

MOKSCAL
een programma voor dichotome Mokken-schaalanalyse

Rob R. Meijer en Klaas Sijtsma *)

Samenvatting

In deze bespreking wordt een korte inleiding gegeven op het Mokken model. Tevens worden enige methoden vermeld om te onderzoeken of empirische gegevens aan de eisen van het model voldoen. Daarna wordt het programma MOKSCAL besproken. Aan de orde komen algoritmen die in het programma worden gebruikt, werking van het programma, kwaliteit van de uitvoer en kwaliteit van de handleiding. Tenslotte wordt geëvalueerd in hoeverre MOKSCAL iets toevoegt aan bestaande Mokken-schaalanalyse software.

*) Vrije Universiteit, Vakgroep Arbeids- en Organisationspsychologie, De Boelelaan 1081, 1081 HV Amsterdam. Tel. 020-5485502

1. Technische gegevens

Naam	: Mokscal 2.0
Release	: April, 1989
Auteurs	: J. Kingma, T. Taerum
Leverancier	: Computing Services, University of Alberta Edmonton Alberta T6G 2E1, Canada
Besturingssysteem	: MS-DOS
Processor	: 8088, 8086 (XT) met coprocessor AT, harde schijf niet noodzakelijk
Omvang programma	: 268 Kbytes
Prijs	: alleen verzendkosten

2. Theoretische achtergrond

Mokken (1971) stelt twee nonparametrische item response modellen voor: het model van monotone homogeniteit en het model van dubbele monotonie.

Het model van monotone homogeniteit is gebaseerd op drie assumpties: lokaal stochastische onafhankelijkheid, eendimensionele meting en monotonie in de latente trek. Lokaal stochastische onafhankelijkheid wil zeggen dat het antwoordgedrag van een respondent op een willekeurig item niet wordt beïnvloed door zijn/haar antwoordgedrag op de andere items in de test. Eendimensionele meting houdt in dat de antwoorden op de items door een gemeenschappelijke psychologische eigenschap worden verklaard. Monotonie in de latente trek houdt in dat naarmate iemand een hogere waarde op de schaal van de latente trek heeft, hij/zij met grotere waarschijnlijkheid een item positief beantwoordt. Wanneer het model van monotone homogeniteit een goede beschrijving van de data geeft, kunnen personen uit de betreffende populatie worden geordend naar hun meetwaarde op de latente trek.

Het model van dubbele monotonie is gebaseerd op de drie assumpties van het model van monotone homogeniteit plus de assumptie van invariante itemordening (op knopen na) in iedere subpopulatie van de populatie waarvoor het model geldt. Wanneer het model van dubbele monotonie past bij de data kunnen zowel personen als items worden geordend naar respectievelijk de latente trek en de itemmoeilijkheid.

Om te onderzoeken of empirische gegevens kunnen worden beschreven met het model van monotone homogeniteit, gebruikt Mokken (1971) Loevingers (1948) H-coëfficiënt. Mokken laat zien dat bij gegeven monotone homogeniteit geldt dat $0 \leq H \leq 1$. Hierbij wil $H=0$ zeggen dat het antwoordgedrag van de

respondenten op de items past bij de situatie waarin er geen verband is tussen de kans op een correct antwoord en de meetwaarde op de latente trek (voor tenminste $k-1$ van de k items in de test). $H=1$ wil zeggen dat het antwoordgedrag van alle respondenten overeenkomt met het antwoordgedrag zoals dat wordt verwacht onder het Guttman (1950) model. Aangezien een positieve H -waarde geen voldoende voorwaarde is voor monotone homogeniteit, stelt Mokken (1971) voor praktisch gebruik een ondergrens $H=0.3$ voor: naarmate H een hogere waarde aanneemt is er meer zekerheid dat de data voldoen aan het model van monotone homogeniteit. Beneden deze waarde wordt de ordening van personen te onnauwkeurig geacht. Verder spreekt Mokken van een zwakke schaal als $0.3 \leq H < 0.4$, van een matige schaal als $0.4 \leq H < 0.5$ en van een sterke schaal als $0.5 \leq H \leq 1$. De kwaliteit van een individueel item i kan worden geëvalueerd met de coëfficiënt H_i en van paren van items (i,j) met de coëfficiënt H_{ij} . Mokken heeft de steekproevenverdeling voor H , H_i en H_{ij} afgeleid: met behulp van de asymptotisch normaal verdeelde grootheid Δ^* kan worden getoetst of deze coëfficiënten gelijk zijn aan nul tegen het alternatief dat ze groter zijn dan nul. Ook bestaat de mogelijkheid om te toetsen of H in twee of meer populaties identiek is. De hiervoor gebruikte grootheid T is asymptotisch chi-kwadraat verdeeld.

