

INTRODUCTIE	11
Hoofdstuk 1	
VOORBEREIDING: DE EUROPEAN SOCIAL SURVEY	13
1.1 Toegang tot de ESS	13
1.2 De ESS downloaden	14
1.3 De ESS gebruiksklaar maken	15
Hoofdstuk 2	
INLEIDING TOT ODDS EN ODDS-RATIO'S	19
2.1 Proporties, kansen of probabiliteiten	19
2.2 Conditionele kansen	20
2.3 Odds	21
2.4 De odds-ratio	24
2.5 Eigenschappen van odds en odds-ratio's	24
2.6 Odds en odds-ratio's bij meerdere categorieën	26
2.7 Deze analyse in SAS	28
2.8 Deze analyse in STATA	29
Hoofdstuk 3	
THEORETISCHE ACHTERGROND BIJ LOGISTISCHE REGRESSIE	33
3.1 Lineaire regressie en een dichotome afhankelijke	33
3.2 Van lineaire naar logistische regressie	41
3.3 De functionele vorm van de logistische regressie	43
3.4 Assumpties van logistische regressie	44
3.5 De acht stappen van logistische regressieanalyse	49
3.6 Waarschuwing voor de lezer	51
3.7 Deze analyse in SAS	52
3.8 Deze analyse in STATA	54

Hoofdstuk 4	
ANALYSE 1: EÉN ONAFHANKELIJKE VARIABELE OP INTERVAL-NIVEAU	57
4.1 Inleiding	57
4.2 Syntax	57
4.3 Output 1: Voorbereidende analyses	58
4.4 Output 2: Specificatie van het model	60
4.5 Output 3: Controle van het model	61
4.6 Output 4: Interpretatie van de parameters	71
4.7 Deze analyse in SAS	74
4.8 Deze analyse in STATA	79
Hoofdstuk 5	
ANALYSE 2: CATEGORIALE VARIABELEN OPNEMEN	83
5.1 Inleiding	83
5.2 Het probleem van ordinale en nominale variabelen in regressie	84
5.3 Syntax	88
5.4 Output 1: Voorbereidende analyses	89
5.5 Output 2: Specificatie van het model	90
5.6 Output 3: Controle van het model	91
5.7 Output 4: Interpretatie van de parameters	93
5.8 Deze analyse in SAS	104
5.9 Deze analyse in STATA	110
Hoofdstuk 6	
ANALYSE 3: INTERACTIE-EFFECTEN INVOEREN	113
6.1 Inleiding	113
6.2 Interactie-effecten interpreteren	114
6.3 Syntax	116
6.4 Output	117
6.5 Deze analyse in SAS	127
6.6 Deze analyse in STATA	132
Hoofdstuk 7	
ANALYSE 4: STAPSGEWIJZE (HIËRARCHISCHE) LOGISTISCHE REGRESSIES UITVOEREN	135
7.1 Syntax	136
7.2 Output	136
7.3 Deze analyse in SAS	142
7.4 Deze analyse in STATA	144

Hoofdstuk 8	
DE LINEARITEITSASSUMPTIE TESTEN	145
8.1 De schending vaststellen	146
8.2 De schending verhelpen	147
8.3 Deze assumptie testen en verhelpen in SAS	152
8.4 Deze assumptie testen en verhelpen in STATA	153
Hoofdstuk 9	
ANDERE REGRESSIEPROBLEMEN DETECTEREN 1: MULTICOLLINEARITEIT	155
9.1 Het probleem vaststellen	156
9.2 Het probleem verhelpen	158
9.3 Multicollineariteit opsporen in SAS	159
9.4 Deze analyse in STATA	159
Hoofdstuk 10	
ANDERE REGRESSIEPROBLEMEN DETECTEREN 2: OUTLIERS, HEFBOOMPUNTEN EN INVLOEDRIJKE CASES	161
10.1 Het probleem vaststellen	161
10.2 Het probleem verhelpen	167
10.3 Outliers, hefboompunten en invloedrijke opsporen in SAS	169
10.4 Outliers, hefboompunten en invloedrijke opsporen in STATA	174
Hoofdstuk 11	
ANDERE REGRESSIEPROBLEMEN DETECTEREN 3: NUL-CELLEN EN SCHEIDING IN DE DATA	177
11.1 Complete scheiding in de data	178
11.2 Problemen met nul-cellen	181
11.3 Problemen met nul-cellen verhelpen	183
11.4 Nulcellen en scheiding in de data opsporen in SAS	187
11.5 Nulcellen en scheiding in de data opsporen in STATA	190
Hoofdstuk 12	
ANALYSE 5: MULTINOMIALE LOGISTISCHE REGRESSIE	193
12.1 Inleiding	193
12.2 Syntax	195
12.3 Output 1: Specificatie van het model	197
12.4 Output 2: Controle van residuen	198
12.5 Output 3: Controle van het model	199

