

Analysetechnieken voor multivariate analyse van verkeersveiligheidsgegevens

D-93-11

S. Oppe

Leidschendam, 1993

Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV

Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV
Postbus 170
2260 AD Leidschendam
Telefoon 070-3209323
Telefax 070-3201261

Inhoud

1. *Inleiding*
2. *HOMALS*
3. *PRINCALS*
4. *CANALS*
5. *WPM*

Literatuur

Bijlagen 1 t/m 3

1. Inleiding

De meer klassieke technieken voor multivariate analyse, zoals variantie- en covariantie-analyse, factoranalyse, discriminantanalyse, multiple lineaire regressie en canonische correlatie, zijn vaak niet toepasbaar bij de analyse van problemen in de verkeersveiligheid.

Voor een belangrijk deel heeft dit te maken met het niveau waarop de gegevens zijn gemeten. Sommige kenmerken zijn echte metingen (zoals snelheden van voertuigen of de leeftijd van bestuurders), soms is alleen de rangorde van gegevens bekend, maar niet een meetwaarde (bijvoorbeeld: grote, middelgrote en kleine gemeenten), soms slechts onderscheid in klassen zonder dat daaraan direct een ordening kan worden toegekend (bijvoorbeeld het nominale kenmerk: wijze van verkeersdeelname). Meestal komen mengvormen voor.

Een bijkomend probleem is dat vaak kenmerken worden onderzocht in relatie tot aantallen ongevallen. In dat geval is dus niet sprake van een meting, maar van een telling. Lineaire modellen zijn dan niet adequaat gebleken. Meestal worden dan log-lineaire modellen gebruikt. Het bekendste programma voor log-lineaire analyse is GLIM. Dergelijke methoden zijn nog niet ontwikkeld voor analyses waarin de kenmerken worden gerelateerd aan meerdere reeksen ongevallen tegelijk.

De afdeling Datatheorie van de Rijksuniversiteit Leiden heeft modellen ontwikkeld die wel kunnen worden toegepast in bovengenoemde situaties. Deze modellen kunnen worden gebruikt als het meetniveau van één of meer kenmerken ordinaal of nominaal is. Ook aantallen kunnen met deze programma's worden geanalyseerd.

Het belangrijkste kenmerk van deze technieken is dat gezocht wordt naar een zodanige weergave van de klassen van een kenmerk op een metrische schaal, dat een optimale oplossing van klassieke technieken wordt bereikt. Naast de gebruikelijke informatie over de samenhangen tussen kenmerken, krijgt men eveneens van elk kenmerk een schaling van de klassen.

Men zou bijvoorbeeld kunnen doen alsof van echte metingen slechts de rangorde bekend is, om vervolgens achteraf na te gaan of de door het programma gereconstrueerde schaal klopt met de feitelijke waarden. Als er tussen twee kenmerken een logaritmisch verband bestaat, dan zal het programma de klassen van het ene kenmerk logaritmisch (of het andere kenmerk exponentieel) herschalen om de beste beschrijving voor hun samenhang te geven. Op deze manier is het dus ook mogelijk om niet-lineaire relaties tussen kenmerken te onderzoeken. Dit is tevens de manier om aantallen ongevallen te analyseren. Bij toepassing in de praktijk is gebleken dat de aantallen ongevallen inderdaad logaritmisch worden geschaald.

De Afdeling Datatheorie van de Rijksuniversiteit Leiden heeft ook programma's ontwikkeld voor gegeneraliseerde principale componenten-analyse PCA (soms beter bekend onder de naam factoranalyse). Het meest gebruikte programma HOMALS, is echter door de combinatie met de schaling veel ruimer toepasbaar dan voor problemen die met de klassieke PCA werden opgelost. Zo wordt het programma veelvuldig gebruikt bij de analyse van enquêtegegevens om samenhangen tussen responsecategorieën (responseprofielen) op te sporen of als multidimensionele schaaltechniek.

Daarnaast is het programma PRINCALS ontwikkeld dat meer geëigend is voor PCA-toepassingen.

Voor multiple lineaire regressie, discriminantanalyse, canonische discriminantanalyse en canonische correlatie-analyse is het programma CANALS ontwikkeld.

Deze programma's hebben vooral hun nut bewezen als exploratieve technieken, dus bij het onderzoeken van relaties tussen vele kenmerken waarvan de onderlinge structuur niet bekend is.

De hierboven genoemde technieken zijn bedoeld voor de analyse van objecten zoals ongevallen, slachtoffers etc. die elk een aantal kenmerken bezitten.

Soms echter worden deze objecten eerst ingedeeld in zogenaamde kruistabellen voordat ze worden geanalyseerd. De analyse-eenheid is dan niet een object met zijn kenmerken, maar een combinatie van kenmerken en het bijbehorend aantal objecten. De meest gebruikte techniek voor de analyse van (meerdimensionele) kruistabellen is de zogenaamde log-lineaire analyse. Voor het verkeersveiligheidsonderzoek is het echter vaak nodig om op de aantallen een correctie toe te passen, bijvoorbeeld voor de mate van blootstelling aan het verkeer. Samen met de SWOV heeft de afdeling Datatheorie van de Rijksuniversiteit Leiden een variant van de log-lineaire analysetechniek ontwikkeld waarin dit mogelijk is. Deze techniek heeft de naam 'Weighted Poisson Models' (WPM).

De hierbovengenoemde technieken zullen kort worden beschreven. Voor een uitgebreide behandeling van HOMALS, PRINCALS en CANALS wordt verwezen naar Gifi (1990). Een beschrijving van WPM is te vinden in De Leeuw & Oppe (1976). In de bijlagen zijn enkele voorbeelden opgenomen van de behandelde technieken.

2. HOMALS

HOMALS is een programma voor homogeniteitsanalyse. Als de analyse-eenheid bijvoorbeeld ongevallen is (dit kunnen ook voertuigen zijn, of slachtoffers of enquêtes), dan zoekt HOMALS naar een afbeelding van alle ongevallen in een ruimte van een dimensionaliteit die door de onderzoeker wordt opgegeven. De klassen van ongevallenkenmerken kunnen in diezelfde ruimte worden afgebeeld, als de zwaartepunten van de ongevallen die tot die klassen behoren. Kenmerken waarvan de klassen ver uiteen liggen discrimineren goed tussen ongevallen. Klassen van verschillende kenmerken die dicht bij elkaar liggen komen vaak gezamenlijk voor. Clusters van kenmerken kunnen dus worden gebruikt om homogene groepen ongevallen te definiëren.

Zoals ook bij de klassieke PCA geldt, zijn de oplossingsdimensies waarop de ongevallen worden afgebeeld die combinaties van de oorspronkelijke kenmerken die de gegevens het best weergeven.

Kort samengevat geeft een homogeniteitsanalyse een eenvoudig beeld van complexe relatiestructuren. Dit beeld wordt bepaald door de keuze van kenmerken die worden onderzocht. De objecten (ongevallen) worden oorspronkelijk afgebeeld als punten in een ruimte die wordt opgebouwd uit dimensies die ieder een klasse van een kenmerk representeren. De oplossing betreft dan een afbeelding in een deelruimte waarbij zo weinig mogelijk van de oorspronkelijke informatie verloren gaat. De oplossing wordt beschreven door de volgende getallen:

- Getallen die aangeven hoe goed de objecten (hier ongevallen) zijn af te beelden in een kleinere ruimte dan de ruimte waarin ze oorspronkelijk zijn afgebeeld. Deze getallen (de 'eigenwaarden') geven aan hoeveel procent van de totale variantie wordt verklaard door de dimensies die voor de eenvoudige beschrijving zijn gekozen.
- Getallen die aangeven hoe goed de gekozen dimensies overeenkomen met de oorspronkelijke kenmerken. Deze getallen (de 'discriminatie-maten') geven aan hoeveel procent van de variantie van de objecten op het oorspronkelijke kenmerk overblijft na projectie op de nieuwe dimensie.
- Getallen die aangeven waar de individuele objecten in de nieuwe ruimte worden afgebeeld. Deze getallen (de objectscores) zijn in feite de coördinaten in de nieuwe ruimte van de punten die de objecten representeren. Objecten in het centrum van de afbeelding zijn 'doorsnee' objecten en hebben dus veel gemeenschappelijke kenmerken, objecten aan de rand wijken het meest af.
- Getallen die aangeven welke objectscore de objecten in een klasse van een kenmerk gemiddeld hebben op een dimensie. Voor elk kenmerk waarin de objecten geclassificeerd zijn geldt dat deze getallen (de 'categoriescores' of de 'categoriekwantificaties') veel van elkaar verschillen als de klassen goed van elkaar zijn te onderscheiden, dus weinig overlap vertonen. Dit is het geval bij goed discriminerende kenmerken. Klassen met weinig observaties zijn over het algemeen niet erg stabiel. Dit betekent dat bij een andere steekproef andere waarden gevonden kunnen worden. Klassen van verschillende, goed onderscheidende kenmerken die dicht bij elkaar liggen hebben veel objecten gemeen. Zij representeren homogene groepen objecten.

