

— De opzet van een MDS-onderzoek

12.1 ASPECTEN VAN EEN MDS-ONDERZOEK

Zoals we in de vorige hoofdstukken gezien hebben, vormen meerdimensionale schaaltechnieken een familie van modellen en technieken voor de analyse van allerlei typen data in diverse soorten datamatrices. Het komt voor dat MDS-technieken worden toegepast op ‘bestaande data’, dat wil zeggen, op data die in eerder onderzoek verzameld zijn, vaak zonder dat het oorspronkelijk de bedoeling was de gegevens met MDS te analyseren. Anderzijds worden MDS-technieken gebruikt in onderzoek dat speciaal is opgezet om gegevens te verkrijgen die met MDS geanalyseerd zullen worden. In het eerste geval is de belangrijkste vraag of de beschikbare data zich inderdaad lenen voor MDS-analyse en zo ja, welke modellen het meest geschikt zijn. In het tweede geval gaat het er vooral om het onderzoek zodanig op te zetten dat de verzamelde data geschikt zijn om door middel van MDS-analyse een antwoord te geven op de *vraagstelling* die aan het onderzoek ten grondslag lag. In dat geval hangt de toepasbaarheid van MDS dus niet alleen af van de data, maar ook van de *vraagstelling*.

In alle empirisch wetenschappelijk onderzoek, of het nu gebruikmaakt van MDS of niet, is een aantal aspecten te onderscheiden die van belang zijn voor de opzet en uitvoering van zo’n onderzoek. Deze aspecten zijn: vraagstelling, selectie van stimuli en condities, selectie van proefpersonen of andere observatie-eenheden, procedure van dataverzameling waaronder taken en instructies, selectie van analysemethode, uitvoering van analyse, interpretatie van resultaten, en rapportage. In dit hoofdstuk zullen deze aspecten, soms kort en soms meer in detail, worden besproken.

12.2 VRAAGSTELLING

Globaal zou men vier redenen kunnen noemen om in een onderzoek MDS-technieken toe te passen. In de eerste plaats kan men vooral geïnteresseerd zijn in het maken van een plaatje, dat wil zeggen, een mooie, eenvoudige afbeelding van een veelheid van gegevens die zonder MDS niet in één oogopslag te interpreteren zijn. Op deze manier wordt MDS puur toegepast als grafische techniek, als *ordinatiemethode* (zie Everitt, 1980). Hierbij gaat men er zonder meer van uit dat de verzamelde data zich lenen voor de desbetreffende MDS-analyse, de nadruk valt op de interpretatie van het plaatje.

Een tweede reden om MDS-analyse toe te passen is het willen berekenen van *schaalwaarden*. In dit soort toepassingen gaat het niet alleen (of niet vooral) om een plaatje, maar om de coördinaten van de stimuli op de dimensies. Zulke schaalwaarden kunnen opgevat worden als *operationalisaties* van de gevonden dimensies. Ze kunnen gebruikt worden om de relaties van die dimensies met andere variabelen te bestuderen. Bijvoorbeeld: we kunnen nagaan of de schaalwaarden van misdrijven gecorreleerd zijn met de minimum en maximum strafmaten uit het Wetboek van Strafrecht. Of: correleren de schaalwaarden van geursensaties met de concentraties van bepaalde chemische stoffen? Ook kunnen schaalwaarden later gebruikt worden om meetinstrumenten op te stellen. Bijvoorbeeld: gegeven de schaalwaarden van een aantal politieke uitspraken kunnen we iemand karakteriseren in termen van de gemiddelde schaalwaarde van de uitspraken waar die persoon het mee eens is. Voor dit soort toepassingen is het natuurlijk wel belangrijk dat er een goede *fit* is van de MDS-oplossing met de data.

Een derde reden om MDS-technieken te gebruiken is om na te kunnen gaan wat de structuur van de onderzochte stimuli of objecten precies is. Welke overeenkomsten en verschillen zijn er tussen een aantal stimuli? Welk afstandsmodel en welke dimensies liggen er aan zulke (dis)similarities ten grondslag? Hoe moeten deze dimensies geïnterpreteerd worden? Is er een *a priori* theoretische structuur (een hypothese) die voorspelt hoe de resultaten eruit zouden moeten zien? En als er data van verschillende observatie-eenheden zijn: is er sprake van individuele verschillen en zo ja, kunnen die herleid worden tot één onderliggende configuratie van de objecten? Bij veel van het MDS-onderzoek in de sociale wetenschappen gaat het om dit soort inhoudelijke vraagstellingen, die vaak vooral exploratief maar ook wel confirmatief onderzocht worden. Juist bij dergelijk onderzoek moet er een goede aansluiting bestaan tussen vraagstelling, dataverzameling en analyse. In de rest van dit hoofdstuk wordt vooral aan dit type onderzoek aandacht besteed.

In veel MDS-onderzoek worden data geanalyseerd die bestaan uit beoordelingen door proefpersonen van de onderlinge gelijkenis van stimuli. Bij dit soort beoordelingen moet de proefpersoon een cognitieve taak verrichten; de waargenomen relaties tussen de stimuli zijn *cognities* die onderdeel zijn van de *cognitieve schema's* van de respondent. Daarom wordt een MDS-oplossing ook wel een *cognitieve kaart* genoemd. Dit soort toepassingen van MDS zijn uiteraard

zeer belangrijk, maar – zoals we gezien hebben – het is bepaald niet de enig zinvolle manier waarop MDS gebruikt kan worden.

Een vierde reden om MDS-analyse toe te passen is om verschillende technieken, modellen en algoritmen met elkaar en met nieuwe varianten te vergelijken.

Vaak gaat men daarbij uit van bestaande data waarvan de structuur bekend is of verzamelt men nieuwe gegevens over stimuli waarvan men mag verwachten dat ze een bekende structuur hebben (bijvoorbeeld de kleurencirkel).

Het zal duidelijk zijn dat het onderscheid tussen bovengenoemde ‘redenen’ nogal kunstmatig is. Met name de eerste drie typen onderzoek zullen nauwelijks in een ‘zuivere vorm’ voorkomen. Als onderzoekers geïnteresseerd zijn in plaatjes, dan zijn ze dat omdat ze de structuur van stimuli willen beschrijven. En om schaalwaarden zinvol toe te kunnen passen, moeten de dimensies waarop de schaalwaarden ‘gemeten’ zijn een goede representatie van de objecten zijn. Kortom, als er al sprake is van een onderscheid, dan is dat graduueel en niet absoluut.

12.3 STIMULI EN CONDITIES

Wanneer een onderzoeker een inhoudelijke vraagstelling wil beantwoorden en geen gebruikmaakt van bestaande data, dan moet er veel aandacht besteed worden aan de selectie van een *representatieve verzameling* stimuli. Daarbij zijn van belang het aantal en de aard van de stimuli en de manier waarop de selectie plaatsvindt.

Het aantal stimuli

Om met een voor de hand liggende opmerking te beginnen: het aantal stimuli moet niet te klein en mag niet te groot zijn, maar hoe klein of hoe groot dit aantal moet zijn, hangt van zeer veel inhoudelijke en praktische omstandigheden af. Dat het aantal stimuli niet te klein moet zijn, spreekt vanzelf als men wil dat de onderzochte stimuli representatief zijn voor een veel groter domein van mogelijke stimuli. Dat is des te belangrijker als stimuli in dit domein in veel opzichten van elkaar zouden kunnen verschillen. Immers, in Hoofdstuk 7 hebben we naar voren gebracht dat r , het aantal dimensies van een oplossing waaraan men enige stabiliteit wil kunnen toeschrijven, ten hoogste gelijk is aan $(m - 1)/4$ (waarbij m het aantal stimuli is). Als we verwachten dat de stimuli op vier of misschien wel vijf dimensies variëren, zouden we er dus minstens 21 in ons onderzoek moeten betrekken.

Hoewel een groot aantal stimuli vaak wenselijk is, hangt het van de dataverzamelingsprocedure af of dit ook mogelijk is. Als het gaat om onderzoek waarin respondenten de stimuli paarsgewijs moeten beoordelen op gelijkenis of preferentie, dan krijgt iedere proefpersoon $m(m - 1)/2$ stimulusparen voorgelegd, een aantal dat al gauw te groot is. Bijvoorbeeld: 21 stimuli leiden tot 210 paarsgewijze vergelijkingen. Om proefpersonen een dergelijk aantal beoordelingen te laten maken, moet men speciale voorzieningen treffen,

zoals herhaalde rustpauzes, meerdere onderzoekssessies of incomplete designs. Schiffman e.a. stellen dat

it is rarely feasible, from the viewpoints of sensory fatigue, interest, and personal schedules, to retain subjects for more than 1 ½ hours per session. A shorter session is preferable. We recommend that no more than 55 verbal or visual judgments be attempted in one session. These can be accomplished in 1 hour or less. From the outset we wish to emphasize the importance of gathering data in a very sound experimental fashion. Hastily gathered data under uncontrolled conditions seldom yield results which can be replicated (1981, p. 20).

Later in dit hoofdstuk zullen we nog op deze kwestie terugkomen.

De aard van de stimuli

Hoewel het natuurlijk veel uitmaakt of men respondenten concepten (zoals misdrijven, bijdragen aan wetenschappelijke artikelen, persoonlijkheidskenmerken) laat beoordelen of paren chemische stimuli (in een geurexperiment), toonhoogten of gekleurde lichtjes, in alle gevallen geldt dat de gebruikte stimuli op een evenwichtige manier van elkaar moeten verschillen. Het gaat daarbij om de juiste balans tussen *heterogeniteit* en *homogeniteit*. Dit is het gemakkelijkt te verduidelijken door een voorbeeld.

Stel, onze stimuli bestaan uit misdrijven die we paarsgewijs op gelijkenis willen laten beoordelen. We geven de respondenten een beoordelingsschaal die loopt van 1 (maximaal verschillend) tot 7 (identiek) om hun gelijkenisoordelen op uit te drukken. Als een van de stimuli die we gebruiken het misdrijf moord is terwijl de andere stimuli zakkenrollen, zwartrijden, fietsenstelen en vernieling zijn, dan zal het verschil tussen moord en de andere misdrijven alle andere verschillen vrijwel zeker domineren. De paren met moord erin zullen waarschijnlijk gelijkenisscores 1 krijgen, terwijl de andere paren scores van 5, 6 of 7 zullen ontvangen. Hetzelfde zal gebeuren in een onderzoek naar de onderlinge gelijkenissen tussen een groep kunstenaars die op één na uit schilders bestaat en één componist bevat. Kortom, één van de stimuli is veel te afwijkend, vergeleken met de rest. De totale groep stimuli is te heteroog of, anders gezegd, een deel van de stimuli is veel te homogeen.

Wat we dus willen is een verzameling stimuli die homogeen is omdat ze tot één verzameling van vergelijkbare objecten behoren, en aan de andere kant voldoende heteroog is om relevante en subtiele verschillen tussen die objecten naar voren te laten komen. Vandaar dat het van het grootste belang is hoe de stimuli geselecteerd worden.

Selectie van stimuli

In principe geldt voor de selectie van stimuli hetzelfde als voor de selectie van respondenten: we willen een representatieve *steekproef* van objecten uit een veel en veel grotere populatie (domein, universum) van potentiële stimuli. En – eveneens in principe – kunnen we daarvoor alle bekende methoden toepassen, bijvoorbeeld *aselecte (random)* of *gestratificeerde steekproeftrekking*. Het

probleem is dat men daartoe zou moeten beschikken over een complete lijst met mogelijke stimuli (te vergelijken met telefoonboeken en registers van de Burgerlijke Stand bij het selecteren van personen). Voor sommige onderwerpen, waaronder beroepen, zijn zulke lijsten inderdaad beschikbaar. Coxon en Jones (1979), bijvoorbeeld, vermelden de *Classification of Occupations 1970* van het *Office of Population, Censuses and Surveys* en een *Dictionary of Occupational Titles* van het *Department of Employment*, beide uit het Verenigd Koninkrijk.

In de vs selecteerde Cattell (1946) uit woordenboeken meer dan drieduizend woorden die gebruikt kunnen worden om persoonlijkheidseigenschappen te beschrijven. Brokken (1978) deed hetzelfde in Nederland en kwam tot een lijst van 1200 van zulke woorden. Op veel terreinen bestaan er atlanten, encyclopedieën, compendia en andere naslagwerken die allerlei opsommingen, classificaties en taxonomieën bevatten.

Niet altijd zijn zulke lijsten van tevoren beschikbaar en moet een onderzoeker eerst zelf empirisch een uitputtende lijst van stimuli proberen op te stellen. Bijvoorbeeld: om een zo compleet mogelijke lijst van pijnbeschrijvende woorden te krijgen hebben Verkes, Van der Kloot en Van der Meij (1989) verschillende methoden toegepast. Onder andere verspreidden zij in de wachtkamers van artsen en fysiotherapeuten vragenlijstjes waarop patiënten verzocht werd hun pijn te beschrijven. Ook vroegen zij een groep van artsen, fysiotherapeuten en verpleegkundigen over een periode van twee weken de woorden te noteren waarmee hun patiënten pijnklachten aanduiden. Falbo (1977) liet een aantal proefpersonen opstellen schrijven ('*How I get my way*') over de strategieën die zij toepasten om hun zin te krijgen. Deze opstellen bevatten in totaal 346 van zulke strategieën, waarvan 91 procent door drie experts herleid kon worden tot 16 basistypen. Van deze basistypen werden vervolgens data over de onderlinge nabijheid verzameld.

Als men over een uitgebreide – al of niet zelf samengestelde – lijst van potentiële stimuli beschikt, dan blijft het de vraag hoe men hieruit de stimuli moet selecteren om in een MDS-onderzoek te gebruiken. Heel vaak is dit een tamelijk subjectieve beslissing van de onderzoekers. Al of niet op grond van hun expertise hebben onderzoekers min of meer intuïtieve ideeën over de overeenkomsten en verschillen tussen stimuli en over de daaraan ten grondslag liggende dimensies. De steekproef die zij kiezen, bestaat dan uit objecten die naar hun verwachting evenwichtig verdeeld zijn over die dimensies.

Naast intuïtieve manieren om stimuli te selecteren worden er soms ook meer systematische methoden toegepast om stimuli te selecteren of te construeren. Bijvoorbeeld: in hun onderzoek naar *The images of occupational prestige*, selecteerden Coxon en Jones (1979) beroepen die thuishoorden in de vier kwadranten die gevormd worden door de factoren opleiding (relatief hoog versus relatief laag) en gerichtheid van het werk (voornamelijk op mensen versus voornamelijk op gegevens en machines). Deze *factoriële* manier om stimuli te selecteren heeft ook potentiële bezwaren. Null (1980) is van mening dat

in most cases it is not advisable to factorially choose the stimulus set. If an experimenter varies size ..., color ..., shape..., and orientation ... of geometric figures, he or she should not be surprised if the subjects use size, color, shape, and orientation alone in making their judgments (p. 278).

