

De relatie tussen achtergrondkenmerken en schoolloopbaanvariabelen binnen COROP-gebieden nader onderzocht

A.C. Meester, J. de Leeuw

Vakgroep Datatheorie FSW/RUL

SAMENVATTING

In Meester & De Leeuw (1984) is aan de hand van gegevens uit het CBS-SMVO-cohort de relatie tussen sociale achtergrond en geslacht enerzijds en de schoolloopbaan anderzijds in 40 COROP-gebieden beschreven. Tevens werd geprobeerd demografische kenmerken van COROP-gebieden aan de gegevens uit het SMVO-cohort te relateren. Met behulp van loglineaire analyse werd de hypothese onderzocht dat regionale herkomst en schoolloopbaan onafhankelijk zijn gegeven het sociaal milieu waaruit de leerlingen afkomstig zijn. Het resultaat suggereert dat regionale verschillen in schoolloopbaan gering zijn ten opzichte van verschillen in sociale herkomst; een uitkomst die bevestigd wordt door partiële canonische correlatie analyse. Aanvullend worden deze resultaten door middel van variantieanalyse nader genuanceerd.

INLEIDING

In Meester & De Leeuw (1984) werd de relatie tussen sociaal milieu en geslacht enerzijds en de schoolloopbaan in het voortgezet onderwijs anderzijds in verschillende regio's onderzocht en werd getracht een verband te leggen tussen kenmerken op regionaal niveau en deelname aan en schoolloopbaan in het voortgezet onderwijs. Het onderzoek werd door de Stichting Voor Onderzoek van het Onderwijs (SVO-project 1085) gesubsidieerd.

Dit artikel beoogt een aanvulling te geven op de analyses die m.b.t. het eerste gedeelte van de vraagstelling werden uitgevoerd. De relatie tussen schoolloopbaan- en regionale kenmerken blijft verder buiten beschouwing. Van belang is alleen te vermelden dat de beschikbaarheid van gegevens op het niveau van z.g. COROP-gebieden* een van de redenen was om de keuze op dit regionale niveau te bepalen.

In de aangehaalde publikatie wordt nader toegelicht hoe technische en organisatorische problemen verhinderden, een regionaal niveau te kiezen dat dicht bij de individuele gegevens lag. In dit verband worden ook meer algemene belemeringen van het zg. 'multi-level' onderzoek belicht (Meester & De Leeuw, 1984, 1.2 en 2.1).

Voor de exploratie van het eerste gedeelte van de vraagstelling werden kruistabellen samengesteld van geslacht, sociaal milieu, COROP-gebiedsnummer en elk van de schoolloopbaanvariabelen afzonderlijk. Deze tabellen werden loglineair geanalyseerd. Een soortgelijke procedure werd in Meester & De Leeuw (1983) toegepast, maar dan met de score op een intelligentietest in plaats van de regionale aanduiding. Geslacht en sociaal milieu werden daarbij in beide gevallen tot één interactieve variabele samengevoegd.

In de eerste publikatie werd uiteengezet dat het loglineaire model met alle 2e-orde interacties daarin opgenomen, zo geïnterpreteerd kan worden dat de relatie tussen schoolloopbaan en sociale achtergrond in alle COROP-gebieden dezelfde is. Dit model heeft voor elk van de schoolloopbaanvariabelen een uitstekende passing. Maar ook het (sub)model met alleen de interacties schoolloopbaan/milieu-sexe en COROP/milieu-sexe, dat opgevat kan worden als toetsing van van de hypothese dat COROP en schoolloopbaan onafhankelijk zijn, gegeven sociaal milieu en geslacht, heeft steeds een alleszins acceptabele fit.

Uit de analyses trokken we de conclusie dat voor het verloop van de schoolloopbaan het ouderlijk milieu van veel groter belang is dan de regionale herkomst. Dit resultaat werd bovendien bevestigd door de uitkomsten van een correlatie-analyse.

Over het gehele bestand genomen is de canonische correlatie tussen milieu-sexe en een gewogen combinatie van de schoolloopbaanvariabelen .352. Als nu de 'invloed' van regio uit deze associatie wordt weggezuiverd verandert de correlatie nauwelijks (.351). We weten nu dat het ouderlijk milieu veel meer geassocieerd is met de individuele loopbaan in het voortgezet onderwijs dan de regionale herkomst, maar niet noevél meer. Een mogelijke indruk daarvan zouden we kunnen krijgen door in de canonische correlatie de 'omgekeerde' procedure toe te passen en de invloed van het ouderlijk milieu verwijderen uit de correlatie tussen schoolloopbaan en COROP-gebied. COROP-gebied beschouwen we echter als een nominale variabele die niet goed lineair is te transformeren. De correlatie met een dergelijke variabele is daarom niet goed te interpreteren.

