

DE VERKEERSONVEILIGHEID IN DE PROVINCIE NOORD-BRABANT IX B

Het relatie-onderzoek: Onderzoekopzet en methode van onderzoek

R-81-39

Ir. H. Hoek & drs. S. Oppe

Leidschendam, 1981

Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV

SAMENVATTING

Het relatie-onderzoek in Noord-Brabant heeft tot doel relaties tussen ongevallen en wegsituaties op te sporen, die als mogelijke verklaringen voor de onveiligheid kunnen worden opgevat. Behalve de totale onveiligheid wordt daarbij vooral de onveiligheid van de onderzoekgebieden beschouwd. De onderzoekgebieden betreffen ongevalstypen die in Noord-Brabant veel voorkomen. Dit zijn: kop-staart- en flankongevallen, ongevallen bij schemer en duisternis, ongevallen op kruispunten en enkelvoudige ongevallen. Een ander onderzoekgebied is de verdeling van de ongevallen over de dagen van de week.

Het relatie-onderzoek heeft betrekking op rijks- en provinciale wegen buiten de bebouwde kom in Noord-Brabant. Voor deze wegen is een grote hoeveelheid gegevens verzameld (SWOV, 1980).

In dit rapport wordt ingegaan op de opzet van het onderzoek, de gehanteerde analysetechnieken en de mogelijkheden om met behulp van deze analysetechnieken tot maatregelen te komen.

De interesse gaat bij dit onderzoek vooral uit naar de relaties tussen zeer veel ongevallen en weg- en verkeerskenmerken. Aangezien deze relaties niet te voren bekend zijn, is een methode van onderzoek gekozen, waarbij als vanzelf de belangrijkste relaties tussen een groot aantal kenmerken naar voren komen.

Deze manier van werken brengt met zich mee dat men niet vooraf een beperkt aantal hypothesen formuleert die men vervolgens toetst, maar dat men het materiaal min of meer voor zich laat spreken. Wel bepalen de ideeën die er vooraf bestaan in belangrijke mate de kenmerken die in de analyses betrokken worden.

Het relatie-onderzoek bestaat uit drie delen, waarin verschillende soorten locaties worden onderzocht, namelijk: wegvakken van 200 m ("raaien"), kruispunten en wegverbindingen van grotere lengte ("strengen").

De analysetechnieken die voor dit soort onderzoek ter beschikking staan zijn factoranalyse en canonische analyse, maar dan geschikt gemaakt voor kwalitatieve gegevens. Deze technieken, die respectievelijk HOMALS en CANALS heten, zijn in dit onderzoek gebruikt voor het analyseren van een gecomprimeerde vorm van het verzamelde materiaal.

HOMALS wordt gebruikt voor het vinden van een structuur binnen één groep kenmerken. CANALS biedt de mogelijkheid de relatie tussen twee of drie groepen kenmerken te bepalen. Omdat de interesse hier uitgaat naar de relatie tussen ongevallen enerzijds en de weg- en verkeerskenmerken anderzijds, is CANALS de meest gebruikte analysetechniek. Deze techniek kan worden gebruikt om een canonische correlatie-analyse of een meervoudige-analyse uit te voeren op kwalitatieve gegevens. Bij de analyse worden aan de categorieën van de kenmerken getalswaarden toegekend, zodanig dat de canonische correlatie (of meervoudige correlatie) maximaal is. We kunnen het toekennen van getallen aan de categorieën van de kenmerken zien als een "herschaling" van deze kenmerken. De analyse op deze zgn. categorie-scores kan dan worden beschreven als het oplossen van het klassieke probleem van meervoudige lineaire regressie of canonische correlatie.

Om de gevonden samenhangen goed te interpreteren kan dan ook gebruik worden gemaakt van stapsgewijze meervoudige-regressie-analyse van de categorie-scores en/of het berekenen van partiële correlaties. Beide technieken zijn klassieke hulpmiddelen om vast te stellen of in een gevonden samenhang de weg- en verkeerskenmerken elk een ander deel dan wel hetzelfde deel van de ongevallen verklaren.

De betrouwbaarheid van het gevonden resultaat kan worden bepaald door middel van een Bootstrapstudie. Daarbij worden een aantal aselechte steekproeven met teruglegging uit het onderzoekmateriaal getrokken en worden de resultaten van de analyses over deze steekproeven onderling en met het oorspronkelijke bestand vergeleken.

Doordat de analyse voor iedere locatie aangeeft of de combinatie van weg- en verkeerskenmerken als veilig dan wel als onveilig moet

worden beschouwd, zijn de analyseresultaten direct bruikbaar voor het bepalen van maatregelen. Er kan een lijst met de meest gevaarlijke locaties met bijbehorende kenmerken uitgedraaid worden, waarmee direct bepaald kan worden welke onveilige weg- of verkeerskenmerken aanwezig zijn. Indien een maatregel wordt overwogen, dan hangt het door de analyse voorspelde effect van de maatregel op de veiligheid hangt af van de hoogte van de canonische correlatie in de analyse (hoe "goed" is de oplossing). De juistheid van die voorspelling hangt behalve van de hoogte van de correlatie ook af van de juiste interpretatie van het analyseresultaat en van de "juiste" werking van de maatregel.

INHOUD

<u>Voorwoord</u>	7
1. <u>Inleiding</u>	9
2. <u>De opzet van het relatie-onderzoek</u>	11
2.1. Doel van het onderzoek	11
2.2. De onderzoekfilosofie	11
2.3. De uitvoering van het onderzoek	12
2.4. De inventarisatie en de bestandsopbouw	13
2.5. De analyses	14
3. <u>Methode van onderzoek</u>	16
3.1. Factoranalyse	17
3.2. Canonische analyse	17
4. <u>De gebruikte analysetechnieken</u>	20
4.1. De schaling van kenmerken	20
4.2. HOMALS	20
4.3. CANALS	21
4.4. De relatie met andere analysetechnieken	22
5. <u>Een theoretisch voorbeeld van een CANALS-analyse</u>	24
6. <u>Geometrische weergave van lineaire-regressie-oplossingen</u>	28
7. <u>Interpretatie van samenhangen en betrouwbaarheid van de oplossingen</u>	34
7.1. Interpretatie van samenhangen	34
7.1.1. Stapsgewijze multiple regressie	35
7.1.2. Partiële correlaties	36
7.1.3. De schaling van de criteriumvariabelen	38
7.2. Betrouwbaarheid van de oplossingen	39

8.	<u>Analyseresultaten en maatregelen</u>	41
8.1.	De analyses in relatie tot de locaties	41
8.2.	Keuze van locaties voor maatregelen	43
8.3.	Het zoeken naar maatregelen voor de gekozen locaties	44
8.4.	Beschrijving van een mogelijke procedure	45
8.5.	De effectiviteit van maatregelen	47
9.	<u>Een praktijkvoorbeeld</u>	49
9.1.	De presentatie van de analyseresultaten	49
9.2.	Interpretatie van de gevonden analyseresultaten	49
9.2.1.	De analyseresultaten	49
9.2.2.	Multiple regressie en partiële correlatie	50
9.2.3.	De beoordeling van de schaling	52
9.3.	De betrouwbaarheid van de oplossing	52
9.4.	Analyseresultaat met betrekking tot de locaties	55
9.5.	De verdere aanpak	56
9.6.	Effectiviteit van de maatregelen	57
10.	<u>Slotopmerkingen</u>	58
	<u>Literatuur</u>	60
	<u>Afbeeldingen</u>	61
	<u>Bijlagen</u>	

VOORWOORD

De Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV heeft begin 1975 een onderzoek ingesteld naar de verkeersonveiligheid in de provincie Noord-Brabant. Dat onderzoek heeft als resultaat opgeleverd dat er onvoldoende aanwijzingen zijn om Noord-Brabant als de meest onveilige provincie van Nederland te bestemmen. Wel kon worden geconcludeerd dat de verkeersonveiligheid in Noord-Brabant in vergelijking met die in de andere provincies groot is.

Naar aanleiding hiervan is in Noord-Brabant een meer gedetailleerd onderzoek ingesteld. Hieruit zijn een aantal aandachtsgebieden naar voren gekomen, waarop de verkeersonveiligheid in Noord-Brabant significant verschilt van die in de rest van Nederland (SWOV, 1976a).

Uit de aandachtsgebieden is een selectie gemaakt van gebieden die in aanmerking kwamen om er nader onderzoek naar te verrichten, de zgn. onderzoekgebieden (SWOV, 1976b).

Het vervolgonderzoek (fase 2) is begin 1977 gestart in opdracht van het Provinciaal Bestuur van Noord-Brabant en de Minister van Verkeer en Waterstaat. Op basis van dit onderzoek zullen maatregelen of aanknopingspunten daartoe worden aangegeven.

Een belangrijk deel van fase 2 wordt gevormd door het zogenaamde relatie-onderzoek, dat relaties tussen ongevallen en weg- en verkeerskenmerken probeert te vinden. Dit onderzoek heeft betrekking op rijks- en provinciale wegen buiten de bebouwde kom. Van deze wegen zijn ten behoeve van het relatie-onderzoek grote aantallen weg- en verkeerskenmerken verzameld (SWOV, 1980).

Het voor u liggende rapport beschrijft de opzet van het relatie-onderzoek en de onderzoeksmethode die gebruikt wordt.

Het relatie-onderzoek bestaat uit drie delen, die betrekking hebben op de analyse van respectievelijk wegvakken van 200 m, kruispunten en wegverbindingen. Het eerste deel, de zogenaamde "analyse raaien", is inmiddels voltooid. De rapportage daarvan verschijnt tegelijk met dit rapport. De volgende delen van het

relatie-onderzoek, de analyse van kruispunten en langere wegvakken (strengen genaamd), worden in de loop van 1981 en 1982 uitgevoerd, waarna een integraal eindrapport zal verschijnen.

In het kader van het onderzoek naar de verkeersonveiligheid in de provincie Noord-Brabant zijn inmiddels al verschillende rapporten verschenen. Deze betreffen min of meer geïsoleerde deelproblemen, dan wel voorstudies ten behoeve van het relatie-onderzoek. In bijlage 1 is een overzicht van alle tot nu toe verschenen rapporten opgenomen. In bijlage 2 is de onderlinge samenhang tussen de deelonderzoeken aangegeven.

1. INLEIDING

Doel van dit rapport is het verschaffen van informatie die nodig is voor een goede interpretatie van de analyseresultaten van het relatie-onderzoek. Het is geschreven, omdat de opdrachtgever bij het interpreteren van de analyseresultaten behoefte had aan een toelichting. De gestelde vragen zijn als uitgangspunt gebruikt bij het opstellen van dit rapport. Als gevolg daarvan worden niet alle aspecten evenwichtig behandeld: soms wordt een zeer gedetailleerde uitwerking gegeven van technische aspecten, soms wordt alleen globale informatie verschaft. De nadruk ligt in dit rapport op het verschil tussen CANALS en de klassieke correlatietechnieken. Bij de lezer wordt een zekere kennis van de klassieke technieken verondersteld.

Na een bespreking van de opzet van het relatie-onderzoek en de gevolgde methode van onderzoek (hoofdstuk 2 resp. 3) wordt in hoofdstuk 4 ingegaan op de gebruikte analysetechnieken: HOMALS en CANALS. Ook het voor deze analysetechnieken specifieke aspect van de schaling van klassen kenmerken komt hier aan de orde.

In hoofdstuk 5 wordt de output van het computerprogramma uitvoerig beschreven aan de hand van een theoretisch voorbeeld.

In de hoofdstukken 6 en 7 komen aspecten aan de orde die niet specifiek zijn voor HOMALS en CANALS, maar die in het algemeen van belang zijn bij canonische-correlatie-analyses en multiple-regressie-analyses.

Relatief veel aandacht wordt in hoofdstuk 6 besteed aan de geometrische weergave van kenmerken in de vorm van vectoren. De hoeken tussen de vectoren geven de relaties tussen de kenmerken aan. Dat dit onderwerp hier zo uitvoerig wordt behandeld, komt omdat het in weinig tekstboeken terug te vinden is. Toch is een dergelijke geometrische weergave een belangrijk hulpmiddel om inzicht te krijgen in de analyseresultaten.

In hoofdstuk 7 komt de interpretatie van samenhangen en de betrouwbaarheid van de oplossingen aan de orde. Voor de interpretatie wordt gebruik gemaakt van stapsgewijze multiple regressie en

partiële correlaties. Beide hulpmiddelen worden beknopt besproken, waarna ook nog wordt ingegaan op de schaling van de criteriumvariabelen. Eveneens in dit hoofdstuk worden drie methoden besproken om de betrouwbaarheid van de oplossingen vast te stellen.

In hoofdstuk 8 wordt uitgebreid ingegaan op de betekenis van de CANALS-oplossing voor de onderzoekobjecten (de locaties), waarna aangegeven wordt hoe de analyseresultaten vertaald kunnen worden in maatregelen.

Hoofdstuk 9 bevat een voorbeeld dat laat zien hoe de kennis uit de hoofdstukken 2 t/m 8 in de praktijk van het relatie-onderzoek kan worden toegepast.

In hoofdstuk 10 worden nog enkele slotopmerkingen gemaakt.

2. DE OPZET VAN HET RELATIE-ONDERZOEK

2.1. Doel van het onderzoek

Het onderzoek naar de verkeersonveiligheid in Noord-Brabant heeft tot doel te komen tot maatregelen die de verkeersveiligheid in Noord-Brabant vergroten. Om deze maatregelen te kunnen kiezen is kennis nodig over de oorzaken van de ongevallen.

De oorzaken van ongevallen kunnen van zeer verschillende aard zijn. Ze kunnen gelegen zijn in het gedrag van de weggebruiker, het gedrag van het voertuig, de hoeveelheid en samenstelling van het verkeer, de kenmerken van de weg en zijn omgeving, en in omstandigheden zoals het weer. Het onderzoek Noord-Brabant richt zich op maatregelen die door de wegbeheerder genomen kunnen worden. Zulke maatregelen zullen doorgaans gericht zijn op het wegnemen van ongevalsoorzaken die gelegen zijn in weg- en verkeerskenmerken. Daarom zal de gewenste kennisvermeerdering vooral betrekking moeten hebben op de relatie tussen ongevallen en weg- en verkeerskenmerken. Ongevalsoorzaken van andere aard spelen daarbij alleen indirect een rol.

De relaties tussen ongevallen en weg- en verkeerskenmerken kunnen een indicatie geven van de ongevalsoorzaken. De te nemen maatregelen zullen zich moeten richten op het wegnemen van deze ongevalsoorzaken. Het relatie-onderzoek heeft ten doel de genoemde relaties vast te stellen voor rijks- en provinciale wegen buiten de bebouwde kom in Noord-Brabant.

2.2. De onderzoekfilosofie

Bij de aanvang van een onderzoek naar de relatie tussen ongevallen en weg- en verkeerskenmerken is vanzelfsprekend reeds een hoeveelheid kennis en ideeën over deze relaties aanwezig. De vraag is dan hoe deze in het onderzoek een rol moeten spelen. Een zeer formele manier is om een aantal hypothesen te formuleren en deze te toetsen. Daarbij zullen meestal geen zeer complexe relaties onderzocht

kunnen worden en is de mogelijkheid om op nieuwe ideeën te komen gering. Daarentegen is de zekerheid over de gevonden resultaten vrij groot. Men kan echter ook volstaan met enige beschouwingen over de mogelijke (complexen van) oorzaken en een onderzoeksmethode gebruiken die de aanwezige structuren blootlegt. De beschouwingen over de mogelijke oorzaken bepalen dan (mede) de te analyseren kenmerken en zijn een hulpmiddel bij de interpretatie. Op deze wijze kunnen ook meer complexe relaties gevonden worden en kunnen nieuwe ideeën ontstaan. De zekerheid over de gevonden resultaten is echter niet groot, zodat ze soms aanvullend getoetst moeten worden. Omdat in dit geval vooral kennisvermeerdering over complexe relaties tussen meerdere kenmerken noodzakelijk was, lag de laatstgenoemde wijze van onderzoek voor de hand.

2.3. De uitvoering van het onderzoek

In de uitvoering van het relatie-onderzoek kunnen de volgende fasen worden onderscheiden:

- het verzamelen van de benodigde gegevens;
- de controle en bewerking van de ruwe gegevens tot analyseerbare gegevens;
- het ontwikkelen van een adequate onderzoeksmethode;
- het uitvoeren van de analyses;
- de interpretatie en het trekken van conclusies;
- de rapportage;
- het aanbevelen van maatregelen.

In bijlage 3 is dit schematisch weergegeven.

Het verzamelen en bewerken van de gegevens en het kiezen van een analysemethode hebben parallel plaatsgevonden. De gekozen analysemethode is daarna tijdens de uitvoering van de eerste analyses verder geoperationaliseerd.

Dit rapport geeft een beschrijving van de gebruikte onderzoeksmethode, de bijbehorende analysetechnieken en de wijze waarop de resultaten geïnterpreteerd moeten worden. Ook wordt aangegeven op welke wijze vanuit de onderzoeksresultaten tot maatregelen gekomen kan worden.

Voordat we daar in de volgende hoofdstukken nader op ingaan, zullen we voor de volledigheid een korte beschrijving geven van de andere fasen van het relatie-onderzoek, waarover afzonderlijke rapportages zijn verschenen of nog zullen verschijnen.

2.4. De inventarisatie en de bestandsopbouw

De eerste activiteiten binnen het relatie-onderzoek waren erop gericht de benodigde gegevens over de ongevallen en de weg- en verkeerskenmerken te verkrijgen. Hiertoe was een uitgebreide inventarisatie "in het veld" nodig. Een daartoe geschikte inventarisatiemethode moest echter eerst nog ontwikkeld worden.

De methode, die de raaimethode wordt genoemd, is ontwikkeld door DHV Raadgevend Ingenieursbureau BV, in samenwerking met PWS Noord-Brabant, de Regionale Directie Noord-Brabant en de Dienst Verkeerskunde van de Rijkswaterstaat en de SWOV. De inventarisatie is uitgevoerd door medewerkers van de Provinciale Waterstaat en de Regionale Directie van Rijkswaterstaat. Wegvakken van 200 m en kruispunten zijn afzonderlijk geïnventariseerd.

De gehele bestandsopbouw is uitgevoerd door DHV. De geïnventariseerde gegevens zijn uitvoerig gecontroleerd. Daarmee zijn de wegkenmerken, de verkeerskenmerken en de ongevallen, die alle drie op verschillende wijze zijn geïnventariseerd, zodanig bewerkt dat ze per locatie aan elkaar gekoppeld konden worden.

Dit leverde twee bestanden op die de basis vormen voor het onderzoek:

- een "raaienbestand" bestaande uit wegvakken van 200 m met de wegkenmerken en de verkeerskenmerken van die wegvakken en de gegevens van de ongevallen die op die wegvakken zijn gebeurd;
- een kruispuntenbestand dat soortgelijke informatie levert maar dan voor kruispunten met bijbehorende takken.

Uit deze basisbestanden zijn verschillende analysetapes voor verschillende soorten analyses gemaakt (SWOV, 1980).

2.5. De analyses

Kruispunten en wegvakken verschillen zo sterk van elkaar dat het niet mogelijk is ze in één onderzoek gezamenlijk te beschouwen. Daarom moet het relatie-onderzoek apart voor kruispunten en voor wegvakken worden uitgevoerd. Door de gebruikte inventarisatiemethodiek en de noodzaak om zeer plaatselijke problemen van meer structurele problemen te kunnen onderscheiden, bleek het bovendien gewenst zowel wegvakken van zeer korte lengte als weggedeelten van grote lengten (routes) afzonderlijk te onderzoeken. Dit leidt tot een driedeling van het relatie-onderzoek, te weten:

1. de analyse raaien, waarbij wegvakken van tweehonderd meter onderzocht worden;
2. de analyse kruispunten;
3. de analyse strengen, waarbij weggedeelten van grotere lengte als één geheel onderzocht worden, routes tussen twee steden. De term "strengen" is gebaseerd op het aaneenrijgen van basiseenheden van honderd meter tot grotere eenheden.

Deze drie delen van het relatie-onderzoek worden na elkaar uitgevoerd.

De deelonderzoeken Raaien en Kruispunten hebben grotendeels dezelfde opbouw (zie bijlage 3). Op basis van beschouwingen over mogelijke ongevalsoorzaken en algemene ideeën over de onveiligheid worden uit het basismateriaal de relevante kenmerken geselecteerd, zonodig hercodeerd en vervolgens op tape gezet. De gegevens van deze eerste analysetape worden geanalyseerd. Wanneer de resultaten van die analyses nieuwe vragen oproepen, waardoor andere kenmerken relevant worden, of wanneer een nadere detaillering van de kenmerken nodig is, wordt een tweede tape gemaakt. Op basis van de twee series analyses worden vervolgens conclusies getrokken en kunnen ideeën over te nemen maatregelen worden ontwikkeld.

Bij het deelonderzoek Strengen worden routes geformeerd door aaneenrijgen van de raaiavakken en kruispunten die binnen de betreffende route liggen. De kenmerken van eenheden binnen een streng worden daarbij omgevormd tot gegevens van de streng (bijvoorbeeld aantal kruispunten, percentage van de lengte die verlicht is).

Welke strengen worden geformeerd en welke kenmerken worden meegenomen, wordt bepaald door de wegstructuur, de gereden routes en de ideeën over daarmee samenhangende en andere structurele oorzaken voor onveiligheid.

Wanneer alle drie de deelonderzoeken zijn afgerond, bestaat een volledig beeld (kennis en nieuwe ideeën) over de locale oorzaken van ongevallen (op raaien en kruispunten) en de structurele oorzaken van ongevallen (strengen). Pas dan kan de maatregelkeuze plaatsvinden (zie het onderste deel van bijlage 2).

3. METHODE VAN ONDERZOEK

Bij de beschrijving of analyse van de verkeersonveiligheid heeft men te maken met zeer veel factoren van diverse aard. Uit onderzoek blijkt telkens dat de invloeden die de verschillende weg- en verkeerskenmerken op de veiligheid uitoefenen, niet van elkaar los te zien zijn. Voor sommige weg- en verkeerskenmerken geldt dat ze onderling sterk samenhangen. Het is dan niet duidelijk of ze elk hetzelfde deel van de ongevallen verklaren dan wel elk een ander deel. Voor andere kenmerken geldt dat ze een gezamenlijke relatie met ongevallen hebben. In beide gevallen zou de gecombineerde relatie met ongevallen niet worden opgespoord als de kenmerken geïsoleerd in de analyse werden betrokken.

Zoals reeds eerder gesteld, is dit onderzoek juist gericht op de bijdrage van onderling samenhangende kenmerken aan de verklaring van de onveiligheid. Er moest dus gebruik worden gemaakt van technieken die de onderlinge samenhang tussen een groot aantal kenmerken (ook wel de onderliggende structuren genoemd) kunnen vinden. De technieken die ons daartoe ter beschikking staan zijn factoranalyse en canonische analyse.

Een complicatie is dat bij de klassieke canonische analyse en factoranalyse een aantal aannamen wordt gedaan. De belangrijkste zijn dat de kenmerken lineair samenhangen en minimaal op "intervalniveau" gemeten zijn. Aan deze eisen wordt niet voldaan door de gegevens van het onderzoek. Meer recente analysemodellen maken het echter mogelijk ook een canonische en factoranalyse uit te voeren als er niet echt gemeten is, maar de gegevens voor elk kenmerk in vaak willekeurige klassen zijn ingedeeld. Nominale of ordinale kenmerken worden eerst geschaald, om daarna te worden behandeld als metrische kenmerken. Bij de schaling van kenmerken worden de getallen voor de klassen zodanig gekozen dat een zo eenvoudig mogelijke beschrijving van het lineaire regressiemodel ontstaat. De strengere eis dat de kenmerken ook nog lineair moeten samenhangen met het criterium, vervalt dan; deze eis wordt ondervangen door de schaling, zodat ook niet-lineaire relaties

tussen de verklarende kenmerken en het criterium kunnen worden onderzocht.

Voor nadere informatie wordt verwezen naar Oppe (1980) en Gifi (1981). De verderop beschreven programma's voor deze soort analyses zijn ontwikkeld door de afdeling Datatheorie van de RU Leiden.