Deze opmerking is alleen relevant als men via MDS nieuwe dimensies op zou willen sporen waar men zelf nog niet aan gedacht heeft. Vaak echter kan een onderzoeker *a priori* een heleboel dimensies opnoemen, maar gaat het onderzoek erom welke dimensies door naïeve proefpersonen worden ‘waargenomen’ of ‘gebruikt’. Dat laatste is ook het geval in de *facet-theorie*, die in Blok 12.1 besproken wordt. Een andere methode is Kelly’s *repertory grid* (1955) die verderop in dit hoofdstuk wordt behandeld.

Eén belangrijke opmerking moet hier nog gemaakt worden. Niet alleen moeten de stimuli representatief zijn voor het universum van alle stimuli, ze moeten ook representatief zijn voor het repertoire van kennis en gedragingen van de onderzochte proefpersonen. Het heeft geen enkele zin om stimuli te gebruiken die voor de meeste respondenten volstrekt onbekend zijn of waar zij slechts gebrekkige ervaring mee hebben.

BLOK 12.1 FACET-THEORIE

Meer dan ‘alleen maar’ een methode om stimuli te selecteren is facet-theorie (FT) een complete methodologie, een theorie voor het doen van sociaal-wetenschappelijk onderzoek (Borg & Shye, 1995).¹ De FT-methodologie bestaat uit drie onderdelen: design, data-analyse en zogenaamde correspondentiehypothesen die de verbinding tussen ontwerp en analyse vormen. De kerngedachte van FT is dat de fenomenen die men wil onderzoeken, in meerdere opzichten van elkaar verschillen en dus volgens verschillende gezichtspunten of aspecten kunnen worden ingedeeld, geassocieerd. Zulke aspecten heten in de FT *facetten*. Een facet is een verzameling categorieën waarin objecten, taken of personen op een zinvolle manier gepartitioneerd kunnen worden. Bijvoorbeeld: een populatie van personen is in te delen (te stratificeren) volgens het facet geslacht (Borg & Shye, 1995).

Naast facetten die personen in klassen indelen, zijn er facetten die onderscheidingen tussen stimuli aanbrengen. Het kenmerkende van de FT is nu dat onderzoekers *a priori* een aantal facetten bedenken of aanwijzen die volgens hen op het onderwerp van toepassing zijn. Hierdoor willen zij

¹ Om die reden is het onmogelijk om deze theorie kort te bespreken zonder haar geweld aan te doen. Daar komt bij dat FT ‘in a sense ... is a rather idiosyncratic approach, because it relies on a specific language that not many people outside the International Facet Theory Society are using’ (De Leeuw, 1995, p. ix).

structuur aanbrengen in een tot dan toe ongestructureerd onderzoeksdo-
mein (Borg & Shye, 1955). Bijvoorbeeld: in een onderzoek naar meningen
over misdaad kunnen we het universum van misdrijven in eerste instantie
indelen naar de volgende drie facetten:

- *Slachtoffer* (F_1). Wie lijdt schade van het misdrijf: (a) een persoon, (b)
een groep, (c) een instelling, of (d) de overheid?
- *Schade* (F_2). Wat voor schade wordt er veroorzaakt: (a) lichamelijke,
(b) geestelijke, (c) materiële, of (d) financiële?
- *Geweld* (F_3). Is er sprake van geweld: (a) ja of (b) nee?

Deze lijst van facetten kan gebruikt worden om richting te geven aan de
selectie van stimuli. Samen hebben de drie facetten $4 \times 4 \times 2 = 32$ combi-
naties van categorieën (zogenaamde *structuples*) en de onderzoekers moe-
ten nu voor iedere combinatie een geschikt misdrijf zoeken of
construeren. Bijvoorbeeld: voor structuple (c, d, b) ‘oplichting van een
verzekeringsbedrijf’, voor (a, a, a) ‘iemand wordt in elkaar geslagen’, voor
(d, c, b) ‘het stadhuis wordt beklad’.

Merk op dat de categorieën van sommige facetten nog verder verfijnd
kunnen worden, bijvoorbeeld door bij F_1 toe te voegen of het slachtoffer
een man of een vrouw is. F_2 bevat een aantal categorieën die elkaar niet
uitsluiten. Een slachtoffer kan immers zowel lichamelijke, geestelijke,
materiële als financiële schade oplopen. F_2 kan dus beter vervangen wor-
den door vier nieuwe facetten geestelijke schade (ja/nee), lichamelijke
schade (ja/nee), enzovoort.

Het voordeel van deze facet-aanpak is dat we al van tevoren beschikken
over informatie die ons kan helpen bij de interpretatie van de analysere-
sultaten. In een MDS-plaatje van 32 op bovengenoemde manier geconstru-
eerde misdrijven kunnen we bij elk stimuluspunt de categorieletters van
de drie facetten noteren (zie Hoofdstuk 7). Daardoor krijgen we inzicht of
en zo ja, op welke manier de ruimte van de MDS-oplossing gepartitioneerd
wordt door de drie facetten.

Als we met behulp van FT een aantal stimuli geconstrueerd hebben, moe-
ten we ze voorleggen aan een aantal respondenten, bijvoorbeeld om ze
paarsgewijs te laten beoordelen op ernst of gelijkenis. Ook kunnen we ze
laten sorteren of laten beoordelen op een aantal beoordelingsschalen. In
het typische FT-onderzoek gebeurt meestal het laatste, zij het ook weer op
een bijzondere FT-manier. Stel dat we bij een groep respondenten willen
nagaan op welke manieren de daders van misdrijven volgens hen het best
aangepakt zouden kunnen worden. We kunnen dan nog een of meer facet-
ten definiëren die zulke manieren van aanpak beschrijven. Bijvoorbeeld
Aanpak van dader (F_4): (a) gevangenisstraf, (b) huisarrest, (c) maatschap-
pelijke dienstverlening, (d) geldboete, (e) schadevergoeding aan slacht-
offer, (f) smartegeld voor slachtoffer, (g) heropvoeding van dader. Daarbij

is het nog van belang of het gaat om *Recidive* (F_5): (a) eerste misdrijf van de dader, (b) recidive. De 32 structuples van misdrijven worden dan uitgebreid tot $32 \times 7 \times 2 = 448$ combinaties. Elke structuple zouden we in vraagvorm kunnen voorleggen aan een aantal respondenten. Bijvoorbeeld: 'In hoeverre vindt u gevangenisstraf noodzakelijk voor een dader die als eerste misdrijf een verzekeringsbedrijf heeft opgelicht?'. De antwoorden kunnen dan geregistreerd worden op een beoordelingsschaal die loopt van 1 'volstrekt onnodig' tot 9 'absoluut noodzakelijk'. Deze antwoordmogelijkheden zijn de elementen van een *respons*-facet (R_1). Bovenstaande vraag is één van de 448 vragen die gegenereerd kunnen worden uit wat in de FT een *mapping sentence* genoemd wordt. Een van de vele mogelijke formuleringen van zo'n mapping sentence luidt in dit voorbeeld: Respondent $\{p\}$ vindt het $\{element\ van\ R_1\}$ dat een dader met kenmerk $\{element\ van\ F_5\}$ die een misdrijf gepleegd heeft met als slachtoffer $\{element\ van\ F_1\}$ en schade $\{element\ van\ F_2\}$ en daarbij $\{element\ van\ F_3\}$ geweld heeft toegepast, bestraft wordt met maatregel $\{element\ van\ F_4\}$.

Zoals gezegd, genereert deze mapping sentence 448 vragen. Dat zijn er te veel om in één onderzoek op te nemen. De onderzoekers zullen dus een of andere selectie moeten toepassen om een onderzoekbaar aantal te krijgen. Een voor de hand liggende mogelijkheid is om slechts één maatregel (bijvoorbeeld huisarrest) te onderzoeken. We houden dan 64 vragen over. Stel nu dat er n respondenten zijn die deze 64 vragen beantwoorden. In de typische FT-aanpak worden de antwoorden van de proefpersonen dan gebruikt om correlaties tussen de vragen te berekenen en vervolgens wordt de correlatiematrix met niet-metrische MDS geanalyseerd (in Hoofdstuk 13 wordt nader ingegaan op MDS van correlatiecoëfficiënten). Dit levert dus een meerdimensionale afbeelding van structuples die aan de hand van hun elementen geïnterpreteerd kunnen worden.

De grondslagen van FT zijn zo'n veertig jaar geleden bedacht door Louis Guttman, een naam die al enkele malen in dit boek is voorgekomen. Een van de artikelen waarin hij de FT uit de doeken doet is Guttman (1971). Andere literatuur is te vinden in Borg en Shye (1995).

Selectie van condities

Meestal worden de gegevens in een MDS-onderzoek onder één conditie verzameld die voor alle respondenten dezelfde is. Dat hoeft natuurlijk niet. Het is mogelijk om dezelfde verzameling stimuli onder verschillende omstandigheden aan verschillende groepen respondenten voor te leggen om later na te gaan of deze omstandigheden van invloed zijn op de MDS-configuraties van de verschillende groepen. De omstandigheden vormen dan de condities van een experiment; het zijn de niveaus van een experimenteel gevarieerde onafhankelijke variabele, een *treatment factor*. De condities kunnen bestaan uit verschil-

lende tijdstippen waarop het onderzoek wordt uitgevoerd (bijvoorbeeld het beoordelen van de gelijkenis tussen stimuli door patiënten vóór en na therapie), maar ook uit verschillende opdrachten die aan de respondenten gegeven worden (bijvoorbeeld het paarsgewijs beoordelen van misdrijven met betrekking tot enerzijds de ernst van de gevolgen voor het slachtoffer en anderzijds de ernst van de gevolgen voor de maatschappij). Ook kan men denken aan verschillende vormen van aanbidding van de stimuli of de taken.

Eén voorbeeld van verschillende aanbiedingscondities is het onderzoek van Miller en Nicely (1955) die data verzamelden over de verwarring tussen zestien Engelse medeklinkers. De proefpersonen in dit experiment luisterden naar vrouwelijke sprekers die in gerandomiseerde volgorde éénlettergrepige klanken voorlezen; hen werd gevraagd om na elke voorgelezen lettergreep op te schrijven welke medeklinker zij gehoord meenden te hebben. Er waren 17 experimentele sessies, waarin telkens de spraaktransmissie op verschillende wijze werd verstoord. Naast de standaardconditie met een frequentiebereik van 200 tot 6500 Hz en een signaal/ruisverhouding van 18 dB waren er vijf condities met lagere signaal/ruisverhouding (12, 6, 0, -6, -12 dB), zes condities waarin geluiden boven een bepaalde frequentie werden uitgefilterd en vijf condities waarin geluiden beneden een bepaalde frequentie niet werden doorgelaten. In elk van de zeventien experimentele condities werd geteld hoe vaak de 16 uitgesproken medeklinkers correct werden herkend en hoe vaak zij met elkaar werden verward.

De selectie van condities wordt uiteraard volledig bepaald door de inhoudelijke hypothesen die onderzoekers op een bepaald terrein erop nahouden. Daarover valt hier verder weinig te zeggen. De belangrijkste richtlijn is dat de verschillen tussen de condities zodanig gekozen worden dat ze sterk genoeg zijn om eventuele effecten aan te kunnen tonen.

12.4 SELECTIE VAN RESPONDENTEN

Wat de selectie van respondenten betreft, beperkt dit boek zich tot enkele algemeenheden. Het belangrijkste is natuurlijk dat de onderzochte personen representatief zijn voor de populatie waarnaar men wil generaliseren. Willen we uitspraken doen over vrouwen of over mannen; alleen over hoogopgeleide jongeren of ook over anderen? Het ligt voor de hand dat we onze steekproef van respondenten aan de hand van dat soort vragen moeten samenstellen. Ook bij de selectie van respondenten gaat het om een goede balans tussen heterogeniteit en homogeniteit van de steekproef. De steekproef moet homogeen zijn, in die zin dat er alleen exemplaren uit de beoogde populatie in zijn opgenomen. Aan de andere kant moet de steekproef heterogeen zijn, in die zin dat de relevante verschillen tussen de exemplaren van de populatie ook in de steekproef vertegenwoordigd zijn.

Bovengenoemde vorm van heterogeniteit is des te belangrijker, als we op zoek zijn naar individuele verschillen tussen personen. Willen we zulke verschillen

in een MDS-oplossing aan kunnen tonen, dan moeten de respondenten ook in voldoende mate op allerlei relevante eigenschappen van elkaar verschillen. Liever dan één aselechte steekproef te trekken en af te wachten in welke mate de respondenten van elkaar verschillen, is het in dit type onderzoek beter om specifieke subpopulaties te definiëren waaruit men aselechte steekproeven trekt. Bijvoorbeeld: bij onderzoek naar kleurenperceptie door normaalzienden en personen met bepaalde vormen van kleurenblindheid is het aan te bevelen afzonderlijke steekproeven van normaalzienden en van kleurenblinden te trekken, anders loopt men het risico dat er in de groep respondenten geen of nauwelijks mensen met kleurenblindheid zitten.

12.5 DATAVERZAMELINGSPROCEDURES

De procedures waarmee men data kan verzamelen zijn uiteraard afhankelijk van het soort data dat men uiteindelijk wil analyseren. Om in de volgende paragrafen enige systematiek aan te brengen zouden we uit kunnen gaan van de indeling van Carroll en Arabie (1981; zie Hoofdstuk 3) waarin onderscheid gemaakt wordt tussen tweeweg/éénmodale, drieweg/tweemodale, tweeweg/tweemodale en drieweg/tweemodale nabijheidsdata. In de praktijk worden de verschillende datamatrices van de derde weg van een driewegmatrix meestal op precies dezelfde manier verzameld: bijvoorbeeld doordat men één en dezelfde dataverzamelingsmethode niet bij één, maar bij verschillende proefpersonen toepast. Of: de verschillende datamatrices bestaan uit hetzelfde soort observaties die op verschillende tijdstippen, eventueel onder verschillende condities, gedaan zijn. Het maakt voor de manier van dataverzameling niet uit of we de observaties eenmalig of met replicaties vastleggen. Daarom kunnen we hier een eenvoudiger indeling gebruiken, waarin we slechts onderscheid maken tussen tweeweg/éénmodale en tweeweg/tweemodale datamatrices, beide al of niet met replicaties.

