

MDS van drieweg/tweemodale data

9.1 MDS VAN VIERKANTE DRIEWEGDATA

In het vorige hoofdstuk is een variant van het klassieke CMDS-probleem besproken, namelijk MDS van één asymmetrische nabijheidsmatrix. Van een andere variant is er sprake als er niet één, maar verschillende vierkante, symmetrische nabijheidsmatrices geobserveerd zijn. We hebben dan een zogenaamde *drieweg/tweemodale matrix* of – afhankelijk van de interpretatie – een *tweeweg/éénmodale matrix met replicaties*. Zo'n matrix kan het beste voorgesteld worden als een cake die in plakjes is gesneden. Ieder plakje is een vierkante, symmetrische nabijheidsmatrix; de plakjes vormen samen de derde weg, de replicaties. Meestal bestaat de derde weg uit verschillende personen, verschillende condities, verschillende gelegenheden of verschillende tijdstippen. Als we de totale driewegmatrix aanduiden met O , dan vormt ieder plakje een matrix O_k waarbij $k = 1, \dots, n$ met n het aantal plakjes of replicaties. Iedere matrix O_k bevat observaties o_{ijk} die de nabijheid tussen object i en object j weergeven zoals die geobserveerd is in replicatie k . Dit soort data kunnen globaal op vier verschillende manieren geanalyseerd worden: na aggregatie, apart, als replicaties en met een individuele-verschillenmodel.

Analyse na aggregatie

Iedere drieweg/tweemodale datamatrix is opgebouwd uit een aantal afzonderlijke tweeweg/éénmodale nabijheidsmatrices. De simpelste aanpak om dit soort data te analyseren bestaat uit aggregatie: men maakt er toch maar weer één matrix van, bijvoorbeeld door de driewegmatrix te middelen over de derde weg, zodat $O^* = \sum_k O_k / n$. De nieuwe tweewegmatrix O^* wordt dan als standaard CMDS-probleem geanalyseerd. Het zal duidelijk zijn dat hiermee informatie verloren gaat: aan de oplossing kunnen we niet zien of deze een goede

overeenstemming heeft met de data uit de afzonderlijke matrices.

Deze aanpak is alleen dan zinvol, als we kunnen aannemen dat de nabijheidsrelaties van alle afzonderlijke matrices dezelfde onderliggende structuur hebben en dat de overeenkomstige elementen in de verschillende matrices alleen door toevallige fluctuaties van elkaar verschillen. Ook moeten de data onconditioneel zijn (dus niet matrix- of rijconditioneel) want anders zijn de getallen in de ene matrix niet te vergelijken met de getallen in een andere matrix.

Optellen is dan niet zinvol. Bovendien moeten de data minstens op intervalniveau gemeten zijn; het zonder meer middelen van ordinale data kan onzinnige resultaten opleveren. Beter is het om matrixconditionele en ordinale nabijheden te aggregeren door middel van een op de Thurstonemethode gebaseerde bewerking. Deze methode wordt in Blok 9.1 uitgelegd.

BLOK 9.1 HET AGGREGEREN VAN ORDINALE NABIJHEIDSGEGEVENS

Als voorbeeld nemen we twee matrices met *ordinale* gegevens over de onderlinge ongelijkheid (dissimilarities) van vier objecten, A, B, C en D. Deze gegevens zijn met vette cijfers weergegeven in de onderdiagonale gedeelten van de tabellen 9.1 a en b. De cursieve getallen boven de diagonaal zijn dezelfde gegevens na omzetting in *rangnummers* (daarover later).

Tabel 9.1a en b Matrices met ordinale gegevens over dissimilarities van vier objecten

	A	B	C	D		A	B	C	D
A		–	3	1	6	A	–	4	1
2.5									
B	2 3		2	4	B	3 4	–	2.5	5
C	1 0	1 2		5	C	1 1	2 3	–	6
D	4 6	3 4	4 5	–	D	2 3	4 5	5 6	–

Zowel uit de onderdiagonale als uit de bovendiagonale helften van deze tabellen kunnen we twee nieuwe matrices afleiden die slechts uit enen en nullen bestaan en die allebei evenveel rijen en kolommen hebben als er paren objecten zijn (in dit voorbeeld zes). Het getal 1 in cel (i, j) van zo'n matrix geeft aan dat het i -de objectenpaar het j -de objectenpaar qua afstand domineert. Bijvoorbeeld: onder de diagonaal van de linker tabel zien we dat $o_{AB} > o_{AC}$. In de afgeleide matrix moet er dus een 1 komen te staan in de cel die gevormd wordt door rij AB en kolom AC. In de cel met rij AC en kolom AB komt dan een 0. De bij het voorbeeld behorende dominantiematrices staan hieronder.

Tabel 9.2a en b Dominantiematrices, afgeleid van tabel 9.1a en b

	AB	AC	AD	BC	BD	CD	
AB	–	1	0	1	0	0	2
AC	0	–	0	0	0	0	0
AD	1	1	–	1	1	1	5
BC	0	1	0	–	0	0	1
BD	1	1	0	1	–	0	3
CD	1	1	0	1	1	–	4

	AB	AC	AD	BC	BD	CD	
AB	–	1	1	1	0	0	3
AC	0	–	0	0	0	0	0
AD	0	1	–	.5	0	0	1.5
BC	0	1	.5	–	0	0	1.5
BD	1	1	1	1	–	0	4
CD	1	1	1	1	1	–	5

Merk op dat de tweede nabijheidsmatrix twee *ties* bevat ($o_{AD} = o_{BC} = 23$). Daarom redeneren we dat o_{AD} in de helft van de gevallen groter is dan o_{BC} en dat o_{BC} in de andere helft van de gevallen groter is dan o_{AD} , zodat we in de bijbehorende dominantiematrix de cellen (BD, CD) en (CD, BD) beide de waarde .5 toekennen.

De rijtotalen van bovenstaande dominantiematrices geven weer hoe vaak het objectenpaar van de betreffende rij de andere paren qua afstand overtrof. Deze totalen zijn gelijk aan de rangnummers uit de bovendiagonale nabijheidsmatrices min 1. De rangorde van de rijtotalen is dus dezelfde als de rangorde van de nabijheden in de oorspronkelijke matrices.

Tellen we de enen en nullen van de verschillende dominantiematrices bij elkaar op, dan ontstaat een nieuwe matrix met frequenties die aangeven hoe vaak het ene paar een grotere afstand heeft dan het andere paar: zie hieronder.

Tabel 9.3 Matrix met de sommen van de elementen van tabel 9.2a + b

	AB	AC	AD	BC	BD	CD	
AB	–	2	1	2	0	0	5
AC	0	–	0	0	0	0	0
AD	1	2	–	1.5	1	1	6.5
BC	0	2	.5	–	0	0	2.5
BD	2	2	0	2	–	0	6
CD	2	2	1	2	2	0	9

De rijtotalen van deze matrix (die de som is van de rijtotalen van de afzonderlijke dominantiematrices) kunnen we dan gebruiken om een nieuwe (geaggregeerde) rangordening van de objectafstanden te bepalen. Merk op dat deze rijtotalen gelijk zijn aan de som min n (hier: 2) van de rangnummers uit de overeenkomstige cellen van de oorspronkelijke nabijheidsmatrices. In plaats van dominantiematrices te construeren en op te tellen, kunnen we net zo goed de data omzetten in rangnummers en die bij elkaar optellen en (eventueel) middelen.

Deze aanpak kunnen we nog uitbreiden door de frequenties uit de opgetelde dominantiematrix om te zetten in proporties en vervolgens in z -waarden.

Tellen we deze z -waarden over de kolommen op en delen we ze door het aantal objecten, dan hebben we gebruikgemaakt van het klassieke Thurstonemodel (zie Hoofdstuk 1) om geaggregeerde afstandsdata te berekenen. Dit is de rationale van Torgersons (1958) aanpak om afstands-schattingen af te leiden. We zouden nu kunnen concluderen (maar het hoeft niet!) dat we de afstanden op intervalniveau geschat hebben. Merk op dat sommige 'afstanden' negatief zijn. Daarom moet men ervoor zorgen (bijvoorbeeld via $CUTOFF = -1.0$) ook de negatieve getallen goed in te lezen of eerst allemaal positief te maken door er een getal bij op te tellen. Als de data op intervalniveau geanalyseerd worden, zal het MDS-algoritme eerst een additieve constante moeten berekenen. Bij analyse op ordinaal niveau is dat niet nodig.

De hierboven beschreven aggregatie van ordinale data wordt aangeraden als er getalsmatig grote verschillen tussen de matrices bestaan. In dat geval is het aannemelijk dat de data matrixconditioneel zijn. Dat wil zeggen: de functie die de relatie tussen de observaties in de ene matrix en de 'echte' afstanden beschrijft, is niet dezelfde als de functie die voor de tweede matrix geldt. Waarschijnlijk is er niet zoveel bezwaar tegen het middelen van de onderdiagonale waarden uit de matrices van ons voorbeeld. Voor deze twee matrices geldt dat de rangorde van de gemiddelde observaties

gelijk is aan het gemiddelde van de rangorden. Het zou ook voor kunnen komen dat de getallen in de ene matrix uit de onderdiagonale waarden bestaan, terwijl de getallen in de tweede matrix door de bovendiagonale waarden gevormd worden. In zo'n geval zijn de verschillen zo groot dat het geen zin heeft de twee verzamelingen getallen zomaar te middelen.

