

<b>Introductie</b>	9
Hoofdstuk 1	
<b>VOORBEREIDING: DE EUROPEAN SOCIAL SURVEY</b>	11
1.1 Toegang tot de ESS	11
1.2 De ESS Downloaden	12
1.3 De ESS gebruiksklaar maken	13
Hoofdstuk 2	
<b>INLEIDING TOT SCHAALTECHNIEKEN</b>	15
2.1 Wat is een schaal?	15
2.2 Dimensionaliteit van een schaal testen	18
2.3 Betrouwbaarheid van een schaal testen: betrouwbaarheidsanalyse	20
Hoofdstuk 3	
<b>THEORETISCHE ACHTERGROND BIJ FACTORANALYSE</b>	21
3.1 Assumpties achter factoranalyse	21
3.2 Voorwaarden om tot een goed factormodel te komen	23
3.3 Families van factoranalyse	23
3.4 Een begrippenkader voor factoranalyse	25
3.5 Het principale componentenmodel (PCA)	27
3.5.1 Het bepalen van het aantal componenten	31
3.5.2 Het roteren van componenten	33
3.5.3 Het interpreteren van componenten	35
3.6 Het principale factormodel (PFA)	37
3.6.1 Het bepalen van het aantal factoren	41
3.6.2 Het roteren van de factoren	41
3.6.3 Het interpreteren van factoren	42

3.7	Het verschil tussen PCA en PFA	42
3.8	De verwoording van de items	43
3.8.1	De response set	43
3.8.2	De richting van de items	43

#### Hoofdstuk 4

<b>ANALYSE 1. PRINCIPALE COMPONENTENANALYSE (1 COMPONENT)</b>		45
4.1	Syntax	45
4.2	Output	47
4.2.1	Output 1: Beschrijvende analyse van de variabelen	47
4.2.2	Output 2: Inspectie van de assumpties en de voorwaarden	49
4.2.3	Output 3: Het bepalen van het aantal componenten	53
4.2.4	Output 4: Het interpreteren van de componenten	55
4.3	Deze analyse in SAS	56
4.3.1	Syntax	57
4.3.2	Output	58

#### Hoofdstuk 5

<b>ANALYSE 2. EEN PRINCIPALE COMPONENTENANALYSE (MEERDERE COMPONENTEN)</b>		65
5.1	Syntax	65
5.2	Output	66
5.2.1	Output 1: Beschrijvende analyse van de variabelen	66
5.2.2	Output 2: Inspectie van de assumpties	67
5.2.3	Output 3: Het bepalen van het aantal componenten	70
5.2.4	Output 4: Het interpreteren van het aantal componenten	72
5.3	Deze analyse in SAS	76
5.3.1	Syntax	76
5.3.2	Output	76

#### Hoofdstuk 6

<b>ANALYSE 3. EEN PRINCIPALE FACTORANALYSE</b>		85
6.1	Syntax	85
6.2	Output	86
6.2.1	Output 1: Beschrijvende analyse van de variabelen	86
6.2.2	Output 2: Inspectie van de assumpties	86
6.2.3	Output 3: Het bepalen van het aantal factoren	88
6.2.4	Output 4: Het interpreteren van de factoren	90
6.3	Deze analyse in SAS	91

6.3.1	Syntax	91
6.3.2	Output	92
Hoofdstuk 7		
<b>THEORETISCHE ACHTERGROND BIJ PRINCIPALE COMPONENTENANALYSE VOOR CATEGORIALE DATA</b>		97
7.1	Principes van PRINCALS	97
7.2	Het PRINCALS-model	98
7.2.1	Het bepalen van het aantal componenten	100
7.2.2	Het interpreteren van componenten	101
7.2.3	Over PRINCALS als iteratieve procedure	103
Hoofdstuk 8		
<b>ANALYSE 4. EEN PRINCALS ANALYSE</b>		107
8.1	Syntax	107
8.2	Output	109
8.3	Deze analyse in SAS	116
Hoofdstuk 9		
<b>THEORETISCHE ACHTERGROND BIJ BETROUWBAARHEIDSANALYSE</b>		117
Hoofdstuk 10		
<b>ANALYSE 5. BETROUWBAARHEID</b>		121
10.1	Syntax	121
10.2	Output	122
10.3	Deze analyse in SAS	125
10.3.1	Syntax	125
10.3.2	Output	126
Hoofdstuk 11		
<b>THEORETISCHE ACHTERGROND BIJ HET INTRODUCEREN VAN CONCEPTEN IN MULTIVARIATE ANALYSES</b>		127
11.1	Stapsgewijze aanpak	127
11.2	Factorscores	128
11.3	Somschalen	129
11.3.1	De niet-restrictieve somschaal	130
11.3.2	De restrictieve somschaal	132
11.3.3	De tussenoplossing	132
11.4	Welke keuze: factorscores of somscores?	134
11.5	Dezelfde berekeningen in SAS	136