2. PRINCALS

Zoals gezegd is PRINCALS een programma dat meer lijkt op de klassieke principale componenten-analyse, zij het dat ook hier een herschaling van de kenmerken plaatsvindt. PRINCALS en CANALS onderscheiden zich van HOMALS doordat een variabele in deze programma's slechts éénmaal wordt herschaald, voor alle dimensies tegelijk. In HOMALS wordt voor elke dimensie opnieuw herschaald. De achterliggende gedachte is dat deze programma's meer gebruikt worden voor metrische of geordende gegevens, waarvan de schaling niet bekend is en moet worden gereconstrueerd, of waarbij verondersteld wordt dat de onderlinge relaties tussen de kenmerken niet rechtlijnig zijn, maar kromlijnig (bijvoorbeeld logaritmisch of U-vormig). HOMALS gaat in principe uit van echte nominale gegevens. Elke dimensie laat dan een ander aspect van hetzelfde kenmerk zien. In het HOMALS-voorbeeld wordt van het kenmerk 'conflictype' op de eerste dimensie vooral het onderscheid tussen de verschillende soorten ongevallen van het rijdende verkeer onderling beschreven en op de tweede dimensie het onderscheid tussen voetganger ongevallen en de ongevallen van het rijdende verkeer onderling (zie Bijlage 1, Afbeelding 1.3). In HOMALS kan de schaling niet beperkt worden tot een ordinale of metrische schaal. Een HOMALS-analyse in één dimensie is gelijk aan een PRINCALS-analyse in één dimensie, met nominale schaling van alle kenmerken.

3. CANALS

CANALS is een gegeneraliseerd programma voor canonische correlatie-analyse, waarmee alle mengvormen van metrische en niet-metrische kenmerken kunnen worden geanalyseerd. Bij de klassiek canonische correlatie-analyse worden voor afzonderlijke objecten - zoals wegsecties - zodanige lineaire combinaties van kenmerken in een groep van kenmerken berekend dat een maximale correlatie ontstaat met lineaire combinaties van kenmerken in een tweede groep. Het gaat daarbij dus om die combinatie van kenmerken uit de eerste groep die zoveel mogelijk samenhangt met een combinatie uit de tweede groep. Verschillende (onafhankelijke) oplossingen zijn mogelijk. De oplossingsruimte is dan meer-dimensionaal. De correlatie tussen beide combinaties wordt een canonische correlatie-coëfficiënt genoemd. Bij deze klassieke analyse kan alleen met metrische gegevens gewerkt worden.

Als er in de tweede groep maar één kenmerk is, dan gaat de canonische correlatie-analyse over in multiple lineaire regressie-analyse. Gezocht wordt dan naar die combinatie van kenmerken (bijvoorbeeld wegkenmerken) die een ander kenmerk (bijvoorbeeld ongevallen) zo goed mogelijk voorspelt. De kenmerken uit de eerste groep heten dan de onafhankelijke variabelen, het te voorspellen kenmerk de afhankelijke variabele.

Bij CANALS kunnen ook variabelen worden gebruikt die niet metrisch zijn. Deze nominale en ordinale kenmerken worden op zodanige wijze herschaald dat de canonische correlatiecoëfficiënt zo hoog mogelijk wordt. De combinaties van kenmerken zijn pas na schaling lineair. Door deze schaling zijn er meer mogelijkheden om CANALS te gebruiken. Is de afhankelijke variabele bijvoorbeeld nominaal dan gaat de multiple regressie-analyse over in (canonische) discriminantanalyse. Verder is gebruik van CANALS aan te raden bij regressie-analyse, als de verbanden wel metrisch, maar de relaties tussen de variabelen mogelijk niet lineair zijn. Een nominale of ordinale schaling laat dan de aard van de relatie zien. CANALS is bijzonder goed bruikbaar om bijvoorbeeld die combinaties van weg- en verkeerskenmerken op te sporen die de beste voorspelling geven van (combinaties van) kenmerken van ongevallen. De ordinale transformatie van het aantal ongevallen blijkt dan vaak logaritmisch te zijn. Bij een log-lineaire analyse met bijvoorbeeld GLIM wordt van deze logaritmetransformatie uitgegaan. CANALS laat zien dat dit inderdaad de meest efficiënte beschrijving geeft. Oppe (1992) geeft een voorbeeld van toepassing en een vergelijking van CANALS en GLIM.

5. WPM

Sinds het eind van de jaren zestig heeft een sterke ontwikkeling plaatsgevonden in de analyse van kruistabellen. Kruistabellen zijn tabellen waarin aantallen observaties zijn opgesplitst naar één of meer kenmerken. De belangrijkste kenmerken van de nieuwe analysemethode zijn:

- de klassieke chi-kwadraatanalyse voor de twee-wegtabel is uitgebreid naar tabellen met meer dimensies;
- de totale chi-kwadraatwaarde als maat voor interactie tussen kenmerken is op te splitsen in chi-kwadraten voor deelaspecten van die kenmerken;
- relaties tussen klassen kunnen worden beschreven met specifieke parameters;
- de analyse kan ook worden uitgevoerd op gewogen aantallen observaties.

Deze nieuwe methoden worden vaak aangeduid met de term 'log-lineaire analyse van kruistabellen'. Deze term is ontleend aan de modelvorm waarin de tabelgegevens worden beschreven. Niet de celantallen zelf worden geanalyseerd, maar de logaritme van die aantallen. Omdat de basisveronderstelling voor de analyse van kruistabellen berust op een multiplicatief model (van onafhankelijke kansen voor bijvoorbeeld rijen en kolommen), geldt voor de log-lineaire analyse dat deze als een lineair model kan worden geschreven.

Voor een drie-weg tabel geldt bijvoorbeeld dat de logaritme van het cel-aantal y_{ijk} als volgt geschreven kan worden:

$$m_{ijk} = \lambda + \lambda_i + \lambda_j + \lambda_k + \lambda_{ij} + \lambda_{ik} + \lambda_{jk} + \lambda_{ijk}, \quad (1)$$

waarin m_{ijk} de verwachte waarde is voor $\log(y_{ijk})$ en de λ 's parameters zijn voor de rij-bijdrage (i), de kolom-bijdrage (j), de laag-bijdrage (k) en combinaties daarvan (ij, ik etc.).

Met een volledig model zoals (1) kan een perfecte beschrijving van de tabel worden gemaakt. Door het weglaten van parameters, of door het opleggen van restricties daaraan kan de structuur van de tabel worden onderzocht. Er zijn zeer veel tekstboeken verschenen die deze technieken beschrijven. Het meest geciteerde en zeer uitgebreide standaardwerk is Bishop et al. (1975). Hierin staan ook veel voorbeelden. Log-lineaire analysetechnieken kunnen worden uitgevoerd met behulp van diverse statistische pakketten, zoals SAS, GLIM etc. In deze pakketten wordt echter geen mogelijkheid geboden voor een meer-dimensionele kruistabellenanalyse, waarbij de aantallen kunnen worden gewogen. Dit is juist voor verkeersveiligheidsonderzoek een vaak gestelde eis. Bijvoorbeeld om bij een vergelijking van de veiligheid tussen gemeenten te kunnen corrigeren voor het aantal inwoners, voor de lengte van de meetperiode enz.

WPM is een log-lineaire analysetechniek die deze mogelijkheid wel heeft. Verder heeft het een aantal extra nuttige eigenschappen die bij een analyse van pas kunnen komen. Zoals bij de meeste log-lineaire technieken wordt ook hier aangenomen dat de aantallen Poisson-verdeeld zijn. De onderzoeker kan zelf aangeven hoe hij de relaties tussen de klassen van elk kenmerk beschreven wil zien; voor elk kenmerk in de analyse wordt derhalve om een design matrix gevraagd. Om een tabel te kunnen analyseren is de volgende input nodig:

- keuze uit drie analyse-opties;
- het aantal kenmerken in de tabel en van elk het aantal klassen;
- de observaties per cel;
- de weegfactoren per cel;
- de designmatrixen per kenmerk;
- eventueel specifieke modellen die men getoetst wil zien.

Oorspronkelijk bestond er voor het model alleen een 'minimum-chi-kwadraatoplossing'. Later is een maximale aannemelijkheidsoplossing toegevoegd. Het belangrijkste verschil tussen beide oplossingen is een andere definitie van wat wordt verstaan onder een zo goed mogelijk bij de gegevens passend model. Daarbij komt nog dat in het 'minimum chi-kwadraatmodel' eerst alle parameters worden geschat (dit heet een verzadigd model). Vervolgens wordt nagegaan in hoeverre parameters kunnen worden weggelaten zonder dat de tabel significant slechter wordt beschreven door de resterende parameters in het beperkte model. Bij het 'maximale aannemelijkheidsmodel' worden bij het beperkte model nieuwe parameterschattingen gemaakt voor de resterende parameters. In de praktijk zijn de verschillen in oplossing echter niet groot.

Literatuur

Bishop, Y.M.M.; Fienberg, S.E. & Holland, P.W. (1975). *Discrete multivariate analysis: Theory and practice*. MIT-Press, London.

De Leeuw, J. & Oppe, S. (1976). *Analyse van kruistabellen: Log-lineaire Poisson modellen voor gewogen aantallen*. SWOV, Voorburg.

Gifi, A. (1990). *Non-linear multivariate analysis*. Wiley, New York.

Oppe, S. (1992). *A comparison of some statistical techniques for road accident analysis*. *Accid. Anal. & Prev.* 24, 4: 397-423.