Om te onderzoeken of gegevens passen bij het model van dubbele monotonie stelt Mokken (1971) een visuele inspectie van de P - en $P(0)$ -matrices voor. De P -matrix is van de orde $k \times k$ en bevat als elementen de bivariate proporties positieve antwoorden op tweetallen items. De rijen en kolommen van deze matrix zijn geordend naar oplopende p -waarden (proporties positieve antwoorden) van de k items. Onder het model van dubbele monotonie moet nu gelden dat van boven naar beneden en van links naar rechts de bivariate proporties niet-dalen. De $P(0)$ -matrix is ook van de orde $k \times k$ en bevat de proporties negatieve antwoorden op tweetallen items. Onder het model van dubbele monotonie moeten de rijen en kolommen in deze matrix niet-stijgend zijn. Schendingen in de P - en $P(0)$ -matrix in de steekproef kunnen duiden op het niet passen van het model van dubbele monotonie.

In het laatste decennium is het Mokken model vele malen met succes toegepast (voor een overzicht, zie Sijsma, 1990). Bovendien is het model theoretisch verder uitgediept. Zo ontwikkelde Mokken zijn model voor dichotome data, terwijl Molenaar (1982, 1986) het model uitbreidde voor polytome data. Verder heeft Molenaar (1983) toetsen voorgesteld om steekproefschendingen in de P - en $P(0)$ -matrix te onderzoeken op significantie, om monotone homogeniteit te controleren met behulp van de restscore-bij-items matrix (zie onder) en om dubbele monotonie te controleren met behulp

van de testscore-bij-items matrix. Sijsma en Molenaar (1987) hebben methoden voor de schatting van de betrouwbaarheid zoals voorgesteld door Mokken verder uitgediept en een nieuwe methode voorgesteld. Sijsma (1988) stelt een nieuwe methode voor om gegevens te controleren op monotone homogeniteit en dubbele monotonie. Tevens heeft Sijsma (1988) in het kader van het Mokken model een passingsindex voorgesteld voor personen. Ten slotte stellen Sijsma en Meijer (1991) een nieuwe methode voor om gegevens op dubbele monotonie te controleren.

3. Bestaande Software

De programma's MOKKEN SCALE en MOKKEN TEST (Niemöller, 1980; Niemöller & Van Schuur, 1980) zijn veelgebruikte software programma's voor Mokken-schaalanalyse. Zij zijn alleen ontwikkeld voor mainframe-gebruik en zijn geschikt voor dichotome data. Voor dichotome en polytome data is het programma MSP (recentste release MSP 1.5, zie Debets & Brouwer, 1989) ontwikkeld. MSP is zowel geschikt voor PC- als voor mainframe-gebruik. Voor dichotome data is het programma MOKSCAL (Kingma & Taerum, 1988a; 1988b; 1989) ontwikkeld. MOKSCAL is zowel geschikt voor PC- als voor mainframe-gebruik.

4. Mogelijkheden MOKSCAL

In deze bespreking beperken we ons tot de PC-versie van MOKSCAL. De handleiding (Kingma, TenVergert & Taerum, 1989) richt zich ook alleen tot PC-gebruikers van MOKSCAL.

Het programma bevat een Zoek-, Test- en Robuustheidsprocedure. Met behulp van de bottom-up Zoekprocedure kunnen uit een verzameling items die items worden geselecteerd die een Mokken-schaal volgens het model van monotone homogeniteit vormen. De Testprocedure biedt de gebruiker de mogelijkheid om de schaalbaarheid van a priori gegeven schalen te bepalen, terwijl de Robuustheidsprocedure de mogelijkheid biedt de robuustheid (gelijkheid) van de H-coëfficiënt in verschillende populaties te onderzoeken.

Modelcontrole op monotone homogeniteit kan plaatsvinden door middel van H , H_i en H_{ij} .

Om te controleren of empirische gegevens passen bij het model van dubbele monotonie worden de P- en P(0)-matrix berekend en afgedrukt.

