

F5

LISREL VOOR DUMMIES **EEN STAPSGEWIJZE HANDLEIDING**

Versie 0.1

Harry B.G. Ganzeboom

Vrije Universiteit Amsterdam

23 maart 2009

LISREL VOOR DUMMIES: STAPJE VOOR STAPJE

Stap 1: Schrijf het conceptueel (causaal) model uit

De basis van alle LISREL modellering is het maken van een causaal diagram waarin je via pijlen weergeeft hoe variabelen in causale relaties met elkaar staan. Je moet daarbij een onderscheid maken tussen (A) latente variabelen (de werkelijke kenmerken van de wereld, die je bedoelt te meten) en (B) de geobserveerde variabelen (= hoe de kenmerken van de wereld in je datamatrix terecht zijn gekomen). Om daarvan een causaal diagram te kunnen maken, moet je de volgende zaken weten:

- In welke causale (=tijds)volgorde staan de latente variabelen?
- Welke eerdere variabele (oorzaak) is van invloed op welke latere variabelen (gevolg)?
- Zijn de exogene oorzaakvariabelen (variabelen die beïnvloed worden door geen enkele andere variabele) met elkaar gecorreleerd?
- Welke geobserveerde variabelen (metingen, indicatoren) worden beïnvloed door welke latente variabele?

Stap 2: Bereken een correlatiematrix in SPSS

De databewerking van de geobserveerde variabelen vindt het meest gemakkelijk plaats in SPSS. Zaken om in je SPSS bewerkingen aandacht aan te besteden zijn:

- Bekijk de frequenties en descriptives van alle variabelen waarmee je gaat werken.
- Neem verstandige beslissingen ten aanzien van missing values, uitschieters, meetschalen (bv. 0/1 meting, rangscores, etc.) en de richting van de meting (zorg ervoor dat ‘veel’ en ‘hoog’ hoog scoort en ‘niet’ en ‘weinig’ laag).
- Bekijk missing values in de correlatiematrix en neem een beslissing of je de gegevens listwise of pairwise wilt analyseren. Vaak is het verstandig om beide te doen en de conclusies met elkaar te vergelijken.

Stap 3: Vervoer de correlatiematrix vanuit SPSS naar LISREL

Correlatiematrices kunnen in SPSS op verschillende manieren berekend worden. De standaard [corr] is niet zo handig, het verloopt iets gemakkelijker als je een correlatiematrix genereert via factor of regression. De correlatiematrix die je bij regression krijgt, heeft als handigheid dat je een aparte matrix kunt verkrijgen met de pairwise N.

De correlatiematrix copieer je in Excel. In Excel kun je het aantal decimalen fatsoeneren (neem er ten minste 3) en eventueel fantoom variabelen (bij groepsvergelijking) invoegen.

Copieer de correlatiematrix in Notepad ('kladblok'). Dit verwijdert elke formattering. Sla de correlatiematrix als .txt bestand op in de directory waar je ook je LISREL-syntax opslaat. Je kunt ook de variabelennamen uit Excel naar Notepad meenemen, dat kun je dan weer snel kopiëren naar LISREL.

Stap 4: Het omzetten van het conceptueel model in LISREL code

Het causaal diagram moet nu worden omgezet in een LISREL-diagram. Dit komt er vooral op neer dat je nauwkeurig telt hoeveel ETA/KSI en Y/X variabelen je hebt. De nummering van deze variabelen spelen een grote rol in je LISREL modellering.

Stap 5: Het inlezen van de data (correlatiematrix) in LISREL

De correlatiematrix wordt ingelezen via de DATA regel en daarna een verwijzing naar het bestand die de correlaties bevat. De standaard specificatie voor de DATA regel kan zijn:

DATA NI=kk NO=nnn [NG=gg] [MA=KM]

kk Aantal variabelen in de correlatiematrix
nnn Aantal observaties in de datamatrix (zie SPSS)
gg Aantal te analyseren groepen (meestal 1)

MA=KM betekent dat je een correlatiematrix wilt analyseren. Voor meer gevorderden is er: MA=CM, hetgeen betekent dat je van een covariantiematrix wilt uitgaan. Dit is vooral interessant in geval van meerdere groepen en het veronderstelt dat je inderdaad een covariantiematrix (=correlatiematrix en standaarddeviaties) als data inleest.