SAS/STAT (1990). *Users Guide*. Version 6, Vol. 2. SAS Institute Inc., Cary, NC.

Bijlage 1. Voorbeeld van een HOMALS-analyse

De analyse is toegepast op de ongevallen met dodelijke afloop die in Nederland hebben plaatsgevonden in 1978 en 1982. Dit om de stabiliteit van de beschrijving over een korte tijdperiode te toetsen. Gekozen is voor de kenmerken (met klasse-indeling) van Tabel 1.1.

Voor beide analyses geldt dat 52% van de informatie uit de 20-dimensionele tabel is weer te geven in de driedimensionale deelruimte.

Door toevoeging van meer dimensies krijgt men weliswaar een vollediger, maar tevens onoverzichtelijker beeld. De percentages bedragen voor beide analyses resp. 22%, 16% en 14% voor dimensie 1, 2 en 3.

Ook in alle andere opzichten komen de twee analyses zeer nauw overeen. De volgende uitspraken die zijn afgeleid uit de analyse van 1978 gelden dan ook zowel voor 1978 als voor 1982.

-
1. Dagsoort (zaterdag + zondag; maandag + vrijdag; rest)
 2. Seizoen (december, januari, februari maart, april, mei; juni, juli, augustus; september, oktober, november).
 3. Tijdstip (21 t/m 24 + 0 t/m 6; 7 t/m 8; 9 t/m 15; 16 t/m 20).
 4. Bebouwing (binnen of buiten bebouwde kom)
 5. Snelheidslimiet (<50; 50; <99; 100 km/uur)
 6. Wegbeheerder (gemeente; provincie; rijk)
 7. Plaats (rechte weg; kruispunt; T-, Y-kruisingen + verkeersplein; hoek + bocht)
 8. Verlichting (n.v.t.; geen; wel)
 9. Lichtgesteldheid (dag; schemer; duister)
 10. Weersomstandigheden/toestand wegdek (droog; nat wegdek; regen, sneeuw + mist)
 11. Verhardingssoort (bitumen; rest (klinkers, beton, etc).
 12. Regio (noord + oost; west; zuid)
 13. Grootte gemeente ongeval (> 200.000 inwoners; >50.000 inwoners; rest)
 14. Conflicttype (eenzijdig; voetganger; snel-snel; snel-langzaam; langzaam-langzaam)
 15. Plaats slachtoffer (bestuurder; passagier; voetganger)
 16. Geslacht slachtoffer (man; vrouw)
 17. Leeftijd slachtoffer (0 t/m 5; 6 t/m 11; 12 t/m 17; 18 t/m 35; 36 t/m 64; 65+)
 18. Ernst letsel slachtoffer (dood ter plaats; idem zelfde dag; idem dag + 1; idem dag + 2-5; idem dag + 6-10; idem dag +11-30; opname ziekenhuis)
 19. Wijze verkeersdeelname slachtoffer (pers. auto; vracht-, bestelwagen + bus; motor; bromfiets; fiets; voetganger)
 20. Alcoholgebruik bij één der bestuurders (geen; wel)
-

Tabel 1.1. *Gebruikte kenmerken van ongevallen in Nederland met hun categorieën.*

In Tabel 1.2 zijn de zgn. 'discriminatie-maten' gegeven. Per dimensie geven deze aan in hoeverre de varianten in de scores op een bepaald kenmerk wordt beschreven door die dimensie. Anders gezegd, hoe dicht de

waarde voor een bepaald kenmerk bij één ligt, des te meer representeert die dimensie dat kenmerk; hoe dichter bij nul, hoe minder. Op grond van de combinatie van kenmerken die een hoge discriminatiemaat hebben is het mogelijk een interpretatie aan een dimensie te geven.

| Kenmerken | Dimensies | | |
|------------------------------|-----------|------|------|
| | 1 | 2 | 3 |
| 1. Dagsoort | 0,06 | 0,02 | 0,01 |
| 2. Seizoen | 0,01 | 0,05 | 0,03 |
| 3. Tijdstip | 0,24 | 0,20 | 0,24 |
| 4. Bebouwing | 0,46 | 0,00 | 0,23 |
| 5. Snelheidslimiet | 0,45 | 0,01 | 0,25 |
| 6. Wegbeheerder | 0,22 | 0,00 | 0,14 |
| 7. Plaats (wegsituatie) | 0,15 | 0,12 | 0,02 |
| 8. Verlichting | 0,30 | 0,28 | 0,38 |
| 9. Lichtgesteldheid | 0,21 | 0,29 | 0,32 |
| 10. Weer/toestand wegdek | 0,03 | 0,02 | 0,01 |
| 11. Verhardingssoort | 0,06 | 0,00 | 0,03 |
| 12. Regio | 0,05 | 0,01 | 0,06 |
| 13. Gemeentegrootte | 0,17 | 0,03 | 0,07 |
| 14. Conflicttype | 0,55 | 0,71 | 0,29 |
| 15. Plaats slachtoffer | 0,32 | 0,60 | 0,14 |
| 16. Geslacht slachtoffer | 0,02 | 0,00 | 0,03 |
| 17. Leeftijd slachtoffer | 0,31 | 0,11 | 0,20 |
| 18. Ernst letsel slachtoffer | 0,09 | 0,02 | 0,06 |
| 19. Wijze verkeersdeelname | 0,56 | 0,67 | 0,28 |
| 20. Alcoholgebruik best. | 0,09 | 0,06 | 0,10 |

Tabel 1.2. *Overzicht van de discriminatiematen voor de twintig kenmerken voor elk van die drie dimensies.*

Uit de cijfers van Tabel 1.2 blijkt dat dimensie 1 dominant is als combinatie van kenmerken. Dimensie 2 en 3 hebben niet echt een specifieke betekenis. Ze benadrukken nog eens een bijzonder aspect van dimensie 1. Voor dimensie 2 betreft dit de kenmerken 14, 19 en 15. Voor dimensie 3 zijn de kenmerken 4 en 5 specifiek.

De volgende kenmerken zijn niet van belang voor de samenstelling van homogene groepen:

1. dagsoort
2. seizoen
10. weersomstandigheden/toestand wegdek
11. verhardingssoort
12. regio
16. geslacht slachtoffer
18. ernst letsel slachtoffer
20. alcoholgebruik bestuurder(s)

De kenmerken die in alle drie dimensies belangrijk zijn (in volgorde van belangrijkheid op de eerste dimensie): kenmerk 19, 14, 15, 8, 3 en 9. Omdat de dimensies onafhankelijk zijn, betekent dit dat elke dimensie een ander aspect van het kenmerk weergeeft. We komen hierop terug bij de beschrijving van de categorieën van de kenmerken.

Om wille van de duidelijkheid en eenvoud zullen wij ons verder beperken tot een bespreking van de resultaten voorzover deze betrekking hebben op de eerste en tweede dimensie. Analoge beschrijvingen kunnen ook worden gemaakt voor de gegevens van de andere combinaties van de drie dimensies (bijvoorbeeld dimensie 1 versus dimensie 3).

In Afbeelding 1.1 zijn de discriminatiematen afgebeeld, die de belangrijkheid van de kenmerken voor de twee dimensies aangeven. Deze afbeelding is van geheel andere orde dan Afbeelding 1.2 en 1.3. Daar is sprake van een projectie (doorsnede) van de ruimte waarin de gegevens zijn afgebeeld. In Afbeelding 1.1 is alleen sprake van een visualisering van het belang van de kenmerken voor de dimensies.

We geven nu een voorbeeld van een wijze van interpretatie die aan een dergelijk resultaat kan worden gegeven.

Voor het feit dat de ernst van het letsel niet erg discrimineert, betekent dat de hypothese aannemelijk wordt dat bestrijding van ongevallen met dodelijke afloop ook een teruggang in de letselongevallen betekent. De kenmerken die wel discrimineren zijn met name 14 (type conflict) en 19 (vervoerwijze slachtoffer) en wel in alle drie dimensies, maar vooral op dimensie 1 en 2. Daaraan kan worden toegevoegd kenmerk 15 (wijze van verkeersdeelname: passagier, bestuurder of voetganger).

