

Hoofdstuk 12

Complexere relaties tussen variabelen

1. Inleidende begrippen

Onafhankelijke, intermediaire en afhankelijke variabelen

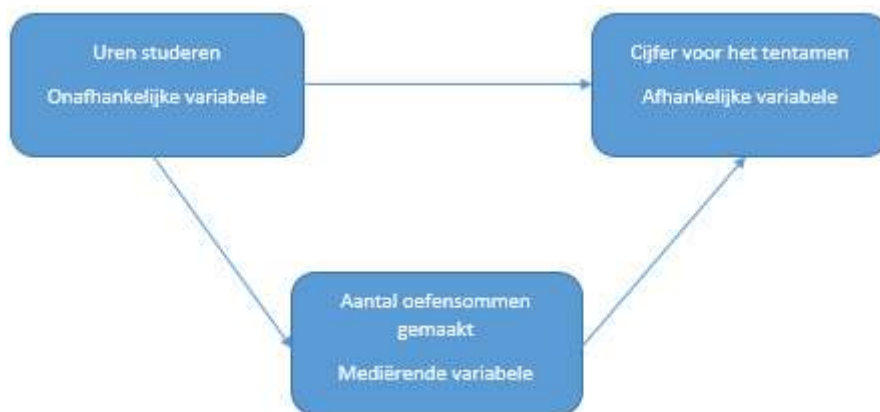
Bij een oorzaak-gevolgrelatie is er altijd sprake van een relatie tussen minstens twee variabelen en een mechanisme dat oorzaak en gevolg met elkaar verbindt. Onafhankelijke variabelen worden in de causale variabelentaal exogene variabelen genoemd. Zij hebben effecten op de endogene variabelen. Endogene variabelen zijn afhankelijke variabelen. In de pad-analyse zijn er meerdere afhankelijke variabelen en is het zo dat een afhankelijke variabele in een deel van het statistisch model ook een effect heeft op andere variabelen in het statistisch model.

Wanneer we de relatie tussen drie of meerdere variabelen bestuderen, spreken we van netwerken van relaties tussen variabelen. De concepten en hun onderlinge relaties kunnen worden samengevat in een **conceptueel model**. Dit wordt soms ook een **conceptueel diagram** genoemd. Wanneer de relaties tussen de variabelen op basis van empirisch onderzoek tot stand komen, spreken we van een **padmodel of een paddiagram**. Het begrip 'structureel model' wordt eveneens gebruikt om deze modellen aan te duiden. Zowel in het conceptuele model als in eenvoudige padmodellen zijn er **3 basisposities voor de variabelen**. **Helemaal rechts** staan de variabelen die verklaard worden door andere variabelen. Dit zijn de **afhankelijke variabelen**. **Helemaal links** staan de **onafhankelijke variabelen**. Dit zijn de **variabelen vanwaar het effect uitgaat**. Tussen deze twee types variabelen in staan de **intermediaire variabelen**. Dit zijn variabelen die verklaard worden door één of verscheidene onafhankelijke variabelen en die tevens verklarend zijn voor de afhankelijke variabelen.

2. Mediatorvariabele

Een mediërende variabele of mediatorvariabele is een variabele die tussen een oorzaak-gevolgrelatie staat en hiermee het effect tussen de oorzaak-gevolgrelatie beter verklaart en de relatie sterker maakt. De mediatorvariabele noemt men vaak een mechanisme waarlangs een externe oorzaak (of beter gezegd: de eigenschap van een gebeurtenis) het gevolg teweegbrengt.

In deze paragraaf gaan we door op het voorbeeld van de student die veel studeert (onafhankelijke variabele) om zo een hoger cijfer te halen voor zijn tentamen (afhankelijke variabele). We voegen nu een mediërende variabele aan het voorbeeld toe. De mediërende variabele staat tussen de onafhankelijke en afhankelijke variabele in (de oorzaak-gevolgrelatie) en stelt je in staat om de oorzaak-gevolgrelatie beter te verklaren. Een mediërende variabele kan erg lastig zijn om te interpreteren en om conclusies aan te verbinden. Daarom onderbouw je een mediërende variabele altijd met een statistische analyse. Zie hieronder een voorbeeld van een mediërende variabele.



Voorbeeld conceptueel model met mediërende variabele

In dit voorbeeld staat de relatie tussen de onafhankelijke variabele ‘Uren studeren’ en de afhankelijke variabele ‘Cijfer voor het tentamen’ centraal. De oorzaak-gevolgrelatie is dat hoe meer uren de student studeert, hoe hoger het cijfer is voor het examen. Nu hebben we de mediërende variabele ‘Aantal oefensommen gemaakt’ toegevoegd. Zoals je in het voorbeeld kunt zien, staat de mediërende variabele tussen de onafhankelijke en afhankelijke variabele. Hoe meer uren een student studeert, hoe meer oefensommen de student maakt en hoe meer oefensommen de student maakt, hoe hoger het cijfer voor het tentamen. Door de mediërende variabele ‘Aantal oefensommen gemaakt’ toe te voegen versterken we de oorzaak-gevolgrelatie tussen de onafhankelijke en afhankelijke variabele.

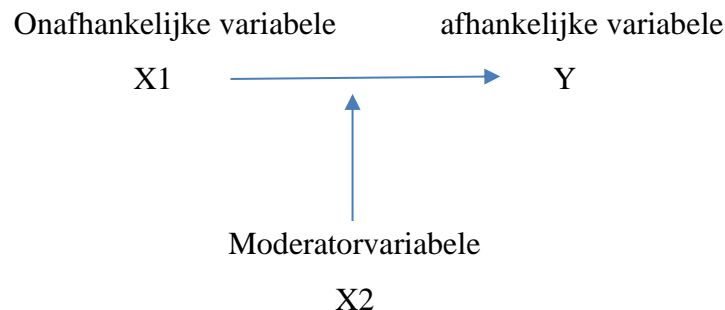
3. Moderatorvariabele of het interactie-effect

Inleiding

In kwantitatief onderzoek verwijst de term ‘interactie’ naar de situatie waarbij het effect van een onafhankelijke variabele X_1 op een afhankelijke variabele Y beïnvloed wordt door een derde onafhankelijke variabele X_2 . Deze laatste variabele wordt de **moderatorvariabele**

genoemd omdat deze variabele de wijze verandert waarop X1 een impact heeft op de afhankelijke variabele Y. Figuur 1 toont hoe een interactie-effect visueel wordt voorgesteld.

Figuur 1: schematische voorstelling van een interactie-effect



In criminologisch empirisch onderzoek is het belangrijk dat de onderzoeker op voorhand een bepaalde theoretische verwachting heeft. We bedoelen hiermee dat de onderzoeker, op basis van een theorie of voorgaand empirisch onderzoek, een bepaalde veronderstelling heeft van welke X-variabele in eerste instantie een effect heeft op een Y-variabele en van welke X-variabele verwacht wordt dat die de relatie wijzigt, dus welke X-variabele potentieel een moderatorvariabele is. Bijvoorbeeld: volgens sociale leertheorieën hebben delinquente vrienden een positief effect op het plegen van delicten bij adolescenten. Schematisch kunnen we deze relatie als volgt noteren:

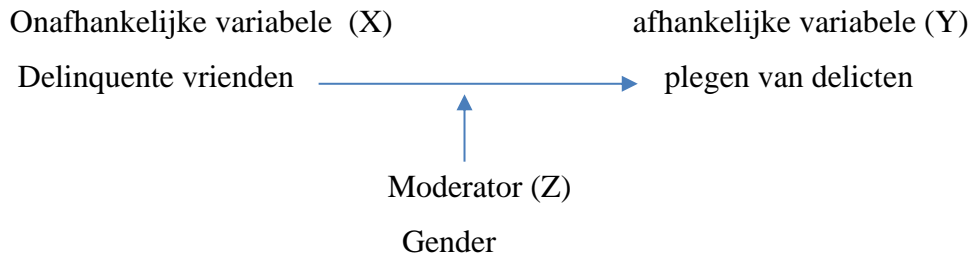
Figuur 2: schematische voorstelling van het hoofdeffect van delinquente vrienden op het plegen van delicten



We verwachten op basis van de sociale leertheorie in de eerste plaats een **hoofdeffect** van delinquente vrienden op het plegen van delicten;

Op basis van deze theoretische veronderstelling (die trouwens in heel veel criminologisch empirisch onderzoek wordt gerepliceerd) zouden we ons kunnen afvragen of de impact van delinquente vrienden op het plegen van delicten voor *elke* adolescent hetzelfde is. Wordt elke adolescent *op dezelfde manier* beïnvloed door delinquente vrienden? Zou het bijvoorbeeld niet kunnen dat er een verschil is tussen jongens en meisjes? Deze vraag stellen betekent dat we ons afvragen of de variabele 'gender' (meisje of jongen-zijn) een impact heeft op de

relatie tussen delinquente vrienden en het plegen van delicten. Als de effecten van delinquente vrienden op het plegen van delicten inderdaad voor jongens en meisjes anders zijn, dan spreekt men van *interactie*. Schematisch kunnen we dit voorstellen als volgt:



De verwachting is dat de relatie delinquente vrienden en het plegen van delicten anders is voor jongens dan voor meisjes.¹. Als we verwachten dat het effect van delinquente vrienden op het plegen van delicten verschillend is voor meisjes en voor jongens dan beschouwen we ‘gender’ als een **moderatorvariabele**. Dus nogmaals: in het geval van interactie onderzoeken we of het effect van X op Y verandert als Z een andere waarde heeft (dat wil zeggen: andere groep, categorie of conditie).

