

Hoofdstuk 6

Inleiding tot de bivariate beschrijvende statistiek

1. Inleiding: causale relaties versus statistische relaties

Hoewel het begrip causaliteit vaak in onze theorievorming voorkomt, kunnen wij **nooit** met statistiek het bestaan van causale relaties bevestigen. We onderzoeken enkel statistische relaties. Statistische relaties zijn nooit waterdichte bewijzen voor causale relaties. Het enige dat we als onderzoekers kunnen doen, is de kenmerken (sterkte, richting en vorm) van een verband nagaan. Causaliteit is een filosofische en theoretische kwestie. De empirische realiteit laat zich dus wel ten dele onderzoeken, op voorwaarde dat we over adequate (geldige) en betrouwbare meetinstrumenten beschikken, maar statistiek biedt nooit waterdichte bewijzen voor causaliteit. Het omgekeerde is echter wel waar: het bestaan van causaliteit impliceert statistische samenhang. En daar zit het belang van statistisch onderzoek naar de relatie tussen variabelen. Omdat David Hume heel overtuigend had beargumenteerd dat een oorzaak-gevolg relatie niet kan worden vastgesteld met het blote oog, hield hij een pleidooi voor het beschouwen van causaliteit als regelmatige samenhang. Door het herhaaldelijk samen observeren van X en Y, zo beargumenteerde hij, besluiten we tot causaliteit. Causaliteit is in deze visie dus een gewoonte van de geest en niets meer dan dat. Enkele prominente statistici, waaronder Karl Pearson, de uitvinder van de correlatiecoëfficiënt, geloofde dat causaliteit weinig belang had en dat wetenschap evengoed kon groeien door kennis te nemen van samenhangen.

Tegenover deze Humeaanse visie staat de visie dat causaliteit **reëel** is. Vroeger dacht men dat causaliteit te maken had met onvermijdelijkheid, met complete willoosheid en met onvoorkomelijkheid. Denk maar aan het deterministische gedachtengoed uit de 19^{de} eeuw, dat kenmerkend was voor de positivistische school. Gaandeweg is men dat **determinisme** gaan omruilen voor **probabilisme**. Immers, de realiteit van het sociaal-wetenschappelijk onderzoek laat zien dat causaal onderzoek altijd leidt tot frequentieverdelingen en samenhangen die nooit perfect zijn. Maar moeten wij perfecte verbanden veronderstellen? Natuurlijk niet. De mens maakt keuzes, i.e. de mens delibereert, maar laat zich bij het maken van keuzes wel leiden door een aantal attitudes, preferenties en dergelijke meer. Keuzes worden ook gemaakt in een bepaalde context. De combinatie hiervan (desires, beliefs, opportunities, constraints) bepalen allemaal of een handeling gesteld wordt. Een causale factor is een factor die een effect teweeg brengt. Een belangrijk onderscheid is het onderscheid tussen:

- ***Voldoende en noodzakelijke oorzaken:*** een factor is een voldoende en noodzakelijke causale factor als het effect wordt teweeggebracht als de factor aanwezig is en het effect niet wordt teweeggebracht als de factor niet aanwezig is. Het criterium van de noodzakelijkheid in combinatie met het voldoende criterium is zeer streng. Zulke factoren vinden we niet in de criminologie.
- ***Onvoldoende en noodzakelijke oorzaken:*** zulke factoren brengen het effect enkel teweeg in combinatie met een reeks van factoren. Dergelijke factoren vormen het noodzakelijke geheel binnen een geheel van factoren die enkel in combinatie een effect teweegbrengen.
- ***Voldoende en niet-noodzakelijke oorzaken:*** deze factoren zijn zelfstandig in staat om het effect teweeg te brengen bij aanwezigheid, maar er zijn ook nog andere factoren die hetzelfde effect kunnen teweegbrengen. In tegenstelling tot de voldoende en noodzakelijke voorwaarden, is het bij de voldoende voorwaarden zo dat de factor niet noodzakelijk is. Voldoende factoren zijn veel realistischer in criminologisch onderzoek: er zijn in de geschiedenis van de criminologie veel kandidaat-voldoende factoren naar voor geschoven. Ze zijn echter op zich niet noodzakelijk, want andere factoren kunnen hetzelfde teweegbrengen.
- ***Contributieve oorzaken of niet-voldoende en niet noodzakelijke oorzaken:*** dit komt dichtst bij de realiteit: het gaat om factoren waarvan we weten dat ze een bijdrage leveren (een verhoging of verlaging van probabiliteiten). In het kader van multicausale problemen, zoals criminaliteitsproblemen, mogen ze nooit alleen bestudeerd worden, maar steeds in interactie met andere factoren.