Hoofdstuk 12	
<b>ANALYSE 6. INTEGRATIE VAN EEN CONCEPT IN EEN MULTIPLE REGRESSIE</b>	139
12.1 Stap 1: Factoranalyse	140
12.2 Stap 2: Betrouwbaarheidsanalyse	141
12.3 Stap 3: De schaal aanmaken	141
12.4 Stap 4: De schaal in een regressiemodel inbrengen	144
12.4.1 Fase 1. Theoretisch-Conceptuele fase	145
12.4.2 Fase 2. Uni- en bivariate verkenning	146
12.4.3 Fase 3. Eerste inhoudelijke inspectie	152
12.4.4 Fase 4. Controle van de assumpties van het regressie- model	153
12.4.5 Fase 5. Multicollineariteit	157
12.4.6 Fase 6. Residuanalyse	157
12.4.7 Fase 7. Tweede inhoudelijke inspectie	159
12.5 De SAS-syntax van deze analyse	161
12.5.1 Stap 1: Factoranalyse	161
12.5.2 Stap 2: Betrouwbaarheidsanalyse	161
12.5.3 Stap 3: De somschaal maken	162
12.5.4 Regressie Fase 2. De uni- en bivariate verkenning	162
12.5.5 Regressie Fase 3. De eerste inhoudelijke inspectie	166
12.5.6 Regressie Fase 4. Controle van de assumpties	166
12.5.7 Regressie Fase 5. Controle op multicollineariteit	167
12.5.8 Regressie Fase 6. Residuanalyse	168
12.5.9 Regressie Fase 7. De tweede inhoudelijke inspectie	168
12.6 De SAS-output van deze analyse	169
Hoofdstuk 13	
<b>WETENSCHAPPELIJK RAPPORTEREN OVER FACTORANALYSE</b>	171
13.1 Inleiding	171
13.2 Bespreking van het voorbeeldmodel van PCA en betrouwbaarheidsanalyse	173
Hoofdstuk 14	
<b>FACTORANALYSE IN SPSS EN SAS</b>	177
14.1 Via de menu's (SPSS)	177
14.2 Via de FACTOR syntax (SPSS)	181
14.3 Via de PROC FACTOR syntax (SAS)	184
Hoofdstuk 15	
<b>8   INDEX</b>	189

# Introductie

Statistiek en multivariate onderzoekstechnieken zijn vaak een nachtmerrie voor studenten. Binnen de sociale wetenschappen zijn statistische vakken niet zelden vakken met lage slaagcijfers en hoge vervloekingsgraden. Minder evident, maar even reëel, is de afkeer van vele onderzoekers van de technische statistische literatuur. Niet zelden laat men gevorderde multivariate technieken in het sociaalwetenschappelijk onderzoek links liggen omdat de gevorderde handboeken té wiskundig of té ontoegankelijk geschreven zijn. Binnen de faculteit Politieke en Sociale Wetenschappen was de Methodengroep jarenlang verantwoordelijk voor het adviseren en assisteren van studenten en onderzoekers bij het toepassen van eenvoudige en complexe multivariate technieken in het dagelijks onderzoek. Of het nu thesisonderzoek is of fundamenteel wetenschappelijk onderzoek, dezelfde vragen komen steeds terug. Daarom werd gestart met het ontwikkelen van een reeks MTSO-INFO. De basisfilosofie van deze reeks was statistische software en statistische analysetechnieken zodanig uit te leggen, dat wie van een absoluut nulniveau start, met een minimum aan inspanning in staat is om gevorderde analyses uit te voeren.

