

Statistiek op de ATARI met ALMO (Allgemeines Lineares Modell)

Teije Euverman

Samenvatting

In dit artikel wordt een algemeen statistiek pakket besproken voor de ATARI. Vrijwel alle gangbare technieken zijn met ALMO mogelijk, maar de bespreking richt zich vooral op het lineaire model en op onderdelen van overige analyses, voor zover deze niet op een standaard manier zijn geïmplementeerd.

1 Produktinformatie

Naam:	ALMO
Versies:	ALMO V3.1 (Een update); ALMO_E (Een nieuwe versie)
Leverancier/	
ontwikkelaar:	Prof. Dr. Kurt Holm, Universität Linz
Privat:	Am Schloßberg 8, A 4060 Leonding, Österreich
Tel:	09-43 732 674711
Configuraties:	ATARI ST/TT/STE, 2 MB; Onder OS/2 is ALMO_E beschikbaar
Prijzen:	Steeds inclusief handboek: Voor ATARI: ALMO_E: DM 498,- ,(demo: DM 20,-); ALMO_V3.1: DM 248,-, (demo: DM 10,-).Onder OS/2: ALMO_E: DM 1200,-, (demo: DM 20,-).Steeds DM 20,- verzendkosten.
Opmerking:	Een Atari ST is te emuleren met GEMULATOR onder MS/DOS en een 386 of 486 processor: prijs: f 498,- (zie bijv. Smit, 1993).

2 Inleiding

Het programma ALMO is oorspronkelijk ontwikkeld voor een mainframe computer en is in 1988 geschikt gemaakt voor de ATARI ST. Inmiddels is er tevens een versie voor de snelle ATARI TT en voor OS/2, waarbij de multi-tasking mogelijkheden ten volle worden benut. Het onderstaand artikel is tot stand gekomen op grond van het rekenen met de voorlaatste versie op een ST met 1 MB intern geheugen en de handleiding voor de nieuwste versie (ALMO_E), die minimaal 2 MB vereist. ATARI kent nog veel aanhangers in Nederland al moet gezegd worden dat het marktbeleid van ATARI aan de zeer magere kant is. Misschien brengt de introductie van de FALCON machine daarin verandering (zie bijv. Kuit, 1993). Meer dan in Nederland, wordt de ATARI bij de oosterburen als een volwaardige computer beschouwd en vele goede programmatuur is dan ook vooral in Duitsland ontwikkeld. ATARI heeft een besturingssysteem dat werkt in een grafische omgeving, hetgeen het gebruik van de machine zo vriendelijk maakt. Vensters behoren dus al tot de

standaarduitrusting en bijvoorbeeld het geavanceerd tekst(formule)verwerkingsprogramma SIGNUM past dan ook het gebruikersvriendelijke WYSIWYG- principe met gemak toe. ATARI wordt ook wel de "poor man's MAC" genoemd en deze vergelijking zou men dan ook op zijn minst nieuwsgierig moeten maken. Het ALMO programma is inderdaad een goedkoop pakket waarmee geavanceerde multivariate analyses kunnen worden uitgevoerd. ALMO is in Oostenrijk ontwikkeld onder supervisie van Prof. Dr. Kurt Holm aan de Universiteit van Linz. Het centrale thema in dit pakket is het lineaire model, maar daarnaast zijn meerdere statistische procedures mogelijk. Voor een snel overzicht zij verwezen naar de paragrafen 4.2 en 4.3. ALMO is geprogrammeerd in C, terwijl in ALMO wordt geprogrammeerd in het Duits.

De bijbehorende handleiding telt 800 pagina's, maar is veelal niet erg informatief m.b.t. de onderliggende statistisch-theoretische keuzes. Men verwijst helaas veel naar lokale literatuur. Waar mogelijk en nuttig, wordt in dit artikel verwezen naar alternatieve, gemakkelijker beschikbare literatuur.

3 Werking van een ALMO programma

Kenmerkend voor ALMO is het feit dat niet de totale datamatrix in RAM wordt bewaard, maar sequentieel record voor record. Het gevolg is dat ALMO een onbeperkt lange datamatrix kan bewerken. De bewerking geschiedt dus per record.

