

In de voorgaande twee hoofdstukken hebben we gekeken naar het verband tussen twee nominale of ordinale variabelen, oftewel naar bivariate analyses. In dit hoofdstuk staan multivariate analyses met nominale en ordinale variabelen centraal, dat wil zeggen dat we meer dan twee variabelen in de analyse zullen betrekken.

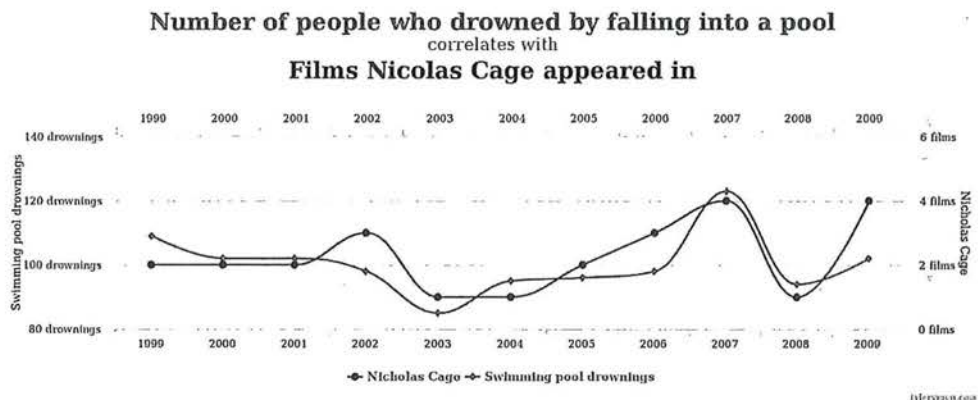
Waarom zou je meer dan twee variabelen in de analyse betrekken? Dat doe je omdat als je een samenhang tussen twee variabelen hebt gevonden, je het inzicht in die samenhang kunt vergroten door naar de rol van een derde variabele te kijken. Het is belangrijk om je te realiseren dat de samenhang tussen twee variabelen niets zegt over *causaliteit*, over oorzaak en gevolg. Om er zeker van te zijn dat veranderingen in de ene variabele (x) de oorzaak zijn van veranderingen in de andere variabele (y) zou je moeten uitsluiten dat een derde variabele verantwoordelijk is voor de veranderingen in zowel x als y . Dat is meestal niet zo eenvoudig. Als er sprake is van een causaal verband, moeten veranderingen in x ook altijd voorafgaan aan veranderingen in y . Zelfs als we dat hebben kunnen vaststellen, is het echter nog altijd mogelijk dat een andere variabele eerst de veranderingen in x heeft veroorzaakt en pas daarna de veranderingen in y . Zelfs als we op theoretische gronden een invloed van x op y verwachten en we daarom kiezen voor een asymmetrische associatiemaat, weten we nog niet zeker dat er een *causaal* verband is. Dat laatste kunnen we alleen aannemelijk maken (dus niet: vaststellen!) door allerlei andere verklaringen uit te sluiten. Als we aan de hand van een Somers' d concluderen dat als peuters meer educatieve spellen op een tablet spelen, zij een hogere woordenschat hebben, dan kunnen we niet zeggen dat deze woordenschat door het spelen van educatieve spellen wordt *veroorzaakt*. We moeten voorzichtig zijn met deze conclusie vanwege twee redenen. Ten eerste kan het zo zijn dat er een derde variabele in het spel is (al dan niet gemeten in het onderzoek) die de resultaten beïnvloedt. Ten tweede geven associatiematen geen uitsluiting over de causaliteit van een verband. We hebben dan als onderzoeker wel bedacht dat in bovenstaand voorbeeld het spelen op de tablet de onafhankelijke variabele was en de woordenschat de afhankelijke variabele, en het is ook intuïtief aannemelijk dat het spelen invloed heeft op de woordenschat, maar statistisch gezien is er geen bewijs voor een causale relatie.

Bij enquêtes en inhoudsanalyses kun je variabelen opnemen waarvoor je kunt controleren of ze de relatie tussen x en y beïnvloeden, maar het is niet mogelijk om alle denkbare van invloed zijnde factoren te meten. Bij een experiment

wordt geprobeerd om de situatie dusdanig te controleren dat een poging tot een uitspraak over causaliteit gedaan kan worden. Dit doen we door bijvoorbeeld een *experiment* in een laboratorium af te nemen waardoor je zeker weet dat *x* voorafgaat aan *y* en storende invloeden beperkt en/of onder controle worden gehouden. Bovendien kunnen we door een controlegroep toe te voegen de uitkomsten van die groep vergelijken met de experimentele groep.