Tweeweg/éénmodale nabijheidsdata

Bij tweeweg/éénmodale nabijheidsdata beschikken we over een matrix met de onderlinge similarities of dissimilarities van één verzameling objecten. Zo'n matrix kan op verschillende manieren verkregen worden. Een (flauw) voorbeeld is de afstandentabel voor steden in Nederland en Europa die in bijna iedere agenda is afgedrukt. Andere voorbeelden zijn de vele correlatiematrices en kruistabellen die in, al dan niet wetenschappelijke, publicaties reeds gepubliceerd zijn. Maar het opsporen van bestaande datamatrices is niet wat we hier onder dataverzameling verstaan. Waar het in dit hoofdstuk om gaat, zijn dataverzamelingsprocedures om met behulp van bestaande documenten, producten of proefpersonen een matrix met (dis)similarities te verkrijgen. Over de rol van proefpersonen en documenten is nog het volgende te zeggen. Zoals we hieronder zullen zien, is er in archieven, bibliotheken en losse publicaties allerlei materiaal te vinden dat als *bron* kan dienen voor diverse vormen

van nabijheidsdata. In dat geval zijn de gegevens afkomstig uit bestaande documenten. Veel documenten zijn niet aan één of meer specifieke personen toe te schrijven, dat wil zeggen: ze bevatten informatie die nauwelijks iets met de auteurs te maken heeft. Een voorbeeld van zulke documenten zijn wetboeken, die wel een beeld van een land geven, maar niet van de personen die de boeken geschreven hebben. Een ander soort documenten, met name brieven, romans, gedichten en ook wetenschappelijke werken, is wel tot personen te herleiden en bevat allerlei expliciete en impliciete informatie over de auteurs. Maar ook in dit geval gaat het om bestaande documenten die onafhankelijk van de onderzoeker geproduceerd zijn.

Een totaal ander type documenten wordt verkregen als een onderzoeker zelf een aantal personen – speciaal voor zijn onderzoek – benadert. De onderzoeker kan die personen vragen hun meningen, gedachten en gevoelens weer te geven, bijvoorbeeld door middel van een opstel. Ook kan een onderzoeker de personen verzoeken een of ander product te vervaardigen (bijvoorbeeld een tekening, een knipselwerk, een kleifiguur) dat door de onderzoeker op verschillende manieren beschreven kan worden. In dat geval is er sprake van documenten of producten die speciaal ten behoeve van de onderzoeker vervaardigd zijn en die gebruikt kunnen worden om nabijheidsgegevens uit af te leiden. De personen die aan het onderzoek meewerken – de proefpersonen – zijn er in dit geval niet van op de hoogte dat het de onderzoeker uiteindelijk om nabijheidsmaten te doen is. Een ander geval is een onderzoek waarin de proefpersonen gevraagd wordt expliciet de onderlinge gelijkenissen van een aantal objecten te beoordelen.

Bestaande documenten

Een type data dat in de laatste decennia al vaker verzameld en met MDS geanalyseerd is, betreft de wederzijdse citaties van wetenschappelijke tijdschriften. Het eerste onderzoek is dat van Coombs (1964) die van tien psychologische tijdschriften noteerde hoe vaak zij in hun artikelen van jaargang 1960 refereerden naar eerdere publicaties in hun eigen tijdschrift en in de andere tijdschriften. Bijvoorbeeld: in de gezamenlijke artikelen van het *Psychological Bulletin* van 1960 kwamen in totaal 1779 referenties aan eerder gepubliceerde artikelen voor. Hiervan waren er 684 die aan een van de tien onderzochte tijdschriften refereerden; 1095 referenties verwezen naar andere tijdschriften. Van de 684 relevante referenties verwezen er 52 naar het *Psychological Bulletin*; de rest refereerde aan de overige negen tijdschriften. De resulterende matrix met tellingen

seems ... on a priori grounds to be a reasonable criterion of similarity. This is very literally a count of the extent to which a journal 'uses' another journal. The more one journal uses another the more like it it must be (Coombs, 1964, p. 463).

Hetzelfde onderzoek heeft Coombs (Coombs, Dawes & Tversky, 1970) herhaald, met jaargang 1964 van acht van de tien eerdere tijdschriften (zouden we de data van de jaargangen 1960 en 1964 achter elkaar zetten, dan ontstaat een

drieweg/tweemodale matrix die geanalyseerd zou kunnen worden om veranderingen in de tijd op te sporen).

Soortgelijke data zijn verzameld door Jaspars en Ackermans (1965) met betrekking tot de referenties in drie jaargangen van een representatieve verzameling psychologische, sociologische en sociaal-psychologische tijdschriften. Het zal duidelijk zijn dat deze referentietellingen asymmetrische matrices opleveren. Als men dat zou willen vermijden, dan kan men als alternatief voor verwijzingen *van* het ene *naar* het andere tijdschrift noteren hoe vaak binnen elk artikel een referentie aan het ene tijdschrift samengaat met een referentie aan een ander tijdschrift. We negeren dan uiteindelijk het tijdschrift van waaruit verwezen wordt en tellen alleen de *naar*-verwijzingen. In dat geval krijgen we vanzelfsprekend een symmetrische matrix met zogenaamde *co-occurrences*, frequenties van samen voorkomen.

Deze methode – het tellen hoe vaak dingen samen voorkomen – is in principe op alle mogelijke documenten en producten toe te passen. Een opmerkelijke toepassing is die van Rosenberg en Jones (1972) die een analyse uitvoerden op beschrijvingen van de personen die voorkwamen in een aantal literaire portretten in het boek *A gallery of women* van de Amerikaanse schrijver Theodore Dreiser. Uit deze portretten extraheerden zij '*every linguistic unit that described a stable characteristic of each person mentioned ...*' (p. 373). In totaal verzamelden zij 6761 van zulke eenheden die door Dreiser aan in totaal 241 personen waren toegeschreven. Hieruit leidden zij 99 categorieën van persoonlijkheidskenmerken (*traits*) af die dezelfde betekenis hadden (bijvoorbeeld *beauty* en *beautiful* werden in één categorie gestopt) en het vaakst (tien of meer keer) voorkwamen. Op deze manier ontstond er een 241×99 -matrix van personen bij categorieën, met frequenties die aangaven hoe vaak een bepaalde categorie gebruikt was om een bepaalde persoon te beschrijven. Uitgaande van deze matrix werden verschillende nabijheidsmaten berekend, waaronder een zogenaamde *direct co-occurrence index* (Rosenberg & Jones, 1972). Deze index is een functie van hoe vaak twee trait-categorieën samen gebruikt zijn in de beschrijving van de 241 personen. In een latere fase van het onderzoek zijn deze indices gebruikt om via MDS een configuratie van trait-categorieën te verkrijgen.

Documenten van proefpersonen

Een toepassing van co-occurrence data is te vinden bij Marx (1987) die acht groepjes van dertig studenten opdracht gaf gedurende zestig seconden vrij te associëren over Duitse politieke partijen (zoals CDU, CSU, FDP). Voor iedere partij werden relatieve frequenties van de associaties vastgesteld (bijvoorbeeld twee procent van de proefpersonen associeerde partij A met kenmerk X) en vervolgens werd voor elk tweetal partijen een overlappingscoëfficiënt berekend doordat men de som nam van de percentages *gemeenschappelijke* associaties. Deze overlappingscoëfficiënten werden daarna met MDS geanalyseerd. In een onderzoek van Heij en Marx (1987) werden 132 studenten uit vijf ver-

schillende landen geïnstrueerd om alle Europese staten die zij kenden onder elkaar op te schrijven. Om uit deze rijtjes subjectieve landkaarten van Europa te reconstrueren, gebruikten de onderzoekers de verschillen in de volgorden waarmee de staten genoemd waren. Als bijvoorbeeld een proefpersoon Nederland als eerste en Duitsland als derde staat opgeschreven had, dan was voor die proefpersoon de ‘afstand’ tussen Nederland en Duitsland gelijk aan twee. Deze afstandsmaten werden opgeteld over de proefpersonen uit één en hetzelfde land. Zo kon voor elk van de vijf nationaliteiten apart via MDS een ‘kaart van Europa’ verkregen worden.

Verwarringsmatrices

Een speciaal geval van een (asymmetrische) co-occurrence matrix is een zogenaamde verwarringsmatrix. Een voorbeeld daarvan zijn de data van het eerder besproken onderzoek van Miller en Nicely (1955). In dat onderzoek werd onder zeventien verschillende condities een datamatrix verzameld met getallen die aangeven hoe vaak elk van de zestien voorgelezen medeklinkers correct herkend is of met een van de andere medeklinkers verward is. De matrix is asymmetrisch, omdat het aantal keren dat medeklinker i voor medeklinker j wordt aangezien niet gelijk hoeft te zijn aan het aantal keren dat j als i wordt waargenomen.

12.6 BEOORDELINGEN DOOR PROEFPERSONEN

In de laatste voorbeelden is de uiteindelijke bron van de nabijheidsdata een groep proefpersonen. Echter, deze personen waren zich er waarschijnlijk niet van bewust dat het in de taken die zij moesten uitvoeren om de nabijheden tussen bepaalde stimuli ging. Dit is anders, als de respondenten expliciet de opdracht krijgen een aantal objecten op onderlinge gelijkenis te beoordelen. Dit soort beoordelingen kan op verschillende manieren verkregen worden: via paarsgewijze vergelijking, met behulp van de triadenmethode, de tetradenmethode, de methode van multiële rangordening en de sorteermethode. Deze methoden zullen we hieronder afzonderlijk bespreken.

Paarsgewijze vergelijkingen

In de methode van paarsgewijze vergelijking worden de stimuli twee-aan-twee aan de respondenten voorgelegd. Aan de proefpersonen kan dan worden gevraagd welke stimulus de ander domineert (bijvoorbeeld groter is, mooier is), of er wordt hun gevraagd met een getal aan te geven in welke mate beide stimuli op elkaar lijken. Een typisch voorbeeld van de *instructie* die proefpersonen krijgen voor het paarsgewijs beoordelen van gelijkenis is door Van der Kloot en Van Herk (1991) gebruikt om nabijheidsdata te verzamelen tussen zestien persoonlijkheidsadjectieven. Elke proefpersoon kreeg een vragenlijst die negen pagina's bevatte met op elke bladzijde een stuk of veertien paren

persoonlijkheidseigenschappen. De eerste bladzijde bevatte onderstaande instructie:

Op de volgende bladzijden ziet u steeds twee eigenschappen staan met daartussen een rijtje getallen die lopen van 1 tot en met 7.

Bijvoorbeeld:

dominant 1---2----3---4---5---6---7 extravert

De bedoeling is dat u aangeeft in hoeverre u vindt dat deze twee eigenschappen op elkaar lijken. Dit kunt u doen door één van de getallen te omcirkelen, namelijk het getal dat het best uw mening weergeeft.

Als u een 7 omcirkelt, dan geeft u hiermee aan dat volgens u de eigenschap dominant veel lijkt op de eigenschap extravert. Een 6 betekent dat de twee eigenschappen iets minder op elkaar lijken, een 5 weer iets minder, enzovoort, tot het cijfer 1, dat aangeeft dat de eigenschappen dominant en extravert volgens u in het geheel niet op elkaar lijken.

Het gaat bij deze taak steeds om uw persoonlijke mening. Er zijn dan ook geen goede of foute antwoorden.

Er zijn minstens vier belangrijke eisen waar een instructie aan moet voldoen (zie Schiffman e.a., 1981). Ten eerste moet er een duidelijk voorbeeld worden gegeven van de taak die aan de proefpersoon wordt aangeboden. Ten tweede moet helder worden uitgelegd hoe de respondent zijn of haar antwoorden moet geven. Ten derde moet worden duidelijk gemaakt dat er geen 'goede' of 'foute' antwoorden zijn.

Ten vierde moet men in of bij de instructie een overzicht geven van alle of de meest uiteenlopende stimuli die gebruikt worden (wat overigens in bovenstaande instructie ontbreekt). Het is namelijk wenselijk dat de respondent een idee krijgt van de variatie die er in de stimuli aanwezig is, zodat hij of zij daarmee rekening kan houden bij het toekennen van numerieke scores. Stel dat de eerste paren die een proefpersoon te zien krijgt, bestaan uit dominant en twistziek, dominant en koud en twistziek en arrogant. De proefpersoon zou dan het idee kunnen krijgen dat alle gebruikte adjectieven minder 'aardige' eigenschappen betreffen. Dominant, twistziek, koud en arrogant zullen in dat geval als minder op elkaar lijkend beoordeeld worden, dan wanneer de proefpersoon weet dat er ook 'aardige' eigenschappen bij zitten. Dominant, twistziek, koud en arrogant zullen in het laatste geval als meer op elkaar lijkend worden waargenomen, namelijk omdat zij vergeleken worden met paren als dominant en warm, en twistziek en gezellig.

Een variant op de opdracht om stimuli op hun gelijkenis te beoordelen, is die waarin de respondenten geïnstrueerd worden aan te geven in hoeverre de eerste stimulus van het paar de tweede impliceert. In onderzoek naar persoonlijk-

heidskenmerken kunnen we de proefpersoon vragen aan te geven ‘hoe groot de kans is dat iemand met eigenschap *A* ook eigenschap *B* bezit’ (zie Hays, 1958; Van der Kloot, 1975). Of, zoals Ekman onder andere vroeg in zijn onderzoek naar kleurperceptie ‘hoeveel procent van Stimulus *A* is ook aanwezig in Stimulus *B*?’ De vraag van Hays naar *conditionele waarschijnlijkheden* van het type $P(B/A)$ en Ekmans ratio-methode leveren rijconditionele nabijheidsdata op, die dus ook als zodanig geanalyseerd moeten worden. Qua conditionaliteit lijken deze data op de gegevens die de methode van multiële rangordeningen oplevert. Merk op dat er bij m stimuli in deze conditionele gevallen $m(m-1)$ paren moeten worden beoordeeld, in plaats van $m(m-1)/2$ paren in het gebruikelijke geval.

Aanbiedingsvolgorde

In een onderzoek waarin de stimuli paarsgewijs worden aangeboden, zijn er twee volgorde-effecten waar men rekening mee moet houden. Het eerste effect betreft de volgorde waarin de twee stimuli binnen een bepaald paar worden aangeboden. Bijvoorbeeld: twistziek – warm of warm – twistziek. Het ligt natuurlijk voor de hand om de stimulusaanbieding zodanig te balanceren, dat elke stimulus in de helft van het aantal paren als eerste en in de andere helft van de paren als tweede gepresenteerd wordt.

Het tweede volgorde-effect betreft de opeenvolging van de verschillende paren. Hier ligt het voor de hand om niet eerst alle paren met twistziek aan te bieden, vervolgens alle overige paren met warm en daarna alle resterende paren met dominant, enzovoort. Ook hier moet een gebalanceerde volgorde gezocht worden die erop neerkomt dat de paren waarin een bepaalde stimulus voorkomt, zover mogelijk van elkaar verwijderd en op gelijke afstanden van elkaar worden aangeboden. Daarbij moeten regelmatig terugkerende patronen worden vermeden. Ross heeft in 1934 al een methode bedacht om de aanbieding van paren en stimuli te balanceren. Cohen en Davison (1973) publiceerden een computerprogramma om zulke Ross-volgorden te genereren.