Uiteraard is ALMO menu gestuurd en de meeste arbeid geschiedt d.m.v. de muis. Er zijn sjablonen beschikbaar van ieder type statistische analyse, zodat bij een eerste gebruik om te oefenen niets hoeft te worden ingetikt. In het sjabloon venster wordt per regel de bedoeling van de syntaxis uitgelegd. Er kunnen veranderingen in aangebracht worden en men kan gaan rekenen en het resultaat in de editor bekijken. Voor iedere analyse bestaan tevens voorbeeldprogramma's die in de editor kunnen worden bewerkt en de gewijzigde dataset kan naast het oorspronkelijke programma bewaard worden. Het bewerken zal meestal wat schrappen en wat geringe toevoegingen betekenen. Het is dus nooit nodig helemaal een eigen programma in te tikken. De editor werkt zeer eenvoudig en met muisklikken komt men in twee types van modi. In de edit-mode worden via klikken in de drop down menu's de gewenste commando's uitgevoerd. Handmatige wijzigingen kunnen in de andere (input)-mode worden aangebracht.

Bewaarde ASCII-files uit tekstverwerkers of database programma's kunnen in ALMO worden geladen. Files uit SPSS met het FIX formaat worden volledig herkend door ALMO. ALMO werkt het snelst met het zg DIREKT-format. Dit is is een ALMO-specifieke, intern binaire code en is dus niet uitwisselbaar. Het FREI-format komt vrijwel overeen met ASCII-code.

Hieronder wordt in het kort de opbouw weergegeven van ALMO. Enerzijds wordt ALMO-code gegeven, anderzijds, cursief, algemene informatie. Sommige code moet voorbeeldgewijs beschouwd

worden, d.w.z. als een mogelijke specifieke invulling van een programmablok (het geheel vormt echter niet een bepaald voorbeeld).

Vereimbare

Variable=10;

Reserveer geheugenruimte voor 10 variabelen. Hierna kunnen diverse voorinstellingen aangegeven worden. Als in te lezen variabelen kunnen ook vectoren en tabellen gelden.

Anfang

Het begin van ieder ALMO-programma. Hier kunnen bijvoorbeeld namen aan variabelen gegeven worden en aan hun waarden. Bijv.:

Name1=Geslacht:mannelijk,vrouwelijk;

Het gedeelte na het =teken is dus het enige wat men zelf in hoeft te tikken in het betreffende voorbeeldprogramma. Deze benoeming kan nu overal in het programma de standaard benoeming v1 vervangen.

Programm=20;

Begin van een subblok genaamd: Programmparameter. Het getal geeft de keuze van de procedure aan. Programma 20 is het algemene lineaire model. Voor een buitenstaander is dit natuurlijk niet meteen duidelijk, maar voor een gebruiker is dit een triviale zaak. Hierna volgen allerlei aanwijzingen relevant voor de betreffende routine. Bijv.:

U_nominale_v=v1,v2;

De U van unabhängig en voor abhängig dan

A_quantitative_v=v3,v4;

Voor de nominale variabelen moeten onder en bovengrenzen worden aangegeven:

Untergrenze 1,2=1,1;

Obergrenze 1,2= 2,3;

Ug en og mag ook. Verkorte vormen zijn veelal mogelijk.

Interaktionen=2;

Verfahren =fitting constants;

Matrix=Quadratsumme;

Ende_programmparameter;

Einde subblok. Bovenstaand subblok mag ontbreken. Dit houdt dan bijvoorbeeld in dat uitsluitend data veranderingen gedefinieerd worden via het volgende blok. In dit tussenblok kunnen ook labels (Marke) aangebracht worden om naar toe te springen.