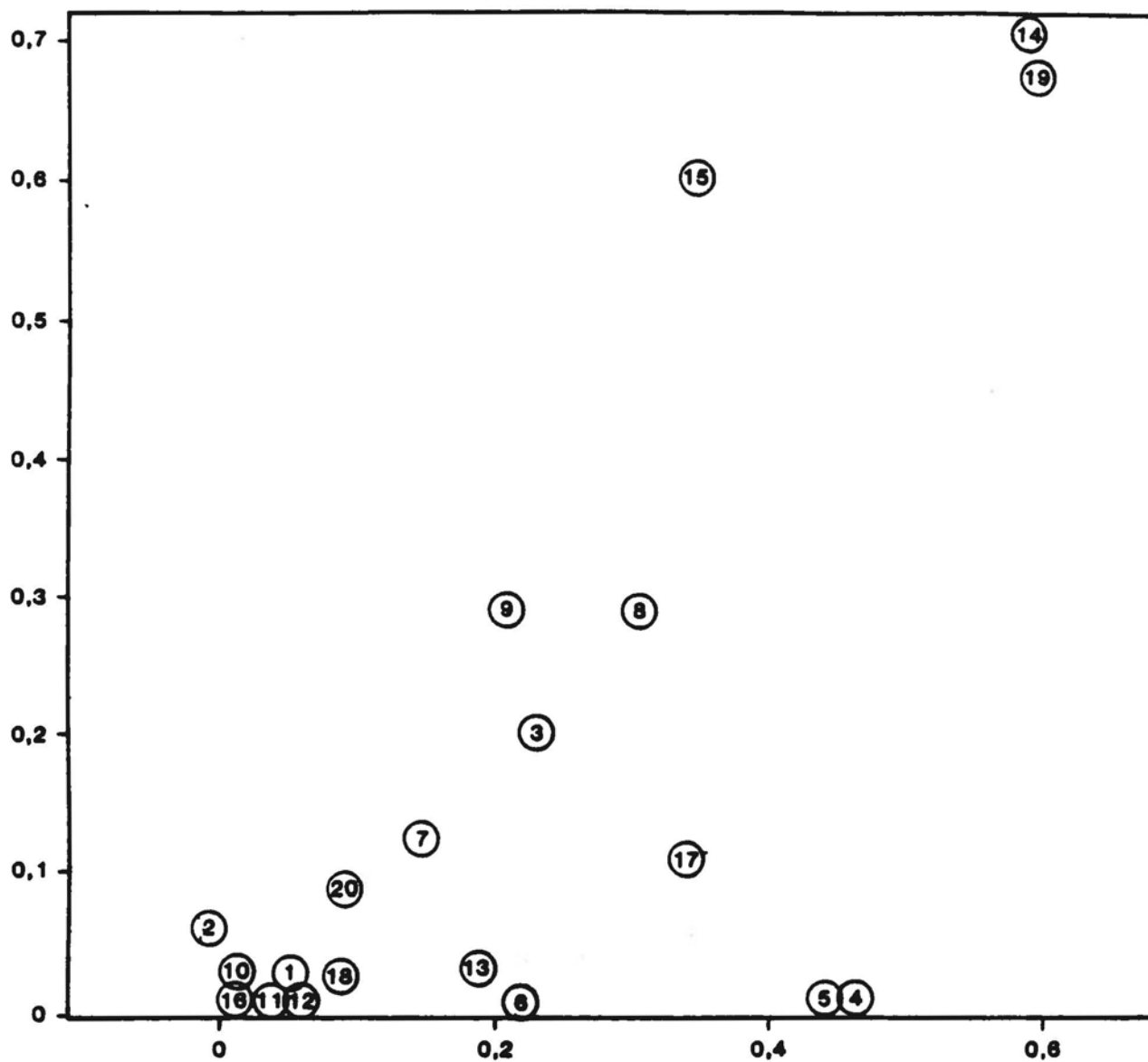
Deze kenmerken hebben veel gemeen. De keuze van bijvoorbeeld kenmerk 14 als hoofddelingskenmerk ligt voor de hand; kenmerk 19 is wellicht beter hanteerbaar.

In Afbeelding 1.2 zijn alle ongevallen afgebeeld op het doorsnijdingsvlak in de ruimte dat hoort bij dimensie 1 en 2. De nummers in de afbeelding corresponderen met de categorieën van kenmerk 14. Een dergelijke afbeelding is er voor ieder kenmerk te maken. De punten liggen telkens op dezelfde plaats in het vlak, alleen de getallen veranderen omdat een klasse van een ander kenmerk wordt aangeduid. De coördinaten van de punten op de beide dimensies worden 'objectscores' genoemd.

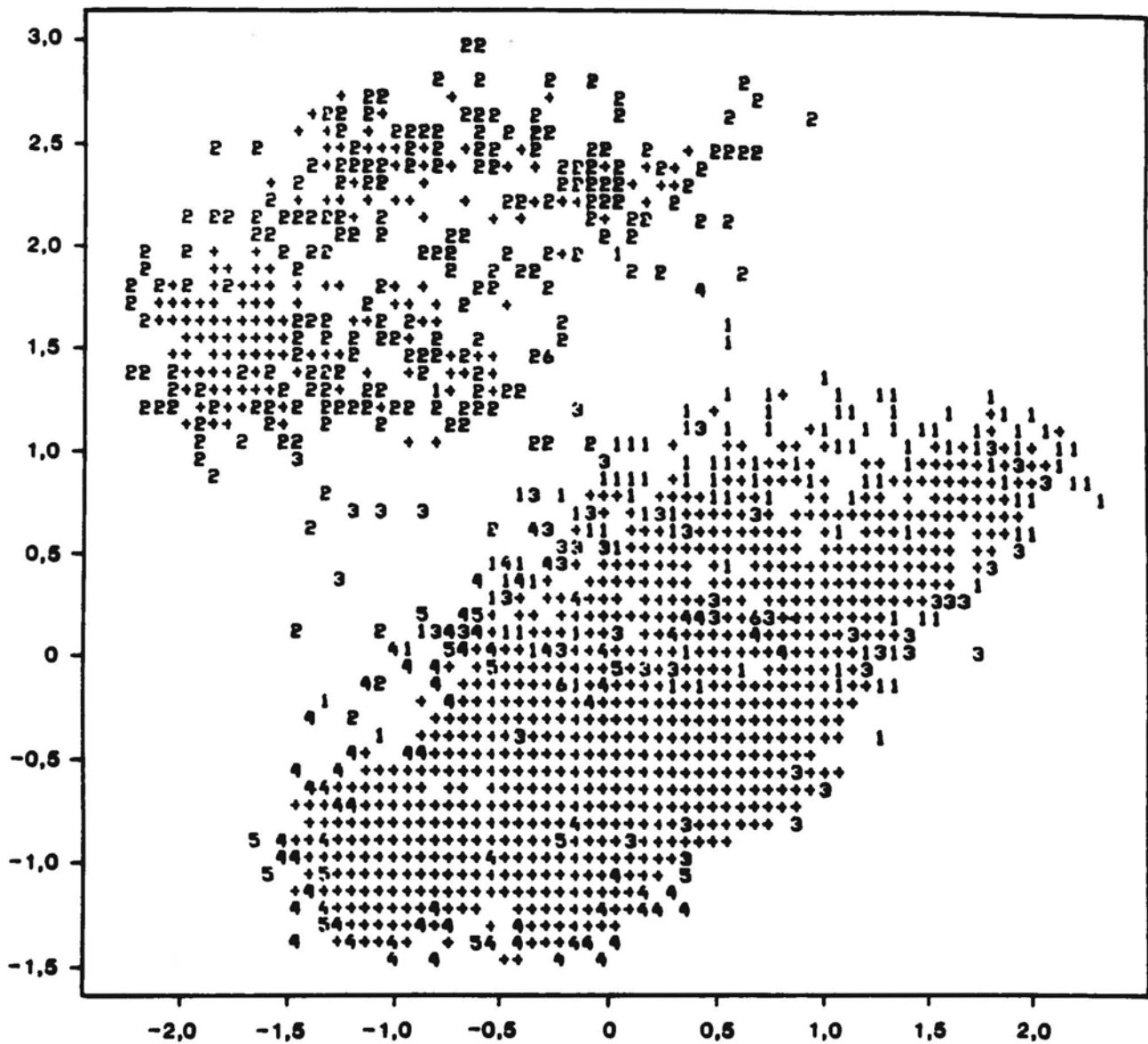
Voor kenmerken die goed discrimineren blijken de objecten uit een zelfde categorie ook dicht bij elkaar te liggen en de categorieën ver uit elkaar. De plussen in de afbeelding geven aan dat er meer objecten op één punt worden afgebeeld. Voor Afbeelding 1.2 geldt dat de plussen links boven ook praktische alle een waarde 2 hebben. Een zelfde constatering geldt ook voor de plussen in het 1-gebied, het 3-gebied en het 4, 5-gebied. De categorieën 4 en 5 onderscheiden zich wel van de andere categorieën, maar verschillen veel minder van elkaar.

Uit Afbeelding 1.2 blijkt duidelijk de discriminatie tussen voetgangersongevallen (code 2) en ongevallen met uitsluitend rijdend verkeer (code 1, 3, 4 en 5). De afbeelding laat zien dat beide groepen ongevallen het best afzonderlijk zijn te bestuderen. Dit kenmerk dient dan ook voor de hoofdonderscheiding te worden gebruikt. Beide groepen zijn dan weer het best te onderscheiden naar de sterk samenhangende kenmerken 4 en 5 (bebouwing en snelheidslimiet). Kenmerken 4 en 5 discrimineren vooral op dimensie 1 en ook wel op dimensie 3, maar niet op dimensie 2.