Toen stilaan duidelijk werd dat de reeks niet alleen binnen de eigen faculteit gebruikt werd maar ook daarbuiten, werd met Acco samengewerkt aan het publiceren van de handleidingen in de reeks *Stap In Statistiek en Onderzoek*. Het uitgangspunt van deze reeks is exact dezelfde als die van MTSO-INFO: een instap bieden voor het werken met statistische software, om zo te komen tot een begrijpen en uitvoeren van gevorderde analysetechnieken. We beperken bewust de hoeveelheid theorie. We geven kort een inleiding over een techniek en gaan dan onmiddellijk aan de slag. Daarbij wordt de wiskundige achtergrond van de technieken in sterke mate uit de weg gegaan. Wie nadien meer wil weten over de techniek, de achtergronden en wie de uitgebreidere toepassingen ervan wil leren kennen, wordt doorverwezen naar handboeken statistiek. De reeks *Stap In Statistiek en Onderzoek* wil bewust geen statistiek-

reeks zijn, maar een praktijkreeks met een lage instapdrempel. Daarom staat de output van de statistische procedures en de programma's die tot die output leiden centraal. Het uiteindelijke doel van een analyse moet immers het begrijpen van de sociale werkelijkheid zijn. Dat begrijpen komt voort uit het correct lezen en interpreteren van statistische output. Toch willen we benadrukken dat het regelmatig consulteren van een handboek statistiek naast deze reeks noodzakelijk is om de finesses die in de reeks aan bod komen beter te begrijpen.

In dit volume staan we stil bij factoranalyse. Factoranalyse is een familie van technieken waaronder veel varianten een plaats vinden. In dit boek behandelen we twee klassieke vormen: de principale componentenanalyse (PCA) en de principale factoranalyse (PFA). Hoewel de principale factoranalyse dominant is in het sociaal wetenschappelijk onderzoek, is het belangrijk om het verschil met de principale componentenanalyse aan te geven. Maar we blijven niet stilstaan bij deze technieken alleen. We plaatsen het gebruik van beide technieken in een ruimer kader waarbij we niet alleen ook de betrouwbaarheidsanalyse bekijken maar ook de plaats van latente factoren (op basis van somschalen of factorscores) in multivariate analyses kaderen. Als voorbeeld kozen we om een schaal te gebruiken in een regressieanalyse om op die manier aansluiting te vinden bij het voorgaande volume in deze reeks. Naast deze brede kijk op de plaats van factoranalyse in het onderzoek, verdiepen we onze blik ook door kort in te gaan op een van de niet-interval varianten van de factoranalyse: Princals. Zo wordt een brug geslagen met de categoriale statistiek en de technieken die daar voor datareductie gebruikt worden (cfr. Homals, Latente Klassenanalyse).

Als auteurs zijn we verschillende mensen erkentelijk voor hun bijdrage aan dit boek. Wij willen dan ook heel uitdrukkelijk Inge Pasteels en Karel Van den Bosch bedanken voor het nalezen en becommentariëren van eerdere versies van dit boek. Ook zijn we Koen Pepermans dankbaar voor zijn vroegere cursusversie van het deel over betrouwbaarheidsanalyse. Het eindresultaat blijft uiteraard de volledige verantwoordelijkheid van de auteurs.

We hopen dat u als lezer vooral *praktisch plezier* beleeft aan deze reeks. Het soort plezier dat ontstaat wanneer je voor de eerste keer een significantietoets meester wordt of een principale factoranalyse geschat hebt.

DIMITRI MORTELMANS  
en BRITT DEHERTOGH

# Vorbereiding: De European Social Survey

Voor alle voorbeelden in dit volume wordt gebruik gemaakt van de European Social Survey. Deze keuze is gemaakt om twee redenen. In de eerste plaats bevat de ESS gegevens die onderzoekers uit diverse disciplines kunnen interesseren. Bovendien biedt het internationale karakter van het onderzoek ook veel mogelijkheden om internationaal vergelijkende voorbeelden te geven. Een tweede reden waarom voor de ESS geopteerd werd, is het vrij beschikbare karakter van de gegevens. Iedere onderzoeker kan gratis de data van de ESS gebruiken voor het eigen onderzoek. Dat betekent dat iedereen de voorbeelden uit deze reeks zelf thuis of op het werk kan reproduceren op reële gegevens en bovendien met dezelfde resultaten tot gevolg.

