

Inhoud

Introductie	9
Hoofdstuk 1	
VOORBEREIDING: 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
Hoofdstuk 2	
THEORETISCHE ACHTERGROND BIJ REGRESSIE	15
2.1 Inleiding	15
2.2 De regressievergelijking	17
2.3 Regressiemodellen beoordelen (R^2)	21
2.4 Assumpties van regressierekenen	25
2.5 De acht stappen van regressieanalyse	31
2.6 Waarschuwing voor de lezer	34
Hoofdstuk 3	
ANALYSE 1. EEN INTERVAL-ONAFHANKELIJKE	35
3.1 Inleiding	35
3.2 Syntax	35
3.3 Output 1: Voorbereidende analyses	36
3.4 Output 2: Regressieanalyse	41
3.5 Deze analyse in SAS	44
Hoofdstuk 4	
ANALYSE 2. MEERVOUDIGE REGRESSIE (MEERDERE ONAFHANKELIJKEN)	51
4.1 Inleiding	53 5

4.2	Syntax	54
4.3	Output	54
4.4	Deze analyse in SAS	61

Hoofdstuk 5

ANALYSE 3. CATEGORIALE VARIABELEN OPNEMEN	63
--	----

5.1	Inleiding	63
5.2	Syntax	70
5.3	Output	72
5.4	Deze analyse in SAS	76

Hoofdstuk 6

ANALYSE 4. ONAFHANKELIJKEN HERSCHALEN	81
--	----

6.1	Inleiding	81
6.2	Syntax	84
6.3	Output	86
6.4	Deze analyse in SAS	89

Hoofdstuk 7

ANALYSE 5. STAPSGEWIJZE (HIËRARCHISCHE) REGRESSIES UITVOEREN	93
---	----

7.1	Inleiding	93
7.2	Syntax	94
7.3	Output	98
7.4	Deze analyse in SAS	104

Hoofdstuk 8

ANALYSE 6. INTERACTIE-EFFECTEN INVOEREN	109
--	-----

8.1	Inleiding	109
8.2	Syntax	110
8.3	Output	111
8.4	Deze analyse in SAS	119

Hoofdstuk 9

ASSUMPTIES TESTEN 1A. DE LINEARITEITSVEREISTE (NIET-LINEAIRE REGRESSIE 1)	123
--	-----

9.1	De schending vaststellen	124
9.2	De schending verhelpen (Deel 1)	129
9.3	Deze assumptie testen en verhelpen in SAS	139

Hoofdstuk 10	
ASSUMPTIES TESTEN 1B. DE LINEARITEITSVEREISTE (NIET-LINEAIRE REGRESSIE 2)	145
10.1 De schending verhelpen (Deel 2)	145
10.2 Box-cox transformaties in SAS	153
Hoofdstuk 11	
ASSUMPTIES TESTEN 1C. DE LINEARITEITSVEREISTE (NIET-LINEAIRE REGRESSIE 3)	159
11.1 De schending verhelpen (Deel 3)	159
11.2 Spline regressie in SAS	169
Hoofdstuk 12	
ASSUMPTIES TESTEN 2. DE NORMALITEITSVEREISTE	173
12.1 De schending vaststellen	173
12.2 De schending verhelpen	179
12.3 Deze assumptie testen en verhelpen in SAS	184
Hoofdstuk 13	
ASSUMPTIES TESTEN 3. DE ONAFHANKELIJKHEIDSVEREISTE	189
13.1 De schending vaststellen	189
13.2 Deze assumptie testen in SAS	191
Hoofdstuk 14	
ASSUMPTIES TESTEN 4. DE HOMOSCEDASTICITEITSVEREISTE	193
14.1 De schending vaststellen	193
14.2 De schending verhelpen	202
14.3 Deze assumptie testen en verhelpen in SAS	207
Hoofdstuk 15	
ANDERE REGRESSIEPROBLEMEN DETECTEREN 1. MULTICOLLINEARITEIT	215
15.1 Het probleem vaststellen	216
15.2 Het probleem verhelpen	217
15.3 Multicollineariteit opsporen in SAS	223
Hoofdstuk 16	
ANDERE REGRESSIEPROBLEMEN DETECTEREN 2. OUTLIERS, HEFBOOMPUNTEN EN INVLOEDRIJKE CASES	227
16.1 Het probleem vaststellen	227

16.2	Het probleem verhelpen	237
16.3	Outliers, hefboom punten en invloedrijke cases opsporen in SAS	238
Hoofdstuk 17		
STAPSGEWIJS OVERZICHT VAN REGRESSIEANALYSE		245
17.1	Fase 1. Theoretisch–Conceptuele fase	245
17.2	Fase 2. Uni- en bivariate verkenning	246
17.3	Fase 3. Eerste inhoudelijke inspectie	250
17.4	Fase 4. Controle van de assumpties van het regressiemodel	252
17.5	Fase 5. Multicollineariteitsanalyse	262
17.6	Fase 6. Residu-analyse	264
17.7	Fase 7. Tweede inhoudelijke inspectie	268
17.8	De SAS-syntax van dit voorbeeld	273
Hoofdstuk 18		
WETENSCHAPPELIJK RAPPORTEREN OVER REGRESSIEANALYSE		281
18.1	Inleiding	281
18.2	Voorbeeld van de bespreking van het voorbeeldmodel	283
Hoofdstuk 19		
REGRESSIEANALYSE IN SPSS EN SAS		287
19.1	De menu-aanpak in SPSS	287
19.2	De REGRESSION syntax (SPSS)	290
19.3	De PROC REG syntax (SAS)	293

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 staat regressieanalyse centraal. Regressie is een veelgebruikte techniek in de sociale wetenschappen die door veel onderzoekers echter op een minimalistische (en foute) wijze toegepast wordt. De basis van de techniek is immers relatief eenvoudig te begrijpen en softwarepakketten als SPSS of SAS maken het uitvoeren van een regressie tot iets van luttele seconden. Toch dient de onderzoeker regressie met de nodige omzichtigheid te behandelen. Regressie is immers niet ongevoelig aan inbreuken op de basisassumpties van de techniek. Om die reden wil dit boek regressieanalyse op een ruime manier behandelen: de basistechnieken om op een snelle wijze een regressiemodel te schatten maar ook de uitgebreide controle op de geldigheid van het geschatte regressiemodel. Dit alles wordt stap voor stap uitgelegd aan de hand van concrete voorbeelden, afkomstig uit de European Social Survey. Op die manier wordt de drempel verlaagd om ook de moeilijkere elementen uit de regressieanalyse aan te pakken.

Als auteurs zijn we verschillende mensen erkentelijk voor hun bijdrage aan dit boek. Wij willen dan ook heel uitdrukkelijk Inge Pasteels, Laurent Snoeckx, Karel van den Bosch, Jeremi Van Gorp en Kaat Vanseer bedanken voor het nalezen en becommentariëren van eerdere versies van dit boek. 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 fittend logistisch regressiemodel geschat hebt.

DIMITRI MORTELMANS
en BRITT DEHERTOGH

HOOFDSTUK 1

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 webbrowswer 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*.

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. Klik¹ 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 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 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 ESS1e05_1.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 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 kunt dan steeds de autoexec.sas opnieuw uitvoeren om een nieuwe kopie van het databestand aan te maken.