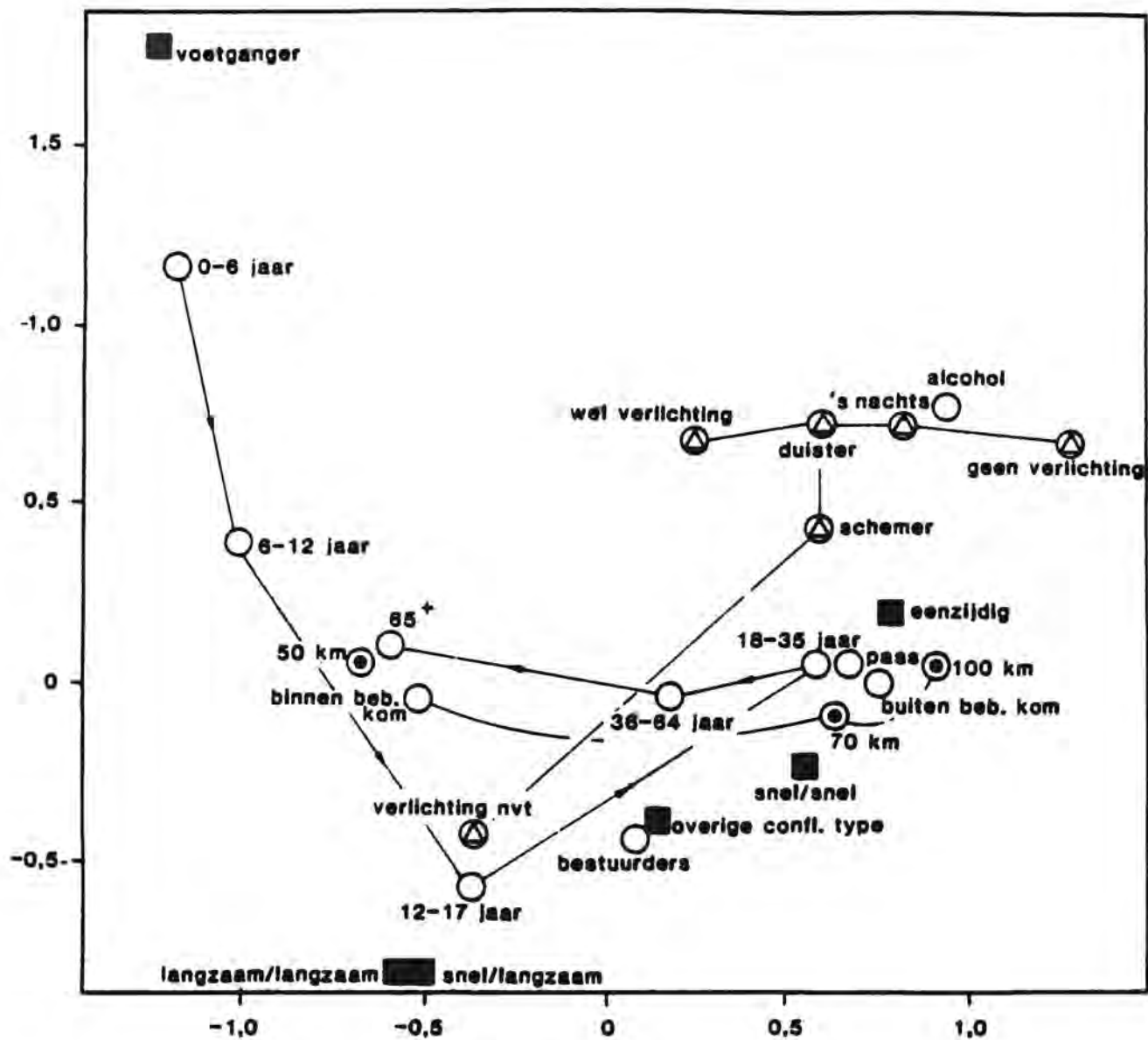
In Afbeelding 1.3 is een beeld te vinden van categorieën die bij elkaar horen. De punten zijn afgebeeld in dezelfde ruimte als bij Afbeelding 1.2. Ze geven aan waar de gemiddelden van alle punten uit een zelfde klasse liggen. Afbeelding 1.3 heeft voor kenmerk 14 van Afbeelding 1.2 slechts vijf waarden: het gemiddelde voetgangerspunt (code 2) binnen de wolk



Afbeelding 1.1. Plot van de discriminatiematen voor de verschillende kenmerken uit Tabel 2 in de eerste twee oplossingsdimensies.



Afbeelding 1.2. Plot van de objectcores voor kenmerk 14: conflicttype, in de eerste twee oplossingsdimensies (code 1: eenzijdig; code 2: voetgangers; code 3: snel-snel; code 4: snel-langzaam; code 5: langzaam-langzaam).



Afbeelding 1.3. Plot van de categoriekwantificaties voor de belangrijkste kenmerken in de eerste twee oplossingsdimensies.

van punten die corresponderen met de voetgangersongevallen en het gemiddelde punt voor de andere categorieën 1, 3, 4 en 5 in de andere puntenwolk. Afbeelding 1.3 is dus een combinatie van gemiddelden voor verschillende kenmerken. Punten die klassen van diverse kenmerken weergeven die dicht bij elkaar liggen corresponderen met veel objecten die beide klassekenmerken bezitten. Het beeld wordt echter vertekent door de indeling in de twee hoofdgroepen. Sommige punten zijn daardoor eerder gemiddelden van twee zwartepunten. Dit geldt vooral voor kenmerk 4 en 5.

Duidelijk is te zien hoe de leeftijdscategorieën telkens verschuiven van de voetgangersproblematiek, naar de fiets en bromfiets en vervolgens via het snelverkeer weer in de richting van het langzame verkeer. Toch discrimineert leeftijd niet erg sterk zoals blijkt uit Afbeelding 1. Dit kan voor deelproblemen natuurlijk wel het geval zijn. De nachtelijke problematiek wordt zichtbaar rechts boven in Afbeelding 1.3. Deze problematiek is relatief belangrijk. De punten in het centrum van de plot zijn, zoals te verwachten was, de categorieën van minder belangrijke variabelen.

Het specifieke van dimensie 3 hangt zoals gezegd samen met dezelfde kenmerken als in dimensie 1 en dimensie 2 aan de orde kwamen.

De kenmerken zijn in volgorde van belangrijkheid:

8. verlichting (vooral: n.v.t. versus brandend)

9. lichtgesteldheid (vooral: dag versus duister)

14. conflicttype (voetganger + snel/snel versus een zijding + snel/langzaam + langzaam/langzaam)

19. wijze verkeersdeelname (voetganger + personen + vrachtauto versus motor, fiets, bromfiets)

5. snelheidslimiet (snel versus langzaam)

3. tijdstip (ochtend versus middag versus nacht)

4. bebouwing (buiten versus binnen bebouwde kom)

Ook hier lijkt een consistente richting aangegeven te worden.

Uitsplitsing in de twee groepen: voetgangersongevallen en ongevallen met rijdend verkeer kan het beeld nog verduidelijken. Een nadere analyse van de groepen apart lijkt een volgende stap in de beschrijving van de samenhang in problematiek.

Bijlage 2. Voorbeeld van een PRINCALS- en een CANALS-analyse

Dit voorbeeld is overgenomen uit Oppe (1992). Het betreft een analyse die is toegepast op gegevens van 78 vier-armige rotondes in Engeland. De gegevens zijn per arm geanalyseerd. Tabel 2.1 geeft een overzicht van de kenmerken. Allereerst is een PRINCALS-analyse uitgevoerd om de onderlinge relaties tussen alle 25 kenmerken te onderzoeken. Daarna is een CANALS-analyse uitgevoerd, waarbij de eerste 20 kenmerken zijn daarbij gebruikt om de laatste vijf ongevalskenmerken te voorspellen. Alle kenmerken zijn ordinaal geschaald. In de PRINCALS-analyse wordt geen onderscheid gemaakt tussen de kenmerken van de rotonde en de ongevalskenmerken. Dit is bij CANALS dus wel het geval.

Afbeelding 2.1 geeft de beste PRINCALS-beschrijving in twee dimensies weer. Te zien is dat de ongevalskenmerken A1 t/m A4) duidelijk discrimineren tussen rotondes (de punten liggen ver van het 0-punt af op beide dimensies). Veiligheid is dus duidelijk een belangrijk discriminerend kenmerk en een ongevalsanalyse heeft dus zin. Het ongevalskenmerk A1 (ongevallen tussen voertuigen die de rotonde oprijden met voertuigen op de rotonde) zich onderscheidt van de andere soorten ongevallen (A2 t/m A5), die veel meer op elkaar lijken. A1 hangt vooral samen met kenmerken E,T6,NDA,CID,ICD,QPED en JP. A2 t/m/ A4 vooral met QX, QME, QE, QMC,V,VR,QC en G.

Vooraf de verkeerskenmerken QPE,QMC en QPC hebben met alle ongevallen te maken.

Afbeelding 2.2 geeft een vergelijkbaar beeld van de CANALS-analyse. Hier zijn de ongevalskenmerken belangrijker. Dit komt omdat een CANALS-oplossing de samenhang tussen de ongevalskenmerken en de andere kenmerken van de rotonde maximaliseert. Het beeld is in grote lijnen overeenkomstig. Ook hier weer blijkt A1 te verschillen van de andere ongevalskenmerken.

Op basis van deze analyse-uitkomsten is een vervolganalyse toegepast op A1. Om de stabiliteit van de analyse te toetsen zijn ook aparte analyses uitgevoerd op de even en oneven takken van de rotondes. Uit Afbeelding 2.3 is af te lezen welke kenmerken het belangrijkste zijn voor de verklaring van A1. Verder blijken de resultaten van de deelanalyses in grote lijnen overeen te komen met die van de analyse op de totale groep.