12.6	Output 4: Interpretatie van de parameters	202
12.7	Het model opnieuw schatten om de bijkomende contrasten te identificeren	207
12.8	Corrigeren voor overdispersion	209
12.9	Deze analyse in SAS	210
12.10	Deze analyse in STATA	214

Hoofdstuk 13

ANALYSE 6: ORDINALE LOGISTISCHE REGRESSIE	217	
13.1	Inleiding	217
13.2	Syntax	221
13.3	Output 1: Specificatie van het model	221
13.4	Output 2: Controle van residuen	222
13.5	Output 3: Controle van het model	223
13.6	Output 4: Interpretatie van de parameters	225
13.7	Deze analyse in SAS	228
13.8	Deze analyse in STATA	231

Hoofdstuk 14

ANALYSE 7: PROBIT REGRESSIE	233	
14.1	Inleiding	233
14.2	Syntax	236
14.3	Output	237
14.4	Output 1: Specificatie van het model	237
14.5	Output 2: Controle van het model	238
14.6	Output 3: Interpretatie van de parameters	240
14.7	Deze analyse in SAS	246
14.8	Deze analyse in STATA	248

Hoofdstuk 15

ANALYSE 8: LOGISTISCHE REGRESSIE MET GECLUSTERDE DATA (GEE)	251	
15.1	Inleiding	251
15.2	Syntax	253
15.3	Output	254
15.4	Deze analyse in SAS	256
15.5	Deze analyse in STATA	258

Hoofdstuk 16

STAPSGEWIJS OVERZICHT VAN LOGISTISCHE REGRESSIE	261
--	-----

8 16.1	Fase 1. Theoretisch-Conceptuele fase	261
----------	--------------------------------------	-----

16.2	Fase 2. Uni- en bivariate verkenning	262
16.3	Fase 3. Eerste inhoudelijke inspectie	264
16.4	Fase 4. Controle van de lineariteitsassumptie en problemen met nulcellen	266
16.5	Fase 5. Multicollineariteitsanalyse	268
16.6	Fase 6. Residu-analyse	269
16.7	Fase 7. Tweede inhoudelijke inspectie	270
16.8	De SAS-syntax van dit voorbeeld	272
16.9	De STATA-syntax van dit voorbeeld	277
Hoofdstuk 17		
	WETENSCHAPPELIJK RAPPORTEREN OVER LOGISTISCHE REGRESSIE	281
17.1	Inleiding	281
17.2	De moeilijke vertaling van odds ratios	283
17.3	Voorbeeld van de bespreking van het voorbeeldmodel	286
Hoofdstuk 18		
	LOGISTISCHE REGRESSIE IN SPSS, SAS EN STATA	289
18.1	De menu-aanpak in SPSS	289
18.2	De menu-aanpak in STATA	291
18.3	De LOGISTIC REGRESSION syntax (SPSS)	292
18.4	De PROC LOGISTIC syntax (SAS)	294
18.5	De logistic en logit syntax (STATA)	297
	INDEX	299