3.1. Factoranalyse

Men kan stellen dat bij factoranalyse wordt geprobeerd de onderzoeksobjecten te beschrijven aan de hand van enkele basiskenmerken. Die basiskenmerken kunnen niet rechtstreeks worden gemeten. Ze liggen ten grondslag aan de veelheid van specifieke kenmerken die wel worden gemeten. De basiskenmerken worden ook wel latente trekken of factoren genoemd. De score op zo'n factor is te beschrijven als een lineaire combinatie van de scores op de gemeten kenmerken of, anders gezegd, als een gewogen som daarvan. Het gewicht dat aan elk gemeten kenmerk wordt toegekend, geeft aan in hoeverre dit kenmerk de betreffende factor mede bepaald.

3.2. Canonische analyse

Als we te maken hebben met twee groepen objecten, dan zouden we kunnen nagaan wat de overeenkomst tussen de groepen is en dus via factoranalyse kunnen zoeken naar factoren die aangeven hoe homogeen de kenmerken zijn. Vaak echter zijn we ook geïnteresseerd in verschillen. We zoeken dan een factor die zo goed mogelijk discrimineert tussen beide groepen. Dit gebeurt met behulp van discriminantanalyse. Hierbij wordt gezocht naar de factor waarvoor geldt dat de scores van objecten op die factor zo goed mogelijk corresponderen met de groepsindeling. Vanuit de score op deze factor kunnen we dan voorspellen tot welke groep een object behoort.

Zo kunnen we ons bijvoorbeeld afvragen welke factoren de nachtongevallen van de dagongevallen onderscheiden. Zo kunnen bijvoor-

beeld locaties waar het 's nachts potentieel gevaarlijk is, worden opgespoord. Zijn er meer dan twee groepen, dan spreken we van canonische discriminantanalyse. De groepsindeling kunnen we zien als een extra kenmerk waarop de objecten worden gecategoriseerd. Een kenmerk echter dat op nominaal niveau is gemeten. Zouden we de objecten niet in groepen indelen, maar bijvoorbeeld op intervalniveau meten op het betreffende kenmerk, dan zou de vraagstelling luiden: welke factoren geven een zo goed mogelijke voorspelling van het kenmerk waarin we geïnteresseerd zijn? In dat geval spreken we van multiple regressie-analyse. Er wordt dan gezocht naar die lineaire combinatie van (metrische) kenmerken, ook wel onafhankelijke variabelen genoemd, die een optimale voorspelling geeft van een bepaalde criteriumvariabele, ook wel afhankelijke variabele genoemd.

We kunnen (canonische) discriminantanalyse dus opvatten als een bijzonder geval van multiple lineaire regressie (MLR), namelijk: MLR met één nominale criteriumvariabele. Als we bij MLR niet met één enkele afhankelijke variabele te maken hebben, maar met meerdere metrische criteriumvariabelen, dan spreken we van canonische correlatie-analyse. We zoeken dan naar die factor (of factoren) van de groep van verklarende variabelen die een optimale voorspelling geeft van (of maximale correlatie heeft met) de groep criteriumvariabelen. Het doel is dus het vinden van een beschrijving van de samenhang tussen twee groepen kenmerken door middel van die factoren van elke groep die deze samenhang zo goed mogelijk weergeven. Het gaat er bij alle bovengenoemde methoden om lineaire combinaties te vinden van metrische variabelen, zodanig dat bepaalde relaties optimaal beschreven worden. Bij factoranalyse gaat het om de relaties binnen een groep kenmerken, bij MLR en (canonische) discriminantanalyse om de relaties van een groep kenmerken met een ander kenmerk (dit kenmerk kan ook een groepsindeling van objecten betreffen), bij canonische correlatie-analyse om de relaties van een groep kenmerken met een andere groep kenmerken. Met variantie-analyse behoren deze analysemodellen tot de groep van lineaire-analysemodellen. De aantrekkelijke statistische eigenschappen van het lineaire-analysemodel

hebben geleid tot het formuleren van de genoemde varianten en de bijbehorende statistische toetsen.

4. DE GEBRUIKTE ANALYSETECHNIEKEN

De geoperationaliseerde analysetechnieken voor factoranalyse en canonische analyse, die geschikt zijn om kwalitatieve gegevens te analyseren en daarom in dit onderzoek zijn gebruikt, heten HOMALS, PRINCALS en CANALS.

4.1. De schaling van kenmerken

Zowel bij HOMALS als bij CANALS wordt aan de klasse van de kenmerken een getalswaarde (schaling) toegekend, zodanig dat een optimale beschrijving van de samenhang mogelijk is. Bij HOMALS en bij CANALS kunnen aan deze schaling restricties worden opgelegd. Bij de schaling wordt dan bijv. rekening gehouden met de volgorde van de klassen (ordinale restrictie) of wordt een kenmerk als metrisch opgevat.

De gebruikte CANALS-versie geeft echter slechts één schaling die voor alle dimensies gelijk is, dit in tegenstelling tot HOMALS waarbij voor elke dimensie een eigen schaling van de kenmerken wordt gezocht. Een variant van HOMALS, PRINCALS genaamd, kent ook slechts één schaling.

We zullen niet ingaan op de voor- en nadelen van een keuze voor PRINCALS of HOMALS. Beide analyses zijn gebruikt, afhankelijk van bepaalde praktische omstandigheden.

In principe zijn bij CANALS beide schalingsmogelijkheden aanwezig. Gezien de aard van de problematiek hebben we gekozen voor slechts één schaling.

4.2. HOMALS

De HOMALS-"factoranalyse" in bijvoorbeeld twee dimensies laat zien welke twee schalingen (van de klassen van een kenmerk) het beste een weergave van een grote groep kenmerken door middel van twee nieuwe (basis)kenmerken mogelijk maken. Deze twee nieuwe kenmerken (factoren, dimensies) geven dan aan wat gemeenschappelijk is in de oorspronkelijke kenmerken. Technisch kan factor-

analyse worden gezien als het oplossen van een eigenwaarde - eigenvectoren probleem. De relatieve bijdrage van elke dimensie (eigenvector) aan de beschrijving is dan af te leiden uit de grootte van de bijbehorende eigenwaarde. Uit de correlaties van de oorspronkelijke kenmerken met de dimensies kunnen we zien, welke kenmerken voornamelijk gemeenschappelijk worden weergegeven en welke kenmerken niet. De discriminatiematen van HOMALS zijn te beschouwen als kwadraten van correlaties tussen kenmerken en dimensies.

Met behulp van een HOMALS-analyse kan men de onderlinge samenhang van kenmerken binnen één groep van kenmerken onderzoeken. Toegepast op bijv. de wegkenmerken van een bepaald type weg, komt uit een dergelijke analyse het wegbeeld naar voren: welke kenmerken van de weg komen bij dit wegtype gezamenlijk voor?

We kunnen nu de analyse uitbreiden tot de verkeerskenmerken om te zien of bij een bepaalde combinatie van wegkenmerken ook steeds een zelfde patroon van intensiteiten hoort.

Vervolgens kunnen we ons afvragen of deze structuur ook een relatie heeft met de onveiligheid. Daartoe kunnen we de onveiligheidskenmerken toevoegen aan de groep weg- en verkeerskenmerken.

Een HOMALS-analyse geeft echter geen optimale beschrijving van de relatie tussen weg- en verkeerskenmerken aan de ene kant en de ongevalskenmerken aan de andere kant. HOMALS zoekt naar de overeenkomsten tussen alle kenmerken en niet naar een relatie tussen groepen kenmerken.

4.3. CANALS

Geeft HOMALS inzicht in de onderlinge structuur van kenmerken, het eigenlijke probleem is het vinden van de relaties tussen weg- en verkeerskenmerken enerzijds en ongevalskenmerken anderzijds. Bij een CANALS-analyse wordt deze samenhang tussen twee groepen van kenmerken optimaal beschreven.

De basisvraag bij CANALS is: welke combinatie van kenmerken uit groep 1 (bijvoorbeeld welke combinatie van wegkenmerken) hangt het meest samen met welke combinatie van kenmerken uit groep 2 (bijvoorbeeld welke combinatie van ongevalskenmerken?). Voor elk van deze twee combinaties, die zijn op te vatten als basiskenmerken, kan per observatie-eenheid een canonische score worden berekend uit de gemeten scores.

De canonische correlatie, d.i. de correlatie tussen de canonische scores op beide basiskenmerken, geeft aan hoe sterk die samenhang uiteindelijk is. Meerdere basiskenmerken (canonische assen of dimensies) zijn mogelijk, en leveren verschillende beelden van samenhang tussen de twee groepen.

Naast de schalingen zijn er gewichten voor de kenmerken, die aangeven welke kenmerken het meest hebben bijgedragen aan de oplossing. Beter is het echter om te kijken naar de correlaties tussen de kenmerken en de canonische scores van groep 2. Immers, als bijvoorbeeld twee kenmerken sterk onderling correleren, kan het ene een hoog gewicht krijgen en het andere niet. De correlaties zullen dan wel voor beide hoog zijn. De correlaties met de canonische scores van groep 1 zijn hier minder interessant, omdat groep 2 steeds als criteriumgroep wordt gekozen.

4.4. De relatie met andere analysetechnieken

Hoe verhouden de bij CANALS gevonden canonische correlaties zich tot de resultaten van meer conventionele technieken? We kunnen dat als volgt beschrijven.

Stel dat we de correlatie tussen twee kenmerken x_1 en y_1 berekenen en een correlatie r vinden. Indien we ter voorspelling van y_1 niet alleen x_1 kiezen maar ook andere kenmerken (bijv. x_2 en x_3), vinden we een multiple correlatie R die groter is dan of gelijk aan de absolute waarde van r : de drie kenmerken voorspellen y_1 minstens zo goed als x_1 alleen.

Verder geldt dat ook de (canonische) correlatie r^* die we met CANALS vinden tussen x_1 en y_1 , groter is dan r (er wordt net zo lang met de categorieën van x_1 en y_1 geschoven tot een maximale r wordt gevonden).

Ook geldt dat de multiple correlatie R^* tussen y_1 en de kenmerken x_1 , x_2 en x_3 van CANALS hoger is dan of gelijk aan r en R . Vinden we ook in de tweede set meerdere kenmerken, bijv. y_2 en y_3 dan is de canonische correlatie C in het algemeen nog hoger. In elk geval groter dan of gelijk aan de hoogste R -waarde van y_1 , y_2 of y_3 . In afbeelding 1 is dit in beeld gebracht.

De CANALS-analyse geeft dus de maximaal haalbare samenhang tussen de geanalyseerde kenmerken. Is de bij CANALS gevonden canonische correlatie erg laag, dan heeft de toepassing van andere technieken helemaal geen zin. Er zijn dan alleen zwakke verbanden aanwezig. In dat geval is wellicht een herbezinning op de geformuleerde onderzoeksproblematiek nodig.

5. EEN THEORETISCH VOORBEELD VAN EEN CANALS-ANALYSE

Om enig inzicht te geven in wat CANALS concreet met de onderzoeksgegevens doet, behandelen we hier een voorbeeld met fictieve gegevens. In bijlage 4 is een afdruk van de analyse te vinden. Na enige algemene informatie vinden we de datamatrix afgedrukt (raw data). De datamatrix bestaat uit scores voor objecten op drie onafhankelijke kenmerken (x_1 , x_2 en x_3) en één afhankelijk kenmerk (y). Uit de eerste kolom blijkt dat er 10 objecten zijn, uit de tweede en derde kolom dat x_1 en x_2 twee klassen hebben, uit de vierde kolom dat x_3 drie klassen heeft en uit de laatste kolom dat y eveneens drie klassen heeft. Er is één afhankelijk kenmerk, dus het betreft hier een bijzonder geval van canonische analyse, nl. multiple regressie-analyse. Er is dan ook maar één oplossingsdimensie mogelijk: de oplossing die dat ene y -kenmerk beschrijft m.b.v. de x -kenmerken.

Onder de datamatrix staan de marginale frequenties afgedrukt. Het is nuttig deze frequenties bij de schaling in het achterhoofd te houden. Soms vinden we extreem hoge schaalwaarden voor één categorie van een kenmerk. Het blijkt dan vaak dat die categorie bijv. slechts één observatie telt. In een dergelijk geval is het beter de categorie te verwijderen (bij andere categorie indelen of observatie-eenheid verwijderen).

Onder de frequenties volgt informatie over het meetniveau dat voor de kenmerken is opgegeven. Hier (evenals bij bijna alle kenmerken van het onderzoek) worden de kenmerken ordinaal geschaald.

Na nog wat informatie volgen de schalingen per kenmerk. De normering van de schaling is:

$$\bar{x} = (\sum_i x_i) = 0 \text{ en } \sum_i x_i^2/n = 1.$$

Bijv. voor kenmerk 1: $4 * -1.225 + 6 * .816 = 0$, dus:

$$\bar{x} = 0 \text{ en:}$$

$$4 * (-1.225)^2 + 6 * (.816)^2 = 10.$$

De gegevens staan dus in "standaardvorm".

Dan volgt de matrix met geschaalde waarden (alle categorienummers zijn vervangen door schaalwaarden). Vanwege de grote omvang is deze matrix bij de echte analyses niet afgedrukt.

Van hieraf is de analysebeschrijving metrisch. De oplossing is dus op te vatten als multiple lineaire regressie op de matrix met geschaalde waarden.

In termen van de lineaire algebra is het model nu:

$$y = Xa + e,$$

waarbij y de vierde kolom aanduidt en X de eerste drie kolommen; a geeft de rij van drie regressieparameters weer en e de afwijkingen van het model voor elke y .

De kleinste-kwadratenschatter voor a vinden we (klassiek) door:

$$\hat{a} = (X^T X)^{-1} X^T y;$$

in het geval van standaardscores als:

$$\hat{a} = R^{-1} r,$$

waarbij r de drie correlaties tussen de x -kenmerken en het y -kenmerk weergeeft en R de drie-maal-drie correlatiematrix van de x -kenmerken onderling.

De uit X afgeleide voorspelling \hat{y} van y vinden we als volgt:

$$\hat{y} = X\hat{a}$$

Dus:

$$y = \hat{y} + e;$$

$\sum_i e_i^2/n$, de variantie van de modelafwijkingen, wordt wel de stress genoemd, en is gelijk aan 1 minus $\sum_i \hat{y}_i^2/n$ (de verklaarde variantie) = 1 minus het kwadraat van de canonische correlatie.

In het voorbeeld:

$$1 - .738^2 = .456 \text{ (vgl. stress van de 4e iteratie).}$$

Hier worden voor \hat{a} de volgende waarden (regressiegewichten) gevonden:

$$\hat{a} = R^{-1} r = \begin{bmatrix} 1.000 & -.250 & -.171 \\ -.250 & 1.000 & -.099 \\ -.171 & -.099 & 1.000 \end{bmatrix}^{-1} \begin{pmatrix} .359 \\ -.527 \\ -.444 \end{pmatrix} = \begin{pmatrix} .144 \\ -.538 \\ -.472 \end{pmatrix}$$

De canonische correlatie is dus $\sqrt{r^T R^{-1} r} = \sqrt{.545} = .738$.

Delen we de gevonden gewichten door de canonische correlatie, dan vinden we de gewichten zoals CANALS die afdrukt, echter met omgekeerd teken. Dit komt doordat CANALS een oplossing zoekt voor $-y$ i.p.v. y (zie ook opmerking 1). Voor de op de canonische correlatie genormeerde gewichten w van CANALS geldt dat $w^T r = .738$ direct de canonische correlatie oplevert.

Voor \hat{y} vinden we:

$$\hat{y}_1 = (-.196 * -1.225) + (.725 * .816) + (.643 * .903) = 1.412$$

$$\hat{y}_{10} = (-.196 * .816) + (.725 * -1.225) + (.643 * -1.514) = -2.022$$

Voor de 10 waarden van \hat{y} vinden we zo:

$$\hat{y} = 1.41, .78, 1.18, -.07, -.70, -.54, -.70, -.14, .78, -2.02$$

De canonische correlatie is hier .738. Dit is de correlatie tussen y en \hat{y} , die aangeeft hoe goed de voorspelling is.

Is de canonische correlatie gelijk aan één dan is \hat{y} gelijk aan y en is y dus perfect "verklaard". De stress is dan gelijk aan nul en de variantie in de \hat{y} -waarden (verklaarde variantie) gelijk aan de variantie van y (totale variantie).

Soms wordt met behulp van de schaling een oplossing gevonden die op het eerste gezicht goed of zelfs perfect lijkt maar bij nader inzien triviaal blijkt. Dit is bijvoorbeeld het geval als één of enkele locaties zowel een klasse van een verklarend kenmerk als een klasse van een te verklaren kenmerk gemeen hebben, terwijl dit voor geen van de andere locaties geldt. Door die klassen dan een extreem hoge categoriescore te geven wordt technisch gezien de beste oplossing gevonden. Dit probleem kan worden opgelost door de desbetreffende locaties uit het bestand te verwijderen of door de klassen met een extreem hoge categoriescore te combineren met andere klassen van het kenmerk.

De correlaties tussen de x -kenmerken en $-\hat{y}$ bedragen $-.488$, $.711$ en $.605$. Deze correlaties geven aan hoezeer (de uit de drie x -kenmerken geconstrueerde) \hat{y} samenhangt met de individuele x -kenmerken. Omdat er slechts één y -kenmerk is, vinden we daarna voor de samenhang tussen y en \hat{y} de waarde van de canonische correlatie terug (nu met min-teken).

Over dit voorbeeld valt het volgende op te merken:

Opmerking 1: multiple regressie versus canonische analyse

Als er meerdere y 's zouden zijn, dan zouden onderaan de analyse de correlaties staan van de y 's met Y_b (de combinatie van y 's waarvoor een oplossing was gevonden), zoals er nu een combinatie X_a van x -kenmerken is die de oplossing geeft. Het regressie-model is dan:

$$Y_b = X_a + e \text{ (i.p.v. } y) = X_a + e.$$

Dit is in bijna alle analyses het geval. Hier is Y_b gelijk aan $y * -1 = -y$ en hebben we te maken met multiple regressie als bijzonder geval van canonische analyse.

Opmerking 2: schaling en regressie

Vanwege de duidelijkheid wordt telkens gesproken over schaling los van het regressieprobleem en omgekeerd. Wiskundig is er maar één probleem, met één unieke oplossing. Deze oplossing wordt iteratief gevonden door beurtelings de schaling te verbeteren en het regressieprobleem op te lossen, totdat de unieke schaling en oplossing voor het bijbehorende regressieprobleem is bereikt.

Opmerking 3: het teken van de schaling

We zien dat x_2 en x_3 beide een positief gewicht hebben en x_1 een negatief gewicht. Ook voor de correlaties geldt dit. Voor de schalingen geldt dat ze steeds van laag naar hoog zijn. Voor de interpretatie is het vaak gemakkelijk om schalingen in dezelfde richting te hebben. De oplossing verandert niet als we een schaling om-draaien. Stel dat we het teken van de schaling van kenmerk x_1 om-draaien; $(-1.225, .816)$ wordt dan: $(1.225, -.816)$. De tekens in eerste kolom van de matrix veranderen natuurlijk mee; het gewicht zal van tekens veranderen om de oplossing gelijk te houden ($-.196$ wordt $.196$); \hat{y} verandert niet, wel de bijbehorende correlaties ($-.488$ wordt $.488$ en $-.359$ wordt $.359$).

Als we de schaling van y van teken veranderen, moeten we ook het gewicht en de correlaties van teken veranderen (-1.0 wordt 1.0 , $-.738$ wordt $.738$ en -1.0 wordt 1.0). De oplossing is dan direct een oplossing van y en niet van $-y$.

6. GEOMETRISCHE WEERGAVE VAN LINEAIRE-REGRESSIEOPLOSSINGEN

Zoals we gezien hebben kunnen we de CANALS-analyse voorstellen als een tweestaps proces. Eerst vindt schaling van de kenmerken plaats. Hieruit resulteert een matrix van optimaal geschaalde variabelen. Op deze nieuwe datamatrix, die we opvatten als een matrix met metrische gegevens, passen we vervolgens lineaire regressie toe.

Uitgaande van de $X|y$ -matrix van geschaalde variabelen uit het theoretische voorbeeld, kunnen we nu plots maken van bijv. x_3 tegen y om de correlatie tussen deze kenmerken zichtbaar te maken.

De 10 observaties zijn dan punten (eindpunten van vectoren) in het x_3 - y -vlak. Zo kunnen we ook een afbeelding van de 10 punten maken in de driedimensionale ruimte x_2, x_3, y . En, hoewel we ons dat niet meer kunnen voorstellen, 10 punten in de vierdimensionale ruimte van x_1, x_2, x_3 en y . Op de laatste manier hebben we een perfecte afbeelding van de $X|y$ -matrix. De rijen van deze matrix zijn dan vectoren in een vierdimensionale ruimte van kenmerken. We noemen zo'n afbeelding wel een "puntenwolk-model".

We kunnen de matrix ook andersom afbeelden. In de ruimte van objecten beelden we dan de kenmerken (kolommen) af als vectoren. In het voorbeeld hebben we dan een tiendimensionale ruimte, met daarin vier vectoren. In een dergelijke afbeelding tekenen we meestal de gehele vector en niet alleen het eindpunt. We noemen dit model het "vectormodel".

Omdat we hier met genormeerde kenmerken werken, zijn alle vectoren van lengte 1 ($\sum x_i^2 / 10 = 1$ etc.). De correlatie tussen elk paar vectoren is nu gelijk aan de (richtings)cosinus van de hoek tussen de vectoren. De grootte van de correlatie komt overeen met de lengte van de projectie van de kenmerken op elkaar.

Bij het gewone lineaire-regressieprobleem kunnen we de vector van geobserveerde y -scores schrijven als $y = ax + e$. Meestal is er nog een extra constante, maar in het geval van standaardscores is deze afwezig; ax geeft nu de modelbeschrijving van y , ofwel de waarde die we voor y schatten uit x . We duiden deze aan met \hat{y} .

We kunnen y dus schrijven als $y = \hat{y} + e$ (een rij getallen y is te

schrijven als een uit x afgeleide rij getallen \hat{y} en een rij toevalsfouten of meetfouten e); e en \hat{y} zijn onafhankelijk van elkaar, dus correleren niet met elkaar. Tot zover betreft het een herhaling van wat in hoofdstuk 5 is te vinden. In de vectorafbeelding (afbeelding 2) zijn e en \hat{y} niet gecorreleerd, ze staan loodrecht op elkaar (de cosinus van een hoek van $90^\circ = 0$).

De cosinus van de hoek tussen y en x is gelijk aan de correlatie tussen y en x . De projectie van y op x levert \hat{y} . Het kwadraat van de lengte van deze vector is de proportie verklaarde variantie van y . De projectie op e laat de modelafwijkingen zien en daarmee de proportie niet-verklaarde variantie. Samen "verklaren" ze y volledig. We moeten ons realiseren dat de vectoren in een vlak liggen in de ruimte met observatie-eenheden als coördinaat-assen. Bij het onderhavige voorbeeld hebben we met meer dan één x te maken; \hat{y} ($= Xa$) is nu die vector in de deelruimte waarin x_1 , x_2 en x_3 liggen, die de grootste (multiple) correlatie heeft met y , dus de kleinste hoek met y heeft, dus de grootste projectie heeft op y en dus het meest van y verklaart.

Stel we hebben alleen te maken met x_1 , x_2 en y ; dan valt $\hat{y} = a_1x_1 + a_2x_2$ samen met de projectie van y op het vlak door x_1 en x_2 ; x_1 en x_2 mogen daarbij best met elkaar correleren, zodat de hoek tussen beide $\neq 90^\circ$. De geometrische weergave is te zien in afbeelding 3.

De kwadratensommen van e en \hat{y} zijn nog steeds gelijk aan de kwadratensom van y . Hoe kleiner de hoek tussen y en \hat{y} , hoe meer variantie in de waarden van y verklaard wordt door de x -kenmerken (hoe hoger dus ook de correlatie tussen y en \hat{y}).

