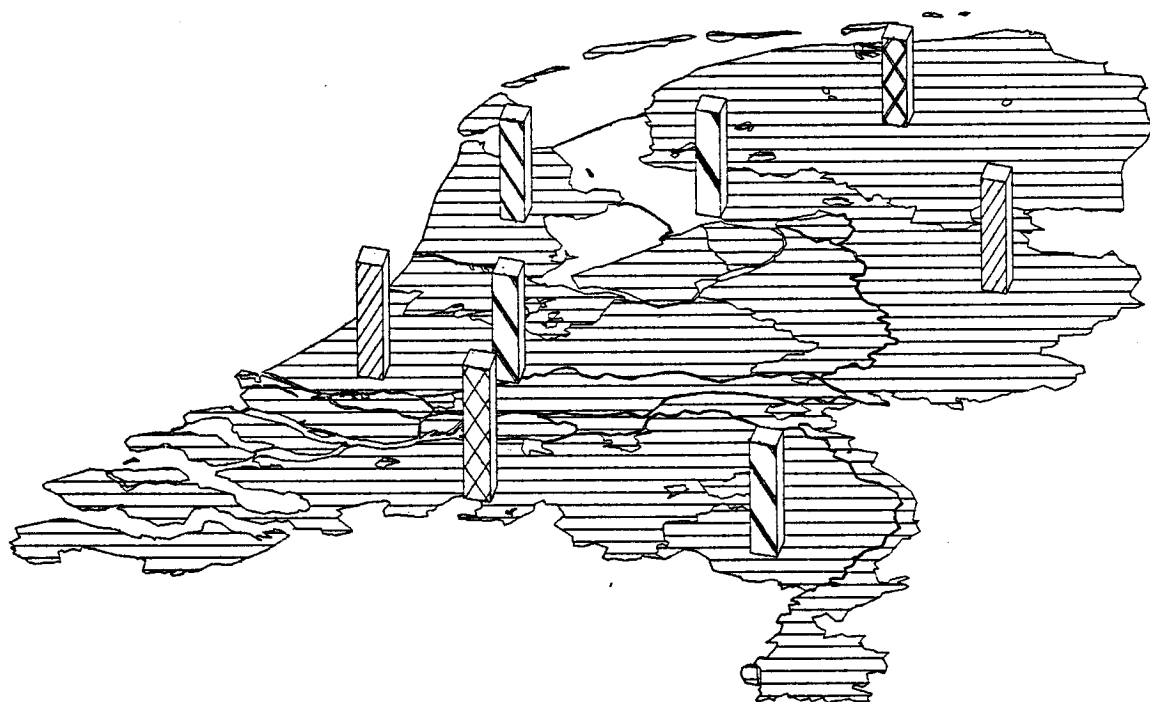


REGIONALE KENMERKEN
EN SCHOOLLOOPBAAN



Adriaan Meester
Jan de Leeuw
Vakgroep Datatheorie
FSW/RUL

REGIONALE KENMERKEN EN SCHOOLLOOPBAAN

September 1984

Adriaan Meester
Jan de Leeuw
Vakgroep Datatheorie
FSW/RUL

Tweede gewijzigde druk. Wij bedanken dr. J. Dronkers en drs. C.M. van Putten voor hun commentaar op de eerste druk.

Dank ook aan Marjon Schouten voor typografie en vormgeving

INHOUDSOPGAVE

	<u>blz.</u>	
1	INLEIDING	7
1.1	Indeling van de rapportage	7
1.2	Problemen van het multi-level onderzoek	8
1.3	Voorafgaand onderzoek naar regionale factoren	10
2	PROBLEEMSTELLING, VARIABELENKEUZE EN ANALYSESTRATEGIE	15
2.1	Regionaal niveau; van gemeente naar COROP-gebied	15
2.2	COROP-gebieden en hun kenmerken	16
2.3	Individuele schoolloopbanen; SMVO-variabelen	19
2.4	Opzet van de analyse	20
3	ANALYSES	23
3.1	Log-lineaire analyse	23
3.2	Analyse van de relaties tussen sociale achtergrond en de schoolloopbaan-variabelen binnen COROP-gebieden	24
3.3	Relatie tussen individuele schoolloopbaan en COROP-kenmerken	27
3.4	Heranalyse van COROP-kenmerken	28
3.4.1	Principale componenten analyse; eerste resultaat	28
3.4.2	Extra variabelen	30
3.4.3	Factoranalyse; 2 factoren	31
3.5	De relatie tussen kenmerken op COROP-niveau en de SMVO-variabelen	36

	<u>blz.</u>
3.5.1 Gemiddelden van gequantificeerde SMVO-variabelen versus COROP-factoren	37
3.5.2 De samenhang van de correlaties tussen SMVO-variabelen en de COROP-factorscores	45
3.6 Canonische correlatie-analyse	52
4 DISCUSSIE EN AFSLUITING	57
4.1 Het intelligentie-onderzoek bij de militaire keuringen	57
4.2 Samenvatting en slotconclusie	59
NOTEN	63
APPENDICES	67
A: Correspondentie-analyse met behulp van spline-functies	
B: PARTALS	
C: Een binomiaal model voor onvolledige gegevens	
LITERATUUR	75
BIJLAGEN	79

1 INLEIDING

Deze studie is de tweede in een serie die verslag doet van secundaire analyses van het door het CBS in het zogenaamde SMVO-cohort verzamelde schoolloopbaanmateriaal. Beide projecten werden door de Begeleidingscommissie Onderzoek Schoolloopbanen (BOS) van het Ministerie van Onderwijs positief beoordeeld en via de Stichting voor Onderzoek van het Onderwijs (SVO) aanbesteed.

Stonden in de voorafgaande aflevering kenmerken van de sociale achtergrond en intellectuele begaafdheid centraal (Meester en De Leeuw, 1983), in het voorliggende rapport worden een aantal pogingen om individuele schoolloopbanen aan regionale kenmerken te relateren besproken. Deze studie bouwt in belangrijke mate voort op de resultaten van de vorige studie. Variabelenkeuze en analysestrategie vertonen grote overeenkomsten. Enigszins simplificerend zou men kunnen zeggen dat het verschil tussen beide onderzoeken bestaat in het vervangen van de onafhankelijke variabele 'intelligentie' door de variabele 'regio'. Verschillen tussen de twee studies komen vooral voort uit het feit dat intelligentie als een numerieke variabele kan worden opgevat en regio als een nominale.

1.1 Indeling van de rapportage

Regionale karakteristieken hebben in het schoolloopbanenonderzoek tot nog toe, i.t.t. het sociaal milieu waaruit leerlingen afkomstig zijn, betrekkelijk weinig aandacht gekregen (Tesser, 1981). Een bespreking van de relevante literatuur zoals die in een van de volgende paragrafen is opgenomen, heeft dan ook noodzakelijkerwijs een vrij summier

karakter. Daaraan voorafgaand wordt een korte schets gegeven van de problemen die zich voordoen wanneer gegevens van verschillend aggregatieniveau in een onderzoek in beschouwing worden genomen.

Een belangrijk punt van overweging is de keuze van het regionaal niveau waarop de analyses zullen worden uitgevoerd. Oorspronkelijk lag het in de bedoeling kenmerken op gemeentelijk niveau met de individuele schoolloopbanen in verband te brengen. Technische en organisatorische problemen dwongen tot overschakeling naar een meer geaggregeerd niveau (het zg. COROP-gebied). In het tweede hoofdstuk waarin ook de keuze en categorisering van variabelen op individueel niveau zal worden verantwoord en de probleemstelling wordt geformuleerd, zal deze overschakeling nader worden toegelicht.

De daaropvolgende hoofdstukken zijn gewijd aan meer en minder geslaagde pogingen individuele en regionale kenmerken aan elkaar te relateren. Uiteindelijk resulteert dat in twee series grafieken waarin de in de data gevonden verbanden (of de afwezigheid daarvan) worden samengevat.

Dit gedeelte van de rapportage wordt afgesloten met het verslag van een analyse waarbij de invloed van de regionale variabele uit de relatie tussen achtergrond en de schoolloopbaan wordt gepartitioneerd.

In het vierde en laatste hoofdstuk wordt alles nog eens samengevat en gesystematiseerd.

Evenals in het voorafgaande analyseproject, maakte ook van deze studie een heranalyse van voor- en na-oorlogse dienstkeuringsgegevens ter ondersteuning van de interpretatie van de resultaten deel uit.

Het verslag van deze heranalyses is eveneens in het vierde hoofdstuk opgenomen.

1.2 Problemen van het multilevel onderzoek

In een commentaar op een recent artikel van Van den Eeden en Saris (1984) zetten De Leeuw en Kreft (1984) uiteen waar het in het zogenaamde 'multilevel'-onderzoek vaak aan schort. Een aantal van de onderzoeken die in de afgelopen jaren onder deze noemer zijn gepresenteerd kan strikt genomen niet zo worden benoemd. De belangrijkste punten uit de argumentatie van De Leeuw en Kreft worden hier kort weergegeven. De term 'multilevel' slaat op het gebruik van gegevens

die op verschillende aggregatieniveaus zijn verzameld. Als we onder een variabele een afbeelding of functie verstaan die een bepaald domein (de verzameling van onderzoekseenheden waarop de variabele is gedefinieerd) heeft, dan doet multilevelonderzoek uitspraken over de relatie tussen variabelen van verschillend domein. Omdat deze domeinen meestal 'genest' zijn (de elementen van een bepaald domein zijn een deelverzameling van de elementen van een minder geaggregeerd niveau) kunnen ze via de eenvoudige relatie van aggregatie of disaggregatie met elkaar in verband gebracht worden.

Wanneer nu, zoals bijvoorbeeld door Dronkers en Schijf (1984) gedaan wordt, variabelen van verschillend niveau onder hetzelfde domein gebracht worden, is van multilevel analyse eigenlijk geen sprake meer. Dronkers en Schijf disaggregeren door de waarden van variabelen op school- en wijkniveau aan individuele leerlingen toe te kennen. Dat zij deze procedure volgen is op zichzelf niet zo verwonderlijk. Het Nederlandse schoolloopbanenonderzoek (Tesser, 1981, en Peschar, 1983, geven daar een overzicht van) heeft zich tot nog toe voornamelijk met analyse op individueel niveau beziggehouden. Ook Bun-Siersma en Spruit (1982) pasten een dergelijke aggregatie/disaggregatiemethodiek toe. In de volgende paragraaf gaan we daar nog nader op in.

Zij het zo, dat wanneer alle variabelen naar één domein worden ge-(dis-)aggregeerd er eigenlijk van multilevelanalyse geen sprake meer is, dan nog moeten we erop bedacht zijn dat op geaggregeerd niveau gevonden relaties niet zonder meer op individueel niveau toegepast kunnen worden. Galtung (1967) noemt dit "fallacy of the wrong level". Hannan (1971a) refereert er in een uitstekend overzicht aan als de "ecologische fout" en het "Robinson effect".

Daarbij zullen correlatiecoëfficiënten bij toenemend aggregatieniveau in het algemeen sterk aan inflatie onderhevig zijn. Al deze bezwaren gelden a fortiori wanneer bij de analyse gebruik gemaakt wordt van causale- of padmodellen. Het 'fitten' van een schoolloopbaanmodel met variabelen op individueel en geaggregeerd niveau vereist op de eerste plaats consistentie, d.w.z. dat de relaties op het geaggregeerde niveau dezelfde vorm hebben als op het individuele. Consistentie is alleen te bereiken als alle relaties in het model (micro-, macro- én aggregatierelaties) lineair zijn (Hannan 1971b, p. 497).

Lineariteit binnen één domein is te benaderen door gebruik van optimale schalingstechniek en het is verheugend dat dit besef in de Neder-

landse onderwijssociologie begint door te breken (Dronkers en Schijf, 1984). Consistentie van het aangepaste model vereist echter meestal zware vooronderstellingen. Hannan (1971b) merkt hierbij op: "Given the state of theory and data in the nonexperimental social sciences, it is highly unlikely that all of the assumptions underlying any one of these techniques will be met in any substantively interesting application" (p. 473).

Dat betekent overigens niet dat van dergelijke analysemethoden maar moet worden afgezien. Realistischer is het de vooronderstellingen zoveel mogelijk te expliciteren en in gedachten te houden dat het kleinste kwadraten-algorithme waar op veel van deze technieken zijn gebaseerd, de neiging heeft zich te concentreren op de grootste effecten in het model. Variabelen met relatief gering effect (en dat zijn regionale variabelen vaak) krijgen op die manier weinig kans zich te manifesteren.

De conclusie die voor het onderhavige onderzoek uit het voorafgaande wordt getrokken, is dat het verstandig is zo weinig mogelijk te aggregeren, bij voorkeur slechts bivariate samenhangen te analyseren en de gebruikte variabelen steeds volgens één of ander rationeel criterium optimaal te herschalen.

We hanteren daarbij een twee-stepsprocedure. Relaties tussen variabelen worden geanalyseerd binnen het domein waarop ze zijn gedefinieerd en pas in tweede instantie worden de verschillende aggregatieniveaus met elkaar in verband gebracht.

1.3 Voorafgaand onderzoek naar regionale factoren

Regionale factoren hebben in het Nederlandse schoolloopbaanonderzoek tot het einde van de jaren zeventig slechts marginale aandacht genoten. Dat wil niet zeggen dat regionale verschillen volledig werden verwaarloosd. Het onderzoek van de provincie Noord-Brabant (1952) en het daarop aansluitende onderzoek van Matthijssen en Sonnemans (1958) hadden het specifieke karakter van die regio als vertrekpunt. Het Talentenproject (Van Heek, 1968) bestond uit deelprojecten in vier verschillende regio's (Friesland, Twente, Noord-Brabant, Amsterdam). Door Ruiter, Eisma en Passenier (1968) zijn de verschillen tussen deze regio's geanalyseerd en zij komen tot de (overigens empirisch niet

erg stevig onderbouwde) conclusie dat de verschillen in VHMO-participatie tussen de vier regio's te herleiden zijn tot verschillen in samenstelling van de beroepsbevolking in deze gebieden. Een recente heranalyse van de gegevens van het Talentenproject (Meester en De Leeuw, 1983) lijkt dat te bevestigen. Collaris en Kropman (1978) beschouwen de invloed van 'macro-sociale' factoren als regio, urbanisatiegraad en godsdienstige gezindte 'minder relevant' (p. 70) en vonden ook geen significante samenhangen met deze variabelen (p. 93).

Verschillen in deelname aan de onderscheiden vormen van voortgezet onderwijs zijn niet los te zien van deelname en organisatie van lager en hoger onderwijs.

Santema (1974) probeert verschillen in deelname aan het voortgezet onderwijs in verband te brengen met de organisatiestructuur van het lager onderwijs op het platteland. Hij vindt (mede op basis van buitenlandse literatuur) allerlei aanwijzingen dat verschillen tussen plattelandscholen verschillen tussen leerlingen 'veroorzaken', die zich dan zouden uiten in de doorstroming naar het v.o.

De doorstroming naar het hoger onderwijs zou in belangrijke mate beïnvloed worden door de afstand tussen woonplaats en universiteitsstad (CBS, 1958; Vergoossen en Wever, 1978; Koornstra, Gerritsen en Wiercx van Rhijn, 1983). Daarnaast wordt enig belang toegekend aan de differentiële deelname aan de hogere vormen van v.o. door de hogere sociale milieus. Kodde (1984) relativeert dit weer enigszins: "De studierichtingkeuze is echter aanzienlijk minder sociaalmilieu gepredetermineerd dan vroeger" (p. 18).

Vliegen en De Jong (1981) hebben de relatie tussen regionale kenmerken en onderwijsparticipatie in Nederland onderzocht met gebruikmaking van gegevens uit de volkstelling van 1971. Voor de 80 zogenaamde nodale gebieden pasten zij een pad-analytisch model aan dat het effect van culturele en structurele kenmerken op de deelname aan het VWO moet schatten. De culturele aspecten blijken in het model van gering belang. De structurele variabelen - regionale woonsituatie, arbeidsmarkt en verstedelijking - 'verklaren' 55% van de variantie in VWO-deelname. Hoe deze 'invloed' wordt uitgeoefend blijft onduidelijk.

Dronkers en Diekerhof (1980) hebben een aanzet gegeven tot meer systematisch onderzoek naar regionale verschillen in individuele schoolloopbaan in het secundair onderwijs. Zij hebben het effect van

locale deelname aan verschillende typen onderwijs geprobeerd vast te stellen aan de Van Jaar tot Jaar-gegevens. Zij gebruikten daarbij een vereenvoudigde versie van het z.g. Wisconsin-model (zie o.a. Dronkers, 1978, Dronkers en De Jong, 1978 en Dronkers en Jungbluth, 1979). Zij meenden interpreteerbare en niet te verwaarlozen effecten van locale participatie-variabelen gevonden te hebben en bepleitten nader onderzoek op dit vlak.

Voor het KASKI was dit resultaat aanleiding een voorstel in te dienen voor een onderzoek dat de samenhang van variabelen op regionaal niveau met de ontwikkeling van individuele schoolloopbanen zou moeten vaststellen m.b.v. de data van het SMVO-cohort. In Bun-Siersma en Spruit (1981, 1982) wordt hierover gerapporteerd.

In eerste aanleg is voor 792 van de 810 Nederlandse gemeenten een aantal kenmerken verzameld. De relaties tussen deze kenmerken zijn door middel van principale componenten analyse onderzocht. Dat resulteerde in een 3-dimensionale structuur waarvan de componenten benoemd werden als 'werkloosheid', 'sociaal-educatieve voorsprong' en 'culturele gelegenheid'. Voor details omtrent de opgenomen variabelen en de gevolgde werkwijze zij verwezen naar de KASKI-publikatie van 1981.

Door De Leeuw (1982a) zijn dezelfde gegevens op verschillende manieren niet-lineair onderzocht. Hij laat zien dat de niet-lineaire oplossingen weliswaar globaal hetzelfde beeld geven als de KASKI-benadering maar veel interessante extra informatie te bieden hebben voor wat betreft de eigenschappen van de variabelen en de structuur van hun relaties. Het blijkt mogelijk de op gemeentenniveau verzamelde kenmerken te beschrijven in een 2-dimensionaal vlak dat bepaald wordt door werkloosheid en gemeentegrootte.

In de vervolg-publikatie (1982) van het KASKI worden de gemeentenkenmerken samen met de schoolloopbaangegevens in één regressiemodel geanalyseerd en vervolgens wordt deze procedure herhaald met vergelijkbare geaggregeerde variabelen op het niveau van nodale gebieden. Voor de analyses worden alle variabelen naar het individuele niveau gedisaggregeerd. In het licht van het in de voorgaande paragraaf gestelde, lijkt deze methodiek een weinig gelukkige keuze. Dat de verschillende modellen die aangepast werden weinig tastbaar resultaat opleverden, terwijl op de werkwijze bovendien nogal wat methodologische kritiek mogelijk is, wekt dientengevolge weinig verwondering.

In paragraaf 2.1 zal nog worden toegelicht dat ook de wijze waarop de SMVO-steekproef is getrokken een analyse als door het KASKI uitgevoerd, bezwaarlijk maakt.

Een laatste publikatie waaraan we in deze paragraaf enige aandacht willen besteden is van het RION (Boon e.a., 1983). In dit onderzoek wordt een sterke relatie gevonden tussen sociaal-culturele achterstand en deelname aan verschillende vormen van voortgezet onderwijs per provincie. Eveneens wordt een verband geconstateerd tussen werkloosheid en de onderwijsdeelname. Naarmate in een provincie minder kans op werk is, kiezen meer jongeren voor lager of middelbaar beroepsonderwijs. De gevonden verbanden zijn overigens moeilijk één-duidig te interpreteren nog afgezien van het feit dat het provinciale niveau voor het onderzoek van regionale verschillen naar onze opvatting eigenlijk te grof is.

In een appendix van de RION-publikatie zijn ook de resultaten opgenomen van een analyse van onderwijsparticipatie per z.g. COROP-gebied. Deze resultaten worden op kruistabel-niveau besproken. Wij zullen deze gegevens in onze analyses op dit regionale niveau betrekken (zie paragraaf 3.4.2).

2 PROBLEEMSTELLING, VARIABELENKEUZE EN ANALYSESTRATEGIE

Zeer in het algemeen kan de probleemstelling van het onderhavige onderzoek worden geformuleerd als: "Hoe is de relatie tussen sociaal milieu en geslacht enerzijds en de schoolloopbaan in het voortgezet onderwijs anderzijds in verschillende regio's en is er een verband tussen kenmerken op regionaal niveau en deelname aan en schoolloopbaan in het voortgezet onderwijs?"

Als we voor de analyse de individuele schoolloopbaan als uitgangspunt nemen ligt de nadruk op het eerste gedeelte van de vraagstelling. Voor het tweede gedeelte is het regionale niveau het uitgangspunt.

Op grond van de onder 1.2 geformuleerde problemen van aggregatie en disaggregatie zou het in de rede liggen voor de regionale kenmerken een aggregatieniveau te kiezen dat zo dicht mogelijk bij de op individueel niveau verzamelde schoolloopbaangegevens ligt. Bij analyse van de SMVO-cohort gegevens zou daarom het gemeentelijk niveau het meest in aanmerking komen. In de oorspronkelijke onderzoeksopzet was dan ook voorzien in een eerste fase waarin een typologie van gemeenten zou worden gemaakt. In een tweede fase zouden schoolloopbanen binnen gemeentetypen worden bestudeerd, om eventuele verschillen vast te stellen en te representeren. Waarom deze opzet uiteindelijk niet is gerealiseerd wordt in de volgende paragraaf uiteengezet.

2.1 Regionaal niveau; van gemeente naar COROP-gebied

Nog afgezien van het teleurstellend resultaat van een analyse op gemeentelijk niveau (Bun-Siersma en Spruit, 1982) en de kritiek die daarop geïnterd kan worden (De Leeuw, 1982a), is een dergelijke analyse

met gebruikmaking van de SMVO-gegevens eigenlijk niet goed mogelijk. De manier waarop de SMVO-steekproef is samengesteld, brengt dat met zich mee. Op 'getrapte' wijze zijn uit het totaal beschikbare scholenbestand (met ongelijke trekkingskans voor de verschillende schooltypen) scholen getrokken en vervolgens binnen scholen één of meer klassen geselecteerd. Dat heeft o.a. tot gevolg dat m.n. in kleinere gemeenten slechts één schooltype kan voorkomen, zodat het lijkt alsof alle kinderen uit één gemeente naar bijv. het LBO of het VWO gaan. Bovendien is voor het SMVO-cohort de woongemeente geregistreerd en niet de gemeente waar de school staat¹⁾. Vooral t.a.v. weinig voorkomende schooltypen en kleinere gemeenten kan discrepantie tussen woon- en schoolgemeente bestaan. Het effect van deze discrepantie op de individuele schoolloopbaan is niet te schatten. Deze overwegingen maken het tamelijk dubieus om kenmerken op gemeentelijk niveau aan individuele schoolloopbanen te relateren.

Een tweede reeks bezwaren tegen de oorspronkelijke analyseopzet is van technische en organisatorische aard. Sommige gegevens (zoals bijv. het aantal bijstandstrekkers per gemeente) zijn geheim. De gemeentelijke indeling verandert van jaar tot jaar en niet alle gewenste gegevens zijn voor de onderzochte periode (1977-1982) of voor dezelfde jaren beschikbaar. Koppeling van de regionale gegevens aan het gemeentenummer zou om deze redenen altijd onvolledig blijven en bovendien erg kostbaar worden.

In overleg met de SVO werd daarom besloten een grovere definitie van regio te gebruiken, nl. het zogenaamde COROP-gebied. We verwachten in dit soort gebieden een betere verdeling van de scholen en minder nadelen van de stratificatie. Door Nijkamp en zijn medewerkers (zie o.a. Nijkamp en Rietveld, 1983 en Brouwer, Nijkamp en Rietveld, 1983) zijn al verschillende typologische analyses op dit regionale niveau uitgevoerd. Met gebruikmaking van niet-lineaire technieken kunnen wij op dit onderzoek voortbouwen.

Additionele voordelen zijn dat deze gebiedsindeling in de cohort-periode ongewijzigd is gebleven en nauwer aansluit bij de regionale planning.

2.2 COROP-gebieden en hun kenmerken

In nauw overleg met de Coördinatie-commissie Regionaal Onderzoeksprogramma heeft het CBS een indeling van Nederland 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 een 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 locatie van de 40 regio's is afgebeeld in fig. 2.1.