Het aantal paren

Het probleem met paarsgewijze aanbieding van stimuli (of het nu over gelijkenis of preferentie gaat) is dat een kleine toename in het aantal stimuli een grote toename in het aantal paren tot gevolg heeft. Bij 10 stimuli zijn er 45 paren, bij 15 stimuli zijn er 105 en bij 20 stimuli zijn er al 190 paren. Afhankelijk van het soort stimuli dat beoordeeld moet worden (concepten, smaken, geuren) ligt het maximum aantal paren dat men een respondent in één sessie kan aanbieden ergens tussen de 50 en de 150. Dat betekent dat men maximaal 11 tot 17 stimuli zou kunnen onderzoeken. In de woorden van Dunn-Raskin (1983): ‘... *it becomes increasingly fatiguing to respond to more than 15 objects that are paired in all possible ways* (p. 16)’. Meer stimuli en dus meer paren vereisen speciale maatregelen.

Een voor de hand liggende maatregel is het onderzoek per proefpersoon over verschillende sessies uit te smeren. De nadelen van deze aanpak liggen even-

eens erg voor de hand. In de eerste plaats is het maar de vraag of alle proefpersonen inderdaad op alle sessies aanwezig zullen zijn. Bovendien zou het kunnen zijn dat allerlei tijdelijke omstandigheden (stemming, lichamelijke conditie) de beoordelingen van een proefpersoon van sessie tot sessie beïnvloeden. Dit is nooit echt uit te sluiten. De enige tegenmaatregelen bestaan eruit de proefpersoon iedere keer opnieuw een goede gestandaardiseerde instructie te geven, de volgorde van de paren volledig over de sessies te randomiseren, en enkele paren stimuli in iedere sessie opnieuw te laten beoordelen. Op die manier kan men tenminste empirisch nagaan of de verschillende sessies van invloed zijn op de beoordelingen.

Een andere aanpak bestaat uit het aanbieden van slechts een deelverzameling van alle paren aan alle proefpersonen. Als men bereid is de verschillende personen als replicaties van elkaar op te vatten, dan kan men verschillende deelverzamelingen van alle paren aan verschillende groepen proefpersonen voorleggen en achteraf de observaties in één matrix samenvoegen. Ook dan is het belangrijk om de verschillende respondentgroepen toch een aantal zelfde paren te laten beoordelen. Men kan dan nagaan of er zich systematische verschillen tussen de groepen voordoen.

In weer een andere aanpak wordt slechts één deelverzameling van de paren aan alle respondenten aangeboden. De matrix met nabijheidsdata is dus *incomplete*, er zijn *missing data*. Het idee hierachter is dat niet alle paren nodig zijn om toch een goede weergave (*recovery*) van de stimulusconfiguratie te krijgen. Er bestaan talloze manieren om een deelverzameling paren te selecteren. Spence en Domoney (1974) hebben echter laten zien dat *random* selectie van stimulusparen door de bank genomen een even goede *recovery* oplevert als andere selectiemethoden. Twee belangrijke factoren die de *recovery* bepalen, zijn de proportie paren die geselecteerd worden en de hoeveelheid *error* in de data. Het ziet ernaar uit dat een *random* selectie van ongeveer zestig procent van de stimulusparen bij één proefpersoon meestal al een goede oplossing geeft. Als er data bij pakweg twintig proefpersonen verzameld worden, dan is zo'n veertig procent van de stimulusparen vaak al voldoende (MacCallum, 1979). Bovendien hebben Ross en Cliff (1964) aangetoond dat als m objecten in een r -dimensionale ruimte liggen, elk punt gelokaliseerd kan worden, als we de afstanden weten tussen dat punt en $r + 1$ andere punten. In totaal hoeven er dan slechts $(m - r - 1)(r + 1) + (r + 1)r/2$ paren beoordeeld te worden.

Een andere aanpak stoelt op de gedachte dat sommige paren minder belangrijk zijn voor de uiteindelijke oplossing dan andere. Met name de paren van stimuli die (zeer) veel op elkaar lijken, dragen minder bij aan de MDS-oplossing dan stimuli die grote 'afstanden' tot elkaar hebben. Vaak hebben onderzoekers van tevoren gefundeerde ideeën over de gelijkenis en de dimensionaliteit van hun stimuli, waardoor ze zich kunnen laten leiden bij de selectie van de paren die wel en die niet worden aangeboden.

Een variant op de laatste methode, die ook bruikbaar is zonder *a priori* kennis over de configuratie, kan worden toegepast als de stimulusparen worden aangeboden met behulp van een computer die tegelijkertijd de oordelen van de

respondent registreert. Het is dan mogelijk om *interactief* data te verzamelen, dat wil zeggen, de computer stemt de aangeboden stimulusparen af op de antwoorden van elke proefpersoon. Een programma waarin dit gebeurt, is het ISIS-algoritme (*interactive scaling with individual subjects*) van Young en Cliff (1972). Een ander interactief programma is ISO (*interactive similarity ordering*) van Young, Null en Sarle (1978) (zie verderop de paragraaf over multiple rangordeningen).

De tetradenmethode

Bij de paarsgewijze beoordeling van gelijkenis wordt de proefpersonen gevraagd aan elk paar stimuli een numerieke waarde toe te kennen. Om deze getallen met MDS te kunnen analyseren, moeten we aannemen dat ze een functie zijn van de afstanden tussen de stimuli in de configuraties die de proefpersonen 'in hun hoofd' hebben. Meestal nemen we daarbij aan dat het om een ordinale relatie tussen observaties en afstanden gaat. Dat wil zeggen dat de relevante informatie in de data bestaat uit de rangorde van de nabijheidsscores. Als $o_{AB} = 7$ en $o_{CD} = 4$ dan weten we – bij similarity-data – dat $d_{CD} > d_{AB}$, de afstand tussen C en D domineert de afstand tussen A en B. De observaties (of hun rangorde) kunnen dus gebruikt worden om een dominantiematrix te construeren die aangeeft welke afstanden groter zijn dan welke andere afstanden. Bijvoorbeeld, bij vier stimuli met geobserveerde similarities $o_{AB} = 7$, $o_{AC} = 1$, $o_{AD} = 6$, $o_{BC} = 2$, $o_{BD} = 5$ en $o_{CD} = 3$, hoort onderstaande dominantiematrix van afstanden.

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

Merk op dat de rijtotalen corresponderen met de rangorde van de afstanden. Dus, in plaats van de rangorde van afstanden te gebruiken om een dominantiematrix te construeren, kunnen we een dominantiematrix van afstanden gebruiken om de rangorde van die afstanden te bepalen.

Een dominantiematrix van afstanden kunnen we verzamelen met de zogenaamde *tetradenmethode*, dat wil zeggen, door proefpersonen *paren van paren* aan te bieden, bijvoorbeeld Paar AB en Paar AC, en te vragen binnen welk van de twee paren de twee stimuli het meest op elkaar lijken. Zo'n oordeel is waarschijnlijk gemakkelijker en daardoor betrouwbaarder dan een numerieke schatting van de mate van gelijkenis. Het voordeel is bovendien dat een complete dominantiematrix inzicht geeft in de consistentie (een vorm van

betrouwbaarheid) van de oordelen. Zie hiervoor ook Blok 12.2, dat ontleend is aan Meerling (1989, p. 182-183).

BLOK 12.2 CONSISTENTIE EN CIRCULAIRE TRIADEN

De enen en nullen en ook de rijtotalen van een dominantiematrix laten zien of de keuzen van een proefpersoon *transitief* zijn. Er is sprake van transitiviteit als voor elk drietal paren geldt dat als $AB > CD$ en $CD > EF$ dan ook $AB > EF$. Als daarentegen $AB < EF$ dan vormen deze drie paren een *circulaire triade*.

Als er m stimuli zijn, dan bestaan er $q = m(m-1)/2$ paren van stimuli. Als de oordelen van een proefpersoon volledig consistent zijn, dan zullen de rijtotalen alle hele waarden van 0 tot en met $q-1$ aannemen. Er is namelijk één paar dat nooit een ander domineert en één paar dat groter is dan alle $q-1$ andere paren. Als de keuzen van een proefpersoon *volledig inconsistent* zijn dan zullen bij een oneven aantal paren alle rijtotalen gelijk zijn aan $(q-1)/2$ en bij een oneven aantal paren gelijk aan $(q-1)/2 \pm 1/2$. Er bestaat dus een relatie tussen de spreiding van de rijtotalen en het aantal circulaire triaden c . Deze relatie is

$$c = \frac{q(q^2 - 1)}{24} - \frac{1}{2} \sum_{j=1}^q (r_j - \bar{r})^2 \quad [12.1]$$

waarin r_j het rijtotaal van rij j is en \bar{r} het gemiddelde van de rijtotalen is. Een bekende maat voor inconsistentie is

$$K = 1 - \frac{\text{aantal circulaire triaden}}{\text{maximum aantal circulaire triaden}} = 1 - \frac{c}{c_{\max}} \quad [12.2]$$

Bij een oneven aantal stimuli is $c_{\max} = q(q^2 - 1)/24$ en bij een even aantal is $c_{\max} = q(q^2 - 4)/24$. Gegeven de waarden van r en q kunnen we K eenvoudig uitrekenen. Als $K = 1$ zijn de dominantieoordelen volledig consistent; is $K = 0$, dan zijn de oordelen volledig inconsistent.

Het probleem met de tetradenmethode is dat het aantal paren van paren dat met elkaar vergeleken moet worden al gauw gigantisch groot wordt. Van tien stimuli kunnen we $(10 \times 9)/2 = 45$ paren vormen, die samen $(45 \times 44)/2 = 990$ paren van paren opleveren. Vijftien stimuli leiden tot 5460 paren van paren. Dit zijn aantallen die men onmogelijk aan een proefpersoon kan aanbieden. Wil men toch de tetradenmethode gebruiken, dan moet men er wel toe over-

gaan om incomplete data te verzamelen. Zo kan men de totale verzameling tetraden gebalanceerd of aselekt opsplitsen en aan verschillende groepen proefpersonen voorleggen van wie de data later worden samengevoegd. Of men moet 'slimme', interactieve procedures gebruiken om alleen de meest noodzakelijke vergelijkingen te presenteren: bijvoorbeeld door aan te nemen dat iemands oordelen transitief zijn.

Als we eenmaal weten dat $AB > CD$ en $CD > EF$, dan is het niet nodig om AB en EF met elkaar te laten vergelijken. Bij transitiviteit mogen we aannemen dat $AB > EF$. Een andere methode om slechts een deelverzameling paren van paren aan te bieden is de triadenmethode die hieronder wordt besproken.

Als we via de tetradenmethode of varianten daarvan voor verschillende proefpersonen een al dan niet complete dominantiematrix van afstanden hebben verzameld, dan kunnen we die op diverse manieren analyseren. Beschikken we over complete matrices, dan kunnen we voor iedere proefpersoon de rijtotalen berekenen en die gebruiken om een rangorde van de afstanden te bepalen. Die rangordecijfers kunnen we per proefpersoon in een vierkante nabijheidsmatrix noteren, zodat er voor de personen gezamenlijk een drieweg/tweemodale data-matrix met ordinale nabijheidsgegevens ontstaat. In plaats daarvan kunnen we de matrices van de verschillende proefpersonen bij elkaar optellen (ook als ze incompleet zijn). Daardoor ontstaat een dominantiematrix met frequenties die aangeven hoe vaak de afstand tussen het ene paar groter gevonden is dan die van een ander paar. Ook nu zouden we de rijtotalen kunnen berekenen en die totalen in een symmetrische nabijheidsmatrix noteren. Maar in plaats daarvan zouden we ook Thurstones methode kunnen toepassen, dat wil zeggen de frequenties omzetten in proporties en z -waarden en daarvan de rijgemiddelden berekenen. Die rijgemiddelden zijn dan schattingen op intervalniveau van de afstanden tussen de stimuli. Dit is de klassieke aanpak, die door Torgerson (1958) is aangepast om in de triadenmethode (zie hieronder) gebruikt te kunnen worden.

De triadenmethode

In de triadenmethode worden de stimuli in drietallen aan een proefpersoon voorgelegd, met de vraag aan te willen geven welke twee stimuli het meest op elkaar lijken en welke twee het minst. Stel, het gaat om het drietal A , B en C en de proefpersoon vindt AB het meest en AC het minst op elkaar lijken. In termen van de dominantiematrix van afstanden weten we nu dat $AC > AB$, $AC > BC$ en $BC > AB$. Eén triade, die van de proefpersoon twee oordelen vraagt, levert dus drie dominantierelaties op. Dat is betrekkelijk efficiënt, vooral omdat het aantal triaden veel kleiner is dan het aantal paren van paren. Men kan een triade opvatten als twee paren van paren met één gemeenschappelijke stimulus in elk paar, en dat zijn er veel minder dan alle mogelijke paren van paren met en zonder gemeenschappelijke stimuli. Bij m stimuli is het totale aantal triaden gelijk aan $m(m-1)(m-2)/6$. Tien stimuli leveren dus 120 triaden en bij vijftien stimuli zijn er 455; nog steeds veel, maar wel beter hanteerbaar.

Stel dat iemand alle vier triaden beoordeelt die er uit vier stimuli te vormen zijn en de volgende antwoorden ‘geeft’. Triade ABC: A en B lijken het meest op elkaar, A en C het minst; Triade ABD: A en B meest, B en D minst; Triade ACD: A en D meest, A en C minst; Triade BCD: B en D meest, B en C minst. De dominantiematrix van afstanden ziet er dan als volgt uit:

	AB	AC	AD	BC	BD	CD
AB	–	0	0	0	0	?
AC	1	–	1	1	?	1
AD	1	0	–	?	0	0
BC	1	0	?	–	1	1
BD	1	?	1	0	–	0
CD	?	0	1	0	1	–

In deze matrix staat in elke rij een vraagteken, wat wil zeggen dat er voor die combinatie van paren geen gegevens verzameld zijn. Dat klopt ook, want het gaat daarbij om tetraden van vier verschillende stimuli en die worden in de triadenmethode niet aangeboden. Dit betekent dus dat deze dominantiematrix niet zonder meer een volledige rangorde van afstanden oplevert als we gewoon de rijtotalen berekenen. Ook de Thurstonemethode levert – als er meerdere proefpersonen zijn – niet zonder meer een complete rangorde van afstanden. Torgerson (1958, p. 263-268) heeft, voortbouwend op Thurstone, een manier bedacht om triadengegevens van meerdere proefpersonen om te zetten in afstandsschattingen. Deze methode wordt ook gedetailleerd behandeld door Meerling (1988, p. 76 - 77).