NB1: merk op dat je bij LISREL zelf moet aangeven op hoeveel cases de correlatiematrix berust. Het programma weet niet hoe groot de N is. De meeste inferentiele statistiek (standard errors, t-values, en fit-statistics) wordt door deze keuze bepaald.

NB2: als je pairwise deletion of missing values hebt gedaan bij het berekenen van ja correlatiematrix, is de keuze van NO tamelijk arbitrair. Verstandige keuzes kunnen zijn:

- De minimaal geobserveerde N in de correlatiematrix
- De mediane of gemiddelde N.

Er bestaan overigens binnen LISREL veel genuanceerdere manieren om met pairwise data om te gaan, dit is slechts een ruwe methode voor de beginner.

LAbels

varnaam varnaam etc /

De LA regel geeft de variabelennamen aan, zoals ze in de output verschijnen en zoals je ze op de SE regel kunt gebruiken. Het is handig om ze via Excel en Notepad te kopiëren uit SPSS. Verder:

- Kies voor variabelennamen 8 tekens of minder.
- Eindig de regel met een /. Dat betekent: hier houdt het op.

NB1: LISREL is wat liberaler bij variabelennamen dan SPSS. Zo mogen variabelennamen met cijfers beginnen. Een spatie is echter niet toegestaan.

NB2: De namen moeten op de volgende regel staan.

Daarna komt de plek waar de correlatiematrix staat:

KM FU file=bestandsnaam

KM betekent hier dat je een correlatiematrix inleest en FU dat die zowel boven als onder de diagonaal waarden heeft.

NB1: Het bestand moet in dezelfde directory staan als de de lisrel syntax.

NB2: Zorg ervoor dat de bestandsnaam en het pad ernaar toe geen ingewikkelde tekens (zoals een spatie of een apostroph) bevat.

Stap 6: SELECT: Het selecteren en herordenen van de variabelen

Met SELECT geef je aan welke variabelen uit de correlatiematrix je wilt analyseren en in welke volgorde je die wilt zien. Je kunt hiervoor gebruik maken van getallen (op basis van de volgorde op de LA regel) maar ook van de variabelen namen. Het laatste is veel handiger (dan wordt het een bewerking van de variabelennamen die je bij LA hebt gebruikt). Heel belangrijk:

- Eerst dien je de Y-variabelen aan te geven, daarna de X-variabelen
- Eindig de regel met een /. Dit betekent: hier houdt het op.

NB1: De namen van de geselecteerde variabelen moeten op de volgende regel te staan.

NB2: Het is handig na de SELECT een comment regel te zetten die de Y en X variabelen nummert:

```
! Y1 Y2 Y3 X1 X2
```

Stap 7: de MODEL specificatie

In de MODEL regel doe je strategische keuzes wat betreft het model. Je geeft eerst aan hoeveel variabelen van elke soort (ETA, KSI, Y, X) je hebt. Vervolgens geef je de uitgangssituaties aan van de acht LISREL matrices.

Stap 7a: Het aantal variabelen op de MODEL specificatie

```
MODEL NY=ny NX=nx NE=ne NK=nk
```

ny De hoeveelheid geobserveerde Y-variabelen
ne De hoeveelheid veronderstelde ETA-variabelen (latent)
nx De hoeveelheid geobserveerde X-variabelen
nk de hoeveelheid veronderstelde KSI-variabelen (latent)

$Ny + nx$ moet samen de hoeveelheid geselecteerde variabelen zijn.

Stap 7b: De uitgangssituatie van de 8 matrices aangeven

Het LISREL model wordt weergegeven in acht matrices, die als volgt aan elkaar gerelateerd zijn:

	Exogeen	Endogeen
Causale Effecten	GA	BE
Meet Effecten	LX	LY
Residuele (co) -varianties	PH	PS

Latente vars		
Residuele (co) -varianties Geobserveerd	TD	TE

De effectmatrices GA, BE, LX en LY hebben als dimensies: hoeveelheid afhankelijke variabelen * hoeveelheid onafhankelijke variabelen. Deze matrices zijn asymmetrisch: GA 2 3 geeft aan hoe ETA2 wordt beïnvloed door KSI2, GA 3 2 hoe ETA3 wordt beïnvloed door KSI2.