Als we x_3 ook gebruiken bij de voorspelling, dan is \hat{y} een vector in de ruimte waarin x_1 , x_2 en x_3 liggen en is e een vector daar loodrecht op. Er verandert niets wezenlijks aan het beeld. Deze ruimtelijke voorstelling is voor een beter begrip wel handig.

Komt er ook nog een y extra bij, dan is de oplossing als volgt te lezen: kies die vector $\hat{y} = Xa$ in de ruimte van de x -kenmerken, die de grootste projectie heeft op het vlak door de twee y -kenmerken. De vector in het Y -vlak waarop \hat{y} wordt geprojecteerd, kunnen we

aanduiden met Y_b . Dit is de canonische as van de tweede set. De hoek tussen $\hat{y} = X_a$ en Y_b correspondeert nu met de canonische correlatie. Het is de hoek tussen beide canonische assen. Laten we de oorspronkelijke x-kenmerken even weg, en tekenen we alleen \hat{y} , dan ziet het beeld er uit als in afbeelding 4.

Er is nog een tweede oplossing mogelijk, die onafhankelijk is van de eerste oplossing. Voor de y-kenmerken ligt deze oplossing vast: het is de vector $Y_{b'}$, loodrecht op Y_b in het vlak door de kenmerken y_1 en y_2 . Bij meer dan twee x-kenmerken zijn er over het algemeen meerdere $X_{a'}$ vectoren loodrecht op X_a en Y_b te vinden. De $X_{a'}$ met de grootste projectie op $Y_{b'}$ geeft de tweede oplossingsdimensie.

We zullen nu nog wat nader ingaan op het meest voorkomende analyse-type in het onderzoek.

Wanneer we zoals in dit laatste voorbeeld uitgaan van een analyse met m onafhankelijke kenmerken (weg- en verkeerskenmerken) en 2 afhankelijke (ongevals-)kenmerken waarvan de schaling vast ligt, dan kunnen we de te analyseren dataset opvatten als een $n \times (m + 2)$ matrix. Daarin hebben n observaties (locaties) een score op $m + 2$ kenmerken. Deze datamatrix ziet er dan als volgt uit:

	kenmerk 1	kenmerk 2	kenmerk m	kenmerk $m+1$	kenmerk $m+2$
locatie 1	$x_{1, 1}$	$x_{1, 2}$	$x_{1, m}$	$x_{1, m+1}$	$x_{1, m+2}$
locatie 2	$x_{2, 1}$	$x_{2, 2}$	$x_{2, m}$	$x_{2, m+1}$	$x_{2, m+2}$
locatie 3	$x_{3, 1}$	$x_{3, 2}$	$x_{3, m}$	$x_{3, m+1}$	$x_{3, m+2}$
.....					
locatie n	$x_{n, 1}$	$x_{n, 2}$	$x_{n, m}$	$x_{n, m+1}$	$x_{n, m+2}$

We kunnen deze matrix van scores op twee manieren afbeelden als vectoren in een ruimte:

1. Als n rijvectoren in een $m + 2$ -dimensionale ruimte (opgespannen door de $m + 2$ kenmerken).
2. Als $m + 2$ kolomvectoren in een n -dimensionale ruimte (opge-

spannen door de n locaties). In dit laatste model zijn de onderlinge samenhangen tussen kenmerken af te leiden uit de hoeken tussen de vectoren. De correlatie tussen kenmerk a en kenmerk b kan worden geïnterpreteerd als een richtingscosinus van de hoek tussen de twee vectoren in het vectormodel die corresponderen met kenmerk a en kenmerk b . In zoverre is het vectormodel een afbeelding van de correlatiematrix die ontstaat als we uit de datamatrix de correlaties berekenen voor elk paar kenmerken. Hierbij moet bedacht worden dat bij correlaties de vectoren genormeerd zijn: alle vectoren hebben lengte 1.

Het canonische probleem dat telkens wordt gedefinieerd is: zoek die combinatie van onafhankelijke kenmerken die maximaal samenhangt met een combinatie van de afhankelijke kenmerken.

Vertaald in termen van het vectormodel betekent dit bij twee afhankelijke kenmerken: zoek die combinatie van onafhankelijke kenmerken (dit is een vector in het vectormodel) die een zo klein mogelijke hoek maakt met het vlak door de twee vectoren van de afhankelijke kenmerken.

In afbeelding 5 is het vlak door k_{m+1} en k_{m+2} (de afhankelijke kenmerken - bijv. totaal aantal ongevallen en aantal dodelijke ongevallen) geschetst. De vectoren van de onafhankelijke kenmerken k_1 t/m k_m zullen in het algemeen niet in datzelfde vlak liggen (indien dat voor een bepaald kenmerk wel het geval is, is er sprake van een perfecte voorspelbaarheid). Wanneer er meer dan drie locaties zijn, kunnen we dit vectormodel niet meer grafisch weergeven (meer dan drie dimensies zijn niet uit te beelden). Daarom geeft afbeelding 5 ook geen echt juist beeld; er is slechts getracht een indruk te geven van de relaties.

Gezocht wordt nu naar een vector in de ruimte waarin de vectoren k_1 t/m k_m liggen. De oplossing is nu die vector, die een lineaire combinatie is van de vectoren k_1 t/m k_m , en waarvan de projectie op het vlak door k_{m+1} en k_{m+2} zo groot mogelijk is. Stel dat in afbeelding 5 de vector a_1 die vector is en a'_1 zijn projectie op het vlak (a'_1 is dan een lineaire combinatie van de vectoren k_{m+1} en k_{m+2}), dan geldt dat de canonische correlatie van de eerste

dimensie van de CANALS-oplossing gelijk is aan de cosinus van de hoek α tussen a_1 en a'_1 .

Om nu te zien in hoeverre de individuele kenmerken k_1 t/m k_m met deze oplossing samenhangen, kan de cosinus van de hoek van elk van deze vectoren met a'_1 worden berekend. Deze correlaties worden vermeld als de correlaties tussen de observatiescores van de eerste set en de canonische scores van de tweede set. Omdat de hoeken tussen k_1 t/m k_m en de projecties ervan op het vlak (k'_1 t/m k'_m) altijd groter dan of gelijk aan de hoek α zijn, zijn deze correlaties altijd kleiner dan of gelijk aan de canonische correlatie. De correlaties tussen de observatiescores van elk kenmerk van de eerste set en de canonische scores van de eerste set komen overeen met de hoeken tussen k_1 t/m k_m en de vector a_1 .

Omdat de canonische assen onafhankelijk van elkaar moeten zijn, zoeken we voor de tweede dimensie een vector a_2 , in de ruimte waarin k_1 t/m k_m liggen, die loodrecht staat op a_1 en a'_1 . Deze vector a_2 is die lineaire combinatie van k_1 t/m k_m , waarvan de hoek β tussen a_2 en zijn projectie a'_2 zo klein mogelijk is onder de gestelde voorwaarden. Omdat β altijd groter is dan α , is de canonische correlatie van de tweede dimensie altijd kleiner dan of gelijk aan die van de eerste dimensie.

Voor een goede interpretatie van de analyses met twee afhankelijke variabelen zijn de correlaties tussen de observatiescores op k_{m+1} en k_{m+2} en de canonische scores van a'_1 en a'_2 van groot belang. Deze corresponderen met de hoeken tussen k_{m+1} en a'_1 respectievelijk a'_2 en met de hoeken tussen k_{m+2} en a'_1 respectievelijk a'_2 . In afbeelding 6 zijn twee voorbeelden gegeven van het vlak door k_{m+1} en k_{m+2} (bij verschillende oplossingen), met behulp waarvan dit nader toegelicht kan worden.

Afbeelding 6a geeft een voorbeeld van een oplossing waarbij de eerste dimensie vooral het ongevalstype k_{m+1} en de tweede dimensie vooral het ongevalstype k_{m+2} representeert. In een dergelijke analyse is voor de eerste dimensie de correlatie tussen de observatiescores van k_{m+1} en de canonische scores bijna gelijk aan één en tussen de observatiescores k_{m+2} en de canonische scores bijna nul. Voor de tweede dimensie geldt het omgekeerde.

Afbeelding 6b geeft een voorbeeld van een oplossing waarbij de eerste dimensie aangeeft wat de beide ongevalstypen gezamenlijk hebben en de tweede dimensie aangeeft waarin de ongevalstypen verschillen. Zowel voor k_{m+1} als voor k_{m+2} geldt dat de correlatie met a'_1 en a'_2 niet gelijk is aan 0 of 1. Bij de eerste dimensie zijn de tekens van de correlaties aan elkaar gelijk, bij de tweede dimensie tegengesteld.

Een uitvoerige beschrijving van de geometrische representatie van multivariate-analysetechnieken is te vinden bij Van de Geer (1967).

7. INTERPRETATIE VAN SAMENHANGEN EN BETROUWBAARHEID VAN DE OPLOSSINGEN

7.1. Interpretatie van samenhangen

Een CANALS-oplossing levert geen kant en klare verklaring van de samenhang tussen kenmerken. De getoonde samenhang moet nog begrijpelijk worden gemaakt. Het beeld moet passen in een theorie. De oplossing moet worden geïnterpreteerd. We gaan er daarbij vanuit dat de oplossing een betrouwbaar beeld geeft van bestaande samenhangen, op zijn minst met betrekking tot de steekproefgegevens die zijn verzameld. Op het vaststellen van deze betrouwbaarheid wordt nog nader ingegaan. Vanwege de eenvoud zullen we ons voorlopig beperken tot de interpretatie van het multiple-regressieprobleem. Bij de interpretatie van gevonden oplossingen zijn de volgende punten van belang:

- Welke kenmerken spelen bij de verklaring van het criterium een rol en welke niet?
- Hoe is de onderlinge samenhang van de kenmerken in die verklaring?
- Hoe zien de schalingen van de relevante kenmerken eruit; past deze schaling bij de verwachting die vanuit een theorie kan worden gegeven? Op de rol van de schalingen in de oplossing is al eerder ingegaan. Kort samengevat kunnen we zeggen dat de schaling van een kenmerk aangeeft hoe dit kenmerk maximaal bijdraagt aan de verklaring van het criterium. De schaling van het criterium speelt bij dit alles nog een aparte rol. We komen daar straks op terug. Wat de samenhang tussen de verklarende kenmerken betreft, is het van belang om te weten of kenmerken elkaar aanvullen bij het geven van de verklaring, danwel hetzelfde aspect van de verklaring laten zien.

Indien we de schaling even beschouwen als een vóórfase in de analyse, dan laat het probleem zich verder als "metrisch" beschrijven. Bij de interpretatie van een MLR-oplossing kunnen er twee benaderingen worden gebruikt. Een eerste benadering is stapsgewijze multiple regressie, een tweede benadering vindt plaats vanuit het begrip partiële correlatie.

7.1.1. Stapsgewijze multiple regressie

Bij multiple regressie gaat het om een totale verklaring van een aantal kenmerken gezamenlijk. Door aan de oplossing nieuwe kenmerken toe te voegen, of uit de oplossing bepaalde kenmerken weg te laten, kan de inbreng van bepaalde variabelen worden vastgesteld. Stapsgewijs kunnen we dan nagaan hoeveel bepaalde kenmerken toevoegen aan een gegeven verklaring, of in hoeverre bepaalde kenmerken de verklaring van andere kenmerken kunnen overnemen.

Een standaardzin is bijv. "wat voegt Y toe aan de verklaring van Z door X". Uitgangspunt hiervoor zijn de correlaties tussen x en z, tussen y en z en tussen x en y. De oplossing resulteert in één correlatiecoëfficiënt, de multiple-correlatiecoëfficiënt, die voor verschillende groepen kenmerken wordt berekend. Vergelijking van de multiple-correlatiecoëfficiënten van verschillende groepen kan de invloed van enkelvoudige kenmerken of groepen van kenmerken in een bepaalde oplossing aan het licht brengen. Om dit wat meer concreet te maken gaan we terug naar het theoretische voorbeeld van hoofdstuk 5.

Met de schaling van de kenmerken die in dat voorbeeld is gevonden, ligt matrix X en vector y vast en is de analyse verder te beschouwen als multiple lineaire regressie (MLR).

Willen we bijvoorbeeld weten of kenmerk x_3 overbodig is, dan kunnen we de derde kolom weglaten en opnieuw een oplossing zoeken voor:

$$\hat{a} = R^{-1} r$$

met de gereduceerde 2 maal 2 matrix $R = \begin{bmatrix} 1.000 & -.250 \\ -.250 & 1.000 \end{bmatrix}$ en

met $r = (-.359, .527)$ i.p.v. $r = (-.359, .527, .444)$. Voor \hat{a} vinden we dus:

$$\hat{a} = \begin{bmatrix} 1.0666 & .2666 \\ .2666 & 1.0666 \end{bmatrix} \begin{pmatrix} -.359 \\ .527 \end{pmatrix} = \begin{pmatrix} -.242 \\ .466 \end{pmatrix}$$

De multiple correlatie R is dan $\sqrt{r^T R^{-1} r} = \sqrt{.332} = .577$. Tot zover dit voorbeeld. Bij het praktijkvoorbeeld (hoofdstuk 9) zullen we hierop nog nader ingaan.

7.1.2. Partiële correlaties

Bij partiële correlaties wordt gezocht naar de invloed van bepaalde kenmerken op de relaties tussen andere kenmerken. Getracht wordt dan om deze invloed uit te schakelen.

Bekeken kan dan worden of bijv. de correlatie tussen y en z geme- ten over de gehele range van x-waarden, gelijk is aan diezelfde correlatie voor de y- en z-waarden die behoren bij een constante waarde van X. Getracht wordt dan de correlatie tussen y en z te corrigeren voor het effect van x. De volgende voorbeeldredenering geeft een beschrijving van dit partiële-correlatieprobleem:

"Dat in de gevonden oplossing het kenmerk verlichting een negatief effect op de veiligheid lijkt te hebben, is een gevolg van het feit dat verlichting vooral op drukke wegen voorkomt. Het positieve effect van verlichting valt weg tegen het negatieve effect van de intensiteit. Voor wegen met gelijke intensiteit zal dit verband positief blijken te zijn."

Technisch gesproken wordt dit: "Als we intensiteit uitpartialiseren, dan vinden we een positieve correlatie tussen verlichting en vei- ligheid in plaats van negatieve correlatie."

Bij partiële correlatie gaat het om de invloed van kenmerken (of groepen van kenmerken) op de individuele relatie tussen andere ken- merken en niet op de invloed ervan op de totaal-oplossing. De corre- laties worden vervangen door gecorrigeerde correlaties.

Op deze partiële correlaties kan eventueel weer een multiple- regressieanalyse worden toegepast, om een gezamenlijke verklaring te geven van kenmerken na correctie. Corrigeren van kenmerken is iets anders dan het weglaten van kenmerken uit een verklaring.

De partiële correlatie tussen y en z (met x uitgepartialiseerd) is gelijk aan:

$$r_{yz \cdot x} = \frac{r_{yz} - r_{yx} * r_{zx}}{(1 - r_{yx}^2) (1 - r_{zx}^2)}$$

$r_{yz.x}$ is de correlatie tussen y en z voor zover deze niet te verklaren is uit x.

In afbeelding 7 is dit grafisch weergegeven. In dit voorbeeld zien we dat er een positieve samenhang is tussen y en x, tussen z en x en tussen y en z. De partiële correlatie tussen y en z, $r_{yz.x}$, is negatief.

Bij MLR kan de vraag worden gesteld, in welke mate de oplossing beter wordt, als we z niet alleen proberen te verklaren met behulp van x maar met behulp met x en y samen. De oplossing is dan de hoek die z vormt met het vlak door x en y i.p.v. de hoek tussen z en x. Is deze hoek veel kleiner, dan draagt y veel bij aan de verklaring van z. Ligt z bijv. in het x-y-vlak, dan maakt het toevoegen van y de voorspelling van z perfect mogelijk. In de grafiek is de partiële correlatie tussen z en y dan -1; zou z tussen x en y in gelegen hebben, dan zou $r_{yz.x} = 1$. Dit is ook noodzakelijk, gezien de perfecte beschrijving van z door x en y. Voor het theoretische voorbeeld geldt bijv., als we het derde kenmerk niet weglaten maar uitpartialiseren, dat:

$$r_{x_1 y \cdot x_3} = \frac{r_{x_1 y} - r_{x_1 x_3} * r_{yx_3}}{\sqrt{(1 - r_{x_1 x_3}^2) (1 - r_{yx_3}^2)}} = \frac{.359 - (-.171) (-.444)}{\sqrt{(1 - .171^2) (1 - .444^2)}} = .321$$

$$r_{x_2 y \cdot x_3} = \frac{r_{x_2 y} - r_{x_2 x_3} * r_{yx_3}}{\sqrt{(1 - r_{x_2 x_3}^2) (1 - r_{yx_3}^2)}} = \frac{-.527 - (-.099) (-.444)}{\sqrt{(1 - .099^2) (1 - .444^2)}} = -.640$$

De partiële correlatie tussen x_1 en y is dus iets lager dan de gewone correlatie, die van x_2 en y wordt nog meer negatief.

In de praktijk is met name de invloed van de intensiteitskenmerken op de diverse relaties van belang. Voor een aantal onduidelijke analyseresultaten zijn deze partiële correlaties berekend. Daarnaast is m.b.v. stapsgewijze regressie-analyse getracht interpretaties te geven.

7.1.3. De schaling van de criteriumvariabelen

De schaling van y speelt een bijzondere rol. Dit blijkt uit het volgende. CANALS werkt met een additief model: $y = a_1x_1 + a_2x_2$ etc. Als de schaling van y lineair is, dan is de oplossing echt additief. Zijn alle x -schalingen ook lineair, dan levert CANALS niets anders op dan MLR zou hebben gedaan.

Zijn de x -schalingen anders, dan gaat de analyse meer in de richting van niet-lineaire multiple regressie. Bij nominaal geschaalde x -waarden ontstaat een soort variantie-analysemodel. Alle genoemde technieken zijn varianten van het lineaire (additieve) model.

Als de schaling van y logaritmisch is, dan betekent dit dat de oplossing eigenlijk multiplicatief is. De CANALS-oplossing geeft dan een multiplicatief model voor y , m.b.v. een additief model voor $\log y$.

Immers als $y = a \times b$, dan is $\log y = \log a + \log b$. Aangezien a en b vrij te kiezen parameters zijn, kan hiervoor een waarde in overeenstemming met $\log a$ resp. $\log b$ worden gekozen. Als we de x -schalingen nu nominaal kiezen, dan is het model verwant met bijv. log-lineaire analysemodellen. CANALS zoekt dan echter een oplossing voor de \hat{y} -waarden die gebaseerd is op de onderlinge relaties tussen de x -kenmerken (dus op de correlatiematrix R) en de correlatie van de x -kenmerken met $\log y$ (de correlaties r).

In de log-lineaire analysemodellen wordt, evenals bij variantie-analyse, ook nog gezocht naar hogere-orde-interacties. Bij kleine aantallen observaties, zoals hier het geval is, is dat echter niet realistisch. Bovendien is de methode dan ook maar toepasbaar bij analyses met weinig kenmerken, maximaal 5 à 6 in de praktijk. Als we de x -schalingen ordinaal kiezen, dan hebben we natuurlijk weer sterkere restricties.

Zo blijkt bij de analyses EG 1 en EG 1A (analyses over het totale aantal ongevallen van het complete analysebestand resp. het bestand van ongevallenlocaties op enkelbaanswegen voor gemengd verkeer) inderdaad sprake te zijn van een logaritme-transformatie. In afbeelding 8 is deze schaling weergegeven op half-logaritmisch papier. De schaling blijkt dan nagenoeg lineair te zijn. CANALS

laat dus zien dat het multiplicatieve model hier de voorkeur verdient boven het additieve. (Dit is één van de sterke kanten van CANALS als analysemodel: achteraf blijkt dit model het meest geschikt te zijn; we hoeven het er niet eerst in te stoppen!). Voor de oplossing heeft dit verder geen consequenties. De \hat{y} -waarden kunnen direct worden gebruikt als indicatie voor de onveiligheid. Deze log-transformatie vinden we niet altijd even mooi terug. In het praktische voorbeeld (vgl. bijlage 10, blad 3) geldt dit wel voor het totale aantal ongevallen, maar niet voor het aantal dodelijke ongevallen. Het lijkt alsof CANALS daar gezocht heeft naar een schaling van het aantal dodelijke ongevallen die zoveel mogelijk onafhankelijk is van het totale aantal ongevallen. Dit is waarschijnlijk ook de reden waarom de tweede oplossingsdimensie is te interpreteren als de dimensie die specifiek is voor het aantal dodelijke ongevallen.

7.2. Betrouwbaarheid van de oplossingen

De aard van de gegevens maakt dat het toepassen van strikt gedefinieerde modellen met sterke stochastische aannamen onrealistisch is. De ruimte die een model als CANALS biedt, maakt het mogelijk een realistischer beeld te krijgen van de onderlinge relaties tussen de vele kenmerken, maar brengt daardoor wel een aantal complicaties met zich mee. Allereerst wordt de interpretatie van de oplossing moeilijker. Men heeft daarbij, zoals al aangegeven, te maken met het effect van de schaling van de kenmerken op de oplossing en met het belang van die kenmerken voor de oplossing. Daarnaast wordt de vraag naar de betrouwbaarheid van de oplossing relevant. Doordat de analyse vrij zwakke eisen stelt aan de onderzoekgegevens, wordt het moeilijker om op grond van statistische theorieën iets te zeggen over de betrouwbaarheid van de oplossing. Toch is het belangrijk om te weten in hoeverre geringe veranderingen in de steekproef, veranderingen in de oplossing met zich mee kunnen brengen. Vrij recent, met name doordat men gebruik kon maken van de computer, zijn er een aantal procedures ontwikkeld die informatie leveren over de betrouwbaar-

heid van oplossingen langs empirische weg. Een zo'n methode, de "jack-knife"methode, knipt telkens een steekproefelement van de rest los en kijkt naar de oplossing van de overige steekproefelementen. Door alle mogelijkheden te bekijken verkrijgt men informatie over de variaties in de oplossing die daarvan het gevolg zijn. Bij de "monte-carlo"-methode wordt uit een welomschreven populatie m.b.v. een welomschreven kansverdeling een aantal random steekproeven getrokken. De analyseresultaten van deze steekproeven worden vergeleken met de kenmerken van de populatie. Bij de "bootstrap"-methode wordt de steekproef van gegevens van een bepaald onderzoek tot populatie verheven en worden uit deze steekproef random steekproeven getrokken met teruglegging. De analyse van deze steekproeven geeft informatie over de stabiliteit van de oorspronkelijke analyse. Als we aannemen dat de onderzoeksteekproef ook een random steekproef is uit de populatie, dan kunnen ook uitspraken worden afgeleid over de betrouwbaarheid waarmee het gevonden beeld van de steekproef de populatie beschrijft.

Bij HOMALS en CANALS lijkt de "bootstrap"-methode het meest aangewezen hulpmiddel voor het vaststellen van de betrouwbaarheid. Hoewel CANALS in de eerste plaats moet worden gezien als een techniek die bruikbaar is in de fase van de hypothese- en theorievorming, kan door gebruikmaking van de "bootstrap"-methode de betrouwbaarheid van bepaalde uitspraken worden onderzocht. Aangezien dergelijke analyses kostbaar en tijdrovend zijn, kan hiervan slechts op beperkte schaal gebruik worden gemaakt. Wanneer men, met name voor het nemen van maatregelen, meer zekerheid wenst over gevonden relaties, dan lijkt het toepassen van dergelijke technieken geboden.

8. ANALYSERESULTATEN EN MAATREGELEN

8.1. De analyses in relatie tot de locaties

Tot nog toe is er bij de interpretatie van de analyseresultaten alleen gekeken naar algemene verbanden tussen kenmerken van locaties. Voor het komen tot maatregelen zijn de individuele locaties uitgangspunt. We zijn dan vooral geïnteresseerd in de betekenis van de analyse voor elke locatie afzonderlijk, of eventueel voor groepen locaties.