Enkele voorbeelden van interactie-effecten tussen variabelen op laag meetniveau

Hierna geven we twee concrete voorbeelden van statistische interacties aan de hand van kruistabellen. De voorbeelden zijn afkomstig uit voorgaand criminologisch onderzoek.²

In het eerste voorbeeld kijken we naar de relatie tussen immigratie-achtergrond (nominale variabele met 2 categorieën: Belg of immigrant) en lidmaatschap van een problematische jeugdgroep (nominale variabele met 2 categorieën : wel of geen lidmaatschap), dus twee variabelen van het nominale meetniveau.

¹ In dit geval formuleren we onze verwachting nog zeer algemeen zonder een uitspraak te doen over de sterkte van het hoofdeffect. Hou steeds voor ogen dat verwachtingen of veronderstellingen die onderzoekers uitspreken altijd gebaseerd zijn op theorie of voorgaand onderzoek. De keuze voor een moderatorvariabele is immers niet altijd eenduidig. Wat voor de ene onderzoeker een moderatorvariabele is, is dat soms niet voor een andere onderzoeker. Vandaar dat het belangrijk is om de keuze goed theoretisch te onderbouwen (of op basis van eerder empirisch onderzoek).

² Zie voor meer informatie: PAUWELS, L., & HARDYNS, W. (2016). *Problematic youth group involvement as situated choice: testing an integrated conditions-controls-exposure model*. Eleven International Publishing.

Tabel 1: Kruistabel lidmaatschap problematische jeugdgroep naar immigratie-achtergrond

Lidmaatschap problematische jeugdgroep naar immigratie-achtergrond					
			Immigratie-achtergrond		
			Belg	Immigrant	Totaal
PROBLEMATISCHE JEUGDGROEP	GEEN LID	Absolute aantal	1041	1235	2276
		Kolompercentage	95.1%	89.4%	91.9%
	WEL LID	Absolute aantal	54	146	200
		kolompercentage	4.9%	10.6%	8.1%
Total		Absolute aantal	1095	1381	2476
		kolompercentage	100.0%	100.0%	100.0%

$\chi^2 = 26.170$; $df = 1$; $p < .001$

Tabel 1 is een kruistabel die de samenhang weergeeft tussen 2 nominale variabelen. In de kolommen staat de variabele ‘immigratie-achtergrond’ die bestaat uit 2 categorieën: Belg en immigrant. In de rijen staat de variabele ‘problematische jeugdgroep’ die ook bestaat uit 2 categorieën: geen lid en wel lid. Herinner je uit hoofdstuk 6 (*bivariate associatiematen voor nominale en ordinale variabelen*) dat we bij het maken van een bivariate kruistabel statistisch een onderscheid maken tussen een onafhankelijke en een afhankelijke variabele en dat de richting van de associatie bepaald wordt door onze theoretische verwachtingen. Herinner je ook dat we in deze cursus de afspraak maakten om de variabele die we beschouwen als onafhankelijke variabele in de kolommen te plaatsen en de afhankelijke variabele in de rijen. Elke cel geeft ons informatie over de twee variabelen. Kijken we naar de conditionele verdelingen (hoe is één van de twee variabelen verdeeld binnen één categorie van de andere variabele) dan zien we voor de categorie Belg dat 95,1% van de respondenten rapporteerden GEEN LID te zijn van een problematische jeugdgroep en 4,9% rapporteerden WEL LID te zijn. Voor de categorie Immigrant rapporteerden 89,4% van de respondenten GEEN LID te zijn van een problematische jeugdgroep en 10,6% rapporteerden WEL LID te zijn. We besluiten dat de frequentieverdeling voor de beide categorieën van de variabelen niet dezelfde is: er is een verschil tussen Belgen en Immigranten in termen van hun betrokkenheid bij een problematische jeugdgroep. Het verband tussen beide variabelen is bovendien significant ($\chi^2=26.170$; $p<.001$).

Om na te gaan of de samenhang tussen immigratie-achtergrond en lidmaatschap problematische jeugdgroep verschillend is voor meisjes en voor jongens kijken we naar tabel 2. Deze kruistabel toont een aantal verschillen. Uit tabel 2 blijkt dat van de Belgische meisjes slechts 3.6% aangaf WEL LID te zijn van een problematische jeugdgroep. Bij de Belgische jongens ligt de proportie duidelijk hoger, respectievelijk 6.4%. Kijken we naar de categorie Immigrant dan merken we dat binnen deze groep 6.9% van de meisjes aangeeft WEL LID te zijn van een problematische jeugdgroep, terwijl deze proportie bij de jongens ook duidelijk hoger ligt, namelijk 14.2% van de Immigrant-jongens geeft aan WEL LID te zijn van een problematische jeugdgroep.

We merken dus duidelijk dat de relatie tussen immigratie-achtergrond en lidmaatschap problematische jeugdgroep verschilt tussen de geslachten. Er zijn meer jongens, zowel in de categorie Belg als Immigrant, die WEL LID zijn van een problematische jeugdgroep, in vergelijking met Belgische meisjes en Immigrant-meisjes.

Tabel 2: kruistabel lidmaatschap problematische jeugdgroep naar immigratie-achtergrond en geslacht

GESLACHT				Immigratie-achtergrond		Totaal
				Belg	immigrant	
MEISJE	PROBLEMATISCHE JEUGDGROEP	GEEN LID	Absolute aantal	541	644	1185
			kolompercentage	96.4%	93.1%	94.6%
		WEL LID	Absolute aantal	20	48	68
			kolompercentage	3.6%	6.9%	5.4%
	Totaal		Absolute aantal	561	692	1253
			kolompercentage	100.0%	100.0%	100.0%
	JONGEN	PROBLEMATISCHE JEUGDGROEP	Absolute aantal	500	590	1090
			kolompercentage	93.6%	85.8%	89.2%
		WEL LID	Absolute aantal	34	98	132
			kolompercentage	6.4%	14.2%	10.8%
	Totaal		Absolute aantal	534	688	1222
			kolompercentage	100.0%	100.0%	100.0%

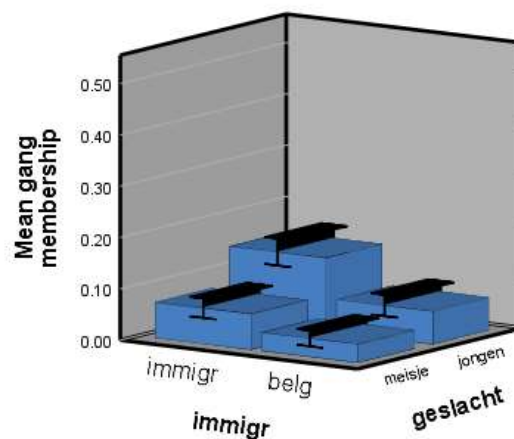
Meisjes: $\chi^2 = 6,861$; $df=1$; $p<.01$

Jongens: $\chi^2 = 19,361$; $df=1$; $p<.001$

In figuur 2 presenteren we de resultaten aan de hand van een drie-dimensionele bar chart waarbij de proportie respondenten die aangeeft WEL LID te zijn van een problematische jeugdgroep duidelijk kan worden afgelezen voor de twee categorieën immigratie-achtergrond en voor zowel jongens als voor meisjes. De figuur spreekt voor zich: de proportie respondenten die aangeeft WEL LID te zijn van een problematische jeugdgroep is het laagst bij de Belgische meisjes en het hoogst bij de immigrant-jongens.