Om de betrouwbaarheid van een schaal te bepalen, berekent het programma de geschatte betrouwbaarheid volgens acht verschillende methoden,

waaronder Mokkens (1971, p. 142-147) methoden 1 en 2, methode MS (Sijtsma & Molenaar, 1987) en de KR-20 (Kuder & Richardson, 1937).

MOKSCAL berekent tevens de punt-biseriële correlatie (item-test correlatie) en de aangepaste punt-biseriële correlatie (item-rest correlatie) voor alle items die aan het programma worden aangeboden alsmede voor deelverzamelingen van items hieruit die te zamen een schaal vormen. Deze correlaties spelen geen directe rol in de schaalconstructie volgens Mokken. Daarentegen spelen ze in de testconstructie volgens de klassieke testtheorie wel een belangrijke rol en kan het handig zijn om over deze extra informatie te beschikken.

5. Bewerking van gegevens

Om het programma te laten draaien dient de gebruiker een file met itemscores en eventuele achtergrondvariabelen en een file met opdrachten (invoer file) te creëren. De invoer file dient gecreëerd te worden in bijvoorbeeld de editor van MS-DOS of een tekstverwerkingspakket en wordt dan weggeschreven als ASCII file. Het programma bevat geen invoerscherm. Alvorens het programma te laten draaien moet de gebruiker aangeven waar hij de uitvoer naar toe wil laten schrijven. De uitvoer kan naar het scherm worden geschreven, naar de printer of naar een uitvoer file. Het is niet mogelijk eerst de uitvoer naar het scherm te sturen om de resultaten te bekijken met het oog op een beslissing over bewaren, afdrucken of vernietigen. Het is dus onmogelijk om interactief met het programma te werken.

Wat betreft het inlezen van de data kunnen twee of meer regels per respondent van de data file worden ingelezen. Tevens biedt het programma mogelijkheden tot datamanipulatie. Zo is het mogelijk subgroepen te onderscheiden op basis van een achtergrondvariabele (bijvoorbeeld, sexe, leeftijd) en het programma apart voor een subgroep te laten draaien (te vergelijken met het 'select if' statement in SPSS), en om variabelen te hercoderen ('recode' statement in SPSS). Bovendien kunnen missing values worden gecodeerd.

6. Werking van het programma

We hebben het programma getest op een Olivetti M250 met mathematische coprocessor. Het programma werd gedraaid op een dataset van 425 personen en 12 items. Een onvolkomenheid is dat niet duidelijk in de handleiding staat dat bij het maken van de invoer file erop gelet moet worden dat bij het

specificeren van de plaats en de naam van de data file het gedeelte achter het '='-teken geheel tussen aanhalingstekens staat. Bijvoorbeeld, als de gegevens file 'mokken1.dat' op de a-schijf staat moet dit worden aangegeven als: **File description File='a:mokken1.dat'**, en niet zoals op sommige plaatsen in de handleiding te lezen staat als: **File description File=a:'mokken1.dat'**. Wanneer deze laatste aanduiding wordt gebruikt krijgt de gebruiker een foutmelding dat de structuur van de data niet klopt.

De Zoekprocedure werd eerst getest op de genoemde data met alle default waarden. Het programma had 9 seconden nodig om de berekeningen uit te voeren. De default waarde voor de ondergrens voor de H-coëfficiënt is in MOKSCAL gelijk aan 0.5. Dit lijkt ons geen gelukkige keuze aangezien Mokken (1971, p. 184) al spreekt van (zwakke) schaalbaarheid wanneer de H-waarden tenminste 0.3 zijn. Een (hoge) ondergrens van 0.5 zal in vele gevallen tot gevolg hebben dat uit een verzameling aangeboden items diverse subschalen worden geselecteerd die uit kleine aantallen items bestaan, en dat items worden weggelaten die wel degelijk een ordening van personen toelaten.

Daarna hebben we het programma eerst laten draaien met als ondergrens voor H de veelgebruikte waarde 0.3 en vervolgens ook met de in het model van monotone homogeniteit minimale waarde 0.0. Het programma bleek zinvolle uitkomsten te geven bij beide ondergrenzen. Het programma gaf zelfs uitkomsten wanneer we een negatieve ondergrens voor H kozen en selecteerde items in de schaal met negatieve H_i -waarden. Aangezien negatieve H_i -waarden niet geoorloofd zijn binnen de Mokken-modellen zou het te verkiezen zijn als het programma dat zou aangeven of zelfs een foutboodschap zou geven (en niet zou draaien).