Een met die van Torgerson verwante aanpak is de volgende. Uit de dominantiematrix die op grond van de triadendata geconstrueerd is, kunnen we een *rij-conditionele* nabijheidsmatrix afleiden. We zien namelijk dat alle paren waar A in voorkomt met elkaar vergeleken zijn. Hetzelfde geldt voor alle paren met B, met C en met D. Kijken we alleen naar de paren met A, dan zien we dat de afstand tussen A en C groter is dan die tussen A en D, welke laatste weer groter is dan de afstand tussen A en B. Deze gegevens en de gegevens van de paren waar respectievelijk B, C en D in voorkomen, kunnen genoteerd worden in onderstaande matrix.

	A	B	C	D
A	–	1	3	2
B	1	–	3	2
C	3	2	–	1
D	1	2	3	–

Elke rij bevat de rangordeningen van de afstanden tussen de kolomstimuli en de rijstimulus. Zo'n rijconditionele matrix kan direct geanalyseerd worden in ALSCAL (zie Hoofdstuk 6).

Wat we ook nog zouden kunnen proberen – maar wat niet altijd helemaal hoeft te lukken – is de gegevens uit bovenstaande matrix zodanig te combineren dat we zoveel mogelijk inzicht krijgen in de volgorde van alle afstanden. In rij A zien we dat $AC > AD > AB$ en in rij D dat $CD > BD > AD$. Combineren we deze twee reeksen, dan krijgen we de *partiële rangordering* ($AC ? (CD > BD) > AD > AB$). In rij C staat $AC > BC > CD$ zodat we nu weten dat $AC > BC > CD > BD > AD > AB$. De gegevens die in rij B staan, blijken met de laatste volgorde te kloppen zodat we nu een complete rangordering van afstanden hebben afgeleid. Deze rangorde kunnen we in een nabijheidsmatrix noteren die vervolgens matrixconditioneel geanalyseerd kan worden. Maar, zoals gezegd: deze combinatorische aanpak van triadendata zal niet altijd lukken. Soms blijven we met een aantal mogelijke partiële rangordeningen zitten.

Een heel primitieve manier, ten slotte, om triadengegevens om te zetten in afstandsschattingen wordt nog beschreven door Dunn-Rankin (1983, p. 14). Iedere keer dat een paar in een triade als minst gelijkend wordt beoordeeld, krijgt het twee punten. Wordt het paar het meest gelijkend gevonden, dan krijgt het nul punten en wordt het paar niet genoemd, dan krijgt het één punt. Toegepast op bovenstaand voorbeeld krijgt AB in totaal nul punten; AC krijgt er vier, AD één, BC drie, BD twee en CD twee.

BLOK 12.3 KELLY'S REPERTORY GRID

Een procedure uit de psychologische diagnostiek waarin de triadenmethode gebruikt wordt, is de *Role construct repertory test* (REP) van Kelly (1955). Hieronder volgt een korte beschrijving, die ontleend is aan Bannister en Fransella (1971) en Bonarius (1980). In de oorspronkelijke vorm van deze techniek wordt een individu gevraagd de namen op kaartjes te schrijven van tussen de vijftien en dertig personen die bepaalde rollen in zijn of haar leven vervullen. Voorbeelden zijn moeder, vader, gehate leraar, echtgenoot of echtgenote, bewonderde persoon, baas, enzovoort. Uit deze kaartjes wordt een aantal drietallen gevormd. Deze drietallen worden één voor één aan het individu voorgelegd, met de vraag aan te willen wijzen welke twee personen *in een belangrijk opzicht duidelijk op elkaar lijken en daarbij verschillend zijn van de derde persoon*. Vervolgens wordt het individu gevraagd in een paar woorden te zeggen in welk *opzicht* de twee uitgekozen personen op elkaar lijken en in welk opzicht de derde persoon van hen verschilt. Op deze manier verbindt het individu aan elke triade een *construct*, dat wil zeggen, een paar kenmerken die als elkaars tegengestelde fungeren (bijvoorbeeld remmend en stimulerend). In de praktijk moet

een individu meestal zo'n vijftien tot twintig triaden beoordelen. Want, volgens Kelly (geciteerd in Bannister & Fransella, 1971, p. 68)

'Experience shows that only persons with the most complex or schizoid outlooks require more than twenty or thirty [triads] to express their repertory of constructs.'

De REP-test levert dus een verzameling voor het individu unieke rolpersonen en een verzameling unieke constructen, die inzicht geven in hoe het individu de wereld om zich heen structureert. Later werd de REP-test uitgebreid tot de *Repertory Grid* (REPGRID) waarin het individu de opdracht krijgt om alle rolpersonen te beoordelen op alle constructen. Een simpele manier waarop dat kan is door een formulier met verschillende rijen en kolommen te maken. De namen van de rolpersonen worden boven de kolommen genoteerd. In elke rij van het formulier zijn drie hokjes omcirkeld; die geven aan welke drie rolpersonen met elkaar vergeleken moeten worden. Als het individu de twee meest op elkaar lijkende personen heeft aangewezen en het construct benoemd heeft, worden de termen van het construct in een nieuwe kolom in de betreffende rij genoteerd. Vervolgens wordt het individu gevraagd ook alle rolpersonen die niet in de triade voorkwamen op het construct te beoordelen, bijvoorbeeld door de personen waarop de ene term van toepassing is een plus te geven en personen waarvoor de andere term geldt, een min. Daarna wordt de volgende triade aangewezen, en wordt bovenstaande procedure herhaald. Dit gebeurt net zo lang tot het individu alle triaden heeft afgewerkt en alle personen op alle constructen heeft beoordeeld. In plaats van plussen en minnen kan men de proefpersoon ook vragen numeriek te beoordelen (bijvoorbeeld op een schaal van 1 tot 7) in welke mate het construct van toepassing is. Het resultaat van de REPGRID is een matrix van constructen (de rijen) bij rolpersonen (de kolommen) die plussen en minnen (om te zetten in enen en nullen) of numerieke scores bevat. Men kan proberen deze matrix direct te interpreteren of eerst met een of andere data-analytische techniek te analyseren (bijvoorbeeld principale-componentenanalyse, HOMALS, MDS). We krijgen daardoor inzicht in de dimensies waarop de rolpersonen variëren en welke inhoudelijke constructen daarbij van belang zijn. Een REP-test of REPGRID hoeft niet uitsluitend over rolpersonen te gaan. Men kan een individu ook vragen andere *elementen* te beoordelen, bijvoorbeeld gedragingen of situaties. In marktonderzoek kan men proefpersonen triaden van producten aanbieden om de bijbehorende constructen te ontdekken.

Los van de diagnostische waarde die een REPGRID heeft, kan het een handige methode zijn om stimuli te genereren voor een MDS-onderzoek. Wish, Deutsch en Kaplan (1976) hebben deze methode toegepast op een verza-

meling interpersoonlijke relaties (de elementen). Op grond van de REPGRID-resultaten konden zij een aantal belangrijke relaties selecteren en vonden zij tegelijkertijd een aantal variabelen (de constructen) waarop ze de relaties konden laten beoordelen.

Multipele of conditionele rangordeningen

Een manier van dataverzameling die een stuk eenvoudiger toe te passen is dan de triaden- en tetradenmethoden is de methode van multipele rangordeningen. Steeds wordt één van de m stimuli als *standaard* genomen en wordt de proefpersonen gevraagd de overige $m - 1$ stimuli te rangordenen naar hun gelijkenis met de standaard. In totaal moet een proefpersoon dus m keer zo'n rangordeningstaak uitvoeren, namelijk één keer voor alle m standaardstimuli.

De gegevens die deze methode oplevert, kunnen direct in een conditionele nabijheidsmatrix genoteerd en als zodanig geanalyseerd worden. Wil men dat niet, dan kan men op een combinatorische manier proberen een complete rangordening van afstanden af te leiden. Ook kan men uit de conditionele nabijheidsmatrix een dominantiematrix van afstanden construeren. Deze ziet er hetzelfde uit als bij de tetradenmethode. Immers, in iedere rij zijn alle cellen leeg in de kolommen van die stimulusparen die geen stimulus gemeenschappelijk hebben met het paar dat bij de rij hoort. Alle manieren waarop men triadendata kan analyseren zijn dus ook in dit geval te gebruiken.

Hoewel de methode van multipele rangordeningen ook met een relatief groot aantal stimuli veel sneller en gemakkelijker is toe te passen dan de triaden- of tetradenmethode, wordt de taak voor de proefpersoon toch wel complex als het aantal stimuli stijgt. Als er twintig stimuli zijn, dan moet een proefpersoon 20 keer 19 afstanden met elkaar vergelijken; samen dus toch al weer 380 vergelijkingen. Daarom is het de moeite waard om methoden te gebruiken waarin niet alle stimuli als standaard worden aangeboden of niet met alle andere stimuli vergeleken hoeven te worden. Bijvoorbeeld: voor de (conditionele) data die beschreven zijn in de paragraaf over de triadenmethode was het voldoende geweest om A, C en D als standaard aan te bieden. De rangordening van de stimuli ten opzichte van deze standaardstimuli was voldoende om een complete rangorde van alle afstanden te vinden. Het was overbodig om ook nog eens stimulus B als standaard te laten beoordelen. Dit kan men natuurlijk niet van tevoren weten. Welke beoordelingen nodig zijn, hangt af van de antwoorden die de proefpersoon gegeven heeft. Young, Null en Sarle (1978) beschrijven een interactieve methode die het mogelijk maakt om met de computer slechts een deel van de beoordelingen aan de proefpersoon voor te leggen. Hun *interactive similarity ordering* (ISO) is gebaseerd op de methode van multipele rangordening. Eén stimulus wordt als standaard genomen en wordt vergeleken met een deelverzameling van de andere stimuli. Afhankelijk van de antwoorden van de proefpersoon worden nieuwe stimuli met de eerste standaard vergeleken of

wordt een nieuwe standaard gekozen. Op het laatst hoeft de proefpersoon per standaard nog maar enkele vergelijkingen te maken en meestal is het niet nodig om alle stimuli als standaard te gebruiken.

Los van deze interactieve methode kan men het aantal oordelen ook beperken door per standaard niet de opdracht *order* $(m - 2)$ of $(m - 1)$ te geven, maar per standaard alleen de k meest of minst gelijkende stimuli te laten rangordenen. De taak van de proefpersoon wordt dan *order* k of $(m - 1)$. Bij twintig stimuli zouden we dus kunnen vragen alleen de tien meest of minst gelijkende stimuli te rangordenen. Het is natuurlijk een empirische vraag hoe groot k precies moet zijn. Ook is het de vraag of we beter naar de minst gelijkende of de meest gelijkende k stimuli moeten vragen. Intuïtief bekeken lijkt minst gelijkend voor MDS de voorkeur te hebben, omdat MDS-oplossingen vooral de grote afstanden goed proberen af te beelden.

Sorteren

Een methode die uitermate geschikt is om gelijkenisgegevens voor grote aantallen stimuli te verzamelen is de sorteermethode. De namen van de stimuli worden op kaartjes geschreven of met behulp van plaatjes afgebeeld. Elke proefpersoon wordt gevraagd deze kaartjes op stapeltjes te leggen. Dat moet zodanig gedaan worden dat stimuli die op elkaar lijken in hetzelfde stapeltje en stimuli die niet op elkaar lijken in verschillende stapeltjes terechtkomen. Deze methode wordt uitgebreid behandeld in Hoofdstuk 13.

Profielafstanden

Iedere matrix met numerieke scores kan gebruikt worden om zogenaamde profielafstanden (zie ook Hoofdstuk 4) tussen de rijen of tussen de kolommen van die matrix te berekenen. Bijvoorbeeld, in een matrix met beoordelingen van objecten door een aantal proefpersonen, kunnen we profielafstanden tussen de objecten of tussen de personen berekenen. Hebben we een matrix van beoordelingen door één proefpersoon van een aantal objecten op verschillende variabelen, dan kunnen we voor die ene persoon profielafstanden berekenen tussen de objecten en tussen de variabelen. Voor profielafstanden wordt meestal de Euclidische afstandsfunctie gekozen, maar alle andere Minkowsky afstandsfuncties kunnen net zo goed gebruikt worden. Afhankelijk van het soort beoordelingen kunnen nog vele afstandsfuncties worden toegepast (zie PROXIMITIES van SPSS). Bestaan de beoordelingen uit positieve getallen, dan kan men ook zogenaamde χ^2 -afstanden bepalen.

Het soort gegevens waar men in deze gevallen over moet beschikken, bestaat doorgaans uit één tweeweg/tweemodale matrix (bijvoorbeeld proefpersonen bij objecten) of uit meerdere tweeweg/tweemodale matrices (bijvoorbeeld objecten bij variabelen bij proefpersonen). In de volgende paragrafen van dit hoofdstuk zullen we heel kort ingaan op het verzamelen van dit soort gegevens. De verschillende analysemogelijkheden worden besproken in Hoofdstuk 13.

12.7 TWEEWEG/TWEEMODALE GEGEVENS

In feite is iedere tabel waarin de categorieën van twee verschillende variabelen met elkaar gekruist worden een tweeweg/tweemodale datamatrix. Niet alleen in de wetenschappelijke literatuur, maar ook in kranten en publicaties als de *Consumentengids* komen ontelbare aantallen van zulke tabellen voor. Die data hoeven dus niet meer verzameld te worden.

Waar het in de volgende paragrafen om gaat zijn drie verschillende soorten tabellen die we al eerder tegen zijn gekomen: tabellen met keuzedata, preferentierangordeningen en beoordelingen van objecten in termen van meerdere variabelen.

Keuzedata

Een tabel met keuzedata is een matrix met als rijen een aantal respondenten en als kolommen een aantal objecten. Hoewel dit de meest gangbare indeling van zo'n matrix is, kan hij natuurlijk ook getransponeerd worden, zodat de rijen de objecten en de kolommen de personen voorstellen. In de cellen van de matrix staan getallen (vaak enen en nullen) die aanduiden of de betreffende persoon het betreffende object al of niet gekozen heeft.