Figuur 2: 3D-voorstelling van interactie tussen immigratie-achtergrond en geslacht

Lidmaatschap problematische jeugdgroep naar immigratie-achtergrond en geslacht



Error Bars: 95% CI

We presenteren een tweede voorbeeld in tabel 3.

In eerste instantie onderzoeken we de relatie tussen de variabelen ‘veelpleger’, die we definiëren als respondenten die minstens 4 gepleegde feiten rapporteren, en gezinsstructuur (bestaande uit twee categorieën: éénouder en twee-oudergezin).

Uit de resultaten blijkt dat 87.6% van de respondenten die opgroeien in een éénoudergezin aangeven ‘GEEN veelpleger’ te zijn tegenover 12.4% die aangeeft ‘WEL veelpleger’ te zijn. Binnen de categorie ‘twee-oudergezin’ zijn de proporties respectievelijk 82.1% (GEEN veelpleger) en 17.9% (WEL veelpleger). We zien dus dat de frequentieverdeling voor de beide categorieën van de variabelen niet dezelfde is: er is een verschil tussen respondenten die opgroeien binnen een éénouder- of twee-oudergezinsstructuur met betrekking tot GEEN/WEL veelpleger. Het verband tussen beide variabelen is significant ($\chi^2 = 8.091$; $p < .01$).

Tabel 3: Kruistabel veelpleger en gezinsstructuur

Veelpleger naar gezinsstructuur			gezinsstructuur		Totaal
			Woont in éénouder gezin	Woont in twee-ouder gezin	
VEELPLEGER (minstens 4 feiten)	GEEN VEELPLEGER	Absolute aantal kolompercentage	1818 87.6%	299 82.1%	
	WEL VEELPLEGER	Absolute aantal kolompercentage	257 12.4%	65 17.9%	322 13.2%
Total		Absolute aantal kolompercentage	2075 100.0%	364 100.0%	2439 100.0%

$\chi^2 = 8.091$; $df=1$; $p<.01$

In tweede instantie vragen we ons af of het verband tussen ‘gezinsstructuur’ en ‘veelpleger’ verschilt naargelang de mate van familiale controle. In tabel 4 presenteren we de frequentieverdelingen van de categorieën gezinsstructuur voor GEEN/WEL veelpleger en LAGE/HOGE familiale controle.

Wat kunnen we aflezen uit de proportionele verdeling ?

Tabel 4: Kruistabel veelpleger naar gezinsstructuur (éénouder – twee-ouders) en familiale controle (laag – hoog)

Veelpleger naar gezinsstructuur en familiale controle

				Gezinsstructuur		
				Woont in éénouder gezin	Woont in twee-ouder gezin	Totaal
FAMILIALE CONTROLE						
LAGE	VEELPLEGER	GEEN	Absolute aantal	997	159	1156
FAMILIALE	(minstens 4 feiten)	VEELPLEGER	kolompercentage	82.0%	73.6%	80.7%
CONTROLE						
		VEELPLEGER	Absolute aantal	219	57	276
			kolompercentage	18.0%	26.4%	19.3%
	Total		Absolute aantal	1216	216	1432
			kolompercentage	100.0%	100.0%	100.0%
HOGHE	VEELPLEGER	GEEN	Absolute aantal	797	138	935
FAMILIALE	(minstens 4 feiten)	VEELPLEGER	kolompercentage	95.9%	95.2%	95.8%
CONTROLE						
		VEELPLEGER	Absolute aantal	34	7	41
			kolompercentage	4.1%	4.8%	4.2%
	Total		Absolute aantal	831	145	976
			kolompercentage	100.0%	100.0%	100.0%

Lage familiale controle: $\chi^2 = 8,277$; $df=1$; $p<.01$ Hoge familiale controle: $\chi^2 = 0,166$; $df = 1$; $p=.683$

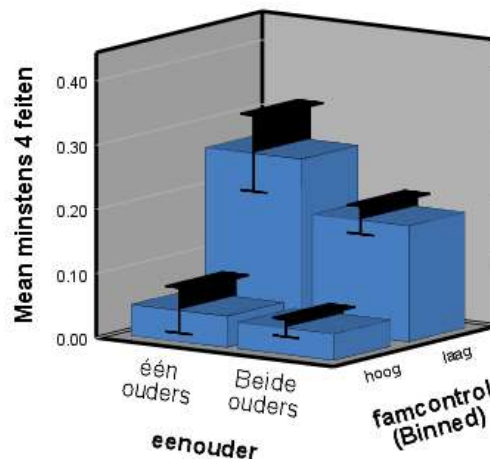
Uit tabel 4 lezen we af dat van de respondenten die opgroeien binnen een éénoudergezin met een lage familiale controle 18.0% veelpleger blijkt te zijn. Deze proportie ligt veel lager voor respondenten uit een éénoudergezin met hoge familiale controle, namelijk 4.1%. Kijken we naar de categorie twee-oudergezin dan merken we dat binnen deze groep 26.4% veelpleger is wanneer er sprake is van een lage familiale controle terwijl slechts 4.8% van de respondenten veelpleger is wanneer er sprake is van hoge familiale controle. We merken duidelijk dat er verschillen zijn in de samenhang tussen gezinsstructuur en veelpleger naargelang de mate van familiale controle.

In figuur 3 presenteren we terug de resultaten aan de hand van een drie-dimensionele bar chart waarbij de proportie respondenten die aangeeft veelpleger te zijn duidelijk kan worden afgelezen voor de twee categorieën gezinsstructuur en mate van familiale controle(hoog/laag).

De visuele voorstelling is zeer duidelijk: de proportie respondenten die aangeeft veelpleger te zijn is het laagst binnen de één- en tweeouder gezinsstructuur met hoge familiale controle en overduidelijk het hoogst binnen de categorie éénoudergezin met een lage familiale controle.

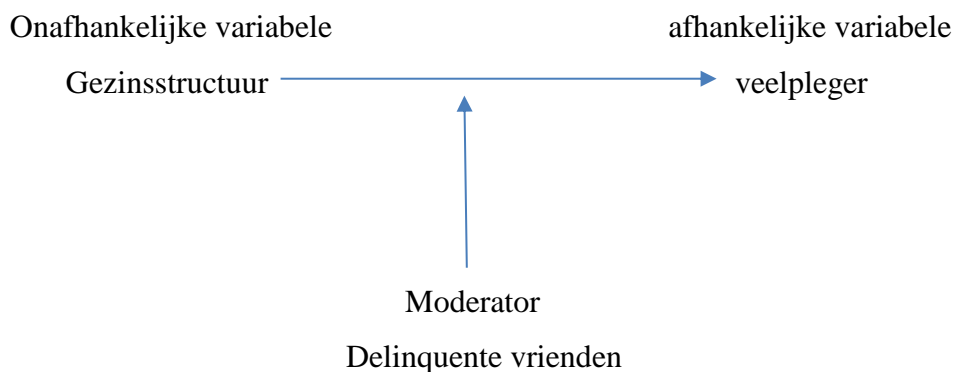
Figuur 3: 3D-voorstelling van interactie tussen gezinsstructuur en familiale controle

Veelpleger naar gezinsstructuur (éénouder - twee-ouders) en familiale controle (laag - hoog)



Error Bars: 95% CI

Tot slot nog een laatste voorbeeld met een moderatorvariabele van het ordinale meetniveau: delinquente vrienden (laag – medium – hoog). We onderzoeken de relatie tussen gezinsstructuur, delinquente vrienden en veelpleger. Laat ons de veronderstelde samenhang eerst conceptueel voorstellen. We veronderstellen een hoofdeffect van gezinsstructuur op de afhankelijke variabele veelpleger. Daarnaast verwachten we dat deze samenhang verschillend zal zijn naar de mate waarin respondenten delinquente vrienden hebben. We veronderstellen dat delinquente vrienden een moderator is in de relatie tussen gezinsstructuur en veelpleger.



Tabel 5 presenteert de frequentieverdelingen van de twee categorieën gezinsstructuur (éénouder en twee-ouders) naar GEEN/WEL veelpleger en drie categorieën delinquente vrienden (laag – medium – hoog). Helemaal onderaan de kruistabel lezen we de totale frequenties af voor wat betreft de samenhang tussen gezinsstructuur en veelpleger (totaal).