Een andere mogelijkheid zou zijn de λ -parameters uit de loglineaire analyses te bestuderen. Interpretatie van deze parameters is echter niet zonder problemen. Alle variabelen loen in loglineaire analyse symmetrisch mee en de λ 's zijn redefinieerd op een schaal die een logaritmische transformatie van verwachte celfrequenties onder een bepaald (sub-) model is. In variantie-analyse worden de variabelen wel asym-

metrisch behandeld en wordt een interpreteerbare onderliggende schaal voor de afhankelijke variabele verondersteld. In onze benadering kan deze schaal bovendien volgens één of ander criterium worden geoptimaliseerd. Om de resultaten uit de eerdere analyses te nuanceren, kiezen we dus voor variantie-analyse met optimale schaling van de afhankelijke variabelen.

DE GEGEVENS

De analyses werden uitgevoerd op de gegevens van het zg. SMVO-cohort. Dit bestand omvat een steekproef van ruim 37.000 leerlingen die in 1977 het lager onderwijs verliet. Door het CBS wordt jaarlijks bijgehouden waar deze leerlingen zich in het voortgezet onderwijs (en na het verlaten daarvan op de arbeidsmarkt) bevinden. Bij de start van het cohort werden ook achtergrondgegevens over deze leerlingen verzameld. We beschrijven het bestand en de in de analyse betrokken variabelen slechts summier, meer informatie is te vinden in CBS (1982) en Meester & De Leeuw (1983, 1984).

In de hier en ook bij de eerder gerapporteerde (loglineaire) analyses wordt de schoolloopbaan beschreven door vier variabelen: het advies van de onderwijzer voor het voortgezet onderwijs (ADV), de schoolkeuze v.o. (KEUS), het na vijf jaar bereikt onderwijsniveau (EIN) en de scores op een verkorte versie van de CITO-toets (PRE). ADV en KEUS hebben als categorieën LBO, MAVO, HAVO en VWO. EIN onderscheidt naast deze hoofdcategorieën ook nog de voortijdige schoolverlaters ('drop-outs') en de leerlingen in het MBO en HBO. Om in de tabellen lege cellen te vermijden werden de oorspronkelijke scores op PRE in 3 klassen (1-32, 33-50 en 51-70 items goed beantwoord) gehercodeerd.

Het ouderlijk milieu (BVA) wordt, voornamelijk op basis van het beroep van de vader, weergegeven in 6 categorieën (arbeiders, boeren en tuinders, kleine zelfstandigen, lagere employeés, middelbare employeés en hogere employeés en vrije beroepen). De categorisering van geslacht (SEX) spreekt voor zichzelf. De regionale variabele (COROP) onderscheidt 40 gebiedseenheden. Na enige selecties bestaat het bestand waarop de analyses worden uitgevoerd uit 29.290 leerlingen.

ANALYSE

Voor de analyse vormen de 4-dimensionale kruistabellen van SEX, BVA, COROP en elk van de schoolloopbaanvariabelen (ADV, KEUS, PRE en EIN) afzonderlijk het uitgangspunt. Dit zijn dus dezelfde tabellen die in Meester & De Leeuw (1984) loglineair werden geanalyseerd. We passen variantie-analyse toe met optimale schaling van de schoolloopbaanvariabelen. De categorieën van deze variabelen worden daarbij nominaal opgevat. Deze procedure is vergelijkbaar met het programma ADDALS (De Leeuw, Young & Takane, 1976).

De kwantificaties van ADV, KEUS, EIN en PRE worden zodanig gekozen dat de variantie binnen de cellen van SEX-BVA-COROP zo klein mogelijk is en de variantie tussen cellen dientengevolge zo groot mogelijk. Met andere woorden de regionale groepen, uitgesplitst naar sociaal milieu en geslacht, worden gemiddeld zo homogeen mogelijk gemaakt in hun verdeling op de schoolloopbaanvariabelen.