We zullen hier nader ingaan op de informatie die de analyse oplevert over de locaties. We gaan daarbij uit van een canonische analyse waarin twee ongevalskenmerken worden "verklaard" met behulp van een groot aantal weg- en verkeerskenmerken. Er kunnen in dit geval maximaal twee dimensies worden gevonden waarop de gezochte samenhang beschreven is. Omdat de beide dimensies niet gecorreleerd zijn, beschrijven ze elk een onafhankelijk deel van de samenhang tussen de groep ongevalskenmerken en de groep weg- en verkeerskenmerken. Elk van de twee dimensies zal in het algemeen een combinatie van beide ongevalskenmerken beschrijven. De oplossing van de ene dimensie kan bijv. het gemeenschappelijke van beide ongevalskenmerken betreffen en die van de andere dimensie het verschil tussen beide kenmerken. Deze gewogen combinaties van ongevalskenmerken worden wel aangeduid met de canonische assen in de ruimte van ongevalskenmerken. Deze canonische assen correleren maximaal met telkens een andere lineaire combinatie van weg- en verkeerskenmerken: de canonische assen van de groep van verklarende variabelen. Tot dusverre is deze beschrijving niet anders dan die welke in hoofdstuk 6 is gegeven. Daar is echter niet ingegaan op de interpretatie van het model voor de individuele locaties. Voor de locaties geldt nu dat de canonische assen kunnen worden opgevat als nieuwe kenmerken, die de onderliggende structuur van de oorspronkelijke kenmerken weergeven m.b.t. de gevraagde verklaring. Deze nieuwe kenmerken kunnen niet direct worden gemeten. Wel kunnen de scores op deze kenmerken voor elke locatie worden berekend uit de scores op de oorspronkelijke ken-

merken. Deze canonische scores geven dan van elke locatie aan in welke mate die locatie het nieuwe kenmerk bezit. Verklaart de canonische as bijvoorbeeld de ernst van de ongevallen, dan laat de canonische score zien in welke mate een locatie kenmerken bezit die met de ernst samenhangen.

Voor elk van de twee oplossingsdimensies geldt dus dat er canonische scores kunnen worden berekend voor de groep weg- en verkeerskenmerken van de locaties en de groep ongevalskenmerken. Als we de correlatie berekenen tussen deze canonische scores van de eerste dimensie, dan vinden we de canonische correlatie terug die bij deze dimensie hoort. Hiermee wordt dan ook duidelijk wat met de canonische correlatie wordt bedoeld. Als we gebruik maken van een puntenwolkvoorstelling, kunnen we de locaties nu plotten als punten in een vlak waarvan de x-as overeenkomt met de canonische as van de ongevalskenmerken en de y-as overeenkomt met de canonische as van de weg- en verkeerskenmerken. Iets dergelijks kan ook worden gedaan voor bijv. de tweede dimensie. Het beeld van deze oplossing is compleet anders dan dat van de eerste dimensie. De canonische scores van de eerste dimensie zijn niet gecorreleerd met die van de tweede dimensie. Dit geldt zowel voor de x-scores en y-scores onderling als voor de x-scores van de ene dimensie en de y-scores van de andere dimensie.

Willen we van elk beeld weten wat het betekent, dan moeten we kijken welke kenmerken correleren met de canonische as. Een dergelijke interpretatie van de oplossing is onontbeerlijk voor het zoeken naar maatregelen die gericht zijn op verbeteringen van de locaties met behulp van de oorspronkelijke kenmerken. Verder moet men zich realiseren dat maatregelen die afgeleid zijn uit de eerste dimensie, geen oplossing geven voor problemen van de tweede dimensie en omgekeerd. In het algemeen echter zijn er ook geen negatieve effecten te verwachten, tenzij het maatregelen betreft die ingrijpen op kenmerken die voor beide dimensies van belang zijn, maar in tegengestelde zin. Zo zou het kunnen voorkomen dat maatregelen die gericht zijn op een reductie van het totale aantal ongevallen een nadelig effect hebben op de ernst van ongevallen. Om na te gaan in hoeverre met een dergelijke moge-

lijkheid rekening moet worden gehouden, is bij de analyses steeds het totale aantal ongevallen als ongevalskenmerk naast het specifieke kenmerk meegenomen.

8.2. Keuze van locaties voor maatregelen

Om aan te geven hoe de locaties worden gekozen die in aanmerking komen voor maatregelen, zullen we ons voor het gemak even beperken tot een oplossing in één dimensie. Stel dat we de canonische scores hebben berekend voor alle locaties. We kunnen dan de canonische scores van de groep weg- en verkeerskenmerken opvatten als voorspellers voor de canonische scores van de groep ongevalskenmerken. Stel dat deze laatste canonische scores kunnen worden opgevat als een score voor de ernst van een ongeval (omdat vooral de kenmerken waarin de ernst van het ongeval is vastgelegd ermee samenhangen). De canonische scores van de groep weg- en verkeerskenmerken geven dan voor elke locatie aan wat de verwachte ongevalsernst is. De canonische scores van de groep ongevalskenmerken geven aan wat de geconstateerde ongevalsernst is op diezelfde locatie. De mate waarin beide waarden overeenstemmen, wordt uitgedrukt in de waarde van de canonische correlatie. Deze overeenstemming wordt door allerlei toevalsfactoren beïnvloed. De oplossing zal niet altijd optimaal zijn, maar zelfs als ze dit wel is en er bijv. inderdaad gesproken kan worden van potentieel gevaarlijke locaties, dan behoeft dit nog niet altijd te betekenen dat er al veel ernstige ongevallen zijn gebeurd. Het is echter wel van belang na te gaan of op zo'n potentieel gevaarlijke locatie maatregelen gewenst zijn. Ook van locaties die niet in de analyse zijn opgenomen kan deze canonische score worden bepaald, als hun weg- en verkeerskenmerken bekend zijn. Deze kunnen dan ook voor maatregelen in beschouwing worden genomen. Bij deze generalisering naar andere locaties moeten we er ons echter van vergewissen dat het gaat om locaties van dezelfde soort. Als dit niet het geval lijkt te zijn, is het verstandig de analyse voor deze locaties over te doen, of na te gaan hoe hoog de canonische correlatie voor deze locaties is.

Om na te gaan welke locaties in aanmerking komen voor een dergelijke beschouwing is het nuttig de locaties te ordenen naar de canonische scores van de groep weg- en verkeerskenmerken. Vooral als er veel locaties in aanmerking komen, zoals in deze analyses het geval is. Locaties met een hoge score zijn te beschouwen als potentieel de gevaarlijkste locaties, locaties met een lage score als potentieel de veiligste locaties. Zijn de locaties eenmaal gesorteerd op hun canonische score, dan zouden we daarna kunnen besluiten om bijv. de 50 gevaarlijkste locaties als rode punten in te tekenen op een wegenkaart, en de 50 veiligste locaties als groene punten op dezelfde wegenkaart. Hiervoor moet men de locatiedecode kennen. Uit zo'n weergave zou kunnen blijken dat veel gevaarlijke locaties op eenzelfde route liggen. In zo'n geval zou gedacht kunnen worden aan meer structurele maatregelen. Voor de overige locaties komen locale maatregelen meer in aanmerking. Een dergelijke benadering wordt bijv. gebruikt bij het onderzoek naar black spots in Groot-Londen. Daar echter zijn dergelijke kaarten slechts gebaseerd op geconstateerde aantallen ongevallen en niet, zoals hier wordt voorgesteld, op de weg- en verkeerskenmerken die als onveilig worden beschouwd. Indien meerdere onveiligheidsaspecten tegelijkertijd in beschouwing worden genomen, dan zal integratie van de gegevens plaats dienen te vinden alvorens tot maatregelen wordt overgegaan. Hier zou dat betekenen dat de kaarten niet voor elk aspect geïsoleerd worden bekeken, maar worden gerelateerd aan de kaarten van andere aspecten.

8.3. Het zoeken naar maatregelen voor gekozen locaties

Om voor de gekozen locaties te bepalen welke maatregelen er zouden moeten worden genomen, is het van belang om van elke locatie te weten welke kenmerken met name die locatie gevaarlijk maken. Behalve een locatiedecode is dus een overzicht nodig van de kenmerken die voor de oplossing het meest relevant zijn. Daardoor zou het ook mogelijk zijn om bij een maatregel die men op grote schaal wil toepassen, te zien welke locaties het meest in aanmerking komen. Deze maatregelen, of pakketten van maatregelen,

moeten dan niet op geïsoleerde kenmerken gericht zijn, maar oplossingsgericht zijn. De oplossing laat over het algemeen zien welke structuur van kenmerken gevaarlijk is. Dit wordt voor een individuele locatie uitgedrukt in de canonische score op de as van de wegkenmerken. De mate waarin deze score door beïnvloeding van de relevante wegkenmerken kan worden omgebogen, bepaalt de effectiviteit van een maatregelenpakket. Maatregelen die ten aanzien van slechts één aspect worden genomen, zullen in de regel minder efficiënt zijn. Voor een aantal locaties kan het natuurlijk voorkomen dat slechts beperkte maatregelen mogelijk zijn. Tot zover hebben we gesproken over een analyse met betrekking tot slechts één veiligheidsaspect. Voor men tot een definitieve keuze van maatregelen voor locaties komt, is het verstandig ook na te gaan of op grond van een andere analyse wellicht ook maatregelen gewenst lijken. Afstemming van dergelijke pakketten van maatregelen op elkaar lijkt dan gewenst. Sommige maatregelen zullen positief beoordeeld kunnen worden met betrekking tot meerdere aspecten, maar andere kunnen op het ene aspect een positief effect en op het andere een negatief effect hebben.

8.4. Beschrijving van een mogelijke procedure

De volgende werkwijze kan worden gevolgd ter bepaling van de locaties waar zich problemen hebben voorgedaan of naar alle waarschijnlijkheid zullen voordoen met betrekking tot een bepaald aspect van de veiligheid.

Deze locaties kunnen worden geordend naar de grootte van de geschatte ongevallenkans die resulteert uit de analyses. Van de locaties met een hoge ongevallenkans wordt nagegaan of deze vooral op bepaalde routes liggen. Ook zal worden bepaald of de gekozen locaties op grond van analyses met betrekking tot andere aspecten van de veiligheid prioriteit dienen te krijgen.

Vervolgens wordt op basis van de interpretatie van de analyse onderzocht, welke maatregelensoorten in aanmerking komen. Van de kenmerken van de specifieke locaties wordt nagegaan of die in ongunstige zin afwijken van bestaande richtlijnen, normen of criteria.

Concreet betekent dit, dat bij een analyse waarin naast het totale aantal ongevallen ook een of meer specifieke ongevalskenmerken zijn meegenomen, de volgende stappen moeten worden gedaan:

1. Bereken voor de geanalyseerde locaties de canonische scores voor de weg- en verkeerskenmerken en de canonische scores voor de ongevalsdimensies. Dit moet voor elke relevante dimensie gebeuren.
2. Sorteert de locaties naar de hoogte van de canonische score voor weg- en verkeerskenmerken voor elke dimensie apart.
3. Druk voor elke dimensie behalve de locatiecode de volgende kenmerken af:
 - canonische score voor de weg- en verkeerskenmerken (sorteercriterium);
 - canonische score voor ongevallen op die dimensie;
 - totaal aantal ongevallen;
 - aantal ongevallen dat behoort bij het geanalyseerde veiligheidsaspect;
 - weg- en verkeerskenmerken (categorie-aanduiding en/of categorie-scores) in de volgorde van belangrijkheid van het betreffende kenmerk voor de oplossing.
4. Teken op kaart van Noord-Brabant in de bovenste 25 of 50 en de onderste 25 of 50 locaties als rode resp. groene stippen. Idem voor de andere analyses.
5. Ga na of deze stippen zich concentreren op bepaalde routes. Dit geeft een indicatie voor structurele maatregelen.
6. Ga voor de gekozen locaties na welke kenmerken tot de hoge canonische scores leiden.
7. Bepaal mogelijke maatregelsoorten op basis van de eerder verrichte analyses en van de specifieke locatiekenmerken; ga na welke beïnvloedbare kenmerken er zijn voor elk van de gekozen locaties.
8. Ga na of deze beïnvloedbare locatiekenmerken in ongunstige zin afwijken van de bestaande richtlijnen, normen of criteria.
9. Bepaal de mogelijke maatregelen en weeg deze tegen elkaar af. Eventuele neveneffecten, ook elders, dienen in beschouwing te worden genomen.
10. Bepaal de uiteindelijke maatregel.

8.5. De effectiviteit van maatregelen

Over de effectiviteit van maatregelen bestaat altijd in meer of mindere mate onzekerheid. Maatregelen kunnen positieve of negatieve neveneffecten hebben op de veiligheid en op andere aspecten van het verkeersgebeuren. Verder zijn er soms uitstralings-effecten van maatregelen die zich onttrekken aan een evaluatie. Juist door deze onzekerheden is het nodig om zo zorgvuldig mogelijk het wel meetbare effect van maatregelen vast te stellen. Bij het toepassen van maatregelen wordt ervan uitgegaan dat de onveiligheidsanalyse juist is geweest. Hier gaan we ervan uit dat we een betrouwbaar beeld hebben gekregen van de relaties tussen de onveiligheid en de kenmerken van de locaties. Verder dat dit beeld niet storend is beïnvloed door factoren die buiten de analyse zijn gebleven. En bovendien dat een juiste interpretatie aan het beeld is gegeven, zodat de echte oorzaken zijn opgespoord en niet schijnbare oorzaken (doordat bijvoorbeeld toevallig met de oorzaak correlerende kenmerken voor de oorzaak van de onveiligheid worden aangezien). Ten slotte wordt ervan uitgegaan dat de maatregel het beoogde effect zal hebben, met andere woorden dat de veronderstelde oorzaak van de onveiligheid zal worden weggenomen.

Wanneer we op grond van de analyse het verwachte effect van een maatregel op de veiligheid berekenen, gaan we ervan uit dat die maatregel het beoogde effect op bijv. de snelheid, de intensiteit enz. behaalt. De maatregel zal dan gericht zijn op een reductie van de onveiligheid, zoals deze tot uitdrukking komt in de canonische score van de beschouwde locatie op de canonische as van de wegkenmerken. De mate waarin de combinatie van wegkenmerken - uitgedrukt in die canonische score - is veranderd, bepaalt de voorspelde reductie in de ongevallenkans. Voor de evaluatie van maatregelen maar ook voor het verifiëren van de interpretatie van een analyse is het noodzakelijk om deze verwachte effectiviteiten te berekenen en te vergelijken met de gevonden resultaten bij en voor-na studie. Bij de beschrijving van het

praktijkvoorbeeld in hoofdstuk 9 zal ook deze uitwerking worden opgenomen.

9. EEN PRAKTIJKVOORBEELD

Om de in het vorige hoofdstuk behandelde methode van onderzoek te verduidelijken bespreken we hier een van de analyses van de analyse raaien. Resultaten van dit deel van het onderzoek zijn te vinden in SWOV (1981).

Als uitgangspunt is gekozen de analyse van het totale aantal ongevallen en het aantal dodelijke ongevallen als functie van de weg- en verkeerskenmerken op enkelbaanswegen met gemengd verkeer (EG IAD). De analyse betreft gegevens van 685 ongevalslocaties. De gegevens werden in relatie tot 22 weg- en 4 verkeerskenmerken geanalyseerd in twee dimensies.

9.1. De presentatie van de analyseresultaten

In bijlage 5 zijn de belangrijkste resultaten van deze analyse in tabelvorm weergegeven. In SWOV (1981) zijn alle analyseresultaten op deze gecomprimeerde wijze weergegeven. De tabel levert voor iedere dimensie de hoogste waarden van de correlaties van de kenmerken met de tweede canonische as. In bijlage 6 zijn deze correlaties voor alle 26 + 2 kenmerken weergegeven.

De tabel bevat bovendien een aanduiding van de schaling van de belangrijkste kenmerken. De volledige schaling van de kenmerken van de eerste set (1 t/m 26) is opgenomen in bijlage 7.

In SWOV (1981) worden van analyses in twee dimensies de resultaten ook vaak in een figuur (plotje) gepresenteerd. De figuur die bij deze analyse behoort, is opgenomen in bijlage 8. Het plotje is een weergave van de projecties op het vlak door k_{m+1} en k_{m+2} .

9.2. Interpretatie van de gevonden analyseresultaten

9.2.1. De analyseresultaten

De eerste oplossingsdimensie blijkt toevallig nagenoeg samen te vallen met kenmerk y_1 (totaal aantal ongevallen) en de tweede nagenoeg met y_2 .

y_1 en y_2 zijn kennelijk zo geschaald, dat ze praktisch niet correleren. Dit wil echter niet zeggen, dat het totale aantal ongevallen niet correleert met het aantal dodelijke ongevallen. Het y -vlak ziet er dan globaal uit als weergegeven in afbeelding 9. Een volledige afbeelding is te vinden in bijlage 8.

We maken dus slechts een kleine fout als we zeggen dat de eerste dimensie het totale aantal ongevallen verklaart en de tweede dimensie het aantal dodelijke ongevallen. De canonische correlaties zijn resp. .412 en .266. Het totale aantal ongevallen is dus beter voorspelbaar dan het aantal dodelijke ongevallen.

Opmerkelijk is dat verlichting kennelijk positief samenhangt met het totale aantal ongevallen (geen verlichting - weinig ongevallen; wel verlichting - veel ongevallen) maar negatief met het aantal dodelijke ongevallen. Dit wordt duidelijker als we de projectie van x_{15} op het vlak door y_1 en y_2 bekijken.

Proberen we nu het analyseresultaat in termen van samenhang tussen de kenmerken te beschrijven, dan valt te constateren dat dodelijke ongevallen o.a. samenhangen met de kenmerken 24, 25, 15, 14, 20, 1 en 12. Wanneer we de schalingen erbij betrekken, betekent dit dat dodelijke ongevallen gebeuren bij:

- een hoog percentage vrachtverkeer;
- een hoge (brom)fietsintensiteit;
- afwezigheid van verlichting;
- wegen met een verhardingsbreedte kleiner dan 6,5 m;
- aanwezigheid van bochten.

Aan de hand van de bijlagen 5 en 8 kunnen we ook voor het totale aantal ongevallen een dergelijke opsomming maken. We laten dat hier achterwege.

9.2.2. Multiple regressie en partiële correlatie

Over het tot nu toe behandelde voorbeeld is geen stapsgewijze multiple regressie of partiële correlatie berekend. Om toch een praktijkvoorbeeld te laten zien maken we even gebruik van een andere analyse, namelijk EG 1A extra. Dat is een analyse over

hetzelfde bestand van ongevalslocaties op wegen voor gemengd verkeer, maar dan met alleen het totaal aantal ongevallen als criteriumvariabele. In bijlage 9, blad 1 zijn naast de oorspronkelijke correlaties met de tweede canonische as, de partiële correlaties met die as berekend onder uitpartialisering van verlichting (15), intensiteiten (23 t/m 26) en beide. Daarbij zijn de belangrijkste verschuivingen nog eens apart aangegeven.

Zo blijkt bijvoorbeeld dat de invloed van reflectoren sterk afneemt als de intensiteit wordt uitgepartialiseerd. Kennelijk komen reflectoren vooral op drukke plaatsen voor. Waar veel verkeer is, gebeuren veel ongevallen. De invloed van de obstakelafstand, lichtmasten, verhardingssoort en erfaansluitingen neemt toe. Een verklaring met behulp van deze kenmerken lijkt daarom beter te zijn dan een verklaring met behulp van reflectoren. Wordt de verlichting uitgepartialiseerd, dan verandert vooral de invloed van de kenmerken zicht, lichtmasten, reflectoren en erfaansluitingen. De invloed van de eerste drie kenmerken neemt af, die van erfaansluitingen neemt toe.

Worden verlichting én intensiteit uitgepartialiseerd, dan zien we dat zicht, obstakelafstand en erfaansluitingen zich volgens verwachting gedragen. Over het kenmerk lichtmasten valt op te merken dat de invloed daarvan bijna even sterk toeneemt als wanneer alleen de verlichting wordt uitgepartialiseerd. Het blijft dus een belangrijk kenmerk.

Tot slot kan nog worden opgemerkt, dat de invloed van het belangrijkste kenmerk - het kenmerk bocht - niet veel verandert door het uitpartialiseren.

In bijlage 9, blad 2 is stapsgewijze multiple regressie over dezelfde analyse uitgevoerd. Conclusie 1 daarbij betekent dat een deel van de ongevallen ontstaat door veel verkeer en een deel door "gebreken" aan de weg. Beide effecten staan los van elkaar. De kenmerken 1, ..., 21 verklaren daarom evenveel wanneer ze apart worden opgenomen in een analyse, als wanneer ze worden toegevoegd aan een verklaring met behulp van intensiteit (23 t/m 26). Voor dit bestand geldt, dat die kenmerken veel meer van de onvei-

ligheid verklaren dan de intensiteiten doen. Dit zou samen kunnen hangen met het feit dat het hier een bestand van ongevalslocaties en niet van alle locaties betreft.

9.2.3. De beoordeling van de schaling

We gaan nu weer terug naar het oorspronkelijke voorbeeld: EG 1AD, de analyse over dodelijke ongevallen.

Om na te kunnen gaan of een bepaald kenmerk in aanmerking komt voor maatregelen, is genuanceerde kennis nodig over de categorieën van dat kenmerk. Daarom zijn behalve de hierboven genoemde globale indicaties voor het effect ook de categoriescores zelf van belang. Deze kunnen vaak het beste beoordeeld worden door in een grafiek de oorspronkelijke categorieën uit te zetten tegen de schaling. In bijlage 10 is dit voor de belangrijkste kenmerken gebeurd. Op blad 1 van deze bijlage staan de categoriescores voor de vijf belangrijkste kenmerken van de eerste dimensie afgebeeld. Op blad 2 die voor de tweede dimensie. Op blad 3 zijn deze afbeeldingen gegeven voor de ongevalskenmerken. De schaling van deze kenmerken maakt vooral onderscheid tussen één of meer dodelijke ongevallen en niet tussen één of geen dodelijk ongeval.

Op basis van de grafieken van bijlage 10, blad 1 en 2 kan men vaststellen welke de "onveilige" categorieën van een wegkenmerk zijn. Dat zijn in principe die categorieën, die bij een positieve correlatie van het kenmerk met het ongevalskenmerk hetzelfde teken hebben als de categorie met de grootste aantallen ongevallen; bij een negatieve correlatie hebben ze juist een tegengesteld teken. Kenmerk 12 correleert positief met het totale aantal ongevallen; de klassen 4, 5 en 6 zijn dan ook als onveilig aan te merken. Kenmerk 1 correleert negatief met het totale aantal ongevallen; hier zijn de klassen 1 en 2 onveilig.

9.3. De betrouwbaarheid van de oplossing

Voor men maatregelen neemt, is het soms gewenst meer informatie te hebben over de vatbaarheid van de oplossing voor toevalsfluc-

tuaties. Voor de gegevens van analyse EG IAD is een "bootstrap"-studie verricht om na te gaan hoe stabiel de gevonden oplossing is. Deze studie is bedoeld om de betrouwbaarheid van de canonische correlatie vast te stellen en om na te gaan hoe stabiel de bijdrage van de individuele kenmerken aan de oplossing is bij vergelijkbare analysegegevens. De gegevens voor elke analyse worden verkregen door een steekproef uit de oorspronkelijke gegevens te trekken.

Daarnaast is ook gekeken naar de hoogte van de canonische correlatie. Deze bepaalt namelijk de betrouwbaarheid van de voorspelling over de onveiligheid van locaties op grond van hun weg- en verkeerskenmerken.

9.3.1. De "bootstrap"-studie

Uit de 685 locaties is tien maal een steekproef getrokken met teruglegging (locaties kunnen dus ook meerdere malen in de steekproef terechtkomen, maar ze kunnen ook geheel ontbreken). Ook de steekproeven omvatten telkens 685 locaties. De steekproeven worden met CANALS geanalyseerd en de resultaten van deze analyses worden vervolgens vergeleken met de resultaten van de oorspronkelijke analyse.

Over deze steekproeven zijn de gemiddelde canonische correlatie en de gemiddelde correlaties van de kenmerken met de canonische assen berekend. In bijlage 11 zijn voor beide dimensies de gemiddelde correlaties met de tweede canonische as gegeven, evenals de bijbehorende betrouwbaarheidsgrenzen. Deze waarden zijn berekend met behulp van de door Fisher voorgestelde z-transformatie.