7.1 Interpretatie

Wanneer we een samenhang tussen twee variabelen vinden, betekent dat dus niet dat er een oorzakelijk verband tussen de twee variabelen bestaat. Op internet vinden we verschillende (grappige) voorbeelden van zeer sterke verbanden van variabelen die in werkelijkheid niets met elkaar te maken hebben.¹ Zo zien we dat er een zeer sterk positief verband bestaat tussen het aantal films waar Nicholas Cage in speelt en het aantal verdrinkingen in een zwembad.



Figuur 7.1 Correlatie tussen aantal films met Nicholas Cage en aantal verdrinkingen in een zwembad

In werkelijkheid is het vrij onwaarschijnlijk dat het aantal films met Nicholas Cage de oorzaak is van het aantal verdrinkingen in een zwembad (of andersom: dat het aantal verdrinkingen in een zwembad van invloed is op het aantal films waarin Nicholas Cage speelt). Er is dan sprake van een *schijnsamenhang*. Zo'n schijnsamenhang, die je ook wel een *spurieuze samenhang* noemt, wordt meestal veroorzaakt door andere variabele(n) of is soms gewoon toeval. Het is waarschijnlijker dat er een andere oorzaak is voor het aantal verdrinkingen in een zwembad dan het aantal films met Nicholas Cage, bijvoorbeeld de temperatuur in dat jaar. Omdat het in 2007 warmer was dan in 2008, waren er dat jaar meer zwempartijen en daarom meer verdrinkingen. Dat er in 2007 ook meer films met de acteur waren, is dan toeval maar geen oorzaak. Bij controle door de variabele temperatuur blijkt de eerder gevonden samenhang tussen verdrinkingen in een zwembad en het aantal films met Cage spurieus te zijn: een schijnverband.

Ook is het mogelijk dat je bij controle voor een derde variabele constateert dat het eerder gevonden verband voor sommige groepen sterker is dan voor andere groepen. In dat geval is er geen sprake van een schijnsamenhang. De samenhang is er echt, maar je kunt de eerder gevonden samenhang tussen twee variabelen wel verder *specificeren* door gebruik te maken van een derde variabele. Als je in een onderzoek onder werknemers in een bedrijf een samenhang vindt tussen de hoogte van het inkomen en het aantal jaar dat iemand in dienst is, zou je kunnen concluderen dat het inkomen toeneemt naarmate iemand langer in dienst is. Als je die samenhang controleert voor de variabele geslacht, is het mogelijk dat de samenhang tussen inkomen en dienstjaren bij mannen veel sterker is dan bij vrouwen. Je hebt dan het eerder gevonden verband gespecificeerd en een *interactie* met de variabele geslacht aangetoond. Je conclusie is dan dat langer in dienst zijn bij mannen tot een grotere inkomenstoename leidt dan bij vrouwen. Er is ook nog de mogelijkheid dat je een verband tussen twee variabelen had verwacht dat niet of nauwelijks aanwezig blijkt te zijn. Na controle voor een derde variabele kan blijken dat er voor afzonderlijke groepen wel een verband is. Als voor die groepen de verbanden tussen de twee variabelen tegengesteld zijn, wordt het verband voor alle groepen tezamen *versluierd*.

7.1.1 Spurious samenhang

Als de samenhang tussen twee variabelen hoog is maar het onwaarschijnlijk is dat de twee variabelen elkaar beïnvloeden en je ook geen goede verklaring hebt voor de samenhang, is het verstandig te zoeken naar een derde variabele die de samenhang wel kan verklaren. Laten we een fictief onderzoek naar ijsconsumptie en het aantal verdrinkingen als voorbeeld nemen. Er blijkt een sterke samenhang te bestaan tussen de hoeveelheid ijs die per dag wordt gegeten en het aantal verdrinkingen per dag. De onderzoekseenheden zijn hier 'dagen' (zie tabel 7.1).