De term 'gekozen' wordt hier in de ruimste betekenis van het woord gebruikt. Gekozen kan slaan op boeken die scholieren voor hun literatuurlijst kiezen, televisieprogramma's die men bekeken heeft, landen die men zou willen bezoeken, adjectieven die men op zichzelf van toepassing vindt, thema's die in iemands dromen aanwezig zijn, enzovoort, enzovoort. Gezien de oneindige verscheidenheid aan onderwerpen en vragen die men aan iemand kan voorleggen, zijn er nauwelijks algemene aanwijzingen te geven hoe men zulke data in de praktijk precies moet verzamelen. De dataverzameling kan bijvoorbeeld gebeuren met een eenvoudige *check-list* (kruis s.v.p. aan welke kranten u regelmatig leest), met open vragen (schrijf s.v.p. op wie uw lievelingscomponisten zijn), of met vragen (bent u het ermee eens dat ...) gevolgd door de antwoordcategorieën 'ja' en 'nee' (omcirkel s.v.p. het antwoord van uw keuze). In de bovenstaande voorbeelden vullen de respondenten zelf een vragenlijst in of beantwoorden persoonlijk de vragen van een interviewer. Daarnaast kunnen iemands gedragingen die door de onderzoeker worden geobserveerd als keuzedata geregistreerd worden. De onderzoeker noteert bijvoorbeeld of iemand in een groepsdiscussie wel eens suggesties gaf, vragen stelde, lachte, boos werd, enzovoort. Ook kan de keuzematrix het resultaat zijn van een inhoudsanalyse (door de onderzoeker) van brieven, opstellen, antwoorden op open examenvragen, bijdragen aan gesprekken of producten als tekeningen.

Preferentierangordeningen

Een ander type tweeweg/tweemodale data bestaat uit preferentierangordeningen. Er zijn *grosso modo* drie manieren om zulke gegevens te verzamelen: via

kwantitatieve beoordeling van stimuli, via opdrachten om stimuli te ordenen en via paarsgewijze vergelijking.

Kwantitatieve beoordeling. In plaats van na te gaan welke objecten wel en welke objecten niet door respondenten gekozen worden, kan men proefpersonen ook vragen hoe graag zij elk object zouden willen hebben. Het gaat dan om een kwantitatieve beoordeling van de waarde van de objecten. Men zou respondenten bijvoorbeeld kunnen vragen hoeveel (geld, tijd, moeite) zij voor elk object over hebben. Of men kan proefpersonen laten aangeven hoe sterk zij het met bepaalde meningen eens zijn of in welke mate een bewering op hen van toepassing is. Of een observator kan turven hoe vaak een bepaald gedrag door iemand wordt vertoond.

In al deze gevallen ontstaat een matrix van personen bij objecten, met in de cellen niet alleen maar enen en nullen maar een hele reeks van getallen die de sterkte van de relatie tussen persoon en object uitdrukken. Als we van elke proefpersoon zulke scores omzetten in rangnummers, of als we ervan uitgaan dat de getallen niet meer dan ordinale informatie bevatten, dan kunnen we hier spreken van een matrix met preferentierangordeningen.

Rangordenen van stimuli. Een gangbare manier om preferentierangordeningen te verzamelen bestaat uit het aanbieden aan de respondent van een lijst van alle stimuli. De respondent wordt daarbij gevraagd achter de meest geprefereerde stimulus (de mooiste, de grootste, de meest van toepassing zijnde) een 1 te noteren, achter de op één-na-meest geprefereerde stimulus een 2, achter de op twee-na-meest geprefereerde stimulus een 3, enzovoort. In Coombs' (1964) terminologie heet dit een *order m - 1 of m-opdracht*. Hoewel dit in principe een heel duidelijke opdracht is, doen er zich in de praktijk vaak wel problemen voor, met name wanneer er veel (meer dan vijf, meer dan tien?) stimuli geordend moeten worden. Hoe meer stimuli, hoe moeilijker het wordt om een complete rangordening te geven. Het is maar de vraag of een proefpersoon bij dertig stimuli dertig verschillende gradaties van preferentie kan onderscheiden en betrouwbaar weergeven. Vaak hebben personen wel een duidelijke volgorde van voorkeuren voor de drie, vier of vijf stimuli die ze het meest prefereren, maar zien ze eigenlijk geen verschillen tussen de stimuli die daarna komen. Of: de proefpersonen hebben wel een duidelijke volgorde van voorkeuren voor enkele van de meest geprefereerde en voor enkele minst geprefereerde stimuli, maar vinden het ontzettend moeilijk een rangordening onder de middenmoot aan te brengen. Daarom kan het nuttig zijn om varianten van de *order* opdracht te gebruiken, bijvoorbeeld: rangorden de bovenste k uit m , of rangorden de bovenste k en onderste p stimuli. Hoe groot k en p precies moeten zijn, kan het best met behulp van *pilot studies* worden nagegaan. Een andere mogelijkheid is proefpersonen eerst te vragen welke objecten zij zouden kiezen en hen vervolgens alleen de gekozen objecten te laten rangordenen; of vragen 'welke stimulus prefereert u het meest?', 'welke het minst?', 'welke het één-na-meest?', 'welke het één-na-minst?', enzovoort, totdat de proefpersoon zegt het niet meer te weten.

Paarsgewijze vergelijkingen. De meest informatieve manier om preferentierangordeningen te verzamelen is door middel van paarsgewijze vergelijking. De

objecten worden twee-aan-twee aangeboden en de proefpersoon moet bij elk paar aangeven welk object de meeste voorkeur heeft (of: het grootst is, het mooist is, het ergst is). De antwoorden van een proefpersoon worden genoteerd in een dominantiematrix waarvan de rijtotalen de rangorde van de stimuli weergeven. Het voordeel van deze procedure is dat hij niet alleen een rangorde van de stimuli oplevert, maar bovendien inzicht geeft in de mate waarin de voorkeuren van een proefpersoon *consistent* zijn. Aan de hand van de dominantiematrix (en de rijtotalen) kunnen we opsporen in hoeverre iemands voorkeuren *transitief* zijn. In Blok 12.2 is deze kwestie al behandeld.

Problemen van paarsgewijze vergelijking

De problemen van de methode van paarsgewijze vergelijkingen zijn in dit hoofdstuk al uitgebreid behandeld in de paragraaf over paarsgewijze vergelijking van paren. Eén zo'n probleem betreft de aanbiedingsvolgorde van de paren en de volgorde van de stimuli binnen elk paar. Om beide volgorden te balanceren kan men de paren in *random* volgorde aanbieden en ervoor zorgen dat elke stimulus in (nagenoeg) de helft van de paren als eerste en in de andere paren als tweede wordt gepresenteerd. Een vollediger balancering krijgt men door een van de door Ross (1934) ontwikkelde schema's te gebruiken.

Het grootste probleem van paarsgewijze vergelijking is echter dat, zelfs bij een middelmatig aantal stimuli, het aantal paren al gauw heel erg groot is. Bij twintig stimuli horen 190 paren; uit dertig objecten zijn 435 paren te vormen, enzovoort. Voor dit probleem bestaan verschillende oplossingen, die afhangen van het niveau waarop men de data wil gebruiken: *geaggregeerd* of *individueel*. Als de data van een groep respondenten geaggregeerd worden (zoals bij de toepassing van Thurstones methode om schaalwaarden te berekenen) kan men het aantal paren per proefpersoon reduceren door de verzameling paren in blokken op te delen en die door verschillende subgroepen van personen te laten beoordelen. Later kunnen de gegevens dan worden samengevoegd in één dominantiematrix met frequenties (zie Tabel 1.1 in Hoofdstuk 1). Drie andere methoden (Torgerson, 1958, p. 191) om het aantal paren te verminderen zijn de volgende:

- 1 Gebruik een beperkt aantal stimuli (bijvoorbeeld de helft) als *standaard* en laat de andere stimuli alleen met deze standaardstimuli vergelijken.
- 2 Voor als een onderzoeker al ruwweg de rangorde van de stimuli kent: laat alleen die paren met elkaar vergelijken die qua rangorde dicht bij elkaar liggen. Op die manier ontstaat een incomplete dominantiematrix met alleen observaties rond de hoofddiagonaal. Een alternatief is de stimuli aan de hand van hun *a priori* rangorde op te splitsen in een aantal overlappende subgroepen. Elke subgroep kan dan apart geschaald worden en later kunnen de schaalwaarden worden samengevoegd.
- 3 Verdeel de stimuli in enkele niet-overlappende subgroepen. Laat de stimuli binnen elke subgroep met elkaar vergelijken én laat alle stimuli vergelijken met een aantal stimuli die als standaard zijn gekozen. Hierdoor kunnen de schaalwaarden uit de subgroepen tot één schaal herleid worden.

Bovenstaande methoden zijn niet alle drie te gebruiken, als we voor ieder individu afzonderlijk een volledige voorkeursrangordening willen bepalen. De enige methode die in dit geval bruikbaar is, is de tweede: het aanbieden van paren uit overlappende subgroepen stimuli waarvan de volgorde grofweg bekend is. Als we dan aannemen dat de voorkeuren van een proefpersoon transitief zijn, dan kunnen we de observaties die in de dominantiematrix ontbreken meestal gemakkelijk aanvullen. Met deze methode is het mogelijk om alle proefpersonen een van tevoren vastgestelde lijst van paren voor te leggen. Het alternatief bestaat uit het toepassen van *interactieve* methoden. Daarin worden de eerdere beoordelingen van een proefpersoon gebruikt om – onder aanname van transitiviteit – uit te maken of latere paren overbodig zijn of niet.

Beoordelingen van stimuli op beoordelingsschalen

In het overgrote merendeel van sociaal-wetenschappelijk onderzoek wordt op de een of andere manier wel ergens gebruikgemaakt van beoordelingsschalen (*rating scales*). Het klassieke voorbeeld is het onderzoek van Osgood, Suci en Tannenbaum (1957) die de *semantische differentiaal* ontwikkelden. Iedere proefpersoon krijgt een verzameling stimuli voorgelegd en moet door middel van een getal aangeven in welke mate een verzameling kenmerken op de stimuli van toepassing is. De beoordelingsschalen kunnen *bipolair* zijn (bijvoorbeeld mooi 1–2–3–4–5 lelijk) of *unipolair* (bijvoorbeeld mooi: nee 1–2–3–4–5 ja). Dit is een vorm van dataverzameling die ‘het altijd wel doet’, ook al doen zich hier ook wel problemen bij voor (zie bijvoorbeeld Hoogstraten, 1995). Voorbeelden van beoordelingsschalen, hun toepassingen en hun problemen zijn te vinden in vrijwel elk boek over de opzet van sociaal-wetenschappelijk onderzoek (bijvoorbeeld Meerling, 1989; Swanborn, 1994). Beoordeling van stimuli op rating scales kan in principe veel informatie opleveren. Per proefpersoon beschikken we over een matrix met numerieke gegevens die de relaties tussen objecten en eigenschappen weergeven. Twee voor de hand liggende manieren waarop we dergelijke observaties kunnen interpreteren zijn (a) als een voor iedere proefpersoon afzonderlijk verzamelde matrix met ‘voorkeursrangordeningen’ van objecten voor kenmerken, en (b) als een voor ieder kenmerk afzonderlijk verzamelde matrix met ‘voorkeursrangordeningen’ van proefpersonen voor objecten. Deze data kunnen in hun oorspronkelijke driewegvorm of na aggregatie tot tweewegdata geanalyseerd worden. We kunnen de data onder meer gebruiken om allerlei profielafstanden uit te rekenen, bijvoorbeeld, tussen objecten per proefpersoon, tussen variabelen per proefpersoon, tussen objecten per schaal en tussen schalen per object. Deze en enkele andere analysemethoden worden in Hoofdstuk 13 nader besproken.

12.8 DE UITVOERING VAN DE ANALYSES

Gegeven het aantal wegen en modi van de geobserveerde data en rekening houdend met conditionaliteit en meetkenmerken zal men een beperkt aantal schaaltechnieken overhouden, waarmee men de gegevens kan analyseren. Welke technieken precies voor welk type data gebruikt kunnen worden staat uitgebreid beschreven in de Hoofdstukken 3, 6 tot en met 11, en 13 van dit boek; dat hoeft hier niet opnieuw behandeld te worden. Wel is het over het algemeen de moeite waard om één bepaalde dataset niet uitsluitend op één manier te analyseren. Dat geldt vooral als de analyse een explorerend karakter heeft. Vaak kan men een beter inzicht in de eigenschappen van een dataset krijgen als men die data op verschillende manieren analyseert. Door te zien welke uitkomsten invariant zijn en welke uitkomsten van analyse tot analyse verschillen krijgt men een beter idee wat de optimale oplossing moet zijn. Hieronder volgt een opsomming van de aspecten van een analyse die men in ieder geval zou moeten variëren:

- 1 De *dimensionaliteit van de oplossing*. Het is uit den boze om altijd maar voor iedere dataset een tweedimensionale configuratie te kiezen zonder na te gaan of dit inderdaad de beste oplossing is. Dit betekent dat men de *default* instellingen van HOMALS, PRINCALS en ALSCAL in SPSS zelf moet veranderen. Om te kunnen vaststellen wat (naar alle waarschijnlijkheid) de onderliggende dimensionaliteit van de data is, moet men nagaan wanneer de stress bij een toenemend aantal dimensies nagenoeg niet meer verbetert.
- 2 Het *meetniveau*. In het algemeen zal bij een metrische analyse de *fit* lager zijn (de stress is groter) dan bij een niet-metrische analyse van dezelfde dataset. Het kan soms de moeite waard zijn om dezelfde data zowel op ratio-, interval- als op ordinaal niveau te analyseren. Wanneer er zich grote verschillen voordoen, pleit dat aan de ene kant ervoor om uiteindelijk het 'laagste' niveau te kiezen, maar aan de andere kant kan een analyse op 'hoger' niveau soms wel eens veel duidelijkere configuraties opleveren. Dat kan met name het geval zijn als de niet-metrische oplossing tekenen van degeneratie vertoont, dat wil zeggen, als de verschillende punten in een klein aantal compacte clusters samenklonteren.
- 3 Het *meetproces*. Als de data ordinaal zijn en *ties* bevatten, kan men die ties in stand willen houden of toestaan dat de ties verbroken worden. De laatste methode (Kruskals *primary approach*) levert in het algemeen wel een lagere stress op, maar hoeft niet altijd een beter resultaat te geven. Ook het verbreken van ties kan vormen van degeneratie in de hand werken.
- 4 De *conditionaliteit van de data* volgt soms direct uit de manier waarop de data verzameld zijn. In dat geval heeft 'experimenteren' weinig zin. Soms echter zijn er verschillende keuzen mogelijk, bijvoorbeeld in het geval er meerdere matrices met gelijkenisdata verzameld zijn. Het ligt meestal voor de hand om dan een matrixconditionele analyse uit te voeren, maar onconditioneel of rijconditioneel analyseren kan ook.