1 Marke

Zeichenvariable=v11;

Hierna volgt de z.g. Lese-schleife. Bijv:

```
Lese v1:25  
aus datei1 'daten.dir'  
format direkt  
leer_zu Ende;
```

De extensie wijst op het format. Hier dus van het type direkt, d.w.z. in een interne binaire machinecode, hetgeen direkte (niet-sequentiële) toegang verschaft. Andere formaten zijn mogelijk. Het FIX format van SPSS wordt door ALMO herkend. Schrijven naar een file en afdrukken gaat analoog. Hierna kunnen allerlei herbewerkingen van de variabelen plaatsvinden. Een aantal voorbeelden:

```
v8(1:5=1; 6:10=2)
```

Variabele 8 wordt hier dus binair gehercodeerd.

```
v103 = sin (log v42)  
H2(Zufall 0:1)
```

Er staan automatisch hulpvariabelen ter beschikking. In het laatste voorbeeld wordt aselect getrokken uit een uniforme verdeling. Een normale verdeling is ook mogelijk en voor simulatiedoeleinden ook bivariate trekkingen. Naast hulpvariabelen staan ook toegevoegde variabelen ter beschikking met o.a. gemiddelden en standaarddeviaties. Een varabele kan worden gerepresenteerd in een aantal dummies:

```
v401:v405 (dummy v48)
```

In dit geval bevat v48 6 verschillende waarden; 5 dummies zijn dus voldoende. Er zijn voorwaardelijke toewijzingen en loops (schleifen) mogelijk. Een variabele kan logische waarden 1 en 0 krijgen:

```
v101=v40 gleich v41 und v50 gleich v51
```

Hier krijgt variabele 101 de waarde 1 als variabele 40 dezelfde waarde heeft als variabele 41 en als variabele 50 dezelfde waarde heeft als 51. De bewerkingen geschieden sequentieel per record (datenvector). Met het al dan niet bewerkte record wordt naar het programma gesprongen met:

```
Gehe_in_programm
```

Dan de volgende datavector halen:

```
Gehe_zu Lese
```

Ieder record vervangt dus steeds het oude en dit gaat zo door tot het einde van de dataset en dan een sprong aan het einde van de Lese-Schleife naar:

```
Ende
```

Het einde van het blok dat begon met Anfang. Het programma kan data direct na deze commando's inlezen.

ALMO kent verder filebewerkingen en matrix behandelingen.

4 Statistische procedures in ALMO

4.1 Het algemene lineaire model

Het geïmplementeerde algemene lineaire model is niet het gegeneraliseerde lineaire model, waarin een functie van een verwachte waarde van een random variabele uit de exponentiele familie lineair gemodelleerd wordt (zie bijv. McCullagh & Nelder, 1991). In ALMO is er sprake van een generalisatie van het lineaire regressiemodel naar een algemeen lineair kansmodel waarin men afhankelijke en onafhankelijke variabelen kan kiezen van kwalitatief meetniveau. Logit en probit analyses zijn wel mogelijk. Een categorische data analyse volgens het log-lineaire model echter niet. Een kwalitatieve generalisatie van lineaire regressie kent problemen (zie bijv. Aldrich & Nelson, 1984). De toepassing van kleinste kwadraten en de standaard statistische inferentie in ALMO is dan ook bezwaarlijk in deze gevallen.

