

# Inhoudsopgave

<b>Voorwoord</b>	<b>v</b>
<b>Lijst van figuren</b>	<b>xiii</b>
<b>Lijst van tabellen</b>	<b>xvii</b>
<b>1 Inleiding</b>	<b>1</b>
1.1 Causaliteit	1
1.2 Causaliteitsonderzoek: het klassieke experiment	6
1.3 Waarom een experiment soms niet haalbaar is	12
1.4 Methodologische problemen van niet-experimenten	20
1.5 Organisatie van dit boek	25
1.5.1 De voorbeelddataset	27
1.5.2 SPSS en Stata	29
1.5.3 Notatie	31
<b>2 Regressie-analyse</b>	<b>33</b>
2.1 Multipele regressie	33
2.1.1 Het model van multipele regressie	33
2.1.2 Model fit	36
2.1.3 Aantal en keuze van predictoren	37
2.1.4 Toetsen	39
2.2 Assumpties	40
2.3 Regressie met een dichotome afhankelijke variabele: logistische regressie	42
2.4 Regressie met een afhankelijke <i>count</i> -variabele	49
2.4.1 Poisson regressie	51
2.4.2 Negatieve binomiaalregressie	58
2.4.3 <i>Zero-inflated</i> toevoeging aan modellen	62
2.5 Regressiemodellen met nog anders verdeelde afhankelijke variabelen	64

2.6	Nog wat statistische basics	65
2.6.1	Variantie en <i>bias</i>	66
2.6.2	Consistentie en efficiëntie van schatters	68
<b>3</b>	<b><i>Propensity score matching</i></b>	<b>71</b>
3.1	Het idee van matching	71
3.1.1	Een voorloper van <i>propensity score matching</i> : Mahalanobis <i>metric matching</i>	73
3.2	<i>Propensity score matching</i>	76
3.2.1	Overwegingen bij het gebruik van <i>propensity score matching</i>	78
3.3	Diverse matchingsmethoden	78
3.3.1	<i>Nearest neighbour matching</i>	79
3.3.2	Caliper en <i>radius matching</i>	81
3.3.3	<i>Stratification of interval matching</i>	83
3.3.4	Kernel matching	83
3.3.5	De <i>performance</i> van de verschillende matchingsmethoden	87
3.4	Controle op robuustheid en sensitiviteitsanalyse	88
3.5	Voorwaarden voor <i>propensity score matching</i>	89
3.5.1	Grote steekproeven	90
3.5.2	Een voldoende <i>area of common support</i>	90
3.5.3	Niet te veel missende waarden	91
3.5.4	De juiste <i>functional form</i> van het logistische regressiemodel	91
3.5.5	Geen (resterende) ongemeten <i>bias</i>	92
3.6	Voorbeeld: Behandeling en zedenrecidive	92
3.7	Zelf geknutselde <i>propensity score</i> analyse met SPSS	95
3.7.1	Stap 1: logistische regressie	96
3.7.2	Stap 2: matchen	99
3.7.3	Stap 3: analyse van de gemaakte groepen	103
3.8	<i>Propensity score</i> analyse met Stata: caliper matching	103
3.9	<i>Propensity score</i> analyse met Stata: kernel matching	111
3.10	Software	112
3.11	Uitbreidingen	112
3.12	Verder lezen	114
<b>4</b>	<b>Instrumentele variabelen</b>	<b>117</b>
4.1	Inleiding	117
4.2	Instrumentele variabelen	119
4.2.1	De methode van instrumentele variabelen	119
4.2.2	Als oplossing voor verstoringen van variabelen	122
4.2.3	Als oplossing voor simultane causaliteit	128