We hebben bij een ondergrens van $H=0.3$ de numerieke waarden voor H , H_i , H_{ij} en Δ^* vergeleken met de waarden die worden berekend door het programma MOKKEN SCALE. Er waren geen verschillen.

Na de eerste zoekprocedure worden voor de items die zijn geselecteerd de P- en P(0)-matrix gegeven, alsmede de betrouwbaarheid per item en de betrouwbaarheid van de totaalscore van de geselecteerde items. Daarna wordt de zoekprocedure uitgevoerd voor de items die niet tot de eerste schaal zijn toegelaten. De zoekprocedure wordt herhaald totdat er geen schaalbare items over zijn.

Nieuw in vergelijking met bestaande Mokken-schaalanalyse software is dat met een '>'-teken in de P-matrix en een '<'-teken in de P(0)-matrix significante schendingen van de verwachte ordeningen worden aangegeven. Volgens de handleiding wordt bij een steekproefschending in de P- en P(0)-matrix met een Runs test getoetst of deze schending significant is.

Hiervoor wordt verwezen naar Molenaar (1983). Molenaar toetst dergelijke schendingen echter met de McNemar toets (of voor kleine aantallen met de Binomiaal toets). Het is dus onduidelijk wat hier precies wordt getoetst en of de gevolgde procedure wel adequaat is.

Voor de betrouwbaarheidsschattingen methode 1 en 2 volgens Mokken en methode MS volgens Sijsma en Molenaar (1987) is het niet precies duidelijk wat MOKSCAL berekent. Bij methode 1 worden twee schattingen gegeven, terwijl volgens Mokkens theorie een betere schatting wordt verkregen door deze twee afzonderlijke schattingsmethoden te combineren. Deze combinatie wordt in de uitvoer wellicht gegeven onder het kopje 'Mokkn Min', maar dat is verder niet duidelijk. Bij methode 2 is in de uitvoer sprake van een interpolatie- en een extrapolatiemethode; methode 2 is echter nadrukkelijk gebaseerd op interpolatie (zie voor technische details, Mokken, 1971, p.147) en van extrapolatie is beslist geen sprake. Bij methode MS (Sijsma & Mol in de uitvoer) worden een 'lower' en een 'upper' schatting van de betrouwbaarheid gegeven. Volgens Sijsma & Molenaar (1987) dient deze methode echter uit te monden in een enkele schatting. Verder is het opmerkelijk dat MOKSCAL in totaal 7 betrouwbaarheidsschattingen afdruckt (KR-20 niet meegerekend), terwijl de Mokken-benadering voor dichotome items er maar drie kent. Van twee van deze drie schattingen is bekend dat dat ze bij benadering zuiver (Sijsma & Molenaar, 1987) zijn. Waarom er dan nog extra schattingen worden gegeven, die (1) suboptimaal zijn in termen van zuiverheid en (2) nergens gedocumenteerd zijn, is ons niet duidelijk. Tevens zij opgemerkt dat MOKSCAL per item gegevens (bivariate proporties voor twee onafhankelijke replicaties van het item) afdruckt die voor de uiteindelijke betrouwbaarheidsschattingen weliswaar nodig zijn, maar die voor de gebruiker alleen maar verwarrend zijn omdat ze niet gedocumenteerd zijn. Tenslotte wordt ook nog de itembetrouwbaarheid ('Reliab index'?) gegeven. Ook hier geldt weer dat niet duidelijk is hoe deze betrouwbaarheid wordt geschat en hoe de gebruiker deze informatie moet gebruiken bij de evaluatie van de passing van het model of de bruikbaarheid van de uiteindelijke schaal.

Van belang is dat de betrouwbaarheidsschattingen volgens Mokkens methoden 1 en 2 en Methode MS veronderstellen dat de items voldoen aan het model van dubbele monotonie. In de uitvoer van het programma worden de betrouwbaarheidsschattingen voor de gehele aangeboden verzameling van items gegeven, dus ongeacht de passing van het model van dubbele monotonie. Dit werkt wellicht een onjuist gebruik van de coëfficiënten in de hand.