In wat volgt leggen we uit waar de ESS te vinden is en hoe de gegevens die in dit boek gebruikt worden, gedownload kunnen worden.

### 1.1 Toegang tot de ESS

De ESS is terug te vinden op de Noorse website van het ESS-netwerk. Open bijgevolg een webbrowser en surf naar volgend adres:

<http://www.europeansocialsurvey.org/>

Om de ESS te gebruiken, vraagt men om je eerst te registreren. Na de gratis registratie kan je op de site de gegevens rechtstreeks downloaden.

Om je te registreren, ga je naar het data-archief. Kies in het menu op de hoofdpagina voor “**Archive & Data**”. Vaak staat de link naar het archief ook apart op de hoofdpagina:

<http://ess.nsd.uib.no/>

Je kan de registratieprocedure starten door links in het menu te kiezen voor “Register New User”. Vul je gegevens in en druk op *Register*.

Home ESS Round 1 ESS Round 2 ESS Cumulative

Online browsing and analysis

ESS EduNet

- ▶ Guide to Online Browsing
- ▶ ESS Main Site
- ▶ ESS Conditions of Use
- ▶ Contact ESS Data Archive
- ▶ Register New User

**User registration**

The boxes marked with red are mandatory. If you are a private individual, please enter Private in the institution box.

**Title:**

**Surname:**

**Firstname:**

**Mail:**

**Institution:**

**Department:**

**Country:**

**Type of activity:**


2005 copyright © Norwegian Social Science Data Services [essdata@nsd.uib.no](mailto:essdata@nsd.uib.no)

Onmiddellijk na het registreren, krijg je de boodschap dat je als nieuwe gebruiker opgenomen bent. Nu kan je de website binnengaan en de data downloaden.

## 1.2 De ESS Downloaden


Om in te loggen, typ je vanaf het registratiescherm je emailadres in het log-in-vak (rechtsboven) in. Klik op **Log In**. Het vak verdwijnt en de boodschap verschijnt dat je ingelogd bent.

ESS is een zeer goed gedocumenteerd project. Onder het submenu “*Survey documentation*” en “*Fieldwork documents*” kan je alle methodologische documenten over het project bekijken en downloaden. Als je zelf met de ESS aan de slag wil gaan, bekijk dan zeker eens de documentatie over de weegvariabelen en de vragenlijst. We zijn nu geïnteresseerd om de geïntegreerde data van de eerste golf van de ESS te downloaden. Die zijn te vinden onder het submenu “**Data download**” bij het tabblad “**ESS Round 1**”.

ESS biedt de data aan in twee formaten: SAS en SPSS. Indien je het hele boek zal doornemen, heb je beide formaten nodig. We starten met het databestand in SAS-formaat. Klik<sup>1</sup> daarom op het SAS symbool  achter “ESS1 – inte-

1. ESS brengt op regelmatige basis nieuwe versies uit van de data. Het is dus mogelijk dat er ondertussen een latere versie beschikbaar is op de website. Deze nieuwe versies geven doorgaans slechts minimale verschillen met de in dit boek gebruikte versie.



grated file, edition 6.0". Vervolgens doe je hetzelfde met het SPSS-formaat (icoon: ).

Als de download succesvol was, staan er twee gecomprimeerde databestanden op je harde schijf met de naam: ESS1e06.sas.zip en ESS1e06.spss.zip.

Dit bestand kan je decomprimeren met Winzip of (als je Windows XP of Windows Vista bezit) openen in Windows Verkenner.

Eens dat gedaan is, krijg je volgende lijst van bestanden:

1. De ESS-gebruiksovereenkomst:

ESSConditionsOfUse.txt

2. Het SPSS Bestand:

ESS1e06.por

3. De SAS Bestanden:

ESS1.sas

ESS1e06sas.por

ESS1formats.sas

ESS1miss.sas

ESS1standards.sas

Als je deze bestanden op je harde schijf hebt, ben je klaar om het boek aan te vatten. Voor de voorbeelden in het boek, gaan we er van uit dat deze bestanden allemaal in de directory C:\ESS staan. Uiteraard kun je de bestanden in andere directories plaatsen, maar dan moet je het pad steeds aanpassen aan je eigen situatie.