Nu wordt eerst de specifieke aanpak besproken in het geval er nominale en ordinale variabelen aan de orde zijn. Een nominale variabele wordt uiteengelegd in k binaire variabelen en indien het een afhankelijke variabele betreft, volgt dientengevolge een multivariate aanpak van het bijbehorende lineaire model. De geschatte waarden voor de afhankelijke variabelen worden dan geïnterpreteerd als voorspelde kansen. Het optreden van een ordinale variabele leidt tot een verandering van de datamatrix. De n waarnemingen worden dan paarsgewijs behandeld en de resulterende matrix is een $n(n-1)/2$ bij p matrix, waarbij p het aantal variabelen voorstelt. Ter herinnering: alleen een enkele rij van een datamatrix is in RAM opgeslagen. Als er sprake is van een kwantitatieve afhankelijke variabele en een ordinale onafhankelijke variabele levert de geschatte waarde een interpretatie op van het voorspelde verschil voor ieder paar onderzoekseenheden. Is de afhankelijke variabele nominaal, dan wordt dat dus een kansuitspraak over het verschil, en bij eveneens een ordinale afhankelijke variabele volgt een ordinale kansuitspraak. Een dergelijke paarsgewijze behandeling ligt ten grondslag aan een produkt-moment-correlatieachtige interpretatie van Kendall's tau-b (zie bijv. Yule & Kendall, 1968). Het geïmplementeerde algemene model staat dus meerdere afhankelijke en onafhankelijke variabelen toe van diverse schaaltypes. Dit levert een taxonomie van modellen op, waarvan een aantal correspondeert met traditionele technieken: (M)ANOVA, regressieanalyse, covariantieanalyse, waarbij de afhankelijke variabelen kwantitatief en theoretisch continu zijn. In deze gevallen zal de gebruikelijke statistische inferentie geen problemen opleveren indien de waarden realisaties zijn van normaal verdeelde variabelen. Problemen kunnen verwacht worden bij bijvoorbeeld discrete kwantitatieve variabelen. In een dergelijk geval ligt een Poisson verdeling meer voor de hand (zie bijv. Santner & Duffy, 1989). Van belang is het gevolg van de universele aanpak bij kwalitatieve afhankelijke variabelen. Dan is er sprake van van kwalitatieve (multivariate) regressieanalyse, variantieanalyse en kovariantieanalyse. De aanpak betekent dus het toepassen van het lineaire

kansmodel. Problemen bij dit model komen voort uit de inherente heteroscedasticiteit. Gelukkig is er een gewogen aanpak mogelijk in ALMO met het doel dit probleem op te lossen. Een ander probleem, maar volgens de programmeurs een zelden optredend verschijnsel, is de restrictie dat de voorspelde waarden in (0,1) moeten liggen. Als er sprake is van één nominale afhankelijke variabele, ligt een logit of probit analyse meer voor de hand. Natuurlijk is de generieke aanpak zoals de programmeurs deze voorstaan aantrekkelijk, of misschien verleidelijk, omdat er een algemeen principe aan ten grondslag ligt. Het lijkt alsof de programmeurs de kritiek wel kennen, maar een grote rol toekennen aan een universeel type interpretatie van de resultaten. Dit blijkt uit het belang dat men hecht aan de PRE-coëfficiënt (proportional reduction of error). Hieronder wordt een formule gegeven in termen van kwadratensommen (SS =sums of squares) gegeven, d.w.z. voor één afhankelijke variabele y en q onafhankelijke variabelen x_i :

$$PRE^2 = \frac{SS_E(\text{in het model met } x_1, \dots, x_k) - SS_E(\text{in het model met } x_1, \dots, x_q)}{SS_E(\text{in het model met } x_1, \dots, x_k)},$$

waar x_1, \dots, x_k een deelverzameling is van x_1, \dots, x_q (en vooraan geordend).

De coëfficiënt geeft de proportionele foutenreductie weer indien x_{k+1}, \dots, x_q aan het model worden toegevoegd. Bij kwantitatieve variabelen zijn de (multiële) correlatie en partiele (multiële) correlatie specifieke gevallen (uiteeraard na het trekken van de vierkantswortel). Bekende associatiematen geldig voor variabelen van kwalitatief of gemengd kwalitatief/kwantitatief meetnivo volgen het PRE-stramien. De phi-coëfficiënt, de punt-biseriële coëfficiënt en de eta-coëfficiënt zijn specifieke gevallen. Ook Kendall's tau-b kan volgens het PRE-stramien geïnterpreteerd worden. De PRE-coëfficiënt kan ook niet-in-de-literatuur-benoemde correlatiecoëfficiënten opleveren. Dit gebruik van de PRE-coëfficiënt leidt tot gemakkelijk interpreteerbare resultaten. Bij statistische inferentie is terughoudendheid op zijn plaats. Als meerdere afhankelijke variabelen optreden wordt in de PRE-coëfficiënt de foutenkwaadratensommen SS_E vervangen door de determinanten van de relevante residu matrices. Een interessant gevolg is de invariantie in de verkregen waarde bij de rolverwisseling van één y en q x -variabelen naar de rol van deze y als een onafhankelijke variabele en de q x -variabelen als afhankelijke variabelen. Hierin wordt de symmetrie van ook een multiële correlatie herkend. PRE^2 is gelijk aan 1 minus Wilks' lambda. Deze relatie impliceert dat in ALMO voor het model met polytome afhankelijke variabele (een multivariaat geval) en kwantitatieve onafhankelijke variabelen de waarde van Wilks' lambda dezelfde is als die verkregen bij (canonische) discriminantanalyse. Uiteeraard wordt hiermee niet beweerd dat de regressieaanpak in ALMO equivalent is met (canonische) discriminantanalyse. De equivalentie geldt, zoals bekend, wel voor een dichotome afhankelijke variabele (zie bijv. Tatsuoka, 1971). Bovendien geldt de invariantie van Wilks' lambda bij de ALMO