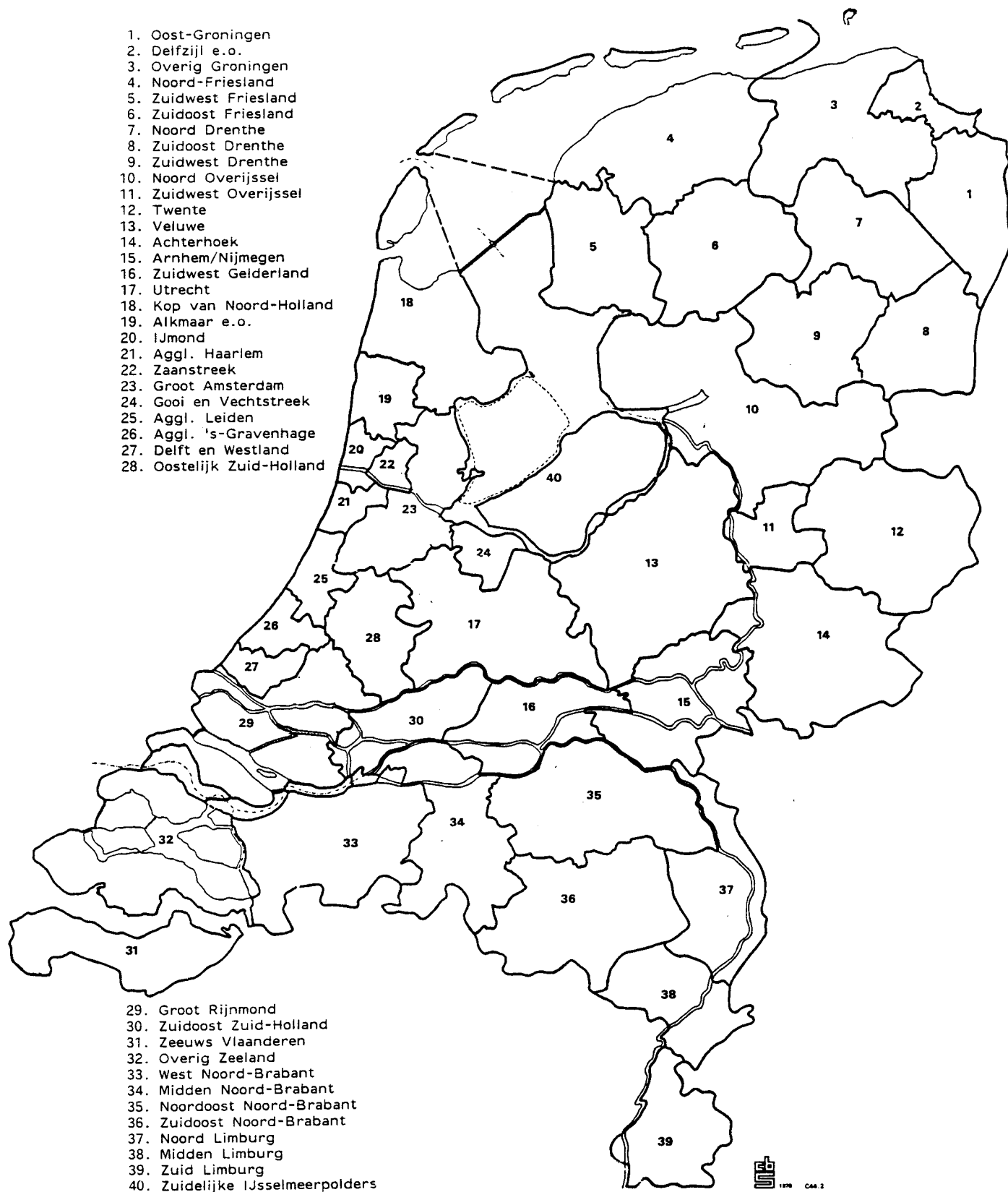
Van Veenendaal (1981) heeft voor elk van de COROP-gebieden 13 indicatoren²⁾ berekend, te weten:

- Sociaal-economische kenmerken
 - 1: Inkomen per hoofd van de bevolking na belastingen
 - 2: Werkloosheidspercentage
 - 3: Vermogen per hoofd van de bevolking ('welvaart')
 - 4: Index voor de kosten van levensonderhoud (i.h.b. huur)
- Kenmerken van het natuurlijk milieu
 - 5: Bevolkingsdichtheid
 - 6: Natuurgebied als percentage van de totale oppervlakte
 - 7: Industrialisatie-index
 - 8: Emissie-index voor vervuiling
- Kenmerken van de infrastructuur
 - 9: Dichtheid van het wegennet
 - 10: Index voor sport- en culturele voorzieningen per hoofd
 - 11: Index voor onderwijsvoorzieningen per hoofd
 - 12: Afstand t.o.v. de Randstad
 - 13: Index voor medische voorzieningen per hoofd

De basisgegevens zijn uit verschillende CBS-statistieken afkomstig en hebben betrekking op verschillende jaren tussen 1974 en 1979.

De kenmerken zijn genormeerd op een schaal van 1 tot 100.

Figuur 2.1: Indeling van Nederland in 40 COROP-gebieden.



2.3 Individuele schoolloopbanen; SMVO-variabelen

De operationalisatie van de sociale achtergrond en de schoolloopbaan van de leerlingen in het SMVO-cohort is volledig analoog aan die in de voorgaande studie (Meester en De Leeuw, 1983).

Voor nadere beschrijving en verantwoording van de variabelen verwijzen wij daarnaar en naar CBS (1982).

Het sociaal milieu waaruit de leerlingen afkomstig zijn (BVA) wordt weergegeven door het beroepsniveau van de vader in 6 klassen:

1. Arbeiders (geschoold en ongeschoold)
2. Boeren en tuinders
3. Zelfstandige middenstand
4. Lagere employées
5. Middelbare employées
6. Wetenschappelijke en vrije beroepen en hogere employées

De noodzaak dit achtergrondskenmerk in de studie te betrekken en de problemen die aan het gebruik van deze milieu-indicator kleven worden in Meester en De Leeuw (1983, pp. 14-15) becommentarieerd. Bovendien hangen volgens Ruiters, Eisma en Passenier (1968) regionale verschillen in deelname aan de hogere vormen van voortgezet onderwijs grotendeels samen met regionale verschillen in de samenstelling van de bevolking naar beroepsgroepen.

De betekenis van sekse blijkt in dit type onderzoek gering te zijn (Meester en De Leeuw, 1983). In het algemeen betrekken we deze variabele slechts marginaal in de beschouwingen. Jongens en meisjes zijn min of meer gelijkelijk over de COROP-gebieden verdeeld (zie tabel (bijlage) 2.1).

In tabellaire (log-lineaire) analyses voegen we, overeenkomstig de in de voorgaande studie gevolgde procedure, BVA en SEXE samen tot één variabele met 6 x 2 categorieën, om eventuele interactie tussen geslacht en sociaal milieu onder controle te houden. Hetzelfde doen we in de afsluitende canonische correlatie analyse.

In Meester en De Leeuw (1983) werden de predictieve waarden van intelligentiescores en schoolprestaties t.a.v. de schoolloopbaan met elkaar vergeleken. Op grond van het daar gevonden resultaat blijven de intelligentiescores buiten beschouwing en worden alleen de totaal-

scores op de CITO-toets (PRE, ingedeeld in 10 categorieën; zie Meester en De Leeuw, 1983, p. 114) in de analyse betrokken.

Het advies van de onderwijzer voor het voortgezet onderwijs (ADV) en de eerste keuze in het v.o. (KEUS) worden, als steeds, in 4 klassen (LBO, MAVO, HAVO en VWO) ingedeeld. Voor ADV handhaven we ook de categorie 5. 'Advies Onbekend'.

Het in 1982 bereikte niveau in het voortgezet onderwijs (EIN) is op dezelfde manier als in het voorgaande onderzoek (Meester en De Leeuw, 1983, p. 115-116) geconstrueerd, met dit verschil dat de variabele nu één jaar meer omvat. De categorie-indeling van EIN-1982 is als volgt:

1. Voortijdige schoolverlaters ('drop-outs')
2. Lager beroepsonderwijs (LBO)
3. Middelbaar algemeen vormend onderwijs (MAVO)
4. Hoger algemeen vormend onderwijs (HAVO)
5. Voorbereidend wetenschappelijk onderwijs (VWO)
6. Middelbaar of hoger beroepsonderwijs (met LBO-diploma of zonder diploma van een voorafgaande opleiding).
7. Middelbaar of hoger beroepsonderwijs (met MAVO- of HAVO-diploma)

Categorie 6 bestaat voornamelijk uit MBO-leerlingen; in categorie 7 zijn de HBO-leerlingen sterker vertegenwoordigd.

Overeenkomstig eerder gevolgde procedures zijn de leerlingen met een ontbrekende score op BVA uitgeselecteerd. Ook de leerlingen waarvan niet bekend is uit welke COROP-gebied ze afkomstig zijn, werden voor de analyses uit het bestand verwijderd. Van het oorspronkelijk SMVO-bestand van 37.280 leerlingen blijven na deze selecties 29.290 gevallen over. De frequentie-verdelingen van BVA, SEXE, PRE, ADV, KEUS en EIN staan in bijlage 2.2.

2.4 Opzet van de analyse

In 1.2 is uiteengezet welke problemen het met zich meebrengt om variabelen van verschillend aggregatieniveau gezamenlijk in één analyse te betrekken. Hoewel verscheidene van de daar genoemde bezwaren (lineariteit van de regressies als noodzakelijke voorwaarde

b.v.) door de in deze studie gehanteerde technieken ondervangen worden, zal in dezen toch uiterste voorzichtigheid worden betracht. In verschillende analyse-stappen worden de aggregatie-niveaus (individueel en regionaal) zoveel mogelijk gescheiden gehouden. Na een eerste exploratie m.b.v. log-lineaire analyse waarbij alle bovenbeschreven variabelen betrokken zijn, analyseren we eerst de schoolloopbaan- en achtergrondgegevens binnen COROP-gebieden om vervolgens tussen COROP-gebieden te vergelijken.

De COROP-gebieden kunnen op die manier dus getypeerd worden door de gemiddelden van de (optimaal gequantificeerde) variabelen en het patroon van hun onderlinge bivariate samenhang (de correlaties). Als de individuele gegevens op deze manier naar het regionale niveau zijn geaggregeerd kan hun relatie met de op COROP-niveau verzamelde data worden bestudeerd.

In Meester en De Leeuw (1983) werd de schoolloopbaan bekeken als functie van intelligentie (na controle voor BVA-SEXE). We willen hier hetzelfde doen met regio als onafhankelijke variabele. Daarvoor is het nodig dat de regio's eerst gequantificeerd worden.

Een mogelijkheid is om de geaggregeerde schoolloopbaan (of een indicator daarvan) rechtstreeks in verband te brengen met elk van de regionale variabelen. Het is echter eenvoudiger om een optimale gewogen combinatie van de COROP-variabelen te berekenen of de COROP-gebieden te beschrijven in termen van één of twee factoren en het verband daarvan met de geaggregeerde individuele gegevens te onderzoeken. Beide strategieën worden beproefd. Na al deze exercities zullen de data en hun onderlinge relaties nog eens meer in detail worden bestudeerd.

Belangrijk is in gedachten te houden dat de aggregatieniveaus zoveel mogelijk gescheiden blijven en dat we in het algemeen slechts bivariate samenhangen (op basis van transformaties die de regressie zoveel mogelijk lineariseren) onderzoeken. In de eerste analyse-stap worden dus, in termen van de probleemstelling, schoolloopbanen naar COROP-gebied gedifferentieerd, in het tweede gedeelte wordt onderzocht hoe regionale kenmerken samenhangen met de deelname aan en de loopbaan in het voortgezet onderwijs. Daarbij wordt het verband met de sociale achtergrond en het geslacht van de leerlingen steeds in de berekeningen mede betrokken.

3 ANALYSES

3.1 Log-lineaire analyse

Om een eerste algemene indruk van de relaties tussen de in de studie betrokken variabelen te krijgen, analyseerden we de kruistabellen van COROP, BVA-SEXE en, respectievelijk PRE, ADV, KEUS en EIN log-lineair. Log-lineaire analyse zoals het in deze studie wordt gebruikt, blijft dicht bij de kruistabel en berust op een minimum aan assumpties. Voor de theorie zij verwezen naar Bishop, Fienberg en Holland (1975), Fienberg (1977) en anderen.

De achtergrondvariabelen BVA en SEXE werden daartoe samengevoegd tot een interactieve variabele met 6×2 categorieën. Als geslachtsverschillen binnen de categorieën van BVA verschillende betekenis hebben, houden we dat op deze manier onder controle. Dezelfde procedure wordt ook in Meester en De Leeuw (1983) toegepast.

In eerste instantie werden dus vier 3-weg tabellen gemaakt. Om daarbij niet al te veel lege cellen te krijgen (de subtabel van COROP en BVA-SEXE heeft immers al 40×12 cellen), werd PRE in 3 categorieën (1-32, 33-50 en 51-70 items goed beantwoord) gehercodeerd en werden de leerlingen waarvan advies- of prestatiescore ontbreken, uit de betreffende analyses verwijderd.

Voor elk van de schoolloopbaanvariabelen vergelijken we de fit van het volledige 2e-orde model met een restrictiever submodel met slechts twee van de drie 2e-orde interacties erin opgenomen. De passing van de modellen voor, respectievelijk PRE, ADV, KEUS en EIN is gegeven in tabel 3.1.

Tabel 3.1: Fit van het volledige 2e orde model COROP, BVA-SEXE en resp. PRE, ADV, KEUS en EIN vergeleken met het submodel dat COROP en schoolloopbaan onafhankelijk zijn, gegeven BVA-SEXE. In de modellen zijn de variabelen aangeduid met de eerste letter van hun naam.

	Model	LRX ²	df.	AIC
a. PRE	[PB][CB]	1819.45	936	- 53
	[PC][PB][CB]	954.73	858	- 761
b. ADV	[AB][CB]	2841.04	1404	33
	[AC][AB][CB]	1533.35	1287	-1041
c. KEUS	[KB][CB]	2450.75	1404	- 358
	[KC][KB][CB]	1473.03	1287	-1101
d. EIN	[EB][CB]	3801.10	2808	-1815
	[EC][EB][CB]	2495.13	2574	-2653

Het submodel met slechts 2 van de drie 2^o-orde interacties erin opgenomen is te interpreteren als toetsing van de hypothese dat COROP en de schoolloopbaan onafhankelijk zijn, gegeven BVA-SEXE (vgl. Meester en De Leeuw, 1983, App.C.; 1984b). Volgens het informatie criterium (AIC) van Akaike (Akaike, 1973) is de fit van het volledige 2^o-orde model altijd beter dan van het submodel met twee interacties. De fit van het submodel is echter volgens de vuistregel dat AIC zo mogelijk kleiner dan nul moet zijn, nog alleszins acceptabel te noemen. Overigens is het volledige 2^o-orde model ook te interpreteren als: de relatie tussen schoolloopbaan en BVA-SEXE is in alle COROP-gebieden dezelfde.¹⁾

De voorlopige conclusie kan zijn dat, gegeven de sociale achtergrond van de leerlingen, de regio waar ze wonen geen invloed op hun schoolloopbaan lijkt uit te oefenen. Ofwel dat de regionale verschillen in schoolloopbanen t.o.v. de milieuverschillen te verwaarlozen zijn.

3.2 Analyse van de relaties tussen sociale achtergrond en de schoolloopbaanvariabelen binnen COROP-gebieden.

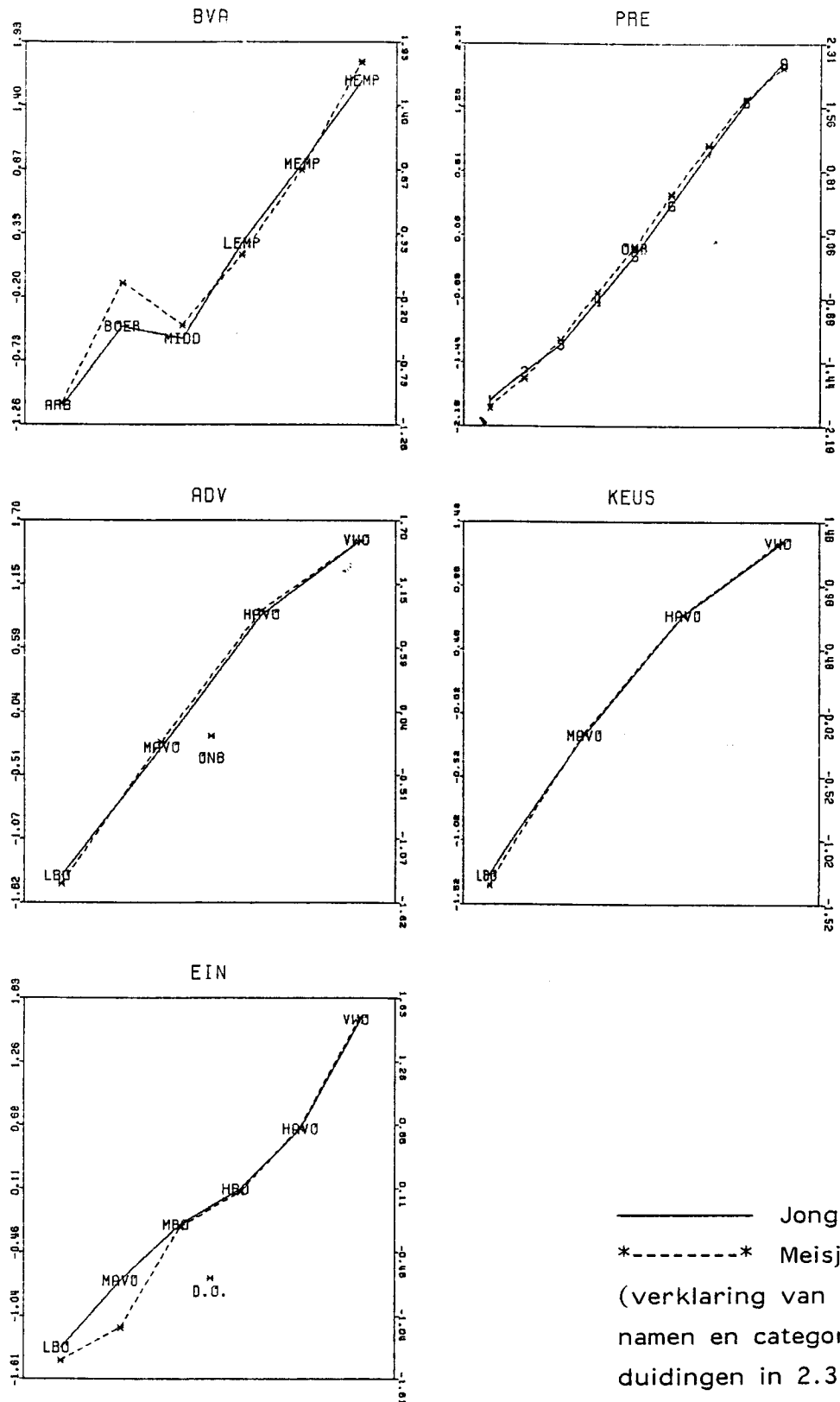
Voor deze analyse werden de gegevens naar COROP-gebied en SEXE gestratificeerd. Dat resulteerde in 40 x 2 subsets. De aantallen jongens en meisjes per regio staan in tabel (bijlage) 2.1. Voor elk van deze subsets werd een correlatiematrix van het sociaal milieu (BVA) en de schoolloopbaanvariabelen berekend. Dit gebeurde met het programma PREHOM²⁾ dat scores aan de categorieën van de variabelen toekent, zodanig dat de regressies zoveel mogelijk gelineariseerd wor-

den. Lineariteit van de regressies is een noodzakelijke voorwaarde om de correlatiecoëfficiënten te interpreteren als een weergave van de bivariate samenhangen. Hoewel de linearisatie vrij goed lukte, waren de gevonden quantificaties van COROP-gebied tot COROP-gebied zo verschillend (met name in de kleinere gebieden) dat een directe vergelijking van de quantificaties en correlaties niet goed mogelijk was. Daarom werden (voor jongens en meisjes apart) over het totale bestand nieuwe quantificaties berekend. Deze quantificaties zijn afgebeeld in figuur 3.1. Het blijkt mogelijk regelmatige transformaties van de variabelen te vinden die voor jongens en meisjes maar heel weinig verschillen. De categorieën 'onbekend' van PRE en ADV worden vrijwel in het nulpunt geplaatst. Op basis van deze quantificaties werden opnieuw 80 correlatiematrices gegenereerd. De variabelen zijn dan dus voor alle COROP-gebieden per geslachts categorie uniform getransformeerd. Per COROP-gebied werden op deze manier, voor jongens en meisjes apart, 10 correlatiecoëfficiënten berekend.

Een vergelijking van de COROP-gebieden naar de bivariate samenhangen is mogelijk door alle 80 x 10 correlatiecoëfficiënten in één grote matrix met 80 rijen en 10 kolommen te verzamelen en op deze tabel correspondentie-analyse toe te passen. Gebieden (naar sexe gedifferentieerd) met een sterk gelijkend patroon in de correlaties worden dan in een 2-dimensionaal vlak dicht bij elkaar geplaatst en ver verwijderd van gebieden met een afwijkend patroon. Als de verschillen tussen de rijen van de tabel geheel aan sexe-verschillen zouden moeten worden toegeschreven en interactie met COROP-gebied afwezig was, zouden alle jongens, evenals de meisjes, dicht bijeen worden geplaatst.

Zowel voor de verschillende als voor de uniforme quantificaties laat de correspondentie-analyse oplossing echter geen systematische samenhang zien.

Figuur 3.1: Optimale quantificaties BVA, PRE, ADV, KEUS en EIN uit PREHOM



3.3 Relatie tussen individuele schoolloopbaan en COROP-kenmerken

Een nieuwe poging om de COROP-gebieden naar schoolloopbaan te differentiëren bestond uit het berekenen, per COROP-gebied en nog steeds voor jongens en meisjes apart, van de correlatie tussen BVA en de som van de schoolloopbaanvariabelen. Uiteraard gebruikten we daarvoor niet de oorspronkelijke categorienummers, maar de uniforme quantificaties uit de eerdere analyse. Op COROP-niveau kan dan de canonische correlatie berekend worden tussen de op dat niveau verzamelde variabelen (zie 2.2) en de typering van COROP-gebieden naar schoolloopbaan. Nu zijn de COROP-variabelen weliswaar numeriek, maar ze hebben nogal merkwaardige verdelingen (De uit Brouwer, Nijkamp en Rietveld (1983) overgenomen tabel staat in bijlage 3.1). Vrijwel al deze variabelen hebben een zeer scheve of meertoppige verdeling. Door 'spline'-functies op de verdelingen te fitten (zie Appendix A) vinden we transformaties die de regressies lineair maken. Dit is, "het kan niet vaak genoeg herhaald worden" (De Leeuw en Kreft, 1984), een noodzakelijke voorwaarde voor de gebruikelijke interpretatie van de (canonische) correlatiecoëfficiënt. Nu lukt het heel goed om de regressies te lineariseren, maar de transformaties van de door de splines gecategoriseerde variabelen zijn onregelmatig en nauwelijks te interpreteren.

Significantie van de gevonden optimale of canonische correlaties kunnen we bestuderen met een eenvoudige permutatietoets. Wanneer we de observaties op één van de twee variabelen willekeurig permuteren (i.e. in een volstrekt willekeurige volgorde zetten), dan kunnen we met deze verstoorde gegevens onze rekenprocedures herhalen. Wanneer nu onze waargenomen correlatie niet wezenlijk hoger is dan de correlaties op basis van verstoorde gegevens, dan kunnen we zeggen dat de waargenomen correlatie niet significant van nul afwijkt. De permutatietoets kan uitgevoerd worden op basis van de permutatieverdeling (zoals bij Fisher's exacte toets voor associatie in 2×2 tabellen), op basis van een normaalverdeelde benadering bij grote steekproeven, of op basis van een Monte Carlo procedure. Zie voor meer details De Leeuw en Van der Burg (1984). We hebben hier de Monte Carlo procedure gebruikt, die door middel van willekeurige permutaties gestoorde data genereert, en vervolgens analyseert. Het blijkt dat een relatief klein aantal van ongeveer tien heranalyses dikwijls al

de gewenste informatie geeft.

De gevonden correlaties blijken deze test niet te doorstaan. Besloten werd het probleem van de andere kant te benaderen en te onderzoeken of de onafhankelijke COROP-variabelen op een zinvolle manier zijn samen te vatten. Daarvan wordt in de volgende paragraaf verslag gedaan.

3.4 Heranalyse van de COROP-kenmerken

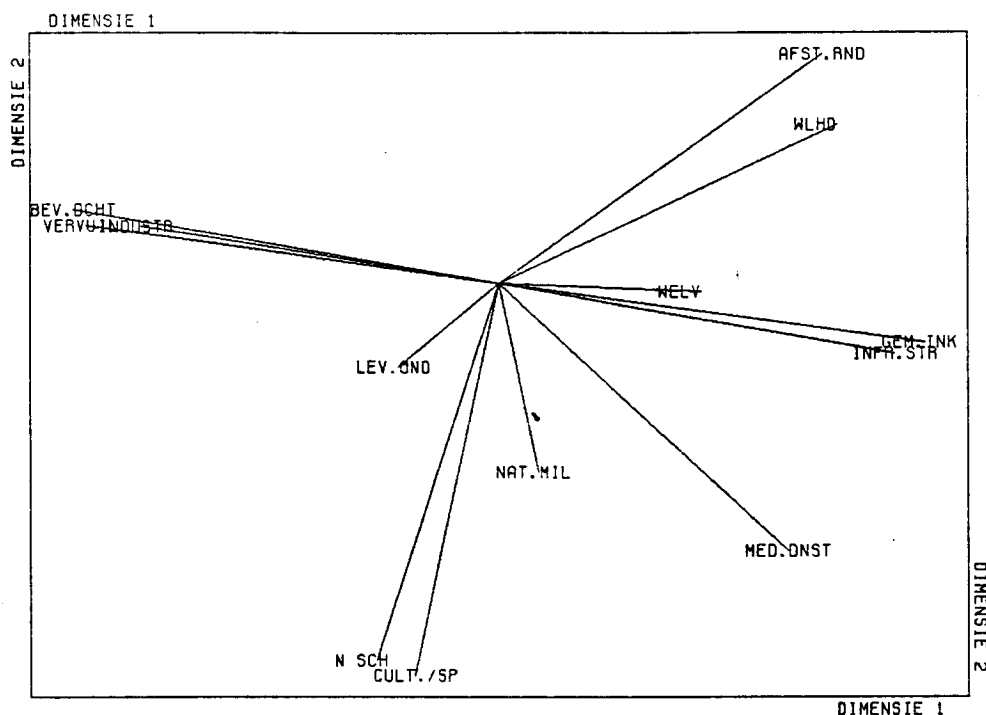
Alvorens de gegevens op COROP-niveau niet-lineair te heranalyseren, berekenden wij eerst de correlaties tussen deze variabelen opnieuw, ter vergelijking met de in Brouwer, Nijkamp en Rietveld (1983) opgenomen correlatiematrix. De verschillen bleken te verwaarlozen. De correlatiematrix staat in tabel (bijlage) 3.2. De matrix is vooral opgenomen ter vergelijking met de matrix van correlaties na optimale transformatie (tabel, bijlage, 3.3). Brouwer e.a. analyseerden deze gegevens met behulp van principale componentenanalyse (PCA). Hun bedoeling daarbij was de resultaten te vergelijken van een analyse waarbij de data als numeriek werden opgevat en een analyse waarbij slechts de rangorde van de gegevens in aanmerking werd genomen. In eerste instantie herhalen wij de analyse van Brouwer e.a. met een programma voor PCA dat scores aan de categorieën van de variabelen toekent, zodanig dat zij (gegeven de ordinaliteitsrestrictie) optimaal worden getransformeerd en vervolgens opnieuw een principale componentenoplossing berekent (Gifi, 1981). De regiokenmerken dienen dan eerst gediscretiseerd te worden. Als discretisatiepunten kiezen we de knooppunten van de spline-functies die we eerder op de verdelingen van de variabelen hebben gefit (zie 3.3 en Appendix A).

