

HOOFDSTUK VIII

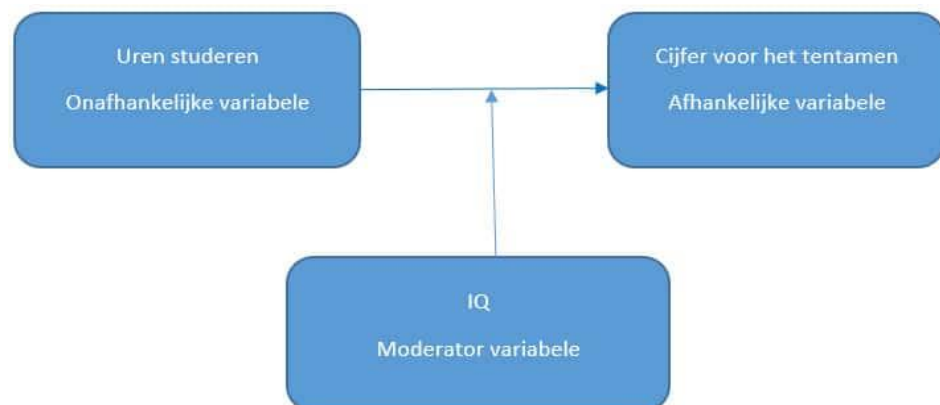
PARTIELE CORRELATIE

1. DOELSTELLINGEN

Op het eind van dit hoofdstuk hebben studenten inzicht in de beperkingen van de bivariate statistiek en is het principe van de statistische controle goed begrepen. Inzicht in de partiële correlatie is nodig aangezien een verband tussen twee variabelen kan verzwakken, wegvallen of versterken na controle voor een derde variabele. Studenten zijn zelf in staat om een partiële correlatie tussen twee variabelen onder statistische controle van een derde variabele te berekenen.

2. TE ONTHOUDEN KERNBEGRIPPEN

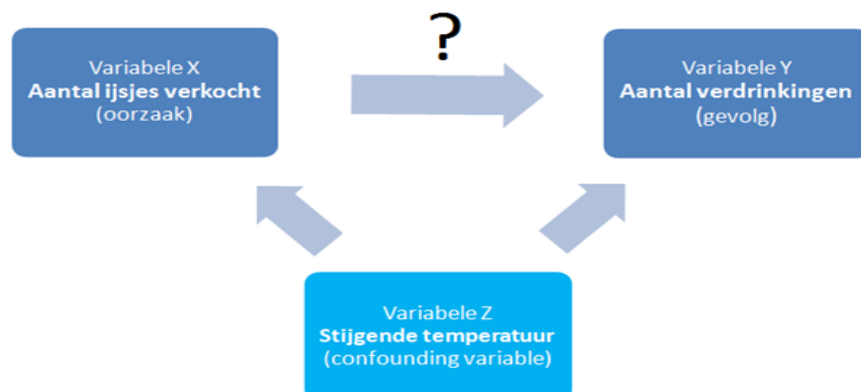
Interactie	Er is sprake van interactie als het effect van een onafhankelijke variabele X op een afhankelijke variabele Y afhangt van de waarde van een andere onafhankelijke variabele Z
------------	---



Opgelet! Niet verwarren met 'controleren voor'
In dat geval wordt er vanuit gegaan dat het effect van x op y min of meer hetzelfde blijft als de controlerende variabele (z) andere waarden aanneemt

Meervoudige regressie	Statistische techniek waarbij een lineair verband wordt berekend tussen 1 afhankelijke variabele en minimum 2 onafhankelijke variabelen
Multivariate analyse	In dergelijke analyses wordt nagegaan of bivariate relaties ook blijven bestaan als rekening wordt gehouden met andere variabelen. Men spreekt in dat verband van 'controleren voor'
Partiële correlatie	De correlatie tussen twee variabelen ontstaat van de invloed van extra variabelen. Een soort netto correlatie
Schijnverband	Er kan sprake zijn van een schijnverband als de oorspronkelijke samenhang tussen een x-variabele en een y-variabele verdwijnt of van teken verandert na rekening te houden met een andere variabele (z)