De residuele covariantiematrixes PH PS TD en TE zijn vierkant, hebben een diagonaal en zijn doorgaans symmetrisch: boven en onder de diagonaal staat hetzelfde. PH 3 4 betekent hetzelfde als PH 4 3.

Als je een matrix niet vermeldt op de MODEL regel, kiest LISREL daarvoor een standaard ('default') waarde (zie Kelloway, tabel 5.1). Het is echter een aanrader ze alle acht expliciet te noemen, en de aanbevolen vorm is daarbij:

```
MO NY=5 NE=3 NX=6 NK=4 ga=fu,fi be=fu,fi lx=fu,fi ly=fu,fi ps=sy,fi
ph=sy,fi te=sy,fi td=sy,fi
```

NB1: Bij de MODEL regel moet de informatie juist op dezelfde regel staan als het woord MO!

NB2: een bijzonder keyword is FIXEDX. Dit geeft aan dat al je exogene variabelen allemaal gelijk zijn aan hun indicatoren. In dat geval behoeft je verder niets meer te vermelden over PH en TD.

Stap 8: Het labelen van de latente variabelen

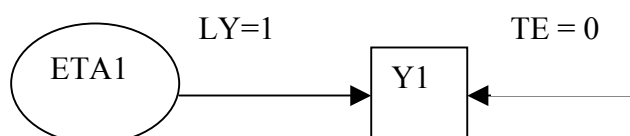
Je kunt namen geven aan de latente variabelen:

```
LE
  NAAM1 NAAM2 /
LK
  NAAM1 NAAM2 /
```

Volg hierbij dezelfde regels als bij **Labels**. Het is prettig om voor de geobserveerde variabelen kleine letters te gebruiken en voor de latente HOOFDLETTERS.

Stap 9: Het specificeren van een meetmodel voor latente variabelen met 1 indicator.

Indien een latente variabele maar door één indicator gemeten wordt, is de gebruikelijke veronderstelling dat deze latente variabelen perfect wordt gemeten door de betrokken indicator:



ST 1 LY 1 1
ST 0 TE 1 1

Of voor een exogene latente variabele:

ST 1 LX 1 1
ST 0 TD 1 1

NB: ST betekent STARTwaarde (niet standaardisatie).

Indien je wilt veronderstellen dat de meting niet perfect is en er dus residuele varantie is, kun je dat aangeven door voor LY een andere constante, lager dan 1 te kiezen en voor de TE de restvariantie op te geven. Het kwadraat van LY en de betrokken TE dienen op te tellen tot de totale variantie van Y, bij gestandaardiseerde geobserveerde variabelen (MA=KM) is dat bv.:

ST .80 LY 1.1
ST .36 TE 1 1

Want $.80^2 + .36 = 1.00$.

Dit heet *correction for attenuation* (Carmines & Zeller, 1979).

Stap 10: Het specificeren van een meetmodel voor latente variabelen met meer dan 1 indicator

Wanneer een latente variabele door meerdere indicatoren wordt gemeten, kunnen de meetrelaties LY (LX) en residuele variaties TE (TD) geschat worden. Er zijn twee manieren om dit op te geven: via standaardisatie en via een referentiemeting.

De meest toegankelijke manier is via stanaardisatie, waar de volgende parameters vrij worden geschat:

FR LY 1 1 LY 2 1
FR TE 1 1 TE 2 2

In dit geval veronderstelt LISREL dat de latente variabele een gestandaardiseerde meeteenheid heeft (zoals de factoren bij factoranalyse in SPSS). Voor exogene variabelen dien je dit echter expliciet aan te geven:

FR LX 1 1 LX 2 2
FR TD 1 1 TD 2 2
ST 1 PH 1 1

Deze aangegeven specificatie brengt LISREL soms in numerieke problemen, in het bijzonder bij meerdere groepen. Een andere systeem is via een referentie-effect:

ST 1 LX 1 1
FR LX 2 1

FR TD 1 1 TD 2 2
FR PH 1 1

Je kiest in deze specificatie een referentie-effect (LX 1 1), dat aangeeft dat de meeteenheid van de latente variabele KS1 gelijk is aan de meeteenheid van geobserveerde variabele X1. De overige geschatte effecten worden dan uitgedrukt ten opzichte van deze referentie. Als je deze specificatie kiest, is er verschil tussen de LISREL estimates (de ongestandaardiseerde oplossing) en de gestandaardiseerde oplossing(en).

