

GRAFISCHE ANALYSE MET SPECTRAMAP V2.0

D. Sikkel *)

Samenvatting

SPECTRAMAP is een programma waarmee een variëteit aan grafische analyses kan worden uitgevoerd die alle gebaseerd zijn op hoofdassen-analyse. De mogelijkheden omvatten correspondentie analyse, principale componenten analyse en "Spectral Map Analysis", een soort principale componentenanalyse op de logaritmes van de ruwe data, waardoor de resultaten een multiplicatieve interpretatie krijgen. Het programma is bedoeld voor gebruikers met weinig mathematische kennis van dit type technieken. De bedoeling van het programma is om door middel van grafische voorstellingen eigenschappen van de data boven water te krijgen die op basis van de cijfers alleen vaak verborgen blijven.

1. Technische gegevens en installatie

Naam: SPECTRAMAP V2.0

Release: 1992

Besturingssysteem: MS-DOS 2.10 of hoger

Minimale configuratie: IBM PC XT met 640 Kb geheugen, harde schijf, tenminste één 3.5" of 5.25" diskette drive, kleurenmonitor

Aanbevolen extra's: EGA of VGA kaart, graphics printer of plotter, voor grote data sets numerieke coprocessor

Prijs: f 3880,-; academische versie (max 30 x 12 tabel): f 1000,-

Handleidingen: een reference manual, een theorieboek en een boek met uitgewerkte voorbeelden

Distributie in Nederland: Smit Consult, Postbus 220, 5150 AE Drunen,
tel. 04163-78125

*) Universiteit van Amsterdam, Faculteit der Politieke en Sociaal-Culturele Wetenschappen, Vakgroep Methoden en Technieken
tel. 020 - 5252091 / 5252089

Installatie geschiedt in principe door de programma's van de diskette naar de harde schijf te kopiëren. Het programma stelt eisen aan de AUTOEXEC.BAT file en de CONFIG.SYS file. Voorbeelden van zulke files die aan de eisen voldoen worden bijgeleverd. Het programma beslaat ongeveer 500Kb aan vrije geheugenruimte. Indien er veel memory resident programma's geladen zijn en/of een netwerkaansluiting veel geheugen in beslag neemt zal Spectramap niet werken. Bij MS-DOS 5.0 en hoger zijn dit soort problemen altijd oplosbaar omdat het systeem en de memory resident programma's in high memory geladen kunnen worden.

2. Doel van het programma

Het doel van SPECTRAMAP is het maken van plaatjes waarin de eigenschappen van een datamatrix tot uiting komen. Een data-matrix kan een kruistabel van verschillende variabelen zijn, maar eventueel ook een matrix van individuele waarnemingen. Daarbij geldt dat het aantal rijen maximaal 250 bedraagt en het aantal kolommen maximaal 50. In het algemeen is het maken van zulke plaatjes een controversieel onderwerp. Wanneer het maken van een grafiek het serieus nadenken over de inhoudelijkheid van het onderwerp gaat vervangen leidt dit tot vervlakking en kwaliteitsverlies van de analyse. Dit is natuurlijk met name het geval bij wetenschappelijk onderzoek waarbij het gaat om het verwerven van objectieve, algemeen geldige kennis. Het overgrote deel van het onderzoek in de praktijk is echter niet wetenschappelijk en dient andere doeleinden dan het cumuleren van objectieve kennis. Zulke andere doeleinden zijn bijvoorbeeld het opdoen van ideeën voor nieuwe produkten, het zoeken op ad hoc basis naar ontwikkelingen in markten, het doorgronden van concurrentieverhoudingen in een bepaalde branche of het simpelweg in kaart brengen van populatiegegevens. In al dit soort gevallen kan het maken van plaatjes grote diensten bewijzen, mits ze ook op een juiste manier worden gebruikt.