Tabel 2.2 geeft een overzicht van de veiligste en onveiligste armen, zoals voorspeld uit de kenmerken van de rotonde, samen met de belangrijkste kenmerken. De laatste kolom geeft de voorspelde veiligheidsscore; kolom 12 en 13 geven de geobserveerde aantallen ongevallen weer. Uit de tabel kan worden afgeleid dat de arm met de ongunstigste score (2.51) vrijwel uitsluitend ongunstige kenmerken heeft. Dit biedt tegelijk aanknopingspunten voor maatregelen.

- JN Junction Number
- AN Arm Number
- Var 1: JP Junction Period in month (max. 72)
- Var 2: T6 Roundabout category

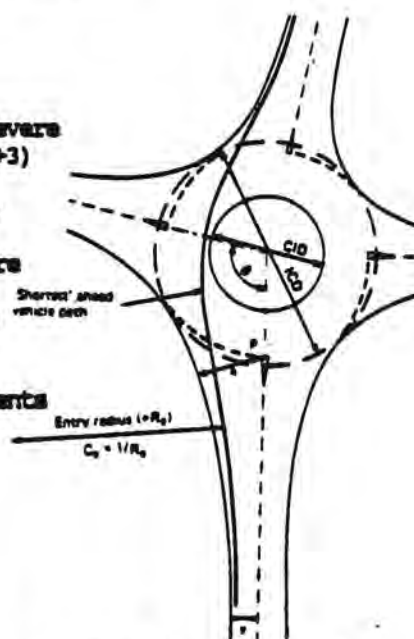
| | 30/40 mph | 50/70 mph |
|------------------|-----------|-----------|
| Small | cat. 1 | cat. 2 |
| Conventional | cat. 3 | cat. 4 |
| Dual carriageway | cat. 5 | cat. 6 |

The following 7 flow variables are measured in average 24 hour flows:

- Var 3: QPE Pedal cycle flow (entering)
- Var 4: QME Motor cycle flow (entering)
- Var 5: QE Total vehicle flow (entering)
- Var 6: QPC Pedal cycle flow (circulating)
- Var 7: QMC Motor cycle flow (circulating)
- Var 8: QC Total vehicle flow (circulating)
- Var 9: QX Total vehicle flow (exiting)
- Var 10: QPED Pedestrian crossing flow (measured in 16 hour flows)
- Var 11: CE Entry Curvature
- Var 12: E Entry width
- Var 13: V Approach width
- Var 14: ANG Angle between arms
- Var 15: G Gradient category, from severe downhill (-3) to uphill (+3)
- Var 16: VR Visibility to the right
- Var 17: ICD Inscribed Circle Diameter
- Var 18: CID Central Island Diameter
- Var 19: AC Approach Curve (just before roundabout)
- Var 20: NDA Nearside Deflection Angle

Accident types:

- Var 21: A1 Entering/circulating accidents
- Var 22: A2 Approaching accidents
- Var 23: A3 Single vehicle accidents
- Var 24: A4 Other accidents
- Var 25: A5 Pedestrian accidents



variables: C_e = var 11, e = var 12,
 v = var 13, θ = var 14,
 ICD, CID = var 17, 18

Tabel 2.1. Kenmerken van de (Engelse) rotondes en hun armen, zoals gebruikt in de PRINCALS- en CANALS-analyses.

LISTING OF MOST IMPORTANT VARIABLES AND DATA FOR SAFE AND UNSAFE ROUNDABOUT ARMS, SORTED ON THE OBJECT SCORES:

- COL 1: JUNCTION NUMBER (JN)
 COL 2: ARM NUMBER (AN)
 COL 3: ROUNDABOUT CATEGORY (TC)
 COL 4: MOTOR CYCLE FLOW (MC)
 COL 5: PEDESTRIAN CROSSING FLOW (PCF)
 COL 6: ENTRY CURVATURE (CE)
 COL 7: ENTRY WIDTH (E)
 COL 8: GRADIENT CATEGORY (G)
 COL 9: CENTRAL ISLAND DIAMETER (CID)
 COL 10: NEAR SIDE DEFLECTION ANGLE (NSDA)
 COL 11: ENTERING/CIRCULATING ACCIDENT CLASS (A1), CORRECTED FOR PERIOD (JP)
 COL 12: ENTERING/CIRCULATING ACCIDENTS (A1), CORRECTED FOR PERIOD (JP)
 COL 13: ENTERING/CIRCULATING ACCIDENTS (A1), NOT CORRECTED FOR PERIOD (JP)
 COL 14: PREDICTED NUMBER OF ACCIDENTS (A1), DIRECT METHOD
 COL 15: PREDICTED NUMBER OF ACCIDENTS (A1), INDIRECT METHOD
 COL 16: OBJECT SCORES

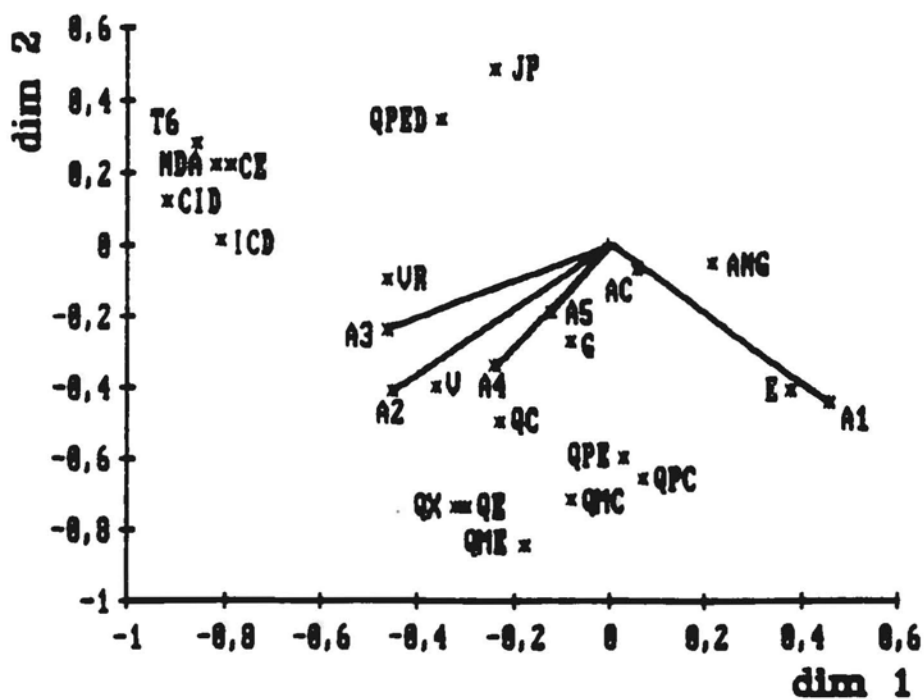
safe roundabout arms:

| | | | | | | | | | | | | | | | |
|-----|---|---|---|---|---|---|---|---|---|---|---|---|-------|------|-------|
| 158 | 1 | 5 | 3 | 4 | 6 | 2 | 1 | 6 | 6 | 2 | 2 | 1 | -4.14 | 0.21 | -1.85 |
| 178 | 3 | 3 | 4 | 4 | 6 | 2 | 5 | 5 | 5 | 1 | 1 | 0 | -2.68 | 0.32 | -1.44 |
| 2 | 4 | 5 | 4 | 1 | 7 | 2 | 4 | 8 | 6 | 1 | 1 | 0 | -1.84 | 0.40 | -1.20 |
| 210 | 4 | 5 | 2 | 3 | 5 | 1 | 2 | 4 | 5 | 2 | 2 | 1 | -1.39 | 0.45 | -1.07 |
| 321 | 4 | 3 | 6 | 2 | 4 | 2 | 4 | 6 | 5 | 2 | 2 | 1 | -1.26 | 0.47 | -1.04 |
| 166 | 3 | 3 | 4 | 2 | 6 | 2 | 4 | 5 | 6 | 1 | 1 | 0 | -1.21 | 0.48 | -1.02 |
| 346 | 4 | 1 | 4 | 6 | 5 | 2 | 5 | 2 | 4 | 1 | 1 | 0 | -1.19 | 0.48 | -1.02 |
| 178 | 1 | 3 | 3 | 5 | 6 | 2 | 4 | 5 | 6 | 1 | 1 | 0 | -1.16 | 0.48 | -1.01 |
| 158 | 2 | 5 | 3 | 5 | 5 | 2 | 3 | 6 | 5 | 2 | 2 | 1 | -1.13 | 0.49 | -1.00 |
| 166 | 1 | 3 | 3 | 3 | 6 | 2 | 4 | 5 | 6 | 1 | 1 | 0 | -1.04 | 0.50 | -0.98 |
| 22 | 4 | 3 | 6 | 4 | 4 | 2 | 4 | 4 | 3 | 1 | 1 | 0 | -1.04 | 0.50 | -0.98 |
| 293 | 4 | 1 | 5 | 2 | 6 | 3 | 1 | 3 | 5 | 1 | 1 | 0 | -0.98 | 0.51 | -0.96 |
| 210 | 2 | 5 | 2 | 2 | 5 | 1 | 4 | 4 | 4 | 2 | 2 | 1 | -0.96 | 0.51 | -0.95 |
| 85 | 2 | 4 | 1 | 9 | 7 | 2 | 4 | 6 | 6 | 1 | 1 | 0 | -0.94 | 0.52 | -0.95 |
| 81 | 1 | 4 | 2 | 9 | 7 | 2 | 5 | 6 | 6 | 1 | 1 | 0 | -0.84 | 0.53 | -0.92 |
| 82 | 2 | 6 | 2 | 9 | 7 | 2 | 4 | 4 | 5 | 1 | 1 | 0 | -0.81 | 0.53 | -0.91 |
| 395 | 4 | 5 | 8 | 4 | 4 | 3 | 2 | 5 | 6 | 1 | 1 | 0 | -0.79 | 0.54 | -0.91 |
| 2 | 3 | 5 | 6 | 3 | 6 | 2 | 4 | 8 | 5 | 2 | 2 | 1 | -0.81 | 0.54 | -0.91 |
| 85 | 4 | 4 | 1 | 9 | 7 | 2 | 3 | 6 | 6 | 1 | 1 | 0 | -0.79 | 0.54 | -0.90 |
| 59 | 4 | 6 | 3 | 9 | 7 | 2 | 6 | 5 | 2 | 1 | 1 | 0 | -0.71 | 0.55 | -0.88 |
| 350 | 2 | 2 | 2 | 9 | 3 | 2 | 2 | 1 | 2 | 1 | 1 | 0 | -0.69 | 0.55 | -0.88 |