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 statistiekreeks 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 logistische regressie centraal. Logistische regressie is een regressietechniek voor dichotome afhankelijke variabelen. Dit type variabele mag niet in een lineaire regressie ingevoegd worden. Wie vertrouwd is met de lineaire regressie, zal merken dat categoriale variabelen als afhankelijken in een regressiemodel een aparte uitdaging vormen. Om die reden wil dit boek logistische regressie op een ruime manier behandelen: de basistechnieken om op een snelle wijze een model te schatten maar ook de interpretatie van de parameters in termen van odds en oddsratio's. 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 moeilijker elementen uit de logistische regressie aan te pakken. Naast het basismodel gaat dit volume immers ook in op logistische regressie met nominale en ordinale variabelen, op probitregressie en op logistische regressie met geclusterde data. Deze uitbreidingen vormen doorgaans slechts een kleine wijziging aan het basismodel en bieden een ruime waaier aan alternatieven op de klassieke logistische regressie. Alle voorbeelden in dit volume worden zoals steeds uitgewerkt met SPSS. Voor onderzoekers die SAS of STATA geruiken, wordt de syntax van deze voorbeelden aangeboden zodat zij de analyses in de eigen software kunnen herhalen.

Als auteur ben ik verschillende mensen erkentelijk voor hun bijdrage aan dit boek. Ik wil dan ook heel uitdrukkelijk Inge Pasteels en Danny Rouckhout bedanken voor het nalezen en becommentariëren van eerdere versies van dit boek. Het eindresultaat blijft uiteraard mijn volledige verantwoordelijkheid.

Ik hoop dat u als lezer vooral *praktisch plezier* beleeft aan deze reeks. Het soort plezier dat ontstaat wanneer je voor de eerste keer een odds ratio meester wordt of een fittend logistisch regressiemodel geschat hebt.

DIMITRI MORTELMANS

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

The screenshot shows the 'User registration' page on the ESS EduNet website. The page has a navigation bar at the top with 'Home', 'ESS Round 1', 'ESS Round 2', and 'ESS Cumulative'. On the left, there is a sidebar with 'Online browsing and analysis' and 'ESS EduNet' logo, followed by a list of links: 'Guide to Online Browsing', 'ESS Main Site', 'ESS Conditions of Use', 'Contact ESS Data Archive', and 'Register New User'. The main content area is titled 'User registration' and contains a note: 'The boxes marked with red are mandatory. If you are a private individual, please enter Private in the institution box.' Below this, there are several input fields: 'Title' (with 'mr.' entered), 'Surname' (with 'Met De Pet' entered), 'Firstname' (with 'Jan' entered), 'Mail' (with 'janmetdepet@hotmail.com' entered), 'Institution' (with 'Antwerp University' entered), 'Department' (with 'Sociology' entered), 'Country' (with 'Belgium' selected in a dropdown), and 'Type of activity' (with 'Student' selected in a dropdown). At the bottom of the form are two buttons: 'registerer' and 'reset'. A copyright notice at the bottom of the page reads '2005 copyright © Norwegian Social Science Data Services 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 e-mailadres 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. Afhankelijk van welk programma je gebruikt, kan je een van beide downloaden. We starten met het databestand in SAS-formaat. Klik¹ daarom op het SAS symbool (📄) achter

1. ESS brengt op regelmatige basis nieuwe versies uit van de data. Het is bijgevolg 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.

“ESS1 – integrated file, edition 6.1”. 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: ESS1e06_1[1].sas.zip en ESS1e06_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:

ESS1e06_1.por

3. De SAS Bestanden:

ESS1e06_1_ms.sas

ESS1e06_1sas.por

ESS1formats.sas

ESS1e06_1_miss.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. 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. Voor **STATA** moet je de data in SPSS of in SAS exporteren en vervolgens in STATA importeren. De ESS wordt niet rechtstreeks in STATA-formaat geleverd. We geven een kort overzicht zodat je de bestanden kan importeren.

1.3.1 SPSS

De meest eenvoudige manier om het ESS-bestand te gebruiken, is via SPSS. De ESS-distributie van SPSS levert zoals gezegd een SPSS-databestand aan in een Portable-formaat. Dat betekent dat de extensie .POR is.