Analyse per matrix

In plaats van één CMDS-analyse van een geaggregeerde matrix kan men ook een CMDS-analyse van elke matrix afzonderlijk uitvoeren. Iedere analyse toont dan de individuele configuratie die bij elke individuele matrix hoort. Er gaat dus geen informatie verloren over eventuele individuele verschillen. Het grote nadeel van deze aanpak is echter dat men over het algemeen over veel afzonderlijke matrices beschikt. Deze matrices leveren dus ook veel afzonderlijke configuraties op, die allemaal geïnterpreteerd en met elkaar vergeleken moeten worden. Voor het vergelijken van de verschillende configuraties kan men weliswaar gebruikmaken van Procrustesanalyse (zie Hoofdstuk 7), maar het vergelijken van meerdimensionale ruimtelijke configuraties is uitermate gecompliceerd, zeker wanneer het om tientallen configuraties gaat.

Analyse als replicaties

Ook deze aanpak berust op de aanname dat de verschillende matrices onderling verwisselbaar zijn en dat de variatie tussen de matrices als *error*, als niet-interessante informatie beschouwd mag worden. Als we desondanks niet willen middelen, dan kunnen we een MDS-oplossing zoeken die zo goed mogelijk bij *alle matrices tegelijkertijd* past. Dit type analyse wordt RMDS genoemd (de *R* staat voor *replicated*). Er zijn drie opties voor zo'n soort analyse:

- 1 Behandel de data *onconditioneel*: alle data worden via dezelfde monotone transformatiefunctie getransformeerd. In dit geval zoeken we coördinaten $\{x_{is}\}$ waarvoor geldt dat de som van gekwadrateerde afwijkingen $\sum_k \sum_i \sum_j \{d_{ij} - f(o_{ijk})\}^2$ zo klein mogelijk is¹. Er kan bewezen worden dat de configuratie van de objecten in dit geval gelijk is aan de oplossing na middeling van de observaties. Echter, de stress is bij RMDS (veel) groter dan bij CMDS van gemiddelden omdat er per afstand d_{ij} nu n pseudo-afstanden $f(o_{ijk})$ zijn die allemaal van d_{ij} kunnen afwijken.
- 2 Behandel de data *matrixconditioneel*: de data van iedere matrix (proefpersoon-gelegenheid-conditie) worden getransformeerd via een eigen transformatiefunctie. In dat geval willen we $\sum_k \sum_i \sum_j \{d_{ij} - f_k(o_{ijk})\}^2$ met behulp van de (s)stressformule minimaliseren.

1 Zoals gezegd (zie Hoofdstuk 6) wordt niet zonder meer de som van de gekwadrateerde afwijkingen geminimaliseerd, maar een functie daarvan: *Stress*₁ of *S-stress*.

- 3 Behandel de data *rijconditioneel*: de data van elke afzonderlijke rij i van elke matrix k krijgen een eigen transformatiefunctie. In dit geval wordt dus $\sum_k \sum_i \sum_j \{d_{ij} - f_{ki}(o_{ijk})\}^2$ door middel van de (s)stressformule geminimaliseerd.

Analyse met een individuele-verschillenmodel

Terwijl RMDS, evenals CMDS na aggregatie, ervan uitgaat dat de afzonderlijke replicaties met elkaar verwisseld mogen worden, is er een andere aanpak waarin de afzonderlijke matrices niet zonder meer als replicaties beschouwd worden. Deze aanpak gaat ervan uit dat de variatie tussen de matrices geen toevallige error is, maar te maken heeft met verschillende structuren van nabijheidsrelaties in de verschillende matrices. Onder deze aanname gebruikt men dus liever een MDS-model dat speciale modelparameters bevat om de verschillen tussen de matrices te beschrijven. Als de derde weg uit personen bestaat kan men zeggen dat de variatie in de derde weg met individueel verschillende gezichtspunten correspondeert. Daarom noemt men een model voor dit soort data een *individuele-verschillenmodel*.

Het bekendste individuele-verschillenmodel is het zogenaamde INDSCAL-model (INDSCAL is een acroniem voor *individual differences scaling*) dat min of meer tegelijkertijd door Horan (1969) en door Carroll en Chang (1970) is voorgesteld. De publicatie van Carroll en Chang heeft veel meer aandacht gekregen, omdat daarin niet alleen het model maar ook een compleet (metrisch) algoritme, het zogenaamde INDSCAL-programma, beschreven werd. Dit individuele-verschillenmodel is gekenmerkt door de gewogen-afstandsfunctie die in de hoofdstukken 4 en 5 besproken is (zie Formule [4.11] en [5.16] tot en met [5.18]).

De analyse van drieweg/tweemodale matrices met behulp van het INDSCAL-model wordt, in navolging van Young (zie Schiffman e.a., 1981) *weighted multidimensional scaling* (WMDS) genoemd. Deze aanpak en het INDSCAL-model worden hieronder in detail besproken.

9.2 MDS MET HET INDSCAL-MODEL

De gewogen-afstandsfunctie waarop het INDSCAL-model berust, ziet er in matrixnotatie uit als

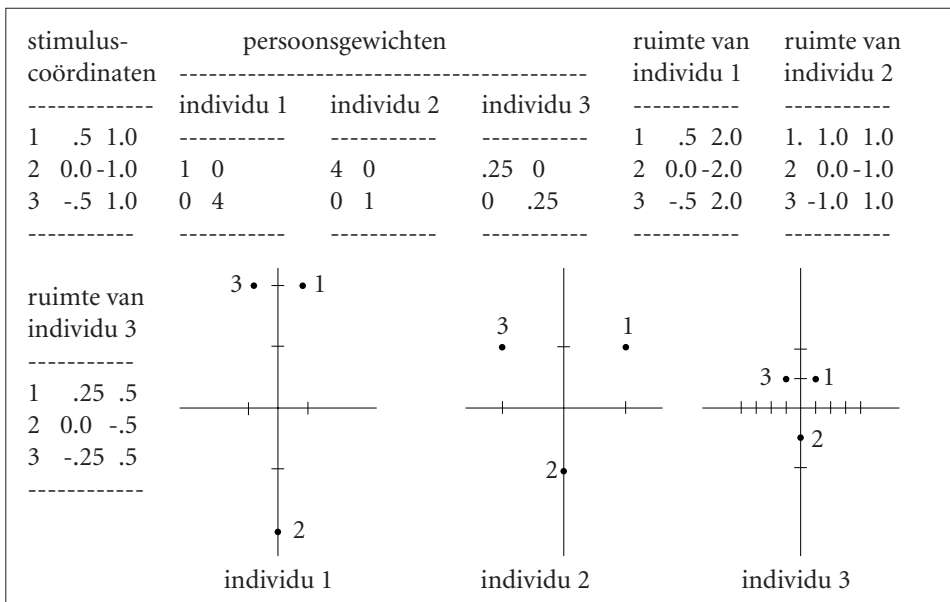
$$\begin{aligned} d_{ijk}^{*2} &= (\mathbf{x}'_i - \mathbf{x}'_j) W_k (\mathbf{x}'_i - \mathbf{x}'_j)' = (\mathbf{x}'_i - \mathbf{x}'_j) W_k^{1/2} W_k^{1/2} (\mathbf{x}'_i - \mathbf{x}'_j)' \\ &= \{(\mathbf{x}^{(k)}_i)' - (\mathbf{x}^{(k)}_j)'\} \{(\mathbf{x}^{(k)}_i)' - (\mathbf{x}^{(k)}_j)'\}' \end{aligned} \quad [9.1]$$

en in som-notatie als [9.2]

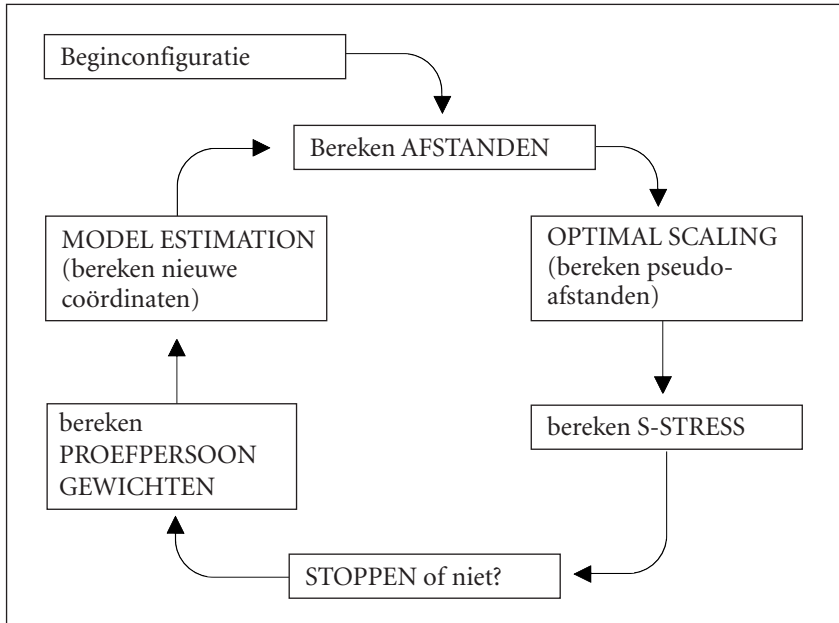
$$d_{ijk}^* = \sqrt{\sum_s w_{ks} (\mathbf{x}_{is} - \mathbf{x}_{js})^2} = \sqrt{\sum_s (w_{ks}^{1/2} \mathbf{x}_{is} - w_{ks}^{1/2} \mathbf{x}_{js})^2} = \sqrt{\sum_s (\mathbf{x}_{is}^{(k)} - \mathbf{x}_{js}^{(k)})^2}.$$

In dit model is er voor ieder 'plakje' van de driewegmatrix een afzonderlijke matrix W_k . Deze matrix is diagonaal, met elementen w_{ks} in de s -de rij en s -de kolom van W_k (eigenlijk zouden we moeten schrijven w_{kss} of $w_{ss}^{(k)}$, maar omdat W_k diagonaal is, is w_{ks} net zo duidelijk om het betreffende element aan te geven). De door Formules [9.1] en [9.2] weergegeven afstandsfunctie houdt in dat voor 'plakje' k alle coördinaten op dimensie s met het gewicht $\sqrt{w_{ks}}$ vermenigvuldigd (gewogen!) worden. Als er in totaal r dimensies zijn, dan zijn er r van zulke gewichten; daarvan kunnen sommige gelijk aan nul zijn. Merk op dat de gewichten w_{ks} nooit negatief mogen zijn, omdat $\sqrt{w_{ks}}$ dan geen reële waarde heeft.