4.2.4	Als oplossing voor meetfouten in de interventievariabele	132
4.2.5	Zwakke instrumenten	137
4.2.6	Overwegingen bij het gebruik van instrumentele variabelen	138
4.3	Voorbeeld: Reclasseringstoezicht en algemene recidive	140
4.3.1	De keuze van de instrumentele variabele	142
4.3.2	Het instrumentele variabelen model	147
4.3.3	Instrumentele variabelen met SPSS	148
4.3.4	Instrumentele variabelen met Stata	151
4.4	Software	153
4.5	Voordelen en beperkingen	156
4.6	Verder lezen	157
<b>5</b>	<b>Het <i>regression discontinuity design</i></b>	<b>161</b>
5.1	Het <i>regression discontinuity design</i>	161
5.2	Illustratie van het <i>regression discontinuity design</i>	163
5.3	Voorbeeld: Behandeling en aantal aanhoudingen	168
5.3.1	Het onderliggend continuüm	171
5.3.2	<i>Regression discontinuity design</i> met SPSS	172
5.3.3	<i>Regression discontinuity design</i> met Stata	182
5.4	Sterke en zwakke punten van het <i>regression discontinuity design</i>	182
5.5	Uitbreidingen	184
<b>6</b>	<b><i>Fixed effects</i> panelmodellen</b>	<b>185</b>
6.1	Inleiding	185
6.2	Analysetechniek voor longitudinale data	185
6.3	Een naïeve analyse van paneldata	189
6.4	Het <i>fixed effects</i> panelmodel	190
6.5	Software	194
6.6	Voorbeeld: Effecten van huwelijk, ouderschap en werk op criminaliteit	198
6.6.1	Periode-effecten: <i>Two-way fixed effects</i> panelmodellen	202
6.6.2	Een <i>random effects</i> panelmodel	203
6.6.3	<i>Fixed effects</i> versus <i>random effects</i> panelmodellen	207
6.7	Beperkingen van <i>fixed effects</i> panelmodellen	209
6.8	Afsluiting	210
6.9	Verder lezen	211
<b>7</b>	<b><i>Trajectory</i> modellen</b>	<b>213</b>
7.1	<i>Trajectory</i> model	214
7.2	Het <i>trajectory</i> model nader beschouwd	217
7.2.1	Voorbeeld van <i>trajectory</i> model	219

7.3	Modelselectie	221
7.4	<i>Posterior probabilities</i> voor groepslidmaatschap	224
7.5	Onafhankelijke variabelen toevoegen aan het <i>trajectory</i> model	226
	7.5.1 Statistische predictoren	226
	7.5.2 Tijdvariërende covariaten	227
7.6	<i>Censoring, missing data</i> en <i>exposure</i>	228
7.7	Voorbeeld: Invloed van werk op criminaliteit	229
	7.7.1 Het <i>trajectory</i> model	230
	7.7.2 Het <i>trajectory</i> model met covariaten	236
7.8	Beperkingen en voordelen	239
7.9	Verder lezen	241
<b>8</b>	<b>Tot besluit</b>	<b>243</b>
<b>Bijlage A</b>	<b>Appendix: Complete SPSS syntax</b>	<b>251</b>
<b>Bijlage B</b>	<b>Appendix: Complete Stata syntax</b>	<b>259</b>
	<b>Bibliografie</b>	<b>269</b>
	<b>Register van onderwerpen</b>	<b>276</b>
	<b>Register van auteurs</b>	<b>280</b>

# Lijst van figuren

1.1	Grafische voorstelling van een direct causaal verband tussen $X$ en $Y$	2
1.2	Grafische voorstelling van een spurieus verband tussen $X$ en $Y$	3
1.3	Grafische voorstelling van een spurieus ketenverband tussen $X$ en $Y$	4
1.4	Grafische voorstelling van multipele oorzaken van een afhankelijke variabele	4
1.5	Grafische voorstelling van een reciproke causale relatie	5
1.6	Grafische voorstelling van een suppressorvariabele	5
1.7	Voorbeeld van vertekening door verschil in motivatie	24
1.8	Voorbeeld van vertekening door verschil in risico	26
2.1	Voorbeeld van regressie bij dichotome afhankelijke variabele	44
2.2	Voorbeeld van de verdeling van politiecontacten in de Nederlandse bevolking	50
2.3	Siméon-Denis Poisson (1781-1840)	53
2.4	Poissonverdeling van kansen op $k = 0, 1, 2, 3, \dots$ bloeiende bloemen voor $\lambda = 2$	54
2.5	Poissonverdelingen met $\lambda = 1, 2, 4$ en $10$	55
2.6	Ladislav von Bortkiewicz (1868-1931)	56
2.7	Negatieve binomiaalverdeling van kansen voor $N = 2, 3, 4, 5, \dots$ pogingen om 2 successen te krijgen, bij $p = 0.4$	60
2.8	Schets van <i>bias</i> en <i>noise</i> van schatters	67
3.1	Prasanta Chandra Mahalanobis (1893 – 1972)	74
3.2	Schematische voorstelling van <i>nearest neighbour matching</i>	80
3.3	Schematische voorstelling van <i>caliper matching</i> en de <i>area of common support</i>	82
3.4	Schematische voorstelling van <i>stratification matching</i>	84
3.5	Kruistabel van de variabelen <code>behandeld</code> en <code>type_dader</code>	93
3.6	Kruistabel van variabelen <code>behandeld</code> en <code>recidive_zeden</code>	95