ESS1e06_1.por

In SPSS kan je dit bestand openen via het menu File/Open. Je kiest de directory waar het bestand staat en duidt in het drop-downmenu “Files of type” de keuze “SPSS Portable (*.por)” aan. Klik het ESS-bestand aan en kies “Open”. SPSS importeert het bestand en je kunt het vervolgens met File/Save opslaan als een echt SPSS bestand (met extensie .sav).

Als je dit in syntax wil doen, gebruik je volgend programma:

```
GET FILE = 'C:\ESS\ess1e06_1.sav'.
DATASET NAME DataSet1 WINDOW = FRONT.
SAVE OUTFILE = 'C:\ESS\ess1e06.sav'/COMPRESSED.
```

We slaan het bestand op als `ess1e06.sav`. Het achtervoegsel `_1` laten we weg zodat de bestandsnaam iets korter wordt.

1.3.2 SAS

Voor SAS wordt niet alleen een databestand geleverd maar ook een aantal hulpbestanden om de labels en formats van het bestand mee in te lezen. In totaal levert ESS vijf bestanden (vier .sas en een .por bestand) waarmee je de data in SAS kan inlezen. Om deze bestanden te laten werken op je eigen computer, zijn er evenwel enkele kleine wijzigingen noodzakelijk aan de syntax die meegeleverd wordt. We geven de wijzigingen ten opzichte van de ESS-standaardbestanden aan in vet. Daarbij gaan we er van uit dat alle bestanden gedownload en ontzipt werden in de directory `C:\ESS` op je harde schijf.

1. Wijzigingen aan `ESS1e06_1_ms.sas`

```
LIBNAME TRANS SPSS "C:\ESS\ESS1e06_1sas.POR";
LIBNAME ESS "C:\ESS";

/*To read MS country files: use following libname and replace XX with country code,
eg. DE for Germany*/
*LIBNAME TRANS SPSS "ESS1XXsas.por";

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

data ESS.ESSe06;
set TRANS._first_;
*run;

/*To create user-defined missing values .a, .b, .c and .d*/
%include "C:\ESS\ESS1e06_1_miss.sas";
```

2. Wijzigingen aan ESS1formats.sas

```

*****
*                               SAS                               *
*                               FORMATS                           *
*                               ESS DATASETS                       *
*****
/*
Comments to Formats and missing values:
* The variables in the data files deposited to NSD
* should have the missing values defined in the
* Data Protocol, i.e. (n)6, (n)7, (n)8 and (n)9.
*/
* a. Permanent formats;
LIBNAME LIBRARY "C:\ESS";
PROC FORMAT LIBRARY = LIBRARY;

```

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

```

options nocenter nodate linesize = 100 pagesize = 30000;
title;
%INCLUDE "C:\ESS\ESS1e06_1_ms.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 (we laten voor de eenvoud de *_1* weg uit de tijdelijke bestandsnaam) 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.

1.3.3 STATA

De ESS gegevens worden enkel in een Portable formaat geleverd voor SAS en SPSS. Om de ESS in STATA te kunnen gebruiken, heb je ofwel SPSS, ofwel SAS nodig om de data om te zetten. We geven hieronder de syntax weer om het bestand naar STATA-formaat om te zetten, vertrekkende vanuit SPSS of vanuit SAS. We gaan er in beide gevallen van uit dat je de instructies hierboven gevolgd hebt en dat het databestand *ess1e06* in C:\ESS staat in ofwel SPSS ofwel SAS-formaat.²

2. Bij de syntax van SAS gaan we uit van het tijdelijke bestand *ess1e06*. Als je het permanente databestand wil omzetten gebruik je *data = ess.ess1e06*.

1. Uitvoeren naar STATA vanuit SPSS

```
save translate outfile = ' C:\ESS\ess1e06.dta'.
```

2. Uitvoeren naar STATA vanuit SAS

```
proc export data = ess1e06 outfile = "C:\ESS\ess1e06.dta";  
run;
```

Om het databestand vervolgens in Stata te openen, gebruik je volgende syntax:

```
set memory 300m  
cd C:\ESS  
use ess1e06
```