12.1% van de respondenten die opgroeien in een tweeoudergezin zijn veelpleger tegenover 17.8% van de respondenten die opgroeien in een éénoudergezin. Het verband tussen de variabelen ‘veelpleger’ en ‘gezinsstructuur’ is significant ($\chi^2 = 8.743$; $p < .01$). Vervolgens gaan we na of de samenhang tussen gezinsstructuur en veelpleger verschillend is naargelang het hebben van delinquente vrienden. ‘Delinquente vrienden’ is een variabele van het ordinale meetniveau en bestaat uit drie categorieën: laag – medium – hoog. Voor wat betreft de categorie: laag niveau van delinquente vrienden zien we geen verschil tussen respondenten uit een éénouder of tweeoudergezin met betrekking tot veelpleger zijn (resp. 1.2% en 1.3%). Voor de categorie: medium niveau van delinquente vrienden zien we dat 7.0% van de respondenten uit een tweeoudergezin veelpleger zijn tegenover 10.4% van de respondenten uit een éénoudergezin. Het verschil bedraagt hier 3.4 procentpunten. In de categorie: hoog niveau van delinquente vrienden is het verschil tussen veelpleger in een éénouder- en een tweeoudergezin 4.1 procentpunten (resp. 42.9% en 38.8%). Besluit: we merken dat de samenhang tussen gezinsstructuur en veelpleger verschillend is naar de mate waarin respondenten delinquente vrienden hebben.

Tabel 5 : Kruistabel veelpleger naar delinquente vrienden en gezinsstructuur

Veelpleger naar delinquente vrienden (laag – medium – hoog) en gezinsstructuur (éénouder – tweeouders)

DELINQUENTE VRIENDEN				Gezinsstructuur		Totaal
				Woont in Twee- oudergezin	Woont in één oudergezin	
Delinquente vrienden : LAAG	VEELPLEGER (minstens 4 feiten)	GEEN VEELPLEGER	Absolute aantal	468	83	551
			kolompercentage	98.7%	98.8%	98.7%
		VEELPLEGER	Absolute aantal	6	1	7
			kolompercentage	1.3%	1.2%	1.3%
	Total		Absolute aantal	474	84	558
			kolompercentage	100.0%	100.0%	100.0%
Delinquente vrienden: MEDIUM	VEELPLEGER (minstens 4 feiten)	GEEN VEELPLEGER	Absolute aantal	1057	147	1204
			kolompercentage	93.0%	89.6%	92.6%
		VEELPLEGER	Absolute aantal	79	17	96
			kolompercentage	7.0%	10.4%	7.4%
	Total		Absolute aantal	1136	164	1300
			kolompercentage	100.0%	100.0%	100.0%
Delinquente vrienden: HOOG	VEELPLEGER (minstens 4 feiten)	GEEN VEELPLEGER	Absolute aantal	252	60	312
			kolompercentage	61.2%	57.1%	60.3%
		VEELPLEGER	Absolute aantal	160	45	205
			Kolompercentage	38.8%	42.9%	39.7%
	Total		Absolute aantal	412	105	517
			kolompercentage	100.0%	100.0%	100.0%

Delinquente vrienden laag: Chi²= 0,003; df=1; p= .954

Delinquente vrienden medium: Chi²= 2,439; df=1; p= .118

Delinquente vrienden hoog: Chi²= 0,566; df=1; p=.452

Interactie in een meervoudige lineaire regressie

Het gebruik van kruistabellen met verschillende niveaus om interactie-effecten op te sporen heeft beperkingen wanneer we het effect van meerdere onafhankelijke variabelen op één afhankelijke variabele onderzoeken, zeker als één of meerdere onafhankelijke variabelen van het metrisch meetniveau zijn.

Een regressiemodel laat toe het effect van verschillende onafhankelijke variabelen op één afhankelijke variabele gelijktijdig te onderzoeken. Als er twee onafhankelijke variabelen zijn dan is de regressievergelijking als volgt:

$$\hat{Y} = \alpha + \beta_1 X_1 + \beta_2 X_2$$

Herhaling van de algemene interpretatie van de regressievergelijking:

\hat{Y} is de voorspelde waarde van de afhankelijke variabele, α is het intercept of de voorspelde waarde voor de afhankelijke variabele als alle onafhankelijke variabelen gelijk zijn aan 0. De richtingscoëfficiënten of regressiecoëfficiënten (β_1 en β_2) geven het effect aan van de onafhankelijke variabelen X_1 en X_2 op de afhankelijke variabele. In een meervoudige regressievergelijking (met meerdere onafhankelijke variabelen) is dat het effect van X op Y gecontroleerd voor de effecten van alle andere onafhankelijke variabelen in het model. Letterlijk is β_1 dan de voorspelde wijziging in Y bij één eenheid wijziging in X_1 onder controle van de effecten van X_2 op Y .

Om een interactie-effect in een regressiemodel te schatten wordt in het model een productterm van de inter-agerende variabelen opgenomen. Als er twee onafhankelijke variabelen zijn, dan ziet de vergelijking eruit als volgt:

$$\hat{Y} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$$

In deze vergelijking worden β_1 en β_2 wel eens hoofdeffecten (*main effects*) genoemd en β_3 het interactie effect.

De interpretatie van het intercept is in deze vergelijking dezelfde: de voorspelde waarde voor de afhankelijke variabele als alle onafhankelijke variabelen gelijk zijn aan 0.

β_1 is het effect van X_1 op Y als X_2 gelijk is aan 0 en β_2 is het effect van X_2 op Y als X_1 gelijk is aan 0. Het is dus duidelijk dat het effect van X_1 op Y afhankelijk is van X_2 en omgekeerd. Tot slot: β_3 geeft het effect weer van de interactieterm.

Laat ons een voorbeeld geven van een meervoudig regressiemodel met als afhankelijke variabele zelf gerapporteerde criminaliteit en als onafhankelijke variabelen geslacht, immigratieachtergrond en cohorte (als controlevariabelen), criminele geneigdheid en lifestyle risk en de interactieterm tussen deze laatste als onafhankelijke variabelen.

Tabel 6: regressiemodel met zelf gerapporteerde criminaliteit als afhankelijke variabele en geslacht, immigratie achtergrond, cohorte, criminele geneigdheid, lifestyle risk en de interactie tussen de laatste twee variabelen als onafhankelijke variabelen.

	Model 1 b/beta (S.E.)	Model 2 b/beta (S.E.)	Model 3 b/beta (S.E.)	Model 4 b/beta (S.E.)
Intercept	7.026*** (.690)	6.485*** (.582)	6.607*** (.545)	5.535*** (.518)
Geslacht	-5.471/-.253*** (.718)	-2.770/-.128*** (.623)	-1.265/-.058* (.599)	-1.868/- .086*** (.562)
Immigratieachtergrond	1.314/.061 (.721)	2.448/.113*** (.611)	2.248/.104*** (.572)	2.416/.112*** (.534)
Cohorte	3.088/.143*** (.716)	.780/.036 (.617)	-.647/-.030 (.591)	-.203/-.009 (.554)
Criminele geneigdheid		5.872/.545*** (.319)	4.708/.437*** (.317)	4.322/.401*** (.298)
Lifestyle risk (Blootstelling aan criminogene settings)			3.551/.330*** (.324)	2.412/.224*** (.320)
Criminele geneigdheid x lifestyle risk (interactieterm)				2.518/.294*** (.227)
R²	.086	.351	.433	.506
Verandering in F- waarde	26.147***	339.023***	119.768***	123.005***