Uit de "bootstrap"-studie blijkt, dat gemiddelden van de canonische correlatie van de steekproeven .47 voor de eerste dimensie en .34 voor de tweede dimensie zijn. (Deze waarden zijn hoger dan die welke werden gevonden bij de oorspronkelijke analyse. We komen hier in paragraaf 9.3.2. nog op terug.)

Uit bijlage 11 blijkt dat de oplossingsdimensie zeer stabiel de omvang en ernst van de ongevallen beschrijven: de gemiddelde corre-

latie van kenmerk 27 met de eerste dimensie is .992 en die van kenmerk 28 met de tweede dimensie is .987. Op grond hiervan mag worden verwacht dat dan ook de samenhang van de weg- en verkeerskenmerken met de dimensies stabiel is.

We zien dat deze stabiliteit voor de eerste dimensie groter is dan voor de tweede dimensie. Dit is gezien de hogere canonische correlatie ook wel te verwachten. De kenmerken 12, 15 en 9 blijven de hoogste samenhang hebben, de kenmerken 1 en 21 geven wat lagere waarden te zien en de kenmerken 14 en 5 hogere waarden. Kenmerk 14, waaraan geen orde-restricties zijn opgelegd, speelt een wat merkwaardige rol. Bij dit kenmerk (evenals bij kenmerk 20) geldt dat de schaling niet altijd even consistent is. Bij de andere kenmerken, die alle ordinaal geschaald zijn, is dat wel het geval. Dit zou tot gevolg kunnen hebben dat door de grotere schalingsvrijheid nu eens een schaling wordt gekozen die meer samenhangt met de eerste dimensie, dan weer een die meer samenhangt met de tweede dimensie. Bij ordinale schaling is deze mogelijkheid veel kleiner.

Voor de tweede dimensie geldt dat de kenmerken 24, 25 en 15 belangrijk blijven, zij het in mindere mate, dat kenmerk 14 gelijk blijft en dat kenmerk 20 verdwijnt. De kenmerken 1, 5 en 7 nemen in betekenis toe.

9.3.2. De hoogte van de canonische correlatie

We hebben gezien, dat de gemiddelde canonische correlatie voor de 10 "bootstrap"-studies - verder aan te duiden als \bar{c}_b - hoger was dan de canonische correlatie van de oorspronkelijke analyse over het analysebestand (c_a). Doordat de "bootstrap"-steekproeven worden getrokken met teruglegging, zijn de bestanden daarvan homogener en is de canonische correlatie dus hoger.

Dit verschil kan nog op een andere wijze worden gebruikt. Het analysebestand is immers niet de werkelijke populatie. Hadden we de werkelijke populatie van alle mogelijke wegstukken van enkelbaanswegen met gemengd verkeer kunnen analyseren, dan was de canonische correlatie (c_p) weer iets lager geweest dan die voor het

analysebestand. Bij het berekenen van effecten van maatregelen moet men zich eigenlijk op de populatiegegevens baseren, en niet direct op de uitkomsten van het analysebestand. Bewezen is, dat het verschil tussen de (niet berekende) analyseresultaten voor de populatie en de resultaten voor het analysebestand gelijk is aan het verschil tussen de resultaten voor het analysebestand en het gemiddelde van de "bootstrap"-resultaten. Voor de canonische correlatie geldt dus in formulevorm:

$c_p - c_a = c_a - \bar{c}_b$, ofwel: $c_p = 2 * c_a - \bar{c}_s$. De canonische correlaties voor de populaties worden dan respectievelijk .35 en .20.

Verwachtingen over de effectiviteit van maatregelen dienen dus gebaseerd te zijn op deze waarden, die erg laag lijken. Een van de redenen hiervoor is, dat in de analyse gebruik is gemaakt van zeer korte weggedeelten, waarop dus een gering aantal ongevallen is gebeurd. Zelfs als we ervan uitgaan dat de ongevallenkans perfect te voorspellen is uit de weg- en verkeerskenmerken, zal de maximale canonische correlatie voor de eerste dimensie daardoor niet veel hoger zijn dan de al genoemde waarde van .35.

Voor iedere locatie afzonderlijk is de voorspelbaarheid inderdaad laag, maar voor meerdere locaties tezamen zal de voorspelbaarheid groter zijn. Als we de canonische correlatie niet zouden berekenen over de afzonderlijke locaties maar over groepen van locaties met dezelfde weg- en verkeerskenmerken, dan zou de canonische correlatie veel hoger zijn. Hoe meer locaties er zijn per groep, des te betrouwbaarder zal de ongevallenkans voor de groep kunnen worden geschat, waardoor de canonische correlatie hoger wordt. De "bootstrap"-analyse laat dan ook zien dat de stabiliteit van de oplossing veel groter is dan de hoogte van de canonische correlatie in eerste instantie doet vermoeden.

Naar de invloed van beide componenten (het aantal ongevallen per locatie en het aantal locaties per groep) op de canonische correlatie wordt nader onderzoek uitgevoerd. Er wordt getracht hierin meer inzicht te krijgen met behulp van de "monte-carlo"-methode. Voorlopige resultaten uit dit onderzoek laten zien dat bij een perfecte samenhang tussen de populatieparameters van beide cano-

nische assen en een vergelijkbaar aantal ongevallen op alle locaties tezamen een gemiddelde canonische correlatie voor de eerste dimensie mag worden verwacht van omstreeks .45. Wanneer de periode waarover ongevalgegevens worden verzameld, zou worden verlengd met een factor 10, dan zou deze correlatie oplopen tot .84. Bij een oneindig lange periode is deze correlatie natuurlijk gelijk aan 1: bij een perfecte samenhang vinden we dan een perfecte voorspelling.

In vergelijking met de op deze wijze geschatte maximale canonische correlatie van .45 duidt een canonische correlatie van .35 op een relatief sterke samenhang tussen de populatieparameters. Er wordt nog nagegaan in hoeverre dit probleem op een meer analytische wijze kan worden benaderd.

9.4. Analyseresultaat m.b.t. de locaties

We zullen ons in dit voorbeeld verder beperken tot de informatie zoals die is verkregen uit de oorspronkelijke analyse. Bij deze analyse zijn voor elke locatie de scores op de beide canonische assen van elke dimensie berekend. De locaties zijn eerst gesorteerd naar de canonische scores op de eerste canonische as van de eerste dimensie. De locaties kunnen nu worden gerangschikt naar de aantallen ongevallen die er (op grond van de weg- en verkeerskenmerken) verwacht mogen worden. In bijlage 12 zijn de locaties met de hoogste en laagste scores op de genoemde canonische as te vinden. Verder zijn afgedrukt de locatiecode, de canonische scores op de tweede canonische as van de eerste dimensie (de gemeten omvang van de veiligheid in plaats van de verwachte omvang), het totale aantal ongevallen, het aantal dodelijke ongevallen en de categoriescores voor de kenmerken die met deze dimensie samenhangen in volgorde van belangrijkheid, t.w. kenmerk 12, 15, 19 en 21. Op grond van de "bootstrap"-analyse zouden we de kenmerken 5 en 14 hieraan kunnen toevoegen. Dit is echter niet gebeurd. Ook zouden naast de categoriescores de categorieën zelf kunnen worden afgedrukt. Deze kunnen echter worden teruggevonden met behulp van bijlage 10. Hierin staan de schalingen voor de

relevante kenmerken van de eerste en tweede dimensie afgebeeld, alsmede die van de beide ongevalskenmerken.

In het algemeen zien we dat de canonische scores op de tweede as van de eerste dimensie sterk samenhangen met het totale aantal ongevallen. De correlatie is .996. De canonische scores op de eerste as hangen samen met die op de tweede as; de canonische correlatie, die deze samenhang aangeeft, is .412. Omdat de canonische scores op de tweede as zo sterk samenhangen met het totale aantal ongevallen, geldt dat de correlatie tussen het totale aantal ongevallen en de canonische scores op de eerste as bijna even hoog is als de canonische correlatie. De aantallen ongevallen van de locaties met hoge scores op de eerste canonische as zijn gemiddeld groter dan die van de locaties met een gemiddelde score en deze zijn weer groter dan die van de locaties met een lage score.

9.5. De verdere aanpak

In dit voorbeeld beschikken we nu over:

- de mogelijke ongevalsoorzaken in termen van weg- en verkeerskenmerken;
 - de onveilige categorieën van die kenmerken;
 - een lijst van locaties met de hoogste (en de laagste) potentiële ongevalskans;
 - informatie over de kenmerken die op deze locaties aanwezig zijn.
- Met behulp van deze informatie kan de in paragraaf 8.4. beschreven werkwijze gevolgd worden om tot maatregelen te komen.

9.6. Effectiviteit van de maatregelen

Als voorbeeld voor het schatten van het effect van de maatregel op de veiligheid kiezen we de eerste locatie m.b.t. de eerste oplossingsdimensie, zoals gegeven in bijlage 12.

Deze locatie, met code 10210716, heeft, zoals uit de score blijkt, twee kruispunten van het type B. Overwogen wordt een oplossing door middel van een onderdoorgang (tunnel). Zoals kan worden af-

gelezen uit de schalingsgrafiek van bijlage 10, blad 1, zou daarmee de score op dit kenmerk worden teruggebracht van 3.486 tot -.280. Willen we hiermee berekenen, wat de daling van de score op de eerste canonische as zou zijn, dan hebben we behalve de schaalwaarden van de categorieën ook het gewicht nodig van kenmerk 21 (kruispunten type B). Dit gewicht bedraagt .125. De gewichten zijn overigens niet vermeld in bijlage 6 of 7, omdat de correlaties voor een interpretatie belangrijker zijn. Voor de score op de eerste canonische as zou dit betekenen dat deze van de waarde 3.699 daalt met $\{3.486 - (-.280)\} \times .125 = .471$ tot 3.228. Stel dat vervolgens het aantal lichtmasten kan worden gehalveerd, zonder dat de verlichtingskwaliteit wordt aangetast, dan vindt een verdere reductie plaats met $(11.180 - .517) \times .229 = 2.442$ tot .786, een waarde die de locatie ergens in het midden van de lijst doet belanden. Voor een schatting van het effect op het totale aantal ongevallen in absolute zin zijn nog wat extra assumpties nodig over de schaling van kenmerk 27. Gaan we bijv. uit van een logaritmische schaling, dan kunnen we de grafiek van kenmerk 27 (bijlage 10, blad 3) vervangen door de best passende logaritmische curve. Bij de waarden 3.699 en .786 van de y-as vinden we voor het verwachte aantal ongevallen op de x-as waarden van ca. 8 en ca. 2. Dit komt dus neer op een reductie van ca. 6 verwachte ongevallen. Zoals al eerder is opgemerkt, moet hierbij wel rekening worden gehouden met de lage voorspelbaarheid van de oplossing voor individuele locaties.

10. SLOTOPMERKINGEN

In dit rapport is de theorie beschreven van de analysetechnieken die bij het relatie-onderzoek worden gebruikt. Aan de hand van een theoretisch en een praktisch voorbeeld is dit nader toegelicht. Ook is ingegaan op de mogelijkheden tot nadere interpretatie en het vaststellen van de betrouwbaarheid.

Verder is ingegaan op de mogelijkheden om tot maatregelen te komen. Met name dit laatste aspect is voor de wegbeheerder van groot belang. Ook al is de techniek gecompliceerd en de interpretatie niet eenvoudig, de uiteindelijke resultaten moeten ingepast kunnen worden in een verkeersveiligheidsbeleid.

De analysemethode levert niet alleen inzicht in de globale relaties tussen de vele ongevalskenmerken en weg- en verkeerskenmerken, zij levert ook hulpmiddelen om te bepalen waar welke maatregelen genomen kunnen worden.

De schaling van de kenmerken geeft aan, welke klassen van de kenmerken onveilig zijn. De lijsten met locaties, gerangschikt naar de hoogte van canonische scores, levert daarbij een overzicht van de locaties die voor maatregelen in aanmerking komen.

Een belangrijk voordeel van deze methode ten opzichte van de meer gebruikelijke black-spotbenadering is dat de onveiligheid van de locatie niet bepaald hoeft te worden op basis van de ongevallen die daar zijn gebeurd. In plaats daarvan kan met behulp van de ongevallen op alle locaties worden vastgesteld, welke locaties een zodanige constellatie van weg- en verkeerskenmerken hebben dat de kans op een ongeval daar groot is. Hierdoor is het mogelijk maatregelen te nemen zonder dat men verzeild raakt in een "als het kalf verdronken is dempt men de put"-benadering. Dit laatste is in feite het geval als men locaties waar veel ongevallen gebeurd zijn aanpakt in de verwachting dat daar in de toekomst ook wel veel ongevallen zullen gebeuren.

Er mag een grotere effectiviteit van de maatregelen worden verwacht, als men uitgaat van een verklaringsmodel dat op basis van alle locaties te zamen verkregen is. Een voordeel daarbij is dat deze verklaring richting geeft aan het zoeken naar oplossingen.

Verder kunnen op basis van de analyse concrete verwachtingen over de effectiviteit van maatregelen worden opgesteld, zodat de toepassing van de methode in de praktijk kan worden geëvalueerd.

LITERATUUR

Geer, J.P. van de (1967). Inleiding in de multivariate analyse. Van Loghem Slaterus, Arnhem, 1967.

Gifi, A. (1981). Non-linear multivariate analysis. Afdeling Data-theorie van de Faculteit der Sociale Wetenschappen. Rijksuniversiteit Leiden, 1981.

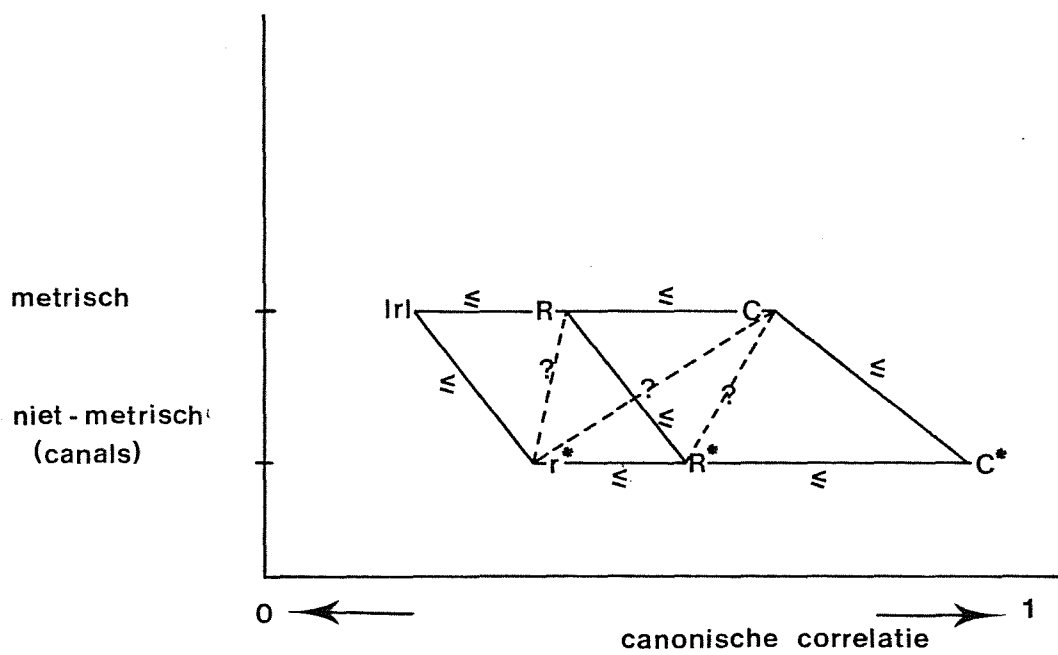
Oppe, S. (1980). Analyse van samenhangen tussen kwalitatieve verkeersveiligheidskenmerken. Verkeerskunde 31 (1980): 364-368.

SWOV (1976a). De verkeersonveiligheid in de provincie Noord-Brabant I en II. Onderzoek Noord-Brabant fase 1a. Een globale vergelijking van de onveiligheid van Noord-Brabant met die van de andere provincies en van geheel Nederland. Onderzoek Noord-Brabant fase 1b. Een beschrijvend onderzoek naar de relatieve onveiligheid in Noord-Brabant in vergelijking met de Rest van Nederland. R-76-5. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV, Voorburg, 1976.

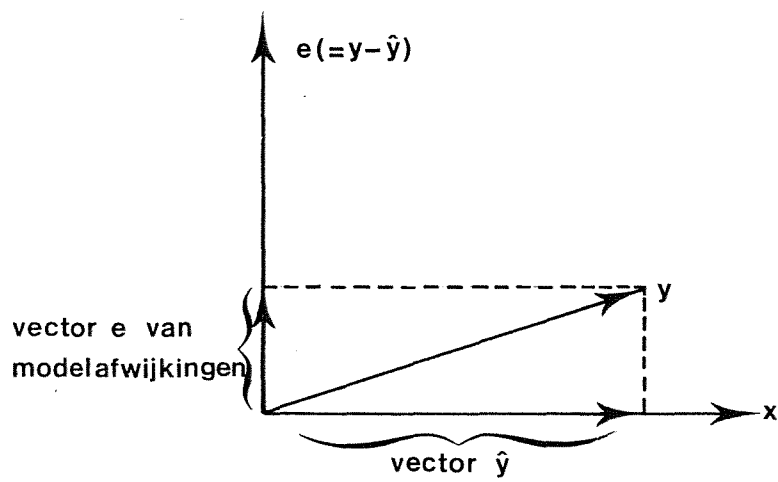
SWOV (1976b). De verkeersonveiligheid in de provincie Noord-Brabant III. Onderzoeksopzet voor het onderzoek Noord-Brabant fase 2. R-76-20. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV, Voorburg, 1976.

SWOV (1980). De verkeersonveiligheid in de provincie Noord-Brabant IX A. Inventarisatie van ongevalgegevens, verkeerskenmerken en wegkenmerken ten behoeve van een onderzoek naar de relatie tussen deze kenmerken. R-80-28 I t/m III. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV, Voorburg, 1980.

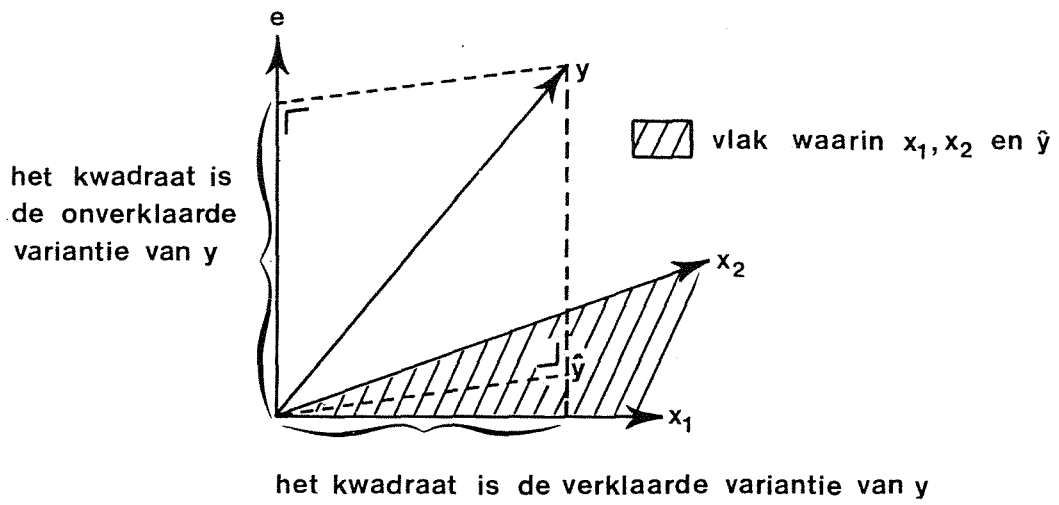
SWOV (1981). De verkeersonveiligheid in de provincie Noord-Brabant IX C. Het relatie-onderzoek: resultaten analyse raaien. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV, Leidschendam, 1981.