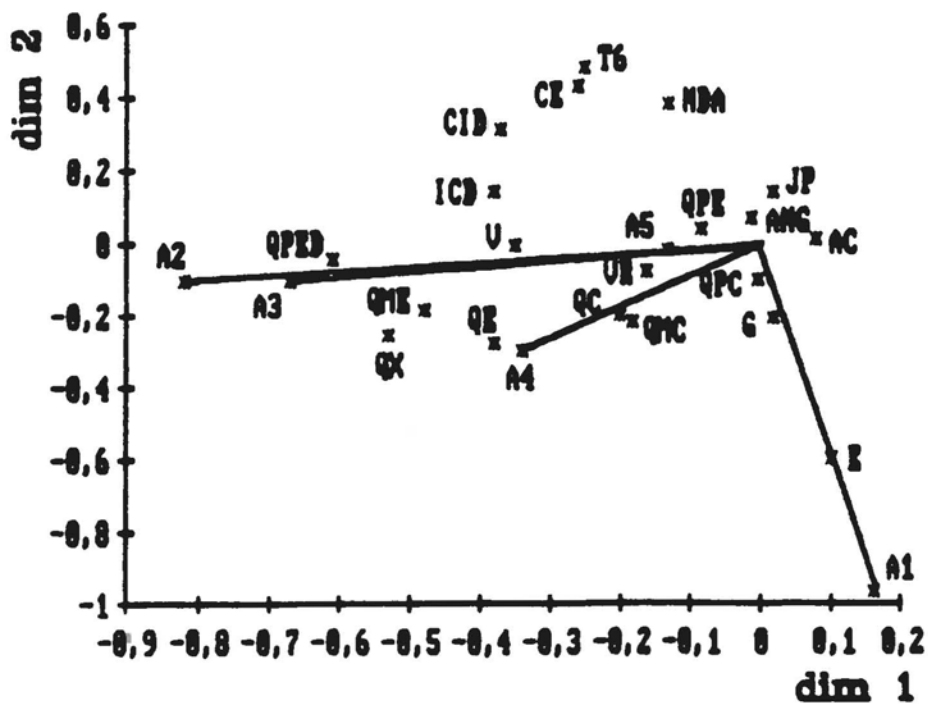
unsafe roundabout arms:

| | | | | | | | | | | | | | | | |
|-----|---|---|---|---|---|---|---|---|---|----|----|----|-------|-------|------|
| 190 | 2 | 1 | 8 | 9 | 3 | 5 | 5 | 2 | 3 | 7 | 5 | 6 | 9.00 | 8.49 | 1.85 |
| 31 | 1 | 4 | 5 | 9 | 2 | 5 | 5 | 6 | 4 | 7 | 7 | 6 | 9.01 | 8.50 | 1.85 |
| 161 | 4 | 1 | 4 | 3 | 2 | 5 | 4 | 3 | 2 | 9 | 10 | 9 | 9.02 | 8.53 | 1.86 |
| 292 | 2 | 1 | 6 | 9 | 4 | 5 | 7 | 5 | 5 | 8 | 7 | 8 | 9.13 | 8.81 | 1.89 |
| 161 | 3 | 1 | 4 | 5 | 2 | 5 | 5 | 3 | 1 | 7 | 7 | 6 | 9.20 | 8.97 | 1.91 |
| 161 | 2 | 1 | 4 | 4 | 1 | 5 | 5 | 3 | 1 | 8 | 9 | 8 | 9.40 | 9.50 | 1.97 |
| 21 | 1 | 1 | 7 | 1 | 2 | 4 | 4 | 2 | 2 | 7 | 6 | 6 | 9.55 | 9.90 | 2.01 |
| 21 | 2 | 1 | 6 | 1 | 3 | 4 | 6 | 2 | 3 | 11 | 14 | 16 | 9.67 | 10.24 | 2.04 |
| 392 | 2 | 1 | 6 | 1 | 3 | 4 | 6 | 2 | 3 | 10 | 3 | 12 | 9.68 | 10.27 | 2.04 |
| 175 | 3 | 1 | 3 | 2 | 1 | 5 | 4 | 1 | 1 | 8 | 9 | 8 | 10.11 | 11.62 | 2.17 |
| 21 | 3 | 1 | 8 | 2 | 2 | 4 | 6 | 2 | 2 | 8 | 8 | 8 | 10.16 | 11.76 | 2.18 |
| 175 | 2 | 1 | 5 | 2 | 1 | 4 | 4 | 1 | 1 | 9 | 10 | 9 | 10.45 | 12.79 | 2.26 |
| 268 | 3 | 2 | 6 | 9 | 1 | 4 | 4 | 2 | 2 | 9 | 10 | 9 | 10.81 | 14.12 | 2.36 |
| 76 | 1 | 2 | 4 | 9 | 3 | 5 | 7 | 3 | 1 | 9 | 10 | 10 | 10.96 | 14.72 | 2.40 |
| 175 | 1 | 1 | 5 | 2 | 1 | 5 | 4 | 1 | 1 | 11 | 16 | 15 | 10.97 | 14.78 | 2.41 |
| 115 | 4 | 2 | 3 | 9 | 3 | 4 | 4 | 4 | 4 | 10 | 7 | 12 | 10.97 | 14.80 | 2.41 |
| 114 | 4 | 2 | 4 | 9 | 3 | 5 | 4 | 4 | 3 | 9 | 7 | 10 | 11.01 | 14.96 | 2.42 |
| 180 | 3 | 1 | 7 | 9 | 4 | 5 | 3 | 2 | 2 | 9 | 7 | 9 | 11.06 | 15.17 | 2.43 |
| 268 | 1 | 2 | 3 | 9 | 1 | 3 | 3 | 2 | 1 | 10 | 12 | 11 | 11.31 | 16.29 | 2.50 |
| 190 | 1 | 1 | 6 | 9 | 4 | 5 | 4 | 2 | 2 | 10 | 9 | 12 | 11.34 | 16.40 | 2.51 |

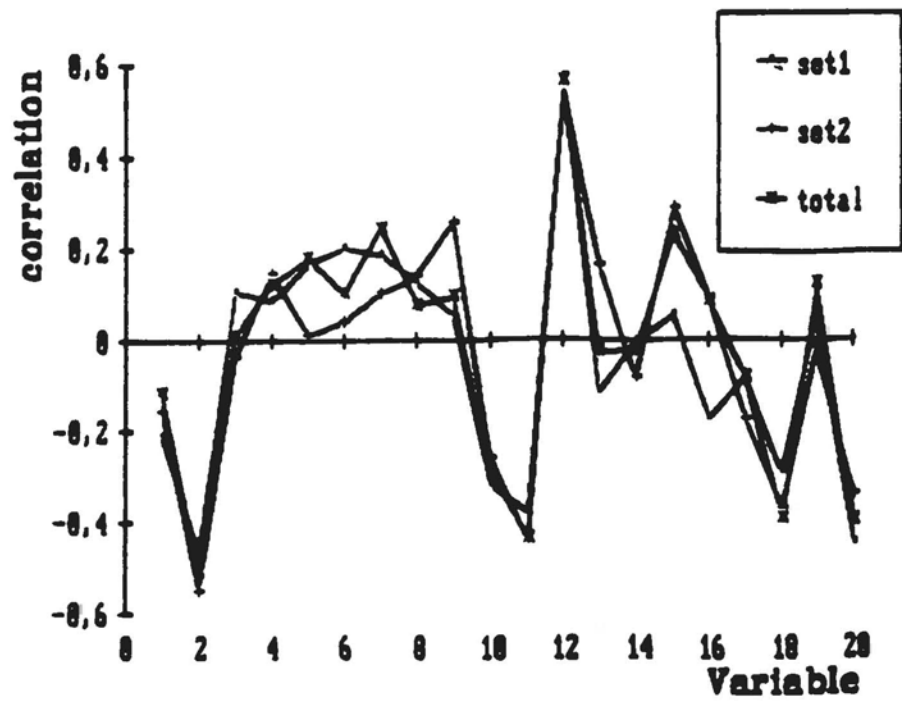
Tabel 2.2. Weergave van de veiligste en onveiligste armen van de rotondes, geordend op de onveiligheidsscore, samen met de belangrijkste kenmerken.