3.4.1 Principale componentenanalyse; eerste resultaat

Heranalyse van de in Brouwer, Nijkamp en Rietveld (1983) opgenomen tabel van genormeerde COROP-variabelen geeft een resultaat dat goed vergelijkbaar is met de PCA-oplossing in genoemde publikatie. Ook wij berekenden een 3-dimensionale oplossing. De drie dimensies hebben eigenwaarden van resp. .436, .166 en .140. De componentenladingen

voor de eerste 2 dimensies zijn weergegeven in fig. 3.2. De correlaties tussen de optimaal getransformeerde variabelen staan in tabel (bijlage) 3.3, de componentladingen in tabel (bijlage) 3.4.

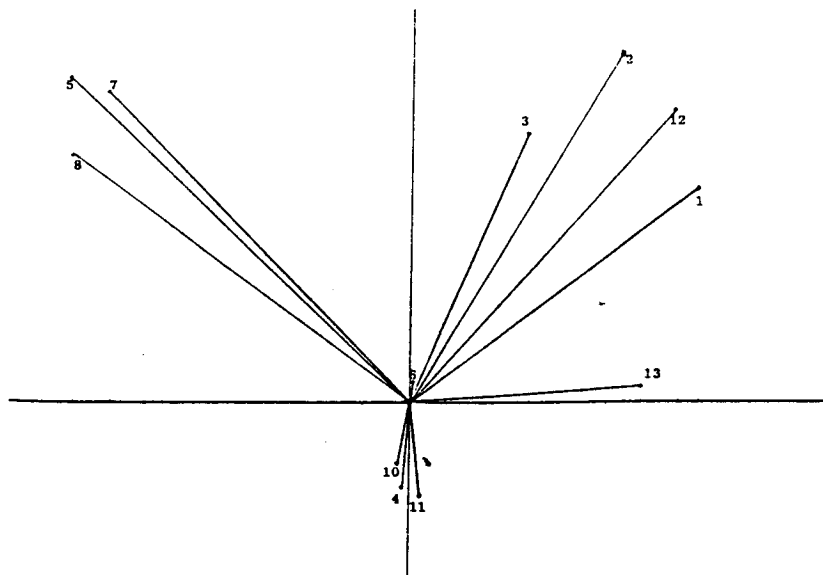
Figuur 3.2: Componentladingen COROP-variabelen op dimensie 1 en 2



De niet-lineaire PCA-oplossing geeft een duidelijker beeld dan de ordinale oplossing die Brouwer e.a. berekenden. Ter vergelijking is figuur 4³⁾ uit hun publikatie overgenomen (fig. 3.3).

Fig. 3.2 laat zien dat de eerste principale component sterk bepaald wordt door inkomen, bevolkingsdichtheid, industrialisatie, luchtvervuiling, dichtheid van het wegennet en (in iets mindere mate) werkloosheid en de afstand tot het 'centrum' van het land. In feite onderscheidt deze component dus de Randstad van de rest van het land. De educatieve en culturele en sport-voorzieningen laden voornamelijk op de tweede component. Samen beschrijven deze componenten 60% van de totale variantie in de 13 variabelen. Vanwege hun geringe ladingen op de eerste twee dimensies en het eveneens geringe belang van de derde component zijn de drie nog niet genoemde variabelen (welvaart, kosten van levensonderhoud en omvang van het natuurgebied) bij verdere analyses niet opgenomen.

Figuur 3.3: Componentladingen COROP-variabelen op dimensie 1 en 2 (volgens Brouwer, Nijkamp en Rietveld, 1983).



3.4.2 Extra variabelen

In Appendix II van het in 1.3 besproken RION-rapport (Boon e.a., 1983) worden de COROP-gebieden vergeleken naar aanbod van onderwijsvoorzieningen, deelname aan het onderwijs en groei in de participatie. In het algemeen gaat volgens de auteur van deze appendix "een hoge respectievelijk lage concentratie scholen steeds gepaard met een hoge respectievelijk lage deelname aan het desbetreffende onderwijsniveau" (Boon e.a., 1983, Appendix II, p. 18). Dit verband is sterker (vooral t.a.v. LBO-scholen) als de opleidingsmogelijkheden gerelateerd worden aan de populatie 12 - 19 jarigen, dan wanneer het aanbod t.o.v. de oppervlakte van de gebieden wordt uitgedrukt.

In de bijlage bij deze appendix zijn tabellen opgenomen van opleidingsmogelijkheden per COROP-gebied, uitgedrukt t.o.v. het landelijk gemiddelde en uitgesplitst naar LBO, MBO, HBO en AVO. De cijfers worden zowel gegeven per 100 km² als gerelateerd aan de bevolking in de categorie 12 - 19 jarigen.

Het leek zinvol de gegevens uit laatstgenoemde tabel in onze verdere analyse met variabelen op COROP-niveau te betrekken. De cijfers heb-

ben betrekking op het schooljaar 1978/'79. De waarden van de toegevoegde variabelen zijn in 5 categorieën gediscretiseerd.

3.4.3 Factoranalyse; 2 factoren

De dataset waarmee verder wordt gewerkt, bestaat, na verwijdering van de drie in 3.4.1 genoemde en toevoeging van de vier 'RION-variabelen' uit 14 variabelen over 40 COROP-gebieden.

We zouden, zoals in 3.4.1 gedaan is, de verschillen tussen COROP-gebieden kunnen weergeven in termen van een aantal principale componenten. Deze componenten zijn lineaire combinaties van de 14 kenmerken die de gebieden beschrijven. De eerste component is die lineaire combinatie die zo veel mogelijk variantie in de 14 kenmerken 'verklaart', de tweede component neemt een zo groot mogelijk deel van de overblijvende variantie voor zijn rekening, enzovoorts.

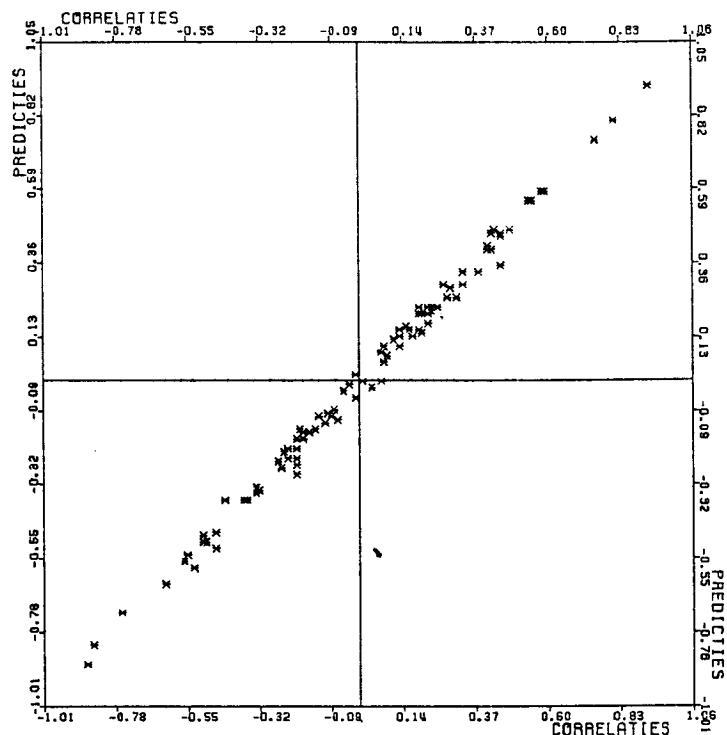
In 3.4.1 werd vastgesteld dat twee componenten voldoende zijn om 60% van de variantie in 13 variabelen te beschrijven en dat de overige componenten niet significant zijn, noch in inhoudelijke noch in statistische zin.

Een andere mogelijke techniek om verschillen of overeenkomsten tussen COROP-gebieden te beschrijven is factoranalyse. Er wordt dan geprobeerd de variantie van ieder van de kenmerken op te splitsen in een portie unieke variantie (behorend bij de variabele zelf) en een portie variantie die een variabele met alle andere in de set gemeenschappelijk heeft (gewoonlijk communaliteit genoemd). Tegelijkertijd proberen we de gemeenschappelijke variantie met zo weinig mogelijk orthogonale componenten te beschrijven.

In de praktijk betekent dit, dat in PCA gepoogd wordt de gehele correlatiematrix te beschrijven als het product van een klein aantal componenten. In factoranalyse zoeken we een zo goed mogelijke beschrijving van alleen de buiten-diagonale elementen van de correlatiematrix en we maken ons geen zorgen over de benadering van de diagonale elementen die immers toch altijd gelijk aan één zijn, en onafhankelijk van de eigenschappen van de gegevens.

We hebben de factoranalyse uitgevoerd met het speciaal voor dit doel geschreven APL-programma FACTALS. Het verschil met gebruikelijke factoranalyse programma's is, als steeds, dat we niet aannemen dat

Figuur 3.4: Correlaties en hun voorspellingen in een 2 factor structuur.



de regressies lineair zijn, en dat scorings voor de variabelen gekozen worden op zo'n manier dat één of ander rationeel criterium geoptimaliseerd wordt. In PRINCALS, onze vorm van principale componenten analyse, werden de schalingen zo gekozen dat de eerste twee componenten een zo groot mogelijk percentage van de totale variantie, d.w.z. van de gehele correlatiematrix beschrijven. In FACTALS moeten de schalingen zo gekozen worden dat twee factoren een zo groot mogelijk gedeelte van de gemeenschappelijke variantie, d.w.z. van alleen de buiten-diagonale elementen van de correlatiematrix, voor hun rekening nemen. Figuur 3.4 laat zien, dat de schalingen inderdaad gekozen kunnen worden op een voor FACTALS uitermate bevredigende manier. De correlaties tussen de geschaalde variabelen (x-as) verschillen vrijwel niet van de op grond van twee factoren voorspelde correlaties (y-as).

Figuur 3.5 laat bovendien zien dat de optimale quantificaties van PRINCALS en FACTALS, als de categorieën nominaal behandeld worden, maar weinig van elkaar verschillen. De variabelen waarvan de schalingen in FACTALS niet monotoon zijn hebben in PRINCALS minder goede fit. Hoewel principale componenten analyse en factoranalyse

dus verschillende technieken zijn, die ook verschillende soorten componenten produceren, lijkt het erop alsof de optimale transformaties van de variabelen voor beide technieken grotendeels overeenkomstig zijn. De relaties tussen de 14 COROP-kenmerken laten zich in het vlak van 2 orthogonale componenten of factoren goed beschrijven.

De factorladingen staan in tabel 3.2 en zijn grafisch weergegeven in figuur 3.6³⁾.

De beide factoren hebben eigenwaarden van 4.92 en 3.28 en beschrijven de gemeenschappelijke variantie volledig (figuur 3.3) en 'verklaaren' respectievelijk 35 en 16% van de totale variantie in de dataset.

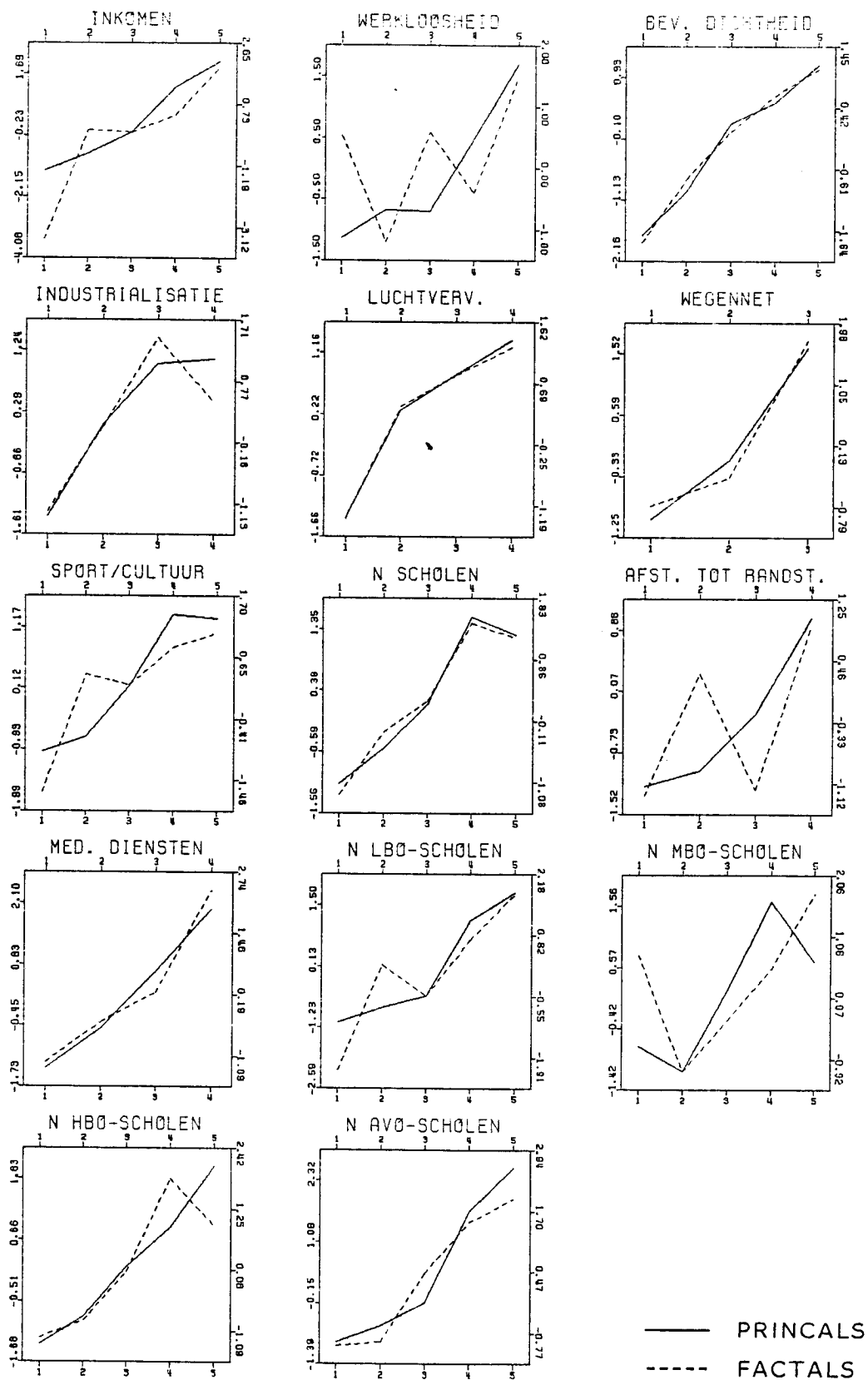
Tabel 3.2: Factorladingen 2 orthogonale factoren; 14 variabelen op COROP-niveau.

Variabele:	Factor 1	Factor 2
1. Inkomen per hoofd van de bevolking	-.63	.05
2. Percentage werkloosheid	-.38	-.12
3. Bevolkingsdichtheid	.99	-.06
4. Industrialisatie	.81	-.14
5. Luchtvervuiling	.93	.01
6. Dichtheid van het wegennet	-.89	-.03
7. Aantal scholen	.12	.55
8. Aantal sport- en culturele voorzieningen	.22	.80
9. Afstand tot de Randstad	-.50	-.27
10. Omvang van medische dienstverlening	-.54	.41
11. Aantal LBO-scholen (t.o.v. het land. gemid.)	.60	.13
12. Aantal MBO-scholen (idem)	.28	.35
13. Aantal HBO-scholen (idem)	-.17	.75
14. Aantal scholen voor AVO (idem)	-.13	.55

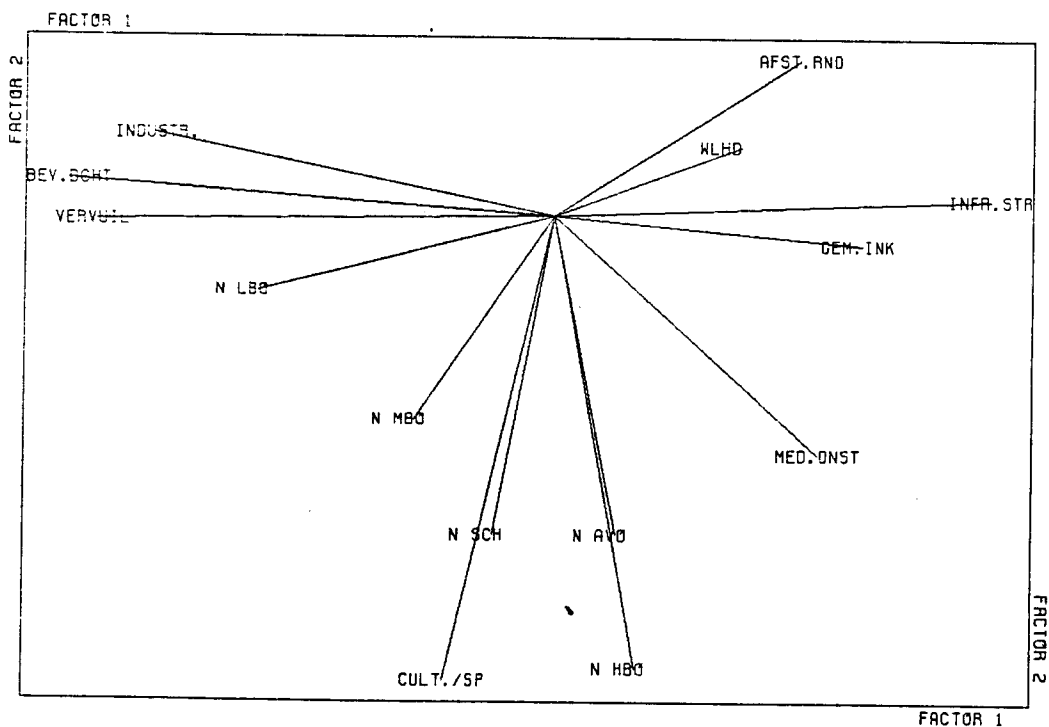
De eerste factor geeft, nog duidelijker dan in de PCA-oplossing van 3.4.1, de tegenstelling tussen het westen en de rest van het land weer. Dit is vooral een tegenstelling in bevolkingsdichtheid. De lading van deze variabele op de eerste factor is .99. De educatieve, sport- en culturele voorzieningen hangen samen met de 2e factor. Opmerkelijk is de positie van het aantal LBO-scholen. Deze variabele correleert vooral met de eerste factor.

Kijken we terug naar de oorspronkelijke gegevens, dan zien we dat met name voor LBO-scholen geldt, dat in relatief dun bevolkte gebieden een relatief hoog aanbod van onderwijsvoorzieningen t.o.v. de populatie 12-19 jarigen bestaat (RION, 1983, Appendix II, p. 16).

Figuur 3.5: Transformaties van 14 kenmerken op COROP-niveau uit PRINCALS en FACTALS (nominale oplossingen).



Figuur 3.6: Factorladingen 2 factoren, 14 variabelen op COROP-niveau.

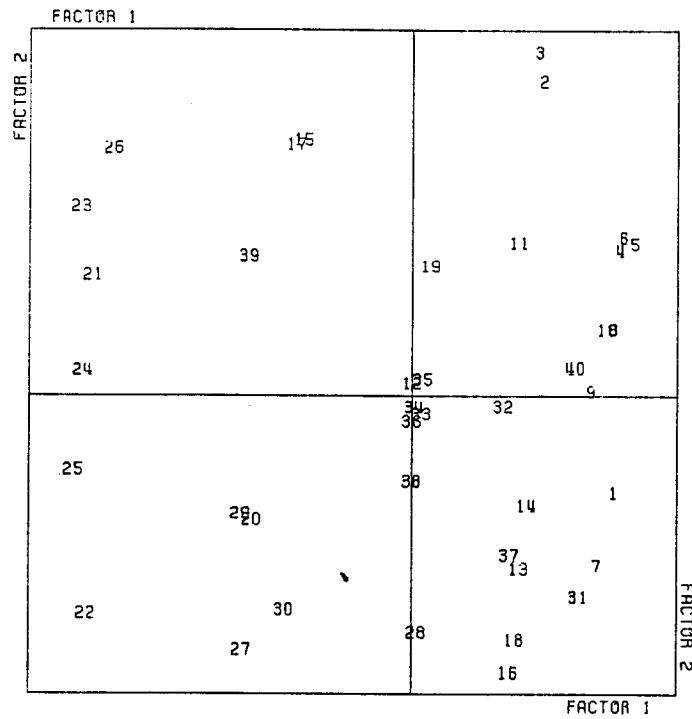


Op basis van de factorladingen en de correlatie matrix kan de score van elk van de COROP-gebieden op de beide factoren worden berekend.

De factorscores zijn weergegeven in figuur 3.7. De nummering van de gebieden correspondeert met die in figuur 2.1. In de volgende paragraaf worden deze factorscores met de gevonden relaties per COROP-gebied in verband gebracht.

Duidelijk is in figuur 3.7 te zien hoe de dunbevolkte gebieden in het noorden en oosten van het land (rechts in de figuur met hoge scores op de eerste factor) onderscheiden worden van de dichtbevolkte, ge-industrialiseerde gebieden met een fijnmazig wegennet in het westen van het land. Bovenaan in de figuur liggen de gebieden met relatief veel educatieve, culturele en sportvoorzieningen, onderaan de in dit opzicht minder bedeelde gebieden.

Figuur 3.7: Factorscores 40 COROP-gebieden



3.5 De relatie tussen kenmerken op COROP-niveau en de SMVO-variabelen.

Door middel van de in de vorige paragraaf besproken factoranalyse hebben we een manier gevonden om de verschillen tussen de COROP-gebieden te quantificeren. Indachtig de steeds gevolgde 2-staps procedure kunnen nu deze verschillen aan de individuele schoolloopbanen worden gerelateerd.

Het onderscheid tussen jongens en meisjes laten we daarbij vervallen. Op de eerste plaats zijn deze verschillen voor wat betreft de door ons gehanteerde variabelen gering (Meester en De Leeuw, 1983). Ten tweede mag aangenomen worden dat de verdeling van jongens en meisjes over de COROP-gebieden betrekkelijk evenredig zal zijn (vgl. tabel, bijlage, 2.1). En bovendien zijn de kenmerken op COROP-niveau ook niet voor de geslachtscategorieën apart berekend, zo dat al doenlijk zou zijn.

Er moeten dus nieuwe quantificaties voor de achtergrond- en schoolloopbaanvariabelen berekend worden. Dat gebeurt op dezelfde manier als in de paragrafen 3.2 en 3.3 met het programma PREHOM dat quantificaties voor de categorieën van de variabelen zoekt, zodanig

dat de regressies zoveel mogelijk worden gelineariseerd.

We berekenen de quantificaties over de gehele dataset (jongens en meisjes samengevoegd). Op basis van deze quantificaties wordt weer per COROP-gebied een correlatiematrix van BVA en de schoolloopbaanvariabelen PRE, ADV, KEUS, EIN en bovendien de som van de gequantificeerde schoolvariabelen (SCH) berekend.

We kunnen dan het verband tussen de gemiddelden en de 15 correlatiecoëfficiënten per COROP-gebied enerzijds en de scores op de eerste en tweede factor anderzijds in beeld brengen.

3.5.1 Gemiddelden van gequantificeerde SMVO-variabelen versus COROP-factoren.

De gemiddelden van de gequantificeerde SMVO-variabelen zijn de figuren 3.8 t/m 3.31 uitgezet tegen de scores op de beide COROP-factoren. De gemiddelden zijn daartoe eerst gestandaardiseerd. De factorscores zijn uiteraard al standaardscores. Omdat deze analyse op COROP-niveau gebeurt, bevat iedere plot slechts 40 punten. Linearisering van de regressie is daardoor niet goed mogelijk (zie ook 3.3). De sterkte van het verband wordt hier weergegeven door de hoek die de in de figuren getrokken regressielijnen met elkaar maken.

Naarmate het verband sterker is, is de kleinste hoek die de regressielijnen maken, scherper. Als de correlatiecoëfficiënt 0 is staan de lijnen loodrecht op elkaar. De voor elk paar van variabelen berekende correlatiecoëfficiënt is steeds in de kop van de figuur vermeld.

Voor de duidelijkheid zij nog eens vermeld dat dun bevolkte, weinig geïndustrialiseerde COROP-gebieden hoge scores op de eerste factor hebben en dat hoge scores op de tweede factor samengaan met een relatief hoog aanbod van educatieve, sport- en culturele voorzieningen. De eerste factor beschrijft ca 2/3 van de gemeenschappelijke variantie in de COROP-kenmerken.

De COROP-gebieden zijn aangeduid met nummers die corresponderen met het in figuur 2.1 (blz. 18) gegeven overzicht.

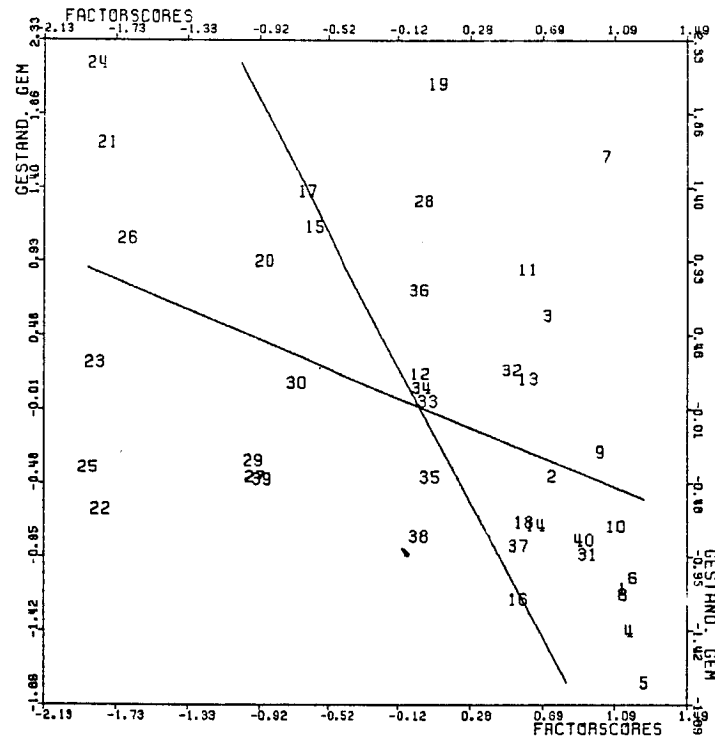
De figuren 3.8, 3.9, 3.12 en 3.14 laten zien dat in de relatief dunbevolkte gebieden buiten de Randstad de milieu-samenstelling gemiddeld lager is, gemiddeld minder hoog op de prestatietoets wordt gescoord

en lagere adviezen worden gegeven en meer voor het lager beroeps-
onderwijs en de MAVO wordt gekozen. In het niveau dat in 1982 in
het v.o. wordt bereikt (fig. 3.16) komt dat niet tot uiting. Mogelijk
is de doorstroming na LBO of MAVO buiten de Randstad groter.