### 1.3 De ESS gebruiksklaar maken

Voor je in dit boek aan de slag kan met de ESS-bestanden, vereisen deze enige bewerking (zie ook Volume 1 Kennismaken met SPSS en SAS) van deze reeks voor een meer uitgebreide toelichting). Voor **SPSS** moet je enkel ESS1e06.por importeren en opslaan als een .sav bestand. Voor **SAS** zijn meerdere bewerkingen op de standaardbestanden van de ESS nodig. We geven een kort overzicht zodat je de bestanden kan importeren in SAS en gebruiksklaar kan maken voor de voorbeelden in dit boek. We geven de wijzigingen ten opzichte van de ESS-standaardbestanden aan in vet.

## 1. Wijzigingen aan ESS1.sas

```
LIBNAME TRANS SPSS "C:\ESS\ESS1e06sas.POR";
LIBNAME ESS "C:\ESS";

/*To create formats (value labels): edit and run ESSformats program*/
%include "C:\ESS\ESS1formats.sas";
run;

data ESS.ESS1e06;
set TRANS._first_;

*run;

/*To create user-defined missing values .a, .b, .c and .d*/
%include "C:\ESS\ESS1miss.sas";
```

## 2. Wijzigingen aan ESS1formats.sas

```
/*Permanent formats*/
LIBNAME LIBRARY "C:\ESS";
PROC FORMAT LIBRARY = LIBRARY;

/*Temporary formats*/
*PROC FORMAT;

%include "C:\ESS\ESS1standards.sas";
```

Tot slot starten we elke sessie van SAS volgende autoexec.sas op:

```
options nocenter nodate linesize = 100 pagesize = 30000;
title;

%INCLUDE "C:\ESS\ESS1.sas";

libname ess "C:\ESS";

data ess1e06;
set ess.ess1e06;
run;
```

In deze autoexec.sas wordt een extra library *ESS* aangemaakt die verwijst naar de directory "C:\ESS". Vermits de ESS-bestanden in deze directory geplaatst werden, kan je ze aanspreken door naar de ESS-library te verwijzen. Tot slot wordt een tijdelijk bestand *ess1e06* aangemaakt op basis van het permanente databestand. Dat is aangewezen zodat je de originele bestanden nooit kwijtraakt, mocht je iets verkeerd doen. Je kunt dan steeds de autoexec.sas opnieuw uitvoeren om een nieuwe kopie van het databestand aan te maken.

## HOOFDSTUK 2

# Inleiding tot schaaltechnieken

### 2.1 Wat is een schaal?

Schalen worden opgesteld om theoretische concepten zoals arbeidstevredenheid of racisme te meten. Niet alle begrippen zijn immers rechtstreeks meetbaar bij respondenten. Veel begrippen zijn te abstract of te ruim om met één enkele vraag volledig omvat te worden. Vragen we aan respondenten hoe racistisch ze zichzelf vinden, dan duiken er verschillende problemen op. Wat verstaan respondenten onder racisme? Racisme kan zoveel vormen en betekenissen aannemen dat de antwoorden van respondenten niet vergelijkbaar zouden zijn. Omdat het een gevoelig thema kan zijn, is de kans op sociaal wenselijke antwoorden bij één enkele vraag naar racisme bovendien groot. Daarom gebruikt men schalen om de mening over een bepaald abstract concept te meten.

Om het complexe begrip te meten worden eerst **indicatoren** geformuleerd die betrekking hebben op houdingen ten aanzien van het begrip. Dit noemt men de fase van **operationalisering**. Men stelt een reeks uitspraken op waar de respondent op moet antwoorden. Wat betreft het *aantal* indicatoren zijn er geen vaste richtlijnen. We nemen aan dat er minstens 3 verwante uitspraken moeten geformuleerd zijn om de inhoud van een latent concept te kunnen dekken.

*Voorafgaand* aan de operationalisering (het meetbaar maken) vindt uiteraard eerst een theoretische fase van definiëring en afbakening van het concept plaats. Dat is de fase die omschreven wordt als de **conceptualisering**. Deze zal in hoofdzaak gestoeld zijn op bestaande kennis (uit wetenschappelijke literatuur). In dit volume van de reeks Stap in Statistiek zullen de fase van conceptualisering en operationalisering niet behandeld worden. Hoe begrippen geconceptualiseerd en geoperationaliseerd worden en op welke manier schalen aangemaakt worden of welke types er bestaan (o.a. Likert, Guttman, Mokken) kan de lezer in meer algemene sociaal wetenschappelijke werken

raadplegen. Dit boek gaat er van uit dat dat werk achter de rug is en dat de lezer indicatoren (items) in zijn database heeft waarmee hij aan de slag kan.