In de criminologie is er sprake van **contributieve** oorzaken. Regelovertredend gedrag is een **multicausaal fenomeen waarbij verklarende factoren bestaan op meerdere niveaus (het individu met zijn persoonlijke eigenschappen en de omgeving met diens eigenschappen)**: er is niet zo maar één oorzaak van regelovertredend gedrag. In tegendeel, vaak is het zo dat er een complex samenspel is tussen causale factoren.

Bovendien moet men een onderscheid maken tussen **singulaire causaliteit** en **veralgemeenbare causaliteit**.

- **Singulaire causatie (token causation):** causatie op het individuele niveau: waarom heeft Jon Jonsson op 27 februari 2013 zijn vrouw doodgeschoten? Het gaat hier om een individuele case studie.
- **Algemene causatie (type causation):** waarom observeren we in een bepaalde groep (een steekproef van jongeren bijvoorbeeld) dat individuen met bepaalde kenmerken vaker gewelddadige delicten plegen?

In het meeste criminologisch verklarend onderzoek gebruiken we de statistiek om veralgemeenbare uitspraken te toetsen. We zijn meer geïnteresseerd in veralgemeenbare theorieën dan in singulaire uitspraken. Dit is begrijpelijk: veralgemeenbare theorieën geven richting aan het beleid. We willen niet enkel voorkomen dat Jan Janssens in de toekomst criminaliteit pleegt, maar we willen er vooral voor zorgen dat individuen met bepaalde kenmerken geen regelovertredend gedrag meer stellen door in te werken op die causale factoren die de sterkste samenhang vertonen in populaties. Let wel: de samenhang moet inhoudelijk en theoretisch motiveerbaar zijn. Dat is de rol van theorie en dat durven praktiserende criminologen wel eens vergeten.

2. Causaliteit op een bierviltje

- Causaliteit als **productie: generatieve causaliteit**: X is een productieve oorzaak van Y als X er in slaagt Y teweeg te brengen. De vraag is alleen hoe. De aanhangers van productieve causaliteitstheorieën leggen de nadruk op de rol van twee kenmerken: (1) **mechanismen** die verklaren op welke wijze de productieve causaliteit plaatsvindt en (2) een zeker “**wetmatig karakter**” (in het Engels: **lawfulness**). Volgens deze invloedrijke theorie brengt een oorzakelijke factor het gevolg teweeg via het in gang zetten van een mechanisme. In de sociale wetenschappen bestaan verschillende mechanismen, sociale mechanismen bestaan in sociale systemen en biosociale mechanismen en psychosociale mechanismen bestaan in individuen (die biosociale systemen zijn). Er kan geen sprake zijn van een mechanisme als een oorzakelijke factor de ene keer iets in gang zet en de andere keer iets anders in gang zet. Het is nu eenmaal eigen aan mechanismen dat er een zekere regelmaat is te bespeuren. Het mag niet gaan om een eenmalige samenhang. Dat “wetmatige” is vaak verkeerd begrepen geweest. Men bedoelt hier zeker geen unicausale deterministische wetmatigheden mee

zoals in de klassieke mechanica. Zo werken wetmatigheden niet in de sociale wetenschappen. Het bestaan van statistische *empirische generalisaties suggereert sociale wetmatigheden*.