In woorden luidt het INDSCAL-model als volgt: bij ieder plakje van de driewegmatrix, dus bij iedere individuele tweeweg/éénmodale nabijheidsmatrix O_k , hoort een eigen, individuele MDS-configuratie X_k die de coördinaten $x_{is}^{(k)}$ bevat. Zo'n individuele configuratie X_k is door middel van een niet-afstandsbehoudende transformatie afgeleid van een *gemeenschappelijke configuratie* X met coördinaten x_{is} (ook wel de *groeps-stimulusruimte* genoemd). De transformatie komt tot stand door navermenigvuldiging van X met $W_k^{1/2}$, die de zogenaamde (*proefpersoon*)gewichten van individu k bevat. De transformatie $XW_k^{1/2}$ houdt in dat de assen van de gemeenschappelijke configuratie selectief worden opgerekt of ingekrompen (door vermenigvuldiging met $\sqrt{w_{ks}}$). De afstand d_{ijk} wordt dus als het ware in een andere configuratie berekend dan de afstand d_{ijh} . Een voorbeeld van dit model wordt weergegeven in Figuur 9.1.



Figuur 9.1 Voorbeeld van een gewogen-afstandsmodel met proefpersoonsgewichten



Figuur 9.2 Het iteratieve MDS-algoritme, met een stap voor de berekening van proefpersoongewichten

Net als in ASCAL (zie Hoofdstuk 8) bestaat het MDS-probleem nu uit het vinden van twee soorten parameters: de coördinaten van de punten in de gemeenschappelijke ruimte en de gewichten die bij de individuele matrices horen. Om deze proefpersoongewichten te berekenen is in ALSCAL weer een extra stap in het iteratieve algoritme opgenomen. Deze stap komt na het transformeren van de observaties in de *optimal scaling*-stap en vóór de verbetering van de configuratie in de *model estimation*-stap (zie Figuur 9.2).

Het berekenen van de proefpersoongewichten

Om de proefpersoongewichten te bepalen wordt, net als bij het ASCAL-model (zie Hoofdstuk 8), gebruikgemaakt van multiële regressie. Dat gaat als volgt. Voor elke stimulus i geldt volgens Formule [9.2] dat

$$d_{ijk} = \sqrt{\sum_s [w_{ks}(x_{is} - x_{js})^2]} \text{ zodat } d_{ijk}^2 = \sum_s [w_{ks}(x_{is} - x_{js})^2].$$

Omdat in het ideale geval de afstanden d_{ijk} gelijk zijn aan de pseudo-afstanden $f(o_{ijk})$, vullen we deze in op de plaats van de d_{ijk} . In het ideale geval geldt dus – bij r dimensies – dat

$$\begin{aligned} (f(o_{ijk}))^2 &= \sum_{s=1}^r (w_{ks}(x_{is} - x_{js})^2) \\ &= w_{k1}(x_{i1} - x_{j1})^2 + w_{k2}(x_{i2} - x_{j2})^2 + \dots + w_{kr}(x_{ir} - x_{jr})^2. \end{aligned} \quad [9.3]$$

Voor ieder individu k zijn er $m(m-1)/2$ van bovenstaande vergelijkingen, zodat we hier te maken hebben met een multi-pele-regressieprobleem waarin we voor elke proefpersoon afzonderlijk de uit iteratie t bekende gekwadrateerde disparities $(f^{(t)}(o_{ijk}))^2$ willen ‘voorspellen’ uit de verschillen tussen de eveneens bekende coördinaten $x_{is}^{(t)}$. De bijbehorende regressiecoëfficiënten zijn dan de optimale kleinste-kwadratenschattingen van de proefpersoongewichten $w_{ks}^{(t)}$ van proefpersoon k .

Als er n proefpersonen zijn, dan moet het programma n keer zo’n multi-pele-regressieprobleem oplossen om voor alle proefpersonen de betreffende gewichten te berekenen. Zijn deze gewichten eenmaal gevonden, dan kunnen via Formule [9.2] weer nieuwe afstanden berekend worden. Daarmee begint het computerprogramma aan een nieuwe iteratie. Gebruikmakend van de aldus berekende gewichten en afstanden worden in de *model estimation*-stap weer nieuwe coördinaten voor de stimuli berekend. Vervolgens worden in de *optimal scaling*-stap weer nieuwe transformaties bepaald die – na berekening van *S-stress* en de beslissing om door te gaan – weer leiden tot verbetering van de proefpersoongewichten. Daarna worden weer nieuwe coördinaten berekend, enzovoort, net zolang totdat er een optimale oplossing is gevonden. Het iteratieve algoritme bestaat nu dus uit drie stappen, die, alternerend, drie kleinste-kwadratenproblemen oplossen:

- Stap 1 neem de laatste coördinaten en proefpersoon-gewichten en bepaal in de *optimal scaling*-stap de optimale transformaties van de data, de pseudo-afstanden of disparities.
- Stap 2 ga uit van de laatste coördinaten en pseudo-afstanden en bepaal via regressie de optimale proefpersoongewichten.
- Stap 3 ga uit van de laatste coördinaten, pseudo-afstanden en proefpersoongewichten en bepaal in de *model estimation*-stap de nieuwe, optimale coördinaten.

Conditionaliteit

Wat betreft de conditionaliteit van de driewegmatrix zijn er weer drie opties: *onconditioneel*, *matrixconditioneel* en *rijconditioneel*. Het verschil is dat men in het onconditionele geval $\sum_k \sum_i \sum_j \{d_{ij} - f(o_{ijk})\}^2$, in het matrixconditionele geval $\sum_k \sum_i \sum_j \{d_{ij} - f_k(o_{ijk})\}^2$ en in het rijconditionele geval $\sum_k \sum_i \sum_j \{d_{ij} - f_{ki}(o_{ijk})\}^2$ door middel van de (s)stressfunctie minimaliseert.

9.3 INTERPRETATIE VAN EEN INDSCAL-OPLOSSING

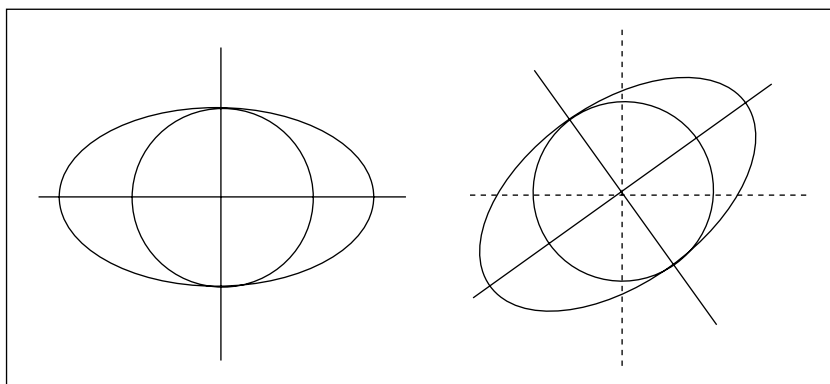
Een voorbeeld van een ALSCAL-analyse met het INDSCAL-model wordt in Blok 9.2 besproken, samen met een RMDS-analyse van dezelfde data. Voordat we de precieze aansturing en uitvoer van deze analyses gaan bekijken, moeten we eerst een aantal algemene aspecten van INDSCAL-oplossingen behandelen. Dat gebeurt hieronder.

De vorm van de configuratie

Meetkundig gezien komt het INDSCAL-model erop neer dat de dimensies van de groepsstimulusruimte op verschillende wijze worden opgerekt of ingekrompen. Een configuratie die in de groepsruimte cirkelvormig is, krijgt in de individuele ruimten de vorm van liggende of staande ellipsoïden, al naar gelang de (verhouding van) de gewichten op de eerste en tweede dimensie. Het zal duidelijk zijn dat er een verband bestaat tussen de (relatieve) grootte van de gewichten en de vorm van de configuratie in de groepsstimulusruimte. Is de configuratie cirkelvormig, dan moet de verhouding $w_{k1} : w_{k2}$ groot zijn om een liggende ellips te krijgen. Heeft de groepsconfiguratie daarentegen al de vorm van zo'n ellips, dan zal $w_{k1} : w_{k2}$ (veel) kleiner zijn. Kortom, de verhouding van de gewichten is afhankelijk van de vorm van de configuratie, en omgekeerd. Om een unieke oplossing te krijgen is het gebruikelijk om – net als in het ASCAL-geval – de groepsstimulusruimte te normaliseren. Dat wil zeggen: de groepsstimulusruimte wordt zodanig gestandaardiseerd dat de kwadraten som van de coördinaten op elke dimensie gelijk is aan m , zodat $\sum_i x_{is}^2 / m = 1$ voor elke dimensie s . Daarmee krijgt de configuratie een meer cirkelachtige vorm.