Tabel 7.1 Kruistabel van ijsconsumptie en verdrinkingen per dag (n = 200)

Verdrinkingen \ IJs	0 weinig	1 veel	
1 veel	18	82	100
0 weinig	82	18	100
	100	100	200 (dagen)

We hebben geen idee of ijs eten het aantal verdrinkingen zou beïnvloeden of andersom, en er is sprake van ordinale variabelen, want 'veel' is meer dan 'weinig'. Wanneer we hier gamma zouden berekenen, zou deze uitkomen op 0,91. Onze conclusie zou zijn: er is een zeer sterke, positieve samenhang tussen het aantal ijsjes dat op een dag wordt gegeten en het aantal verdrinkingen dat op een dag plaatsvindt. Op dagen dat er veel ijs wordt gegeten zijn er ook veel verdrinkingen, en als er weinig ijs wordt gegeten zijn er weinig verdrinkingen.

Dit is natuurlijk onzinnige informatie. Waarom zou er zo'n sterk verband zijn tussen deze twee variabelen? Hoe is dat verband te verklaren? Ijsconsumptie kan toch niet leiden tot een toename in het aantal verdrinkingen en verdrinkingen hebben ook geen hogere ijsconsumptie tot gevolg. Het is waarschijnlijk dat er een derde variabele in het spel is die invloed heeft op beide variabelen. We voegen daarom een derde variabele toe: de gemiddelde temperatuur per dag. Voor het gemak maken we onderscheid tussen koude dagen en warme dagen. We zouden nu naar de ijsconsumptie en het aantal verdrinkingen kunnen kijken op warme dagen en op de koude dagen afzonderlijk. We krijgen dan twee verschillende tabellen (tabel 7.2), die samen de (oorspronkelijke) tabel vormen. We noemen deze twee tabellen van zo'n tabelsplitsing *deeltabellen* of *partiële tabellen*. Binnen de deeltabellen wordt de derde variabele (temperatuur) constant gehouden. We hebben dus een tabel voor warme dagen en een tabel voor koude dagen, die samen de eerdere tabel (tabel 7.1) vormen.

Tabel 7.2 Kruistabellen van ijsconsumptie en verdrinkingen op warme en koude dagen

IJs Verdr.	0	1	
1	9	81	90
0	1	9	10
	10	90	100
warme dagen			

IJs Verdr.	0	1	
1	9	1	10
0	81	9	90
	90	10	100
koude dagen			

IJs Verdr.	0	1	
1	18	82	100
0	82	18	100
	100	100	200

Wanneer we voor de partiële tabellen gamma's uitrekenen, komen deze beide uit op 0. Er is dus op warme dagen geen verband tussen de ijsconsumptie en het aantal verdrinkingen, en op koude dagen evenmin. Dit betekent dat de eerder gevonden samenhang een schijnsamenhang is, oftewel: de eerder gevonden samenhang is spurieus. De derde variabele (temperatuur) heeft zowel invloed op de ijsconsumpties als op het aantal verdrinkingen.

Dit kunnen we controleren door te kijken naar de kruistabellen van ijsconsumptie naar temperatuur en verdrinkingen naar temperatuur (tabel 7.3). Voor beide kruistabellen geldt dat er een sterk verband is ($\text{gamma} = 0,99$). Hieruit blijkt dat de temperatuur op een dag zowel samenhangt met de ijsconsumptie als met het aantal verdrinkingen.

Als blijkt dat een derde variabele een eerder gevonden verband verklaart, zijn er twee mogelijkheden. Er is sprake van een *antecedente variabele* of van een *intervenierende variabele* (ook wel *mediërende variabele*).

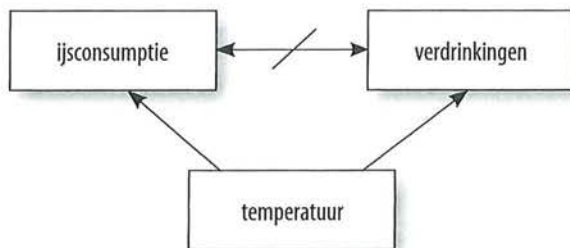
Je spreekt van een antecedente variabele als de derde variabele in de tijd voorafgaat aan de andere twee variabelen. In het voorbeeld ijsconsumptie – verdrinkingen is temperatuur een antecedente variabele, een variabele die voorafgaat aan de ijsconsumptie en aan het aantal verdrinkingen. Wanneer het warmer is,

zwemmen er meer mensen, en wanneer er meer mensen zwemmen, is er meer kans op een verdrinking. Hetzelfde geldt voor de ijsconsumptie: wanneer het warmer is, zullen mensen eerder een ijsje eten dan wanneer het koud is. Deze verbanden zijn verklaarbaar. Dat was bij het eerder gevonden verband tussen ijs eten en verdrinkingen niet het geval (zie figuur 7.2).