NB1: Het kiezen van een referentiemeting is niet hetzelfde als te veronderstellen dat de betrokken indicator de latente variabelen perfect meet. De variantie van de geobserveerde variabelen blijft gelijk aan de variantie van de latente variabelen + de residuele variantie.

NB2: het is gebruikelijk en nuttig om het referentie-effect te kiezen bij de indicator die de geringste residuele variantie heeft.

Stap 11: Het schatten van gecorreleerde (systematische) meetfouten

Systematisch meetfouten kun je laten schatten via de off-diagonale elementen van de TD en TE matrices, Bv.:

FR TD 1 3

Gegeven het symmetrisch karakter van TD en TE is dit hetzelfde als:

FR TD 1 3

Gecorreleerde residuen dienen inhoudelijk gemotiveerd zijn, dat wil zeggen je moet een redelijke theorie hebben waarom ze zouden optreden. Mathematisch gezien zijn slechts een beperkt aantal van dit type termen toegestaan.

NB1: Het specificeren van gecorreleerde residuen is in feite weinig anders dan het veronderstellen van een extra exogene variabele die op twee metingen een identiek effect heeft. Er zijn situaties waarin het gebruik van een extra KS voor het beoogde doel beter bruikbaar is: denk bv. aan het modelleren van een antwoordtendentie die zich in meer dan twee indicatoren voordoet.

NB2: LISREL lijkt geen mogelijkheden te bieden om de residuen van X en Y variabelen te laten correleren. Er bestaat echter een negende matrix, waarin je dit soort effecten kunt aangeven de THETA-DELTA-EPSILON matrix, afgekort als TH

Stap 12: Het schatten van correlaties tussen exogene latente variabelen

Exogene variabelen zijn in observatie-studies gecorreleerd. Deze correlaties dienen gefit te worden en men doet dat meestal door ze te schatten als de off-diagonale elementen van de PH matrix:

FR PH 1 2 PH 1 3 PH 2 3

Omdat de schattingen gelijk zijn aan de geobserveerde correlaties is het woord 'schatting' hier wat overdreven. Je kunt ze ook een ST waarde gelijk aan de geobserveerde correlaties geven:

ST 0.43 PH 1 2

NB1: LISREL heeft een speciale voorziening voor het geval dat alle exogenen gelijk zijn aan hun indicator en met elkaar gecorreleerd zijn: FIXEDX op de MODEL regel. Dat scheelt een hoop typewerk.

NB2: Er zijn ook gevallen denkbaar waarin je off-diagonale elementen van de PH matrix op 0 wilt fixeren. Denk bv. aan de correlatie tussen geslacht en opleiding van ouders, of aan de correlaties tussen twee experimentele variabelen in een multi-factorieel experiment. Ook zijn er gevallen denkbaar dat je elementen in de PH matrix aan elkaar gelijk wilt zetten, bv. de correlatie van vaders opleiding en moeders opleiding met leeftijd.

Stap 13: Verklaarde en onverklaarde variantie in de latente variabelen

Ook latente variabelen hebben verklaarde en onverklaarde variantie. Bij exogene variabelen (KSI) is het percentage onverklaarde variatie per definitie 100% -- dat is de definitie van een exogene variabelen. Een veel gebruikte specificatie is daarom:

ST 1 PH 1 1 PH 2 2 etc.

De diagonale elementen van PHI worden dus gefixeerd (als ze dat al niet zijn) en gestart op 1. Dit betekent dat er geen verklaarde variantie wordt geschat en dat de totale variantie van de betrokken KS variabelen op 1 wordt gezet; deze variabelen zijn dan gestandaardiseerd.

Voor de ETA variabelen is de specificatie van de diagonale elementen van de PS-matrix doorgaans:

FR PS 1 1 PS 2 2 etc.

Stap 14: Structurele causale effecten: GAMMA en BETA

Vervolgens geef aan welke structurele effecten je geschat wilt zien. Voor effecten van exogene variabelen heb je de GA parameters, voor de effecten van andere endogene variabelen de NE:

FR GA 3 1

Geeft aan de invloed van KS1 op ETA3.

FR BE 2 1

Geeft aan de invloed van ETA1 op ETA2.