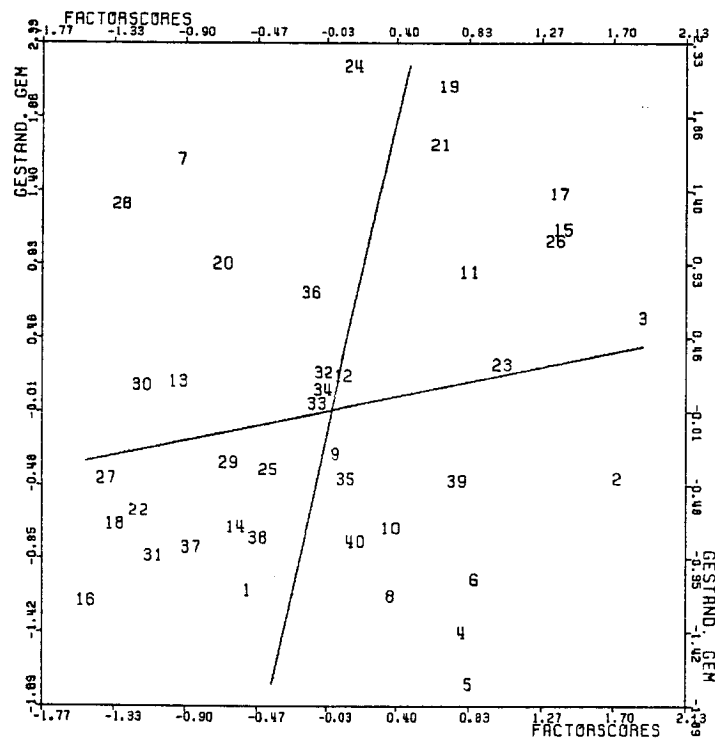
De optelling van de schoolloopbaanvariabelen (figuur 3.18) wordt
voornamelijk door PRE, ADV en KEUS bepaald. Dat resulteert dan
ook in een figuur die sterk overeenkomt met de grafieken voor deze
variabelen afzonderlijk.

Bezien we het verband tussen de SMVO-variabelen en het aanbod van
educatieve en culturele voorzieningen (COROP-factor 2, figuur 3.9,
3.11, 3.13, 3.15, 3.17 en 3.19), krijgen we een ander beeld. In ge-
bieden met een relatief hoog aanbod van deze voorzieningen zijn de
leerlingen meer afkomstig uit de hogere sociale milieus, hebben gemid-
deld wat lagere prestatiescores en behalen na 5 jaar een hoger niveau
in het voortgezet onderwijs. Schooladviezen en -keuzen hangen even-
als de optelling van de loopbaanvariabelen vrijwel niet samen met de
tweede factor. Overigens moet wel in gedachten worden gehouden dat
de tweede factor maar een betrekkelijk gering gedeelte van de varian-
tie in de COROP-kenmerken voor zijn rekening neemt. Dat komt ook
door de sterke onderlinge samenhang van de variabelen die op de
eerste factor laden en het relatief grote aantal van deze demografische
variabelen in de dataset. De door de figuren weergegeven verbanden
zijn echter geen van alle erg sterk. De meest uitgesproken samenhang
is er nog met de milieusamenstelling van de COROP-gebieden.

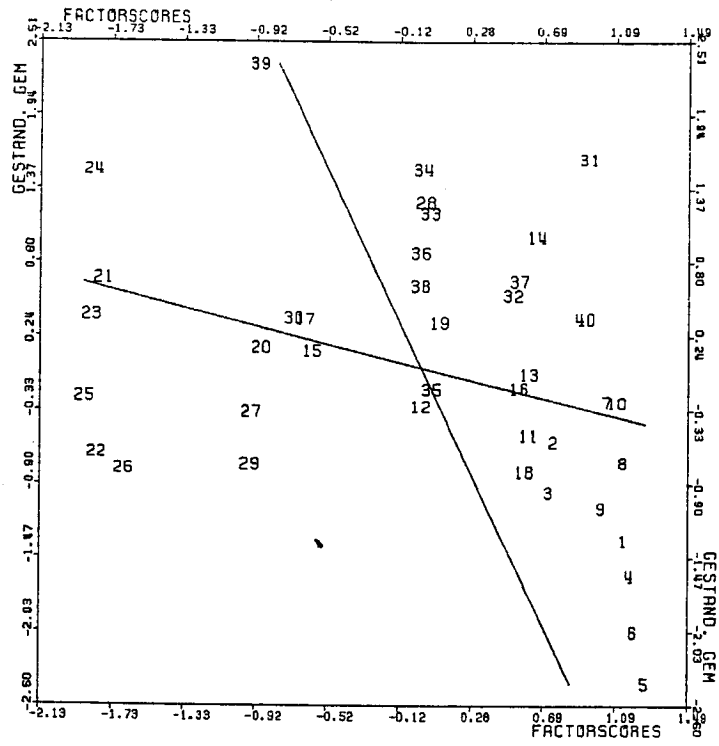
Figuur 3.8: Gemiddelde quantificaties BVA versus factor 1 voor 40 COROP-gebieden; corr.: -.47



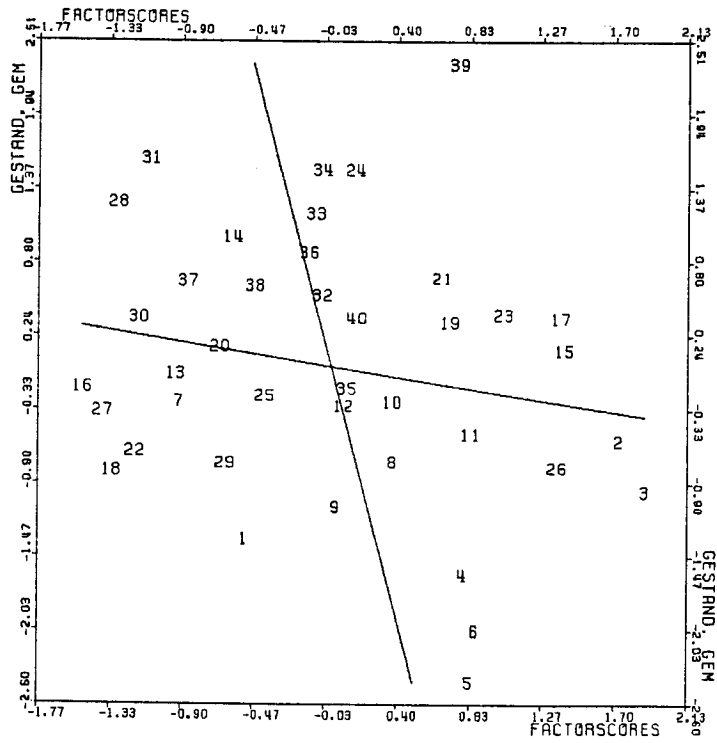
Figuur 3.9: Gemiddelde quantificaties BVA versus factor 2 voor 40 COROP-gebieden; corr.: .22



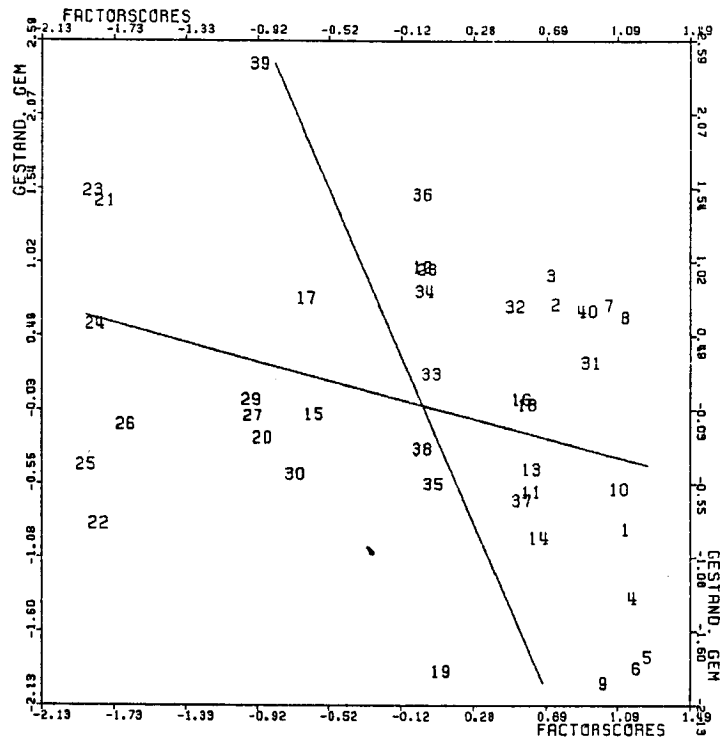
Figuur 3.10: Gemiddelde quantificaties PRE versus factor 1 voor 40 COROP-gebieden; corr.: -0.35



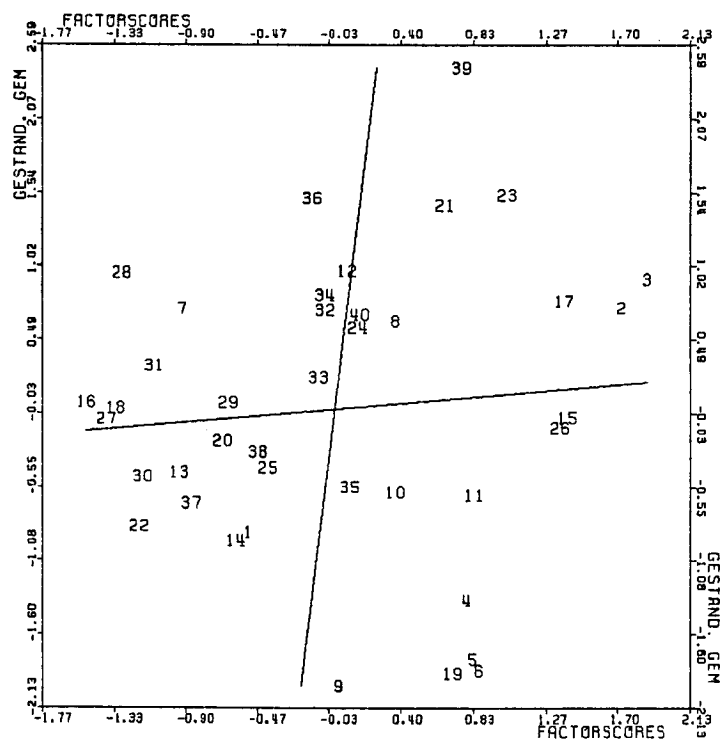
Figuur 3.11: Gemiddelde quantificaties PRE versus factor 2 voor 40 COROP-gebieden; corr.: -0.21



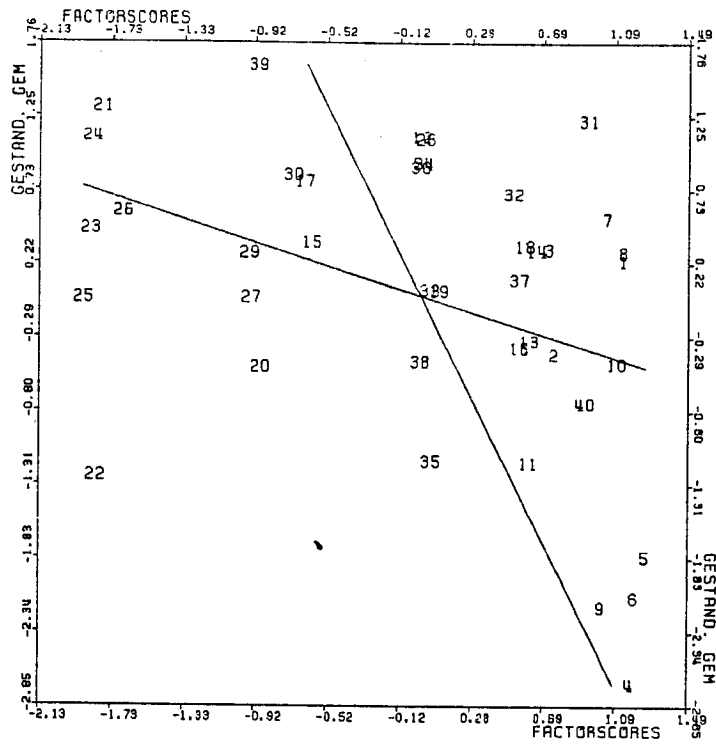
Figuur 3.12: Gemiddelde quantificaties ADV versus factor 1 voor 40 COROP-gebieden; corr.: -0.34



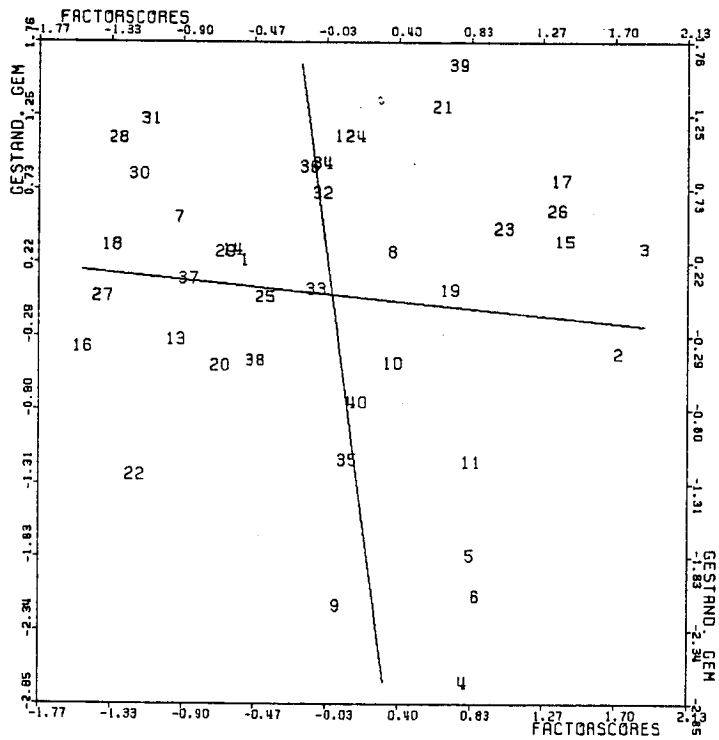
Figuur 3.13: Gemiddelde quantificaties ADV versus factor 2 voor 40 COROP-gebieden; corr.: 0.10



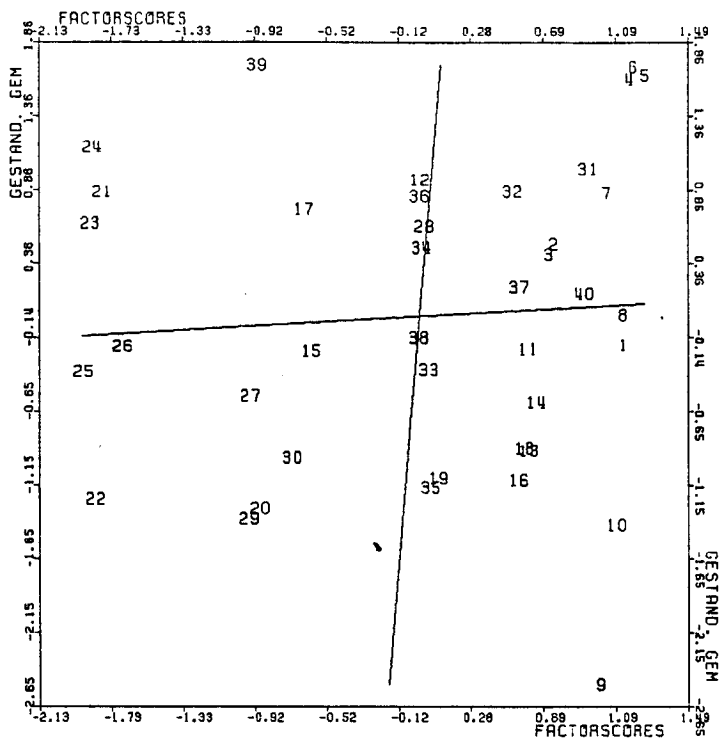
Figuur 3.14: Gemiddelde quantificaties KEUS versus factor 1 voor 40 COROP-gebieden; corr.: $-.39$



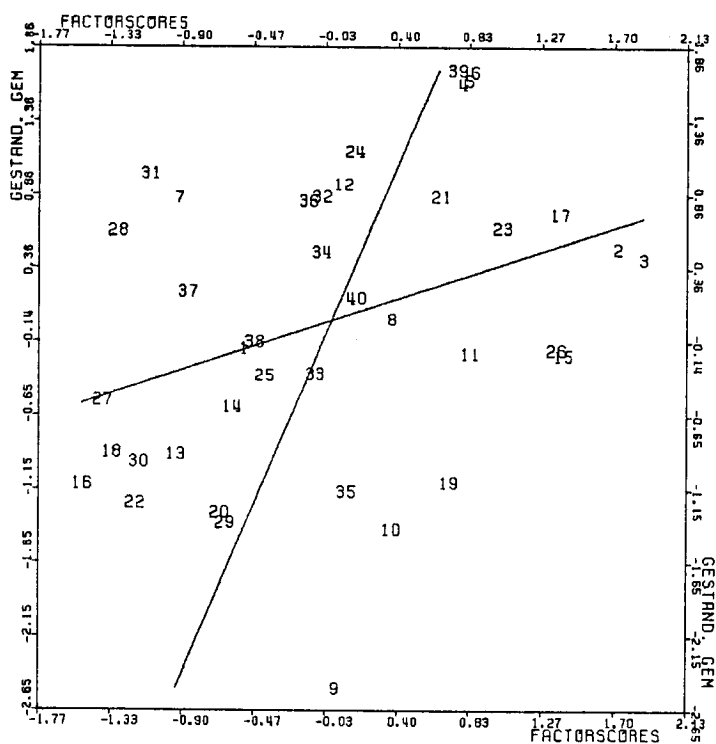
Figuur 3.15: Gemiddelde quantificaties KEUS versus factor 2 voor 40 COROP-gebieden; corr.: $-.12$



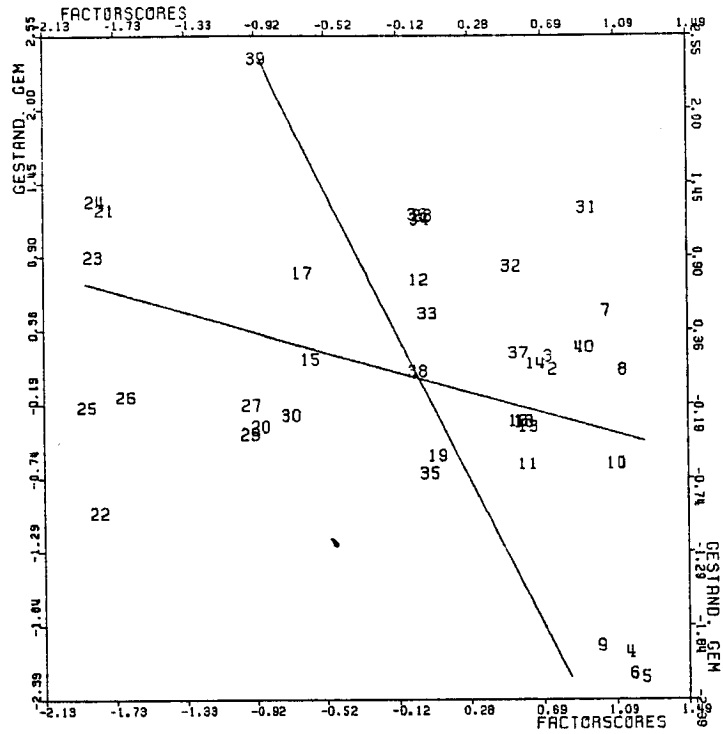
Figuur 3.16: Gemiddelde quantificaties EIN versus factor 1 voor 40 COROP-gebieden; corr.: .07



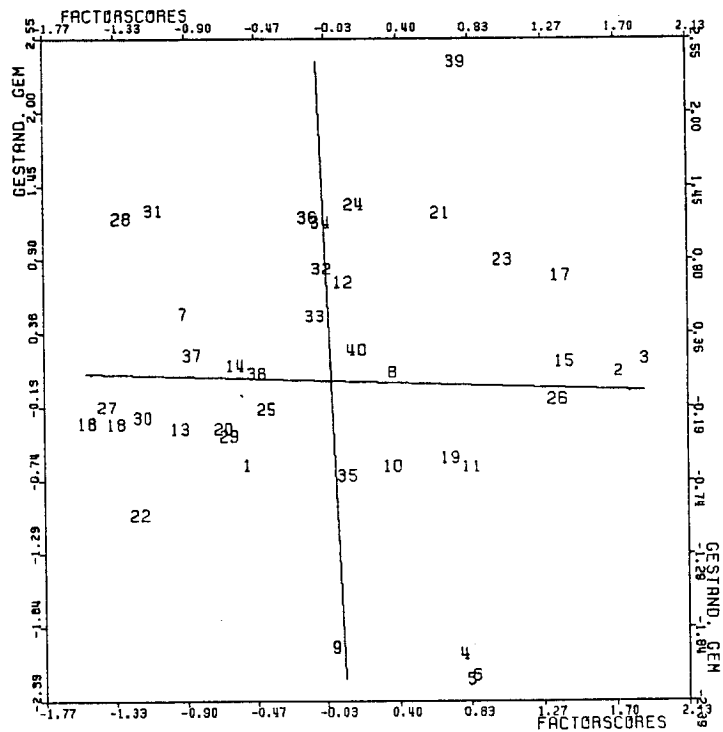
Figuur 3.17: Gemiddelde quantificaties EIN versus factor 2 voor 40 COROP-gebieden; corr.: .37



Figuur 3.18: Gemiddelde quantificaties SCH versus factor 1 voor 40 COROP-gebieden; corr.: $-.38$



Figuur 3.19: Gemiddelde quantificaties SCH versus factor 2 voor 40 COROP-gebieden; corr.: $-.04$



3.5.2 De samenhang van de correlaties tussen SMVO-variabelen en de COROP-factorscores.

Op dezelfde manier als in de vorige paragraaf voor de gemiddelden van de SMVO-variabelen gebeurd is, kunnen ook de onderlinge relaties tussen BVA, PRE, ADV, KEUS, EIN en SCH met de COROP-factoren in verband worden gebracht.

We geven eerst de correlatiematrix tussen optimaal gequantificeerde variabelen over alle leerlingen in het SMVO-bestand in tabel 3.3. Deze tabel geeft een indruk van de (over alle COROP-gebieden gemiddelde) sterkte van de samenhangen.

Tabel 3.3 Correlaties tussen optimaal getransformeerde versies van achtergrond- en schoolloopbaanvariabelen uit het SMVO-bestand.

	BVA	PRE	ADV	KEUS	EIN
BVA	*				
PRE	.24	*			
ADV	.31	.65	*		
KEUS	.33	.63	.82	*	
EIN	.31	.60	.72	.77	*

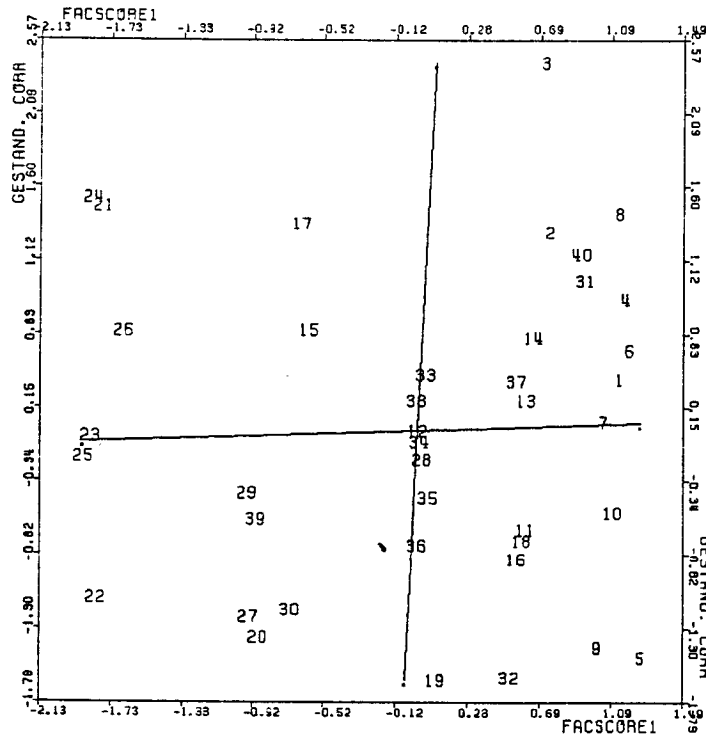
De correlaties per COROP-gebied zijn (net als de gemiddelden in 3.5.1) gestandaardiseerd en uitgezet tegen de scores op de COROP-factoren. De figuren 3.20 t/m 3.29 tonen het verband tussen de correlaties van BVA met de schoolloopbaanvariabelen en de COROP-typeringen. De op zichzelf niet erg sterke samenhang tussen het milieu van herkomst en de schoolloopbaan (zie tabel 3.3) houdt niet systematisch verband met de tegenstelling tussen de Randstad en de rest van het land (figuur 3.20, 3.22, 3.24, 3.26 en 3.28). De regressielijnen staan alle vrijwel loodrecht op elkaar. Wel lijkt er een verband te zijn met COROP-factor 2, het aanbod van educatieve en culturele voorzieningen (de oneven genummerde figuren).

Naarmate in een COROP-gebied meer aanbod van deze voorzieningen is, is de samenhang tussen het milieu van herkomst en de schoolloopbaan sterker. Dit is een moeilijk te interpreteren resultaat. Gebieden die hoog op de tweede factor scoren (waar dus een relatief hoog aanbod van educatieve, culturele en sportvoorzieningen is) zijn Noord-Groningen (inclusief Groningen-stad), Delfzijl e.o., Utrecht, Den Haag e.o.