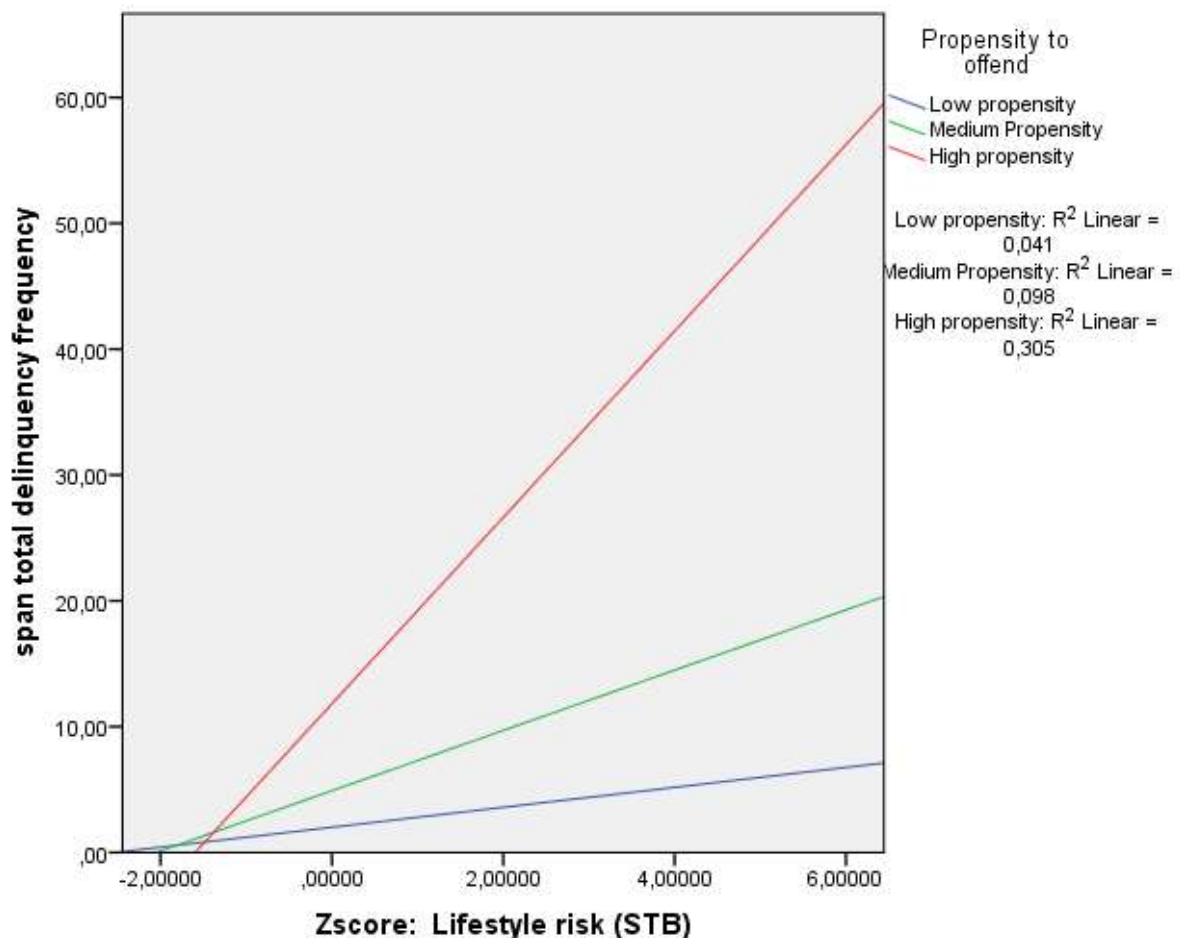
* $p < .05$; ** $p < .01$; *** $p < .001$

Tabel 6 toont de resultaten van een bloksgewijze regressieanalyse. In model 1 worden de coëfficiënten van een aantal controlevariabelen gepresenteerd (geslacht, immigratieachtergrond en cohorte waartoe men behoort). In model twee wordt ‘criminele geneigdheid’ toegevoegd aan de analyse. We zien dat criminele geneigdheid een sterk positief effect heeft op zelf gerapporteerde criminaliteit (Beta .545; $p < .001$). Model 3 voegt de onafhankelijke variabele lifestyle risk (of blootstelling aan criminogene settings) toe. Ook deze variabele heeft een sterk en positief netto-effect op zelf gerapporteerde criminaliteit (Beta .330; $p < .001$). Merk op dat het effect van criminele geneigdheid in model 3 lichtjes daalt maar sterk positief en significant blijft (Beta .437; $p < .001$). Tot slot voegen we in model

4 de interactieterm tussen criminele geneigdheid en lifestyle risk toe. Het interactie-effect is sterk positief en bovendien significant (Beta .294; $p < .001$). Kijken we naar de determinatiecoëfficiënt dan zien we dat toevoeging van de interactieterm de verklaaringswaarde van het model verhoogt met 7.3% (R^2 in model 4 = .506 en R^2 in model 3 = .433). De verandering in de F-waarde die bij de determinatiecoëfficiënt hoort is eveneens significant, hetgeen wil zeggen dat het model met de interactieterm de data beter samenvat dan het statistische regressiemodel zonder de interactieterm. Alle variabelen in het volledige model verklaren 50.6% van de variantie in zelf gerapporteerde criminaliteit. Bemerkt ook nog dat criminele geneigdheid (Beta .401; $p < .001$) en lifestyle risk (Beta .224; $p < .001$) wel een positief netto-effect op zelf gerapporteerde criminaliteit behouden. De netto effecten zijn weliswaar minder sterk maar ze blijven wel significant.

We geven het interactie-effect ook grafisch weer in figuur 4.

Figuur 4: visuele weergave van de interactie tussen criminele geneigdheid (propensity to offend) en lifestyle risk



Figuur 4 geeft het verband weer tussen blootstelling aan criminogene settings (het aantal uur aanwezig op plaatsen met lage sociale cohesie en delinquente vrienden) en het aantal delicten per uur (totale som). Dit voorbeeld is afkomstig uit een unieke studie waarbij de kenmerken van de omgeving werden gemeten aan de hand van een bijzondere methode: de STB- space time budget methode. Dit is een soort van dagboek waarin uur per uur de blootstelling van de respondent aan kenmerken in geografische micro-plaatsen wordt voorgesteld.

We zien dat er een heel duidelijke link is tussen lifestyle risk (of het aanwezig zijn op plaatsen met lage sociale cohesie en delinquente vrienden) en het aantal delicten dat men pleegt, maar we zien ook dat het verband niet voor iedereen hetzelfde is. De regressie- of voorspellingslijnen lopen niet parallel. Voor respondenten met een hoge criminele geneigdheid (high propensity) is de regressierechte veel steiler dan voor de respondenten met medium en lage criminele geneigdheid (medium en low propensity). Voor respondenten met lage criminele geneigdheid is de voorspellingsrechte veel vlakker. De hellingsgraad toont dus de sterkte van het effect van lifestyle risk dat veel sterker is voor respondenten met hoge criminele geneigdheid en zwakker voor respondenten met lage criminele geneigdheid. Concreet betekent dit dat jongeren die laag scoren op het kenmerk ‘geneigdheid tot regelovertreding’ nauwelijks beïnvloed worden door een criminogene omgeving, maar je ziet ook dat onder jongeren die hoog scoren op *propensity to offend* (dit betekent: sterk geneigd zijn regelovertreding als alternatief te zien) een sterke relatie bestaat tussen blootstelling aan criminogene settings en het aantal delicten dat men pleegt per uur in criminogene settings.

4. De pad-analyse

De pad-analyse dankt zijn naam aan de analyse van “paden”, dit zijn statistische paden of manieren waarlangs een variabele een andere beïnvloedt. De klassieke pad-analyse bestaat uit een reeks van structurele vergelijkingsmodellen tussen gemeten variabelen. Een pad-analyse: analyse van structurele modellen waarbij alle variabelen geobserveerd (manifest) zijn. De methode stamt oorspronkelijk uit de biologie en is ontworpen door de bioloog Sewell Wright in 1922. Wanneer de pad-coëfficiënten gestandaardiseerd zijn, kan men ze vergelijken met de gestandaardiseerde regressiegewichten, dus de β -coëfficiënten.

Een structureel model representeert alle causale hypothesen omtrent de patronen van directe en indirecte effecten tussen alle variabelen in een statistisch model. Dit is een zeer vaak gebruikte manier om hypothesen te toetsen uit criminologische theorieën. **Elk theoretisch construct in het model (bvb. zelf-gerapporteerde criminaliteit, ouderlijke controle,**

criminele geneigdheid en blootstelling aan criminogene morele settings,...) wordt gemeten door 1 geobserveerde variabele en correspondeert dus met 1 variabele in de dataset.

Directe en indirecte effecten

Door de opeenvolging van directe effecten kunnen in een pad-analyse op een natuurlijke wijze indirecte effecten worden gemodelleerd (mediatie).

We geven een voorbeeld waarbij de volgorde van de variabelen een rol spelen: x_1 , y_1 , y_2 , y_3 .

X_1 is een onafhankelijke variabele of een **exogene variabele**. Een exogene variabele is een variabele waar geen causale effectrelaties (pijlen) toekomen. Er vertrekken enkel pijlen naar andere variabelen (y_1 en y_3).

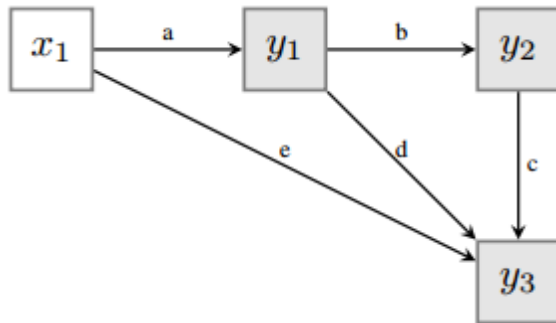
De Y-variabelen (y_1 , y_2 en y_3) zijn allemaal **endogene variabelen**. Dit wil zeggen dat er pijlen toekomen. Ze worden causaal bepaald door een reeks andere variabelen.