Doorgaans zijn deze coëfficiënten het belangrijkste doel van je analyse. Ze geven de regressievergelijkingen tussen de latente variabelen aan.

Stap 15: Residuele covarianties tussen endogene variabelen

Het is ook mogelijk om residuele covarianties tussen de ETA te laten schatten:

FR PS 1 2

Dit geeft aan dat je veronderstelt dat er een onbekende bron van covariatie tussen ETA1 en ETA2 is. Je gebruikt zoiets als je tussen deze twee ETA's geen BE wil veronderstellen, met andere woorden de causale richting van de relatie niet kent.

NB1: Tussen twee ETA's kun je niet tegelijkertijd een BE en een PS laten schatten.

NB2: Je kunt een off-diagonale ETA ook zien als een wederzijds effect:

FR BE 1 2 BE 2 1
EQ BE 1 2 BE 2 1

Stap 16: Equality constraints

Met Equality constraint kun je afdwingen dat twee vrije parameters op dezelfde waarde geschat worden. De syntax is eenvoudig, bv.

EQ LY 1 1 LY 2 1 LY 3 1

Dit betekent dat de drie LY dezelfde waarde zullen krijgen. Dit kost dan maar 1 vrijheidsgraad ipv drie.

Equality constraints [algemeen: constrained estimation] vormen een van de grote meerwaarden van LISREL boven andere vormen van regressie- en factoranalyse. Je gebruikt ze om interessant nulhypotesen over de gelijkheid van effecten te toetsen. Zulke vragen zijn bv. interessant bij meetmodellen (zijn verschillende indicatoren even scherpe metingen?) of vergelijking van effecten van vaders / moeders.

NB1: Equality constraints vereenvoudigen een model. Ze zijn daarom soms ook nuttig om identificatieproblemen op te lossen. Ook verhogen ze de *power* van je model: in een vereenvoudigd model zijn minder parameters te schatten en de SE's kleiner, dus zijn coëfficiënten eerder significant.

NB2: EQ is slechts een van de vormen van constrained estimation die LISREL ter beschikking heeft. Er is hier nog veel meer te doen.

Stap 17: Startwaarden voor ongespecificeerde parameters

Om het iteratief algoritme te kunnen beginnen heeft LISREL startwaarden nodig. Doorgaans kan het programma deze zelf berekenen, maar bij wat ingewikkelde modellen is het nuttig om het programma een handje te helpen, ook bij de vrije [te schatten] parameters. Je kunt van elke parameter een startwaarde opgeven, bv.:

ST .37 BE 2 1

Als 0.37 dicht bij de oplossing zit, zal het LISREL algoritme sneller convergeren. Doorgaans is het heel specifiek opgeven van startwaarden voor de te schatten parameters niet noodzakelijk, maar helpt het volgende:

ST 0.5 all

Dit geeft de startwaarde van alle vrije parameters die tot dan geen startwaarde hebben gekregen.

NB1: Indien het programma convergentieproblemen meldt, kun je de waarde 0.5 wat variëren om te zien of hij voorbij het moeilijke punt komt.

Stap 18: PD: Het Path Diagram

PD geeft een plaatje van het causaal diagram met schattingen, indien een model succesvol geschat (hetgeen overigens niet betekent dat de geschatte waarden de beste oplossing vormen). Zolang het plaatje niet verschijnt, weet je dat er nog iets mis is.

Stap 19: De OUTPUT regel

De LISREL syntax sluit af met een specificatie wat voor soort output je wilt zien. Een goede keuze hier is:

OU ML SC TV rs nd=3 ad=off

Dit betekent:

ML	gebruik maximum likelihood voor schatting
SC	bereken standardized solutions
TV	geeft SE's en T-waaren
RS	Geef geschatte correlatuematrix en residuen
ND=3	Print drie decimalen
AD=OFF	ga door met iteraties ook al lijkt het niets te worden

Voor (veel) meer mogelijkheden, zie Kelloway.

Stap 20: F5

Met F5 zet je LISREL in werking. Dit heeft als automatisch gevolg dat je syntax naar disk wordt geschreven en de eerdere syntax en output daarvan worden overschreven. Wilde je die bewaren, dan had je dat eerder moeten bedenken.

Stap 21: Het lezen van het plaatje

Als het path-diagram te voorschijn komt, is het model schatbaar gebleken. Dat betekent dat je LISREL syntax werkt, maar niet dat de geschatte getallen de goede zijn.