Tabel 7.3 Kruistabellen van verdrinkingen en ijsconsumptie naar temperatuur ($n = 200$)

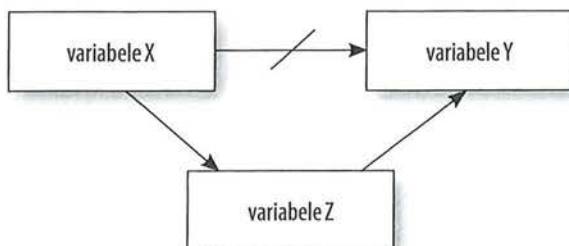
Verdrinkingen \ Temp.	0 koud	1 warm	
1 veel	10	90	100
0 weinig	90	10	100
	100	100	200 (dagen)

IJs \ Temp.	0 koud	1 warm	
1 veel	10	90	100
0 weinig	90	10	100
	100	100	200 (dagen)



Figuur 7.2 Spurieuus verband door antecedente variabele

Het kan ook voorkomen dat de derde variabele beïnvloed is door een van de twee oorspronkelijke variabelen en vervolgens de andere variabele beïnvloedt. Het is dan geen antecedente, maar een interveniërende variabele. De derde variabele Z (die we hier een *mediator* noemen) gaat nu niet vooraf aan X en Y , maar Z wordt beïnvloed door X en Z beïnvloedt vervolgens Y , waardoor er een schijnverband tussen X en Y ontstaat (zie figuur 7.3). Dit wordt ook *mediatie* genoemd.



Figuur 7.3 Spurieuus verband met interveniërende variabele

Je vindt bijvoorbeeld een sterke samenhang tussen sekse en het maandelijks inkomen. Mannen hebben een hoger inkomen dan vrouwen. Het is mogelijk dat dit verband verdwijnt als je controleert voor parttime of fulltime werken. Vrouwen werken vaker parttime en hebben daardoor een lager inkomen dan mannen. Parttime of fulltime werken is in dit voorbeeld de interveniërende variabele: sekse (X) → parttime of fulltime werken (Z) → inkomen (Y).

Of een variabele antecedent of interveniërend oftewel mediërend is, moet je meestal op basis van gezond verstand bepalen. Een mediator (zoals in figuur 7.3) moet bijvoorbeeld altijd zowel onafhankelijk als afhankelijk kunnen zijn, omdat deze zowel beïnvloed wordt (door X) als invloed uitoefent (op Y).

Hoe ziet tabelsplitsing er nu uit in SPSS? Eerst bekijken we de kruistabel en de samenhang (gamma) tussen ijsconsumptie en het aantal verdrinkingen zonder de variabele temperatuur (tabel 7.4). Deze tabel is gelijk aan tabel 7.1. Gamma is inderdaad 0,91. Er is een zeer sterke, positieve samenhang tussen ijsconsumptie en het aantal verdrinkingen.

Tabel 7.4 Kruistabel van ijsconsumptie en verdrinkingen (SPSS-output)

Verdrinkingen * ijs Crosstabulation

			ijs		Total
			0 weinig	1 veel	
verdrinkingen	1 veel	Count	18	82	100
		% within ijs	18,0%	82,0%	50,0%
	0 weinig	Count	82	18	100
		% within ijs	82,0%	18,0%	50,0%
Total		Count	100	100	200
		% within ijs	100,0%	100,0%	100,0%

Symmetric Measures

		Value	Asymptotic Standardized Error	Approximate T	Approximate Significance
Ordinal by Ordinal	Gamma	,908	,032	11,779	,000
N of Valid Cases		200			

Vervolgens splitsen we deze tabellen naar deeltabellen. In deze tabellen is de temperatuur constant gehouden. Er is apart naar de warme en koude dagen gekeken. In de onderste tabel is ook de oorspronkelijke, bivariate tabel te zien. In het onderste deel van tabel 7.5 (*Symmetric Measures*) is te zien dat gamma zowel op warme dagen als op koude dagen 0 is. Er is geen enkel verband tussen ijsconsumptie en het aantal verdrinkingen als de temperatuur constant wordt gehouden. De eerder gevonden samenhang is spurieus, en wordt veroorzaakt door de temperatuur (antecedent).