Y_1 wordt rechtstreeks bepaald door x_1 . Y_2 wordt indirect bepaald door x_1 en direct door y_1 . Y_3 wordt bepaald door het rechtstreekse effect van x_1 , het rechtstreekse effect van y_1 en het rechtstreekse effect van y_2 .

Naast al deze rechtstreekse effecten zijn er nog een reeks onrechtstreekse effecten. Dit zijn mediërende effecten of onrechtstreekse effecten. De onrechtstreekse effecten zijn causale paden waarlangs een variabele een effect uitoefent op een andere variabele, maar nooit rechtstreeks, altijd via de impact op een tussenliggende variabele.

We geven een voorbeeld door alle indirecte effecten te bespreken die uit de figuur af te leiden zijn.

- Er is een indirect effect vast te stellen van x_1 op y_3 , met name een indirect effect dat verloopt via y_1 en y_2 naar y_3 , maar ook een effect via y_1 naar y_3 .
- e = het rechtstreekse effect van x_1 op y_3
- a = het rechtstreekse effect van x_1 op y_1
- b = het rechtstreekse effect van y_1 op y_2
- c = het rechtstreekse effect van y_2 op y_3
- d = het rechtstreekse effect van y_1 op y_3



We kunnen dit pad-model ook uitschrijven aan de hand van een reeks vergelijgingsmodellen (equations). Dit doen we door de regressievergelijking uit te schrijven voor elke endogene variabele. Bij het uitschrijven vermelden we enkel de rechtstreekse effecten. De onrechtstreekse effecten zijn af te leiden uit de vergelijkingen. Wanneer we de vergelijkingen tekenen, krijgen we zicht op de mogelijke manieren waarop de variabelen met elkaar in verband staan.

Laat x_1 de sociale controle in het gezin zijn, y_1 de criminele geneigdheid en y_2 de blootstelling aan criminogene morele settings. Sociale controle heeft effecten op zelf-gerapporteerde criminaliteit via de impact op de criminele geneigdheid, maar heeft ook een rechtstreeks effect op de zelf-gerapporteerde criminaliteit. De criminele geneigdheid heeft op diens beurt ook een rechtstreeks effect op de zelf-gerapporteerde criminaliteit, maar ook onrechtstreeks via de blootstelling aan criminogene morele settings.

$$Y_1 = ax_1 + e_1$$

$$Y_2 = by_1 + e_2$$

$$Y_3 = ex_1 + dy_1 + cy_2 + e_3$$

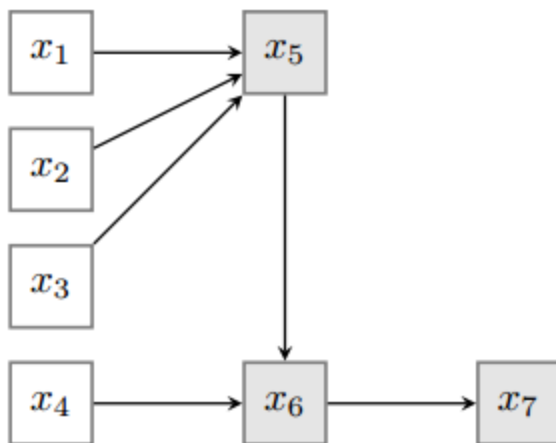
5. Berekening van de totale en indirecte effecten in de pad-analyse

Een totaal effect = alle rechtstreekse effecten + alle onrechtstreekse effecten

- **Het totale effect van x_1 op y_1** = de bivariate correlatie (x_1, y_1) of de β -coëfficiënt in de regressie van y_1 op x_1 .
- **Het totale effect van x_1 op y_2** = Het rechtstreekse effect van x_1 op y_2 (= 0 want er is geen rechtstreekse pijl) + het onrechtstreekse effect van x_1 op y_2 via y_1 en dat is gelijk aan $a*b$.

- Het totale effect van x_1 op y_3 = de som van het rechtstreekse effect van x_1 op y_3 en de diverse onrechtstreekse effecten. Hier hebben we een direct effect (e), een binnenste onrechtstreeks effect (van x_1 op y_3 via y_1) ($a*d$) en een buitenste onrechtstreeks effect (van x_1 op y_3 via y_1 en y_2) ($a*b*c$). Dus de berekening van het totale effect = $e + a*d + a*b*c$. Nog steeds geldt dat er geen onrechtstreeks effect is van x_1 op y_3 via y_2 omdat er geen rechtstreeks effect is van x_1 op y_2 .

6. Nog een voorbeeld van een pad-model



Overzicht

In deze pad-analyse zijn er meerdere exogene variabelen. De exogene variabelen zijn $x_1 - x_4$, want hier komen geen pijlen toe.

Er zijn meerdere endogene variabelen. Een aantal endogene variabelen heeft een dubbel statuut: x_5 - x_6 = mediators = ze mediëren de effecten van de onafhankelijke variabelen.

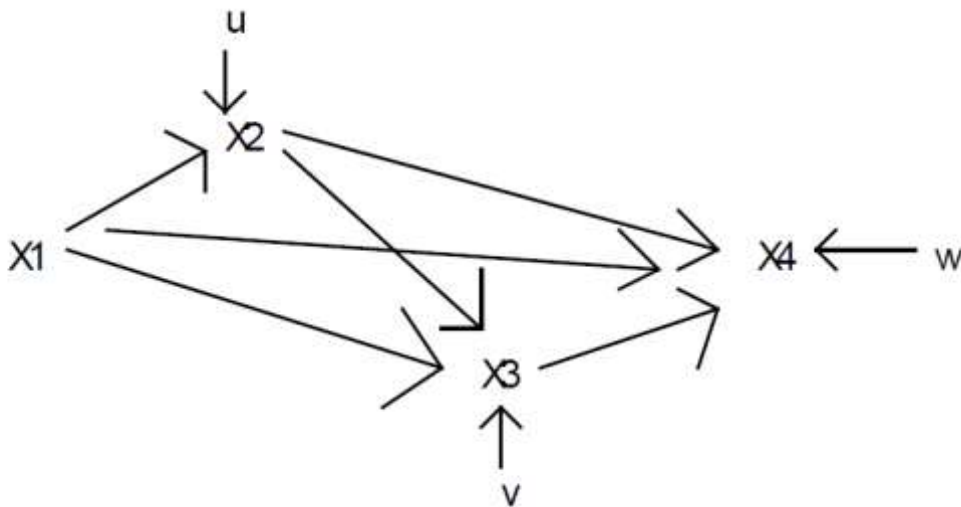
x_7 : uitkomstvariabele of hoofdvariabele.

We kunnen $x_1 \rightarrow x_7$ met een omweg bereiken.

Ook dit pad-model is een voorstelling van de realiteit: het is de hypothese van de onderzoeker, die zich door een theorie heeft laten inspireren.

Laten we de rekenoefening eens maken aan de hand van een intussen klassiek geworden voorbeeld uit de sociologie, uit de cursus van Otis Dudley Duncan, één van de klassieke handboeken over pad-analyse.

Ziehier een reeks van causale effecten.



Laten we dit pad-model eens onder de loop nemen en elk effect bespreken. Om de paden te identificeren kunnen we twee manieren volgen: we werken van de oorzaken naar de gevolgen of we werken omgekeerd, van de gevolgen naar de oorzaken. Het hierboven getekend pad-model is een recursief model, dit wil zeggen dat er geen feedback loops zijn, of geen wederzijdse causale effecten.

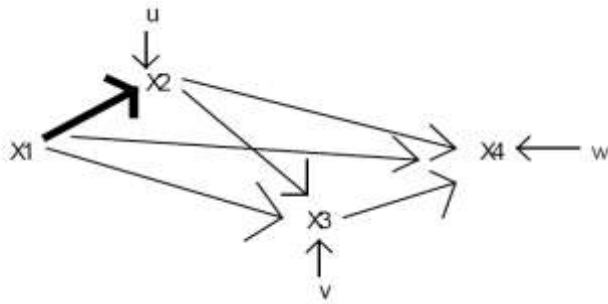
Er is een direct effect van x_1 op x_2 .