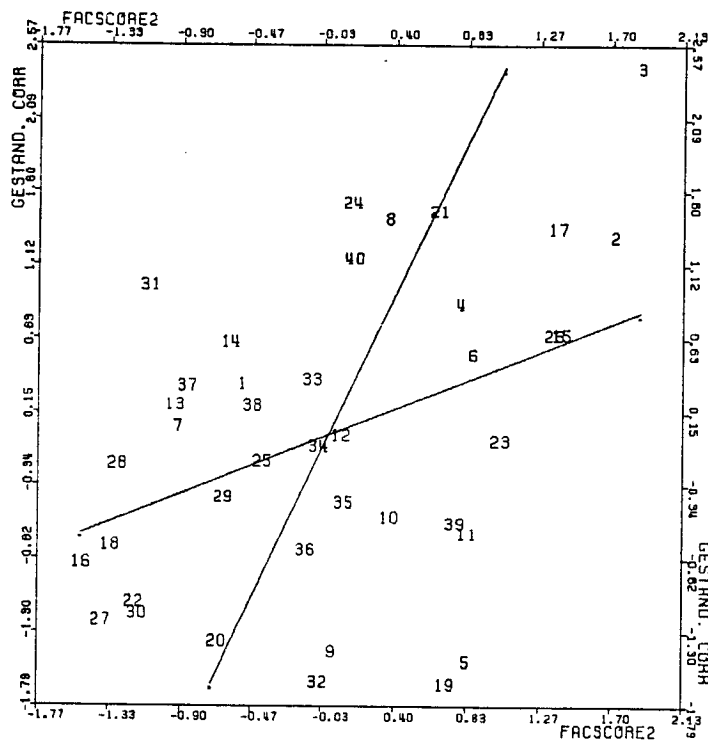
en Arnhem en Nijmegen. Lage scores op de tweede factor hebben de Betuwe, de kop van Noord-Holland, het Westland, de Zaanstreek, Oostelijk Zuid-Holland en de Krimpenerwaard. Op Delfzijl e.o. na, is de milieu-samenstelling in de eerstgenoemde gebieden hoger dan het gemiddelde; op oostelijk Zuid-Holland na, is in de laatstgenoemde gebieden de milieu-samenstelling gemiddeld laag. In de volgende paragraaf wordt de samenhang tussen sociaal milieu en geslacht enerzijds en de schoolloopbaanvariabelen anderzijds in de verschillende COROP-gebieden, als aanvulling op de loglineaire analyses die in 3.2 werden gerapporteerd, nader onderzocht. De overige grafieken die de correlaties tussen de schoolloopbaanvariabelen onderling tegen de COROP-factorscores laten zien, zijn nauwelijks interessant. Voor de volledigheid zijn ze opgenomen in bijlage 3.5. De samenhangen zijn zwak tot afwezig en lijken in enkele gevallen door 'uitbijters' te worden bepaald. Voor COROP-gebied nr. 40 (Zuidelijke IJsselmeerpolders) bijvoorbeeld, is dat in het grootste gedeelte van de grafieken het geval. De voor dit gebied over een slechts gering aantal leerlingen (84) berekende correlaties wijken sterk af van die voor de andere gebieden die onderling in dit opzicht slechts geringe spreiding vertonen.

De voorlopige conclusie van paragraaf 3.1 dat de verschillen tussen de COROP-gebieden voornamelijk bepaald worden door verschillen in de milieucompositie van de gebieden, lijkt door het hier gevonden resultaat te worden bevestigd.

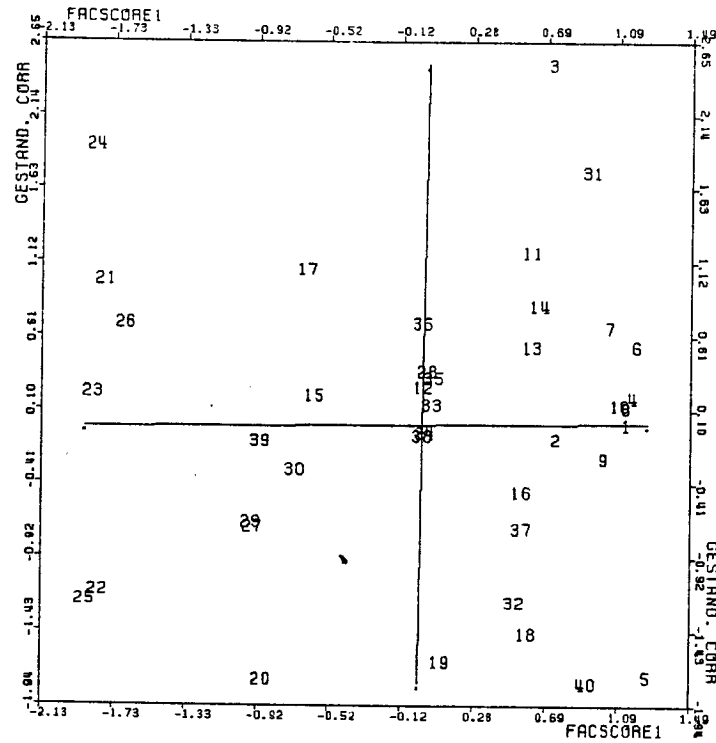
Figuur 3.20: Correlatie BVA-PRE versus factor 1, 40 COROP-gebieden;
corr.: .04.



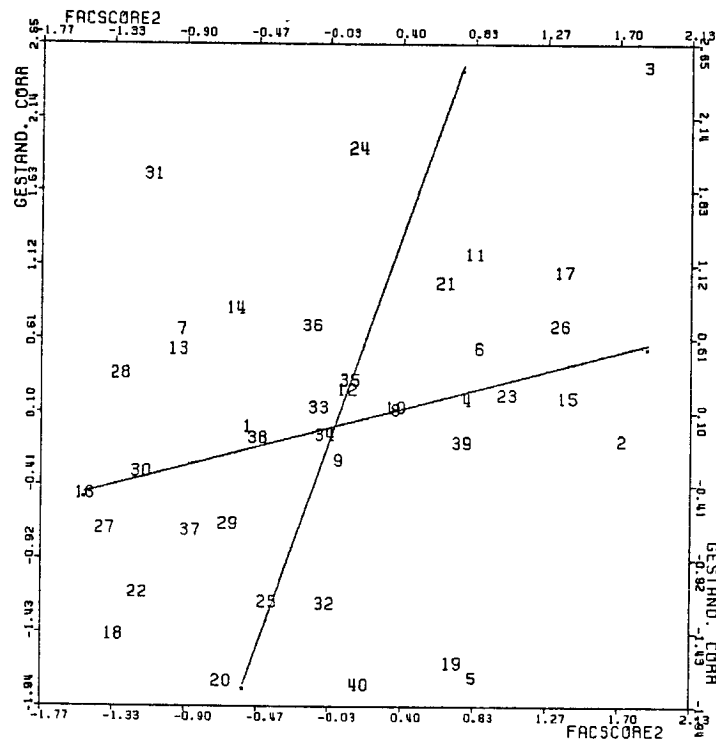
Figuur 3.21: Correlatie BVA-PRE versus factor 2, 40 COROP-gebieden;
corr.: .43.



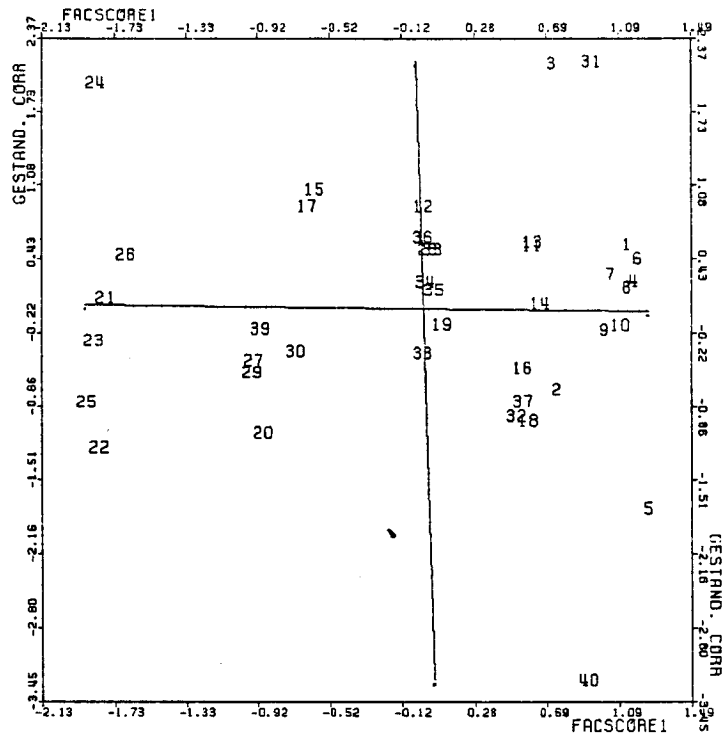
Figuur 3.22: Correlatie BVA-ADV versus factor 1, 40 COROP-gebieden; corr.: .01.



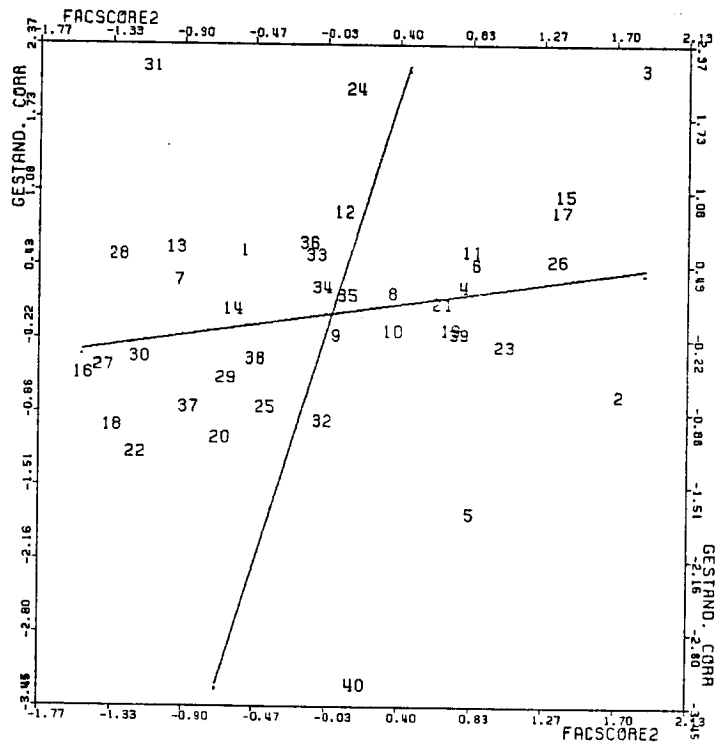
Figuur 3.23: Correlatie BVA-ADV versus factor 2, 40 COROP-gebieden; corr.: .31.



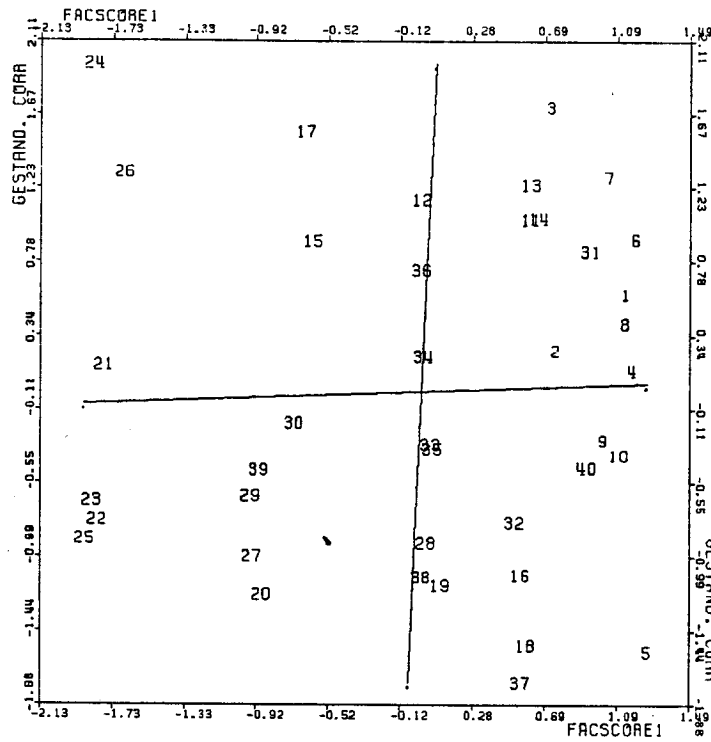
Figuur 3.24: Correlatie BVA-KEUS versus factor 1, 40 COROP-gebieden; corr.: -.02.



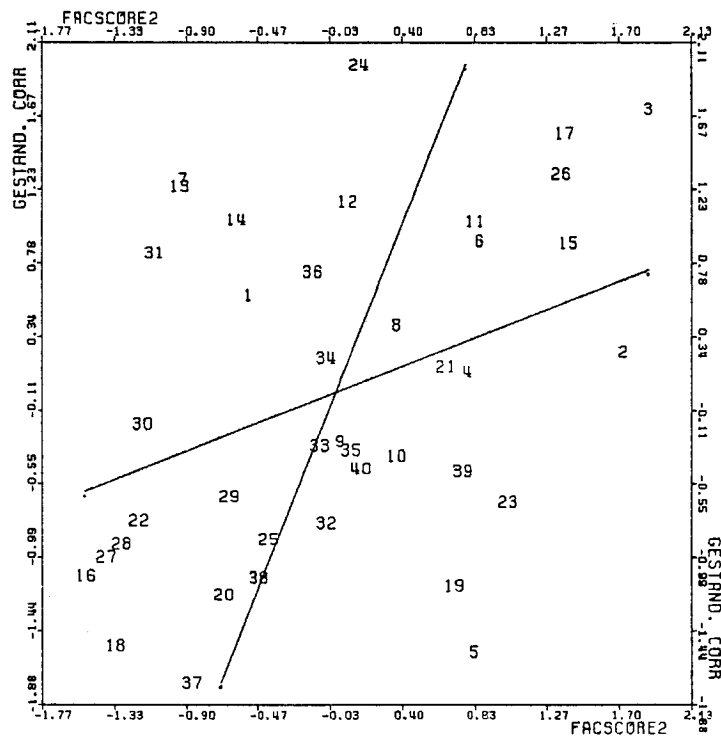
Figuur 3.25: Correlatie BVA-KEUS versus factor 2, 40 COROP-gebieden; corr.: .21.



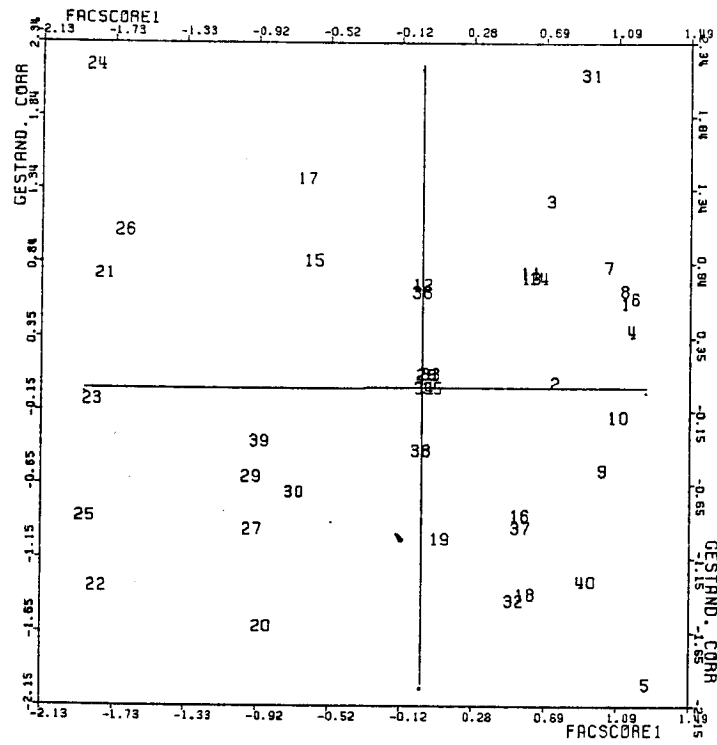
Figuur 3.26: Correlatie BVA-EIN versus factor 1, 40 COROP-gebieden; corr.: .04.



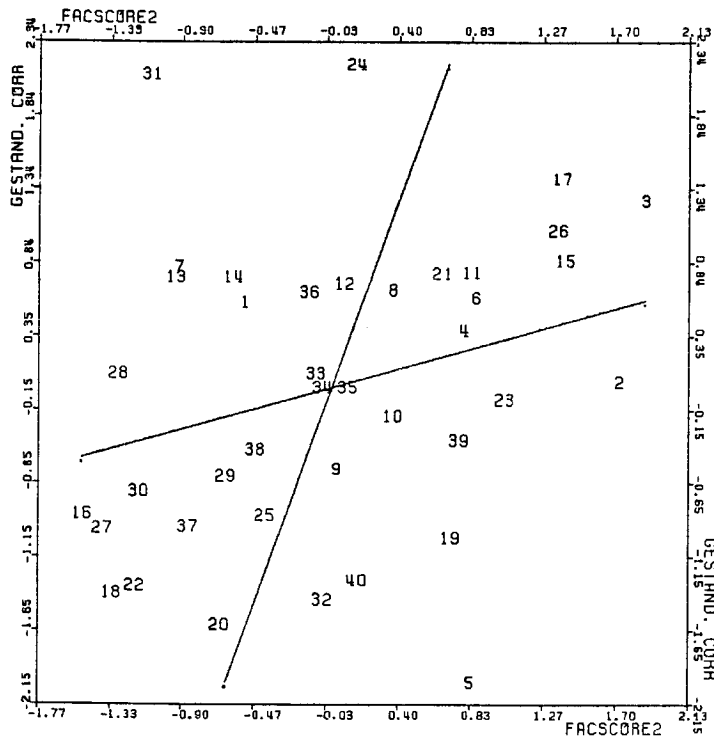
Figuur 3.27: Correlatie BVA-EIN versus factor 2, 40 COROP-gebieden; corr.: .40.



Figuur 3.28: Correlatie BVA-SCH versus factor 1, 40 COROP-gebieden; corr.: .00.



Figuur 3.29: Correlatie BVA-SCH versus factor 2, 40 COROP-gebieden; corr.: .32.



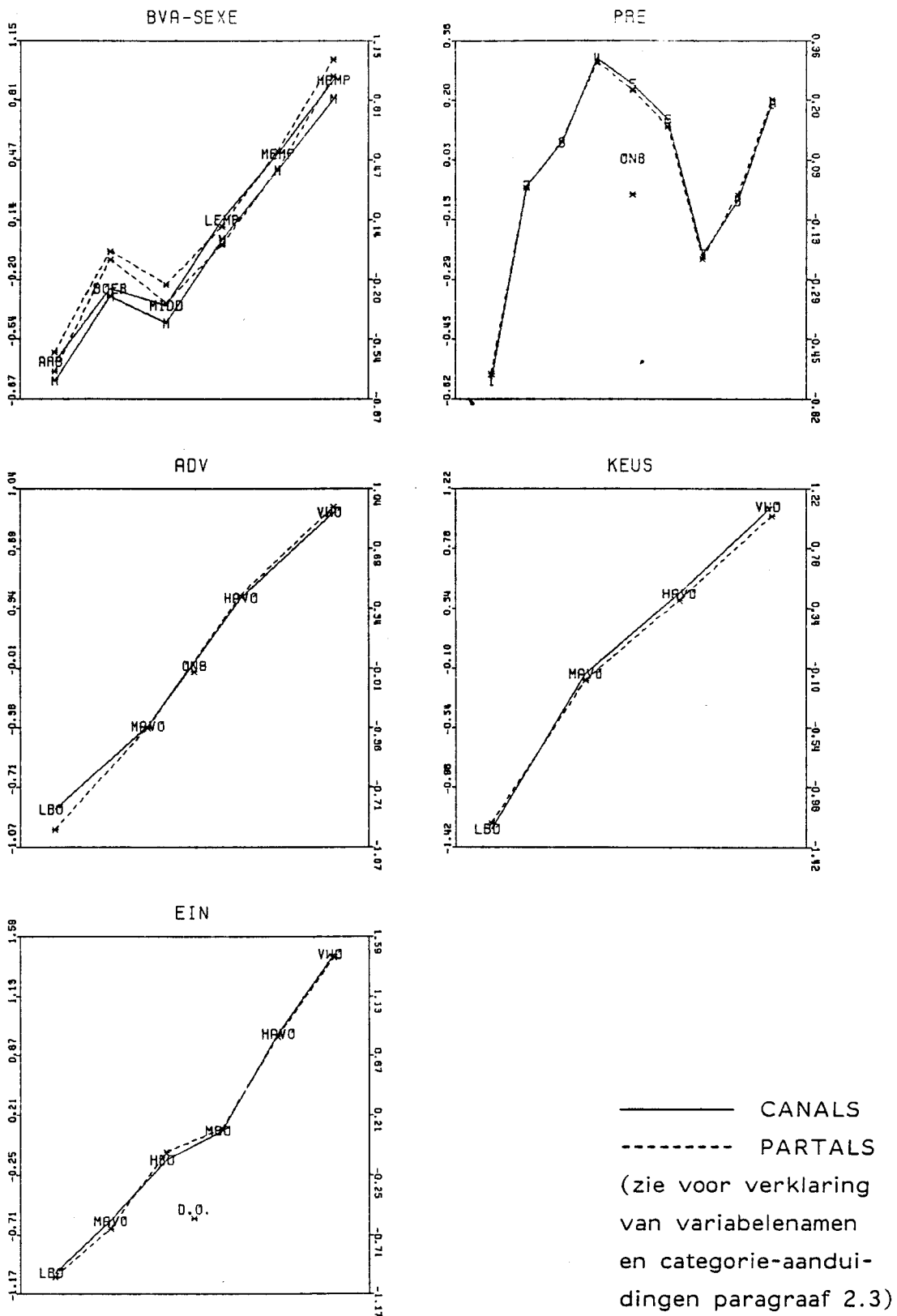
3.6 Canonische correlatie-analyse

In de paragrafen 3.2 en 3.5 hebben we al de samenhang tussen achtergrondvariabelen en schoolloopbaan bestudeerd, daarbij rekening houdend met de regionale variabelen. Net als in eerdere analyses (paragraaf 3.1; Meester en De Leeuw, 1983) worden BVA en SEXE tot één interactieve variabele samengevoegd.

Eerst wordt nu de samenhang tussen milieu-sexe en de vier schoolloopbaanvariabelen wat meer in detail bekeken. Dit gebeurt met behulp van het programma CANALS (Van der Burg en De Leeuw, 1983; Van der Burg, 1983). In 3.2 kozen we de quantificaties van de categorieën van ieder van de variabelen op zo'n manier dat de variabelen zo lineair mogelijk werden, dat wil zeggen zodanig dat de regressies zo goed mogelijk gelineariseerd werden. Dit werd gedaan m.b.v. het programma PREHOM. (zie noot 1). In 3.4 kozen we quantificaties die het mogelijk maakten de totale variantie (PRINCALS) of de gemeenschappelijke variantie (FACTALS) van de regio-variabelen zo volledig mogelijk in twee dimensies te representeren. En nu kiezen we quantificaties op zo'n manier dat de correlatie tussen de achtergrond- en schoolloopbaanvariabelen, de zogenaamde canonische correlatie van de beide sets, zo groot mogelijk wordt. Het mag dus duidelijk zijn dat de keuze van de quantificatie afhangt van de specifieke vraagstelling die we hebben, evenals bijvoorbeeld de keuze van het model of de techniek van de vraagstelling afhangt. Men kan overigens aantonen, dat onder bepaalde voorwaarden, die in dit soort onderzoek dikwijls behoorlijk goed opgaan, al dit soort optimale quantificatie technieken dezelfde quantificatie van de categorieën opleveren (met steekproefafwijkingen, vanzelfsprekend). Vergelijk De Leeuw (1982b).

De categorie-quantificaties uit CANALS staan in figuur 3.30. Wanneer deze figuur wordt vergeleken met figuur 3.1 (de transformaties uit PREHOM) dan is goed te zien hoe de variabelen vrijwel hetzelfde worden gequantificeerd. De deelfiguur voor PRE lijkt daarbij nogal af te wijken, maar nauwkeurige bestudering van de schaal (langs de verticale as) laat zien dat de verschillen tussen de categorieën miniem zijn. Verschillen in scores op de prestatietoets dragen in deze analyse weinig bij.

Figuur 3.30: Optimale transformaties van SMVO-variabelen uit CANALS en PARTALS



Niettemin is het, gegeven de vraagstelling, onbevredigend dat de CANALS analyse op geen enkele manier rekening houdt met regio. Dit probleem is in principe heel eenvoudig op te lossen. Immers, als voor ieder van de 40 COROP-gebieden een afzonderlijke CANALS-oplossing wordt berekend, dan wordt als het ware voor regio gecorrigeerd. Vergelijking van de canonische correlatie binnen COROP-gebieden met de canonische correlatie over het totale bestand (deze is .3525) laat zien in hoeverre variatie tussen COROP-gebieden een rol speelt in het bepalen van de correlatie.

Helaas werkt dit eenvoudige plan maar zeer ten dele. Het is uitvoerbaar voor sommige van de grote gebieden. Zo vinden we bijvoorbeeld voor gebied nr. 25 (653 leerlingen) een correlatie van .364 en voor gebied nr. 23 (1657 leerlingen) een correlatie van .369 bij transformaties die veel lijken op de in figuur 3.30 afgebeelde. Maar bij de interpretatie van deze bevindingen moet men rekening houden met het feit dat de gevonden correlatie niet alleen de sterkte van de samenhang, maar ook de grootte van de steekproef weerspiegelt. Bij kleine steekproeven is er grotere kans op 'kanskapitalisatie', waarbij toevallige eigenschappen van de steekproef tot grote correlatiecoëfficiënten kunnen leiden die per definitie weinig betrouwbaar zijn. (Bijvoorbeeld: gebied nr. 1: 281 leerlingen, canonische correlatie .486 of gebied nr. 2: 140 leerlingen, canonische correlatie .630). Voor kleine COROP-gebieden vinden we dus hoge canonische correlaties, die gepaard gaan met niet-intuïtieve quantificaties van de categorieën. Net als in 3.2 toont dit aan dat analyse met optimale schalingsmethoden per COROP-gebied teveel parameters probeert aan te passen met een daaruit voortvloeiend gebrek aan stabiliteit. In 3.3 en 3.5 hebben we dit voorkomen door één enkele categorie-quantificatie per variabele te gebruiken voor alle COROP-gebieden. Hier wordt iets dergelijks gedaan. We hebben, speciaal voor dit onderzoek, een vorm van partiële niet-lineaire canonische correlatie analyse ontwikkeld, die het verband tussen achtergrond en schoolloopbaan beschrijft, gecorrigeerd voor COROP-gebied. De techniek wordt in appendix B kort beschreven.