behandeling van een dichotome afhankelijke variabele en een polytome onafhankelijke variabele, en die lambda berekend bij de overeenkomstige variant van correspondentieanalyse. Voor overige bestaande relaties tussen de diverse analysetechnieken zij hier verwezen naar Lebart e.a. (1984).

Het is dus zeer aantrekkelijk om het regressiemodel te generaliseren. Problematisch wordt het bij de statistische inferentie. Men moet zich bijvoorbeeld voortdurend afvragen of Wilks' lambda (gedefinieerd als een ratio) wel altijd een F-verdeling (na transformatie) heeft voor de diverse specifieke gevallen van het algemen lineaire model. In ALMO worden bovendien in al die gevallen regressiecoëfficiënten berekend, waarop F-toetsen worden losgelaten, hetgeen natuurlijk op z'n minst zeer aanvechtbaar is.

Concluderend kan gesteld worden dat in beschrijvende statistische zin de waarden afleidbaar uit de PRE-coëfficiënt goed interpreteerbaar zijn, maar dat de gehanteerde verdelingen van de optredende schatters niet universeel gelden. Bovendien bestaan er andere, theoretisch onderbouwder correlatiecoëfficiënten voor ordinale, polytome of dichotome variabelen, te weten, de polychorische, tetrachorische, polyseriële en biseriële coëfficiënten waarvan bijvoorbeeld in LISREL gebruik wordt gemaakt.

Na de voorafgaande uiteenzetting omtrent de grondgedachte in ALMO, worden nu in het kort de mogelijkheden van het geïmplementeerde lineaire model beschouwd zonder te letten op de schaalproblematiek. Dit houdt in dat voor het gemak gedacht moet worden aan het standaard lineaire model in de zin van Searle (1971). De mogelijkheden zijn legio. Het programma berekent kleinste kwadraten schatters bij de gewenste manier van hiërarchische uitpartialisering. ALMO kent standaardprocedures zoals UNIQUE en SEQUENTIAL in SPSS. Ook in het overgeparametriseerde model (het niet-volledige rang model) worden in eerste instantie regressiecoëfficiënten bepaald op grond van een dummy X -matrix van volledige kolom-rang door in de oorspronkelijke X -matrix redundante kolommen weg te gooien. Deze coëfficiënten vormen dus één van de mogelijke oplossingen van de normaal-vergelijkingen en zijn derhalve geen schatters van schatbare functies. De schatbare functies, d.w.z. de effecten of lineaire combinaties van de verkregen coëfficiënten, zijn helaas niet eenvoudig in z'n algemeenheid te achterhalen (hiervoor heeft men de beschikking nodig van de bijbehorende lokale literatuur) in ALMO. Men interpreteert de vector met schatbare functies, zeg a , in termen van de bijdragen door de oorspronkelijke dummies in X . Een dergelijke a kan worden verkregen door deze te definiëren als $a = GX'Xb_0$, waar G een gegeneraliseerde inverse is van $X'X$ en b_0 de vector is met de regressiecoëfficiënten (zie Searle, 1971, p. 185). Hier is dus sprake van een verschil met bijv. de aanpak van Finn (1974), waar X getransformeerd wordt op grond van de lineair onafhankelijke schatbare functies (contrasten) (zie Finn, 1974, p. 220). Tot zover over de berekening van de effecten, waarbij dus ook hier voor de veralgemenisering van het model hetzelfde stramien gevolgd wordt.