Wanneer we paden bestuderen, dan geldt de volgende regel:

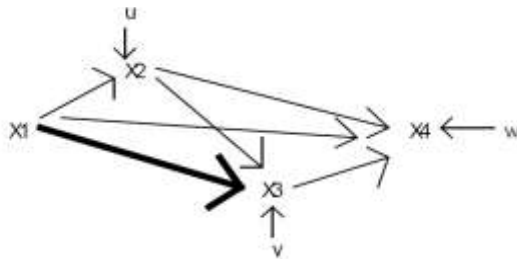
Ga na hoe je van de variabelen aan de linkerkant naar de variabelen aan de rechterkant geraakt. Elke pijl wijst op een route.

- Als je één variabele kan terugkeren en dan kan stoppen, heb je een direct effect geïdentificeerd.
- Als je twee variabelen kan terugkeren en niet opnieuw terug vooruit kan, heb je een exclusief indirect effect geïdentificeerd.
- Als je een variabele kan vinden die een pijl trekt naar twee variabelen die verderop in het model staan, heb je een gemeenschappelijke oorzaak ontdekt.

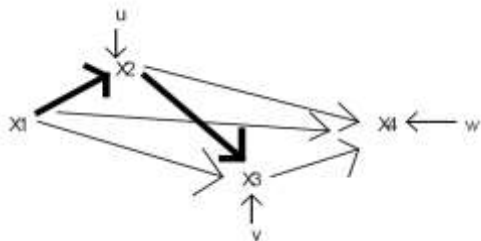
We leggen het principe van de pad-analyse uit aan de hand van een voorbeeld waarbij we wijzen op de verschillende effecten tussen x_1 en x_4 .



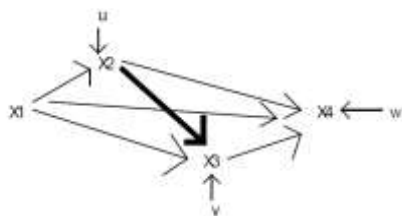
In de figuur hierboven werd het directe effect van x_1 naar x_2 in de verf gezet. De variabele “ u ” is de aliënatiecoëfficiënt, ofwel de proportie onverklaarde variantie in x_2 op basis van x_1 .



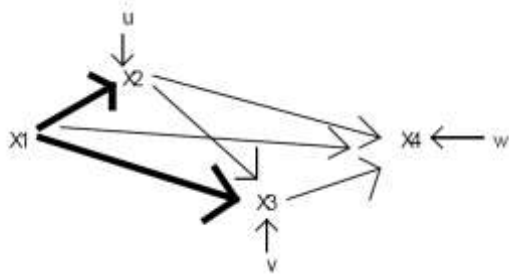
Vervolgens zie je dat er ook een rechtstreeks effect is van x_1 op x_3 . Ook dat is een rechtstreeks effect.



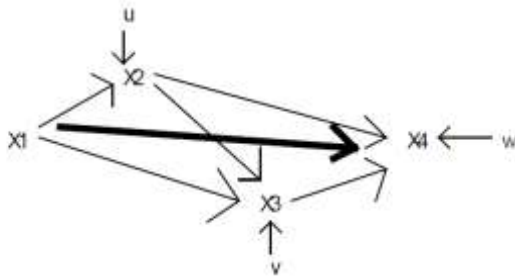
Hierboven zie je dat er niet enkel een rechtstreeks effect van x_1 naar x_3 bestaat, maar dat je ook van x_1 naar x_3 geraakt via x_2 . Dit is een indirect effect.



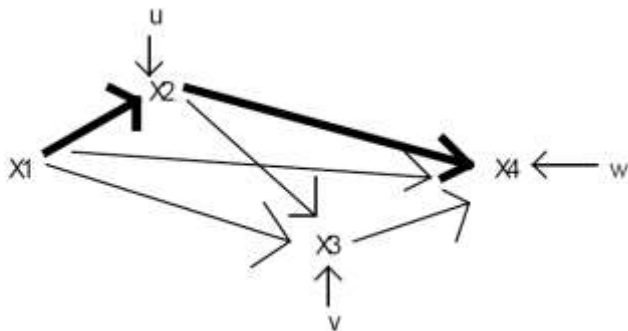
Het effect van x_2 op x_3 is ook een rechtstreeks effect. x_3 kan verklaard worden vanuit de rechtstreekse effecten van x_1 en x_2 . De variabele “ v ” verwijst hier naar de onverklaarde variantie in x_3 , die niet door x_1 en x_2 kan worden verklaard.



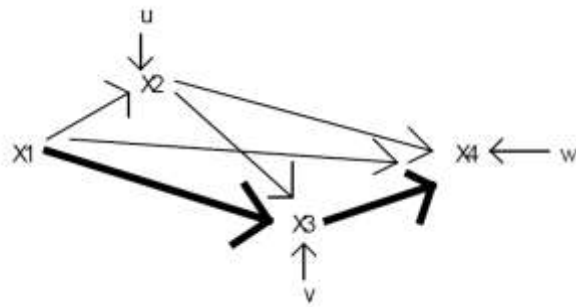
We zien echter ook dat x_1 kan gelden als een partiële gemeenschappelijke oorzaak van de relatie tussen x_2 en x_3 . We zeggen partiële gemeenschappelijke oorzaak, omdat er ook nog een rechtstreeks effect bestaat van x_2 op x_3 . Indien er geen pijl kon getrokken worden van x_2 naar x_3 , dan zou de samenhang tussen x_2 en x_3 volledig te wijten zijn aan het feit dat x_2 en x_3 een gemeenschappelijke oorzaak kennen.



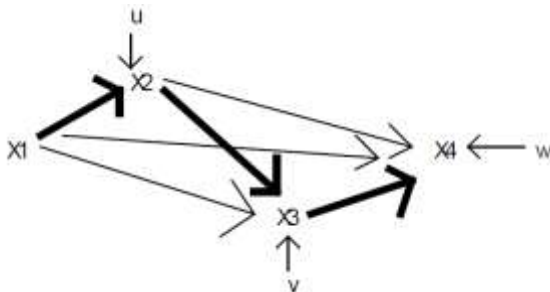
Verder is er ook een effect dat rechtstreeks gaat van x_1 naar x_4 . De variabele “ w ” is de niet verklaarde variantie in x_4 . Dit is de variantie in x_4 die niet kan worden verklaard door de rechtstreekse effecten van x_1 , x_2 en x_3 .



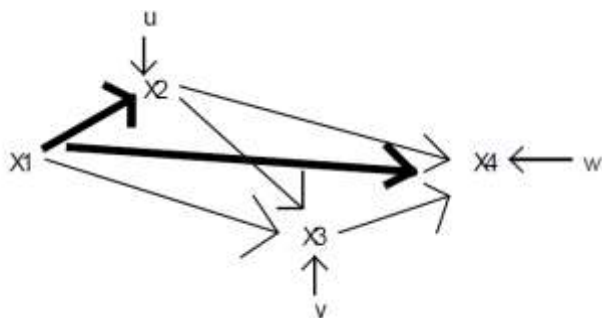
Er is naast het rechtstreeks effect van x_1 op x_4 ook een onrechtstreeks effect van x_1 op x_4 via x_2 .



Er is naast het rechtstreeks effect van x_1 op x_4 ook een onrechtstreeks effect van x_1 op x_4 via x_3 .

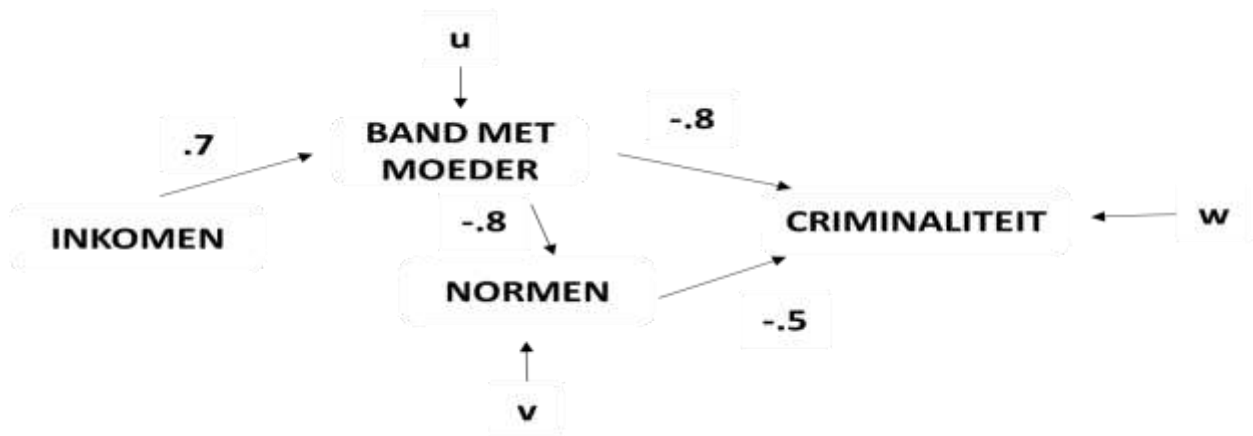


Verder is er nog een manier om van x_1 naar x_4 te geraken: namelijk via x_2 naar x_3 en zo naar x_4 .