De stand van de assen

Zoals gezegd, kan men iedere individuele configuratie opvatten als een transformatie $XW_k^{1/2}$ van de groepsstimulusruimte X . De transformatiematrix $W_k^{1/2}$ veroorzaakt een *differentiële dilatatie* van de configuratie X . In tegenstelling tot een centrale dilatatie (zie Hoofdstuk 5) is zo'n differentiële dilatatie niet-afstandsbehoudend. Dat is ook de bedoeling: we gebruiken de transformatie $W_k^{1/2}$ juist om per individuele configuratie verschillende afstanden te creëren. Eveneens in tegenstelling tot een centrale dilatatie maakt het bij een differentiële dilatatie uit welke assen men nu van welk gewicht voorziet. Het gewicht $w_{ks}^{1/2}$ geldt niet voor zomaar een willekeurige richting, maar voor één precies vastgelegde richting in de ruimte. Dit is te zien aan de hand van de tekeningen in Figuur 9.3. In de linker tekening (a) is een liggende ellips afgebeeld die verkregen kan worden door een cirkel in horizontale richting uit te rekken en in verticale richting in te krimpen. Zouden we de horizontale en verticale as eerst roteren, bijvoorbeeld tot de stand in tekening (b), dan kan men nooit dezelfde ellips als in (a) krijgen door alleen maar de assen te wegen. Differentiële weging van de nieuwe assen levert een andere ellips op. Dit aspect van het INDSCAL-model wordt het *gebrek aan rotatievrijheid* genoemd. Bij een INDSCAL-oplossing hoort een vast, gefixeerd, assenstelsel dat men niet willekeurig mag roteren. Bij de interpretatie moet men dus uitgaan van de unieke oriëntatie die de assen in de INDSCAL-oplossing hebben meegekregen.



Figuur 9.3 Het effect van weging voor en na rotatie

Van bovenstaande eigenschap is bij de lancering van het INDSCAL-model betrekkelijk veel ophef gemaakt. Het gebrek aan rotatievrijheid is namelijk geen nadeel, maar een voordeel: we hoeven nu niet langer te zoeken naar nieuwe richtingen die optimaal interpreteerbaar zijn. De enig mogelijke assen die optimaal in het model passen, hebben we uit de analyse verkregen. Misschien hebben we dus wel de ‘ware onderliggende dimensies’ gevonden!

Het is echter maar de vraag of men aan deze eigenschap zoveel waarde moet hechten. Rotaties – zeker als ze niet al te groot zijn – geven natuurlijk wel andere gewichten, maar het is niet altijd zo dat de goodness-of-fit van het model meteen drastisch omlaaggaat. Bovendien is de verzameling gewichten die men in de analyse vindt afhankelijk van de verzameling matrices die men geanalyseerd heeft. Als die verzameling matrices een steekproef is uit een populatie (zoals bij proefpersonen het geval is) moet men nog maar afwachten of men bij andere steekproeven dezelfde oriëntatie van de assen zal vinden. Met andere woorden, het is de vraag hoe stabiel de gevonden richtingen uiteindelijk zijn.

Standaardisatie van de gewichten

Bij de interpretatie van de gewichten $\{w_{ks}\}$ gaat het vooral om hun onderlinge *verhouding*, die aangeeft hoe de verschillende dimensies aan de afstanden tussen de punten bijdragen. Het gaat om de *relatieve nadruk* op de verschillende dimensies, weergegeven door de verhouding $w_{k1} : w_{k2}$ (in het tweedimensionale geval). Wil men individuen met elkaar vergelijken, dan moet men niet de afzonderlijke gewichten² van de verschillende proefpersonen met elkaar vergelijken, maar de verschillende verhoudingen van de gewichten.

2 Een uitzondering doet zich voor als de data onconditioneel geanalyseerd zijn: in dat geval heeft het wel zin om afzonderlijke gewichten van de ene matrix met die van een andere te vergelijken (zie MacCallum, 1977).

Omdat de absolute grootte van de gewichten arbitrair is, kan men de gewichten op zo'n manier kiezen dat zij extra informatie bevatten. In ALSCAL worden de gewichten van elke individuele matrix zodanig gestandaardiseerd dat hun kwadraten som gelijk is aan de bijbehorende RSQ_k , dus:

$\sum_s w_{ks}^2 = RSQ_k$. De grootte van de gewichten geeft daardoor inzicht in de mate waarin een bepaalde nabijheidsmatrix O_k adequaat beschreven wordt door het INDSCAL-model.

Gewichten als individuele verschillen

Aangezien de gewichtenmatrix W_k kenmerken beschrijft van de nabijheidsmatrix O_k van individu k , hangen de betreffende gewichten mogelijk samen met andere kenmerken van individu k . Zulke samenhangen kunnen voor een groep proefpersonen op verschillende manieren onderzocht worden:

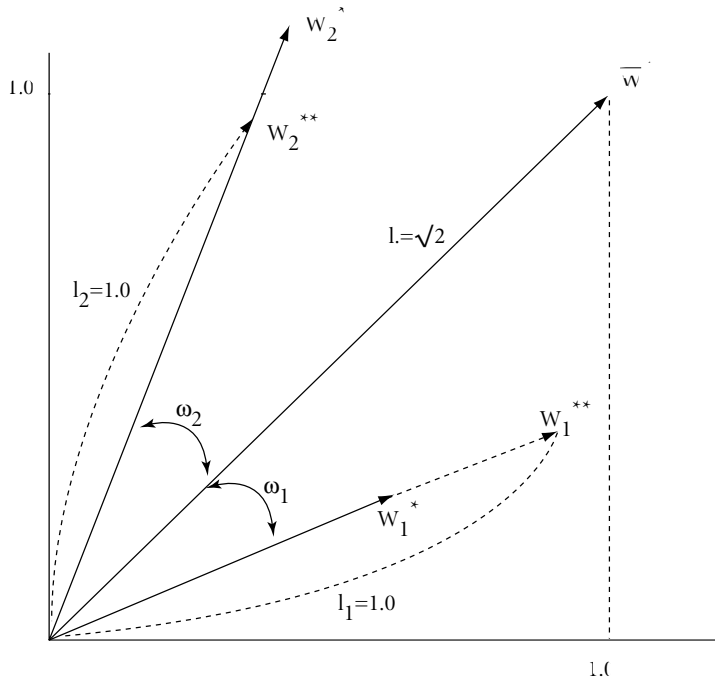
- 1 Door middel van multipele regressie en canonische correlatie van de $n \times r$ -matrix met gewichten $\{w_{ks}\}$ enerzijds met een of meer externe variabelen die eigenschappen van de individuen meten. Dat kan in twee richtingen gebeuren. We kunnen bijvoorbeeld het stemgedrag van een groep individuen proberen te voorspellen uit de gewichten die zij toekennen aan de assen van een configuratie van politieke partijen. Of: we kunnen die INDSCAL-gewichten juist proberen te voorspellen uit hun sekse, leeftijd, opleiding en economische status.
- 2 Door de individuen in groepen in te delen en univariate of multivariate variantie-analyse (ANOVA en MANOVA) op de gewichten toe te passen.
- 3 Door middel van *analysis of angular variation*, een ANOVA-achtige analyse van *richtingsgegevens* (zie Schiffman e.a., 1981; Jones, 1983).

Bij de twee eerstgenoemde analyses moet men bij voorkeur wel eerst een transformatie van de gewichten toepassen, enerzijds omdat de ruwe gewichten afhankelijk zijn van de *fit*, anderzijds omdat de gewichten op de verschillende dimensies op een niet-lineaire manier van elkaar afhankelijk zijn (zie Jones, 1983). Twee voorbeelden van zulke transformaties (zie Young, 1994) zijn de *weirdness index* en de *flattened weights*, die hieronder behandeld worden.

De weirdness-index

Zoals gezegd gaat het bij de interpretatie van INDSCAL-gewichten om de onderlinge verhoudingen van de gewichten op de verschillende dimensies. In plaats van univariate analyses op de gewichten van de afzonderlijke dimensies, of multivariate analyse op de gewichten van de verschillende dimensies tegelijkertijd, kan men de *verhoudingen van de gewichten* als variabele nemen. Laten we ons hier beperken tot het tweedimensionale geval. We kunnen dan het quotiënt w_{k1}/w_{k2} nemen om het 'gedrag' van individu k te beschrijven. Dit quotiënt is echter geen handige maat, omdat het heel groot kan worden als w_{k2} in de buurt van nul komt. Young (1994) beschrijft een andere index om individuele verschillen weer te geven, de zogenaamde *weirdness index*. De berekening daarvan gaat als volgt:

- 1 Bereken voor de hele groep het gemiddelde gewicht op de eerste dimensie, $\bar{w}_{.1} = \sum_k w_{k1}/n$, en het gemiddelde gewicht op de tweede dimensie, $\bar{w}_{.2} = \sum_k w_{k2}/n$ (algemeen: $\bar{w}_{.s} = \sum_k w_{ks}/n$).
- 2 Herschaal de oorspronkelijke gewichten door ze te delen door het bijbehorende gemiddelde. Dat geeft: $w_{ks}^* = w_{ks}/\bar{w}_{.s}$ (merk op dat het gemiddelde van deze nieuwe gewichten per dimensie gelijk is aan 1). Deze nieuwe gewichten kunnen we tegen elkaar uitzetten in een grafiek (zie Figuur 9.4) waarin we vanuit de oorsprong een lijn trekken naar het punt (1,1). Deze lijn, die hoeken van 45 graden met de horizontale en de verticale assen maakt, komt dus overeen met de gemiddelde richting van de groep proefpersonen. Daarom noemen we die lijn de gemiddeldenvector. Trekken we vanuit de oorsprong lijnen naar de punten met coördinaten w_{k1}^* en w_{k2}^* (de proefpersonen dus!), dan zien we het volgende. Proefpersonen met gewichten die gelijk zijn aan de gemiddelde gewichten van de groep vallen op de 45 graden-lijn. Proefpersonen met grotere gewichten voor de eerste dimensie dan voor de tweede komen onder die lijn terecht en proefpersonen met grotere gewichten voor de tweede dimensie dan voor de eerste komen boven die lijn uit. De hoek tussen de proefpersoonvector en de 45 graden-lijn laat dus zien hoe extreem een proefpersoon een bepaalde dimensie benadrukt. In de volgende stappen wordt deze hoek berekend.
- 3 In deze stap worden de gewichten van de proefpersonen opnieuw herschaald, en wel zodanig dat $w_{ks}^{**} = w_{ks}^* / \{(w_{k1}^{*2} + w_{k2}^{*2})^{1/2}\}$. Daardoor wordt de lengte van de herschaalde gewichtsvector van alle proefpersonen gelijk aan 1.00.
- 4 Aangezien $w_{k1}^{**} + w_{k2}^{**} = (1.0) \times w_{k1}^{**} + (1.0) \times w_{k2}^{**}$ is $w_{k1}^{**} + w_{k2}^{**} = l_k \cos(\omega_k) = (\sqrt{2.0}) \cos(\omega_k)$ zodat $\omega_k = \cos^{-1}\{(w_{k1}^{**} + w_{k2}^{**})/\sqrt{2.0}\}$. Deze hoek kan maximaal 45 graden boven en maximaal 45 graden onder de gemiddeldenvector uitkomen. De *weirdness* index z_k is nu gedefinieerd als het quotiënt van ω_k gedeeld door 45 graden, dus: $z_k = \omega_k/45 = [\cos^{-1}\{(w_{k1}^{**} + w_{k2}^{**})/\sqrt{2.0}\}]/45$.
Op deze manier ontstaat een index die tussen 0 (de individuele gewichtenvector valt samen met de gemiddeldenvector) en 1.0 (de individuele vector valt samen met een van de assen) varieert. De berekening van de *weirdness* index kan gemakkelijk gegeneraliseerd worden naar het geval dat r meer dan twee dimensies zijn. De *weirdness*-index kan men correleren met andere kenmerken van de individuen.



Figuur 9.4 Afbeelding van de herschaalde gewichtsvectoren en hun relatie tot de bijbehorende weirdness indices

‘Flattened weights’

Als we de individuele gewichten afbeelden als vectoren in de ruimte, dan wordt het verschil tussen twee individuen het best weergegeven door de hoek die hun vectoren onderling maken. Young (1994) beschrijft een procedure om in het tweedimensionale geval de individuen op een lijn af te beelden, zodanig dat de verschillen tussen de punten op de lijn zinvol corresponderen met de hoeken tussen de individuele gewichtenvectoren. In het driedimensionale geval worden de individuele vectoren afgebeeld op een plat vlak, enzovoort. De procedure is als volgt: deel de gewichten van de afzonderlijke dimensies door hun som en beschouw verder alleen de eerste $r - 1$ dimensies. Noemen we de platgemaakte gewichten p_k , dan zijn die gelijk aan de getallen op de eerste dimensie, in afwijking van hun gemiddelde. Dus, in het tweedimensionale geval:

$$p_k = \frac{w_{k1}}{w_{k1} + w_{k2}} - \frac{1}{n} \sum_{k=1}^n \left(\frac{w_{k1}}{w_{k1} + w_{k2}} \right).$$

In het driedimensionale geval is

$$p_{k1} = \frac{w_{k1}}{w_{k1} + w_{k2} + w_{k3}} - \frac{1}{n} \sum_{k=1}^n \left(\frac{w_{k1}}{w_{k1} + w_{k2} + w_{k3}} \right)$$

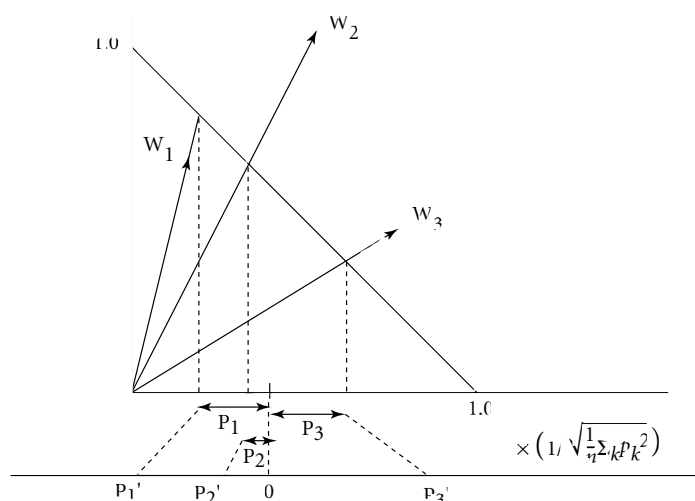
en

$$p_{k2} = \frac{w_{k2}}{w_{k1} + w_{k2} + w_{k3}} - \frac{1}{n} \sum_{k=1}^n \left(\frac{w_{k2}}{w_{k1} + w_{k2} + w_{k3}} \right).$$

Vervolgens worden deze p -gewichten zodanig gestandaardiseerd dat hun kwadraatsom gelijk aan n wordt. Dus:

$$p_k^* = \frac{p_k}{\sqrt{\frac{1}{n} \sum_k p_k^2}}.$$

Een grafische afbeelding van deze procedure staat in Figuur 9.5.



Figuur 9.5 Een grafische afbeelding van platgemaakte gewichten

BLOK 9.2 RMDS EN INDSICAL: VERONDERSTELDE RELATIES TUSSEN LANDEN

In een onderzoek van Jaspars, Van de Geer, Tajfel en Johnson (1972) wilde men van 120 leerlingen in de tweede tot en met zesde klassen van een basisschool te weten komen wat de door hen veronderstelde relaties tussen de landen Amerika, Duitsland, Engeland, Frankrijk, Nederland en Rusland waren. De landen werden op kaartjes paarsgewijs aangeboden, en de leerlingen werd gevraagd aan te geven of de landen 'goed met elkaar konden opschieten'. De leerlingen konden op deze vraag de antwoorden 'ja' of 'nee' geven. De onderzoekers telden voor ieder paar landen per klas het aantal leerlingen dat 'ja' geantwoord had en deelden dit aantal door het totaal aantal leerlingen in die klas. Op deze manier werden vijf verzamelingen proporties verkregen, die een maat zijn voor de psychologische nabijheid van de zes landen. Van elke proportie is door de onderzoekers .50 afgetrokken. De resulterende getallen staan in Tabel 9.1.

Tabel 9.1 Veronderstelde relaties tussen zes landen (proporties positieve relaties minus .50)^a

	Klas				
	2	3	4	5	6
Nederland - Amerika	.125	.239	.333	.407	.500
Nederland - Engeland	.188	.428	.303	.417	.500
Nederland - Frankrijk	.104	.178	.156	.261	.120
Nederland - Duitsland	-.094	-.031	-.053	.229	.359
Nederland - Rusland	-.136	.031	-.239	-.208	-.185
Amerika - Engeland	.208	.333	.125	.333	.413
Amerika - Frankrijk	.208	.083	.167	.042	.152
Amerika - Duitsland	.042	-.208	-.125	.125	.109
Amerika - Rusland	.167	.083	-.125	-.292	-.326
Engeland - Frankrijk	.167	.083	.250	.167	.022
Engeland - Duitsland	-.083	.000	-.042	.125	-.022
Engeland - Rusland	-.042	.125	-.042	-.250	-.239
Frankrijk - Duitsland	.000	-.042	-.125	.000	.022
Frankrijk - Rusland	.208	-.042	.000	.000	.022
Duitsland - Rusland	.125	.083	-.042	-.083	-.152

^a ontleend aan Jaspars et al., 1972, p. 361, Table 4).

We hebben nu dus vijf matrices met gegevens over de veronderstelde nabijheden tussen zes landen. Deze gegevens zijn op twee manieren met ALSCAL geanalyseerd: als RMDS- en als WMDS-probleem.

Replicated MDS van vijf nabijheidsmatrices voor zes landen

Hieronder geven we een afdruk van de SPSS-commando's voor een RMDS-analyse van bovengenoemde data met ALSCAL.

```
TITLE 'RMDS van Jaspars, Van de Geer, Tajfel en Johnson, 1972'.
```

```
DATA LIST TABLE/ NED AME ENG FRA DUI RUS 1-36.
```

```
SUBTITLE 'Replicated MDS'.
```

```
BEGIN DATA.
```

```
.500
.125 .500
.188 .208 .500
.104 .208 .167 .500
-.094 .042 -.083 .000 .500
-.136 .167 -.042 .208 .125 .500
.500
.239 .500
.428 .333 .500
.178 .083 .083 .500
-.031 -.208 .000 -.042 .500
.031 .083 .125 -.042 .083 .500
.500
.333 .500
.303 .125 .500
.156 .167 .250 .500
-.053 -.125 -.042 -.125 .500
-.239 -.125 -.042 .000 -.042 .500
.500
.407 .500
.417 .333 .500
.261 .042 .167 .500
.229 .125 .125 .000 .500
-.208 -.292 -.250 .000 -.083 .500
.500
.500 .500
.500 .413 .500
.120 .152 .022 .500
.359 .109 -.022 .022 .500
-.189 -.326 -.239 .022 -.152 .500
```

```
END DATA.
```

```

ALSCAL VARIABLES=NED TO RUS
/SHAPE=SYMMETRIC
/LEVEL=ORDINAL(SIMILARITY)
/CONDITION=MATRIX
/MODEL=EUCLID
/CRITERIA CONVERGE(.00001) ITER(100)
      DIMENSIONS(2) CUTOFF(-1.00)
/PRINT DATA HEADER
/PLOT=ALL.