Een ander probleem vormen de waarden van de punt-biseriële en aange-

paste punt-biseriële correlatie. We vonden bij een van de 12 items een waarde voor de punt-biseriële correlatie groter dan een. Berekening van de item-rest correlatie voor de twaalf items met SPSS leverde andere getallen op dan met MOKSCAL. Het is duidelijk dat de vermelde correlaties in MOKSCAL niet de item-test en item-rest correlaties zijn en dat ze wellicht beter niet gebruikt kunnen worden.

Opgemerkt zij dat MSP en MOKKEN SCALE dezelfde zoekprocedure kennen als MOKSCAL, met ongeveer dezelfde uitvoer: MOKSCAL biedt op dit terrein dus niets nieuws.

De bepaling van de a priori schaalbaarheid van de 12 items met de Testprocedure leverde geen nieuwe problemen op. De numerieke waarden van H kwamen overeen met die volgens MOKKEN SCALE.

Een andere mogelijkheid om monotone homogeniteit te controleren is met behulp van de restscore-bij-items matrix. Deze matrix bevat de proportie personen die een positief antwoord op een item geven conditioneel op de restscore (= totaalscore minus score op het betreffende item). Met behulp van deze restscore-bij-items matrix, waarin de kolommen zijn geordend naar oplopende restscore en de rijen zijn geordend naar afnemende itemmoeilijkheid kunnen we de assumptie van monotonie in de latente trek onderzoeken. Gegeven dat de het model van monotone homogeniteit geldt, moeten de proporties in elke kolom in deze matrix van boven naar beneden niet-dalen. Of verstoringen van deze ordening te wijten zijn aan steekproeffluctuaties kan worden getoetst in een twee-bij-twee tabel van restscore-bij-itemscore met de Fisher exact toets (zie Molenaar, 1982, 1983). MOKSCAL biedt (evenals MOKKEN SCALE en MOKKEN TEST) echter geen restscore-bij-items matrix hetgeen deze extra controle op monotone homogeniteit onmogelijk maakt.

Om de robuustheid van H over verschillende populaties te testen verdeelden we de totale groep op basis van een dichotome variabele. Het programma rekende zonder problemen de statistische grootheid T uit. Wel geeft het programma nogal veel uitvoer omdat zowel voor de totale groep als voor de subgroepen de scorepatronen, P- en P(0)-matrix, H_{ij} -matrix en betrouwbaarheidsgegevens worden uitgerekend. De optie ROBUST van MOKSCAL levert overigens grotendeels dezelfde informatie op als het programma MOKKEN TEST (Niemöller & Van Schuur, 1980).

7. Uitvoer

Voorafgaand aan de Zoek- Test- als Robuustheidsprocedure geeft het program-

ma in de uitvoer standaard (default) eerst een overzicht per item welk itemalternatief als het positieve alternatief wordt beschouwd, de waarde van het significantieniveau voor het toetsen of een item past bij de schaal, de waarde voor de opgegeven of gebruikte (default) ondergrens voor H en het aantal items dat in de procedure wordt geanalyseerd. Daarna wordt per item het positieve alternatief nog een keer genoemd, wordt vermeld om welke procedure het gaat (Test, Zoek of Robuust) en worden de eerste vier cases van de datafile afgedrukt en wordt het totaal aantal cases vermeld.

Vervolgens worden alle verschillende responsepatronen (patronen van enen en nullen op de k aangeboden items) die in de steekproef voorkomen afgedrukt met hun frequentie, gesorteerd naar toenemende totaalscore. Tevens wordt de kans op het positief beantwoorden van een item gegeven de totaalscore geplot.

Ook worden de P- en P(0)-matrix voor alle aangeboden items afgedrukt en worden significante schendingen aangegeven met een '>'-teken en een '<'-teken. Als schendingen zich echter uitstrekken over drie of meer items - wat duidt op een relatief ernstige schending - wordt dit echter niet aangegeven. Niet erg fraai is bovendien dat tussen de P- en P(0)-matrix de reeds besproken betrouwbaarheidsresultaten worden afgedrukt.

Voor de aangeboden itemverzameling geeft het programma een matrix waarin de H_{ij} -waarden voor alle itemparen worden vermeld. Numerieke H_{ij} -waarden beneden de ondergrens worden aangegeven met een ster (*). De totale H-waarde wordt niet vermeld.

