden het waarschijnlijk ongehoord om de mathematische complexiteit te minimaliseren in een handboek multivariate analyse. Terecht bestaat de vrees voor de opmars van een nieuw fenomeen, met name het ontstaan van een generatie van criminologen die enkel aan de hand van statistische verwerkingspakketten, syntax of de eenvoudige command interface, complexe technieken uitvoert in een mum van tijd, zonder notie te hebben van wat er achter de schermen gebeurt. Deze generatie zou dan technieken gebruiken zonder zicht te hebben op statistische assumpties en vooral zonder zicht te hebben op de consequenties van schendingen van statistische assumpties. Deze vrees is onterecht, want de auteurs wringen zich behoorlijk in bochten om abstracte statistische parameters en uitkomsten in gewone mensentaal om te zetten. Het is precies hier dat het boek een hiaat opvult in het criminologieonderwijs. Kwantitatieve criminologische toepassingen zijn ofwel heel rudimentair en beperkt tot bivariate kruistabellen ofwel heel complex (structurele vergelijkingsmodellen en multilevel-toepassingen). Criminologen die een behoorlijke intermediaire kennis van statistiek hebben, vond je lange tijd nauwelijks, tenzij deze ook een bijko-

* Prof. Dr. Lieven Pauwels is verbonden aan de Vakgroep Strafrecht en Criminologie van de Universiteit Gent en is co-directeur van de onderzoeksgroep Sociale VeiligheidsAnalyse.

mende opleiding in een andere sociaal-wetenschappelijke discipline hadden gevolgd, zoals daar opnieuw zijn de psychologie of sociologie. Met dit boek hebben criminologen die daar de behoefte toe hebben een goed gereedschap in handen om zich in te wijden in de wereld van de multivariate analyse. Deze eerste uitgave draagt zeker bij tot het bereiken van een intermediair niveau toegepaste statistiek, en dan is het een kunst om sommige details weg te laten en zoals de auteurs doen, te verwijzen naar de meer gespecialiseerde literatuur en tegelijkertijd toch trachten uit te leggen onder welke condities het verstandig is voor een logistische regressie te kiezen (i.e. als je afhankelijke variabele categorisch is).

Inhoudelijk is dit boek ook sterk bepaald door de (Nederlandse) belangstelling voor predictieonderzoek en onderzoek naar recidive van delictplegers. Technieken zoals de survival analyse en Cox-regressie verdienen een prominente plaats in het handboek, terwijl andere technieken dan weer in mindere mate aan bod komen.

