

Introductie	9
Vorbereiding: DE EUROPEAN SOCIAL SURVEY	11
1.1 Toegang tot de ESS	11
1.2 De ESS downloaden	12
1.3 De ESS gebruiksklaar maken	13
Deel 1: DATABEHEER IN SPSS	15
2.1 Inleiding	15
2.2 Werken met bestanden	16
2.2.1 <i>Bestanden samenvoegen</i>	16
2.2.1.1 <i>Het concateneren van bestanden (Add Cases)</i>	17
2.2.1.2 <i>Het mergen van databestanden (Add Variables)</i>	19
2.2.2 <i>Bestanden aggregeren</i>	20
2.2.3 <i>Bestand opsplitsen</i>	25
2.3 Werken met variabelen	29
2.3.1 <i>Berekeningen op variabelen uitvoeren</i>	30
2.3.1.1 <i>Eenvoudige berekeningen: constanten en operatoren</i>	30
2.3.1.2 <i>Functies gebruiken in berekeningen</i>	33
2.3.1.3 <i>Toepassing 1: Het omkeren van een variabele</i>	36
2.3.1.4 <i>Toepassing 2: Variabelen dichotomiseren</i>	38
2.3.2 <i>Het voorkomen van waarden tellen (COUNT)</i>	40
2.3.3 <i>Voorwaardelijk databeheer</i>	43
2.3.3.1 <i>Voorwaarden stellen met DO IF</i>	43
2.3.3.2 <i>Voorwaarden stellen met IF</i>	45
2.3.4 <i>Overbodige code vermijden: lussen</i>	46

2.3.4.1	De lus DO REPEAT	46
2.3.4.2	Vectoren aanmaken	48
2.3.4.3	De lus LOOP	49
2.3.4.4	Alternatieve manieren om het aantal loops te bepalen	52
2.3.5	<i>Categoriale variabelen hercoderen (RECODE)</i>	53
2.3.5.1	Herocoderen met syntax	54
2.3.5.2	Herocoderen met de menustructuur	57
2.3.6	<i>Continue variabelen hercoderen</i>	60
2.3.6.1	Verdeling volgens gelijke klassenbreedte	61
2.3.6.2	Verdeling volgens de nested-means-methode	64
2.3.6.3	Klassenverdeling op theoretische basis	67
2.3.6.4	Fouten bij klassenverdelingen	68
2.3.7	<i>Tekstvariabelen hercoderen</i>	69
2.3.7.1	Methode 1: Tekstvariabelen omzetten naar numerieke variabelen	69
2.3.7.2	Methode 2: tekstvariabelen herwerken met behulp van Excel	71
2.4	Werken met cases	77
2.4.1	<i>Data sorteren</i>	77
2.4.2	<i>Cases selecteren</i>	79
2.4.2.1	Met menustructuur	80
2.4.2.2	Met syntax	83
2.4.3	<i>Cases wegen</i>	84
2.5	Toepassing van databeheer 1: datacleaning	85
2.5.1	<i>Cleaning op zicht</i>	86
2.5.2	<i>Inhoudelijke cleaning</i>	87
2.5.2.1	STAP 1a: De fouten selecteren (FILTER)	87
2.5.2.2	STAP 1b: De fouten zichtbaar maken (SUMMARIZE)	88
2.5.2.3	STAP 2: Fout verbeteren (DO REPEAT)	90
2.6	Toepassing van databeheer 2: werken met een huishouden-grid	91
2.6.1	<i>STAP 1: Voorbereidingen treffen</i>	92
2.6.2	<i>STAP 2: Maak vectoren aan</i>	94
2.6.3	<i>STAP 3: Selecteer de kinderen en hun geboortjaar uit de grid</i>	94
2.6.4	<i>STAP 4: Maak de leeftijd van het oudste en het tweede kind aan</i>	96
2.6.5	<i>STAP 5: Maak de duurvariabele aan</i>	97
2.7	Toepassing van databeheer 3: een persoon-periode bestand aanmaken	97
6 2.7.1	<i>Voorstelling van het persoon-periodebestand</i>	98

2.7.2	<i>Constructie van het bestand</i>	100
2.7.2.1	STAP 1: Voorbereidingen treffen	100
2.7.2.2	STAP 2: Maak het persoon-periodebestand aan	102
 Deel 2: DATABEHEER IN SAS		107
3.1	Inleiding	107
3.2	Werken met bestanden	107
3.2.1	<i>Nieuwe databestanden maken op basis van bestaande databestanden</i>	107
3.2.2	<i>Bestanden samenvoegen</i>	109
3.2.2.1	Het concateneren van bestanden	109
3.2.2.2	Het mergen van bestanden	109
3.2.3	<i>Bestand opsplitsen</i>	111
3.2.3.1	Het BY-bevel	111
3.2.3.2	Het CLASS-bevel	112
3.2.4	<i>Bestanden aggregeren</i>	115
3.3	Werken met variabelen	121
3.3.1	Berekeningen op variabelen uitvoeren	121
3.3.1.1	Eenvoudige berekeningen: constanten en operatoren	122
3.3.1.2	Functies gebruiken in berekeningen	125
3.3.1.3	Toepassing 1: Een variabele omkeren	128
3.3.1.4	Toepassing 2: Variabelen dichotomiseren	129
3.3.2	<i>Het voorkomen van waarden tellen</i>	131
3.3.3	<i>Voorwaardelijk databeheer</i>	133
3.3.3.1	Voorwaarden stellen met een SELECT-groep	133
3.3.3.2	Voorwaarden stellen met IF-THEN	134
3.3.4	<i>Overbodige code vermijden: lussen</i>	136
3.3.5	<i>Categoriale variabelen hercoderen</i>	138
3.3.5.1	Herocoderen met FORMATS	138
3.3.5.2	Herocoderen met voorwaarden	139
3.3.6	<i>Continue variabelen hercoderen</i>	141
3.3.6.1	Verdeling volgens gelijke klassenbreedte	142
3.3.6.2	Verdeling volgens de nested-means-methode	144
3.3.7	<i>Tekstvariabelen hercoderen</i>	146
3.3.7.1	Methode 1: Tekstvariabelen omzetten naar numerieke variabelen	146
3.3.7.2	Methode 2: Herwerken met behulp van Excel	148
3.4	Werken met cases	151
3.4.1	<i>Data sorteren</i>	151
3.4.2	<i>Cases selecteren</i>	153