SPECTRAMAP is ontwikkeld in het bedrijf Janssen Pharmaceutica en heeft dan ook als voornaamste doelgroep onderzoekers in het bedrijfsleven. Deze groep potentiële gebruikers zou inderdaad profijt kunnen hebben van het concept achter SPECTRAMAP. Immers, veel onderzoekers hebben moeite met het communiceren van de stap van soms gecompliceerde datastructuren naar

beleidsimplicaties, zeker waar het management van bedrijven weinig gevoel voor onderzoek heeft. Figuren kunnen, mits optimaal gepresenteerd, helpen bij discussies binnen bedrijven. Het stellen van vragen als: "waarom ligt punt A dicht bij punt B?" of "Wat zouden we moeten doen om de produkten A en B in de figuur wat verder uit elkaar te krijgen?" kunnen discussies aanscherpen en tot vruchtbare ideeën leiden. De vraag is natuurlijk wel of de wijze waarop SPECTRAMAP wordt aangeboden inderdaad tot het gewenste resultaat leidt.

3. Mogelijkheden van het programma

3.1 Algemeen

De eerste versie van SPECTRAMAP is zeven jaar oud; bij de tweede versie is de basis-structuur niet veranderd, maar zijn er wel vele mogelijkheden voor de gebruiker toegevoegd om verschillende varianten van de grafische voorstellingen te maken. Inmiddels heeft de concurrentie natuurlijk ook niet stilgezeten. De grote pakketten als SPSS en SAS hebben in de tussentijd opties als correspondentieanalyse, HOMALS en PRINCALS er bij gekregen en ook zijn er andere gespecialiseerde pakketten op de markt. Toch kunnen we veilig stellen dat SPECTRAMAP zo veel unieks heeft dat het een eigen marktpositie inneemt. De hoofdmodulen van het programma zijn de volgende

1. data input
2. analyse
3. plot file library
4. data file library
5. data transfer
6. plot attributes
7. peripherals

De modulen 3 tot en met 7 dienen voor data-management en output en voldoen aan de normen die men van een commercieel pakket mag verwachten. De eerste module is bestemd voor het invoeren van labels en data. Deze module bevat uitgebreide data-manipulatie mogelijkheden zoals het verwijderen, invoegen, kopiëren en samenvoegen van rijen en kolommen. Een interessante mogelijkheid is het transformeren van gegevens, bijvoorbeeld het optellen van een constante, de logaritme nemen, de wortel trekken of numerieke gegevens vervangen door ranggetallen. De invoermodule, die dus tevens voor data-transformatie dient, is hiermee zeer flexibel.

Tabel 1. Output van de lijst met opties die in SPECTRAMAP zijn gebruikt

PROCEDURE NAME		: General	
CONVERSION		: Transposition	: None
		: Re-expression	: None
GROUPING	Rows	: grp 1	: 1-9
		: grp 2	: 10-18
	Columns	: grp 1	: 1-13
DEFINITION		: Positioning	: None
		: Rejection	: None
WEIGHTING	- Rows	: Constant	
	- Columns	: Constant	
TRANSFORMATION		: Logarithms	: Yes
		: Closure	: None
		: Centering	: Rows+Columns
		: Normalization	: Global
SCALING		: Symmetrical	: Singular Values (.5,.5)
TRANSLATION	- Rows	: No	
	- Columns	: No	
SYMBOLS	Areas	- Rows	: Marginal Mean
		- Maximal Size	: 15.0 (mm)
		- Maximal Value	: 394.38
		- Minimal Value	: 0
		- Columns	: Marginal Mean
		- Maximal Size	: 15.0 (mm)
		- Maximal Value	: 397.78
		- Minimal Value	: 0
	Labels	- Rows	: Name
		- Columns	: Name
VIEW	Frame	- Rows	: Functional
		- Columns	: Functional
	Zoom	- Rows	: 100 (%)
		- Columns	: 100 (%)
	Mask	- Rows	: None
		- Columns	: None
	Labels	- Rows	: All
		- Columns	: All
LINES		- Rows	: --1 5--> Cal 5 Mrks
			: --10-14--> Cal 5 Mrks
		- Columns	: None
DISPLAY	Axes	: Horizontal	: Factor 1
		: Vertical	: Factor 2
		: Depth	: Factor 3
	Rotation Angles	: Compute	: None
		About Horizontal Axis	: 0 (deg)
		About Vertical Axis	: 0 (deg)
		About Depth Axis	: 0 (deg)
	Reflection	: Horizontal	: None
		: Vertical	: None
		: Depth	: None