Spurieuze relatie Schijnrelatie
 Een bivariaat verband is het gevolg van een gemeenschappelijke oorzaak



Statistische controle Elke variabele die gecorreleerd is met de onafhankelijke variabele en mede bepalend kan zijn voor de score op de afhankelijke variabele is een storende variabele. Controleren voor een storende variabele betekent dat deze constant wordt gehouden en enkel wordt gekeken naar de variatie in de afhankelijke variabele en in de onafhankelijke variabele die samenhangt met eenzelfde niveau in de controlevariabele

Suppressie-effect Als de samenhang tussen x-variabele en y-variabele sterker wordt na controle voor één of meer andere variabele, dan is er sprake van suppressie of onderdrukking

3. STATISTISCHE SYMBOLEN EN FORMULES

partiële correlatie: verband tussen twee variabelen X en Y gecontroleerd voor een derde variabele Z

$$r_{XY \cdot Z} = \frac{r_{XY} - (r_{XZ})(r_{YZ})}{\sqrt{1 - r_{XZ}^2} \times \sqrt{1 - r_{YZ}^2}}$$

OEFENINGEN

1. Hieronder vind je de SPSS output van de bivariate lineaire regressie waarin het aantal arrestaties na gevangenisstraf gedefinieerd wordt in functie van het aantal jaren gevangenisstraf in een steekproef van 20 ex-gedetineerden. (hypothetische data).

Descriptive Statistics

	Mean	Std. Deviation	N
arrestatiesNA	3,10	2,360	20
gevangenisstraf	3,15	,988	20

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,716 ^a	,512	,485	1,694

a. Predictors: (Constant), gevangenisstraf

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2,283	1,295		-1,763	,095
	gevangenisstraf	1,709	,393	,716	4,346	<,001

a. Dependent Variable: arrestatiesNA

- Hoeveel procent van de variatie in het *aantal arrestaties na gevangenisstraf* kan verklaard worden op basis van *aantal jaren gevangenisstraf*?
- Wat is de verwachte toename in *aantal arrestaties na gevangenisstraf* met elk bijkomend jaar *gevangenisstraf*?

- Een bivariaat model voor arrestaties na gevangenisstraf dat stelt dat aantal jaren gevangenisstraf de enige impactfactor is, is echter een twijfelachtig model. Gezond verstand zegt ons immers dat dit bivariate model niet correct is gedefinieerd, er zijn immers nog andere factoren die het aantal arrestaties kunnen beïnvloeden. Als dit waar is - dat relevante factoren die verband houden met het aantal jaren gevangenisstraf uit het model zijn weggelaten, dan kan de regressiecoëfficiënt een zeer misleidende schatting geven van het effect van gevangenisstraf op arrestaties. We voegen daarom een derde variabele toe aan ons model: *aantal arrestaties voor gevangenisstraf*.

	Aantal arrestaties na gevangenisstraf	Aantal jaren gevangenisstraf	Aantal arrestaties voor gevangenisstraf
Aantal arrestaties na gevangenisstraf	1		
Aantal jaren gevangenisstraf	.72	1	
Aantal arrestaties voor gevangenisstraf	.76	.63	1

a) Hoe interpreteer je de relatie tussen *aantal arrestaties voor gevangenisstraf* en *aantal arrestaties na gevangenisstraf*?

b) Is *aantal arrestaties voor gevangenisstraf* derhalve een relevante factor om mee te nemen in het regressiemodel ter verklaring van *aantal arrestaties na gevangenisstraf*?

c) Hoe weten we of de relatie tussen ‘*aantal jaren gevangenisstraf*’ en ‘*aantal arrestaties na gevangenisstraf*’ niet gewoon het gevolg is van het feit dat zij die een langere gevangenisstraf uitzitten over het algemeen al een zwaarder verleden hebben (vertaald als een *hogere aantal arrestaties voor gevangenisstraf*)?