Het resultaat van de operationalisering is dus een schaal die bestaat uit meerdere indicatoren over hetzelfde concept. Deze schaal heeft als kenmerk dat er een bereik van antwoordmogelijkheden wordt aangeboden. De respondenten krijgen in een survey verschillende stellingen (die we de **items** (of indicatoren) van de schaal noemen) voorgelegd die ze op een puntenschaal beantwoorden. In ESS worden bijvoorbeeld verschillende vragen gesteld over immigranten, met telkens een antwoordschaal van volledig eens tot volledig oneens. Men vraagt de respondenten om op alle afzonderlijke items te antwoorden. De 7 items hebben betrekking op de houding ten aanzien van immigranten en zijn na elkaar gesteld in de vragenlijst (in het ESS-bestand gaat het over volgende variabelen: Imwgdnw, Imhecop, Imfljob, Immunplv, Imsmrgt, Imscrlv, Imacrlv).

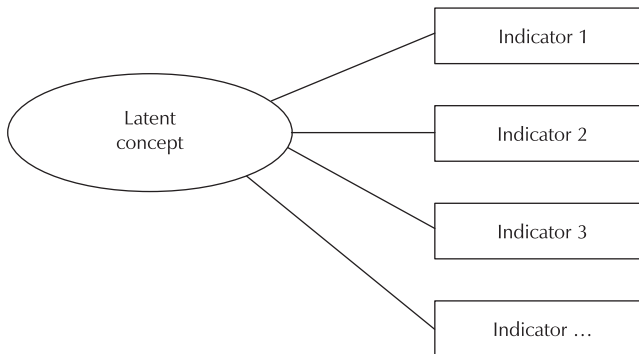
*Wilt u aangeven in hoeverre u het eens (+) of oneens (-) bent met elk van de volgende uitspraken?*

	++	+	±	-	--	Weet niet
Gemiddelde wedden en lonen worden over het algemeen lager doordat mensen hier komen wonen en werken	1	2	3	4	5	8
Mensen die hier komen wonen en werken benadelen in het algemeen de economische vooruitzichten van de armen meer dan die van de rijken	1	2	3	4	5	8
Mensen die hier komen wonen en werken, helpen jobs in te vullen waar er een tekort is aan arbeidskrachten	1	2	3	4	5	8
Als mensen die hier zijn komen wonen en werken langdurig werkloos zijn, zouden zij gedwongen moeten worden te vertrekken	1	2	3	4	5	8
Mensen die hier zijn komen wonen zouden dezelfde rechten moeten krijgen als iedereen	1	2	3	4	5	8
Als mensen die hier zijn komen wonen een ernstige misdaad begaan, moeten zij gedwongen worden te vertrekken	1	2	3	4	5	8
Als mensen die hier zijn komen wonen een, om het even welke misdaad begaan, moeten zij gedwongen worden te vertrekken	1	2	3	4	5	8

Een schaal zou je dus kunnen omschrijven als een 'meetlat' die je naast een bepaald concept van de sociale werkelijkheid legt om dat concept te meten.

Het is een techniek waarbij een onderzoeker op basis van een aantal items (indicatoren) een abstract begrip zoals een houding, een gevoel, een attitude en dergelijke tracht te kwantificeren. Deze kwantificering zien we bij bovenstaande vraag in de scores die telkens aan de antwoordmogelijkheden gekoppeld zijn. Zo krijgt iemand die helemaal eens heeft geantwoord de laagste score (1). Respondenten met een lage score in het databestand kunnen zich perfect vinden in de uitspraak. Een score van 5 betekent dat ze het er helemaal niet mee eens zijn.