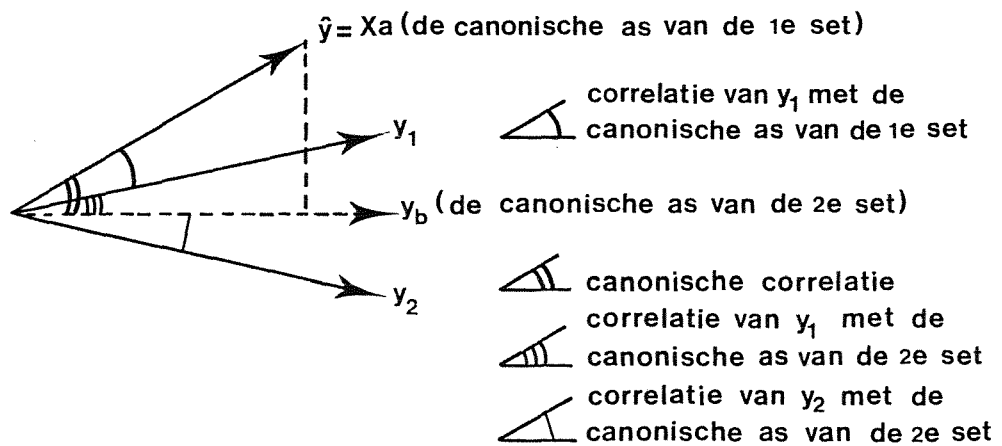
Afbeelding 1



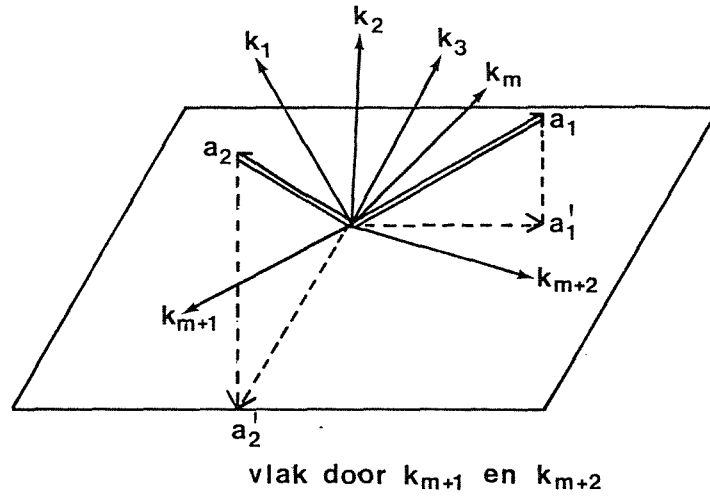
Afbeelding 2



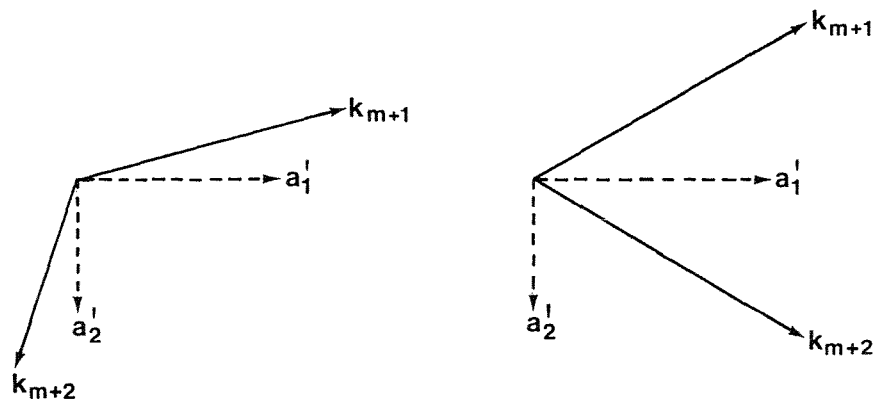
Afbeelding 3



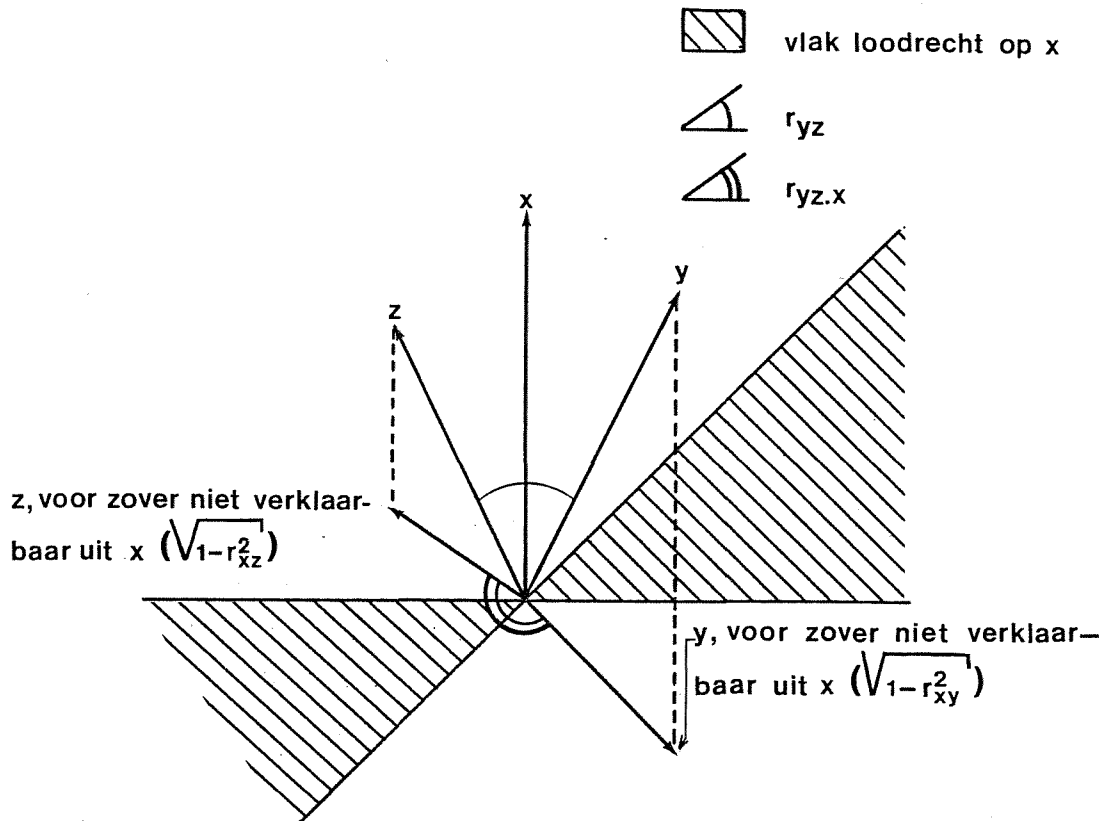
Afbeelding 4



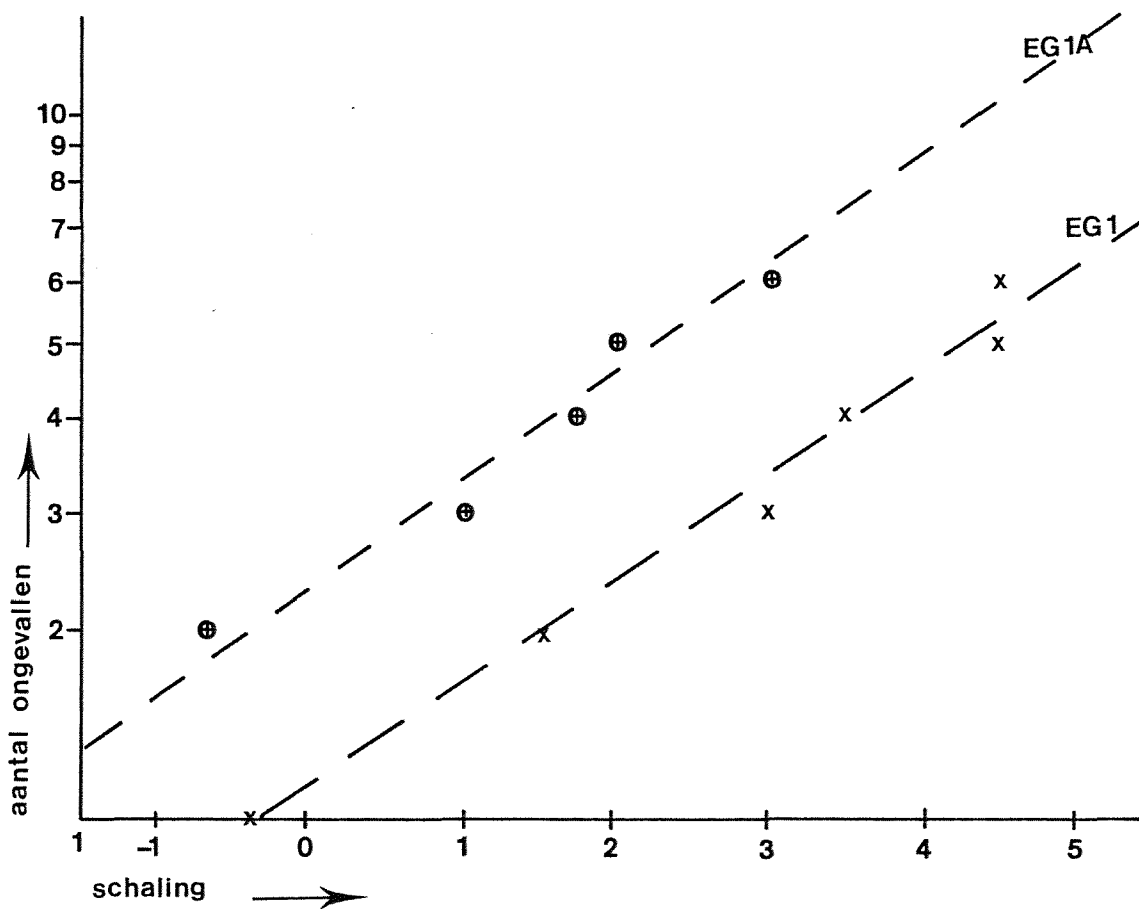
Afbeelding 5



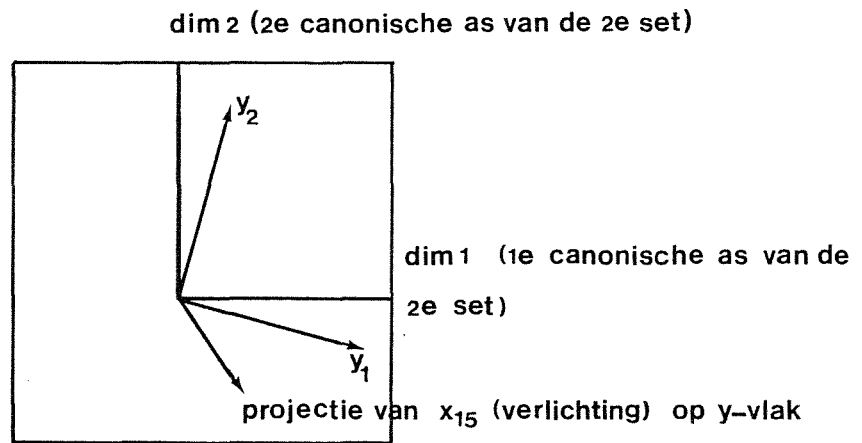
Afbeelding 6 a + b



Afbeelding 7



Afbeelding 8



Afbeelding 9

LITERATUUR SWOV-ONDERZOEK 974.023 en 124.1.

DE VERKEERSONVEILIGHEID IN DE PROVINCIE NOORD-BRABANT

A. Rapporten verschenen t/m 31-12-1981

De verkeersonveiligheid in de provincie Noord-Brabant I en II. Onderzoek Noord-Brabant fase 1a: Een globale vergelijking van de onveiligheid van Noord-Brabant met die van de andere provincies en van geheel Nederland; Onderzoek Noord-Brabant fase 1b: Een beschrijvend onderzoek naar de relatieve onveiligheid in Noord-Brabant in vergelijking met de Rest van Nederland + Tabellen, Afbeeldingen en Bijlagen. R-76-5 I + II. SWOV, mei 1976. 96 + 238 blz.

De verkeersonveiligheid in de provincie Noord-Brabant III. Onderzoeksofzet voor het onderzoek Noord-Brabant fase 2. R-76-20. SWOV, november 1976. 85 blz.

De verkeersonveiligheid in de provincie Noord-Brabant IV. Het aspect stroefheid in het verkeersonveiligheidsonderzoek in Noord-Brabant. R-78-17. SWOV, april 1978. 53 blz.

De verkeersonveiligheid in de provincie Noord-Brabant V. Onderzoek met betrekking tot enkelvoudige ongevallen in Noord-Brabant. R-79-36. SWOV, april 1979. 118 blz.

De verkeersonveiligheid in de provincie Noord-Brabant VI. Onderzoekverslag van het onderzoek Noord-Brabant fase 2, stap 1: Aanvullende ongevalsanalyses. R-79-37. SWOV, december 1979. 75 blz.

De verkeersonveiligheid in de provincie Noord-Brabant VIIA. Onderzoekverslag van het onderzoek Noord-Brabant fase 2, stap 2A: Vergelijking van weg- en verkeerskenmerken van wegvakken voor gemengd verkeer buiten de bebouwde kom in Noord-Brabant en de Rest van Nederland. R-79-44. SWOV, november 1979. 51 blz.

De verkeersonveiligheid in de provincie Noord-Brabant VIIB. Onderzoeksverslag van het onderzoek Noord-Brabant fase 2, stap 2B: Vergelijking van de verkeersprestatie op kruispunten in Rijkswegen buiten de bebouwde kom in Noord-Brabant en de Rest van Nederland. R-80-53. SWOV, november 1980. 38 blz.

De verkeersonveiligheid in de provincie Noord-Brabant VIII. Verslag van het vóóronderzoek voor een vergelijkend risico-onderzoek in Noord-Brabant. R-80-15. SWOV, maart 1980. 59 blz.

De verkeersonveiligheid in de provincie Noord-Brabant IXA. Inventarisatie van ongevallengegevens, verkeerskenmerken en wegkenmerken ten behoeve van een onderzoek naar de relatie tussen deze kenmerken. R-80-28 I t/m III. SWOV, mei 1980. 38 + 87 + 108 blz.

De verkeersonveiligheid in de provincie Noord-Brabant IXB. Het relatie-onderzoek: Onderzoekopzet en methode van onderzoek. R-81-39. SWOV, november 1981. 84 blz.

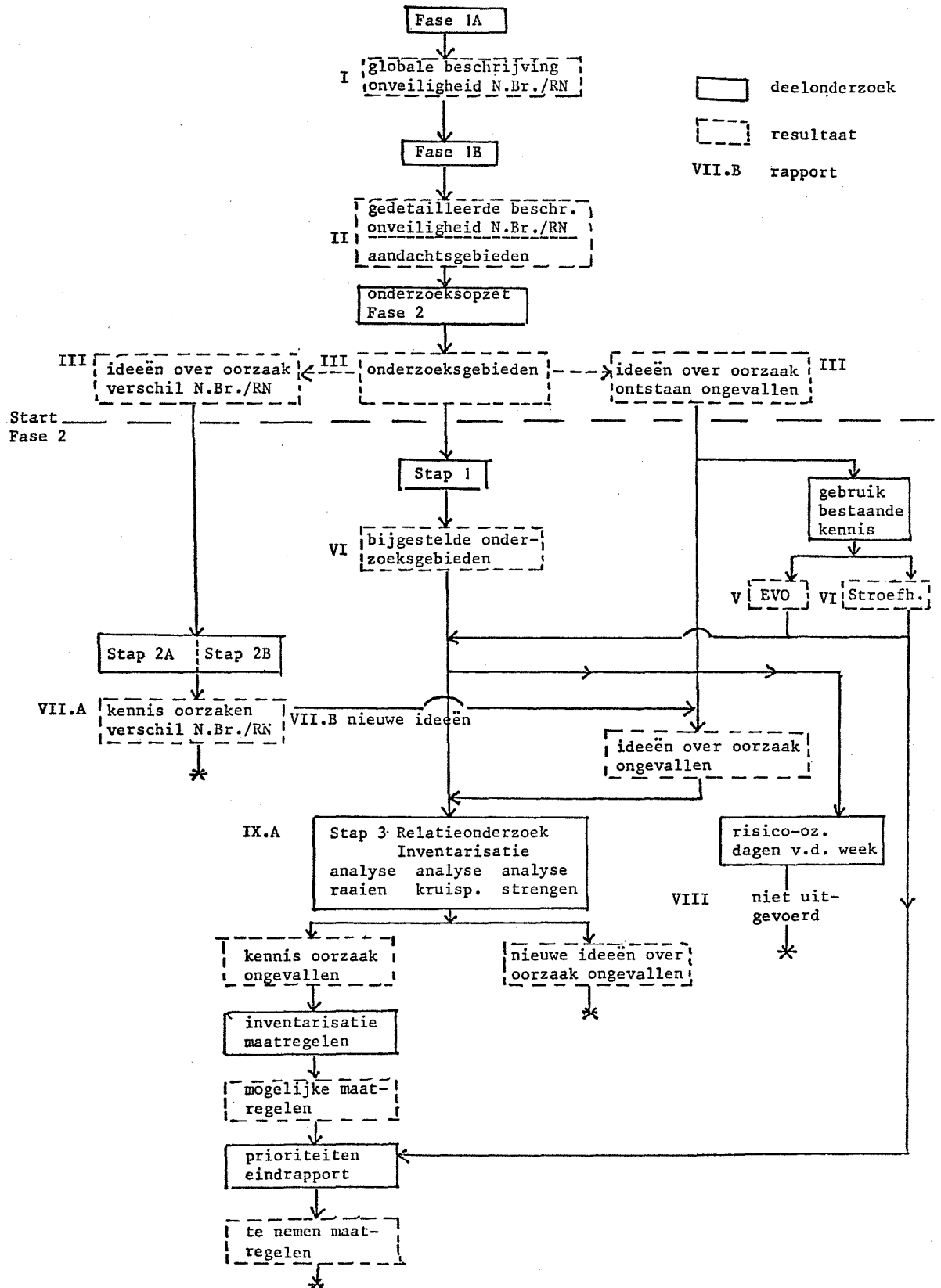
De verkeersonveiligheid in de provincie Noord-Brabant IXC. Het relatie-onderzoek: Resultaten deelonderzoek Analyse raaien + Tabellen en Figuren + Bijlagen. R-81-40 I t/m III. SWOV, november 1981. 84 + 90 + 107 blz.

B. Overzicht verschenen en toekomstige rapporten en onderzoekverslagen met hun "roepnaam"

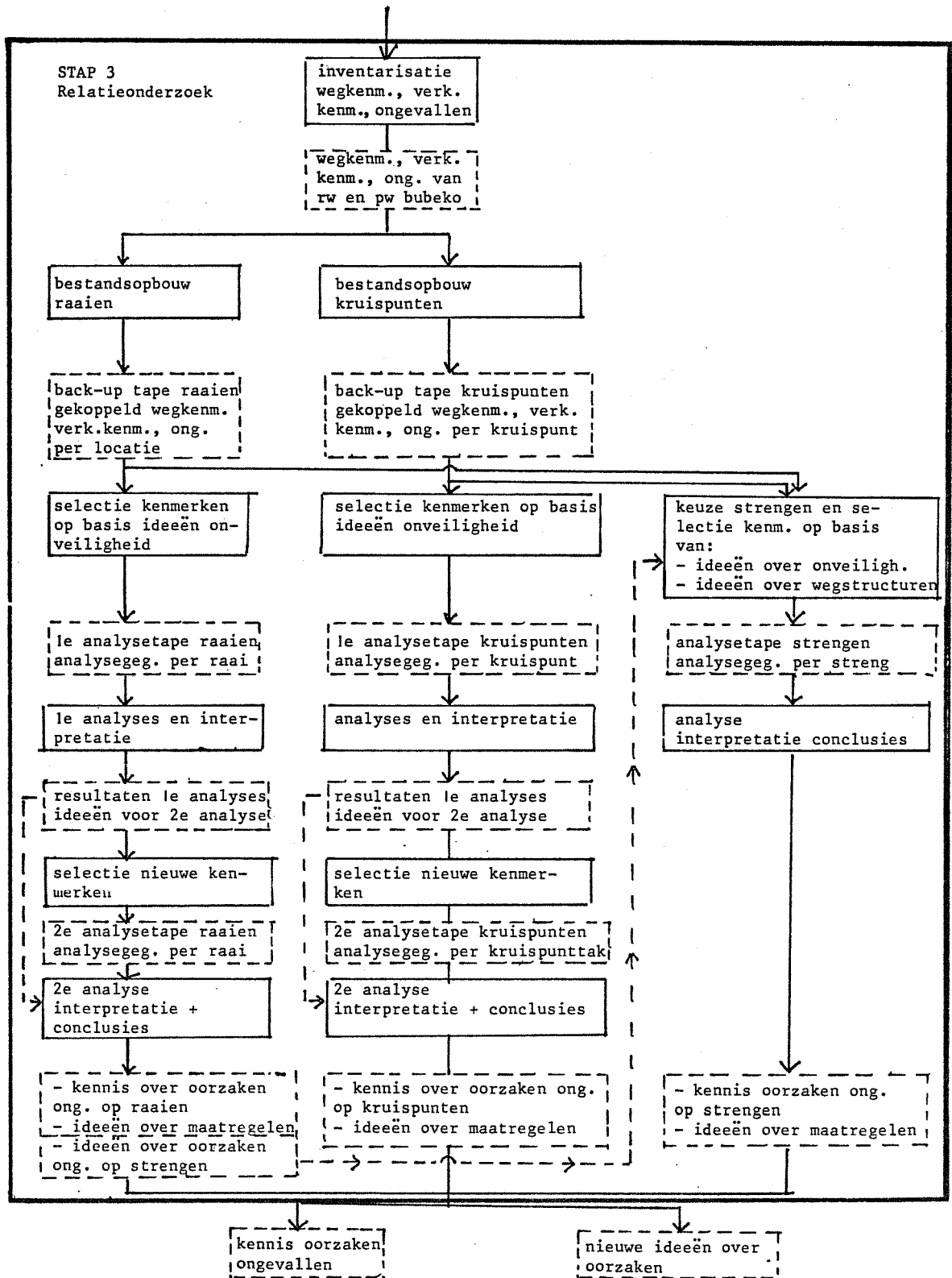
Rapport I.	Noord-Brabant fase 1a
Rapport II.	Noord-Brabant fase 1b
Rapport III.	Onderzoeksopzet Noord-Brabant fase 2
Rapport IV.	Noord-Brabant fase 2 Stroefheid in Noord-Brabant
Rapport V.	124.1.10 Noord-Brabant fase 2, Enkelvoudige ongevallen
Onderzoeksverslag VI.	124.1.30 Noord-Brabant fase 2, Stap 1
Onderzoeksverslag VIIA.	124.1.40 Noord-Brabant fase 2, stap 2A
Onderzoeksverslag VIIB.	124.1.45 Noord-Brabant fase 2, stap 2B
Rapport VIII.	124.1.20 Dagen van de week
Rapport IXA.	Noord-Brabant fase 2, 124.1.51 Inventarisatieverslag
Rapport IXB.	Noord-Brabant fase 2, Onderzoekopzet en -methode
Rapport IXC.	Noord-Brabant fase 2, 124.1.52 Analyse raaien
Rapport IXD.	Noord-Brabant fase 2, 124.1.53 Analyse kruispunten
Rapport IXE.	Noord-Brabant fase 2, 124.1.54 Analyse strengen
Rapport X.	Eindrapport fase 2
(Rapport IX. e.v.	Eventuele evaluatie fase 3)

Het eventuele geïntegreerd verslag is een soort brochure op basis van VI en VIIA. en VIIB. en behoeft geen apart nummer. Over maatregelen wordt geen afzonderlijk rapport uitgebracht.

STRUCTUUR VERKEERSVEILIGHEIDSONDERZOEK NOORD-BRABANT



OPBOUW VAN HET RELATIE-ONDERZOEK



JOB TITLE

DEMONSTRATIE VAN CANALS M.B.V. EEN TECHNISCH VOORBEELD.

DATA SPECIFICATIONS:

NUMBER OF OBJECTS 10
 NUMBER OF VARIABLES IN THE FIRST SET 3
 NUMBER OF VARIABLES IN THE SECOND SET 1
 MAXIMUM NUMBER OF CATEGORIES FOR ONE VARIABLE 3

ANALYSIS SPECIFICATIONS:

NUMBER OF DIMENSIONS 1
 MAXIMUM NUMBER OF ITERATIONS 20
 CRITERIUM STRESS DIFFERENCE 0.001000000
 CRITERIUM WEIGHT INCREASE 0.000100000

I/O SPECIFICATIONS:

INPUT FROM LOGICAL UNIT 5
 OUTPUT OF RAW DATA (1=ALL OBJECTS, 0=10 OBJECTS) 1
 OUTPUT OF OPTIMALLY SCALED VARIABLES TO UNIT 6
 OUTPUT OF STRESS FOR EVERY ITERATION (Y=1,N=0) 1
 OUTPUT OPTIMAL SCORES, WEIGHTS, CORRELATIONS TO UNIT 1
 OUTPUT CANONICAL SCORES (0=NO OUTPUT,
 1=PRINT&PLOT OUTPUT, GT 1=PRINT&PLOT&OUTPUT TO UNIT
 -1=PRINT OUTPUT ONLY, LT -1=OUTPUT TO -UNIT ONLY) 1
 PLOT OPTIMAL SCORES (0=NO PLOTS, 1=ONE OR MORE PLOTS) 0

REQUIRED SPACE FOR ARRAYS AND PROGRAM IS : 71K BYTES (18101 HALF WORDS)
 RAW DATA , ROWS=OBJECTS , COLUMNS=VARIABLES 10 OBJECTS ARE PRINTED

1	1	2	3	1
2	2	2	2	1
3	1	2	2	1
4	1	1	3	2
5	2	1	2	2
6	2	2	1	2
7	2	1	2	3
8	1	2	1	3
9	2	2	2	3
10	2	1	1	3

VARIABLE NUMBER MARGINAL FREQUENCIES OF THE SUCCEEDING CATEGORIES

1	4	6	
2	4	6	
3	3	5	2
4	3	3	4

VARIABLE NUMBER	MAXIMUM CATEGORY	MEASUREMENT LEVEL	NUMBER OF MISSING ENTRIES	PLOT OPTIMAL SCORES
1	2	1 = ORDINAL	0	0
2	2	1 = ORDINAL	0	0
3	3	1 = ORDINAL	0	0
4	3	0 = NOMINAL	0	0

FORMAT

(4I4)

ITERATION	STRESS	AFTER CHANGING	NO ITERATIONS WEIGHTS
1	0.5447034937	WEIGHTS B	1
	0.5421263596	WEIGHTS A	5
2	0.4871308289	WEIGHTS B	2
	0.4654747197	WEIGHTS A	5
3	0.4609279107	WEIGHTS B	2
	0.4564941602	WEIGHTS A	5
4	0.4558792311	WEIGHTS B	2

THE FIRST STRESS IS THE STRESS OF A COMPLETE NUMERICAL SOLUTION OF THE DATA
 (WITH A FIXED MISSING CATEGORY FOR EVERY VARIABLE)
 FIRST SET:

ORIGINAL SCORES, MARGINAL FREQUENCIES AND OPTIMAL SCORES OF VARIABLE NO	1	(ORDINAL)
1	4	-1.225
2	6	0.816

ORIGINAL SCORES, MARGINAL FREQUENCIES AND OPTIMAL SCORES OF VARIABLE NO 2 (ORDINAL)

1	4	-1.225
2	6	0.816

ORIGINAL SCORES, MARGINAL FREQUENCIES AND OPTIMAL SCORES OF VARIABLE NO 3 (ORDINAL)

1	3	-1.514
2	5	0.547
3	2	0.903

SECOND SET:

ORIGINAL SCORES, MARGINAL FREQUENCIES AND OPTIMAL SCORES OF VARIABLE NO 4 (NOMINAL)

1	3	-1.526
2	3	0.589
3	4	0.703

THE OPTIMALLY SCALED VARIABLES; ROWS ARE OBJECTS, COLUMNS ARE VARIABLES

-1.225	0.816	0.903	-1.526
0.816	0.816	0.547	-1.526
-1.225	0.816	0.547	-1.526
-1.225	-1.225	0.903	0.589
0.816	-1.225	0.547	0.589
0.816	0.816	-1.514	0.589
0.816	-1.225	0.547	0.703
-1.225	0.816	-1.514	0.703
0.816	0.816	0.547	0.703
0.816	-1.225	-1.514	0.703

CANONICAL CORRELATIONS FOR EACH DIMENSION

0.738

VARIABLE WEIGHTS OF THE FIRST SET FOR EACH DIMENSION

1	-0.196
2	0.725
3	0.643

VARIABLE WEIGHTS OF THE SECOND SET FOR EACH DIMENSION

4 -1.000

CORRELATIONS BETWEEN THE OPTIMALLY SCALED VARIABLES OF THE FIRST SET AND THE CANONICAL AXES OF THE FIRST SET FOR EACH DIMENSION

1	-0.488
2	0.711
3	0.605

CORRELATIONS BETWEEN THE OPTIMALLY SCALED VARIABLES OF THE SECOND SET AND THE CANONICAL AXES OF THE FIRST SET FOR EACH DIMENSION

4 -0.738

CORRELATIONS BETWEEN THE OPTIMALLY SCALED VARIABLES OF THE FIRST SET AND THE CANONICAL AXES OF THE SECOND SET FOR EACH DIMENSION

1	-0.359
2	0.527
3	0.444

CORRELATIONS BETWEEN THE OPTIMALLY SCALED VARIABLES OF THE SECOND SET AND THE CANONICAL AXES OF THE SECOND SET FOR EACH DIMENSION

4 -1.000

CANONICAL SCORES FOR EACH DIMENSION OF THE FIRST RESPECTIVELY THE SECOND SET

1	1.413	1.526
2	0.784	1.526
3	1.184	1.526
4	-0.067	-0.589
5	-0.696	-0.589
6	-0.542	-0.589
7	-0.696	-0.703
8	-0.141	-0.703
9	0.784	-0.703
10	-2.021	-0.703

THE OPTIMALLY SCALED VARIABLES HAVE BEEN WRITTEN TO UNIT 6 WITH FORMAT (10F8.3). THE ROWS ARE OBJECTS AND THE COLUMNS ARE VARIABLES.