Na al deze resultaten te hebben gegeven, begint het programma pas aan de gevraagde Zoek-, Test- of Robuustheidsprocedure. Aangaande de Zoekprocedure worden voor elke stap de volgende resultaten vermeld voor de tot dan geselecteerde items: het significantieniveau voor het toetsen of een item bij de schaal past en de bij het significantieniveau behorende z-waarde. De H-waarde, de statistische grootte Δ^* , proporties goed van de reeds geselecteerde items, H_i -waarden en de item- Δ^* . Per stap wordt tevens vermeld welke items zijn verworpen in de zoekprocedure. Als de zoekprocedure is afgesloten worden van de niet geselecteerde items de proportie goed, de H_i -waarden ten opzichte van de items in de schaal en de item- Δ^* vermeld.

Gebruikt de onderzoeker de Testprocedure dan wordt grotendeels dezelfde uitvoer verkregen als bij de Zoekprocedure.

In de uitvoer van de Robuustheidsprocedure geeft MOKSCAL de responsepatronen met de relatieve frequenties, de P- en P(0)-matrices, de H_{ij} -matrix en betrouwbaarheidsresultaten voor de gehele groep en voor de onderscheiden subgroepen. Bovendien wordt nu voor de gehele groep en de

subgroepen een matrix gegeven waarin onder andere opgenomen de H-coëfficiënt en de variantie van \hat{H} en \hat{H}_i ($i=1, \dots, k$). Vooral het afdrukken van de responsepatronen kan zorgen voor veel uitvoer. Het is echter mogelijk dit te onderdrukken.

8. De Handleiding

De ons toegestuurde handleiding bestond uit twee losse gedeelten: de installatiehandleiding en de handleiding waarin staat wat de mogelijkheden van het programma zijn.

Installatiehandleiding. Hierin wordt uitgelegd hoe het programma te installeren en te laten draaien. Ook voor niet ingewijden in het MS-DOS besturingssysteem lijkt deze handleiding goed te volgen. Wel staan er zowel in de installatiehandleiding als in de handleiding storende tikfouten; zo wordt het MS-DOS commando om een directory aan te maken gespeld als MDIR in plaats van MKDIR.

Handleiding. In de handleiding wordt eerst weer vermeld hoe het programma te laten draaien. Slordig is dat hier niet wordt vermeld dat het mogelijk is de uitvoer naar een uitvoer file te laten schrijven (terwijl dit wel degelijk mogelijk is). Daarna wordt ingegaan op de structuur van de invoer file: per record worden de verplichte en optionele opdrachten behandeld die de gebruiker moet/kan intikken.

Na een listing van de bijgeleverde voorbeeld invoer files, wordt een overzicht gegeven van de gebruikte algoritmen. Dit gedeelte van de handleiding bevat naast tikfouten, veel anderssoortige slordigheden en fouten. Een willekeurige greep. Onder het kopje SEARCH PROCEDURE wordt vermeld dat na deze procedure de gevormde schaal moet worden geëvalueerd ('has to be evaluated') aan de hand van de P- en P(0)-matrix. Dit is onjuist. De P- en P(0)-matrix zijn geschikt om te controleren of het model van dubbele monotonie geldt voor de data. Wanneer de onderzoeker genoeg neemt met het zwakkere model van monotone homogeniteit is het echter niet nodig de data te controleren met behulp van de P- en P(0)-matrix. Verder wordt gerefereerd aan Molenaar (1983) die een Runs test zou gebruiken om schendingen in de P- en P(0)-matrix te toetsen op significantie terwijl Molenaar hiervoor de McNemar toets gebruikt.

De uitvoer zou het percentage en het aantal respondenten voor elk van de 2^k mogelijke responsepatronen geven. Bedoeld wordt dat elk in de steekproef voorkomend responsepatroon (aantal meestal aanzienlijk kleiner dan 2^k) wordt vermeld met het aantal respondenten dat dit patroon heeft gegenereerd.

Tot slot zij opgemerkt dat de handleiding onvolledig is. Zo wordt bijvoorbeeld niet vermeld dat er (default) een plotje wordt gemaakt van de kans op het goed beantwoorden van een item conditioneel op de totaalscore. Tevens wordt niet beschreven wat 'delta chi' voorstelt, die vermeld wordt boven iedere stap in de zoekprocedure (ons ook niet bekend uit Mokkens theorie). Ook wordt niet uitgelegd wat de 'p-sub' matrix voorstelt die wordt afgedrukt in de uitvoer van de TEST optie.