Tabel 7.5 Tabelsplitsing ijs – verdrinkingen naar warme en koude dagen (SPSS-output)

Verdrinkingen * ijs * temperatuur Crosstabulation

temperatuur				ijs		Total
				0 weing	1 veel	
1 warme dag	verdrinkingen	1 veel	Count	9	81	90
			% within ijs	90,0%	90,0%	90,0%
		0 weing	Count	1	9	10
			% within ijs	10,0%	10,0%	10,0%
	Total		Count	10	90	100
			% within ijs	100,0%	100,0%	100,0%
0 koude dag	verdrinkingen	1 veel	Count	9	1	10
			% within ijs	10,0%	10,0%	10,0%
		0 weing	Count	81	9	90
			% within ijs	90,0%	90,0%	90,0%
	Total		Count	90	10	100
			% within ijs	100,0%	100,0%	100,0%
Total	verdrinkingen	1 veel	Count	18	82	100
			% within ijs	18,0%	82,0%	50,0%
		0 weing	Count	82	18	100
			% within ijs	82,0%	18,0%	50,0%
	Total		Count	100	100	200
			% within ijs	100,0%	100,0%	100,0%

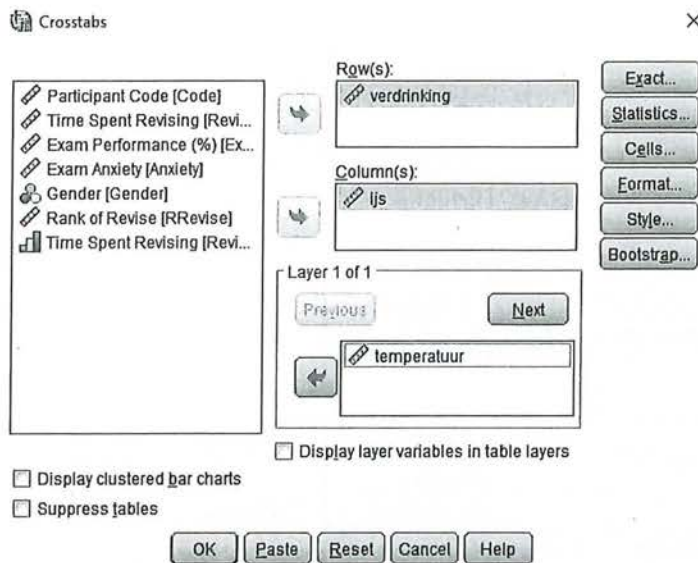
Symmetric Measures

temperatuur				Value	Asymptotic Standardized Error	Approximate T	Approximate Significance
0 koude dag	Ordinal by Ordinal	Gamma	Zero-Order	,000	,556	,000	1,000
	N of Valid Cases			100			
1 warme dag	Ordinal by Ordinal	Gamma	Zero-Order	,000	,556	,000	1,000
	N of Valid Cases			100			
Total	Ordinal by Ordinal	Gamma	Zero-Order	,908	,032	11,779	,000
			First Order Partial	,000			
	N of Valid Cases			200			



Tabelsplitsing gebruik je bij kruistabellen. De eerste stappen van tabelsplitsing zijn dan ook dezelfde als bij het maken van een kruistabel. Via *Analyze* → *Descriptive Statistics* → *Crosstabs* kies je eerst de variabelen voor de kolommen en de rijen. Vervolgens plaats je in de *Layer* de derde variabele.

Via *Statistics* kun je nu kiezen voor een bijpassende associatiemaat.



Figuur A Crosstabs-venster: toevoegen van Layer

Kader 7.1

Uiteraard kun je in principe alle associatiematen bij tabelsplitsing gebruiken. Welke associatiemaat je kiest, is afhankelijk van het meetniveau van de variabelen waartussen het bivariate verband wordt berekend. De variabele waar je op 'splitst' kan een nominaal meetniveau hebben, maar dat heeft geen consequenties voor de berekening van de partiële verbanden.

7.1.2 Specificatie

Bij tabelsplitsing is het ook mogelijk dat je na controle voor een derde variabele niet constateert dat het oorspronkelijke verband spurieus was, maar dat er wel iets anders aan de hand is.