Belangrijkste resultaten van de analyse EG 1AD

BESTAND ANALYSE DIMENSIE CORRELATIE	ONG LOC ENKELB. WEGEN VOOR GEMENGD VERK.			
	I		EG 1AD	
		0.412	II	SCHALING
		0.266		
1 verhardingsbreedte	3	-0.125		(-0.030) 1→2→Rest
2 vrijliggende fietspaden				
3 parallelweg				
4 aanliggende fietspaden				
5 zicht				
6 obstakelafstand				
7 bomen				
8 bossages				
9 lichtmasten	4	0.121		(-0.046) →
10 vangrail				
11 bebouwing				
12 bocht	1	0.251		(0.023) 1,6→5,4,3,2
13 vert. helling				
14 verharding		(0.022)	4	-0.061 3→4→2→1
15 verlichting	2	0.159	3	-0.083 1→Rest
16 reflectoren				
17 voorrang				
18 max. snelheid				
19 inhaalverbod				
20 parkeerverbod		(-0.059)	5	0.063 3→2→1
21 kruisp. type B	5	0.102		1,2→3→4
22 erfaansl.				
23 I				
24 WRMO		(-0.040)	1	0.096 →
24 P				
25 VMO		(0.025)	2	0.088 →
25 I				
26 WRBO				
26 P				
26 VBO				
tot. aant. ong.	1	0.996	2	-0.085 →
dod. ong.	2	0.259	1	0.966 1→2→3

Correlations between the observation scores of the first set and the canonical scores of the second set for each dimension.

	DIM 1	DIM 2
1	-0.125	-0.030
2	-0.043	-0.013
3	0.010	-0.011
4	0.064	0.018
5	-0.035	0.035
6	-0.058	0.010
7	0.020	0.035
8	-0.062	0.047
9	0.121	-0.046
10	-0.065	-0.001
11	-0.037	-0.018
12	0.251	0.023
13	-0.017	-0.022
14	0.022	-0.061
15	0.159	-0.083
16	-0.080	-0.041
17	0.068	-0.008
18	-0.008	-0.046
19	-0.043	-0.019
20	-0.059	0.063
21	0.102	-0.047
22	-0.037	-0.044
23	0.024	-0.019
24	-0.040	0.096
25	0.025	0.088
26	-0.059	0.007

correlations between the observation scores of the second set and the canonical scores of the second set for each dimension

27	0.996	-0.085
28	0.259	0.966

Correlaties van de 26+2 kenmerken met de tweede canonische as voor dimensie 1 en 2 van de CANALS-analyse van het totale aantal ongevallen en het aantal dodelijke ongevallen op de ongevallenlocaties van wegen met gemengd verkeer.

schaling van de
kenmerken

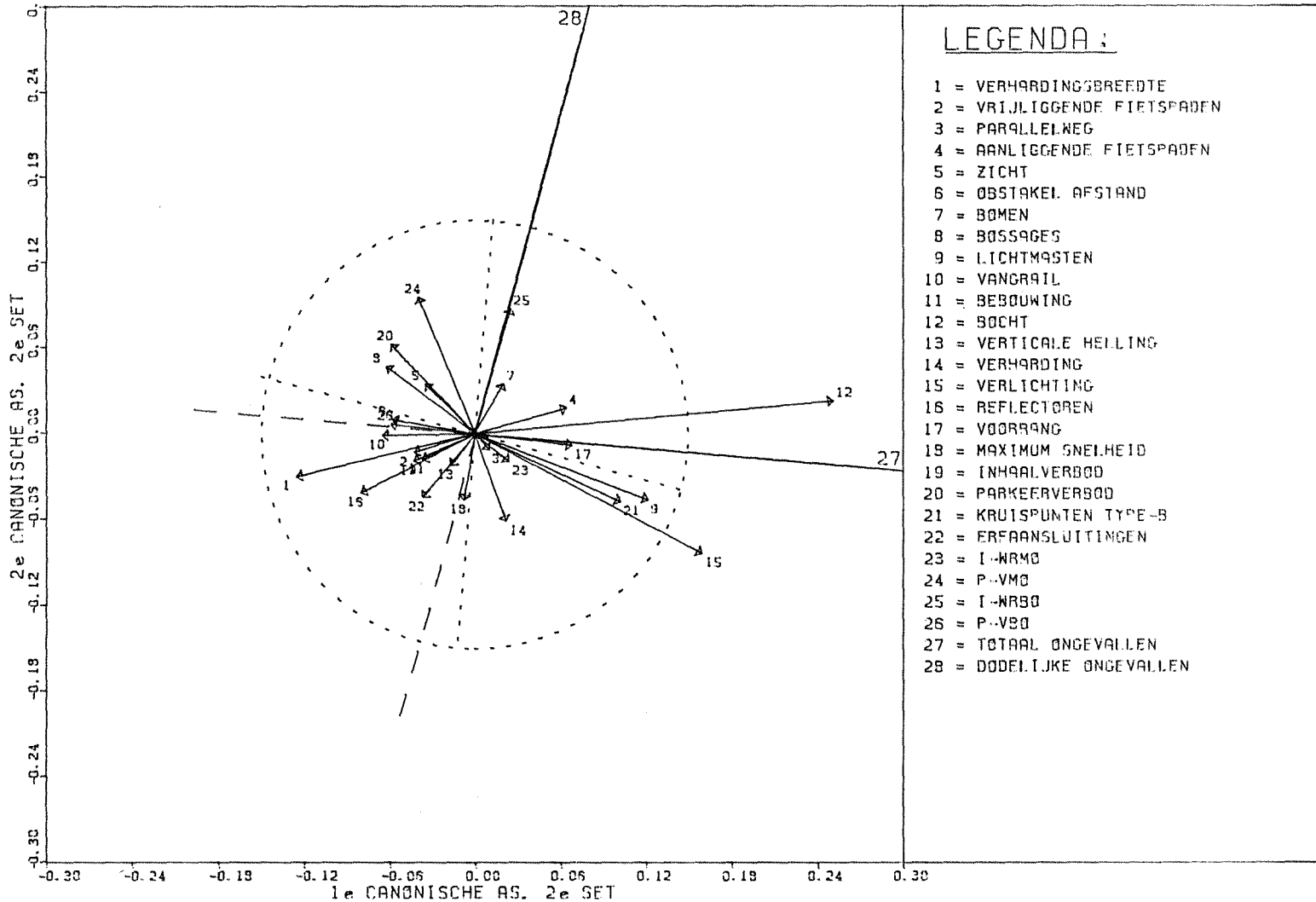
FIRST SET:

ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	1	(ORDINAL)
1	-0.806	
2	0.146	
3	1.563	
4	1.563	
5	1.563	
6	1.563	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	2	(ORDINAL)
1	-0.206	
2	4.844	
3	4.844	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	3	(ORDINAL)
1	-0.134	
2	7.489	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	4	(ORDINAL)
1	-0.274	
2	3.225	
3	3.721	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	5	(ORDINAL)
1	-0.351	
2	-0.351	
3	-0.351	
4	-0.351	
5	-0.351	
6	2.852	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	6	(ORDINAL)
1	-1.260	
2	0.564	
3	1.203	
4	1.203	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	7	(ORDINAL)
1	-1.254	
2	-0.379	
3	-0.379	
4	-0.379	
5	1.063	
6	1.063	
7	1.063	
8	3.428	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	8	(ORDINAL)
1	-0.194	
2	-0.194	
3	3.674	
4	7.782	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	9	(ORDINAL)
1	-0.223	
2	0.517	
3	0.517	
4	5.336	
5	11.180	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	10	(ORDINAL)
1	-0.086	
2	-0.086	
3	11.662	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	11	(ORDINAL)
1	-0.564	
2	0.803	
3	0.803	
4	2.864	
5	5.580	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	12	(ORDINAL)
1	-0.547	
2	-0.547	
3	1.005	
4	1.929	
5	1.929	
6	1.929	

schaling van de
kenmerken

ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	13	(ORDINAL)
1	-0.191	
2	5.248	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	14	(NOMINAL)
1	-2.055	
2	-0.056	
3	2.183	
4	-0.453	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	15	(ORDINAL)
1	-0.942	
2	1.062	
3	1.062	
4	1.062	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	16	(ORDINAL)
1	-2.895	
2	0.345	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	17	(ORDINAL)
1	-0.416	
2	-0.416	
3	2.405	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	18	(ORDINAL)
1	-5.365	
2	-5.365	
3	0.186	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	19	(ORDINAL)
1	-0.178	
2	5.623	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	20	(NOMINAL)
1	2.699	
2	-0.346	
3	-1.984	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	21	(ORDINAL)
1	-0.280	
2	-0.280	
3	3.486	
4	4.203	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	22	(ORDINAL)
1	-0.507	
2	-0.507	
3	-0.507	
4	-0.507	
5	-0.507	
6	1.973	
7	1.973	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	23	(ORDINAL)
1	-4.984	
2	-4.481	
3	-0.708	
4	0.478	
5	0.875	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	24	(ORDINAL)
1	-5.556	
2	-0.416	
3	0.128	
4	2.693	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	25	(ORDINAL)
1	-2.129	
2	-1.134	
3	-1.134	
4	0.816	
5	1.314	
ORIGINAL AND OPTIMAL SCORES OF VARIABLE NO	26	(ORDINAL)
1	-7.402	
2	-0.430	
3	-0.430	
4	0.351	
5	0.846	

EG 1AD - TOTAAL VERSUS DODELIJK - ONGEVALSLOCATIES



Kenmerk	Oorspr.	P15	P23 t/m 26	P15, 23 t/m 26
1 Verhardingsbreedte	-0.116	-0.118	-0.120	-0.123
2 Vrijliggende fietspaden	-0.043	-0.039	-0.039	-0.034
3 Parallelweg	0.009	0.012	0.024	0.028
4 Aanliggende fietspaden	0.054	0.053	0.075	0.073
5 Zicht	-0.132	-0.118	-0.126	-0.109
6 Obstakelafstand	-0.064	-0.058	-0.123	-0.117
7 Bomen	0.024	0.035	0.007	0.020
8 Bossages	-0.075	-0.071	-0.068	-0.064
9 Lichtmasten	0.133	0.118	0.162	0.156
10 Vangrail	-0.065	-0.068	-0.059	-0.062
11 Bebouwing	0.028	0.011	-0.031	-0.056
12 Bocht	0.256	0.248	0.261	0.253
13 Verticale helling	-0.013	-0.012	0.009	0.010
14 Verharding	-0.072	-0.079	-0.105	-0.114
15 Verlichting	0.066	-	0.079	-
16 Reflectoren	-0.077	-0.054	-0.019	0.019
17 Voorrang	0.065	0.060	0.071	0.065
18 Max. snelheid	0.010	0.012	0.003	0.005
19 Inhaalverbod	-0.047	-0.052	-0.038	-0.043
20 Parkeerverbod	0.065	0.069	0.037	0.040
21 Kruisp. type B	0.100	0.092	0.100	0.090
22 Erfaansluitingen	-0.037	-0.057	-0.082	-0.110
23 I _{WRMO}	0.021	0.015	-	-
24 P _{VMO}	-0.058	-0.059	-	-
25 I _{WRBO}	-0.072	-0.076	-	-
26 P _{VBO}	-0.063	-0.063	-	-
27 Totaal aantal ong.	1.000	1.000	1.000	1.000

Het uitpartialiseren van de intensiteit (23 t/m 26) heeft tot gevolg:

kenmerk	oorspronkelijk	na uitpartialisering
- obstakelafst.	-.064	→ -.123
- lichtmasten	.133	→ .162
- bebouwing	.028	→ -.031
- verharding	-.072	→ -.105
- reflectoren	-.077	→ -.019
- erfaansl.	-.037	→ -.082

Het uitpartialiseren van verlichting (15):

- zicht	-.132	→ -.118
- lichtmasten	.133	→ .118
- reflectoren	-.077	→ -.054
- erfaansl.	-.037	→ -.057

Het uitpartialiseren van beide (15 en 23 t/m 26):

- zicht	-.132	→ -.109
- obstakelafst.	-.064	→ -.117
- erfaansl.	-.037	→ -.110
- lichtmasten	.133	→ .156

Bijlage 9.1. Berekening van partiële correlaties uit een analyse over het totaal aantal ongevallen op wegen voor gemengd verkeer (analyse EG1 extra).

Kenmerken van de oplossing:	can. corr.
alle kenmerken	.411
1,5,6,8,9,12,14,16,20,21,23,24,25,26	.373 \times)
1,5,6,8,9,12,14,16,20,21	.354 +)
23,24,25,26	.117
23,...,26 1,...,21	.117 x)
1,...,21 23,...,26	.354
5,6,8,9,12,21	.298
1,14,16,20	.185
5,6,...,21 1,...,20	.284
1,...,20 5,...,21	.193
23,25	.113
24,26	.063

Conclusies:

1) 23,...,26 verklaren minder maar wel iets anders dan 1,...,21:

$\sqrt{.354^2 + .117^2} = .373$, ofwel de projectie op deelruimte A staat loodrecht op de projectie op deelruimte B.

2) 23,25 verklaren ongeveer hetzelfde als 23,24,25,26 (dus 24 en 26 niet nodig): cc = .113 i.p.v. .117.

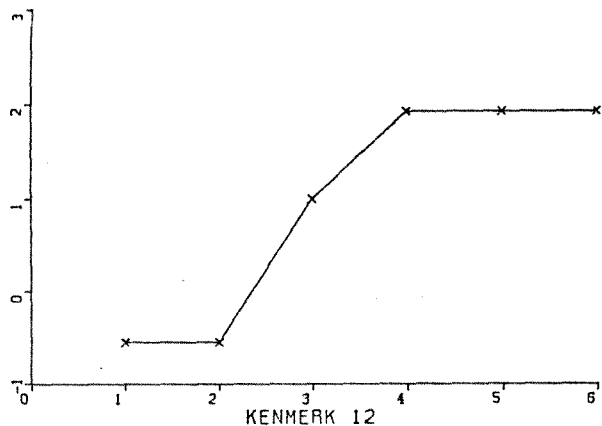
3) 5,6,8,9,12,21 verklaren meer dan 1,14,16,20 en voornamelijk iets anders.

\times) Toevoegen van 15 geeft .380

+) Toevoegen van 15 geeft .361

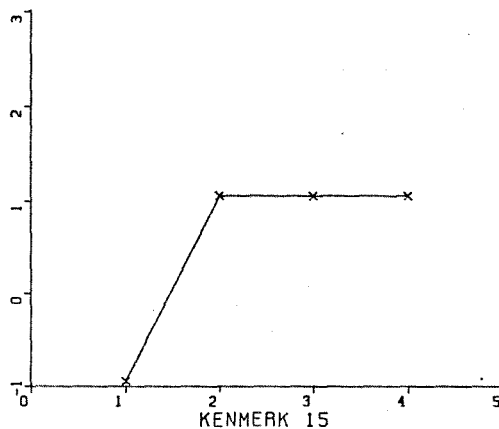
x) B|A moet gelezen worden als: "dat wat B aanvullend verklaart als B wordt toegevoegd aan A". Dit is natuurlijk maximaal gelijk aan dat wat B alleen verklaart. Bij gelijkheid geldt dat de verklaring door B onafhankelijk is van die door A.

Bijlage 9.2. Berekening van stapsgewijze multiple regressie over een analyse over het totale aantal ongevallen op enkelbaans wegen voor gemengd verkeer (analyse EGIA extra).



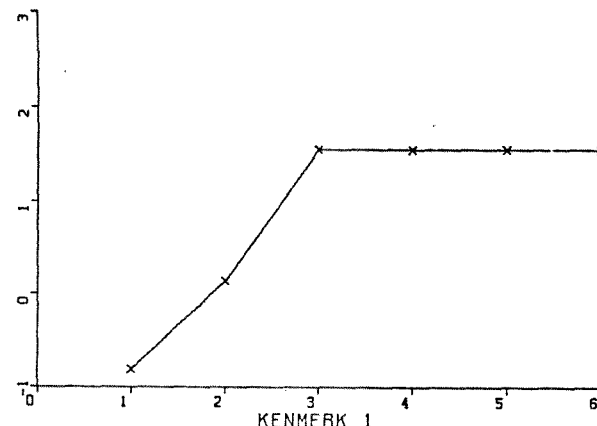
boogstraal

- cat. 1 (-0.547): geen ($R \geq 1000$ m)
- cat. 2 (-0.547): $700 < R < 1000$ m
- cat. 3 (1.005): $500 < R < 700$ m
- cat. 4 (1.929): $400 < R < 500$ m
- cat. 5 (1.929): $200 < R < 400$ m
- cat. 6 (1.929): $R < 200$ m



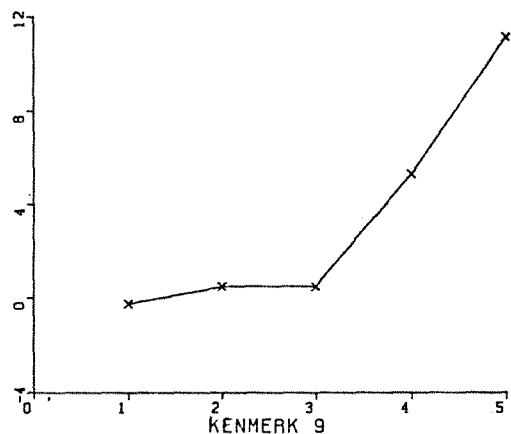
verlichting

- cat. 1 (-0.942): niet aanwezig
- cat. 2 (1.062): gedeeltelijk
- cat. 3 (1.062): oriëntatieverl.
- cat. 4 (1.062): geheel



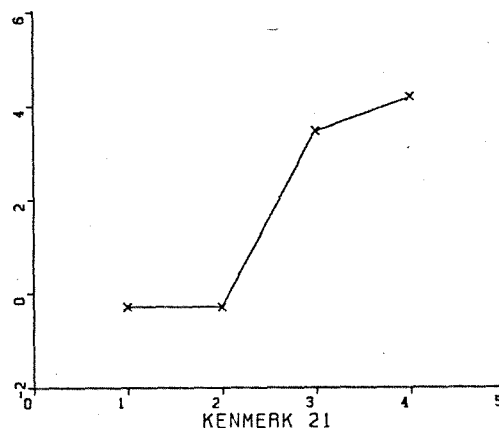
verhardingsbreedte

- cat. 1 (-0.806): $B < 6,0$ m
- cat. 2 (0.146): $6,0 < B \leq 6,5$ m
- cat. 3 (1.563): $6,5 < B \leq 7,0$ m
- cat. 4 (1.563): $7,0 < B \leq 7,5$ m
- cat. 5 (1.563): $7,5 < B \leq 8,0$ m
- cat. 6 (1.563): $B > 8,0$ m



lichtmasten (langs beide raai-
vakken aan beide zijden)

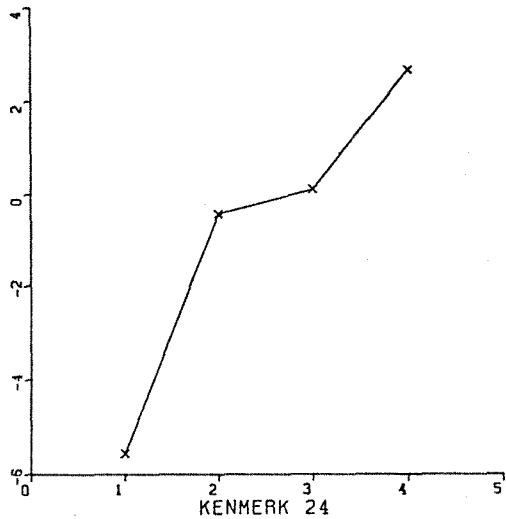
- cat. 1 (-0.223): geen
- cat. 2 (0.517): een
- cat. 3 (0.517): twee
- cat. 4 (5.336): drie
- cat. 5 (11.180): vier



aantal kruispunten type B

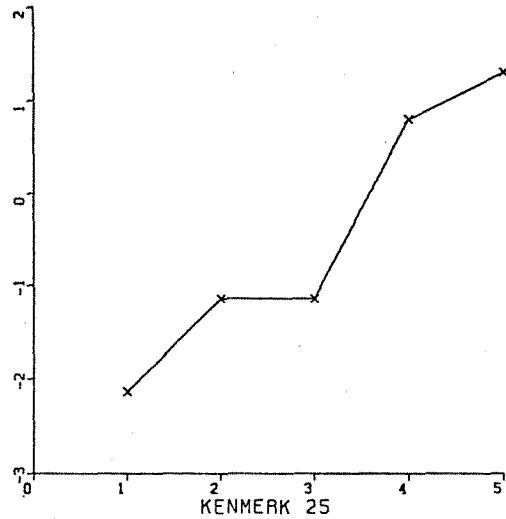
- cat. 1 (-0.280): geen
- cat. 2 (-0.280): een
- cat. 3 (3.486): twee
- cat. 4 (4.203): drie

categorie scores voor de belang-
 rijkste wegkenmerken van de
 eerste dimensie



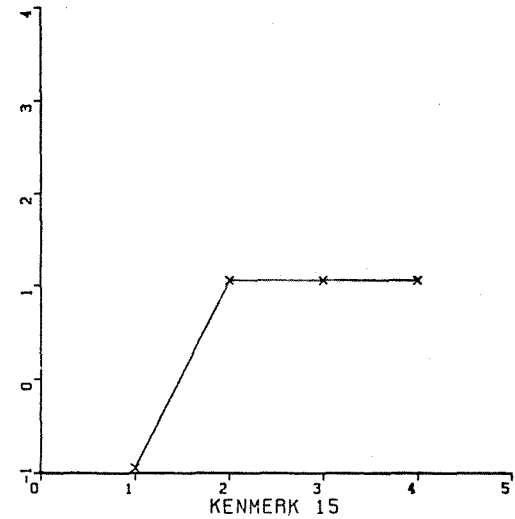
percentage vrachtverkeer

- cat. 1 (-5.556): nul
- cat. 2 (-0.416): $0 < p \leq 10$
- cat. 3 (0.127): $10 < p \leq 20$
- cat. 4 (2.693): $p > 20$