Bij gebruikmaking van deze methode die we PARTALS noemen, vinden we een partiële canonische correlatie van .3516 en categoriequantificaties die in figuur 3.30 met gebroken lijnen verbonden zijn. Duidelijk is te zien dat de quantificaties van CANALS en PARTALS vrijwel het-

zelfde zijn. In de figuur voor BVA-SEXE zouden eventuele verschillen tussen de geslachten tot uitdrukking moeten komen. Ook die zijn vrijwel afwezig. Een gering verschil tussen CANALS- en PARTALS-transformaties zit in de categorie 'boeren en tuinders'. Dat moet aan de relatief geringe vulling van deze categorie worden toegeschreven.

Het bij deze analyse gevonden resultaat bevestigt nogmaals duidelijk hetgeen we in eerdere instantie met loglineaire analyse vonden.

4 DISCUSSIE EN AFSLUITING

Het duidelijk resultaat van de in het vorige hoofdstuk gerapporteerde statistische exercities maakt het mogelijk dit afsluitende hoofdstuk betrekkelijk kort te houden.

Alvorens nog eens op de resultaten in te gaan wordt in de volgende paragraaf voor het 'historisch perspectief' enige aandacht besteed aan de heranalyse van dienstkeuringsgegevens uit de jaren 1925, 1932, 1952 en 1982.

4.1 Het intelligentie-onderzoek bij de militaire keuringen¹⁾

In Meester en De Leeuw (1983, 1984a) wordt verslag gedaan van heranalyses van dienstkeuringsgegevens over twee vóór- en vier na-oorlogse jaren. Deze analyses spitsten zich toe op het verband tussen intelligentiescores en bereikt opleidingsniveau. Geconcludeerd kon worden dat het intelligentie-onderzoek bij de militaire keuringen, globaal genomen, het opleidingsniveau van de gekeurden meet en dat dit verband voor de betrokken jaren betrekkelijk invariant is.

In de publikaties van Groenewegen (1926), CBS (1935) en Demologisch Instituut (1957) zijn ook tabellen opgenomen m.b.t. de regionale herkomst van de keurlingen, d.w.z. de keurings- of indelingsraad waar de keuring heeft plaatsgevonden. Dat laatste geldt ook voor het in 1982 verzamelde bestand waarin intelligentiegegevens van gekeurden en hun vaders zijn bijeengebracht²⁾. Het aantal keuringscentra is in de loop der jaren sterk teruggebracht. Dit bemoeilijkt de vergelijking, maar dat is naar onze mening niet van wezenlijk belang. In Groenewegen (1926) en CBS (1935) over de dienstkeuringsgegevens

van 1925, respectievelijk 1932, staan tabellen van gemiddelde intelligentiescores uitgesplitst naar onderwijsniveau en de indelingsraad waar de jongens gekeurd zijn. Bovendien zijn de aantallen waarover de gemiddelden berekend zijn, bekend. Door een binomiale verdeling binnen de cellen te veronderstellen kan de complete tabel op basis van de gemiddelden worden geschat.

Binnen het volledige binomiaal model dat in appendix C nader wordt toegelicht, toetsen we het meer restrictieve model dat regionale herkomst en bereikt opleidingsniveau onafhankelijk zijn, gegeven de intelligentiescores.

Dit model past voor de 1925- en de 1932-data goed op de gegevens ($\chi^2 = 63.25$ met d.f. = 80, respectievelijk $\chi^2 = 307.29$ bij d.f. = 225). De parameters voor regio die de regionale verschillen in intelligentie weergeven, laten een vrij grote tegenstelling zien tussen de overwegend agrarische gebieden in Groningen, Drenthe en de Brabantse Kempen en de Peel enerzijds en de stedelijke gebieden in Noord-Holland, Zuid-Holland en Utrecht anderzijds. Ook voor de recente data past dit model goed. Als we likelihood van het binomiaal (sub-) model vergelijken met die van de geobserveerde tabel dan vinden we: $\chi^2 = 2756$ met d.f. = 1596. De passing van het binomiaal model is dus ook in dit geval waar we bovendien de geobserveerde tabel kennen, acceptabel. Dat sterkt ons in de overtuiging dat de aanname van de binomiale verdeling voor de voor-oorlogse gegevens een alleszins redelijke was.

De regio-parameters verschillen voor de 1982-data nauwelijks. Eerder bestaande intelligentieverschillen tussen de onderscheiden keuringsraden die, gelet op de sterke samenhang tussen opleiding en intelligentiescores (Meester en De Leeuw, 1984a), evengoed als verschillen in bereikt opleidingsniveau kunnen worden beschouwd zijn in het recente materiaal niet meer terug te vinden.

Graag zouden wij deze argumentatie nog verder hebben onderbouwd aan de hand van een soortgelijke analyse voor de gegevens uit 1952. In Demologisch Instituut (1957) is echter niet een voor dit doel geschikte tabel opgenomen. Correspondentie-analyse van de in die publikatie opgenomen tabellen van (gediscretiseerde) intelligentiescores tegen een regionale verdeling in economisch geografische gebieden laat wel de eerder gemelde tegenstelling tussen stedelijke gebieden in het Westen en agrarische gebieden in het Noorden en

Oosten van het land zien. Verdergaande conclusies kunnen uit deze correspondentie-analyse niet worden getrokken. De eigenwaarde van de dimensie waarop 'regio' en 'intelligentie' worden geschaald kan niet, als gebruikelijk, als de bovengrens voor de canonische correlatie worden opgevat, omdat schalingen van een nominale variabele als 'regio' niet te interpreteren zijn.

4.2 Samenvatting en slotconclusie

In dit analyseproject onderzochten we de vraagstelling of kenmerken van regio's samenhangen met de schoolloopbaan van leerlingen in die regio's en of de schoolloopbaan van leerlingen in het voortgezet onderwijs onderscheiden kan worden naar een regionale indeling als rekening wordt gehouden met de sociale achtergrond en het geslacht van de leerlingen.

De analyses werden uitgevoerd met gebruikmaking van de gegevens die sinds 1977 door het CBS verzameld zijn (en nog steeds verzameld worden) in het zogenaamde SMVO-cohort onderzoek. Het SMVO-cohort omvat in totaal 37.280 leerlingen, waarvan, na enige selectie, 29.290 in de hier gerapporteerde analyses meedoen.

Enigszins problematisch is de keuze van het regionaal niveau dat in de analyses in beschouwing wordt genomen. In het algemeen zou het aangewezen zijn een niveau te kiezen dat zo dicht mogelijk bij het individueel niveau waarop de SMVO-gegevens zijn verzameld, ligt. De wijze waarop de SMVO-steekproef is getrokken legt hier echter beperkingen op. In paragraaf 2.1 wordt toegelicht hoe de keuze voor de regionale indeling in 40 zogenaamde COROP-gebieden uiteindelijk tot stand is gekomen. De keuze van variabelen die de schoolloopbaan en de sociale achtergrond van de leerlingen beschrijven, is op dezelfde manier bepaald als in Meester en De Leeuw (1983).

In het licht van het in paragraaf 1.2 gestelde is het enige 'multilevel' gegeven dat het analyse bestand bevat, het COROP-gebiedsnummer. Een dergelijke nominaal op te vatten variabele wordt niet verondersteld schaalbaar te zijn. In de analyses dient dus een zekere omzichtigheid betracht te worden, ook al omdat het multilevel-onderzoek vele methodologische valkuilen kent.

In belangrijke mate kunnen deze valkuilen vermeden worden door de

verschillende aggregatie-niveaus zoveel mogelijk gescheiden te houden, de relaties tussen de variabelen bij voorkeur te analyseren op het niveau waarop ze verzameld zijn, en de analyses zo eenvoudig mogelijk te houden, zodat geen moeilijk vol te houden assumpties hoeven te worden gedaan.

Bij wijze van eerste exploratie van het materiaal werden tabellen samengesteld van COROP-gebied, sociaal milieu (geïndiceerd door het beroepsniveau van de vader) en elk van de schoolloopbaanvariabelen (prestatiescore, advies van de onderwijzer voor het voortgezet onderwijs, eerste keuze in het v.o. en het na 4 jaar bereikt opleidingsniveau) afzonderlijk.

Log-lineaire analyse van deze 4 drie-dimensionale tabellen liet zien dat de interacties tussen sociaal milieu en de schoolloopbaan variabelen onafhankelijk zijn van COROP-gebied, ofwel dat de relaties tussen het beroepsniveau van de vader en de schoolloopbaan, afgezien van een schaalfactor, in alle COROP-gebieden dezelfde is.

In een tweede serie analyses werden de onderlinge relaties tussen door Van Veenendaal (1981) op COROP-niveau verzamelde kenmerken geanalyseerd.

Deze kenmerken lieten zich goed beschrijven in een 2-dimensionaal vlak waarvan de eerste as de dichtbevolkte, sterk geïndustrialiseerde stedelijke gebieden in de Randstad en Utrecht onderscheidt van de rest van het land. De tweede dimensie onderscheidt de COROP-gebieden naar de mate waarin er educatieve, culturele en sportvoorzieningen aanwezig zijn.

De structuur die op het niveau van COROP-gebieden wordt gevonden, vertoont grote gelijkens met wat de De Leeuw (1982a) op gemeentelijk niveau vaststelde. Het belangrijkste onderscheid tussen regionale eenheden, of we daarvoor nu COROP-gebied of gemeente kiezen, wordt gevormd door het aantal mensen dat er woont. Het lijkt redelijk om aan te nemen dat op tussenliggende aggregatieniveaus (nodaal of economisch geografisch gebied) een vergelijkbaar resultaat zal worden gevonden.

De verschillen tussen de COROP-gebieden lieten zich nu quantificeren door de scores op de beide gevonden factoren.

De variabelen uit het SMVO-bestand die in een eerder stadium van het onderzoek gequantificeerd werden, konden nu met de COROP-factorscores in verband worden gebracht. Deze samenhangen zijn in een

aantal grafieken (figuur 3.8 t/m 3.29 en figuur, bijlage, 3.5.1 t/m 3.5.20) weergegeven. De associatie tussen de gemiddelden van de SMVO-variabelen, c.q. hun onderlinge correlaties en de COROP-factorscores is niet erg sterk en in veel gevallen vrijwel nihil. Bovendien kunnen deze associaties ook nog geïnfleteerd zijn (vgl. Langbein, 1977).

Er is dus geen relatie te leggen tussen de schoolloopbaanbeschrijvingen op individueel niveau en de kenmerken van de COROP-gebieden. Bij de beschrijving van de COROP-gebieden lag de nadruk op globale demografische kenmerken.

In een canonische correlatie analyse, tenslotte, is nogmaals het verband tussen sociale achtergrond (gecombineerd met het geslacht van de leerlingen), de schoolloopbaan en COROP-gebied onderzocht. Over het gehele bestand genomen is de correlatie tussen milieu-sexe en een gewogen combinatie van de schoolloopbaanvariabelen .352. Wanneer nu de 'invloed' van regio uit deze associatie wordt weggezuiverd verandert de correlatie slechts één punt in de derde decimaal (.351).

Dit resultaat is consistent met de uitkomsten van de eerder uitgevoerde loglineaire analyses. Regionale verschillen in individuele schoolloopbanen vallen vrijwel volledig samen met verschillen in de samenstelling van de beroepsbevolking per regio.

Deze al door Ruiter, Eisma en Passenier (1968) op basis van ander materiaal en een minder fijnzinnige analysemethodiek getrokken conclusie krijgt door de resultaten van dit project een stevig empirisch fundament.

Het resultaat wordt verder ondersteund door wat we in de heranalyse van de dienstkeuringsgegevens hebben gevonden: de voor de tweede wereldoorlog nog bestaande regionale verschillen in 'intelligentie' die voor een belangrijk gedeelte verschillen in bereikt opleidingsniveau weergeven, zijn in de na-oorlogse periode vrijwel volledig opgeheven.

NOTEN

Noten hoofdstuk 2

1. Via het schoolnummer zou dit gegeven in principe te achterhalen zijn. Echter, het CBS is niet geneigd een dergelijk gegeven beschikbaar te stellen uit overwegingen van bescherming van de privacy van de respondenten.
2. We geven de variabelen slechts summier weer. De basisgegevens, de herkomst ervan en de berekeningswijze zijn te vinden in Van Veenendaal (1981).

Noten hoofdstuk 3

1. Stel \underline{S} is één van de schoolloopbaanvariabelen, \underline{B} is de interactieve achtergrondvariabele en \underline{C} is de regiovariabele COROP. \underline{S} neemt de waarden S_i ($i = 1, \dots, I$) aan, \underline{B} de waarden B_j ($j = 1, \dots, J$) en \underline{C} de waarden C_k ($k = 1, \dots, K$). Als we π_{ijk} schrijven voor de theoretische kans (in de populatie) dat een willekeurig gekozen leerling waarden (S_i, B_j, C_k) heeft, dan is het algemene model te formuleren als $\pi_{ijk} = \alpha_{ij} \beta_{ik} \gamma_{jk}$. De parameters voor de interactie tussen schoolloopbaan en achtergrond staan in de matrix A , met elementen α_{ij} . Wanneer we de β_{ik} voor COROP k in de diagonale matrix B_k verzamelen en de γ_{jk} in de diagonale matrix C_k , dan zijn volgens het model de theoretische kansen op (S_i, B_j) in C_k gelijk zijn aan $P_k = B_k A C_k$. Dat wil zeggen: in ieder COROP-gebied worden de schoolloopbaan-achtergrond associaties A , die steeds dezelfde zijn, geschaald met COROP-specifieke rij- en kolomfactoren. Een andere formulering is dat alle kruisverhoudingen tussen achtergrond en schoolloopbaan onafhankelijk zijn van COROP-gebied.

In formule kan dit uitgedrukt worden als:

$$(\pi_{ijk} \pi_{uvk}) / (\pi_{ivk} \pi_{ujk}) = (\pi_{ijl} \pi_{uvl}) / (\pi_{ivl} \pi_{ujl})$$

2. Het programma PREHOM combineert niet-lineaire principale componenten analyse (PRINCALS) en multiple correspondentieanalyse (HOMALS). Techniek en programma worden besproken in Bekker (1982).

3. Om één of andere reden is variabele 9 (dichtheid van het wegenet) niet opgenomen of weggevallen.
4. Om wille van de vergelijkbaarheid van de figuren zijn voor deze grafiek de factorladingen met -1 vermenigvuldigd, zodat de vectoren in dezelfde richting wijzen als voor de PCA-oplossing.

Noten hoofdstuk 4

1. De resultaten van de heranalyse van de keuringsgegevens m.b.t. regionale verschillen worden hier globaal weergegeven. In een binnenkort te verschijnen artikel zal op deze analyses gedetailleerd worden ingegaan (Meester en De Leeuw, 1984c).
2. Over de samenhang tussen de intelligentiescores van de keurlingen van 1982 en die van hun vaders zal binnenkort worden gepubliceerd in De Leeuw, Vroon, Meester en Leyten (1984).

APPENDICES

Appendix A: Correspondentieanalyse met behulp van spline-functies

In dit rapport hebben we vooral PREHOM gebruikt om de bivariate regressies tussen m variabelen zo goed mogelijk te lineariseren. Met m variabelen zijn er $\frac{1}{2}m(m-1)$ bivariate regressies, en in het algemeen is het niet mogelijk om transformaties of quantificaties van de categorieën van de variabelen te vinden die alle regressies exact lineariseren. Er is een compromis oplossing nodig. In Meester en De Leeuw (1983, hoofdstuk 3) lieten we, aan de hand van een heranalyse van dienstkeuringsgegevens, zien dat exacte linearisatie mogelijk is in het geval dat er slechts twee variabelen zijn. De linearisaties worden gevonden met behulp van correspondentieanalyse (zie ook Meester en De Leeuw, 1983, appendix A).

Correspondentieanalyse is een geschikte techniek wanneer de twee variabelen waarvan we de correlatie willen bestuderen categorisch zijn, met een niet al te groot aantal categorieën. Hoe groot 'groot' is, wordt onder andere bepaald door het aantal individuen in de steekproef. Het is daarbij vooral van belang dat de kruistabel van de twee variabelen behoorlijk gevuld is. Wanneer de variabelen een groot aantal mogelijke waarden hebben (of wanneer de steekproef relatief klein is), dan zal correspondentieanalyse geen betrouwbare informatie opleveren. We kunnen dat onderzoeken met technieken die stabiliteit van de oplossing proberen te beschrijven, maar we zullen het ook dikwijls zien aan categoriequantificaties die heel anders uitvallen dat we verwachtten (sterke afwijkingen van monotoniciteit, die niet of nauwelijks te interpreteren zijn).

Een bijkomend nadeel van correspondentieanalyse treedt op bij toepassing op numerieke variabelen (die dikwijls natuurlijk zeer veel categorieën hebben). In het geval van hoofdstuk 3 worden de 40 COROP gebieden bijvoorbeeld met een aantal continu variërende variabelen beschreven, die in principe alle waarden tussen 0 en 100 kunnen aannemen (ze zijn genormeerd). Ieder van de variabelen heeft daardoor ongeveer 40 categorieën, een kruistabel van twee variabelen heeft 1600 cellen, waarvan er 1560 leeg zijn, en er 40 de frequentie één hebben. Correspondentieanalyse levert een triviaal resultaat, dat bovendien

volledig instabiel is. Een mogelijke oplossing is om de waarden van de variabelen te groeperen in intervallen. Die intervallen definiëren dan de nieuwe categorieën, waarvan er vanzelfsprekend veel minder zijn. Het is gemakkelijk in te zien dat bij deze procedure de gevonden optimale transformatie van de variabele een stap-functie is. Immers alle COROP gebieden in een interval van een variabele krijgen dezelfde optimale categorie quantificatie. Wanneer we dus oorspronkelijke tegen getransformeerde variabelen uitzetten, vinden we een functie die binnen intervallen constant is, maar tussen intervallen verschilt. Een (niet noodzakelijk monotone) stap-functie dus.

Voor continu variërende variabelen is het gebruik van een stap-functie, en dus van een klassieke correspondentieanalyse na discretisering, niet erg aantrekkelijk. De continuïteit van de variabele gaat immers verloren, en het is wel erg streng om gelijkheid binnen intervallen te eisen. Anders gezegd: weinig continue functies kunnen goed benaderd worden door een stap-functie met weinig stappen. Om dit nadeel te ondervangen kunnen we splines gebruiken. Voor een wat algemenere inleiding in splines verwijzen we naar De Leeuw, Van Rijckevorsel, en Van der Wouden (1981) en naar Van Rijckevorsel (1982). Voor onze doeleinden volstaat het om op te merken dat de splines die wij gebruiken bestaan uit aan elkaar geplakte lijnstukjes. Anders: we eisen nu van de transformatie dat hij in iedere interval (tussen twee 'knooppunten') lineair is, en dat hij over de gehele lijn continu is (dat wil zeggen dat in de knooppunten de stukken aan elkaar vastzitten). De transformatie wordt dus een op de knooppunten geknikte rechte.

Het is nu eenvoudig om de door ons gebruikte vorm van correspondentieanalyse te definiëren. Gegeven zijn de continue variabelen, en de gekozen knooppunten. Gezocht worden 'geknikte lineaire' transformaties die een zo groot mogelijke correlatie opleveren. In het klassieke geval (bij gebruik van stap-functies) kan men aantonen dat de optimale transformaties de beide regressies lineariseren. Bij gebruik van splines is dit niet meer het geval, althans niet meer exact.

Appendix B: PARTALS

Om de in 3.6 gebruikte PARTALS methode uit te leggen, moeten we eerst een korte inleiding geven op CANALS, het programma voor niet-lineaire canonische correlatie rekening. Er zijn twee sets van variabelen, waarvan voor het gemak wordt aangenomen dat ze allemaal als nominale variabelen behandeld worden. Er zijn n individuen. Wanneer we de variabelen in de eerste set quantificeren, en vervolgens de gequantificeerde variabelen lineair combineren, dan ontstaat er één of andere vector van n getallen. Andere quantificaties, of dezelfde quantificaties met andere gewichten voor de lineaire combinatie, geven andere vectoren in \mathbb{R}^n . Maar hoe we de quantificaties en de combinaties ook kiezen, de resulterende vectoren zullen altijd liggen in een deelruimte van \mathbb{R}^n , waarvoor de indicatormatrices (de 'dummy' coderingen) van de variabelen een omspannende verzameling vormen. Dus: alle vectoren van lengte n die we kunnen verkrijgen door eerst quantificatie en dan combinatie vormen een deelverzameling van \mathbb{R}^n , die identiek is met de deelruimte die omspannen wordt door alle lineaire combinaties van de kolommen van de 'dummies' behorende bij de variabelen. Er zijn twee sets van variabelen, en die definiëren dus twee deelruimtes van \mathbb{R}^n . In CANALS vinden we nu twee vectoren x en y , met x in de deelruimte gedefinieerd door de eerste set en y in de deelruimte gedefinieerd door de tweede set, op zo'n manier dat x en y een zo klein mogelijke hoek maken (een zo groot mogelijk correlatie hebben).

En nu PARTALS. Hierbij is er sprake van een derde deelruimte van \mathbb{R}^n , in het voorbeeld gedefinieerd door de variatie binnen COROP gebieden. Het gaat hierbij om alle vectoren die binnen ieder COROP gebied tot nul optellen, dat wil zeggen in afwijkingen van hun bijbehorend COROP-gemiddelde staan. In de vorm van PARTALS die hier gebruikt is, worden x en y weer in, respectievelijk, de achtergrond- en schoolloopbaan-deelruimtes gekozen, op zo'n manier dat hun orthogonale projecties op de derde (binnen-COROP) deelruimte een zo klein mogelijke hoek hebben. We gaan hier niet verder op de technische details in, maar laten verdere interpretatie aan de geometrische fantasie van de lezer over. Details, uitbreidingen, en variaties staan in

De Leeuw (1984). De definitie is dus, nogmaals, vind vectoren x en y in de deelruimte L_X en L_Y op zo'n manier dat hun projecties op de deelruimte L_Z een zo klein mogelijke hoek maken. Die minimale hoek, een invariant van het drievoud (L_X, L_Y, L_Z) , noemen we de maximale partiële canonische correlatie.

Appendix C: Een binomiaal model voor onvolledige gegevens

In Groenewegen (1926) en CBS (1935) zijn tabellen opgenomen met als rijen keuringsraden of regionale herkomst naar economisch geografisch gebied, als kolommen bereikt opleidingsniveau, en als celinhoud gemiddelde score op een intelligentietest. De aantallen observaties waarop de gemiddelden gebaseerd zijn, zijn bekend, de variantie binnen de cellen is dat niet. Intelligentie is weergegeven als het aantal test-items dat goed beantwoord is.

Stel er zijn N opgaven. Als de variabele regio waarden i ($i = 1 \dots I$) en opleidingsniveau waarden j ($j = 1 \dots J$) aanneemt, dan nemen we aan dat de kans dat iemand in cel (i, j) k items goed beantwoord heeft, gelijk is aan:

$$\pi_{k|ij} = \binom{N}{k} \alpha_{ij}^k (1-\alpha_{ij})^{N-k}$$

We veronderstellen dus dat de items onafhankelijk zijn, gegeven de regio en het opleidingsniveau.

De kans dat iemand in cel (i, j) terecht komt is β_{ij} .

De kans dat een keurling in cel (i, j, k) terecht komt is dan:

$$\pi_{ijk} = \binom{N}{k} \alpha_{ij}^k (1-\alpha_{ij})^{N-k} \beta_{ij}$$

De likelihood van een geobserveerde drieweg-tabel P_{ijk} wordt dan:

$$\begin{aligned} L &= \sum_i \sum_j \sum_k P_{ijk} \ln \pi_{ijk} \\ &= \sum_i \sum_j P_{ij*} \ln \beta_{ij} + \sum_i \sum_j P_{ij*} \{ \underline{m}_{ij} \ln \alpha_{ij} + (N - \underline{m}_{ij}) \ln(1 - \alpha_{ij}) \} \end{aligned}$$

In deze formule is P_{ij*} de (bekende) cel-marginaal en

$$\underline{m}_{ij} = \frac{\sum_{k=1}^N k P_{ijk}}{P_{ij*}}$$

ofwel het gemiddelde in cel (i, j) , eveneens bekend. De maximum likelihood schatters van α_{ij} en β_{ij} zijn eenvoudig $\hat{\alpha}_{ij} = \underline{m}_{ij} / N$ en $\hat{\beta}_{ij} = P_{ij*}$.

Dit binomiale model beschouwen we als het volledige model. Omdat een parametrische vorm voor de verdeling binnen de cellen wordt verondersteld die het mogelijk maakt de complete verdeling d.m.v. het gemiddelde te beschrijven, kan het model op basis van de relatief gebrekkige informatie volledig geschat worden.

Binnen het volledige model zijn we geïnteresseerd in het meer restrictieve $\ln \frac{\alpha_{ij}}{1-\alpha_{ij}} = \mu + \gamma_i + \delta_j$, ofwel op een logit-schaal is er geen inter-

actie tussen regio en scholingsniveau voor het bepalen van de intelligentieverdeling. Dit restrictieve model wordt aan de beschikbare gegevens getoetst.