- **Causaliteit als tegenfeitelijkheid.** Een causale factor brengt een gevolg teweeg, *dat niet zou zijn teweeggebracht, indien er geen interventie of manipulatie had plaatsgevonden*. Deze visie is zeer populair onder experimentele criminologen. *Causaliteit moet een verschil maken*. De belangrijkste vertegenwoordiger van de tegenfeitelijke causaliteitstheorie is James Woodward. Tegenfeitelijke causaliteit is interessant met betrekking tot de reflectie over criminaliteitspreventie: hadden we criminaliteit kunnen voorkomen indien bepaalde zaken anders waren gelopen? We kunnen door de manipulatie van de oorzaak, het gevolg teweegbrengen. Onder experimentele criminologen is deze tegenfeitelijke causaliteitstheorie van Woodward zeer populair. Opgelet: soms weten we dat we door een bepaalde manipulatie iets teweeg brengen, maar verstaan we het mechanisme niet.
- **Causaliteit als robuuste statistische afhankelijkheid:** Statistici hebben in de jaren 1960 als reactie op beschuldigingen van determinisme causaliteit gedefinieerd als robuuste causale afhankelijkheid, maar robuust diende geïnterpreteerd te worden in termen van probabilistisch. Statistisch onderzoek leidt altijd tot frequentieverdelingen en stochastische relaties.

In de statistiek spelen we op veilig en gaan we enkel iets zeggen over robuuste afhankelijkheden. De robuuste afhankelijkheidstheorie zegt vereenvoudigd:

- X is een oorzaak van Y als en slechts als X bestaat (de probabilliteit van X is groter dan nul).
- X is een oorzaak van Y als X temporeel eerder komt dan Y (het tijdsmoment T1 komt voor T2), er een statistische relatie is tussen X en Y: de probabilliteit van Y gegeven X is groter dan de probabilliteit van Y “tout court”. Anders gezegd: de conditionele probabilliteit is groter dan de marginale probabilliteit.
- X is een oorzaak van Y als er geen schijnverband is (“no spurious relation”: de statistische relatie tussen X en Y verdwijnt niet als er gecontroleerd wordt voor storende controlevariabelen ook wel confounders genoemd, ook wel de “ceteris paribus” conditie genoemd).

Probabilistische causaliteitstheorieën leggen het robuuste statistische asymmetrische verband uit, niet het causale verband. Maar ze hebben een belangrijke troef gehad: ze hebben de angst voor determinisme van kwantitatief onderzoek verdreven en de weg geruimd voor causaal onderzoek.

Laat ons nu een voorbeeld geven van de verschillende inhoudelijke interpretaties die aan causale verbanden gegeven worden. Inhoudelijk is het belangrijk de mechanismen van de samenleving te begrijpen, maar daartoe is statistiek maar een hulpmiddel.

- An slaagt voor het examen statistiek omdat ze een vrouw is. Vrouwen hebben een statistisch hogere kans om te slagen voor statistiek dan mannen (robuust verband geslacht-studieresultaten).
- An slaagt voor het examen statistiek omdat ze ervoor gestudeerd heeft (studeren= mechanisme)
- An slaagt voor het examen statistiek om ze door een mentor werd getraind (trainen= interventie, die op zich inzichten verbetert).

3. Symmetrische en asymmetrische relaties tussen variabelen

Theoretisch gezien moeten we in de bivariate statistiek een onderscheid maken tussen **asymmetrische** relaties en **symmetrische** relaties. **Dit onderscheid is zeer belangrijk omdat het de keuze van de associatiematen zal bepalen.** Bij een **asymmetrische** statistische relatie wordt inhoudelijk verondersteld dat de ene variabele een “*causale*” invloed uitoefent op de andere. Een asymmetrische relatie is een dependente relatie. Er is dependentie of afhankelijkheid: Y is afhankelijk van X betekent zoveel als Y wordt mee beïnvloed door X. Nog anders gesteld: *X is een oorzaak van Y*. Een criminologisch voorbeeld: het hebben van slechte vrienden leidt tot het plegen van meer criminele handelingen. Bij **asymmetrische** relaties is er steeds een **onafhankelijke** variabele en een **afhankelijke** variabele. De afhankelijke variabele is steeds afhankelijk van de onafhankelijke variabele. In het voorbeeld is het plegen van criminele handelingen de afhankelijke variabele en het hebben van slechte vrienden de onafhankelijke variabele. **De afhankelijke variabele wordt steeds voorgesteld met de hoofdletter Y en de onafhankelijke variabele wordt steeds voorgesteld met de hoofdletter X.** We kunnen bijkomend een onderscheid maken tussen een **rechtstreeks** (direct) en een **onrechtstreeks** (indirect) effect. Een rechtstreeks effect betekent dat een onafhankelijke variabele een rechtstreekse invloed heeft op de afhankelijke variabele. Een onrechtstreeks effect betekent dat een veranderlijke X geen rechtstreeks effect heeft op de afhankelijke variabele Y, maar onrechtstreeks, via de invloed op een andere variabele Z.