3.2 Analyse

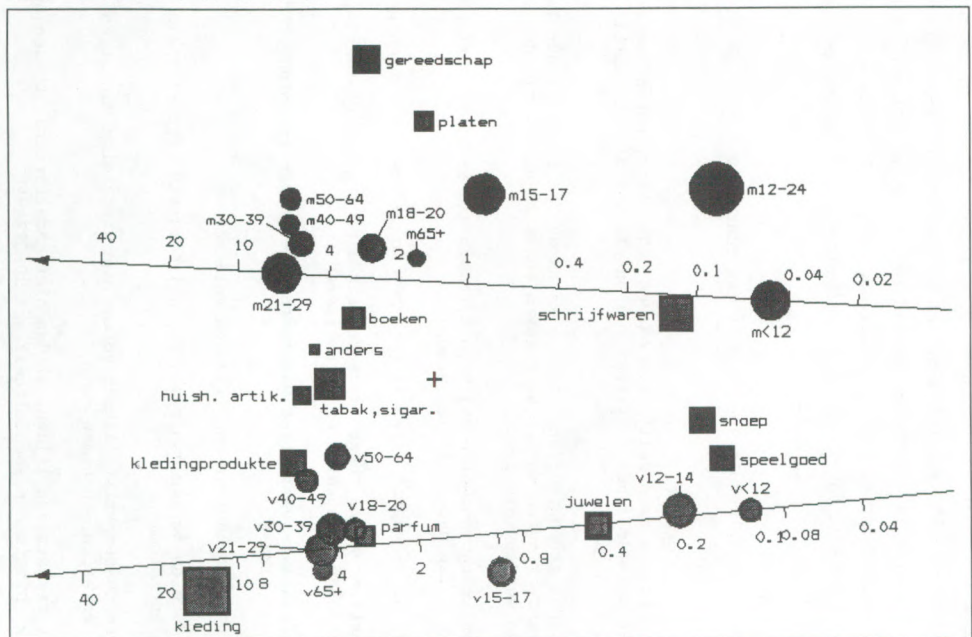
De kern van het programma is de module voor analyse. Hierin wordt aangegeven hoe de figuren en uit moeten komen te zien en worden ook de feitelijke berekeningen uitgevoerd. In deze module komt men een groot aantal opties tegen waarvan verschillen in instelling tot verschillende vormen van analyse leiden. Er zijn uiteraard default-instellingen. Deze leiden tot het uitvoeren van Spectral Map Analysis. De gebruikte opties kan men achteraf als lijst laten wegschrijven. Tabel 1 is een voorbeeld van zo'n lijst en zal als leidraad dienen bij de bespreking van de verschillende opties. Het voorbeeld dat wordt gebruikt is een bekende tabel uit de literatuur: winkeldiefstallen naar leeftijd en geslacht, zie Israels (1987).

Wanneer men de analyse module ingaat komt men achtereenvolgens de volgende opties tegen

- **Conversie.** Dit is de mogelijkheid om een tabel te transponeren en de celwaarden x in de tabel te transformeren als $1/x$ of $c-x$ waarbij c een constante is.
- **Grouping.** Het groeperen van rijen en kolommen die in de figuur later met verschillende kleuren worden aangegeven en waarnaar ook bij de volgende opties kan worden verwezen.
- **Definition.** Aangeven welke rijen en kolommen als passieve variabelen meedoen en welke helemaal niet meedoen.
- **Weighting.** Wijze waarop rijen en kolommen in de analyse worden gewogen.
- **Transformation.** Keuze tussen directe waarde of log-transformatie, delen door marginalen, centreren en normaliseren.
- **Scaling.** Relatieve varianties van factoren en rijen en kolommen ten opzichte van elkaar.
- **Translation.** Verschuiven van rijen en kolommen.
- **Symbols.** Vorm en omvang van symbolen, bijvoorbeeld evenredig aan de omvang van rijen en kolommen.
- **View.** Keuze tussen verschillende vormen van aanduiding van rijen en kolommen, maskeren, inzoomen etc.
- **Lines.** Het trekken van lijnen, al dan niet gecalibreerd, tussen punten in de grafiek. Dit is een specialiteit van SPECTRAMAP
- **Display.** Keuze welke van de eerste 4 factoren worden getoond. Hierbij is handmatig drie-dimensionale rotatie mogelijk.