9. Conclusie

Het programma MOKSCAL biedt de gebruiker weinig nieuwe mogelijkheden boven de bestaande programmatuur MOKKEN SCALE, MOKKEN TEST en MSP. Omdat het programma alleen geschikt is voor dichotome data kan het het beste worden vergeleken met MOKKEN SCALE en MOKKEN TEST. De PC-versie van MOKSCAL biedt dezelfde mogelijkheden als MOKKEN SCALE en MOKKEN TEST, die alleen op mainframe zijn te draaien, plus nog wat aanvullingen zoals uitgebreide betrouwbaarheidsgegevens, item-rest correlaties en toetsing van schendingen in de P- en P(0)-matrix. Zoals hierboven is betoogd zijn het juist deze aanvullingen op de bestaande programmatuur die vraagtekens oproepen of domweg fouten bevatten.

Literatuur

- Debets, P. & Brouwer, E. (1989). User's Manual MSP: a program for Mokken Scale Analysis for Polychotomous Items, version 1.50. Iec PROGAMMA, Groningen, The Netherlands.
- Guttman, L. (1950). The basis for scalogram analysis. In S.A. Stouffer, L. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star & J.A. Clausen (Eds.), Measurement and Prediction. Princeton: Princeton University Press.
- Kingma, J. & Taerum, T. (1988a). A Scaling program based on a nonparametric item response model: MOKSCAL. Psychometrika, 53, 299-300.
- Kingma, J. & Taerum, T. (1988b). A FORTRAN 77 program for a nonparametric item response model: the Mokken scale analysis. Behavior Research Methods, Instruments & Computers, 20, 471-480.
- Kingma, J. & Taerum, T. (1989). SPSS-X procedure and standalone programs for the Mokken scale analysis: A nonparametric item response model. Educational and Psychological Measurement, 49, 101-136.
- Kingma, J., TenVergert E. & Taerum T. (1989). MOKSCAL version 2.0: A

- computerprogram for a nonparametric item response theory model. Department of Computer Services, University of Alberta, Canada.
- Kuder, G.F. & Richardson, M.W. (1937). The theory of estimation of test reliability. Psychometrika, 2, 151-160.
- Loevinger, J. (1948). The technique of homogeneous tests compared with some aspects of 'scale analysis' and factor analysis. Psychological Bulletin, 45, 507-530.
- Mokken, R.J. (1971). A Theory and Procedure of Scale Analysis. The Hague: Mouton/New York, Berlin: de Gruyter.
- Molenaar, I.W. (1982). Mokken scaling revisited. Kwantitatieve Methoden, 3, no. 8, 145-164.
- Molenaar, I.W. (1983). Rasch, Mokken en Schoolbeleving. In S. Lindenberg & F.N. Stokman (Eds.), Modellen in de Sociologie. Deventer: Van Loghum Slaterus.
- Molenaar, I.W. (1986). Een vingeroefening in item response theorie voor drie geordende antwoordcategorieën. In G.F. Pikkemaat & J.J.A. Moors (Eds.), Liber Amicorum Jaap Muilwijk. Groningen: Econometrisch Instituut.
- Niemöller, B. (1980). Mokken Scale. STAP user's manual (Volume 4). Amsterdam: University of Amsterdam Press.
- Niemöller, B. & Van Schuur, W.H. (1980). Mokken Test. STAP user's manual (Volume 4). Amsterdam: University of Amsterdam Press.
- Sijtsma, K. (1988). Contributions to Mokken's Nonparametric Item Response Theory. Amsterdam: Free University Press.
- Sijtsma, K. (1990). Betekenis en bruikbaarheid van het Mokken model. Nederlands Tijdschrift voor de Psychologie, 45, 77-85.
- Sijtsma, K. & Meijer, R.R. (1991). Invariante itemordening in tests en studietoetsen: een eenvoudige onderzoeksmethode. In J. Hoogstraten & W.J. Van der Linden (red.), Methodologie (Proceedings van de Onderwijsresearchdagen 1991). Amsterdam: SCO Kohnstammfonds.
- Sijtsma, K. & Molenaar, I.W. (1987). Reliability of test scores in nonparametric item response theory. Psychometrika, 52, 79-97.