Voor het overige wordt hier vermeld dat ALMO toepasbaar is op alle gangbare types van

onderzoeksontwerpen, grofweg te vatten onder onvolledige ontwerpen, geneste ontwerpen, herhaalde metingen (multivariate en univariate optie) en die waarbij het zelf definiëren van de foutenmatrix nodig is (bijv. bij random effecten). Verder o.a. polynome regressie, toetsen op homoscedasticiteit, op homogeniteit van regressiecoëfficiënten, toetsen van variantieheterogene modellen.

Bij dezelfde programma-aanroep kan men logit en probit analyses uitvoeren volgens het minimum chi-kwadraat principe (zie Aldrich & Nelson, 1984, hoofdstuk 4). Bij meer dan twee ordinale categorieën van de afhankelijke variabele wordt de laatste categorie als referentiecategorie genomen (het z.g. baseline-category logit model). De optie van cumulatieve logits wordt eveneens geboden (zie bijv. Agresti, 1990, hoofdstuk 4). Zoals vermeld is een algemene log-lineaire aanpak niet opgenomen in ALMO. Tenslotte wordt opgemerkt dat in ALMO padanalyse mogelijk is op grond van recursieve regressieanalyses.

4.2 Standaard multivariate analyses

Hieronder worden in alfabetische volgorde routines beschouwd die in ALMO buiten het algemene lineaire model een aparte behandeling krijgen (ook al kunnen ze soms onder het algemene model vallen: zie 4.1).

1 Canonische correlatieanalyse

2 Classificatie

Middels een voorafgaande discriminantanalyse wordt op grond van een realisatie van een random vector groepslidmaatschap voorspeld.

3 Clusteranalyse

Het agglomeratieve hiërarchische algoritme middels diverse linkage procedures en met behulp van vele bekende associatiematen en metrieken. (zie voor een overzicht bijv. Aldenderfer & Blashfield, 1984). K-means clustering methode en de probabilistische versie: Latente profielanalyse (zie Lazarsfeld & Henry, 1968).

4 Correspondentieanalyse

De tweeweg variant wordt behandeld als canonische correlatieanalyse met de representatie van de twee nominale variabelen als dummies (zie bijv. Lebart e.a., 1984, pp. 79-81). De meerweg variant wordt gezien als een vorm van nominale factoranalyse (men bedoelt principale componentenanalyse), maar wordt berekend als de bekende MCA-procedure op grond van de z.g. Burt-matrix (zie Lebart e.a., hoofdstuk 4).

5 Discriminantanalyse

6 Factoranalyse

De volgende mogelijkheden zijn geïmplementeerd: Principale componentenanalyse, principale factoranalyse, alfa factoranalyse, canonische factoranalyse, image factoranalyse (zie bijv. Mulaik,

1971). Rotaties: Quartimin (oblique), Varimax (orthogonaal). Bij nominale variabelen kan worden gerekend volgens de block-diagonale methode van McDonald (1969) of volgens het model van multiële correspondentieanalyse. Factoranalyse op dummies kan ook, maar wordt gelukkig door de programmeurs afgeraden. Een heuristische ordinale variant is gebaseerd op de paarsgewijze vergelijkingen zoals beschreven in 4.1, d.w.z. een analyse op de tau-b correlatiematrix. Overigens zijn analyses voor (ordinale) discrete variabelen in bijvoorbeeld LISREL theoretisch beter onderbouwd.

7 Schaaltechnieken

Paarsgewijze vergelijkingen volgens Thurstone's case V; berekening van de voorlopige afstandsmatrix (afstanden op een additieve constante na) op grond van de volgende data collectiemethoden: (meervoudige) rangordening, tetraden- en triadenmethode, rangordening van paren; berekening van een afstandsmatrix op grond van oordelen verkregen met betrekking tot verschillende dimensies (zie bijv. ook Meerling, 1981). De klassieke niet-metrische multidimensionale schaaltechniek volgens Kruskal (1964) kan eveneens uitgevoerd worden.

4.3 Overige analyses

1 Standaard statistische toetsen

Parametrische toetsen op gemiddelden en varianties, waaronder de Welch-toets voor gemiddelden bij onafhankelijke steekproeven en de Ferguson-toets voor varianties bij afhankelijke steekproeven (zie bijv. Sachs, 1984).