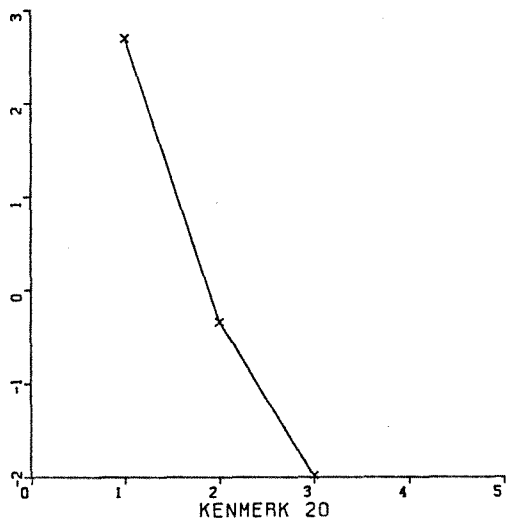
intensiteit (brom)fietsverkeer

- cat. 1 (-2.129): $I = 0$
- cat. 2 (-1.134): $0 < I \leq 118$
- cat. 3 (-1.134): $118 < I \leq 352$
- cat. 4 (0.816): $352 < I \leq 1054$



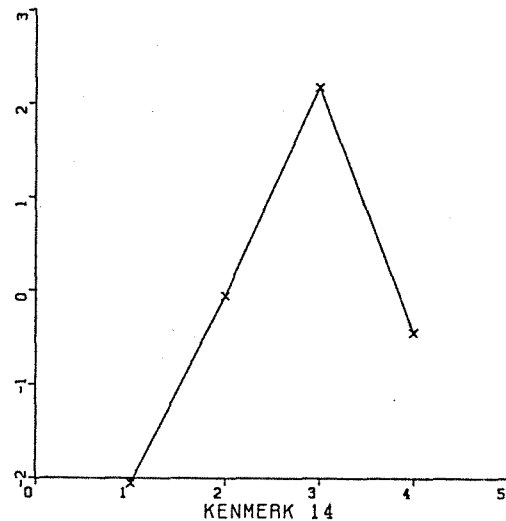
verlichting

- cat. 1 (-0.942): niet aanwezig
- cat. 2 (1.062): gedeeltelijk
- cat. 3 (1.062): oriëntatieverl.
- cat. 4 (1.062): geheel



parkeerverbod

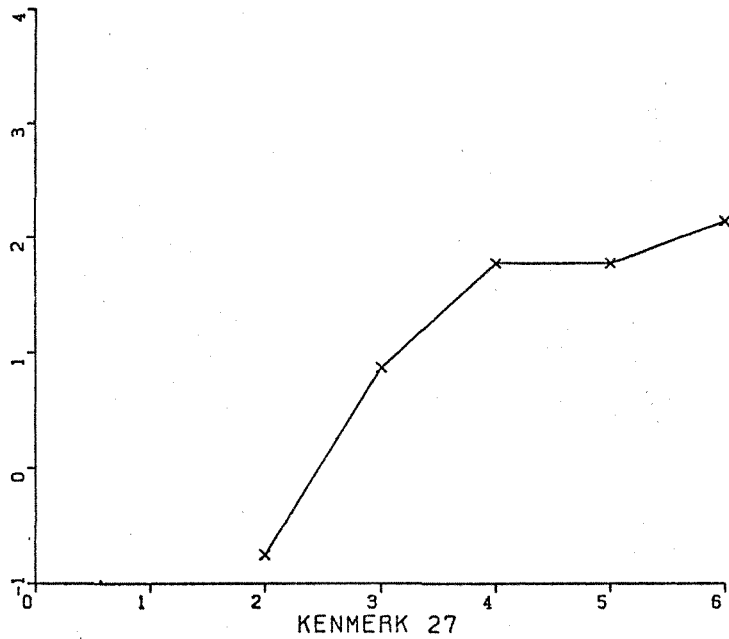
- cat. 1 (2.699): geen
- cat. 2 (-0.346): voorrangsweg
- cat. 3 (-1.984): parkeer- of stop-verbodsbord



verharding

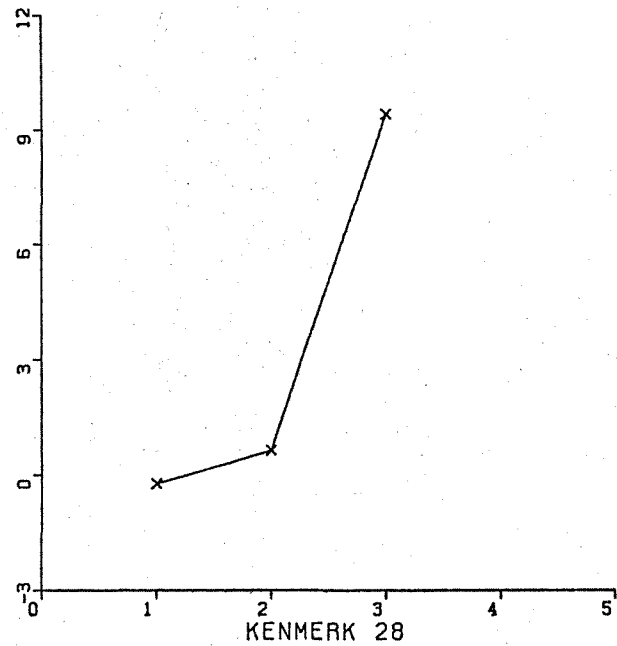
- cat. 1 (-2.055): beton
- cat. 2 (-0.056): asfalt
- cat. 3 (2.183): klinkers
- cat. 4 (-0.453): keien/tegels

categorie scores voor de belang-
 rijkste wegkenmerken van de
 tweede dimensie



totaal aantal ongevallen

cat. x: aantal ongevallen minus 1



dodelijke ongevallen

cat. x: aantal dodelijke ongevallen
minus 1

categoriescores
ongevalskenmerken

dim 1				dim 2			
	ondergrens	gemidd.	bovengrens		ondergrens	gemidd.	bovengrens
can.corr.	0.456	0.470	0.484	can.corr.	0.310	0.336	0.362
kenmerk	ondergrens	gemidd.	bovengrens	kenmerk	ondergrens	gemidd.	bovengrens
1	-0.114	-0.087	-0.059	1	-0.079	-0.056	-0.032
2	-0.071	-0.040	-0.009	2	-0.012	0.013	0.038
3	-0.014	0.004	0.022	3	-0.024	0.000	0.024
4	0.040	0.073	0.106	4	0.014	0.041	0.068
5	-0.145	-0.114	-0.083	5	0.025	0.064	0.103
6	-0.075	-0.042	-0.009	6	-0.035	-0.013	0.009
7	-0.010	0.017	0.044	7	0.026	0.071	0.166
8	-0.071	-0.059	-0.047	8	-0.027	0.002	0.031
9	0.114	0.157	0.198	9	-0.041	0.006	0.053
10	-0.060	-0.052	-0.044	10	-0.026	-0.018	-0.010
11	-0.068	-0.031	0.006	11	-0.036	-0.012	0.012
12	0.248	0.268	0.288	12	-0.038	-0.003	0.032
13	-0.003	0.026	0.055	13	-0.045	-0.025	-0.005
14	0.052	0.107	0.160	14	-0.095	-0.066	-0.037
15	0.073	0.119	0.166	15	-0.089	-0.064	-0.039
16	-0.110	-0.075	-0.040	16	-0.049	-0.020	0.009
17	0.042	0.060	0.077	17	-0.004	0.014	0.032
18	-0.020	0.009	0.038	18	-0.035	-0.011	0.013
19	-0.071	-0.038	-0.005	19	-0.050	-0.034	-0.018
20	-0.099	-0.050	-0.001	20	-0.014	0.017	0.048
21	0.036	0.069	0.102	21	-0.067	-0.042	-0.017
22	-0.031	0.006	0.043	22	-0.047	-0.016	0.015
23	-0.005	0.028	0.061	23	-0.046	-0.030	-0.014
24	-0.060	-0.035	-0.010	24	0.021	0.058	0.095
25	-0.037	-0.008	0.021	25	0.030	0.059	0.088
26	-0.063	-0.030	0.003	26	-0.050	-0.015	0.020
27	0.978	0.992	0.997	27	-0.146	-0.014	0.119
28	0.056	0.204	0.343	28	0.960	0.987	0.996

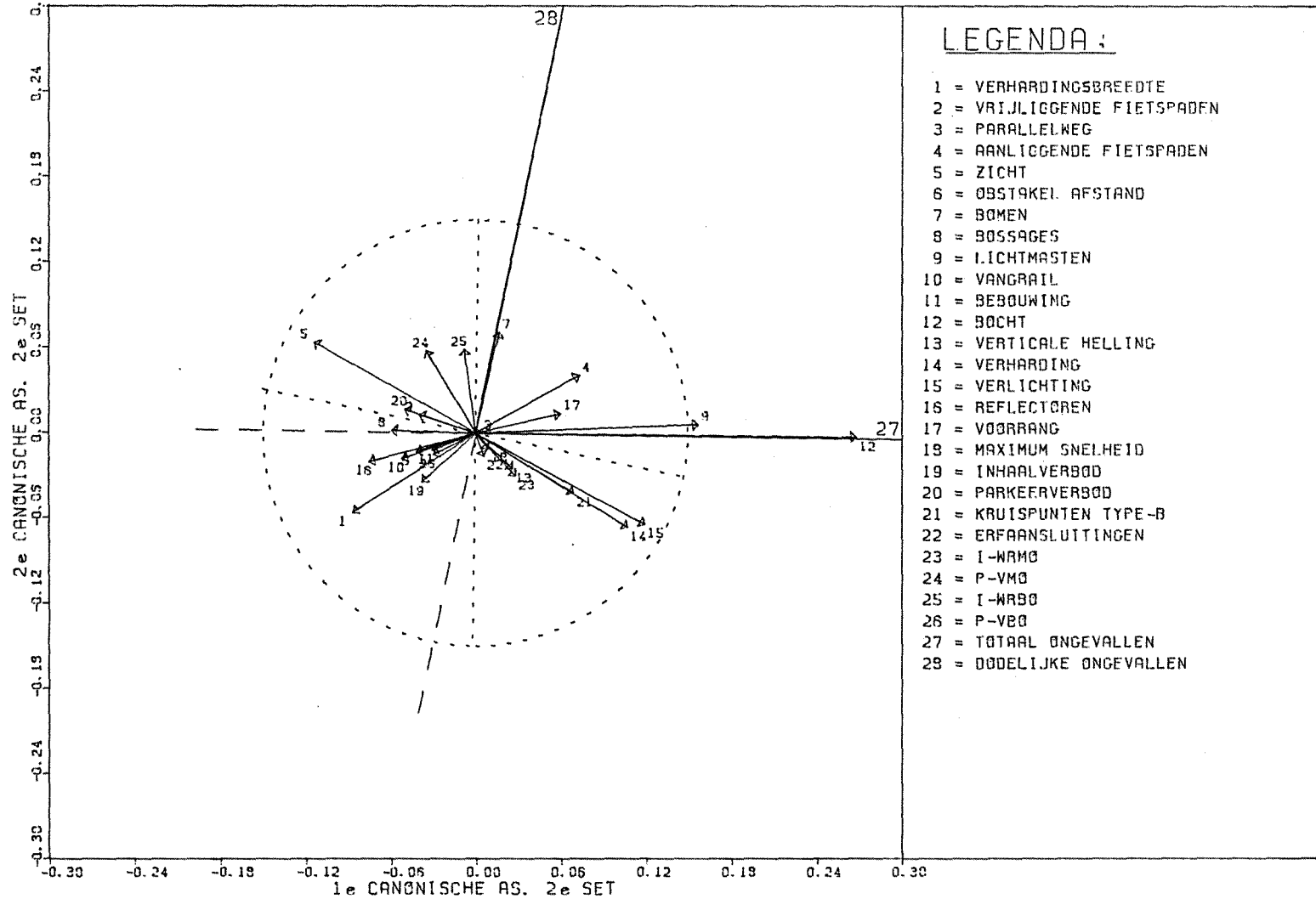
$$z = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right)$$

ondergrens gebaseerd op de waarde $\bar{z} - 1.96 \frac{s_z}{z}$

bovengrens " " " " $\bar{z} + 1.96 \frac{s_z}{z}$

Gemiddelde correlaties met betrouwbaarheidsgrenzen voor EG 1 AD, gebaseerd op 10 bootstrap-analyses.

EG 1 AD - BOOTSTRAPS-GEMIDDELDE - TOTAAL VS DODELIJK - ONG. LOC.



eerste dimensie

locatie- code	score wegkenmerken- as	score ongevallen- as	letsel- ongevallen + 1	dodelijke ongevallen + 1	hoogstraal	verlichting	verhardings- breedte	lichtmasten	kruispunten type B
locaties met hoogste canonische scores van de eerste set:									
10210716	3.699	1.735	4	1	-0.547	1.062	-0.806	11.180	3.486
10210405	3.197	0.845	3	1	1.929	1.062	0.146	5.336	-0.280
10208437	3.052	-0.756	2	1	1.929	1.062	-0.806	0.517	3.486
10208436	2.945	1.810	4	2	1.929	1.062	-0.806	0.517	4.203
10205126	2.859	0.920	3	2	1.929	1.062	0.146	-0.223	-0.280
10205117	2.720	0.845	3	1	1.929	1.062	-0.806	-0.223	-0.280
10201623	2.590	-0.756	2	1	1.929	1.062	0.146	5.336	3.486
10208404	2.572	1.735	4	1	1.929	1.062	-0.806	0.517	-0.280
10210418	2.493	1.810	5	2	1.929	1.062	-0.806	0.517	-0.280
10205672	2.420	0.845	3	1	1.929	1.062	-0.806	-0.223	-0.280
10210854	2.414	1.735	4	1	1.929	1.062	1.563	0.517	-0.280
10200417	2.296	1.735	4	1	-0.547	1.062	0.146	11.180	-0.280
10210415	2.250	0.845	3	1	1.929	1.062	-0.806	0.517	-0.280
10204427	2.236	0.845	3	1	1.929	1.062	-0.806	-0.223	-0.280
10205664	2.188	-0.756	2	1	1.929	1.062	-0.806	-0.223	-0.280
10204420	2.185	0.920	3	2	1.929	1.062	-0.806	-0.223	3.486
10210403	2.181	0.845	3	1	1.929	1.062	-0.806	0.517	-0.280
10201868	2.174	0.845	3	1	1.929	-0.942	0.146	-0.223	-0.280
10202556	2.139	0.845	3	1	1.929	1.062	-0.806	0.517	-0.280
10201019	2.090	0.920	3	2	-0.547	1.062	-0.806	5.336	-0.280
10200419	2.057	0.845	3	1	1.929	1.062	1.563	5.336	-0.280
10206377	2.041	2.922	6	3	1.929	-0.942	1.563	0.517	-0.280
10210428	2.032	0.920	3	2	1.929	1.062	-0.806	0.517	-0.280
10202010	2.011	-0.756	2	1	1.929	1.062	-0.806	-0.223	-0.280
10211227	2.008	1.735	4	1	1.929	1.062	-0.806	0.517	-0.280
10208381	2.006	-0.756	2	1	1.929	1.062	-0.806	0.517	-0.280
10201980	2.002	0.845	3	1	1.929	1.062	-0.806	-0.223	-0.280
10210844	1.947	2.094	6	1	1.929	-0.942	1.563	-0.223	-0.280
10210834	1.940	0.920	3	2	1.929	-0.942	0.146	-0.223	-0.280
10204987	1.919	2.563	4	3	1.929	1.062	-0.806	-0.223	-0.280

locaties met laagste canonische scores van de eerste set:

10210223	-1.556	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10204860	-1.570	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10206189	-1.575	-0.756	2	1	-0.547	1.062	1.563	-0.223	-0.280
10210393	-1.584	-0.756	2	1	-0.547	1.062	1.563	-0.223	-0.280
10209004	-1.671	-0.756	2	1	-0.547	1.062	-0.806	-0.223	-0.280
10202430	-1.689	-0.756	2	1	1.929	1.062	1.563	-0.223	-0.280
10202418	-1.691	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10206117	-1.770	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10202428	-1.791	-0.756	2	1	1.005	-0.942	1.563	-0.223	-0.280
10210388	-1.804	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10210692	-1.816	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10130352	-1.827	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10210335	-1.858	-0.681	2	2	-0.547	1.062	1.563	0.517	-0.280
10209612	-1.876	-0.756	2	1	-0.547	1.062	1.563	0.517	-0.280
10208991	-1.927	-0.756	2	1	-0.547	1.062	-0.806	-0.223	-0.280
10209650	-1.955	0.845	3	1	-0.547	-0.942	0.146	-0.223	-0.280
10204492	-1.957	-0.756	2	1	-0.547	-0.942	-0.806	-0.223	-0.280
10202411	-2.116	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10202395	-2.124	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10211478	-2.223	-0.756	2	1	-0.547	1.062	1.563	0.517	-0.280
10130379	-2.227	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10208990	-2.230	-0.756	2	1	-0.547	1.062	-0.806	-0.223	-0.280
10208963	-2.230	-0.756	2	1	-0.547	1.062	-0.806	-0.223	-0.280
10208966	-2.230	-0.756	2	1	-0.547	1.062	-0.806	-0.223	-0.280
10209610	-2.264	-0.756	2	1	-0.547	1.062	1.563	-0.223	-0.280
10206253	-2.299	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10202406	-2.334	-0.756	2	1	-0.547	-0.942	1.563	-0.223	-0.280
10209641	-2.437	-0.756	2	1	-0.547	-0.942	0.146	0.517	-0.280
10209646	-2.606	-0.756	2	1	-0.547	-0.942	0.146	-0.223	-0.280
10202126	-2.667	-0.756	2	1	-0.547	1.062	1.563	-0.223	-0.280

tweede dimensie

locatie- code	score wegkenmerken as	score ongevallen- as	letsel- ongevallen + dodelijke ongevallen +	boogstraat	verlichting	verhardings- breedte	lichtmasten	kruispunten type B
locaties met hoogste canonische scores van de eerste set:								
10204992	4.957	-0.014	2 1	2.693	0.816	-0.942	2.699	-0.056
10204994	4.957	-0.443	3 1	2.693	0.816	-0.942	2.699	-0.056
10204987	4.094	9.066	4 3	2.693	0.816	1.062	2.699	-0.056
10204985	4.063	9.304	3 3	2.693	0.816	1.062	2.699	-0.056
10204984	3.794	-0.014	2 1	2.693	0.816	1.062	2.699	-0.056
10208951	3.488	-0.014	2 1	2.693	0.816	1.062	-0.346	2.183
10204792	3.328	-0.014	2 1	2.693	0.816	-0.942	-0.346	-0.056
10201603	3.286	0.867	2 2	0.128	0.816	-0.942	-0.346	-0.056
10204788	3.195	0.438	3 2	2.693	0.816	-0.942	-0.346	-0.056
10204790	3.128	-0.443	3 1	2.693	0.816	1.062	-0.346	-0.056
10206395	2.961	-0.443	3 1	2.693	0.816	-0.942	-0.346	-2.055
10210925	2.911	-0.778	6 1	2.693	-1.134	-0.942	2.699	-0.056
10210930	2.844	-0.014	2 1	2.693	-1.134	-0.942	2.699	-0.056
10201606	2.733	9.304	3 3	0.128	0.816	-0.942	-0.346	-0.056
10210921	2.534	-0.014	2 1	2.693	-1.134	-0.942	2.699	-0.056
10210920	2.534	-0.014	2 1	2.693	-1.134	-0.942	2.699	-0.056
10211647	2.441	-0.443	3 1	0.128	1.314	1.062	2.699	-0.056
10206402	2.394	-0.014	2 1	2.693	0.816	1.062	-0.346	-2.055
10204942	2.332	0.867	2 2	0.128	0.816	-0.942	2.699	-0.056
10206387	2.256	-0.443	3 1	0.128	0.816	-0.942	-0.346	-2.055
10204934	2.245	0.867	2 2	0.128	0.816	-0.942	2.699	-0.056
10208004	2.199	-0.014	2 1	0.128	-1.134	-0.942	-0.346	-0.056
10204492	2.196	-0.014	2 1	0.128	-1.134	-0.942	-0.346	-0.056
10206380	2.157	-0.014	2 1	0.128	0.816	-0.942	-0.346	-2.055
10206385	2.157	-0.014	2 1	0.128	0.816	-0.942	-0.346	-2.055
10206382	2.157	-0.014	2 1	0.128	0.816	-0.942	-0.346	-2.055
10206384	2.157	-0.014	2 1	0.128	0.816	-0.942	-0.346	-2.055
10210923	2.153	-0.014	2 1	2.693	-1.134	-0.942	2.699	-0.056
10206377	2.146	8.969	6 3	0.128	0.816	-0.942	-0.346	-2.055
10209026	2.138	0.867	2 2	0.128	0.816	-0.942	2.699	-0.056
locaties met laagste canonische scores van de eerste set:								
10201324	-1.508	0.867	2 2	-0.416	1.314	1.062	-0.346	-0.056
10208389	-1.512	-0.443	3 1	0.128	-1.134	1.062	-0.346	-0.056
10208412	-1.516	-0.443	3 1	0.128	-1.134	1.062	-0.346	-0.056
10208436	-1.568	0.199	4 2	0.128	-1.134	1.062	-0.346	-0.056
10201665	-1.571	-0.014	2 1	0.128	-1.134	1.062	-0.346	2.183
10202097	-1.623	-0.014	2 1	-0.416	-1.134	-0.942	-0.346	-0.056
10202556	-1.690	-0.443	3 1	-0.416	-1.134	1.062	-0.346	-0.056
10201651	-1.691	-0.014	2 1	0.128	-1.134	1.062	-0.346	-0.056
10208399	-1.692	-0.682	4 1	0.128	-1.134	1.062	-0.346	-0.056
10206174	-1.694	-0.682	4 1	-0.416	-1.134	1.062	-0.346	-0.056
10200417	-1.715	-0.682	4 1	0.128	0.816	1.062	-0.346	-0.056
10203438	-1.731	-0.443	3 1	0.128	0.816	1.062	-0.346	2.183
10202416	-1.760	-0.014	2 1	-0.416	-1.134	1.062	-0.346	2.183
10202363	-1.787	-0.443	3 1	-0.416	-1.134	1.062	-0.346	2.183
10201626	-1.805	-0.443	3 1	0.128	-1.134	1.062	-0.346	-0.056
10206167	-1.806	-0.443	3 1	-0.416	-1.134	-0.942	-0.346	-0.056
10206172	-1.806	-0.014	2 1	-0.416	-1.134	-0.942	-0.346	-0.056
10202096	-1.813	-0.443	3 1	-0.416	-1.134	-0.942	-0.346	-0.056
10211055	-1.843	-0.014	2 1	-0.416	-1.134	1.062	2.699	2.183
10206081	-1.868	-0.014	2 1	0.128	1.314	1.062	-0.346	-0.056
10203272	-1.887	-0.443	3 1	0.128	0.816	-0.942	-0.346	-2.055
10201583	-1.921	-0.014	2 1	0.128	-1.134	1.062	-0.346	-0.056
10202391	-1.924	-0.014	2 1	-0.416	-1.134	1.062	-0.346	2.183
10205111	-1.930	-0.014	2 1	-5.556	-2.129	1.062	-0.346	2.183
10202401	-2.009	-0.014	2 1	-0.416	-1.134	1.062	-0.346	2.183
10203441	-2.210	-0.014	2 1	0.128	0.816	1.062	-0.346	2.183
10206554	-2.269	-0.014	2 1	0.128	-1.134	1.062	-0.346	-0.056
10205106	-2.625	-0.443	3 4	-5.556	-2.129	1.062	-0.346	2.183
10201623	-2.845	-0.014	2 1	0.128	-1.134	1.062	-0.346	-0.056
10200221	-3.443	-0.014	2 1	-0.416	0.816	1.062	-0.346	-0.056