```

De opdracht /CUTOFF (-1.00) is nodig om negatieve getallen tussen -1 en 0 in te kunnen lezen. Getallen kleiner dan -1 worden door ALSCAL als missing data beschouwd. Zonder deze toevoeging neemt ALSCAL aan dat alle getallen kleiner dan 0.0 missing data zijn. Hieronder volgen delen van de ALSCAL-uitvoer.

Alscal Procedure Options

Data Options-

```

Number of Rows (Observations/Matrix).      6
Number of Columns (Variables) . . . . . 6
Number of Matrices . . . . . 5
Measurement Level . . . . . Ordinal
Data Matrix Shape . . . . . Symmetric
Type . . . . . Similarity
Approach to Ties . . . . . Leave Tied
Conditionality . . . . . Matrix
Data Cutoff at . . . . . -1.000000

```

Model Options-

```

Model . . . . . Euclid
Maximum Dimensionality . . . . . 2
Minimum Dimensionality . . . . . 2
Negative Weights . . . . . Not Permitted

```

Output Options-

```

Job Option Header . . . . . Printed
Data Matrices . . . . . Printed
Configurations and Transformations . . . . . Plotted
Output Dataset . . . . . Not Created
Initial Stimulus Coordinates . . . . . Computed

```

Algorithmic Options-

```

Maximum Iterations . . . . . 100
Convergence Criterion . . . . . .00001
Minimum S-stress . . . . . .00500
Missing Data Estimated by . . . . . Ulbounds
Tiestore . . . . . 75

```

Allereerst zorgt ALSCAL ervoor dat de ingelezen datamatrices worden afgedrukt: de Raw (unscaled) Data for Subject 1 tot en met de Raw (unscaled) Data for Subject 5. Daarna volgt de eigenlijke analyse met onderstaande uitvoer.

Iteration history for the 2 dimensional solution (in squared distances)

Young's S-stress formula 1 is used.

Iteration	S-stress	Improvement
1	.29782	
2	.27168	.02615
3	.26273	.00895
4	.26019	.00254
5	.25975	.00044
6	.25971	.00003
7	.25973	-.00002

Iterations stopped because S-stress improvement is less than .000010

Stress and squared correlation (RSQ) in distances RSQ values are the proportion of variance of the scaled data (disparities) in the partition (row, matrix, or entire data) which is accounted for by their corresponding distances.

Stress values are Kruskal's stress formula 1.

Matrix	Stress	RSQ	Matrix	Stress	RSQ
1	.301	.551	2	.172	.852
3	.204	.785	4	.130	.913
5	.131	.912			

Averaged (rms) over matrices Stress = .19776 RSQ = .80272

Zoals hierboven blijkt, heeft ALSCAL in zeven iteraties een twee-dimensionale oplossing met $S\text{-stress} = .25793$ gevonden, die de nabijheidsdata van de vijf matrices gezamenlijk zo goed mogelijk beschrijft. Het gaat hier dus om een *compromis*-oplossing waarin elke afstand d_{ij} zo veel als mogelijk moet overeenstemmen met vijf pseudo-afstanden $f_k(o_{ijk})$ (NB: f_k omdat de matrices hier matrixconditioneel behandeld worden; de observaties van elke matrix worden getransformeerd via een eigen transformatiefunctie). De waarde van $S\text{-stress}$ is in dit geval eigenlijk een gemiddelde, een zogenaamde *root mean square*:

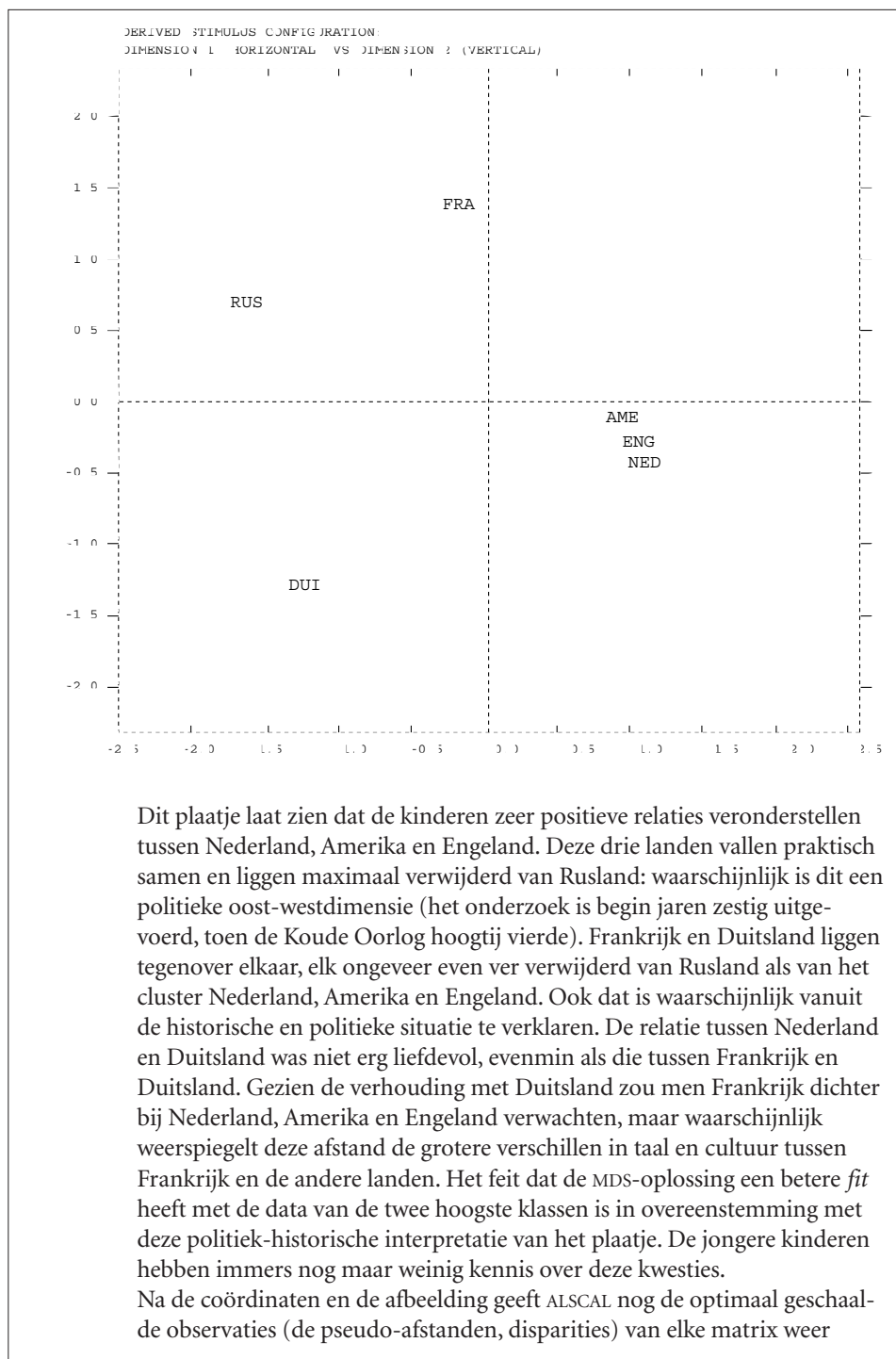
[9.4]

$$S\text{-stress} = \left[\frac{1}{n} \sum_{k=1}^n \left(\frac{\sum_{i=1}^m \sum_{j=1}^m \{d_{ij}^2 - (f_k(o_{ijk}))^2\}}{\sum_{i=1}^m \sum_{j=1}^m \{d_{ij}^2\}} \right) \right]^{.5}$$

Naast bovenstaande $S\text{-stress}$ berekent ALSOAL voor iedere matrix afzonderlijk Kruskals $Stress_1$ en RSQ en ook de *root mean square* van Kruskals $Stress_1$ (hier .198). We zien dat de oplossing beter bij de laatste twee matrices (van klas 5 en 6) past dan bij de eerste drie (van klas 2, 3 en 4). Het ligt voor de hand te concluderen dat oudere kinderen consistentere patronen van relaties tussen de landen 'zien'. Daarom, en gegeven het feit dat het hier om een compromis gaat, is een totale Kruskal $Stress_1$ van .198 een redelijk goede oplossing te noemen. De coördinaten van de landen en de bijbehorende afbeelding zijn hieronder weergegeven.