3.7	Odds ratio van de variabelen <code>behandeld</code> en <code>recidive_zeden</code>	96
3.8	Voorspellers en hun significantie uit de logistische regressie	98
3.9	Boxplot van <i>propensity scores</i> in de behandelde en onbehandelde groep	99
3.10	SPSS-bestand met verplaatste, gesorteerde variabele <code>PRE_1</code> , en variabele <code>caliper_1</code>	100
3.11	SPSS-bestand gesorteerde <i>propensity scores</i> (donkerder kolom) en daarvoor variabele <code>caliper_1</code> met gematchten	102
3.12	Kruistabel van <code>behandeld</code> en <code>recidive_zeden</code> alleen voor de propensity gematchten	104
3.13	Odds ratio van <code>behandeld</code> en <code>recidive_zeden</code> alleen voor de propensity gematchten)	104
3.14	Resultaat van logistische regressie met Stata	106
3.15	Resultaat van <code>caliper (.05) propensity score matching</code> met Stata	108
3.16	Resultaat van <code>caliper (.01) propensity score matching</code> met Stata	108
3.17	Overzicht van matching bij <code>caliper (.05) propensity score matching</code> met Stata	109
3.18	Overzicht van matching bij <code>caliper (.01) propensity score matching</code> met Stata	109
3.19	Resultaat van kernel matching met Stata met 22% trimming	113
4.1	Philip Green Wright (1861-1934)	120
4.2	Voorbeeld van verstorende variabele bij verband tussen roken en gezondheid	123
4.3	Instrumentele variabelen model voor de relatie tussen rookgedrag en gezondheid	124
4.4	Sewall Green Wright (1889-1988)	126
4.5	Voorbeeld van verstorende variabele bij verband tussen werk en jeugdcriminaliteit	127
4.6	Instrumentele variabelen model voor de relatie tussen werk en jeugdcriminaliteit	127
4.7	Voorbeeld van simultane causaliteit bij verband tussen het aanbod en de prijs van spruitjes	129
4.8	Instrumentele variabelen model voor de relatie tussen het aanbod en de prijs van spruitjes	130
4.9	Voorbeeld van simultane causaliteit bij verband tussen de prijs van rogge en vermogenscriminaliteit	131
4.10	Instrumentele variabelen model voor het verband tussen de prijs van rogge en vermogenscriminaliteit	132