LITERATUUR

- Akaike, H., 'Information theory and an extension of the maximum likelihood principle', in: B.N. Petrov and F. Cšaki (eds.): 2nd Int. Symposium on Information Theory Akademia Kiado, Budapest, 1973.
- Bekker, P., 'Relaties tussen varianten van niet-lineaire principale componenten analyse', doctoraalscriptie vakgroep Methoden en Technieken in de Psychologie, Leiden, 1982.
- Bishop, Y.M.N., Fienberg, S.E. en Holland, P.W., 'Discrete Multivariate Analysis: Theory and Practice', Cambridge, Mass., MIT Press, 1975.
- Boon, e.a., 'De relatie tussen onderwijs en arbeidsmarkt in regionaal perspectief', RION, Haren, 1983.
- Brouwer, F., Nijkamp, P. en Rietveld, P., 'Een overzicht van methoden voor de behandeling van kwalitatieve geografische gegevens', in: D. op 't Veld, T. van der Smagt en H. Timmermans (red.), 'Waar harde data ontbreken', Antwerpen, 1983, p. 1-34.
- Bun-Siersma, B.P.G.M. en Spruit, L.G.M., 'Analyse van gemeentekenmerken', Voortgangsbericht II in het kader van het onderzoek 'Regionale kenmerken en individuele schoolloopbanen', KASKI, Den Haag, 1981.
- Bun-Siersma, B.P.G.M. en Spruit, L.G.M., 'Regionale factoren en individuele schoolloopbanen', KASKI-rapport nr. 370, Den Haag, 1982.
- Burg, E., van der, 'Canals user's guide', Leiden, vakgroep Datatheorie, 1983.
- Burg, E. van der en Leeuw, J. de, 'Non-linear Canonical Correlation', British Journal of Mathematical and Statistical Psychology, 30, 1983, p. 54-80.
- CBS, 'Intelligentieverhoudingen in Nederland', Groningen, 1935.
- CBS, 'Analyse van de interregionale verschillen in deelneming aan hoger onderwijs', Statistische en econometrische onderzoeken nr. 1, Zeist, 1958.

- CBS, 'Schoolloopbaan en herkomst van leerlingen bij het voortgezet onderwijs. Deel 2: Cohort 1977, Schoolkeuze'. Staatsuitgeverij, 's-Gravenhage, 1982.
- Collaris, J.W.M. en Kropman, J.A., 'Van jaar tot jaar: Onderzoek naar de school- en beroeps carrière van jongens en meisjes die in 1965 het lager onderwijs verlieten'. Tweede fase. Nijmegen, 1978.
- Demologisch Instituut, 'Begaafdheidsonderzoek en intelligentiespreiding Deel II, Intelligentie en arbeidspotentieel', Utrecht, 1957.
- Dronkers, J., 'Manipuleerbare variabelen in de schoolloopbaan'. In: J.L. Peschar en W. Ultee (red.), 'Sociale stratificatie'. Boeknummer Mens en Maatschappij, Deventer, 1978.
- Dronkers, J. en Diekerhof, E., 'Het effect van lokale deelname aan verschillende typen secundair onderwijs op individuele schoolloopbanen'. In: Tijdschrift voor Onderwijsresearch, 5, 1980, 6, p. 250-270.
- Dronkers, J. en Jong, U. de, 'Jencks en Fägerlind op zijn Hollands; een aanzet tot de studie van de relaties tussen milieu, intelligentie, onderwijs, beroep en inkomen'. In: Sociologische Gids, 25, 1978, p. 4-30.
- Dronkers, J. en Jungbluth, M.M.M., 'Schoolloopbaan en geslacht. De invloed van geslacht op de schoolloopbaan van het Nederlands primair en secundair onderwijs'. In: Amsterdams Sociologisch Tijdschrift, 6, 1979, p. 6-50.
- Dronkers, J. en Schijf, H., 'Buurten, scholen en individuele onderwijsloopbanen; een beter analyse-model?', Paper Sociologendagen 1984.
- Eeden, P. van den en Saris, W.E., 'Empirisch onderzoek naar multi-level uitspraken', Mens en Maatschappij, 59, 1984, p. 165-178.
- Fienberg, S.E., 'The Analysis of Cross-Classified Categorical Data', Cambridge, Mass., MIT Press, 1977.
- Galtung, J., 'Theory and Methods of Social Research', New York, 1967.
- Gifi, A., 'Non-linear Multivariate Analysis', vakgroep Datatheorie, Leiden, 1981.
- Groenewegen, H.Y., "Het onderzoek naar de algemene praktische intelligentie bij de keuringsraden in 1925". In: De militaire spectator, 95, 1926, p. 634-645.
- Hannan, M.T., 'Aggregation and Disaggregation in Sociology', Lexington, Mass., 1971a.
- Hannan, M.T., 'Problems of Aggregation', in: H.M. Blalock (ed.) Causal Models in the Social Sciences, Chicago, 1971b.
- Heek, F. van, 'Het verborgen talent', Meppel, 1968.

- Kodde, D.A., 'Wie kiest er voor hoger onderwijs?', tussenverslag van project 'de vraag naar hoger onderwijs', Researchmemorandum 8401, KU Nijmegen, 1984.
- Koornstra, M.J., Gerritsen, C.W.A. en Wiercx van Rhijn, E.M.F., 'Invloed van regionaal-demografische ontwikkelingen op prognose aantal eerstejaars studenten per instelling van wetenschappelijk onderwijs', RU Leiden, 1983.
- Langbein, L.I., 'Schools or Students: Aggregation Problems in the study of Student Achievement', Evaluation Studies Review Annual, 2, 1977, p. 270-298.
- Leeuw, J. de, 'Gemeentetypen en lineariteit', discussiestuk voor de SISWO-overleggroep Longitudinaal school- en beroepsloopbaanonderzoek, Leiden, vakgroep Datatheorie, 1982a.
- Leeuw, J. de, 'Nonlinear Principal Components Analysis', in: H. Cassinus, P. Ettinger, R. Tomassone (eds.), Compstat 1982. Wien, Physika Verlag, 1982b.
- Leeuw, J. de, 'PARTALS in OVERALS', Interne notitie vakgroep Datatheorie, Leiden, september 1984.
- Leeuw, J. de en Burg, E. van der, 'Permutation Distributions of Canonical Correlations', interne notitie vakgroep Datatheorie, Leiden, 1984.
- Leeuw, J. de en Kreft, I., 'Over multilevel analyse', commentaar op: Van den Eeden en Saris, 'Empirisch onderzoek naar multilevel uitspraken', interne notitie vakgroep Datatheorie, Leiden, 1984.
- Leeuw, J. de, Rijckevorsel, J. van en Wouden, H. van der, 'Nonlinear Principal Components Analysis with B-splines', in: Methods of Operations Research, vol. 43, Hain Verlag, 1981.
- Leeuw, J. de, Vroon, P.A., Meester, A.C. en Leyten, J.L., 'Correlation between Raven Progressive Matrices-test scores of fathers and sons', ter publikatie aangeboden, 1984.
- Matthijssen, M.A.J.M. en Sonnemans, G.J.M., 'Schoolkeuze en schoolsucces bij vmo en ulo in Noord-Brabant'. In: Paedagogische publikaties, 3, 1958, Tilburg.
- Meester, A.C. en Leeuw, J. de, 'Intelligentie, sociaal milieu en de schoolloopbaan', vakgroep Datatheorie FSW/RUL, Leiden, 1983.
- Meester, A.C. en Leeuw, J. de, 'Over het intelligentie-onderzoek bij de militaire keuringen vanaf 1925 tot heden', in: Mens en Maatschappij, 59, 1984a, 1, p. 5-26.
- Meester, A.C. en Leeuw, J. de, 'De regressie van intelligentie-scores op schoolloopbaanvariabelen onder constanthouding van sociale achtergrond en geslacht', verschijnt binnenkort, 1984b.
- Meester, A.C. en Leeuw, J. de, 'Regionale verschillen in intelligentie bij de militaire keuringen 1925-heden', wordt binnenkort ter publikatie aangeboden, 1984c.

- Nijkamp, P. en Rietveld, P., 'Kwalitatieve Statistiek en Econometrie, Een toepassing op regionale verschillen in Nederland', Kwantitatieve Methoden, 10, 1983, p. 55-76.
- Peschar, J., 'Schoolloopbanen in het voortgezet onderwijs', Den Haag, Staatsuitgeverij, 1983.
- Provinciaal Bestuur van Noord-Brabant, 'Rapport over een onderzoek naar de stand van het gewoon lager onderwijs in Noord-Brabant'. 1957.
- Ruiter, R., Eisma, M. en Passenier, J., 'Factoren die de regionale verschillen in deelneming aan het VHMO in Nederland beïnvloeden'. In: F. van Heek e.a., 'Het verborgen talent', Meppel, 1968.
- Rijckevorsel, J. van, 'Canonical Analysis with B-splines', in: H. Cassinus, P. Ettinger, R. Tomassone (eds.), Compstat 1982, Wien, Physika Verlag, 1982.
- Santema, M., 'Onderwijsbeleid en regio', Leiden, Stafleu, 1974.
- Tesser, P., 'Schoolloopbaanonderzoek in Nederland. Een trendstudie naar de ontwikkelingen in het empirisch onderzoek tussen 1960 en 1980'. Nijmegen, 1981.
- Veenendaal, W.M. van, 'Regionale welvaart in Nederland', Economische Faculteit, VU Amsterdam, 1981.
- Vergoossen, Th.W.M. en Wever, E., 'Regio, onderwijs en K.U.', Geografisch en Planologisch Instituut K.U.N., Nijmegen, 1978.
- Vliegen, J.M. en Jong, U. de, 'Onderwijs in Nederland', Monografieën en Volkstelling 1971, 12, Nijmegen, 1981.

BIJLAGEN

Bijlage 2.1: Verdeling van de SMVO-leerlingen over 40 COROP-gebieden uitgesplitst naar geslacht (na uitselecteren BVA en COROP-missing)

GEBIED	JONGENS	MEISJES	TOTAAL
1. Oost-Groningen	130	151	281
2. Delfzijl e.o.	70	70	140
3. Overig Groningen	340	364	704
4. Noord-Friesland	411	374	784
5. Zuidwest Friesland	139	172	311
6. Zuidoost Friesland	184	226	410
7. Noord Drenthe	170	176	346
8. Zuidoost Drenthe	166	163	329
9. Zuidwest Drenthe	153	190	343
10. Noord Overijssel	424	415	839
11. Zuidwest Overijssel	112	148	260
12. Twente	587	532	1119
13. Veluwe	470	573	1043
14. Achterhoek	401	443	844
15. Arnhem/Nijmegen	688	687	1375
16. Zuidwest Gelderland	180	208	388
17. Utrecht	801	879	1680
18. Kop van Noord-Holland	331	334	665
19. Alkmaar e.o.	242	275	517
20. IJmond	202	210	412
21. Aggl. Haarlem	266	274	540
22. Zaanstreek	174	166	340
23. Groot Amsterdam	821	836	1657
24. Gooi en Vechtstreek	206	220	426
25. Aggl. Leiden	306	347	653
26. Aggl. 's-Gravenhage	622	645	1267
27. Delft en Westland	219	252	471
28. Oostelijk Zuid-Holland	317	312	629
29. Groot Rijnmond	1009	1057	2066
30. Zuidoost Zuid-Holland	424	482	906
31. Zeeuws Vlaanderen	123	103	226
32. Overig Zeeland	221	202	423
33. West Noord-Brabant	550	623	1173
34. Midden Noord-Brabant	469	516	985
35. Noordoost Noord-Brabant	518	482	1000
36. Zuidoost Noord-Brabant	617	732	1349
37. Noord Limburg	299	289	588
38. Midden Limburg	198	231	429
39. Zuid Limburg	621	660	1281
40. Zuidelijke IJsselmeerpolders	44	46	90
Totaal	14.225	15.065	29.290

Bijlage 2.2: Frequentie-verdelingen van achtergrond- en schoolloop-
baanvariabelen in het SMVO-bestand (na uitselecteren BVA-
en COROP-missing)

1. Sociaal milieu waaruit de leerlingen afkomstig zijn (BVA)

	Frequentie	%
1. Arbeiders	10303	35.2
2. Boeren en tuinders	1704	5.8
3. Zelfstandige middenstand	2919	10.0
4. Lagere employées	4105	14.0
5. Middelbare employées	6452	22.0
6. Hogere employées en vrije beroepen	3807	13.0
<hr/>		
Totaal	29290	100 %

2. Prestaties op de verkorte versie van CITO-toets (45 taal- en 25 reken-
opgaven)

	frequentie	%
1 - 20 goed	879	3.0
21 - 26 goed	1766	6.0
27 - 32 goed	3000	10.2
33 - 38 goed	3808	13.0
39 - 44 goed	4245	14.5
45 - 50 goed	4238	14.5
51 - 56 goed	3591	12.3
57 - 62 goed	2362	8.1
63 - 70 goed	901	3.1
score onbekend	4500	15.4
<hr/>		
	29290	100 %

3. Advies van de onderwijzer voor het v.o.

	frequentie	%
LBO	6560	22.4
MAVO	11323	38.7
HAVO	6225	21.3
VWO	4350	14.9
onbekend	832	2.9
<hr/>		
Totaal	29290	100 %

4. Eerste keuze in het v.o.

	frequentie	%
LBO	7809	26.7
MAVO	10694	36.5
HAVO	3316	11.3
VWO	7471	25.5
<hr/>		
Totaal	29290	100 %

5. In 1982 bereikt niveau in het v.o.

	frequentie	%
zonder diploma van school	2657	9.1
LBO-diploma	4695	16.0
MAVO-diploma	3254	11.1
HAVO-diploma	5597	19.1
VWO	4866	16.6
MBO/HBO (met LBO- of geen diploma)	3152	10.8
MBO/HBO (met MAVO- of HAVO diploma)	5069	17.3
<hr/>		
Totaal	29290	100 %

Bijlage 3.2: Correlaties tussen 13 gestandaardiseerde COROP-kenmerken

	1	2	3	4	5	6	7	8	9	10	11	12
1. Inkomen	*											
2. Werkloosheid	.64	*										
3. Welvaart	.60	.53	*									
4. Levensonderhoud	-.19	-.11	.02	*								
5. Bev. dichtheid	-.46	-.37	-.23	.17	*							
6. Nat. milieu	.29	.41	.29	.28	.01	*						
7. Industrialisatie	-.43	-.30	-.19	-.17	.92	.02	*					
8. Luchtvervuiling	-.48	-.40	-.24	-.16	.94	.01	.90	*				
9. Wegennet	.77	.37	.37	-.06	-.66	.16	-.62	-.67	*			
10. Cult./Sportvoorz.	.05	-.15	.12	.20	.19	-.04	.28	.21	.04	*		
11. Aantal scholen	-.04	-.17	.12	.07	-.05	.07	-.05	.02	-.02	.35	*	
12. Afst. tot de Randst.	.79	.74	.38	-.16	-.59	.08	-.51	-.59	.61	-.09	-.17	*
13. Med. diensten	.55	.26	.46	-.01	-.44	.19	-.48	-.39	.38	.09	.17	.43

Bijlage 3.3: Correlaties tussen 13 optimaal getransformeerde variabelen op COROP-niveau

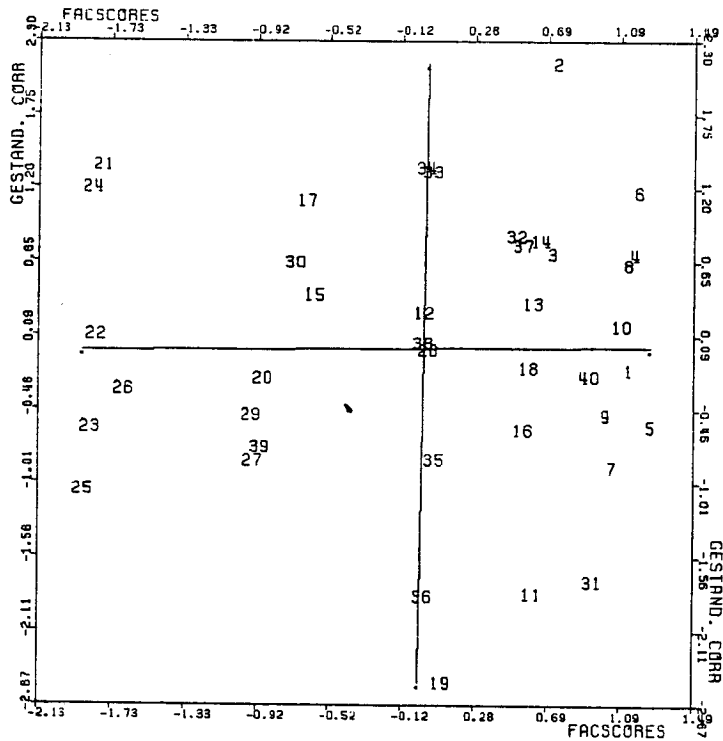
	1	2	3	4	5	6	7	8	9	10	11	12
1. Inkomen	*											
2. Werkloosheid	.69	*										
3. Welvaart	.57	.49	*									
4. Levensonderhoud	-.21	-.17	-.27	*								
5. Bev. dichtheid	-.82	-.53	-.25	.15	*							
6. Nat. milieu	.20	.13	.15	-.24	-.03	*						
7. Industrialisatie	-.64	-.39	-.12	-.06	.80	.20	*					
8. Luchtvervuiling	-.76	-.52	-.18	.15	.97	-.02	.76	*				
9. Wegennet	.75	.46	.16	.03	-.89	-.04	-.83	-.85	*			
10. Cult./Sportvoorz.	-.04	-.42	-.12	.25	.01	.13	.02	.06	-.00	*		
11. Aantal scholen	-.14	-.35	.02	-.02	.16	.31	.21	.18	-.20	.53	*	
12. Afst. tot de Randst.	.66	.69	.29	-.36	-.55	.04	-.34	-.53	.46	-.45	-.45	*
13. Med. diensten	.66	.33	.40	-.11	-.57	.36	-.47	-.56	.49	.18	.25	.17
Eigenwaarden:	.4363	.1661	.1403	.0672	.0552	.0373	.0287	.0216				
	.0171	.0151	.0090	.0046	.0015							

Bijlage 3.4: Componentladingen 3-dimensionale PCA-oplossing 13 variabelen op COROP-niveau

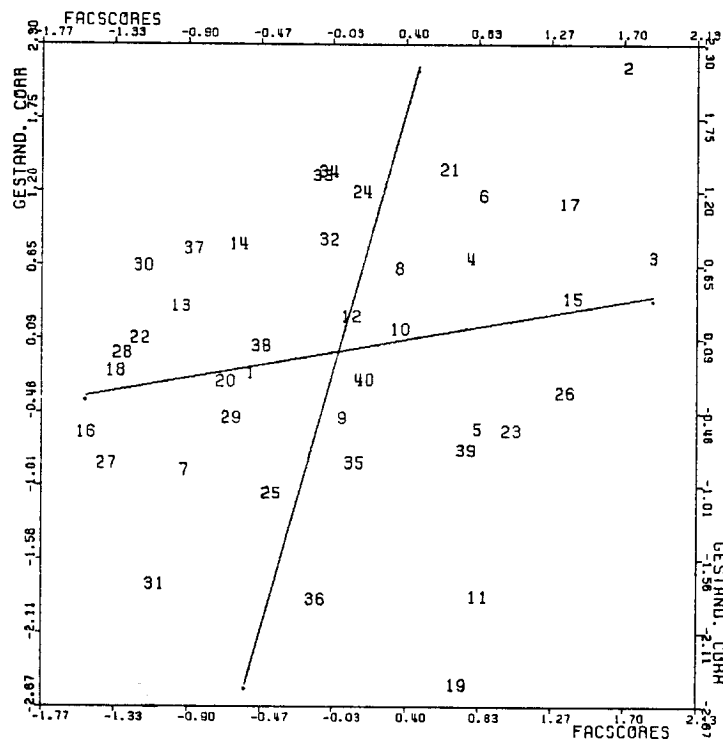
Variabele	dim. 1	dim. 2	dim. 3
1. Inkomen	.926	-.121	.165
2. Werkloosheid	.736	.327	.271
3. Welvaart	.440	-.016	.558
4. Levensonderhoud	-.220	-.172	-.616
5. Bevolkingsdichtheid	-.930	.153	.191
6. Nat. milieu	.087	-.388	.653
7. Industrialisatie	-.782	.118	.463
8. Luchtvervuiling	-.902	.121	.213
9. Wegennet	.860	-.144	-.353
10. Cult. en sportvoorzieningen	-.184	-.807	-.180
11. Aantal scholen	-.268	-.777	.254
12. Afstand tot de randstad	.704	.472	.206
13. Med. diensten	.628	-.551	.214

Bijlage 3.5: Samenhangen tussen schoolloopbaanvariabelen gerelateerd aan COROP factorscores

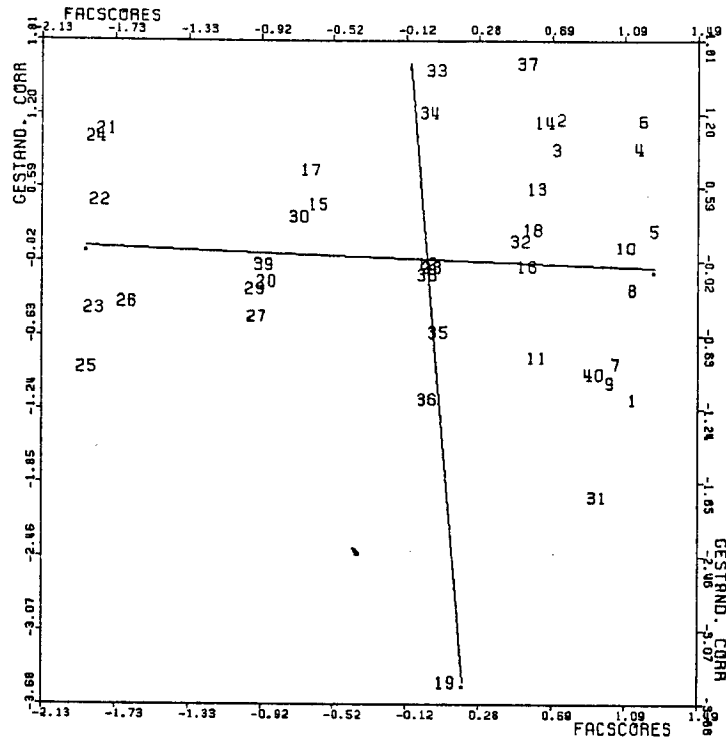
Figuur 3.5.1: Correlatie PRE-ADV versus factor 1, 40 COROP-gebieden; corr.: .01.



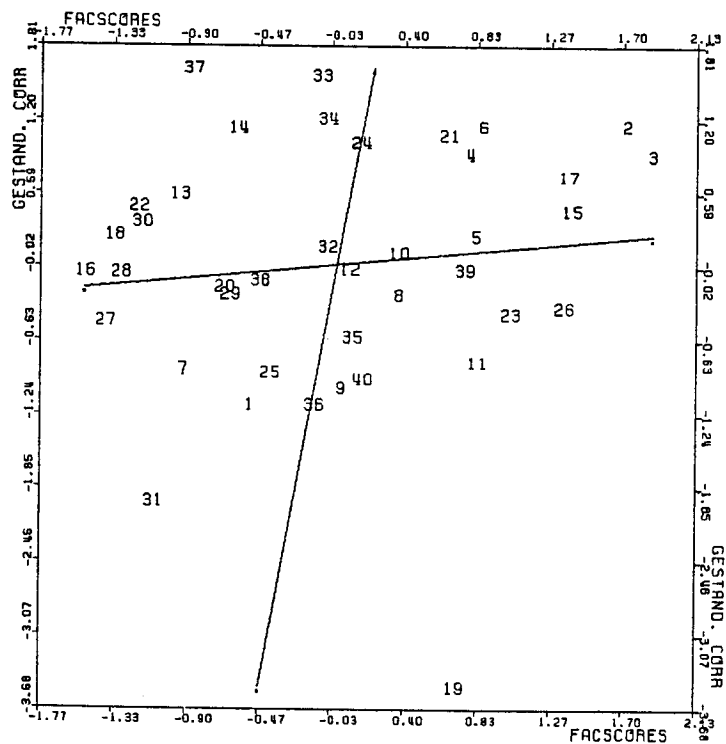
Figuur 3.5.2: Correlatie PRE-ADV versus factor 2, 40 COROP-gebieden; corr.: .22.



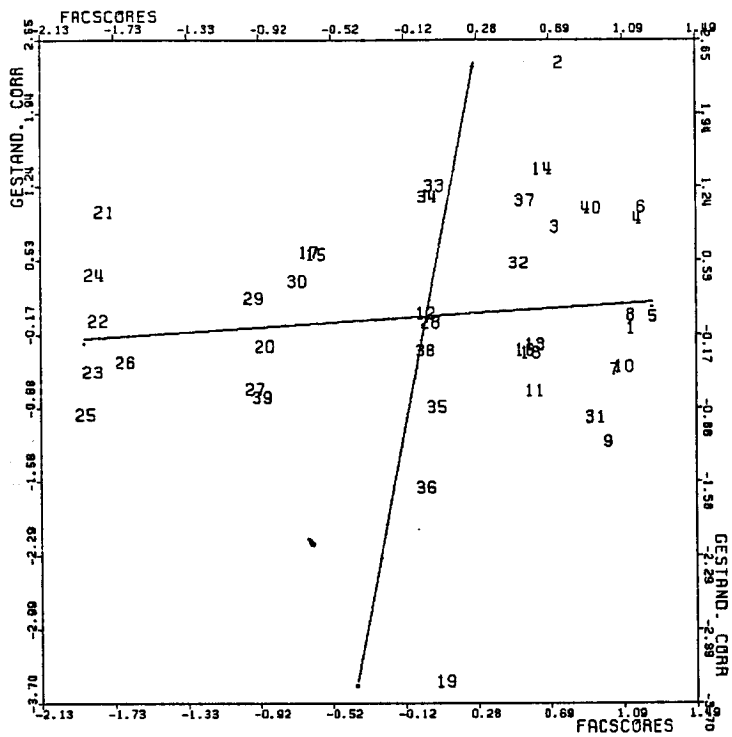
Figuur 3.5.3: Correlatie PRE-KEUS versus factor 1, 40 COROP-gebieden; corr.: -.06.