Configuration derived in 2 dimensions
Stimulus Coordinates

Stimulus Number	Stimulus Name	Dimension	
		1	2
1	NED	1.1028	-.3699
2	AME	.9859	-.1008
3	ENG	1.0199	-.3501
4	FRA	-.2083	1.3699
5	DUI	-1.2307	-1.3064
6	RUS	-1.6695	.7573



gevolgd door het Shepard-diagram (niet in SPSS for WINDOWS) en grafieken van de lineaire *fit* en de transformatiefunctie(s) (zie Hoofdstuk 6). Deze uitvoer geven we hier niet weer. In ALSICALPC en ALSICAL voor SPSS-X worden in de genoemde drie grafieken de gegevens van alle matrices tegelijkertijd afgebeeld. Dat is alleen dan zinvol als de analyse onconditioneel gedaan is. Is de analyse matrixconditioneel uitgevoerd, dan zou men de genoemde drie grafieken voor iedere matrix afzonderlijk moeten maken. Aan de gecombineerde grafieken is namelijk niet te zien of er matrices zijn met een perfecte, zeer goede of juist zeer slechte *fit*, en zo ja, welke matrices dat dan zijn. Deze mogelijkheid bestaat ten dele in de ALSICAL-versie van SPSS voor WINDOWS; alleen daarin kunnen de transformatiefuncties van de afzonderlijke matrices worden opgevraagd (door middel van /PLOT=ALL).

INDSCAL-analyse van de nabijheden tussen zes landen

Dezelfde vijf matrices zijn ook geanalyseerd volgens het INDSCAL-model. De aansturing van ALSICAL was als volgt:

```
ALSICAL VARIABLES=NED TO RUS
/SHAPE=SYMMETRIC
/LEVEL=ORDINAL (SIMILARITY)
/CONDITION=MATRIX
/MODEL=INDSCAL
/CRITERIA CONVERGE(.00001) ITER(100)
      DIMENSIONS(2) CUTOFF(-1.00)
/PRINT DATA HEADER /PLOT=ALL.
```

Deze opdrachten resulteren na 39 iteraties in een tweedimensionale oplossing met $S\text{-stress} = .20157$. Deze waarde is lager dan die van de RMDS-oplossing ($S\text{-stress} = .25973$). Dat moet ook wel omdat deze oplossing veel meer parameters heeft dan in het RMDS-geval. Voor elke matrix zijn er twee gewichten geschat, waardoor een betere overeenstemming tussen de (getransformeerde) observaties o_{ijk} van elke matrix en de bijbehorende afstanden d_{ijk} ontstaat. Het verschil in $S\text{-stress}$ is overigens niet erg groot, wat betekent dat de verschillen tussen de vijf matrices niet véél beter met behulp van het gewogen-afstandsmodel beschreven kunnen worden. Wel zien we dat de Kruskal stress van de laatste twee matrices aanzienlijk gedaald is; voor deze twee gaat het INDSCAL-model dus kennelijk beter op. Hieronder volgen delen van de ALSICAL-uitvoer.

ITERATION HISTORY FOR THE 2 DIMENSIONAL SOLUTION (IN SQUARED DISTANCES)
YOUNGS S-STRESS FORMULA 1 IS USED.

ITERATION	S-STRESS	IMPROVEMENT
0	0.23475	
1	0.22648	
2	0.20925	0.01724
..
..
37	0.20159	0.00001
38	0.20158	0.00001
39	0.20157	0.00001

ITERATIONS STOPPED BECAUSE
S-STRESS IMPROVEMENT LESS THAN 0.000010

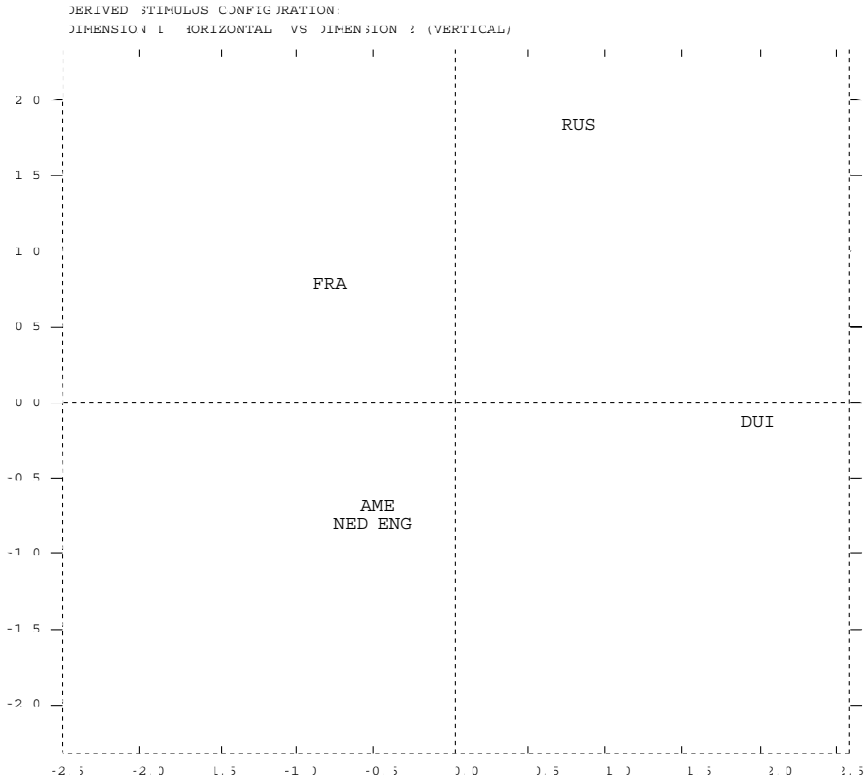
MATRIX	STRESS	RSQ	MATRIX	STRESS	RSQ	MATRIX	STRESS	RSQ
1	0.305	0.654	2	0.270	0.827	3	0.146	0.910

MATRIX	STRESS	RSQ	MATRIX	STRESS	RSQ
4	0.028	0.997	5	0.029	0.997

AVERAGED (RMS) OVER MATRICES
STRESS = 0.194 RSQ = 0.877

CONFIGURATION DERIVED IN 2 DIMENSIONS
STIMULUS COORDINATES

STIMULUS NUMBER	STIMULUS NAME	PLOT SYMBOL	DIMENSION	
			1	2
1	NED	1	-0.7531	-0.8322
2	AME	2	-0.6022	-0.7322
3	ENG	3	-0.5881	-0.8250
4	FRA	4	-0.7363	0.6812
5	DUI	5	1.8837	-0.1870
6	RUS	6	0.7960	1.8952



In de bovenstaande afbeelding heeft de configuratie een wat minder ellips-achtige vorm en wat meer de vorm van een cirkel. Dat ligt aan de standaardisatie van de coördinaten. Deze is hier zodanig dat de kwadraten van de coördinaten op beide dimensies gelijk aan 6.0 is. Een ander verschil met de RMDS-oplossing is dat de horizontale as gespiegeld is en daarna (samen met de verticale as) rechtsom geroteerd is. Dit verandert echter niets aan de inhoudelijke interpretatie.

ALSCAL berekent voor elke individuele matrix twee gewichten, de bijbehorende *weirdness-index* en *flattened weights*.

SUBJECT WEIGHTS MEASURE THE IMPORTANCE OF EACH DIMENSION TO EACH SUBJECT.

SQUARED WEIGHTS SUM TO RSQ. A SUBJECT WITH WEIGHTS PROPORTIONAL TO THE AVERAGE WEIGHTS HAS A WEIRDNESS OF ZERO, THE MINIMUM VALUE. A SUBJECT WITH ONE LARGE WEIGHT AND MANY LOW WEIGHTS HAS A WEIRDNESS NEAR ONE. A

SUBJECT WITH EXACTLY ONE POSITIVE WEIGHT HAS A WEIRDNESS OF ONE, THE MAXIMUM VALUE FOR NONNEGATIVE WEIGHTS.

SUBJECT WEIGHTS					
DIMENSION					
SUBJECT	PLOT	WEIRD-	1	2	(RSQ)
NUMBER	SYMBOL	NESS			
1	1	0.4216	0.7314	0.3447	(.654)
2	2	1.0000	0.9094	0.0000	(.827)
3	3	0.1964	0.7796	0.5498	(.910)
4	4	0.6815	0.2556	0.9652	(.997)
5	5	0.6857	0.2522	0.9659	(.997)

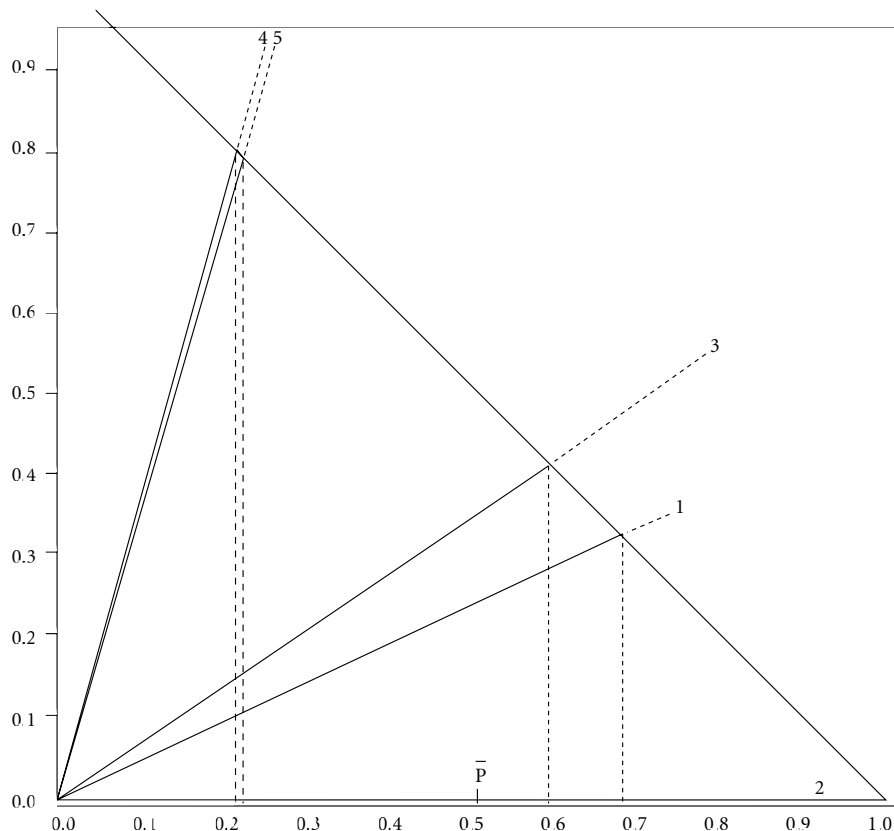
OVERALL IMPORTANCE
OF EACH DIMENSION: 0.4197 0.4571