Het plaatje bevat twee goodness-of-fit statistics: de CHI2 met geassocieerde overschrijdingskans en de RMSEA. De CHI2 moet zo laag mogelijk zijn en in het ultieme geval niet-significant. De RMSEA moet dichtbij .05 of lager zijn. Helaas bevat het PD niet de overschrijdingskans van de RSEA.

In het plaatje moet je in eerste aanleg checken of je bedoelde model goed in de syntax terecht is gekomen. Het helpt daarbij om bij Estimates en Model de verschillende opties te doorlopen.

Als je plaatje erg op spaghetti lijkt, kan je model wel goed zijn, maar ben je onhelder geweest in je programmering. Door je variabelen consequenter te nummeren, kun je veel duidelijkheid scheppen.

NB1: Het PD plaatje is in feite ook het werkscherm voor interactieve vormen van LISREL gebruik. Blijf hiervan weg, dit is voor de Playmobil generatie.

Stap 22: Het lezen van de output

Na het plaatje begin je de output te bestuderen. Achtereenvolgens treft je hierin aan:

- Een afdruk van je syntax.
- (Wanneer de schatting niet gelukt is:) Alle te schatten matrices met daarin de genummerde vrije parameters.
- (Wanneer de schatting niet gelukt is:) Alle te schatten matrices met daarin de initiële schatting van vrije parameters.
- De LISREL estimates, tezamen met standard errors en t-values.
- De Goodness of Fit statistics
- De Fitted Covariance Matrix en Fitted Residuals en een heleboel meer.
- De Standardized Solution en de Completely Standardized Solution

Als er problemen zijn, zoek je in eerste instantie naar Errors en Warnings. De Estimates en Fitted Residuals geven verder vaak waardevolle suggesties over waar de schoen wringt.

Stap 23: De LISREL Estimates

De Lisrel-estimates bevatten de getallen waarin je uiteindelijk geïnteresseerd bent, de regressie-coëfficiënten BE en GA en de factorladingen (meetrelaties) LX en LY, alsmede de residuële (co)varianties en de bijbehorende proporties verklaarde variantie voor de latente en de geobserveerde variabelen. Ze hebben allemaal een SE en een T-waarde ($T = \text{parameter} / \text{SE}$). Een parameter is significant als $T < -2$ of $T > 2$.

In je uiteindelijke verslag vermeld je de coëfficiënten zoals je dat ook bij een gewone regressie en factoranalyse zou doen.

NB1: De “Reduced Form Estimates” kun je doorgaans overslaan.

Stap 24: De FIT statistics

De Goodness of Fit statistics zijn er vele. Kelloway besteedt er veel aandacht aan, zie aldaar. Wat mij betreft gaat het om twee zaken:

- De Chi-2 met bijbehorende vrijheidsgraden. Deze dient zo klein mogelijk te zijn en het liefst niet-significant. Niet-significantie hier betekent dat de empirische correlatiematrix slechts binnen toevalsgrenzen afwijkt van een veronderstelde populatiematrix van de vorm zoals die staat bij “Fitted Covariance Matrix”.

- De RMSEA en de bijbehorende Test of Close Fit. De RMSEA is een gewogen functie van de residuele correlaties en de N waarop de correlaties berusten. De normwaarde voor de RMSEA is .05. De test geeft aan of de afwijking tussen Fitted Covariance Matrix en de empirische correlatiematrix groter is dan deze normwaarde. Als deze test niet significant is, kun je het model als bevredigend beschouwen.

NB1: Er zijn twee soorten CHI-2. Ze liggen dicht bij elkaar voor redelijk goed passende modellen.

NB2: Significantie-testen van LISREL modellen zijn een dubieuze zaak. Ze sturen aan op een al-dan-niet beslissing daar waar het bij uitstek gaat om mate-van. De inferentiele statistiek is gevoelig voor afwijkingen van multinormaliteit in de data en vanzelfsprekend voor de keuze die je bij NO=nnn hebt gemaakt.

NB3: De index die je verder het meest in de literatuur ziet is de AGFI. Deze dient boven de 0.95 te zijn.