4.11	Voorbeeld van <i>endogeneity bias</i> door meetfouten bij verband tussen de culturele participatie van ouders en kinderen	133
4.12	Instrumentele variabelen model voor het verband tussen de culturele participatie van ouders en kinderen	134
4.13	<i>Endogeneity bias</i> bij het verband tussen geweldscriminaliteit en huizenprijzen	135
4.14	Instrumentele variabelen model voor het verband tussen geweldscriminaliteit en huizenprijzen	136
4.15	Frequenties van categorieën reclasseringstoezicht	141
4.16	Verdeling van reclasseringstoezicht	141
4.17	Logistische regressie van recidive op de lengte van reclasseringstoezicht	142
4.18	Fitmaten van logistische regressie van recidive op de lengte van reclasseringstoezicht	142
4.19	<i>Endogeneity bias</i> bij het verband tussen de lengte van het reclasseringstoezicht en recidive	143
4.20	Lengte toezicht per type zedendader	144
4.21	Lengte van reclasseringstoezicht per arrondissement	146
4.22	Regressie-analyse in de eerste stap: fitmaten	148
4.23	Regressie-analyse in de eerste stap: <i>F</i> -toets	149
4.24	Regressie-analyse in de eerste stap: regressiemodel	149
4.25	Regressie-analyse in de tweede stap: fitmaten	150
4.26	Regressie-analyse in de tweede stap: regressiemodel	150
4.27	Stata-output van <code>ivprobit</code>	153
4.28	Stata-output van <code>ivreg2</code> : eerste deel van de 2SLS	155
4.29	Stata-output van <code>ivreg2</code> : tweede deel van de 2SLS	155
5.1	Donald Campbell (1916-1996)	164
5.2	Illustratie van <i>regression discontinuity design</i> : aantallen nieuwe delicten tegen snelheidsovertredingen	165
5.3	Bepalen van het effect van de interventie op de <i>threshold</i>	166
5.4	Voorbeeld van een niet toevallige verdeling om de <i>threshold</i>	167
5.5	Voorbeeld van misspecificatie van regression discontinuity model	168
5.6	Donald Thistlethwaite (1922-1997)	169
5.7	Histogram van de variabele ‘aanhoudingen’	170
5.8	<i>t</i> -toets voor verband tussen aantal aanhoudingen en behandeling	171
5.9	Descriptives van de variabele <code>JSOAP</code>	172
5.10	Scatterplot van ‘aanhoudingen’ en <code>JSOAP</code>	173
5.11	Histogram van de variabele <code>JSOAP</code>	175

5.12	Resultaat van de regressie-analyse voor de onbehandelde zedendelinquenten	177
5.13	Resultaat van de regressie-analyse voor de behandelde zedendelinquenten	178
5.14	Resultaat van de regressie-analyse voor de behandelde zedendelinquenten	179
5.15	Resultaat van de directe methode voor RDD analyse	181
6.1	Longitudinale data aantal delicten naar werk	189
6.2	Grafische illustratie van <i>fixed effects</i> schatting	196
6.3	Resultaat logistisch <i>fixed effects</i> panelmodel	200
6.4	Resultaat <i>two-way</i> logistisch <i>fixed effects</i> panelmodel	204
6.5	Resultaat logistisch <i>random effects</i> panelmodel	206
6.6	Resultaat Hausman-test	208
7.1	Gemiddeld verloop van de frequentie van delicten plegen	215
7.2	Verloop van de frequentie van delicten plegen voor twee homogene subgroepen	216
7.3	Daniel Nagin (1948)	218
7.4	Trajecten van criminele carrières bij hoog-risico jongeren (Van der Geest et al., 2011): Adolescence-limited daders (AL), Laاتبloeiers (LB), Laag frequente dalers (LFD), Hoog frequente dalers (HFD), Hoog frequent chronische daders (HFC)	220
7.5	Coëfficiënten van het model met één groep	231
7.6	Grafische weergave van de oplossing	232
7.7	Coëfficiënten van het model met vijf groepen	233
7.8	Grafische weergave van de gevonden trajecten voor het vijf groepen-model	234
7.9	<i>Mean group probabilities</i> van het model met vijf groepen	235
7.10	Coëfficiënten van het model met vijf groepen met IQ als covariaat	238
7.11	Coëfficiënten van het model met vijf groepen met werk als tijdvariërend covariaat	240



# Lijst van tabellen

2.1	Regressiemodellen met linkfunctie en assumptie over errortermen	62
3.1	Dossiervariabelen in SPSS bestand VM-JSO.sav	94
4.1	Criminologische studies met instrumentele variabelen	159
6.1	Voorbeeld van longitudinale data	188
6.2	Voorbeeld van longitudinale data inclusief individu-specifieke gemiddelden	193
6.3	Voorbeeld van longitudinale data inclusief individu-specifieke gemiddelden en gecentreerde variabelen	195
6.4	Evaluatie resultaat Hausman-test	208
7.1	Behandelkenmerken per trajectgroep Van der Geest et al. (2007)	221
7.2	Criteria om het beste model te kiezen bij <i>trajectory</i> analyse	224
7.3	BIC-waarden bij oplopend aantal groepen	232
8.1	Overzicht van gebruikte technieken	245