Niet-parametrische toetsen, waaronder de exacte Fisher-toets, de exacte Freeman-Halton-toets, en toetsen op variaties: de Siegel-Tukey-toets, de Mood-toets, en de Shorak-toets. De ordinale Ridits-toets is eveneens mogelijk (zie bijv. Agresti, 1990). Vele associatiematen kunnen worden berekend: phi-coëfficiënt, Pearson's C, Tschuprow's T, Cramer's V, de asymmetrische lambda-coëfficiënt (eveneens een toepassing van de PRE-coëfficiënt), Goodman en Kruskal's gamma, Kendall's tau-b, Spearman rho (zie bijv. tabel 2.2 en 2.3 in Coxon, 1982).

Verdelingstoetsen.

Tabelanalyse. Bij meervoudige tabelanalyse wordt een 'overall' chi-kwadraat of een 2I-toets berekend, waarna via z.g. configuratie-frekwentieanalyse afzonderlijke bijdragen in de chi-kwadraat worden getoetst (Krauth & Lienert, 1973). Deze partitionering geschiedt dus niet zoals bijvoorbeeld beschreven in Agresti, 1990.

2 Tijdreeksanalyse

Men rekent alleen in het tijd-domein, waarbij het argument gehanteerd dat er een equivalentie bestaat met het frekwentie-domein (zie bijv. Gottman, 1984). De Box-Jenkins procedure ontbreekt eveneens, met het argument van het verlies aan informatie door de Box-Jenkins transformatie. De

regressiebenadering wordt gebruikt, maar men wijst op de tekortkomingen hiervan. Het voortschrijdend gemiddelde, (partiële) (kruis)-autocorrelaties of covarianties kunnen worden berekend, waarmee eventueel een AR of een MA proces van de eerste of tweede orde kan worden geïdentificeerd. Vanuit een generiek oogpunt kan ALMO het algemene lineaire tijdreeks model bewerken. Dit model bevat voor de verklaring van een autoregressief proces van de afhankelijke variabele een constante, een polynome trend, een seizoen component, meerdere onafhankelijke variabelen en een autoregressieve component van storingstermen.

3 Latente trek en structuur modellen

Het Rasch model met simultane schatting van de persoons- en itemparameter. Een vorm van kansmatige ontvouwing (probabilistic unfolding) door op grond van paarsgewijze vergelijkingen van items in het Rasch model de itemparameter te vervangen door het middelpunt van een itempaar, waarna bij een geaccepteerd model schaalwaarden voor items worden afgeleid naast de persoonsparameters (zie voor alternatieve modellen Post (1992)). Latente structuur analyse met het lineaire traceline model, het latente afstandsmodel (met als speciaal geval de scalogramanalyse), het normaal-ogief model en het latente klas model. De schattingsprocedures (o.a. kleinste kwadraten) van deze modellen zijn onder meer beschreven in Fischer (1974, hoofdstuk 11) en Torgerson (1958, hoofdstuk 13).

4 Sociometrische analyse

Interpersoonlijke voorkeuren worden in weergegeven in een sociomatrix met diverse indices voor groepscohesie (klikdetectie).

5 Berekening van een overgangsmatrix

6 Betrouwbaarheid volgens de klassieke testtheorie

5 Evaluatie

Bij deze evaluatie wordt vooral de ATARI gebruikersgroep voor ogen gehouden en er wordt van uit gegaan dat voor deze groep al of niet via MS/DOS of MAC emulaties bekendheid bestaat met statistiekprogramma's onder dergelijke systemen.

ALMO is zeer gebruikersvriendelijk. Er is nauwelijks enige studie vereist om meteen aan de slag te kunnen. Voorbeeldprogramma's maken snel duidelijk hoe de aanroep van een analyse gestructureerd is. Hoewel er aan gewerkt wordt, zijn de grafische toepassingen echter nihil.