- 5 Het *afstandsmodel*. Bij drieweg-tweemodale nabijheidsdata wordt tegenwoordig haast standaard het INDSCAL-model gebruikt. Dat is dikwijls niet nodig. Als dezelfde data als R(eplicated)MDS-probleem geanalyseerd worden, blijkt vaak dat er nauwelijks meer stress optreedt dan bij analyse volgens het W(eighted)MDS-model. Dat kan ofwel betekenen dat er weinig of geen individuele verschillen tussen de matrices bestaan, ofwel dat die individuele verschillen niet of nauwelijks door het gewogen-afstandsmodel verklaard worden.
- 6 De *beginconfiguratie*. Omdat de meeste MDS-algoritmen het gevaar lopen in een lokaal minimum terecht te komen, zou men zoveel mogelijk moeten uitsluiten dat dit het geval is. Een van de gangbare methoden daarvoor is de analyse een groot aantal keren te herhalen met steeds andere beginconfiguraties. Een aantal van deze beginconfiguraties kan random gekozen zijn, maar het is ook raadzaam om - als men daarover beschikt - een initiële configuratie te gebruiken die overeenkomt met wat men op theoretische of empirische gronden zou verwachten. Als alle analyses steeds naar dezelfde oplossing convergeren, is het onwaarschijnlijk dat men met een lokaal minimum te doen heeft.

Bij al deze variaties is het belangrijk om steeds het Shepard-diagram te laten afbeelden en na te gaan of bepaalde observaties veel slechter dan andere 'gefit' worden. Ook moet men altijd de grafiek van de transformaties bekijken. Daar aan kan men zien of de transformatiefunctie geen degeneratie-achtige kenmerken heeft. Idealiter moet de transformatiefunctie een vloeiend (*smooth*) verloop hebben met weinig horizontale stukken en liefst zonder plotselinge verticale stijgingen.

12.9 INTERPRETATIE

Een van de belangrijkste beslissingen bij de interpretatie van MDS-oplossingen betreft het aantal dimensies dat men aan moet houden. Daarover is het meeste al in Hoofdstuk 7 gezegd. Om de dimensionaliteit te bepalen moeten we in eerste instantie afgaan op de hoogte van de stress, het stressverloop en de verwachte of via simulatie geschatte hoeveelheid error in de data. Wat echter vaak de doorslag zal geven, is de vraag hoeveel dimensies we inhoudelijk kunnen interpreteren.

In Hoofdstuk 7 hebben we gezien dat MDS-configuraties op verschillende manieren inhoudelijk geïnterpreteerd kunnen worden. Aan de ene kant gebeurt dat vaak door visuele inspectie, waarbij allerlei intuïtieve kennis, ideeën en hypothesen over de objecten als het ware in de oplossing geprojecteerd worden. Aan de andere kant kan men over externe gegevens beschikken en die met de MDS-oplossing in verband brengen. Dat kan grafisch (bijvoorbeeld met behulp van gelijkheidscontouren) en statistisch (bijvoorbeeld door middel van regressie-analyse).

In het algemeen is het aan te bevelen om extra (externe) gegevens te verzamelen over de onderzochte stimuli en eventueel ook over de proefpersonen. Dat

geeft een betere basis voor de interpretatie en laat minder ruimte voor speculatie. In de praktijk zijn er overigens allerlei verschillende manieren waarop men externe gegevens zou kunnen verzamelen en gebruiken. Een aantal wordt hieronder kort besproken.

Objectieve en subjectieve eigenschappen

Soms zijn in MDS-onderzoek van de onderzochte stimuli objectieve eigenschappen bekend. Bijvoorbeeld: bij onderzoek van kleurenstimuli zijn golflengte, helderheid en verzadiging bekend; bij (markt)onderzoek van auto's zijn dat prijs, motorvermogen en afmetingen. Van voedings- en genotmiddelen zijn allerlei ingrediënten bekend (hoeveelheid suiker, vet, alcohol, enzovoort) en van misdrijven staan de strafmaten in het Wetboek van Strafrecht. Al dat soort gegevens kan men grafisch of statistisch in de MDS-oplossingen proberen terug te vinden.

In veel gevallen zijn er van de onderzochte stimuli helemaal geen objectieve eigenschappen bekend of zelfs denkbaar. Wat zijn bijvoorbeeld de objectieve eigenschappen van bedenken, leiding geven, analyseren, data-verzamelen en schrijven ten behoeve van een wetenschappelijk artikel? Op welke variabelen verschillen die activiteiten van elkaar en welke waarden nemen zij aan op die variabelen? Antwoorden op die vragen zijn niet objectief te verkrijgen. Het beste wat we kunnen doen is zelf een aantal mogelijke variabelen bedenken en een groep proefpersonen vragen de stimuli op die variabelen (numeriek) te willen beoordelen. Op die manier krijgen we dus een drieweg datamatrix met externe gegevens, die de scores bevat die verschillende proefpersonen op verschillende eigenschappen aan de objecten hebben toegekend. Bij deze aanpak moet men rekening houden met:

- 1 De *proefpersonen*. Wie moet men vragen de stimuli te beoordelen? Moeten dat dezelfde individuen zijn die ook de data voor de MDS-oplossing geleverd hebben of juist niet? Het antwoord hierop hangt af van het doel van het onderzoek. Willen we 'alleen maar' de configuratie van de stimuli onderzoeken zoals die door deze ene specifieke groep proefpersonen wordt waargenomen, dan ligt het voor de hand om dezelfde proefpersonen ook de beoordelingen te laten geven. Maar als we willen generaliseren naar andere groepen personen, dan hebben we een sterker argument, als de MDS-configuratie van ene groep verklaard kan worden in termen van beoordelingen van de andere groep. Een interessante vraag is dan of de personen van wie we de beoordelingen van de stimuli vragen even 'naïef' moeten zijn als de oorspronkelijke groep of juist veel meer experts op het betreffende onderwerp zouden moeten zijn. Bijvoorbeeld: in een MDS-onderzoek naar misdrijven zouden we externe beoordelingen kunnen vragen aan advocaten, rechters en officieren van justitie. Op die manier krijgen we inzicht in hoeverre de MDS-configuratie van leken overeenkomsten vertoont met de oordelen van specialisten.
- 2 *Aggregeren of niet?* Als er geen objectieve gegevens over de stimuli voorhanden zijn en we daarom wel subjectieve beoordelingen moeten verzamelen, zullen in het algemeen een aantal verschillende personen vragen zulke beoordelingen te

geven. Is het nu beter om die beoordelingen op individueel niveau te gebruiken of moeten we de data over de verschillende personen heen aggregeren, bijvoorbeeld door ze te middelen? Dat laatste is het geval als we ervan uitgaan dat de externe beoordelaars in principe allemaal als replicaties van elkaar opgevat mogen worden en dat individuele verschillen, voor zover die er zijn, niet meer dan toevallige, oninteressante fluctuaties betekenen. Als dat zo is, kunnen we de beoordelingen per eigenschap en per stimulus over de beoordelaars middelen en de gemiddelden van de eigenschappen in de MDS-ruimte afbeelden. De in Hoofdstuk 7 gebruikte scores op evaluatie, activiteit en potentie zijn zulke gemiddelden. Ook het voorbeeld dat hieronder in Blok 12.4 besproken wordt, gaat uit van gemiddelden van beoordelingen.

Soms, vooral als we grote systematische individuele verschillen verwachten, met name wanneer de te beoordelen eigenschappen appelleren aan iemands persoonlijke smaak en eigen voorkeuren, kunnen we de beoordelingen beter niet meteen aggregeren. Het kan juist zeer interessant zijn om de beoordelingen van verschillende proefpersonen afzonderlijk in de MDS-configuratie af te beelden. Zouden we dat via regressie-analyse met behulp van het vectormodel doen, dan krijgen we voor elke eigenschap een bundel pijlen. Elke pijl correspondeert met de richting die zo'n eigenschap volgens één bepaalde persoon in de ruimte inneemt. Als voor een bepaalde eigenschap de bundel pijlen weinig spreiding vertoont, dan zouden we die bundel net zo goed kunnen vervangen door één pijl die de gemeenschappelijke richting van de afzonderlijke pijlen aangeeft. Als de afzonderlijke pijlen van een eigenschap echter wijd uiteenwaaieren, dan heeft samennemen natuurlijk weinig zin. Ook bij afbeelding volgens het ideaalpuntmodel kunnen we voor iedere eigenschap en iedere beoordelaar een apart ideaalpunt lokaliseren. Als de verschillende punten die bij een eigenschap horen dicht bijeen liggen, kunnen we net zo goed één gemeenschappelijk punt (de centroide) kiezen. Liggen de punten door de ruimte verspreid, dan moeten we dat niet doen.

Constrained MDS

Eén mogelijkheid om externe gegevens of *a priori* hypothesen in een MDS-analyse mee te nemen is door middel van *constrained* MDS. Waar het om gaat, is dat aan de oplossing verschillende extra eisen worden opgelegd. Die extra eisen kunnen betrekking hebben op (a) de vorm van de configuratie (bijvoorbeeld alle punten moeten op de omtrek van een cirkel, een ellips, een bol liggen), (b) de waarden van de coördinaten (bijvoorbeeld de coördinaten moeten zodanig gekozen worden dat ze maximaal correleren met scores van de objecten op externe variabelen; of: de coördinaten van verschillende punten moeten op bepaalde dimensies aan elkaar gelijk zijn), en (c) de afstanden in de configuratie (bijvoorbeeld de afstanden tussen bepaalde punten moeten gelijk zijn aan elkaar). Een uitgebreide verhandeling over constrained MDS is te vinden in Meulman en Heiser (1983). Een van de weinige computerprogramma's die

sommige constraints in de analyse mee kunnen nemen is PROXSCAL (zie Busing e.a., 1997).

BLOK 12.4 DE EXTERNE INTERPRETATIE VAN WAARDEN

Jones, Senseing en Ashmore (1978) vroegen 258 proefpersonen een lijstje te maken met minstens vijf persoonlijke en sociale levensdoelen (bijvoorbeeld wereldvrede) en minstens zes manieren om zich te gedragen (bijvoorbeeld hulpvaardig) die zij waardevol vonden om na te streven. In totaal schreven de proefpersonen 3458 verschillende woorden of zinnen op. Daarvan werden de 36 meest genoemde voor verder onderzoek geselecteerd. De frequenties waarmee deze 36 waarden samen in de lijstjes voorkwamen, werden als indicatie van nabijheid gebruikt. Uit deze *co-occurrence* indices werden vervolgens afstandsmaten afgeleid volgens een methode die onder andere door Rosenberg voor dit soort data vaak werd toegepast (zie de paragrafen over sorteergegevens in Hoofdstuk 13). De resulterende afstandsmatrix werd daarna met MDS geanalyseerd, waarbij oplossingen in één tot en met vijf dimensies berekend werden. Om die oplossingen te interpreteren vroegen de onderzoekers groepjes studenten om deze 36 waarden te beoordelen op één tegenstellingspaar (bijvoorbeeld doel versus middel; actief versus passief). Er waren acht tegenstellingsparen waarop de waarden door acht groepjes studenten (112 studenten in totaal) beoordeeld werden. Per groepje werden de beoordelingen gemiddeld en de gemiddelden van de waarden op de afzonderlijke tegenstellingsparen werden door middel van multiële regressie als vectoren in de diverse oplossingen afgebeeld. De coördinaten van de waarden op de oplossingsdimensies waren daarbij de predictoren, de scores op de tegenstellingen waren de afhankelijke variabele. Voor iedere tegenstelling werden dus vijf multiële correlaties berekend, met respectievelijk, één, twee, drie, vier en vijf predictoren. Uit de multiële correlatiecoëfficiënten (R) van de verschillende oplossingen bleek dat de tweedimensionale configuratie 'optimaal' was. Voor de meeste tegenstellingen was de R van de tweede oplossing veel groter dan die van de eerste. Bovendien lieten sommige R 's een afname zien wanneer men van de tweedimensionale oplossing naar de driedimensionale ging. Ook bleken de R 's vaak vanaf de tweedimensionale oplossing statistisch significant te zijn. De externe gegevens werden in dit onderzoek dus niet alleen gebruikt voor de inhoudelijke interpretatie van de MDS-oplossingen, maar ook en vooral om de dimensionaliteit van de optimale oplossing vast te stellen.

Hoewel deze aanpak in principe interessant en bruikbaar is, is enige voorzichtigheid toch op zijn plaats. Zoals Stevens (1992, p. 106) vermeldt, is bij een R -in-de-populatie van nul de verwachte waarde van R^2 gelijk aan

$E(R^2) = k/(n - 1)$ waarin k het aantal predictoren en n het aantal proefpersonen is.

Vertaald naar bovenstaande toepassing komt k overeen met het aantal dimensies en n met het aantal objecten. Voor de vijfdimensionale oplossing is $E(R^2) = 5/35 = .143$ zodat $E(R) = .378$ is. Weliswaar waren bij Jones e.a. de gevonden multiële correlatiecoëfficiënten flink wat hoger dan deze waarde (ongeveer twee keer zo groot), maar in veel MDS-studies gebruikt men minder dan 36 objecten, zodat de verhouding van het aantal predictoren tot het aantal objecten al gauw zo ongunstig wordt dat de R 's niet veel meer zeggen. Ook kan men zich afvragen of het in dit soort toepassingen wel zin heeft om te toetsen of de uiteindelijke R -waarden statistisch significant zijn. Enerzijds is de verhouding van het aantal predictoren tot het aantal objecten meestal zodanig dat de betreffende toets een lage *power* heeft, anderzijds voldoen de coördinaten niet aan een van de belangrijkste assumpties die statistische methoden doorgaans aan observaties stellen, namelijk dat zij onderling onafhankelijk zijn. Bij MDS zijn de coördinaten van de verschillende stimuli niet onafhankelijk, omdat zij allemaal uit dezelfde nabijheidsdata berekend worden.

12.10 RAPPORTAGE

Een van de doeleinden van wetenschappelijke verslaggeving zou moeten zijn dat men, naast de resultaten van het onderzoek, alle handelingen en beslissingen voor de opzet en uitvoering ervan op zo'n manier beschrijft dat het onderzoek door anderen herhaalbaar is. Dat geldt niet alleen voor de keuze van stimuli, vragenlijsten, instructies, de selectie van proefpersonen en de procedure van dataverzameling, maar ook voor alle keuzen die bij de data-analyse gemaakt zijn. Iedere gebruiker van MDS-technieken zal dus moeten beschrijven hoe de data verzameld zijn, hoe deze (eventueel) tot nabijheidsdata herleid zijn en met welk computerprogramma en met welke opties daarvan zij geanalyseerd zijn. Tot die opties horen het gebruikte afstandsmodel, de manier waarop replicaties behandeld zijn, de behandeling van (eventuele) asymmetrie, het veronderstelde meetniveau, de conditionaliteit en de behandeling van *ties*. Wat de resultaten betreft, moeten in ieder geval de stress-waarden van de verschillende analyses vermeld worden, zodat de lezer zelf kan nagaan of de gekozen dimensionaliteit acceptabel is. Problemen die zich in het iteratieve proces hebben voorgedaan (bijvoorbeeld: het algoritme is niet geconvergeerd; er zijn lokale minima gevonden) moeten uiteraard ook vermeld worden.