Stel dat je een onderzoek hebt uitgevoerd onder tieners. Daaruit blijkt dat het bezoeken van online sociale media slecht is voor het zelfbeeld van tieners. Je bent daarom benieuwd of het zien van bepaalde foto's op deze sociale media (voornamelijk zien van selfies, vakantiefoto's of foto's van sportieve prestaties) verband houdt met het zelfbeeld. De variabele 'soort foto' is nominaal en onafhankelijk, en daarom is lambda of Goodman en Kruskalls tau de meest geschikte associatiemaat.

Tabel 7.6 Krustabel van zelfbeeld naar soort foto (SPSS-output)

Zelfbeeld * Soort_foto Crosstabulation

			Soort_foto			Total
			1 selfie	2 vakantie-foto	3 sportieve prestatie	
Zelfbeeld	1 laag	Count	5	4	1	10
		% within Soort_foto	50,0%	33,3%	12,5%	33,3%
	2 gemiddeld	Count	2	6	3	11
		% within Soort_foto	20,0%	50,0%	37,5%	36,7%
	3 hoog	Count	3	2	4	9
		% within Soort_foto	30,0%	16,7%	50,0%	30,0%
Total		Count	10	12	8	30
		% within Soort_foto	100,0%	100,0%	100,0%	100,0%

Directional Measures

			Value	Asymptotic Standardized Error	Approximate T	Approximate Significance
Nominal by Nominal	Lambda	Symmetric	,189	,170	1,039	,299
		Zelfbeeld Dependent	,211	,175	1,090	,276
		SNS Dependent	,167	,196	,782	,434
Goodman and Kruskal tau		Zelfbeeld Dependent	,083	,067		,308
		SNS Dependent	,082	,067		,316

Er is een zwak verband tussen het soort foto en het zelfbeeld bij tieners ($\lambda = 0,21$, $n = 30$).² De helft van de tieners die voornamelijk selfies zien heeft een laag zelfbeeld, en de helft van de tieners die voornamelijk sportieve prestaties op foto's zien, heeft een hoog zelfbeeld.

Maar zou het verband ook aan andere factoren kunnen liggen? We vermoeden dat dit verband voor meisjes anders kan zijn dan voor jongens. We splitsen daarom de tabel op naar de variabele sekse:

Uit tabel 7.7 blijkt dat dit verband alleen bestaat onder meisjes ($\lambda = 0,43$), maar niet onder jongens ($\lambda = 0,00$). We zien aan de tabel dat er geen meisjes zijn met een 'hoog zelfbeeld' en dat er geen jongens zijn met een 'gemiddeld zelfbeeld'. Waar 71,4% van de meisjes die voornamelijk selfies zien een laag zelfbeeld hebben, heeft 0% van de jongens dat. De meeste jongens hebben een hoog zelfbeeld, maar daarbij maakt het niet zoveel uit welk soort foto's zij voornamelijk op sociale media zien. Het oorspronkelijke verband hebben we nu *gespecificeerd*. We zeggen ook wel dat we een interactie-effect hebben gevonden tussen soort foto en geslacht op het zelfbeeld: het effect van soort foto op zelfbeeld, is voor jongens anders dan voor meisjes.

Tabel 7.7 Tabelsplitsing soort foto – zelfbeeld naar geslacht (SPSS-output)

Zelfbeeld * Soort_foto * Sekse Crosstabulation

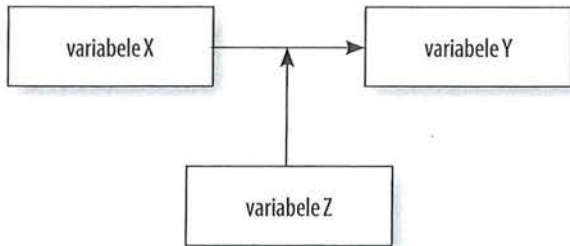
Sekse				Soort_foto			Total
				1 selfie	2 vakantie-foto	3 sportieve prestatie	
1 meisje	Zelf-beeld	1 laag	Count	5	2	0	7
			% within Soort_foto	71,4%	25,0%	0,0%	38,9%
	2 gemiddeld	Count	2	6	3	11	
		% within Soort_foto	28,6%	75,0%	100,0%	61,1%	
Total		Count	7	8	3	18	
		% within Soort_foto	100,0%	100,0%	100,0%	100,0%	
2 jongen	Zelf-beeld	1 laag	Count	0	2	1	3
			% within Soort_foto	0,0%	50,0%	20,0%	25,0%
	3 hoog	Count	3	2	4	9	
		% within Soort_foto	100,0%	50,0%	80,0%	75,0%	
Total		Count	3	4	5	12	
		% within Soort_foto	100,0%	100,0%	100,0%	100,0%	
Total	Zelf-beeld	1 laag	Count	5	4	1	10
			% within Soort_foto	50,0%	33,3%	12,5%	33,3%
	2 gemiddeld	Count	2	6	3	11	
		% within Soort_foto	20,0%	50,0%	37,5%	36,7%	
3 hoog	Count	3	2	4	9		
	% within Soort_foto	30,0%	16,7%	50,0%	30,0%		
Total		Count	10	12	8	30	
		% within Soort_foto	100,0%	100,0%	100,0%	100,0%	