Daarna is de gewogen kwadratensom van de celgemiddelden die gelijk is aan $N \times$ de homogeniteit, opgesplitst in de afzonderlijke bijdragen van SEX, BVA, COROP en de interacties daartussen. Homogeniteit is daarmee gedefinieerd als het gedeelte van de variantie dat door geslacht en sociale en regionale herkomst beschreven wordt. Op die manier kunnen dus de relatieve effecten van de achtergrondvariabelen op ADV, KEUS, EIN en PRE worden geschat.

RESULTATEN

We geven eerst de kwantificaties van de schoolloopbaanvariabelen in tabel 1. In de kop van de tabel staat achter elke variabele tussen haakjes de homogeniteit.

De variabelen blijken zeer regelmatig te worden geschaald. De kwantificaties voor LBO, MAVO, HAVO en VWO liggen voor ADV, KEUS en EIN vrijwel op één lijn, als ze tegen de oorspronkelijke categorienummers zouden worden uitgezet, en dat geldt ook voor de 3 categorieën van PRE. De laatstgenoemde variabele heeft binnen de cellen van SEX x BVA x COROP de grootste variantie (c.q. de kleinste homogeniteit; $\lambda = .1086$). In de overige tabellen bepalen SEX x BVA x COROP 13 à 14% van de variantie in, respectievelijk EIN, KEUS en ADV.

Tabel 1 Kwantificaties schoolloopbaanvariabelen. Tussen haakjes de homogeniteit (λ)

	ADV (0.1418)	KEUS (0.1370)	PRE (0.1086)	EIN (0.1295)
LBO	-1.41	LBO -1.32	1-32 items goed 1.48	drop-outs 0.75
MAVO	-0.24	MAVO -0.20	33-50 items goed 0.07	LBO 1.42
HAVO	0.81	HAVO 0.67	51-70 items goed -1.34	MAVO 0.47
VWO	1.57	VWO 1.36		HAVO -0.75
				VWO -1.64
				MBO 0.63
				HBO 0.00

Een overzicht van de effecten van SEX, BVA en COROP-gebied en hun onderlinge interacties, per schoolloopbaanvariabele, staat in tabel 2. Voor de beoordeling van het relatieve belang van de effecten is de gemiddelde kwadratensom de beste maat, omdat daarin voor het aantal te schatten parameters is gecorrigeerd.

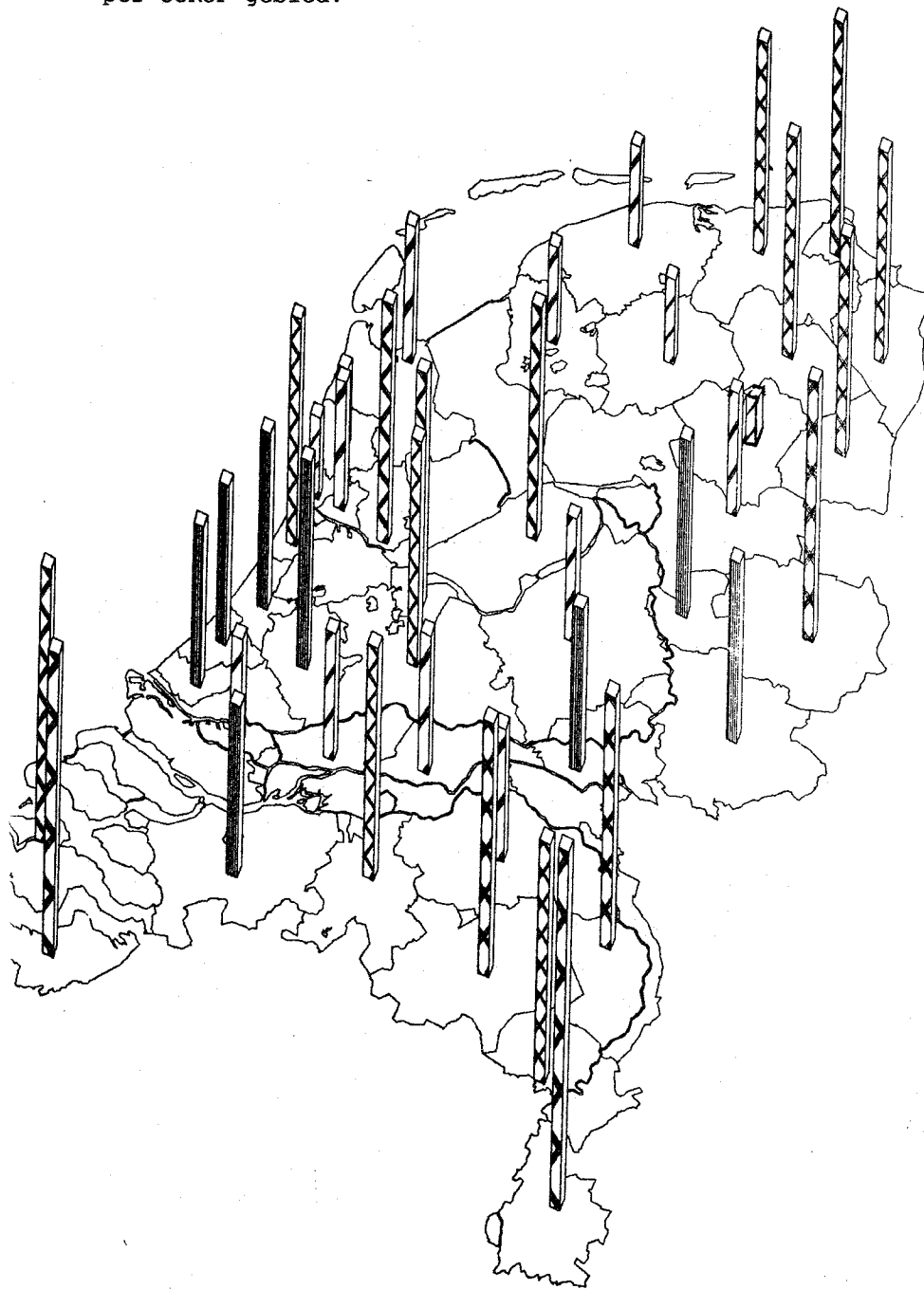
De interacties stellen weinig voor. Ze beschrijven tesamen ten hoogste 1 à 2% van de totale variantie (tussen- plus binnen-cellen variantie) en geen van deze effecten is statistisch significant. We bepalen ons daarom tot de hoofdeffecten.