Stap 25: Fitted Covariance Matrix en Fitted Residuals

De Fitted Covariance Matrix geeft aan wat de geschatte coëfficiënten impliceren over de correlaties tussen de geobserveerde variabelen. Hoeveel ze daarop lijken staat aangegeven bij de Fitted Residuals. Een positieve waarde geeft aan dat een correlatie door het model wordt overschat, een negatieve dat een correlatie wordt onderschat. Je bent in deze matrix niet alleen op zoek naar uitbijters, maar ook naar patronen van residuen met hetzelfde teken.

De Fitted Residuals worden gevolgd door allerlei informatie over gestandaardiseerde residuen, een grafiek en een lijst van grote residuen. Aan deze informatie heb je meestal niet zoveel.

NB1: Residuen geven lang niet altijd direct aan waar het model wringt. De schattingsprocedure heeft de neiging om *lack-of-fit* over zoveel mogelijk residuen uit te smeren. Ook bij regressie-analyse geldt dat een invloedrijk punt lang niet altijd een punt met een groot residu is!

NB2: Meer directe informatie over invloedrijke delen van het model geven de zgn. Modification Indices (te verkrijgen met de parameter MI op de OUTPUT regel). Deze geven aan hoeveel de CHI2 zou dalen als je een bepaalde parameter vrij zou laten. Indien met zorg bestudeerd, kan dit helpen problemen op te sporen.

Stap 26: Standardized Solutions

Er zijn twee soorten standardized solutions:

- Standardized Solution: alleen de latente variabelen zijn gestandaardiseerd.
- Completely Standardized Solution: ook de geobserveerde variabelen zijn gestandaardiseerd.

Als je MA=KM hebt opgegeven en je hebt de meetschaal van de latente variabelen geïdentificeerd via standaardisatie (zie stap XX), dan zijn beide gestandaardiseerde oplossingen gelijk aan de LISREL Estimates.

Stap 27: Het bijstellen van het model

Als je eenmaal een schatbaar model hebt, begint het eigenlijke modellerwerk pas. Via het vrijmaken van meer parameters, het weglaten (op 0 fixeren) van niet-significante parameters en het gebruik van constrained estimation (EQ) werk je naar een model toe dat het beste bij de data past. Je voorkeur ligt daarbij bij een model dat zoveel mogelijk vrijheidsgraden (en dus zo weinig mogelijk geschatte parameters) combineert met een zo laag mogelijke CHI2. Je vergelijkt daarbij een reeks van modellen die op elkaar stapsgewijs volgen doordat er telkens een stapje wordt genomen en getoets wordt of de CHI-2 significant verandert.

- Een significante verhoging van de CHI-2 doet je meestal terugkeren op je pad,
- Een significante verlaging van de CHI-2 is doorgaans een goede stap.

Als je stapjes telkens gaan over 1 parameter en dus een vrijheidsgraad, dan is het goed te weten dat de kritieke waarden van Chi2 voor 1 vrijheidsgraad liggen op 3.84 ($p=.05$) en 6.64 ($P=.01$).

MEER GEAVANCEERDE ONDERWERPEN

- Groepsvergegelijkingen en gestandaardiseerde oplossingen
- Fantoomvariabelen
- Ongestandaardiseerde oplossingen bij analyse van een covariantiematrix.
- Het vergelijken van gemiddelden en het schatten van intercepten.
- Latente groeicurves
- Maximum likelihood estimation van data met MCAR missing values.

Informatie over deze onderwerpen in een volgende versie.

VEEL VOORKOMENDE PROBLEMEN BIJ HET LISREL PRACTICUM

SPSS: Het maken van de correlatiematrix: Hoe doe je dat ook alweer (met factor of regressie).

SPSS: let op het verschil tussen pairwise en listwise deletion of missing values. Dit zie je het mooiste als je regressie gebruikt en aangeeft /DES=CORR N.

DATA: er niet om denken dat als je een 12*12 correlatiematrix inleest, je moet aangeven: NI=12.

SELECT: de Y-variabelen moeten voorop staan, dan pas de X-variabelen. Gebruik de variabelennamen van de LA om de herordening te maken.

SELECT: altijd afsluiten met een slash.

MODEL: Je moet eerst je model tekenen voordat je een beslissing kunt nemen over NE, NY, NK en NX.

MODEL: let erop dat NX+NY altijd gelijk is aan het aantal geselecteerde variabelen op SELECT.