- Wat gebeurt er met de samenhang tussen ‘*aantal jaren gevangenisstraf*’ en ‘*aantal arrestaties na gevangenisstraf*’ als we controleren voor ‘aantal arrestaties voor gevangenisstraf’?
- Vul onderstaande tabel aan met de partiële correlatiecoëfficiënt.
- Interpreteer de bevinding.

CORRELATIES

Controlevariabele		Aantal jaren gevangenisstraf	Aantal arrestaties NA gevangenis
Aantal arrestaties VOOR gevangenisstraf	Aantal jaren gevangenisstraf	1	
	Aantal arrestaties NA gevangenisstraf	??	1

2. Partiële correlatie is een statistische manier om het verband tussen twee variabelen te berekenen als je de invloed van één (of meer) variabele(n) wilt verwijderen. Dit wordt het *corrigeren voor een derde variabele* genoemd. In essentie komt het erop neer dat de waarde van de correlatiecoëfficiënt wordt gecorrigeerd door rekening te houden met de invloed van een derde variabele (of meerdere variabelen).

Je kan de partiële correlatie toepassen als de zero-order correlaties beschikbaar zijn. Dit zijn de bivariate correlaties tussen de hoofdvariabelen.

Hieronder vind je een correlatiematrix met de correlaties tussen drie variabelen.

	NUMERIEKE SCORE = X	VERBALE SCORE = Y	LEEFTIJD = Z
NUMERIEKE SCORE (= X)	1.00		
VERBALE SCORE (= Y)	.97	1.00	
LEEFTIJD (= Z)	.80	.85	1.00

- Bereken $XY.Z$ en vul onderstaande presentatie van de resultaten verder aan.

Presentatie van de resultaten

“Omdat leeftijd zowel met verbale als met numerieke vaardigheid correleert, besloten we om te onderzoeken wat het effect van een correctie voor leeftijd is op de correlatie. Na het uitvoeren van een partiële correlatie neemt de correlatiecoëfficiënt (toe/af), van (= zero-order correlatie tussen numerieke score en verbale score) naar (=waarde van de partiële correlatie). Dit is een (grote/kleine) verandering. Leeftijd heeft (weinig/veel) impact op de correlatie tussen verbale en numerieke scores.”

3. Een onderzoeker vindt voor een grote stedelijke politiezone een correlatie van $-.40$ tussen prestaties op een fysieke bekwaamheidstest (X) en salaris (Y) in een steekproef van 50 politieagenten. Op het eerste zicht zou men hieruit kunnen besluiten dat de politiezone een lager salaris uitbetaalt aan agenten die in fysieke topconditie zijn. Het houdt echter meer steek dat het aantal dienstjaren (Z) een invloed heeft op zowel fysieke bekwaamheid als op salaris.

	FYSIEKE BEKWAAMHEID = X	SALARIS = Y	DIENSTJAREN = Z
FYSIEKE BEKWAAMHEID	1.00		
SALARIS	-.44	1.00	
DIENSTJAREN	-.68	.82	1.00

- Bereken, aan de hand van de mathematische formule, de partiële correlatiecoëfficiënt (notatie= $R_{xy.z}$ = de partiële correlatie van X en Y onder controle van Z).
- Wat is jouw besluit?

4. EXTRA OEFENING

Een onderzoeker is geïnteresseerd in de samenhang tussen werkloosheidsduur en sollicitatie-activiteit onder ex-gedetineerden. Hij interviewt 6 werkloze ex-gedetineerden en bevraagt onder meer :

- Aantal weken dat zij werkloos zijn (X)
- Het aantal sollicitaties dat zij kunnen voorleggen (Y)
- Hun leeftijd (Z).

Je vindt hieronder de resultaten.

AANTAL WEKEN WERKLOOS =X	AANTAL SOLLICITATIES =Y	LEEFTIJD =Z
2	8	30
7	3	42
5	4	36
12	2	47
1	5	29
10	2	56

- Bereken, op basis van een serie regressievergelijkingen, de partiële correlatiecoëfficiënt voor de samenhang tussen 'aantal weken werkloos' en 'het aantal sollicitaties' onder controle van 'leeftijd'.
- Wat is jouw besluit ?