3.4.3	<i>Cases wegen</i>	154
3.5	Toepassing van databeheer 1: datacleaning	155
3.5.1	<i>Cleaning op zicht</i>	155
3.5.2	<i>Inhoudelijke cleaning</i>	156
3.5.2.1	STAP 1: De fouten selecteren en zichtbaar maken	157
3.5.2.2	STAP 2: Fout verbeteren	158
3.6	Toepassing van databeheer 2: Werken met een huishoudengrid	159
3.6.1	<i>STAP 1: Voorbereidingen treffen</i>	159
3.6.2	<i>STAP 2: Maak vectoren aan</i>	161
3.6.3	<i>STAP 3: Selecteer de kinderen en hun geboortjaar uit de grid</i>	162
3.6.4	<i>STAP 4: Maak de leeftijd van het oudste en het tweede kind aan</i>	162
3.6.5	<i>STAP 5: Maak de duurvariabele aan</i>	163
3.7	Toepassing van databeheer 3: Een persoon-periodebestand aanmaken	163
3.7.1	<i>STAP 1: Voorbereidingen treffen</i>	164
3.7.2	<i>STAP 2: Maak het persoon-periodebestand aan</i>	166

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 sociaal-wetenschappelijk onderzoek links liggen omdat de gevorderde handboeken té wiskundig of té ontoegankelijk geschreven zijn. Binnen de faculteit Politieke en Sociale Wetenschappen was de Methodengroep jaren lang 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 en de achtergronden, en wie de uitgebreide toepassingen ervan wil leren kennen, wordt doorverwezen naar handboeken statistiek. De reeks *Stap In Statistiek en Onderzoek* wil bewust geen sta-

tistiekreeks 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.

De keuze voor een bepaald softwarepakket in deze reeks was cruciaal. Ook al zijn er heel gespecialiseerde statistische pakketten op de markt, wij richten ons op de grootste gemene deler. In Vlaanderen en Nederland lijkt de academische onderzoeksmarkt in belangrijke mate gedomineerd te worden door twee grote spelers: SPSS en SAS. Om die reden wordt in de reeks geopteerd om alle technieken uit te leggen met behulp van deze twee pakketten. Daarbij wordt meer aandacht besteed aan SPSS, omdat het gewicht van dit pakket in het dagelijkse gebruik van de onderzoeker naar ons aanvoelen groter is. Toch is SAS op een aantal vlakken krachtiger dan SPSS, zodat wie bepaalde gevorderde analyses wil draaien, in een later stadium geregeld op SAS zal terugvallen. De reeks verbindt zich niet exclusief aan beide softwarepakketten. Wanneer de noodzaak zich opdringt, zullen ook andere pakketten in de reeks aan bod komen. SPSS en SAS zijn echter de hoekstenen van de reeks. Om die reden vangt ze aan met twee volumes die de gebruiker inleiden in SPSS en SAS. Zo wordt een basis gelegd die de onderzoeker nodig heeft om aan de slag te kunnen met de eigenlijke statistische analyses.

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 fittend logistisch regressiemodel geschat hebt.

DIMITRI MORTELMANS
en BRITT DEHERTOGH

Vorbereiding: De European Social Survey

Voor alle voorbeelden in dit volume wordt gebruikgemaakt 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 vrij 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@nsg.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 login-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 daarom op het SAS-symbool  achter “ESS1 – inte-

grated file, edition 5.1". 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:

ESS1e05_1[1].sas.zip en ESS1e05_1[1].spss.zip.

Dit bestand kan je decomprimeren met Winzip of (als je Windows XP 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:

ESS1e05_1.por

3. De SAS-Bestanden

ESS1.sas

ESS1e05_1sas.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 ervan uit dat deze bestanden allemaal in de directory C:\ESS staan. Uiteraard kan 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 van de reeks voor een meer uitgebreide toelichting). Voor **SPSS** moet je enkel ESS2e01.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\ESS1e05_1sas.POR";
LIBNAME ESS "C:\ESS";

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

data ESS.ESS1e05_1;
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 met volgende autoexec.sas op:

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

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

libname ess "C:\ESS";

data ess1e05_1;
set ess.ess1e05_1;
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 *ess1e05_1* aangemaakt op basis van het permanente databestand. Dat is aangewezen zodat je de originele bestanden nooit kwijtraakt, mocht je iets verkeerd doen. Je kan dan steeds de autoexec.sas opnieuw uitvoeren om een nieuwe kopie van het databestand aan te maken.