In de meeste MDS-studies worden de resultaten door middel van grafieken weergegeven. Hoewel zulke grafieken op zich veel inzicht geven, zou het voor de lezer handig kunnen zijn om, behalve afbeeldingen, ook de coördinaten van de punten in een tabel te vermelden. Bij factoranalyse is het immers ook

gebruikelijk om, naast plaatjes, tabellen met factorladingen af te drukken. Ook kan men best – als er slechts één of enkele nabijheidsmatrices geanalyseerd worden – de betreffende tabel(len) met nabijheden weergeven (zoals de correlatietabel waarop factoranalyse wordt toegepast).

In het geval er externe gegevens gebruikt worden voor interpretatie kunnen de resultaten zowel grafisch (bijvoorbeeld gelijkheidscontouren, richtingen in de MDS-configuratie, enzovoort) als in tabelvorm worden weergegeven. In ieder geval moet voor iedere externe variabele de bijbehorende fit (bijvoorbeeld R^2) worden vermeld.

12.11 SIMULATIE

Op verschillende plaatsen in dit boek zijn publicaties genoemd van onderzoek naar de eigenschappen van MDS-technieken die zijn uitgevoerd met behulp van simulatiemethoden. Voorbeelden zijn het onderzoek van Wagenaar en Padmos (1971) en Spence en Graef (1974) waarin verschillende procedures werden onderzocht om de correcte dimensionaliteit van een verzameling nabijheidsdata vast te stellen (Hoofdstuk 7). Andere voorbeelden zijn de studies van Girard en Cliff (1976), Cohen en Jones (1974), MacCallum (1977, 1978), MacCallum en Cornelius (1977), Sherman (1972), Spence (1972), Weinberg en Menil (1993) en Young (1972) naar de *recovery*-prestaties van verschillende MDS-algoritmen onder verschillende omstandigheden.

De basisgedachte achter dit soort onderzoek is als volgt: om de prestaties van een MDS-algoritme te kunnen onderzoeken moeten we met dit algoritme zeer veel verschillende datamatrices analyseren van stimuli waarvan de onderliggende structuur bekend is. Deze data moeten zoveel mogelijk lijken op het soort gegevens dat men doorgaans met het betreffende algoritme analyseert. In de eerste plaats wil dat zeggen dat de afmetingen van de datamatrices (het aantal stimuli) tussen een minimum- en een maximumwaarde moeten liggen. In de tweede plaats betekent het dat de structuur van de data, dat wil zeggen de dimensionaliteit en de vorm van de configuraties, vergelijkbaar moet zijn met wat gebruikelijk is. En in de derde plaats moet de kwaliteit van de data vergelijkbaar zijn met de kwaliteit van de gegevens die doorgaans verzameld worden, bijvoorbeeld gelijkenisoordeelen van proefpersonen. Data voor simulatie-onderzoek worden niet zozeer bij proefpersonen verzameld als wel door onderzoekers *geconstrueerd*. Men spreekt van *gesimuleerde* data. Aan het proces van dataconstructie zijn in dergelijk simulatieonderzoek een aantal facetten te onderscheiden:

- 1 Het *aantal stimuli*. Wagenaar en Padmos (1971), bijvoorbeeld, onderzochten nabijheidsmatrices van 8, 10 en 12 objecten, Spence en Graef (1974) matrices met 12, 18, 26 of 32 stimuli. Girard en Cliff (1976) gingen uit van een configuratie van 50 punten.
- 2 De *dimensionaliteit* van de stimuli. In de meeste toepassingen van MDS (zeker in de sociale wetenschappen) gaat het om oplossingen met één, twee, drie of

vier dimensies. Daarom worden voor simulatie stimulusconfiguraties gekozen met een ‘echte’ dimensionaliteit die varieert van 1 tot en met 4.

- 3 De vorm van de *configuratie*. Na keuze van een aantal (zeg m) punten in een aantal (zeg r) dimensies moet een onderzoeker de betreffende configuratie vastleggen, dat wil zeggen, voor elk punt r coördinaten genereren. Hiervoor worden verschillende procedures gebruikt. Twee mogelijkheden zijn: (a) voor elke dimensie worden m coördinaten aselekt gekozen uit een verdeling met bekende eigenschappen (bijvoorbeeld een normale verdeling met gemiddelde 0 en standaarddeviatie 1.00, of een uniforme verdeling tussen -1.0 en + 1.0); (b) uit een bekende configuratie met r dimensies worden m punten geselecteerd (bijvoorbeeld we noteren de lengte- en breedtegraden van de honderd grootste gemeenten in Nederland en trekken op een of andere manier een steekproef uit deze gemeenten; dat geeft een configuratie in twee dimensies die een realistische vorm heeft).
- 4 Het *error-niveau* van de data. Als we in empirisch onderzoek nabijheidsgegevens verzamelen, dan zullen deze data meestal niet perfect nauwkeurig (betrouwbaar) zijn. Bijvoorbeeld: als een proefpersoon de gelijkenis tussen twee objecten met een 5 beoordeelt (op een schaal van 1-7) dan zou het best kunnen dat de ‘werkelijke’ gelijkenis tussen de twee objecten ergens tussen de 4.5 en 5.5 ligt. In dat geval is de onnauwkeurigheid niet zo groot, maar erger wordt het als de ‘echte’ gelijkenis tussen de 3 en de 7 zou kunnen liggen. Als we aannemen dat de fouten die een proefpersoon maakt random zijn en uit een normale verdeling komen, dan zijn in bovenstaande voorbeelden de standaarddeviaties van de betreffende verdelingen gelijk aan .255 en 1.02. De error-variantie is in het tweede geval dus 16 keer zo groot als in het eerste. Twee belangrijke facetten van gesimuleerde data betreffen de hoeveelheid error die aan de data wordt toegevoegd en de manier waarop dit gedaan wordt. Op deze manieren komen we verderop terug.
- 5 De *toegestane transformatie*. Als een onderzoeker eenmaal de coördinaten van een aantal punten gesimuleerd heeft, dan kan men zonder meer de (Euclidische of andere) afstanden tussen de punten berekenen. Bij niet-metrische MDS gaan we ervan uit dat de nabijheidsdata van een proefpersoon niet meer dan monotone transformaties van de ‘echte’ afstanden tussen de punten zijn. Het is dus niet erg realistisch om het onderzochte MDS-algoritme op een gesimuleerde verzameling echte afstanden toe te passen. Realistischer is het om de berekende afstanden eerst via een of andere monotone functie te transformeren. Dat moet zodanig gebeuren dat de nabijheidsdata die we na transformatie krijgen qua verdeling en variatiebreedte zoveel mogelijk lijken op de scores van echte proefpersonen (bijvoorbeeld: alle nabijheden moeten bestaan uit liefst hele getallen, op een schaal van 1 tot en met 9). Het zal duidelijk zijn dat men een heleboel verschillende transformaties zou kunnen kiezen die allemaal min of meer hetzelfde doen. Eén zo’n transformatie, die door Girard en Cliff (1976, p. 50) beschreven wordt, is $g(d_{ij}) = 5.0 + 4[\arctangens\{.25(d_{ij} - 5)\}]$ (deze maakt overigens geen hele getallen van de nabijheden).

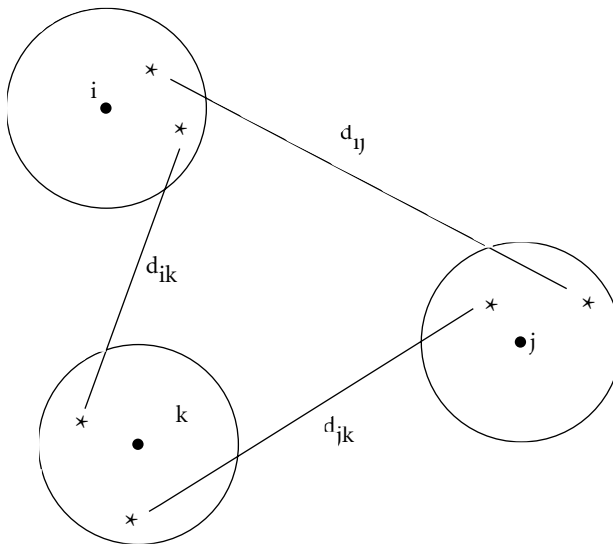
Het simuleren van error

Hierboven werd onder 4 gesproken over twee manieren om error aan de data toe te voegen. In de eerste methode berekent men de echte afstanden tussen de punten in de gesimuleerde configuratie en voegt men daaraan een of andere toevalscomponent toe. Bij Girard en Cliff (1976) werd bij elke afstand d_{ij} een toevallige waarde e_{ij} opgeteld die aselekt getrokken was uit een normale verdeling met gemiddelde nul en standaarddeviatie σ_e (Girard en Cliff gebruikten vijf error-niveaus: .20, .40, .60, .80 en 1.0). Wagenaar en Padmos (1971) daarentegen vermenigvuldigden iedere afstand d_{ij} met een toevallige waarde r_{ij} aselekt getrokken uit normale verdelingen met gemiddelde 1.0 en σ_r gelijk aan .05, .10, .20, .40 en nul (geen error). Het probleem met deze methode van *error op de afstanden* is dat men hierdoor soms afstanden kan krijgen die negatief zijn, waardoor ze niet in alle MDS-programma's zonder meer geanalyseerd kunnen worden.

In de tweede methode om afstanden met error te krijgen wordt de error niet direct aan de afstanden toegevoegd, maar aan de coördinaten van de punten. Met andere woorden, gegeven een configuratie van m punten in r dimensies met coördinaten x_{is} , berekent men afstanden volgens de formule

$$d_{ij}^* = \sqrt{\sum_{s=1}^r [(x_{is} + e_{is/ij}) - (x_{js} + e_{js/ij})]^2} \quad [12.3]$$

Alvorens de afstand tussen i en j uit te rekenen tellen we bij coördinaat x_{is} eerst een toevallige waarde $e_{is/ij}$ op en bij x_{js} een toevallige waarde $e_{js/ij}$. Het subscript is/ij wil zeggen dat het gaat om de coördinaat van punt i op dimensie s , als de afstand tussen Punt i en Punt j berekend wordt. Bij js/ij gaat het om de coördinaat van punt j op dimensie s , als de afstand tussen Punt i en Punt j berekend wordt. Met andere woorden, in iedere afstand waar Punt i in voorkomt, heeft i een verzameling coördinaten die steeds een beetje anders zijn. Hetzelfde geldt voor de andere punten. De getallen $e_{is/ij}$ worden weer aselekt getrokken uit een normale verdeling met gemiddelde nul en standaarddeviatie σ_e . Deze methode werd onder andere toegepast door Spence en Graef (1974), die voor σ_e de waarden .0625, .1225, .25 en 0 (geen error) gebruikten. De verzameling afstanden die berekend worden, zijn wel echte afstanden (ze zijn nooit kleiner dan nul) maar gaan dus iedere keer uit van andere posities van de punten. In twee dimensies variëren die posities in cirkelvormige gebiedjes rond de coördinaten zonder error. Deze manier van afstanden berekenen wordt geïllustreerd door Figuur 12.1.



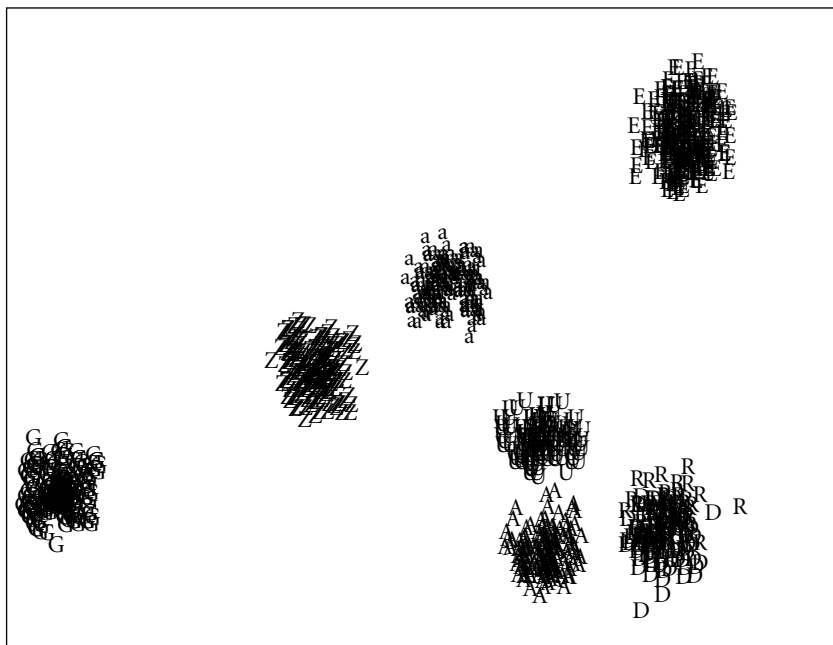
Figuur 12.1 De berekening van afstanden tussen punten met error op de coördinaten

Replicaties

De hierboven beschreven facetten kunnen gecombineerd worden tot een onderzoeksontwerp voor een simulatiestudie. Spence en Graef, bijvoorbeeld, combineerden vier aantallen stimuli met vier dimensionaliteiten en vier error-niveaus. Daarmee creëerden zij 64 condities waarbinnen zij vijf *replicaties* genereerden; een replicatie is een nieuwe matrix van gesimuleerde nabijheidsdata die volgens hetzelfde principe en met dezelfde parameters geconstrueerd zijn. Voor een simulatieonderzoek is het gebruikelijk om (veel) meer dan vijf replicaties te gebruiken.

Om het effect van error te demonstreren, zijn op basis van de coördinaten van de Nederlandse steden uit de niet-metrische ALSCAL-analyse (zie Blok 6.2) nieuwe afstanden berekend, nadat aan de coördinaten getallen waren toegevoegd die aselekt uit een normale verdeling met gemiddelde 0 en $\sigma_e = .10$ getrokken waren. Dit is in totaal honderd keer gedaan. De steden met error op hun coördinaten zijn afgebeeld in Figuur 12.2.

Een beschouwing over de voor- en nadelen van het gebruik van simulatie om MDS-technieken te bestuderen wordt gegeven in het artikel van Spence (1983), dat een uitgebreide literatuurlijst over dit onderwerp bevat.



Figuur 12.2 De lokaties van acht Nederlandse steden, na toevoeging van error aan hun coördinaten