Het boek start met een algemene inleiding waar de auteurs hun uitgangspunten geven tegenover statistiek en waar zij het vakgebiedspecifieke uiteenzetten. Dit inleidende hoofdstuk herneemt de basisprincipes van de univariate en bivariate statistiek. De essentiële idee dat de keuze voor een analysetechniek bepaald wordt door zowel het meetniveau van variabelen en de probleemstelling (dependent versus niet-dependent) wordt hier in een notendop uiteengezet. Het tweede hoofdstuk behandelt methodologische aspecten van multivariaat (survey) onderzoek en onderzoek op secundaire data, met name door de operationalisatie van begrippen uiteen te zetten. Het derde hoofdstuk behandelt de belangrijkste univariate en bivariate analysetechnieken in vogelvlucht. Er wordt een onderscheid gemaakt tussen symmetrische en asymmetrische technieken. Een vierde hoofdstuk legt de verschillen en gelijkenissen tussen technieken uit. Hoewel deze eerste hoofdstukken de basisprincipes van de univariate en bivariate statistiek behandelen, vervangen deze hoofdstukken geen grondige inleiding tot de univariate en bivariate statistiek. De auteurs onderstrepen dit ook. Potentiële geïnteresseerden kunnen hier maar beter rekening mee houden. Lezers wiens basiskennis statistiek onder een grote laag stof zit, kunnen beter eerst een basisboek toegepaste statistiek raadplegen. Vooraleer echt aan de slag te gaan met de systematische behandeling van de multivariate analysetechnieken worden enkele praktische kwesties van datamanagement behandeld in hoofdstuk 5. Dit is niet onbelangrijk, want de leerling-onderzoeker dient niet enkel gewapend te zijn met technieken, maar dient ook de vinger aan de pols van de data te houden vooraleer de multivariate analyses aan te vangen. In dit opzicht is dit boek geslaagd: er wordt opnieuw een brug gelegd tussen kwantitatieve methodologie en dan vooral door via voorbeelden de praktische keuzes die onderzoekers moeten hebben te behandelen: hoe gaat men nu operationele variabelen creëren, hoe komen we van eenvoudige enquêtevragen tot geoperationaliseerde samengestelde begrippen? De procedures hercoderen en transformeren van variabelen worden behandeld. Tevens wordt stilgestaan bij de vraag wat te doen bij missing data, ... De auteurs behandelen heel kort de "missing values analysis" en het "oplappen" van datasets door ontbrekende waarden te schatten op basis van gekende waarden. De klassieke imputatietechnieken dus. Dit stukje is in het handboek wel heel kort uitgelegd. Gelukkig worden bij dit boek ook oefendatasets en bijkomende uitleg via SPSS-output gegeven. Alle multivariate analysetechnieken kunnen worden ingestudeerd aan de hand van deze voorbeeldbestanden en uitgewerkte SPSS-voorbeelden die de studenten ter beschikking worden gesteld. Deze oefenstof omvat reële datasets en bijkomende uitleg bij de syntax-commando's. Deze datasets en de bijbehorende SPSS-aansturing van de analyses zijn gratis te downloaden op de URL van Boom Uitgevers Den

Haag onder de knop "Oefenstof", zodat de gebruiker de voorbeelden zelf stap voor stap kan narekenen en de techniek in de vingers kan krijgen.¹ Deze bijkomende informatie is echt noodzakelijk wil je als lezer het onderste uit de kan halen. De bijkomende SPSS-output bevat namelijk een stap-voor-stap benadering, zodat studenten een zeer goede aansturing krijgen. Hiermee is het boek ook voor zelfstudie geschikt.

Na het praktische hoofdstuk 5 komen drie grote afzonderlijke hoofdstukken die de kern van dit boek uitmaken: de meervoudige regressie-analyse (hoofdstuk 6), de overlevingsduur-analyse (hoofdstuk 7), de principale componenten-analyse (hoofdstuk 8) en exploratieve schaaltechnieken (hoofdstuk 9) komen aan bod. Het hoofdstuk multivariate regressie-analyse gaat behoorlijk in op de grondvoorwaarden voor het uitvoeren van regressie-analyses en behandelt lapmiddelen wanneer regressie-assumpties geschonden zijn, zoals de logaritmische transformatie van variabelen. Aan de hand van voorbeeld-output wordt duidelijk gemaakt hoe je de schending van assumpties zoals lineariteit en homoscedasticiteit kan nagaan. Hieruit mag blijken dat -hoewel de insteek er één is die de mathematische achtergrond beperkt houdt en de praktische toepassing laat primeren- op sommige plaatsen toch wel de nodige zorg wordt besteed aan tal van assumpties waarvan reeds is aangetoond dat de ernstige schending ervan impact kan hebben op de resultaten van statistische analyses. Na een uitvoerige uitleg over de lineaire regressie-analyse wordt ietwat meer summier stilgestaan bij de logistische regressie. Ik vind het jammer dat hier geen SPSS-output en een uitgewerkt voorbeeld wordt besproken, want logistische modellen worden niet enkel heel frequent gebruikt, bij (criminologie)studenten roepen deze modellen vaak vragen op, vooral met betrekking tot model-fit en effectparameters. Een volgende hoofdstuk (hoofdstuk 7) gaat in detail in op de survival-analyse, waarbij het begrip "de overlevingsduur" de basis vormt voor een uiteenzetting over de specifieke problematiek van de analyse van gecensureerde data. De auteurs nemen zich de nodige moeite en tijd om deze techniek uiteen te zetten en eindigen bij de Cox-regressie met als voorbeeld de recidive van delictplegers. In de hedendaagse (Nederlandse) criminologie krijgt risicotaxatie een steeds grotere plaats, dus deze techniek verdient zeker een afzonderlijk hoofdstuk in dit boek. Hoofdstuk 8 besteedt aandacht aan de principale componentanalyse en de schaalconstructie (gesommeerde Likertschaal). De auteurs doen veel moeite om abstracte begrippen zoals eigenwaarden en rotatie in vectorruimtes uiteen te zetten aan de lezer. De voorbeelden komen uit een survey naar de perceptie van strafdoelen en daarmee zijn ze dus ook mooi op de praktijk geschoold.