En tenslotte is er ook een gemeenschappelijke oorzaak vast te stellen: de relatie tussen x_2 en x_4 heeft x_1 als gemeenschappelijke oorzaak. Het verband is echter niet volledig te wijten aan het fenomeen van de gemeenschappelijke oorzaak, want er is nog het overblijvende rechtstreekse effect van x_2 op x_4 .

7. Een (fictief) rekenvoorbeeld op basis van de gestandaardiseerde padcoëfficiënten

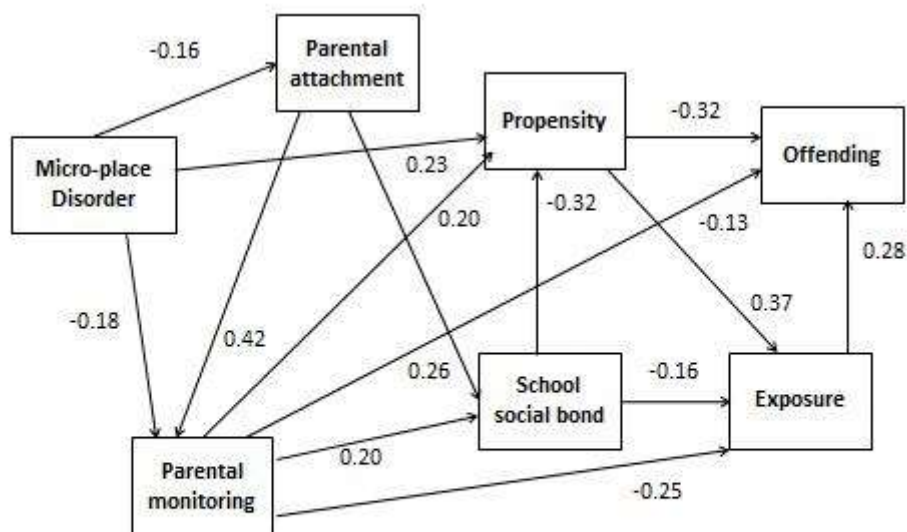


Aan de hand van de methode van Sewell-Wright kan de bivariate correlatie tussen twee variabelen gevonden worden op basis van de gestandaardiseerde padcoëfficiënten.

Correlatiecoëfficiënt	Sewell-Wright methode
R BAND MOEDER; Inkomen = 0.7	Dit is de correlatie tussen band met de moeder en inkomen. <i>Aangezien er enkel een rechtstreeks effect is, is het berekenen van de correlatiecoëfficiënt eenvoudig: deze is gelijk aan de gestandaardiseerde padcoëfficiënt.</i> Dit is het rechtstreeks effect van inkomen op band met de moeder.
R NORMEN; BAND MOEDER = -0.8	Rechtstreeks effect van band met de moeder op normen.
R NORMEN; Inkomen = $-0.8 * 0.7 = -0.56$	Dit is het indirecte effect van inkomen – inkomen beïnvloedt band met de moeder, dat zelf van invloed is op normen.
R CRIMINALITEIT; NORMEN = $-0.5 + -0.8 * -0.8 = 0.14$	De correlatie tussen criminaliteit en normen is gelijk aan <i>het rechtstreekse effect plus de gemeenschappelijke oorzaak</i> . Band met de moeder is een gemeenschappelijke oorzaak van zowel normen als criminaliteit.
R CRIMINALITEIT; BAND MOEDER = $-0.8 + -0.8 * -0.5 = -0.4$	De correlatie tussen criminaliteit en band met moeder is gelijk aan <i>het rechtstreekse effect plus het onrechtstreekse effect</i> .
R CRIMINALITEIT; Inkomen = $-0.8 * 0.7 + -0.5 * -0.8 * 0.7 = -0.28$	Inkomen kent een EERSTE indirect effect op criminaliteit – inkomen is van invloed op band met de moeder, dat op zich van invloed is op criminaliteit.

	Maar er is nog een TWEEDE indirect pad vertrekkend vanuit inkomen. Inkomen is van invloed op band met de moeder, dat op zich van invloed is op normen, dat van invloed is op criminaliteit.
--	--

Een Belgisch criminologisch voorbeeld: jeugddelinquentie vanuit een geïntegreerde theorie



Antwerp Early Adolescence School Survey (2005)

De pad-analyse die we hierboven presenteren komt uit een reël gevoerd onderzoek over de relatie tussen onrechtstreekse oorzaken van regelovertredend gedrag en rechtstreekse oorzaken van regelovertredend gedrag. Het gaat om een analyse van gegevens afkomstig uit een grootschalige studie naar jeugddelinquentie en de rol van de woonomgeving. Het onderzoek eindigde met de vaststelling dat buurten niet zo belangrijk waren als oorspronkelijk werd aangenomen, maar dat de micro-context onrechtstreeks wel relevant is. De micro-context is de plaats zelf (de straat of het straatsegment) en niet de volledige buurt waar men woont. Er werden enkele veronderstellingen gemaakt over de mechanismen die belangrijk zijn bij de verklaring van individuele verschillen in delinquent gedrag en de onderlinge relatie tussen directe en indirecte invloedssferen.

De geïntegreerde theorie (“conditions-controls-exposure theory” van Pauwels en Hardyns) gaat ervan uit dat de micro-context indirect een invloed heeft op regelovertredend gedrag, via

de aantasting van sociale bindingen (attachment) en ouderlijk toezicht (monitoring), maar ook via de invloed op de geneigdheid om regelovertreding als alternatief te zien (propensity). Dat idee lijkt goed stand te houden want je ziet in de figuur, die vergezeld is van gestandaardiseerde pad-coëfficiënten enkel indirecte effecten van de concentratie van overlast in de micro-plaats waar men woont op delinquentie. Er zijn rechtstreekse effecten op attachment, monitoring en propensity. Verder vallen nog een aantal dingen op:

- Er zijn rechtstreekse effecten van parental attachment op parental monitoring en de sociale band met de school. Deze bevindingen kunnen worden geïnterpreteerd vanuit de bindingentheorie. In tegenstelling tot de klassieke analyse van Hirschi, leren we met deze padanalyse veel meer, met name dat er toch verschillende onderlinge relaties zijn tussen de elementen van de sociale band.
- Er zijn een aantal rechtstreekse effecten van monitoring: een rechtstreeks effect op de sociale band met de school, een rechtstreeks effect op propensity en een rechtsreeks effect op blootstelling aan criminogene situaties (de ongestructureerde leefstijl).
- Er zijn slechts twee directe effecten van de sociale band met de school en dat is een effect op propensity en exposure.
- Propensity heeft het sterkste rechtstreekse effect op zelfgerapporteerde criminaliteit en heeft ook een sterk effect op de mate van blootstelling aan criminogene settings.
- Blootstelling aan criminogene settings heeft een rechtsreeks effect op regelovertredend gedrag.

Pad-analyses leveren dus een schat aan informatie op voor zowel de evaluatie van criminologische theorieën als voor het beleid.

8. Leerdoelen

In dit hoofdstuk werd een aanzet gegeven tot het bestuderen van meer complexe relaties tussen variabelen. De techniek die daarvoor gebruikt wordt is de padanalyse. Deze techniek kent zijn oorsprong in de biologie en het was de bioloog Sewell-Wright die een methode bedacht om op basis van de pad-coëfficiënten, die niet meer zijn dan gestandaardiseerde richtingscoëfficiënten, terug te keren naar de correlatiecoëfficiënten. We verwachten van studenten dat ze, als hen een pad-model wordt getoond, de rechtstreekse, onrechtstreekse en totale effecten kunnen berekenen. We verwachten eveneens dat uit een pad-diagram de bivariate correlaties kunnen berekend worden tussen een aantal gevraagde variabelen. Maar

bovenal is het belangrijk dergelijke analyses te kunnen begrijpen wanneer je ze tegenkomt in de kwantitatieve criminologische literatuur.