Tabel 2 Kwadratensommen (SSQ), vrijheidsgraden (DF), percentage variantie (PERC=SSQ/SOM SSQ), de gemiddelde kwadratensommen (MSQ=SSQ/DF) voor hoofdeffecten en interacties (S=geslacht, M=sociaal milieu, R=COROP-gebied)

effect	DF	ADV			KEUS		
		SSQ	PERC	MSQ	SSQ	PERC	MSQ
S	1	2.70	0	2.70	0.22	0	0.22
M	5	2853.50	71	570.70	3112.89	78	622.58
R	39	603.03	15	15.46	384.85	10	9.87
SM	5	18.47	0	3.69	15.54	0	3.11
SR	39	117.92	3	3.02	87.40	2	2.24
MR	195	272.64	7	1.40	249.88	6	1.28
SMR	195	168.11	4	0.86	163.06	4	0.84
totaal	479	4036.36	100	8.43	4013.84	100	8.38
		PRE			EIN		
S	1	18.92	1	18.92	21.65	1	21.65
M	5	1376.64	51	275.33	2955.33	78	591.06
R	39	755.47	28	19.37	342.57	9	8.78
SM	5	11.23	0	2.25	9.38	0	1.87
SR	39	69.04	3	1.77	64.57	2	1.66
MR	195	284.94	11	1.46	264.94	7	1.36
SMR	195	177.16	7	0.91	132.98	4	0.68
totaal	479	2693.40	100	5.62	3791.42	100	7.92

Geslacht hangt weinig samen met ADV en KEUS, meer met PRE en EIN. Afgezien van het effect dat SEX op het na vijf jaar bereikte niveau in het voortgezet onderwijs heeft (meisjes behalen een gemiddeld iets hoger onderwijsniveau) vertonen de verdelingen van de effecten van KEUS en EIN zeer veel overeenkomst.

jur 1 Gemiddeld bereikt opleidingsniveau van leerlingen in
het CBS-SMVO-cohort na 5 jaar voortgezet onderwijs,
per COROP-gebied.



Het hoofdeffect van het milieu van herkomst is voor alle schoolloopbaanvariabelen verreweg het belangrijkste. Op de schoolprestaties is het effect van BVA het geringst en beschrijft ongeveer 5% van de totale variantie. Bij de andere schoolvariabelen is dat ongeveer 10%. Ook COROP heeft op alle schoolloopbaanvariabelen een significant effect. Het duidelijkst is dat bij ADV en PRE, iets geringer is de invloed op KEUS en EIN. COROP neemt 1-3% van de totale variantie voor zijn rekening.