Men heeft erg in de breedte gedacht. In feite is er voor iedere probleemstelling in de sociale wetenschappen wel een analyse beschikbaar in ALMO. Nadeel is dat de diepte in sommige deelgebieden ontbreekt. Bijvoorbeeld: men kan weliswaar categorische datanalyse verrichten, maar in vergelijking met hiervoor geëigende programma's of subroutines zijn de opties beperkt (cf. GLIM, LOGLINEAR of HILOGLINEAR in SPSS). Voor diegenen met programma's als SPSS werken is er

niet veel aanleiding om ALMO aan te schaffen. Voor diegenen die uitgebreide statistiek willen bedrijven in de oorspronkelijke ATARI omgeving is aanschaf zeker de moeite waard. De prijs van dit pakket is uiteraard zeer concurrerend. Met ALMO kunnen tevens analyses worden verricht die minder gangbaar zijn in algemeen statistische pakketten, zoals latente trek en structuur modellen. Men dient bij het gehanteerde lineaire model in ALMO wel de nodige voorzichtigheid betrachten bij de statistische inferentie. Het handboek bij ALMO is qua druk, lay-out en binding slecht uitgevoerd. Bovendien moet men niet verwachten dat er veel statistische informatie geboden wordt. Het handboek is wel zeer informatief m.b.t. de aansturing van de vele programma's.

5 Literatuur

- Agresti, A., (1990), *Categorical data analysis*, New York: John Wiley & Sons.
- Aldenderfer, M.S. & Blashfield, R.K., (1984), *Cluster analysis*, Beverley Hills: Sage.
- Aldrich, J.H. & Nelson, F.D., (1984), *Linear probability, logit, and probit models*, Beverley Hills: Sage.
- Coxon, A.P.M., (1982), *Multidimensional scaling*, London: Heinemann.
- Finn, J.D., (1974), *A general model for multivariate analysis*, New York: Holt, Rinehart & Winston.
- Fischer, G., (1974), *Einführung in die Theorie psychologischer Tests*, Bern: Hans Huber.
- Gottman, J.M., (1984), *Time-series analysis*, Cambridge: Cambridge University Press.
- Krauth, J. & Lienert, G.A., (1973), *Die Konfigurationsanalyse (KFA) und ihre Anwendung in Psychologie und Medizin*, Freiburg: Karl Alber.
- Kruskal, J.B. (1964), Non-metric multidimensional scaling, *Psychometrika*, 29, 115-129.
- Kuit, D. (1993), Nieuw geluid uit de oude doos, *Personal Computer Magazine*, 11, 2, 38-43.
- Lazarsfeld, P.F. & Henry, N.W. (1968), *Latent structure analysis*, Boston: Houghton Mifflin.
- Lebart, L., Morineau, A. & Warwick, K.M., (1984), *Multivariate descriptive analysis*, New York: John Wiley & Sons.
- McCullagh, P. & Nelder, J.A., (1989), *Generalized linear models*, London: Chapman & Hall.
- McDonald, R.P., (1969), A generalized common factor analysis based on residual covariance matrices of a prescribed structure, *Br. J. of Math. and Stat. Psychology*, 22, 149-163.
- Meerling, (1981), *Data-analyse en psychometrie*, Meppel: Boom.
- Mulaik, S. A., (1972), *The foundations of factor analysis*, New York: McGraw-Hill.
- Post, W.J., (1992), *Nonparametric unfolding models*, Leiden: DSWO.
- Sachs, L., (1984), *Applied statistics*, New York: Springer.

- Santner, T.J. & Duffy, D.E., (1989), *The statistical analysis of discrete data*, New York: Springer.
- Searle, S.R., (1971), *Linear models*, New York: John Wiley & Sons.
- Smit, P., (1993), *Gemulator*, Atari ST Nieuws, 54, 11-13.
- Tatsuoka, M.M., (1971), *Multivariate analysis*, New York: John Wiley & Sons.
- Torgerson, W.S., (1958), *Theory and methods of scaling*, New York: John Wiley & Sons.
- Yule, G.U. & Kendall, M.G., (1968), *An introduction to the theory of statistics*, London: Charles Griffin.

Ontvangen: 11-10-1993

Geaccepteerd: 12-1-1994