SPECTRAMAP

winkeliefstallen naar leeftijd en geslacht



© Janssen Pharma 2B-Mrh-93 LNDGS C:47 26 14,87 A:123 R:0 0 0 S:+++ 1 A

Figuur 1. Winkeliefstallen naar leeftijd en geslacht

Duizelt het u inmiddels? Het aantal opties is inderdaad enorm groot en voor de beginner (waarvoor dit programma voornamelijk is bedoeld) absoluut niet te vatten. Dit is niet als kritiek bedoeld. Wanneer het programma slechts zeer beperkte opties had gehad had ik waarschijnlijk geschreven dat er wel wat meer in had gekund. Statistici zitten nu eenmaal voor het dilemma van ofwel een zeer beperkte verzameling methoden aanbieden en daarmee problemen suboptimaal oplossen ofwel veel meer keus aan te bieden met het risico dat de doelgroep het spoor bijster raakt. Bij SPECTRAMAP is gekozen voor een maximaal aantal opties en daarbij een uitgebreide begeleiding. Wie optimaal van het pakket wil profiteren en geen ervaren statisticus bij de hand heeft dient daarom speciaal in het gebruik van dit pakket te worden opgeleid. Zulke opleidingen worden dan ook door de distributeur verzorgd.

Een voorbeeld van de grafische output is figuur 1. Hierin zijn twee lijnen getrokken. Hiermee wordt aangegeven hoe het verloop in winkeldiefstallen is van de groep "21-29 jaar" ten opzichte van "<12 jaar", apart van mannen en vrouwen. Om dit af te kunnen lezen dient men de categorieën gestolen goederen op de lijnen te projekeren. Zo worden schrijfwaren door 21-29 jarige mannen 0.13 maal zo vaak gestolen als door jongens jonger dan 12 jaar. Gereedschap wordt door 21-29 jarige mannen echter 3.5 maal zo vaak gestolen. De omvang van de symbolen is hier evenredig met de omvang van de categorieën. Uit dit voorbeeld blijkt duidelijk de potentiële kracht van de techniek. Het plaatje bevat zeer veel informatie, en je kunt er uren mee brainstormen.

4. Praktische zaken

De eerste versie van het programma is zeven jaar oud en dat is te merken. Inmiddels zijn gebruikers-interfaces redelijk gestandaardiseerd, maar de bediening van het programma is hier niet in meegegaan. Niet dat het ingewikkeld is, het is gewoon anders. Een voorbeeld is de ESC-toets, die in vrijwel elk groot pakket de betekenis "ga één niveau terug" heeft. Bij SPECTRAMAP ga je met de ESC-toets altijd terug naar het hoofdmenu. Een gemis is ook dat de betekenis van de belangrijkste toetsen niet op het scherm staat. Ik, en velen met mij, zijn inmiddels te lui geworden om het op te zoeken in de manual.

De manual geeft aan dat het programma incidenteel een run-time error kan geven en zelfs de computer kan laten hangen. Zonder dat ik het op zijn limieten heb getest is me ook twee maal een run-time error overkomen, beide malen door een niet geïnitieerde pointer. Zo stel je beginnende gebruikers niet op hun gemak.

Bij het pakket horen drie manuals, te weten een reference manual, een boek met voorbeelden en een boek waarin de technieken exact worden uitgelegd. Hoewel dit laatste een grote vooruitgang is ten opzichte van de vorige versie, lijkt mij er ook nog wel behoefte te bestaan aan een duidelijke uitleg voor de beginnende gebruiker wat nu precies de essentiële verschillen in definitie en gebruik zijn tussen correspondentieanalyse, principale componenten analyse en spectramap analyse. De lineaire algebra die (terecht) wordt gepresenteerd zal voor velen onbegrijpelijk zijn.

5. Conclusie

SPECTRAMAP is, ondanks enkele kleine onvolkomenheden, een kwaliteitsprodukt waarin veel expertise en ervaring is verwerkt. Wie echter als bedrijfsonderzoeker voor SPECTRAMAP kiest dient zich te realiseren dat hij zich niet alleen dient in te werken in een groot aantal technische varianten die met het pakket gerealiseerd kunnen worden, maar ook dat hij de mogelijkheden en onmogelijkheden moet leren begrijpen van het analyseren van data op basis van plaatjes. Deze investeringen zijn aanmerkelijk hoger dan alleen de prijs van het pakket.

Referentie

Israels, A.Z., 1987, Eigenvalue Techniques for Qualitative Data, DSWO Press, Leiden.