Wat we meten in de vragenlijst noemen we de **manifeste** variabelen (of indicatoren). Het abstracte concept dat niet direct waarneembaar is, wordt aangeduid als de **latente** variabele. In de statistiek is er een afspraak om steeds een duidelijk grafisch onderscheid te maken tussen manifeste en latente variabelen. De manifeste variabelen (de indicatoren) worden getekend in rechthoeken, de latente in cirkels:



Figuur 1. **Model van latent concept en manifeste indicatoren.**

Het uiteindelijke doel van de technieken die in dit boek behandeld worden, is op basis van een reeks indicatoren uitspraken te kunnen doen over het complexe achterliggende concept. Door de operationalisering wordt het concept ontrafeld in meetbare indicatoren, maar dan weet je als onderzoeker nog steeds niets over de houding van de respondent ten aanzien van het concept als geheel. In de dataset zitten alleen de antwoorden van de respondenten op de afzonderlijke items. Om die te kunnen interpreteren en omvormen tot één maat voor het concept, worden achtereenvolgens twee stappen uitgevoerd. In een eerste stap wordt gekeken of de items wel degelijk één schaal vormen, of meerdere (sub)schalen. Dat is de controle op de **dimensionaliteit**. In een tweede stap wordt naar de interne samenhang gekeken. Dat is **betrouwbaarheidsanalyse** en die bepaalt of je de items in een schaal mag samenvoegen tot één maat. Dat wordt verder uitgelegd in 12.3.

Samengevat krijgen we dan volgend schema:

Datavoorbereiding	Conceptualisering <i>(definiëren en afbakenen van het concept)</i> Operationalisering <i>(het concept ‘meetbaar’ maken door er indicatoren voor op te stellen)</i>
Dataverzameling	Indicatoren voorleggen
Data-analyse	Testen op dimensionaliteit (zie 2.2) Testen op betrouwbaarheid (zie 2.3)  Constructie van de schaal Gebruik van de schaal in verdere analyses

## 2.2 Dimensionaliteit van een schaal testen

De dimensionaliteit van een schaal heeft betrekking op het aantal latente concepten dat aan de basis ligt van de schaal. In theorie vertrekken we van één concept dat wordt geoperationaliseerd aan de hand van een aantal items. Eens we de gegevens hebben verzameld, is het echter maar de vraag of de items die we hebben voorgelegd, allemaal het beoogde concept hebben gemeten. Het kan immers zijn dat wij als onderzoeker een concept willen meten maar dat dat concept niet op die manier in het hoofd van de respondenten zit. Soms gebeurt het dat mensen nog een fijner onderscheid maken dan wij op voorhand in onze conceptualisering en operationalisering hadden voorzien. Of dat het geval is, wordt nagegaan door de dimensionaliteit van een schaal te testen.

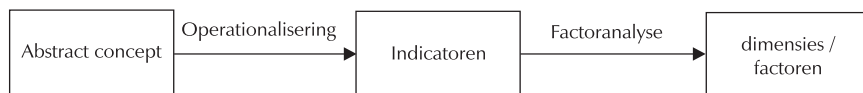
Een voorname manier om dit te testen is door mogelijke ‘samenhang’ tussen de antwoorden op de items te berekenen. Wanneer alle items samenhangen spreken we van een unidimensionele schaal. Er zit één concept (één dimensie) achter de indicatoren vermits de respondenten op gelijklopende wijze hebben geantwoord. Multidimensionele schalen worden gekenmerkt door meerdere (sub)dimensies of (deel)concepten die aan te treffen zijn in de antwoordpatronen.

Een van de voornaamste families van technieken om de dimensionaliteit van een schaal vast te stellen is **factoranalyse**. Dat is een multivariate statistische methode die orde en structuur in data wil krijgen. Het doel is het zo goed mogelijk beschrijven van een groot aantal variabelen (items) door middel

van een kleiner aantal ‘hypothetische’ variabelen (concepten). Dat wil zeggen dat deze techniek probeert om van een groot aantal manifeste variabelen een beperkter aantal latente dimensies te berekenen door de samenhang in antwoordpatronen te bekijken. Daarom wordt factoranalyse ook wel eens een **datareducerende techniek** genoemd. Bij die berekening kan de onderzoeker ook afleiden of zijn oorspronkelijk concept in één schaal (dimensie) gemeten wordt of eerder in verschillende schalen.