Ter illustratie zijn in figuur 1 de regionale verschillen in het na 5 jaar bereikte onderwijsniveau weergegeven. Dit zijn dus de gewogen gemiddelden van EIN voor alle 40 COROP-gebieden (uitgedrukt als afwijkingen van het over-all-gemiddelde; omdat de hoogste categorieën van EIN het laagst getransformeerd worden, zijn ze bovendien voor de duidelijkheid van de figuur met -1 vermenigvuldigd).

De regio's met de zwaarst gearceerde balkjes hebben een waarde dicht in de buurt van het over-all gemiddelde. Kruislings gearceerde balken geven een waarde boven het gemiddelde weer. Enkelvoudige arcering duidt op een waarde daaronder. Naarmate de arcering zwaarder is, is de afwijking van het landelijk gemiddelde groter en de corresponderende balk hetzij langer, hetzij korter. De verschillen tussen de regio's zijn niet erg groot, maar opvallend is dat de gebieden die 'traditioneel' als achterstandsgebieden worden beschouwd (b.v. Oost-Groningen, Drenthe, Limburg) boven het landelijk gemiddelde scoren.

CONCLUSIE

De in Meester & De Leeuw (1984) getrokken conclusie dat de regionale verschillen in schoolloopbaan gering zijn ten opzichte van verschillen in sociale herkomst, kan op grond van deze variantie-analyse nader gespecificeerd worden. De convergentie van resultaten bij loglineaire en variantie-analyse geeft daarbij meer vertrouwen in de validiteit van die conclusie. Wel moeten we bedenken dat we met de drie achtergrondvariabelen tezamen ten hoogste 14% van de variantie in schoolloopbaan kunnen beschrijven.

Bij de prestatiescores neemt het milieu van herkomst ongeveer de helft van dit gedeelte van de spreiding voor zijn rekening, bij de overige schoolloopbaanvariabelen ongeveer driekwart.

Van de overige effecten is het (hoofd)effect van regionale herkomst het belangrijkste en hoewel minder pregnant dan het milieu-effect, is het toch steeds significant. Relatief het sterkst blijkt dat in de schoolprestaties en de onderwijsadviezen, variabelen die het meest met het lager onderwijs te maken hebben. In het voortgezet onderwijs (keuze en niveau na 5 jaar) lijken de relatieve invloeden van sociale en regionale herkomst zich te stabiliseren.

Het na 5 jaar bereikte onderwijsniveau verschilt overigens tussen de regio's niet zo veel. Als we de gemiddelden, per

COROP-gebied, "in kaart brengen" valt het op dat de SMVO-leerlingen in gebieden die vaak als achterstandsgebieden worden beschouwd (b.v. delen van Groningen, Drente en Limburg) boven het landelijk gemiddelde uitsteken.

- * In nauw overleg met de Coördinatie-commissie Regionaal Onderzoeksbeleid heeft het CBS een indeling in 40 functionele regio's gemaakt. Deze regio's worden met verwijzing naar de naam van de commissie meestal aangeduid als COROP-gebieden. Deze regio's omvatten steeds één of meerdere stedelijke werk- en verzorgingscentra die als knooppunten van het maatschappelijk verkeer fungeren. Het verschil met de indeling in zg. nodale gebieden bestaat erin dat de indeling grover is (nodale gebieden zijn er 80) en dat de COROP-gebieden de provinciegrenzen niet overschrijden. De grenzen van de COROP-gebieden zijn globaal in fig. 1 terug te vinden.

REFERENTIES

- CBS, *Schoolloopbaan en herkomst van leerlingen bij het voortgezet onderwijs, deel 2; cohort 1977, schoolkeuze*, 's Gravenhage, Staatsuitgeverij, 1982.
- Meester, A.C. & Leeuw, J. de, "Intelligentie, sociaal milieu en de schoolloopbaan", vakgroep Datatheorie, FSW/RUL, Leiden, 1983.
- Meester, A.C. & Leeuw, J. de, "Regionale kenmerken en schoolloopbaan", vakgroep Datatheorie, FSW/RUL, Leiden, 1984.
- De Leeuw, J., Young, F.W. & Takane, Y., "Additive structure in qualitative data: an alternating least squares method with optimal scaling features", *Psychometrika*, 41, pp 471-503.