In hoofdstuk 9 behandelen de auteurs een reeks exploratieve technieken die met name in bepaalde statistische scholen populair zijn, en elders weer minder: het gaat om exploratieve schaaltechnieken voor nominale variabelen zoals homals (categorische principale component-analyse) en de non-lineaire principale component analyse (princals). Toegegeven, de uiteenzetting is helder, maar hier stel ik mij de vraag of zulke technieken tot een "basispakket" multivariate analyse voor criminologen dienen te behoren. Die mening deel ik niet helemaal met de auteurs. In dat opzicht vind ik het bijvoorbeeld jammer dat de steeds belangrijker wordende veralgemeende lineaire modellen niet afzonderlijk wordt besproken: in de kwantitatieve criminologie worden vandaag de dag heel vaak (naar mijn mening vaker dan de exploratieve technieken voor nominale variabelen) Poisson-gebaseerde regressiemodellen gehanteerd voor de multivariate regressie-analyse van countdata, zoals het aantal delicten waarvan men slachtoffer werd en de criminaliteitsgraad van een gebied. Over details en de invulling

1 <http://www.boomuitgeversdenhaag.nl> (04-12-2008)

van wat in een basispakket dient te behoren hoeft en kan men het zelden helemaal eens zijn. Maar als dit boek is bedoeld voor criminologen, dan meen ik dat de auteurs het in overweging kunnen nemen om in een volgende versie eventueel ook deze vaak gebruikte techniek mee te nemen. Hoofdstuk 10 bevat een inhoudelijk introductie tot de meer complexe vormen van kwantitatieve analysetechnieken die ook in de criminologie worden gebruikt, zoals daar zijn de structurele vergelijkingsmodellen, de trajectory-analysis (een variant van de latente klasse-analyse) en de multilevel analyse van gegevens met geneste structuur. Het is goed dat deze technieken vermeld worden. Men kan deze niet meer weg denken uit de hedendaagse kwantitatieve criminologie en daarom moet stilgestaan worden bij het potentieel van deze technieken.

Een didactisch sterk punt in dit boek is de systematiek die er in zit. Op systematische wijze wordt in elk hoofdstuk aandacht besteed aan het evalueren van de effecten van parameters en de fit van het statistisch model, zodat de lezer steeds goed bij de hand wordt genomen bij het interpreteren van diens eerste analyses. Leuk is ook dat elk hoofdstuk afsluit met een reeks suggesties aan leergierige studenten onder de noemer "verder lezen". De auteurs leggen hiermee de brug naar het meer geavanceerde en abstractere werk.

Samengevat: dit handboek slaat een voortreffelijke brug tussen de mathematische achtergrond van de wereld van de multivariate analyse en de praktijk van het toegepast onderzoek binnen de criminologie. De voorbeelddatasets zijn een absolute must wil men dit handboek gebruiken om een aantal multivariate analysetechnieken op te frissen. Statistiek in de criminologische praktijk toegepast krijgen op een ruimere schaal is een behoorlijke uitdaging en de auteurs zijn hier mijns inzien zeer goed in geslaagd. Catrien Bijleveld is hoogleraar aan de Vrije Universiteit van Amsterdam en senior onderzoeker aan het Nederlands Studiecentrum voor Criminaliteit en Rechtshandhaving en werkte bij het schrijven van dit boek samen Jacques Commandeur, senior onderzoeker bij de Stichting Wetenschappelijk Onderzoek Verkeersveiligheid.