Afbeelding 2.1. Plot van de componentladingen in de twee-dimensionale PRINCALS-oplossing.



Afbeelding 2.2. Plot van de componentenladingen in de twee-dimensionale CANALS-oplossing.



Afbeelding 2.3. Correlaties tussen de verklarende variabelen en het ongevalskenmerk A1, uit de CANALS-analyses van de totale set armen van de rotondes en twee deelsets.

Bijlage 3. Voorbeeld van een log-lineaire analyse met behulp van WPM

Als voorbeeld van een log-lineaire analyse is gekozen voor een drie-weg-tabel. Hierin zijn voor twee jaren bloed/alcoholgegevens uitgesplitst naar geslacht (zie Tabel 3.1). De gegevens zijn afkomstig uit het SWOV-onderzoek naar Rij- en drinkgewoonten (SWOV, 1978).

De aantallen in de tabel betreffen bestuurders van personenauto's die in het najaar van 1975 en 1977 tijdens weekeindnachten van 10 uur 's avonds tot 4 uur 's nachts zijn onderzocht op hun bloedalcoholgehalte (BAG). In Tabel 3.1 zijn de aantallen weergegeven, in Tabel 3.2 de gewichtsfactoren.

| BAG | 1975 | | 1977 | |
|---------|------|-----|------|-----|
| | m | v | m | v |
| <0,2 | 2275 | 448 | 1838 | 452 |
| 0,2-0,5 | 339 | 33 | 350 | 38 |
| 0,5-1,0 | 263 | 11 | 247 | 20 |
| >1,0 | 163 | 10 | 145 | 9 |

Tabel 3.1. Aantal onderzochte bestuurders, uitgesplitst naar geslacht en BAG, voor de jaren 1975 en 1977.

| BAG | 1975 | | 1977 | |
|---------|------|------|------|------|
| | m | v | m | v |
| <0,2 | .275 | .265 | .236 | .233 |
| 0,2-0,5 | .268 | .199 | .251 | .280 |
| 0,5-1,0 | .317 | .229 | .286 | .273 |
| >1,0 | .372 | .556 | .291 | .425 |

Tabel 3.2. Gewichten afgeleid uit de percentages onderzochte bestuurders ten opzichte van het aantal passerende bestuurders.

De gewichten zorgen in dit geval voor een correctie van de aantallen op grond van het percentage van de weggebruikers dat in de steekproef is opgenomen. De samenhang tussen de gewichten en de BAG-klassen is een gevolg van het feit dat in de laat-nachtelijke uren een hoger alcoholgebruik wordt geconstateerd, terwijl dan tevens bijna alle weggebruikers die de onderzoekplaats passeren kunnen worden onderzocht. Voor de vroegere uren is het percentage onderzochten geringer.

Door de aantallen te delen door de gewichten worden gecorrigeerde aantallen verkregen waarop de analyse dient te worden toegepast. In WPM worden de parameters van het log-lineaire model berekend voor deze gewogen aantallen. Van deze parameters, die asymptotisch normaal verdeeld zijn met gemiddelde 0, is tevens een standaardscore berekend die getoetst kan worden tegen de gebruikelijke grenswaarden die gelden voor

de standaard normaal verdeling: $\pm 1,96$ voor het 5% significantieniveau en $\pm 2,58$ voor het 1% niveau. Naast deze parameter zijn chi-kwadraatwaarden berekend voor groepen van parameters overeenkomstig de mogelijke bronnen van interactie. In Tabel 3.3 zijn de diverse interactiebronnen gegeven, met hun chi-kwadraatwaarden en vrijheidsgraden. Tevens zijn de parameters en bijbehorende standaardscores opgenomen.

| Interactiebron | Chi-kwadraat | df | Contrast | Parameter | Standaard score |
|-----------------------|--------------|----|----------|-----------|-----------------|
| jaar x geslacht | 0,21 | 1 | 1 1 0 | 0,08 | 0,46 |
| jaar x BAG | 1,70 | 3 | 1 0 1 | 0,06 | 0,58 |
| | | | 1 0 2 | 0,19 | 1,22 |
| | | | 1 0 3 | -0,05 | -0,24 |
| geslacht x BAG | 102,71 | 3 | 0 1 1 | -1,00 | -10,10 |
| | | | 0 1 2 | -0,58 | -3,65 |
| | | | 0 1 3 | -0,36 | -1,69 |
| jaar x geslacht x BAG | 4,04 | 3 | 1 1 1 | 0,07 | 0,72 |
| | | | 1 1 2 | -0,20 | -1,28 |
| | | | 1 1 3 | 0,12 | 0,55 |

Tabel 3.3. *Analyse van rij- en drinkgewoontegegevens, uitgesplitst naar jaar, geslacht en bloedalcoholgehalte (BAG).*

Bij de interpretatie van de chi-kwadraatwaarden wordt eerst gekeken naar de hoogste-orde-interactie. Indien deze significant is, dan zijn de andere effecten niet goed meer te interpreteren, omdat ze door het hogere-orde-effect beïnvloed worden.

We zien dan de chi-kwadraatwaarde van de interactie voor jaar x geslacht x BAG niet significant is ($X^2 = 4,04$, $df = 3$). De verhouding tussen het drinken van mannen en vrouwen is dus niet veranderd in 1977 ten opzichte van 1975. Indien deze interactie wel significant zou zijn geweest, dan zou een verdere analyse van de relatie tussen BAG en geslacht apart voor 1975 en 1977 dienen te geschieden, of van de relatie tussen BAG en jaar apart voor mannen en vrouwen. Hier is dit niet nodig. Er is ook geen significant verschil te constateren in drinken tussen 1975 en 1977 ($X^2 = 1,70$). Wel vinden we een zeer significant verschil tussen het BAG van mannen en vrouwen over de beide jaren gezamenlijk bekeken ($X^2 = 102,71$). De rij- en drinkgewoonten van mannen verschillen zeer significant van die van vrouwen.

Indien we meer informatie wensen over de bijdrage van de verschillende klassen aan de interactie-effecten, dan is dit voor jaar en geslacht direct duidelijk. Er is slechts een contrast: klasse 1 vs klasse 2. Het effect wordt dus veroorzaakt door het verschil tussen de twee klassen.

Voor het BAG is dit niet zo duidelijk. We kunnen ons afvragen of het effect te maken heeft met het drinken versus het niet-drinken, of eerder is toe te schrijven aan het meer gaan drinken van de drinkers, of misschien aan beide factoren. Om dit te onderzoeken kunnen we de volgende contrasten tussen de BAG-klassen kiezen:

contrast 1: klasse 1 vs klasse 2, 3 en 4 (niet drinken vs drinken);

contrast 2: klasse 2 vs klasse 3 en 4 (weinig drinken vs matig of veel drinken);

contrast 3: klasse 3 vs klasse 4 (matig vs veel drinken).

Deze contrasten zijn onafhankelijk van elkaar te interpreteren en leveren alle informatie ten aanzien van het interactie-effect.

Voor BAG zijn ook andere contrasten mogelijk, bijvoorbeeld:

contrast 4: klasse 1 en 2 vs klasse 3 en 4 (geoorloofd vs ongeoorloofd BAG);

contrast 5: klasse 1 vs klasse 2;

contrast 6: klasse 3 vs klasse 4.

Ook deze contrasten zijn onafhankelijk van elkaar, maar niet van de eerder genoemde contrasten. Ze leveren eveneens alle informatie. De keuze van de contrasten wordt dus vooraf bepaald door de onderzoeker afhankelijk van zijn interesse. Voor de eerstgenoemde contrasten zijn de parameters berekend en vervolgens uitgedrukt in standaardcores.

Voor het interactie-effect geslacht x BAG betreft dit bijvoorbeeld $1 \times 3 + 3$ parameters. Één parameter voor elke combinatie van contrasten. Ook deze zijn gegeven in Tabel 3.3.

We zien dat de grootste parameter wordt gevonden voor combinatie 0 1 1 ($z = -10,10$), daarna voor 0 1 2 ($z = -3,65$), tenslotte voor 0 1 3 ($z = -1,69$). Het teken geeft de richting van het effect aan: mannen drinken vaker, meer en mogelijk zwaarder dan vrouwen.

Uit deze parameters kunnen we tevens parameters afleiden voor de klassen zelf. We vinden dan 2×4 klasseparameters. Deze zijn echter wel afhankelijk van elkaar. De parameters voor de BAG-categorieën van mannen en vrouwen hebben een tegengesteld teken. Verder ligt de laatste parameter vast met de eerste drie parameters: de parameters tellen op tot nul.

Tenslotte kunnen we nog zoeken naar de meest efficiënt beschrijvende model voor de gegevens. Dit blijkt uit bovenstaande gegevens het model te zijn waarin de interacties jaar x geslacht x BAG, jaar x BAG en jaar x geslacht afwezig worden verondersteld. In formulevorm:

$$m_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_{jk}$$

Voor dit model wordt een niet-significante chi-kwadraatwaarde gevonden van 13,10 ($df = 7$).