Statistisch gaat factoranalyse na of er in een groep variabelen hoge correlaties zijn tussen een subgroep van variabelen en lage correlaties met de andere variabelen van de groep. Factoranalyse probeert deze subgroep in een factor samen te brengen. Die factor wordt verantwoordelijk geacht voor het voorkomen van de hoge correlaties in die subgroep. Het achterliggende idee is dat het theoretische concept (de latente variabele) verantwoordelijk is voor het feit dat de variabelen zo sterk samenhangen. De factor probeert dan de samenhang tussen de variabelen te “verklaren” op een regressieachtige wijze. Als het inderdaad zo is dat een achterliggende theoretische factor verantwoordelijk is voor de samenhang tussen enkele variabelen, dan moet de factoranalyse in staat zijn om deze gemeenschappelijke samenhang te extraheren uit die groep variabelen. De centrale vraag die een onderzoeker bijgevolg in factoranalyse probeert te beantwoorden is: *Welk theoretisch concept is verantwoordelijk voor de correlaties die ik in mijn data vind?* In de praktijk gaat de onderzoeker echter omgekeerd te werk. Hij start zijn onderzoek immers met een theoretisch kader en een theoretisch concept. Van daaruit gaat hij indicatoren opstellen waarvan hij verwacht dat ze samen beïnvloed worden door hetzelfde theoretische concept. In de factoranalyse gaat hij vervolgens na of de indicatoren samen inderdaad beïnvloed worden door het vooropgestelde concept.

Wanneer we vertrekken van een theoretisch concept dat we willen meten zal de werkwijze er als volgt uitzien:



Stel dat een onderzoeker over een databestand beschikt met gegevens over de tevredenheid met collega’s, baas, werkuren, promotiekansen, loon enzovoort. In theorie is het abstracte begrip “algemene arbeidstevredenheid”. Door alle mogelijke aspecten van arbeid te bevragen, hoopt de onderzoeker om een algemene maat voor arbeidstevredenheid te kunnen opstellen.

Nadat de gegevens worden verzameld bij de respondenten, kan met een factoranalyse aangetoond worden of de indicatoren gegroepeerd kunnen worden

in één factor (unidimensioneel) of meerdere factoren (multidimensioneel). De onderzoeker gaat ervan uit dat er slechts één dimensie is, maar de factoranalyse zou kunnen aantonen dat de schaal multidimensioneel is. In plaats van één globaal concept van arbeidstevredenheid, moet de onderzoeker dan deelaspecten van arbeidstevredenheid onderscheiden: bijvoorbeeld intrinsieke en extrinsieke arbeidstevredenheid.

### 2.3 **Betrouwbaarheid van een schaal testen: betrouwbaarheidsanalyse**

Een van de eisen die gesteld worden aan een goede schaal is dat ze betrouwbaar is. Dat wil zeggen dat de schaal hetzelfde of quasi hetzelfde resultaat oplevert wanneer we ze opnieuw (en opnieuw en opnieuw) afnemen onder dezelfde omstandigheden. De waarde op de eerste meting zal met andere woorden opnieuw gevonden worden als de meting opnieuw wordt uitgevoerd. Volledige betrouwbaarheid zal je zelden of nooit aantreffen. Er zullen altijd afwijkingen zijn in de antwoordpatronen. In de sociale realiteit is immers zelden een bestudeerd fenomeen ‘perfect’ betrouwbaar. We zijn tevreden als er op elkaar gelijkende waarden worden aangetroffen.

Een andere manier om betrouwbaarheid te omschrijven is door te stellen dat de meting niet verstoord is door ‘toevallige fouten’ (random error). **Toevallige fouten** zijn fouten die optreden omdat een respondent verstrooid is of omdat onze vragen niet duidelijk genoeg gesteld zijn zodat een respondent de ene keer zus en de andere keer zo antwoordt. Om de betrouwbaarheid van een schaal te controleren zouden we idealiter dezelfde groep mensen twee keer na elkaar moeten kunnen bevragen. We kunnen dan kijken of de antwoorden op de tweede meting gelijk zijn aan de antwoorden ervoor. In de praktijk is dit een onbegonnen scenario. Mensen zijn immers vaak niet bereid om twee keer dezelfde vragen te beantwoorden en onderzoeksbudgetten zijn zelden zo groot om dit ook maar te overwegen. Daarom zijn er andere statistische technieken ontwikkeld om betrouwbaarheid te testen. Verder in dit boek (zie 9) zullen we bijvoorbeeld de Cronbach’s alfamaat behandelen die speciaal ontwikkeld is voor dit doel.