Directional Measures

Sekse				Value	Asymptotic Standardized Error	Approximate T	Approximate Significance
1 meisje	Nominal by Nominal	Lamba	Symmetric	,353	,236	1,279	,201
			Zelfbeeld Dependent	,429	,286	1,177	,239
			Soort_foto Dependent	,300	,221	1,177	,239
	Goodman and Kruskal tau	Zelfbeeld Dependent	,315	,190		,069	
		Soort_foto Dependent	,162	,123		,064	
2 jongen	Nominal by Nominal	Lamba	Symmetric	,100	,318	,303	,762
			Zelfbeeld Dependent	,000	,667	,000	1,000
			Soort_foto Dependent	,143	,229	,586	,558
	Goodman and Kruskal tau	Zelfbeeld Dependent	,200	,195		,333	
		Soort_foto Dependent	,092	,101		,363	
Total	Nominal by Nominal	Lamba	Symmetric	,189	,170	1,039	,299
			Zelfbeeld Dependent	,211	,175	1,090	,276
			Soort_foto Dependent	,167	,196	,782	,434
	Goodman and Kruskal tau	Zelfbeeld Dependent	,083	,067		,308	
		Soort_foto Dependent	,082	,067		,316	

Bij specificatie (wat we ook wel *moderatie* of *interactie* noemen) hanteren we de grove richtlijn van een verschil van minimaal 0,1 tussen de associaties van de verschillende groepen, en ten opzichte van de bivariate associatie (de oorspronkelijk sterkte van de associatiemaat voor de tabelsplitsing).

Wanneer we interactie grafisch zouden weergeven in een conceptueel model, zou dat eruit zien als in figuur 7.4:



Figuur 7.4 Conceptueel model met moderatie / specificatie / interactie

Variabele Z noemen we hier de *moderator*, die het effect van X op Y beïnvloedt.

7.1.3 Versluiering

Tot slot kan tabelsplitsing ons ook helpen als we tegen onze verwachting in *geen* verband tussen twee variabelen vinden. Laten we eens kijken naar tabelsplitsing bij een vierkante tabel (2×2) op ordinaal niveau. We kijken naar het verband tussen het opleidingsniveau en algemene kennis (voor dit voorbeeld beide in twee klassen opgedeeld), en hebben de hypothese dat mensen met een hoger opleidingsniveau meer algemene kennis hebben. Een geschikte associatiemaat is dan Somers' d, want de relatie is asymmetrisch. Opleiding is hier de onafhankelijke variabele en kennis de afhankelijke variabele. Wanneer we door SPSS deze maat laten berekenen, levert dit een Somers' d op van $-0,02$. Er is dus geen verband tussen opleiding en kennis. We voegen nu een derde variabele toe, namelijk televisiekijken (weinig/veel), en krijgen de resultaten die in tabel 7.8 staan.

We zien in tabel 7.8 dat het verband voor mensen die weinig televisiekijken anders is dan voor mensen die veel televisiekijken. Televisiekijken is hier dus een moderator. Waar we in eerste instantie geen verband vonden ($d_{yx} = -0,02$), vinden we nu bij beide groepen (mensen die weinig televisiekijken en mensen die veel televisiekijken) wél een verband. Bij weinig televisiekijken is het zo dat er een redelijke negatieve samenhang is tussen opleiding en kennis ($d_{yx} = -0,39$): hoe hoger de opleiding hoe minder de kennis is bij mensen die weinig televisiekijken. Bij mensen die veel televisiekijken is er juist een redelijke positieve samenhang ($d_{yx} = 0,34$): een hogere opleiding leidt bij mensen die veel televisiekijken tot meer kennis. Hoogopgeleiden steken meer op van veel televisiekijken dan laagopgeleiden.