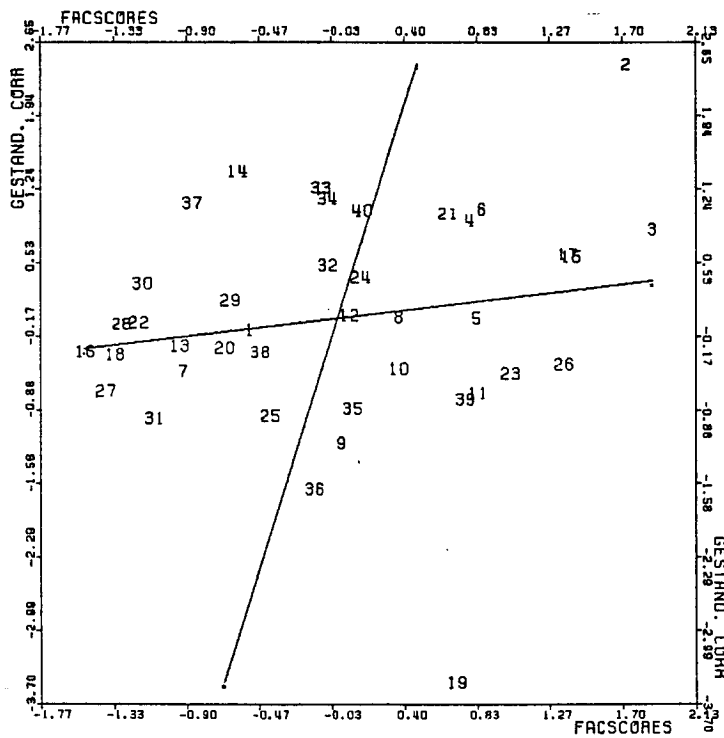
Figuur 3.5.4: Correlatie PRE-KEUS versus factor 2, 40 COROP-gebieden; corr.: .13.



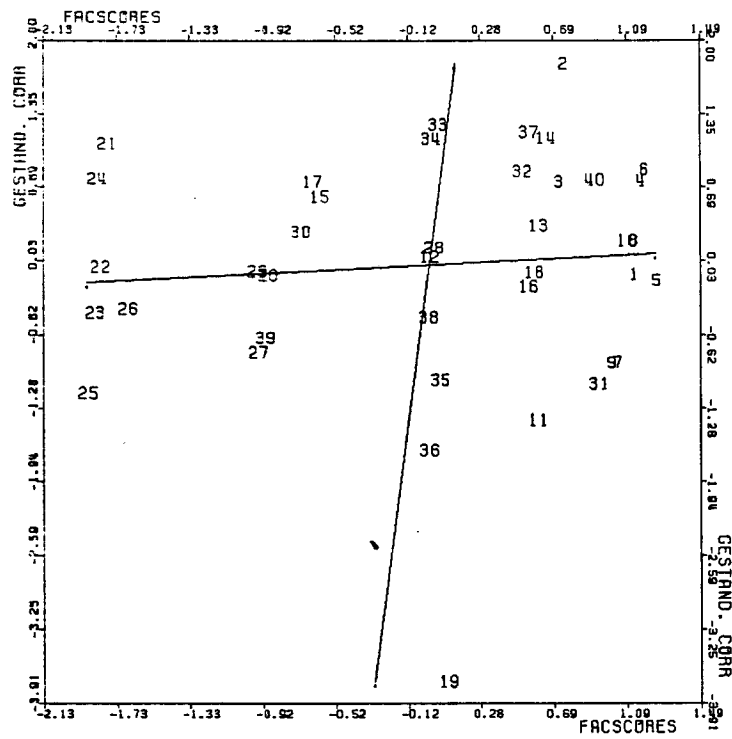
Figuur 3.5.6: Correlatie PRE-EIN versus factor 1, 40 COROP-gebieden; corr.: .11.



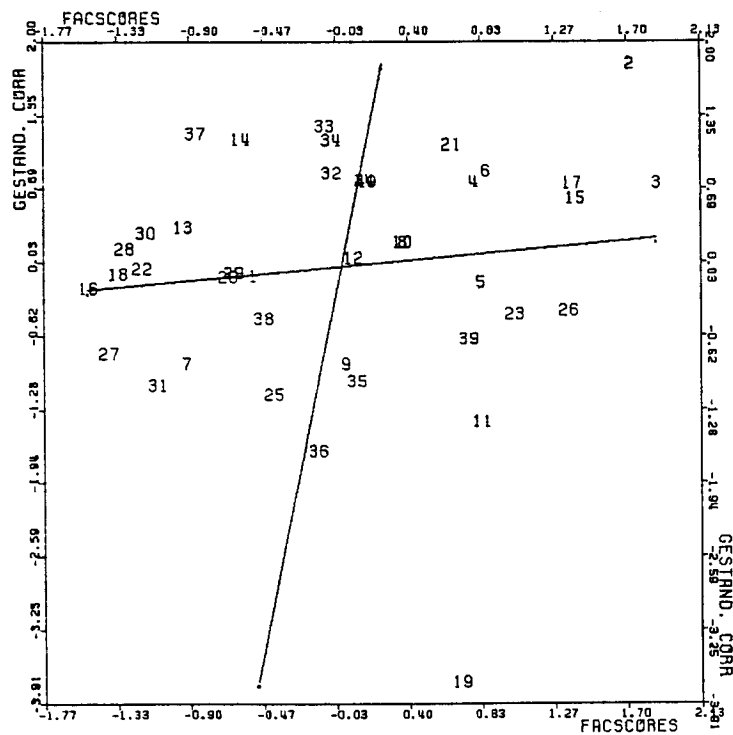
Figuur 3.5.7: Correlatie PRE-EIN versus factor 2, 40 COROP-gebieden; corr.: .20.



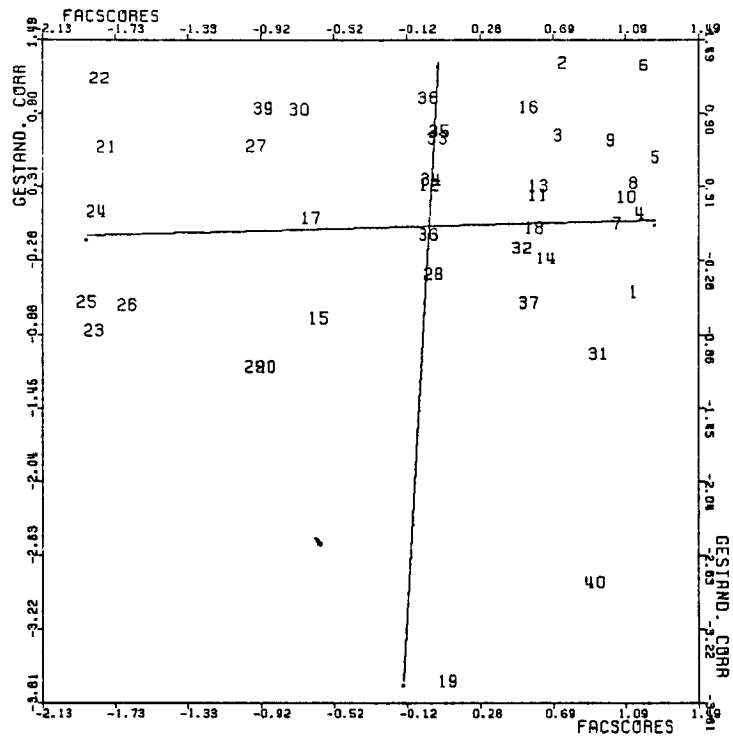
Figuur 3.5.8: Correlatie PRE-SCH versus factor 1, 40 COROP-gebieden; corr.: .08.



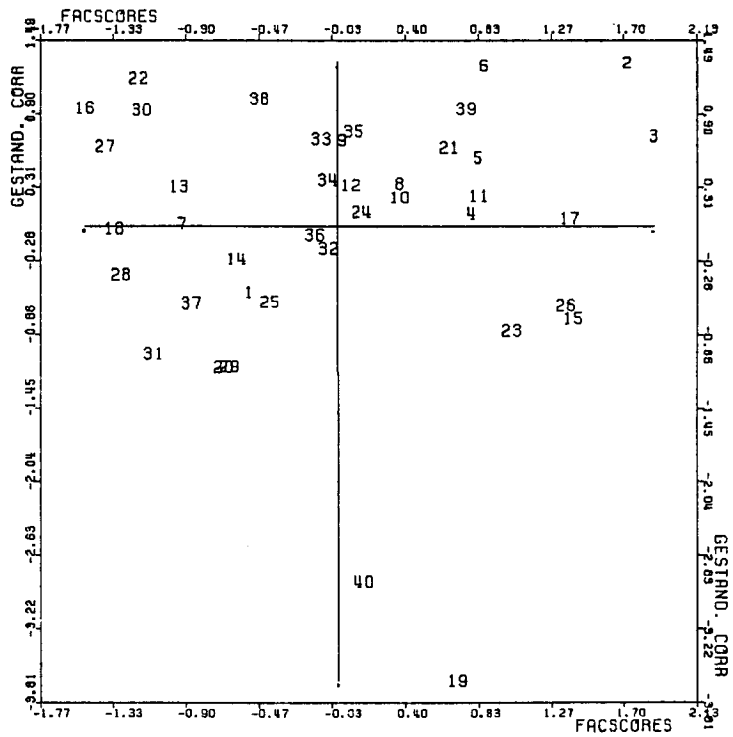
Figuur 3.5.9: Correlatie PRE-SCH versus factor 2, 40 COROP-gebieden; corr.: .14.



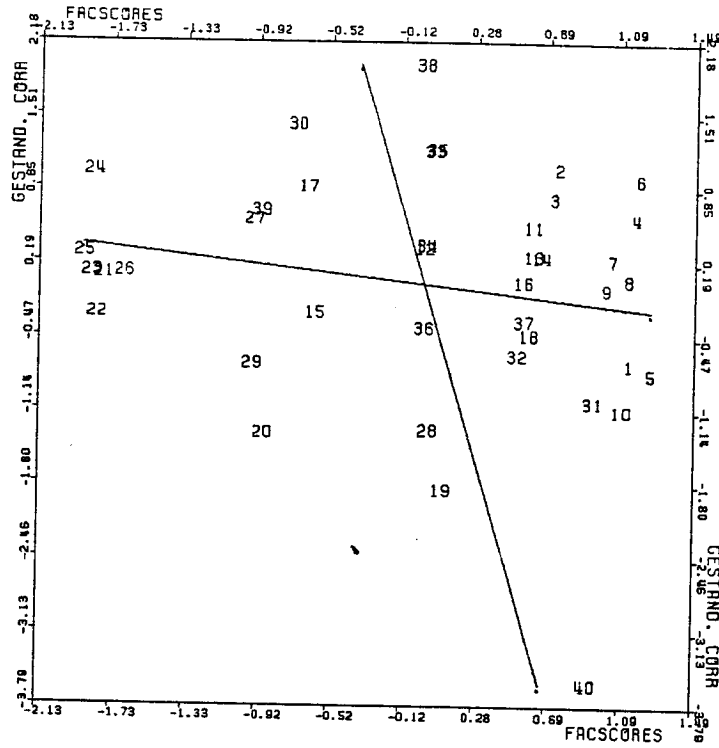
Figuur 3.5.10: Correlatie ADV-KEUS versus factor 1, 40 COROP-gebieden; corr.: .04.



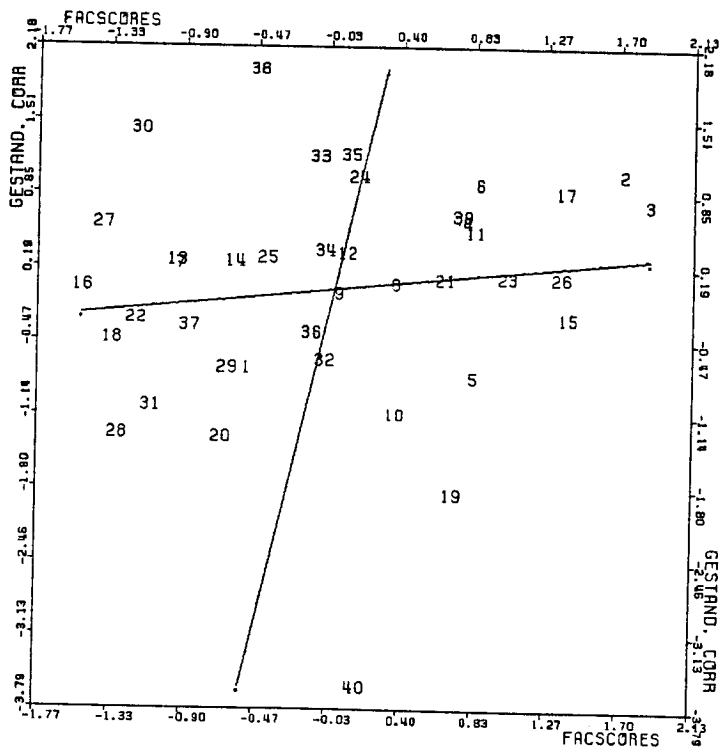
Figuur 3.5.11: Correlatie ADV-KEUS versus factor 2, 40 COROP-gebieden; corr.: -.00.



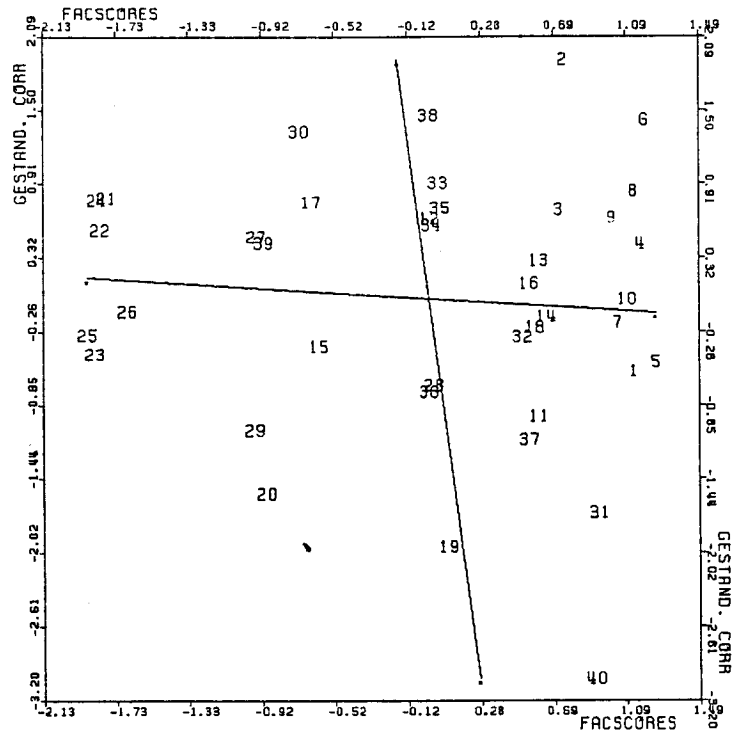
Figuur 3.5.12: Correlatie ADV-EIN versus factor 1, 40 COROP-gebieden; corr.: -.18.



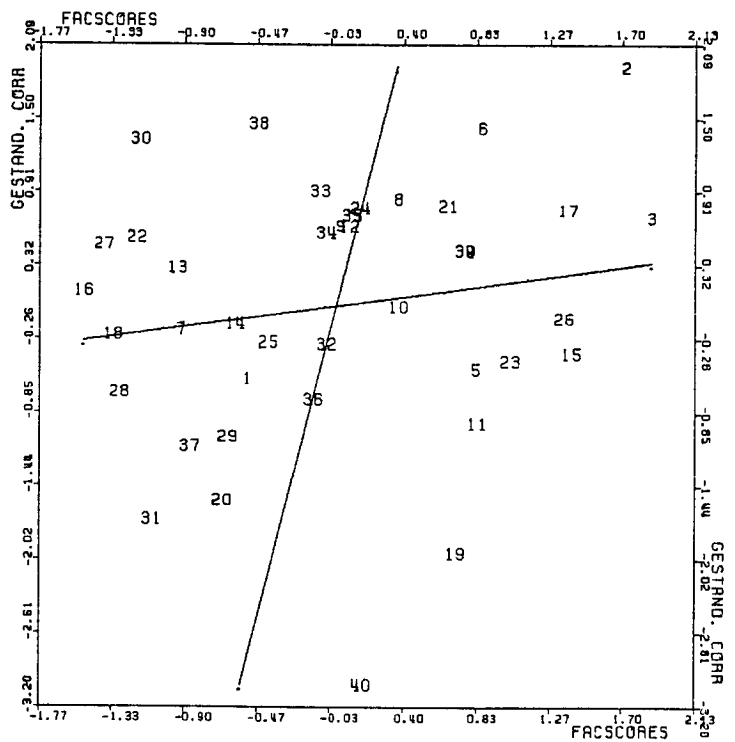
Figuur 3.5.13: Correlatie ADV-EIN versus factor 2, 40 COROP-gebieden; corr.: .15.



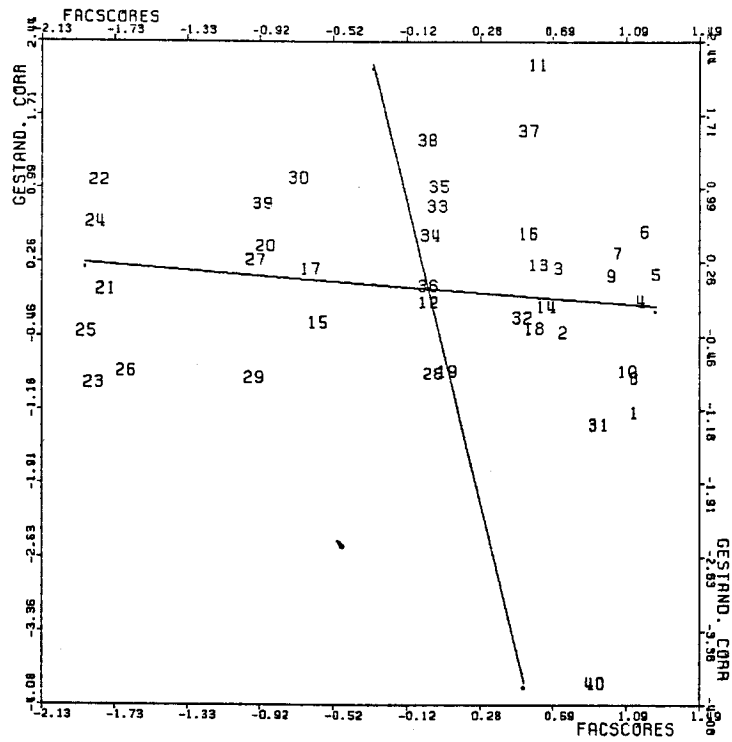
Figuur 3.5.14: Correlatie ADV-SCH versus factor 1, 40 COROP-gebieden; corr.: -.09.



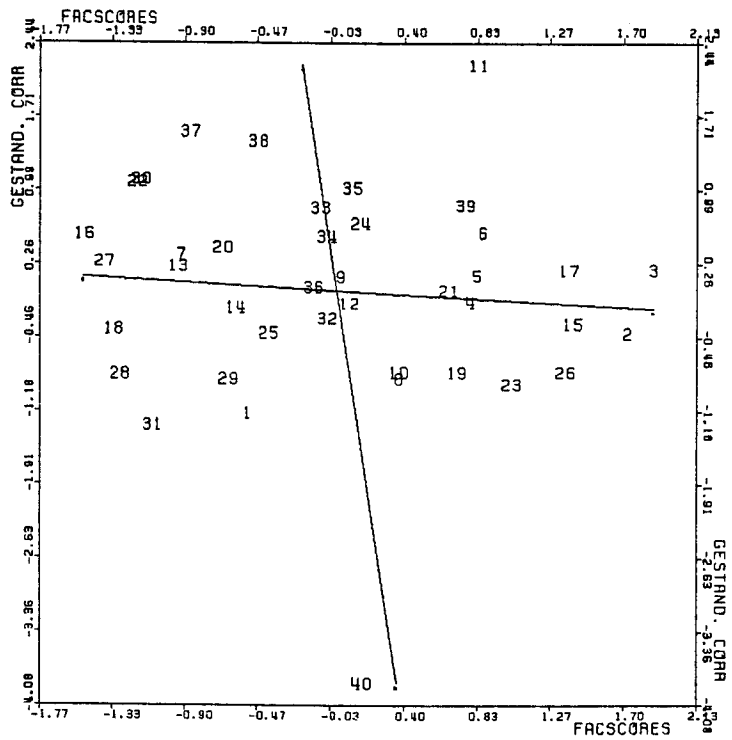
Figuur 3.5.15: Correlatie ADV-SCH versus factor 2, 40 COROP-gebieden; corr.: .18.



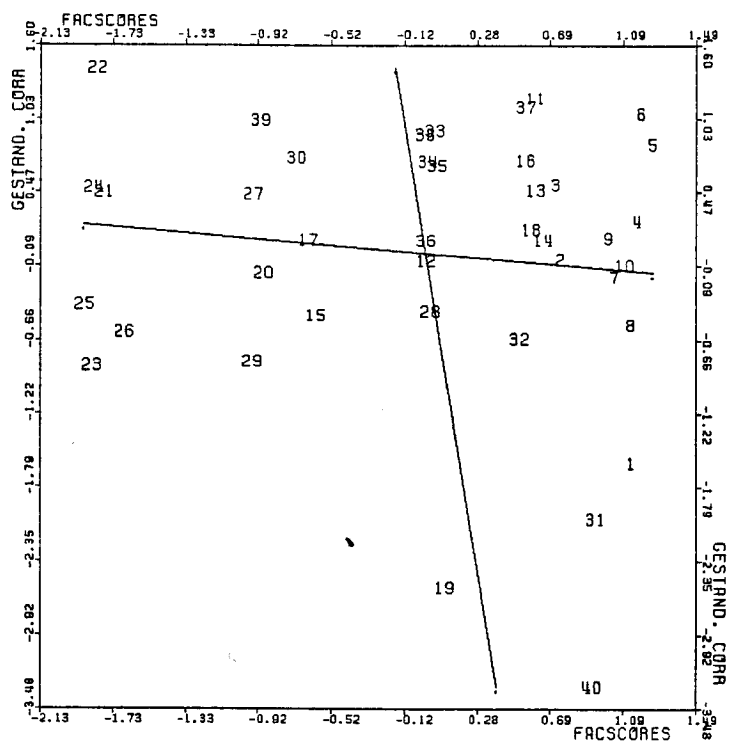
Figuur 3.5.16: Correlatie KEUS-EIN versus factor 1, 40 COROP-gebieden; corr.: $-.13$.



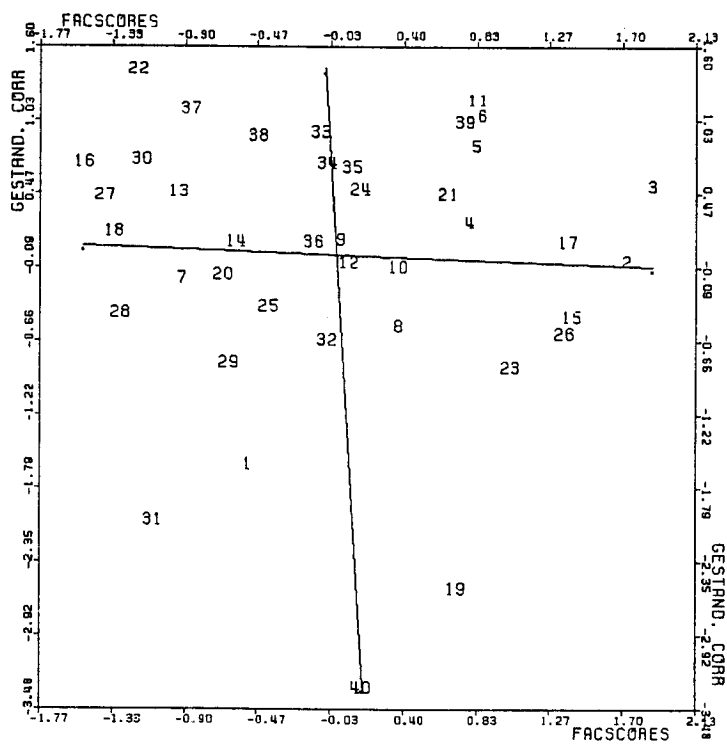
Figuur 3.5.17: Correlatie KEUS-EIN versus factor 2, 40 COROP-gebieden; corr.: $-.09$.



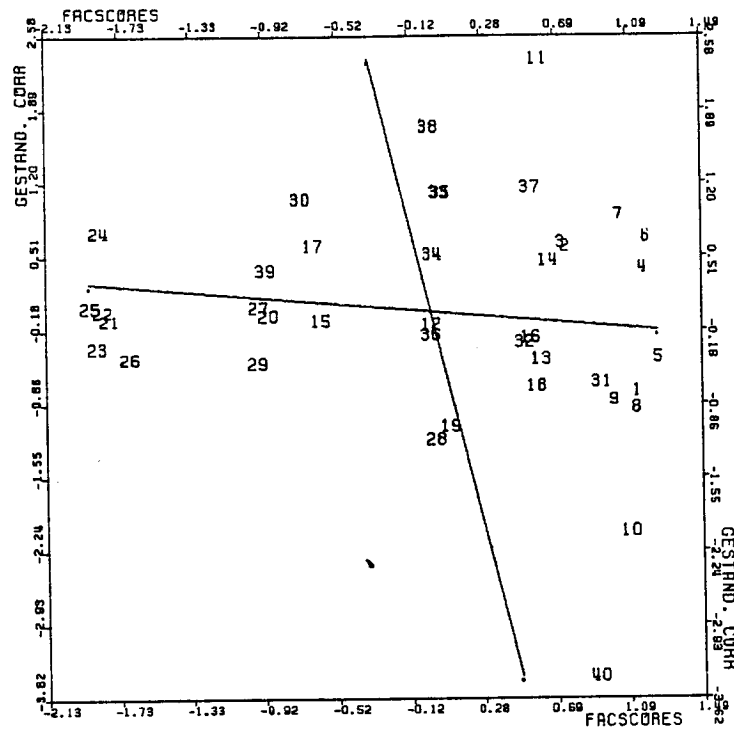
Figuur 3.5.18: Correlatie KEUS-SCH versus factor 1, 40 COROP-gebieden; corr.: $-.12$.



Figuur 3.5.19: Correlatie KEUS-SCH versus factor 2, 40 COROP-gebieden; corr.: $-.05$.



Figuur 3.5.20: Correlatie EIN-SCH versus factor 1, 40 COROP-gebieden; corr.: $-.14$.



Figuur 3.5.21: Correlatie EIN-SCH versus factor 2, 40 COROP-gebieden; corr.: $.08$.