MODEL: geef alle matrixen aan: GA=fu,fi BE=fu,fi LY=fu,fi, LX=fu,fi PS=sy,fi PH=sy,fi, TD=sy,fi TE=sy,fi

MODEL: als je het intypen van de PH matrix moe bent, is een handig alternatief: PH=SY,FR, en op vervolgregels de diagonale elementen van PH te fixeren.

LY: voor latente variabelen met maar een indicator kies je doorgaans LY = 1, TE=0

LY: voor latente variabelen met meerdere indicatoren kies je doorgaans FR LE en FR TE.

LX: datzelfde geldt voor LX.

ST: Soms geeft Lisrel een waarschuwing dat hij geen goede startwaarden kan vinden. Een probaat middel is soms: ST .5 all.

VEEL GESTELDE VRAGEN / VRAGEN DIE VEEL GESTELD MOETEN WORDEN

Wat is nu precies het voordeel van LISREL boven conventionele analyses met betrouwbaarheids-, factor- en regressieanalyse?

- Factoranalyse in SPSS is een explorerende techniek waarin je de uiteindelijke oplossing maar betrekkelijk weinig kunt sturen. Bij LISREL is het een confirmatoire techniek die je de gelegenheid geeft je theorie in een model uit te drukken.
- Moeilijke kwesties in factoranalyse als rotatie en hoeveelheid factoren blijken in LISREL fictieve problemen.
- Betrouwbaarheids- en factoranalyse in SPSS hebben geen inferentiele statistiek, LISREL geeft SE's voor alle geschatte grootheden.
- In SPSS betrouwbaarheidsanalyse schat je alleen maar de (on)betrouwbaarheid (random meetfouten) van je meting, in LISREL corrigeer je voor onbetrouwbaarheid.
- LISREL staat je toe ook systematische meetfouten op te sporen, te modelleren en zo te corrigeren.
- Via constrained estimation kun je je theoretische ideeën modelmatig en toetsbaar voorstellen.
- Bij LISREL kun je partiële effecten op 0 fixeren en zo het model vereenvoudigen.
- Bij LISREL wordt een parameter (regressie- of meetcoëfficiënt) bepaald op basis van alle mogelijke informatie, bij SPSS gebeurt het alleen op basis van het direct relevante deel van de correlatiematrix.
- Multiple groepen geven elegante modellen voor landenvergelijking, maar ook voor missing value patronen.

Maar de twee belangrijkste voordelen van LISREL zijn eigenlijk:

- LISREL dwingt je om je theorieën in een helder geformuleerd causaal diagram neer te leggen. Het leert je om over de werkelijkheid na te denken vanuit causale verklaringen.
- LISREL dwingt je na te denken over meten en de consequenties van meetfouten voor inhoudelijke conclusies.

Waarom moet het zo ingewikkeld, kan het niet wat eenvoudiger?

Je kunt LISREL modellen een stukje eenvoudiger maken door exogene variabelen weg te laten: $NX=0$ $NK=0$. Dan vervallen de GA, PH, LX en TE matrix. Je moet dan echter wel weten hoe je exogene variabelen in het endogene gedeelte (PS, BE, LY en TE) neerlegt. De reden waarom deze lijn in de cursus niet gevolgd wordt, is omdat dan Kelloway onleesbaar wordt.

Een andere mogelijk vereenvoudiging is het gebruik van de SIMPLIS syntax. Dit is een vereenvoudigend alternatief dat wat dichter tegen de SPSS syntax aanligt. Hoe dit moet, kun je vinden in een manual dat zich op de LISREL CD-Rom bevindt. Ik zie er geen voordelen in.

LISREL kent ook een grafische interface voor de playmobiel generatie.

Als LISREL zo goed is, waarom doet niet iedereen het?

LISREL (en soortgelijke programma's) is niet zo populair omdat:

- Het toch nog tamelijk ingewikkeld is om te bedienen en erover te rapporteren.

- LISREL sterk leunt op de veronderstelling dat de wereld bestaat uit metrische variabelen, die met elkaar een multivariaat normale verdeling vormen. Veel onderzoekers vinden (ordinale, nominale) discrete metingen en daarbij behorende modellen realistischer, maar verliezen daarbij uit het oog dat ook (juist) zulke gegevens geplaagd worden door meetfouten.