Tabel 7.8 Tabelsplitsing: kruistabellen van kennis naar opleiding, onder constantheouding van tv-kijken (SPSS-output)

kennis * opleiding * tvkijken Crosstabulation

tvkijken				opleiding		Total
				1 laag	2 hoog	
2 veel	kennis	2 veel	Count	36	35	71
			% within opleiding	53,7%	87,5%	66,4%
	1 weinig	2 veel	Count	31	5	36
			% within opleiding	46,3%	12,5%	33,6%
	Total	2 veel	Count	67	40	107
			% within opleiding	100,0%	100,0%	100,0%
1 weinig	kennis	2 veel	Count	33	10	43
			% within opleiding	63,5%	25,0%	46,7%
	1 weinig	2 veel	Count	19	30	49
			% within opleiding	36,5%	75,0%	53,3%
	Total	2 veel	Count	52	40	92
			% within opleiding	100,0%	100,0%	100,0%
Total	kennis	2 veel	Count	69	45	114
			% within opleiding	58,0%	56,3%	57,3%
	1 weinig	2 veel	Count	50	35	85
			% within opleiding	42,0%	43,8%	42,7%
	Total	2 veel	Count	119	80	199
			% within opleiding	100,0%	100,0%	100,0%

Directional Measures

tvkijken				Value	Asymptotic Standardized Error	Approximate T	Approximate Significance
2 veel	Ordinal	Somers' d	Symmetric	,346	,080	4,115	,000
			kennis Dependent	,338	,080	4,115	,000
			opleiding Dependent	,354	,083	4,115	,000
1 weinig	Ordinal	Somers' d	Symmetric	-,382	,095	-,3,997	,000
			kennis Dependent	-,385	,096	-,3,997	,000
			opleiding Dependent	-,380	,095	-,3,997	,000
Total	Ordinal	Somers' d	Symmetric	-,017	,071	-,242	,809
			kennis Dependent	-,017	,072	-,242	,809
			opleiding Dependent	-,017	,070	-,242	,809

We spreken bij deze vorm van moderatie van een *onderdrukt of versluierd* verband.

7.2 Samenvatting

Wanneer we een verband vinden bij een bivariate analyse, kunnen we niets over de causaliteit zeggen. Het kan namelijk zijn dat er een andere factor in het spel is (die je al dan niet gemeten hebt). Wanneer je deze derde variabele constant houdt, kun je in ieder geval meer genuanceerde informatie geven. Wanneer een derde variabele wordt toegevoegd, noemen we dat een multivariate analyse. In dit hoofdstuk zijn de volgende situaties besproken die zich dan kunnen voordoen:

- Bij controle voor een derde variabele verandert er niets. Dan geldt de geconstateerde samenhang dus ook voor elke waarde van de derde variabele.
- Het verband tussen twee variabelen is een schijnsamenhang die wordt verklaard door een derde variabele. Wanneer een verband spurieus is, is er een derde variabele die de samenhang verklaart. Bij controle voor die derde variabele verdwijnt het eerder gevonden verband. Er is dan sprake van een antecedente variabele of mediatie door een interveniërende variabele. Om te bepalen of het een antecedente of een interveniërende variabele is, moet je zoeken naar de logica in de verbanden tussen de variabelen.
- Een derde variabele specificeert een eerder gevonden verband. Als de samenhang voor een bepaalde groep sterker is dan voor een andere groep, spreken we van specificatie, interactie of moderatie. Door middel van tabelsplitsing, waarbij je de derde variabele constant houdt, kun je bepalen of en hoe een derde variabele de samenhang beïnvloedt. Er moet dan een minimaal verschil van 0,1 zijn tussen de partiële associaties, en met de bivariate associatie.
- Een derde variabele kan het verband tussen twee variabelen onderdrukken/versluieren, waardoor een verwacht verband tussen twee variabelen niet wordt gevonden. Na tabelsplitsing blijkt dan dat voor de afzonderlijke waarden van de derde variabele de verbanden tegengesteld zijn.



Ga naar de website om de opdrachten bij dit hoofdstuk te maken.

Noten

- 1 Zie onder andere de website www.tylervigen.com/spurious-correlations.
- 2 We hadden hier ook voor tau kunnen kiezen.