FLATTENED SUBJECT WEIGHTS		
VARIABLE		
SUBJECT	PLOT	1
NUMBER	SYMBOL	
1	1	0.4756
2	2	1.5392
3	3	0.1657
4	4	-1.0864
5	5	-1.0941

De gewichten laten hier drie dingen zien (a) de ruimte van subject 3 (klas 4) lijkt qua vorm het meest op bovenstaande afbeelding omdat de desbetreffende *weirdness*-index het laagst is; (b) de ruimte van klas 3 is eendimensionaal omdat het gewicht voor de tweede dimensie gelijk aan nul is (in deze ruimte hebben Frankrijk en Duitsland dus de slechtste relatie); (c) de ruimten van klas 5 en klas 6 lijken erg op elkaar omdat zij nagenoeg dezelfde gewichten hebben: de tweede dimensie met de tegenstelling Rusland versus Nederland, Amerika en Engeland is voor deze leerlingen het belangrijkste, hun ruimten zien er dus veel elliptischer uit dan bovenstaande figuur.

Hieronder de grafiek uit ALSCAL waarin de gewichten tegen elkaar zijn uitgezet. In deze figuur zijn ook de *flattened weights* ingetekend. ALSCAL levert ook één of meer grafieken waarin de *flattened weights* tegen elkaar zijn uitgezet. In dit geval is er slechts één reeks *flattened weights*. Een grafiek is hier niet erg zinvol en is daarom niet weergegeven.

DERIVED SUBJECT WEIGHTS: DIMENSION 1 (HORIZONTAL), VS DIMENSIONS 2 (VERTICAL)



Ten slotte drukt ALSICAL het Shepard-diagram (niet in SPSS for WINDOWS), de grafiek van de lineaire *fit* en de grafiek van de transformatiefunctie(s) af. In ALSICALPC en ALSICAL voor SPSS-X worden in de genoemde drie grafieken de gegevens van alle matrices tegelijkertijd afgebeeld. Dat is alleen dan zinvol als de analyse onconditioneel gedaan is. Is de analyse matrixconditioneel uitgevoerd, dan zou men de genoemde drie grafieken voor iedere matrix afzonderlijk moeten maken. Deze mogelijkheid bestaat ten dele in de ALSICAL-versie van SPSS voor WINDOWS; alleen daarin kunnen de transformatiefuncties van de afzonderlijke matrices worden opgevraagd (door middel van /PLOT=ALL).

9.4 ANDERE INDIVIDUELE-VERSCHILLENMODELLEN

IDIOSCAL

Het INDSCAL-model kan nog uitgebreid worden naar het zogenaamde IDIOSCAL-model (Carroll, 1972; zie Hoofdstuk 5). In dit model wordt er per individu en per dimensie niet slechts één gewicht maar een aantal gewichten gepostuleerd. De bijbehorende matrix – laten we die S_k noemen – is dan niet diagonaal zoals W_k , maar kan elke willekeurige orthogonale transformatiematrix zijn. We kunnen nu schrijven $S_k = T_k W_k$ met T_k een orthogonale rotatiematrix en W_k een diagonale gewichtenmatrix. Inhoudelijk komt dit model neer op de veronderstelling dat ieder individu k de assen van de gemeenschappelijke configuratie X eerst op een eigen (idiosyncratische) manier roteert en vervolgens deze assen op een individuele manier weegt (zie ook Hoofdstuk 5).

Reduced rank scaling

Een speciaal geval van IDIOSCAL doet zich voor als een of meer individuele nabijheidsmatrices onderliggende configuraties hebben met (veel) minder dimensies dan de groepsstimulusruimte. Het idee is nu dat elk individu de punten projecteert op een *deelruimte* van de groepsstimulusconfiguratie, vandaar de naam *reduced rank scaling*. De geobserveerde nabijheden zijn dan een functie van de afstanden in deze minder-dimensionale ruimte. In het extreme geval is die ruimte eendimensionaal: de proefpersoon kiest een idiosyncratische richting in de groepsruimte en projecteert daar alle punten op. Ook IDIOSCAL (en dus ook *reduced rank scaling*) is in principe met ALSICAL mogelijk. Daartoe gebruikt men de opdracht /MODEL=GEMSCAL.

Points of view-analysis

Een klassieke manier om individuele verschillen in nabijheidsdata door middel van een analyse zichtbaar te maken is Tucker en Messicks *points of view*-analyse uit 1963. Voor die tijd waren er slechts twee mogelijkheden om een driewegmatrix te analyseren: metrische MDS van de gemiddelde nabijheidsmatrix of metrische MDS van elke nabijheidsmatrix afzonderlijk.

In de methode van Tucker en Messick worden de symmetrische $m \times m$ -matrices van de n individuen omgevormd tot de kolommen van een matrix met $m(m-1)/2$ rijen en n kolommen. Vervolgens wordt op deze matrix een principale-componentenanalyse uitgevoerd, die een aantal componenten (factoren) oplevert waarop de individuen hoge of lage ladingen kunnen hebben. Elk van deze componenten correspondeert met een unieke verzameling nabijheidsdata en wordt daarom een *point of view* genoemd.

Om een *points of view*-oplossing te interpreteren selecteert men enkele proefpersonen die elk een hoge lading hebben op een van de gezichtspunten en nul-ladingen op de andere factoren. Vervolgens verricht men afzonderlijke CMDS-analyses op de nabijheidsdata van de geselecteerde individuen. Aanvankelijk is deze – metrische – methode betrekkelijk veel gebruikt. Hij is echter in

onbruik geraakt toen in 1970 het INDSCAL-programma gelanceerd werd. Nu het tegenwoordig ook mogelijk is om niet-metrische principale componentenanalyses uit te voeren³, verdient deze methode weer meer aandacht dan hij lange tijd gekregen heeft. Aanzetten hiertoe zijn te vinden bij Verboon en Van der Kloot, 1989 en Meulman en Verboon, 1993.

9.5 ASYMMETRISCHE DRIEWEG/TWEEMODALE NABIJHEIDSDATA

De laatste uitbreiding van het klassieke CMDS-geval betreft een combinatie van de twee varianten die in Hoofdstuk 8 en hierboven besproken zijn: het geval dat er een aantal verschillende matrices met asymmetrische nabijheidsdata geobserveerd is. Voor de aanpak van deze data geldt dat we in principe elke mogelijke analyse van asymmetrische data kunnen combineren met elke mogelijke analyse van driewegdata. Behalve in één geval valt daarover weinig nieuws meer te vertellen. Die ene uitzondering is de combinatie die ontstaat door een gewogen-afstandsmodel te gebruiken dat zowel gewichten voor de stimuli (de asymmetrieparameters) als gewichten voor de individuen (proefpersoongewichten) bevat. Het betreffende afstandsmodel is in matrixnotatie

$$\begin{aligned} d_{ijk}^{*2} &= (\mathbf{x}'_i - \mathbf{x}'_j) \mathbf{V}_i^{1/2} \mathbf{W}_k \mathbf{V}_i^{1/2} (\mathbf{x}'_i - \mathbf{x}'_j)' \\ &= (\mathbf{x}'_i - \mathbf{x}'_j) \mathbf{V}_i^{1/2} \mathbf{W}_k^{1/2} \mathbf{W}_k^{1/2} \mathbf{V}_i^{1/2} (\mathbf{x}'_i - \mathbf{x}'_j)' \\ &= (\mathbf{x}^{(k)'}_i - \mathbf{x}^{(k)'}_j) (\mathbf{x}^{(k)}_i - \mathbf{x}^{(k)}_j). \end{aligned} \quad [9.5]$$

In som-notatie is dit [9.6]

$$d_{ijk}^{*2} = \sqrt{\sum_s v_{is} w_{ks} (x_{is} - x_{js})^2} = \sqrt{\sum_s (v_{is}^{1/2} w_{ks}^{1/2} x_{is} - v_{is}^{1/2} w_{ks}^{1/2} x_{js})^2} = \sqrt{\sum_s (x_{is}^{(ki)} - x_{js}^{(ki)})^2}$$

waarin $x_{is}^{(ki)}$ de coördinaat van stimulus i op dimensie s is in een ruimte die bij individu k en object i hoort.

De interpretatie is dezelfde die we eerder bij ASCAL en INDSCAL zijn tegengekomen. De gewichten van de stimuli en van de proefpersonen kunnen onafhankelijk van elkaar geïnterpreteerd worden. Zowel voor de stimulusgewichten als voor de proefpersoongewichten geldt dat het vooral om de verhouding van de gewichten voor de verschillende dimensies gaat. Deze analyse krijgt men in ALSICAL door de opties /SHAPE=ASYMMETRIC /MODEL=AINDS te specificeren.

3 Een geschikt programma hiervoor is het door Gifi (1990) ontworpen programma PRINCALS in de SPSS-module CATEGORIES.