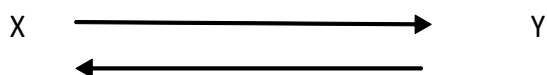
Zo heeft de sociale band van jongeren met diens ouders (variabele X) geen rechtstreeks statistisch effect op de betrokkenheid van jongeren bij crimineel gedrag (variabele Y), maar onrechtstreeks via de invloed van de ouder-kind relatie op de bereidheid van de jongere conforme normen te accepteren (variabele Z). Deze laatste variabele is doorslaggevend en dus het rechtstreekse of directe effect op de afhankelijke variabele Y.

Bij een **symmetrische** relatie kan men op theoretische gronden geen onderscheid maken tussen de beide variabelen. Het enige dat we kunnen zeggen is dat de beide kenmerken *samenhangen of correleren*. Er bestaat een verband, maar we beschikken niet over geldige argumenten om te zeggen hoe de invloed verloopt.

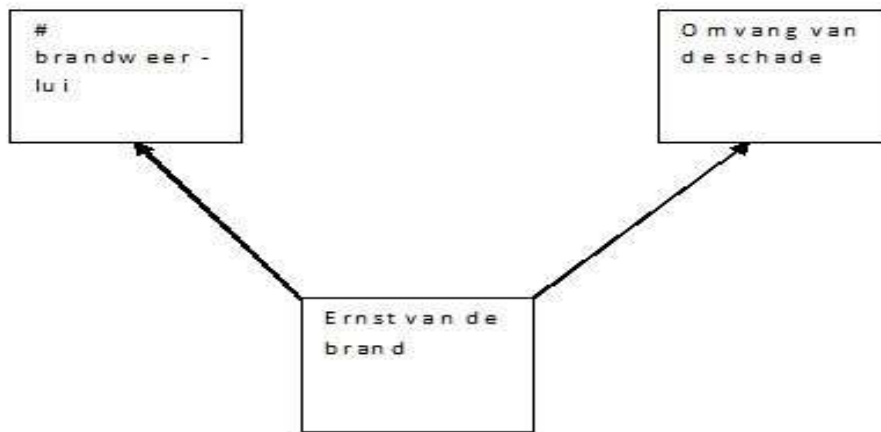
Er zijn 3 theoretische situaties te onderscheiden waarin kan gesproken worden van een symmetrische relatie tussen de variabelen X en Y.



Ten eerste kan er sprake zijn van een **wederzijdse (reciproque) invloed** tussen beide variabelen. In dat geval is er sprake van invloed van X op Y, maar ook van Y op X. Een gekend voorbeeld is de relatie tussen studiemotivatie en prestatie, een gekend criminologisch voorbeeld is het verband tussen de ruimtelijke concentratie van overlast en de ruimtelijke concentratie van criminaliteit op straat. Het kan dat overlast aanzet tot het plegen van criminaliteit, maar het kan dat criminaliteit in de straat op zijn beurt verdere overlast in de hand werkt.



Ten tweede kan er sprake zijn van een **schijnverband**. In dat laatste geval is er sprake van een **gemeenschappelijke oorzaak**.



In bovenstaand voorbeeld stelt men vast dat hoe meer brandweerlui aanwezig zijn op de plaats van de brand, hoe groter de schade is aan het gebouw. Moet men daaruit afleiden dat grote aantallen brandweerlui meer schade veroorzaken dan kleinere aantallen? Nee! Er is een gemeenschappelijke oorzaak die het schijnverband verklaart. Hoe ernstiger de brand, hoe meer brandweerlui opgetrommeld worden, en hoe ernstiger de brand is, hoe groter de schade ook is.

Ten derde kan er een samenhang bestaan tussen twee variabelen omdat deze **indicatoren zijn van hetzelfde begrip**. Zo kan er een sterke samenhang bestaan tussen het percentage buurtbewoners dat slachtoffer wordt van woninginbraak en de politiestatistieken voor woninginbraak, omdat deze beide metingen zijn van hetzelfde begrip. In dat geval vertonen beide metingen meetfouten en is de doorsnede het gevolg van het gemeenschappelijke achterliggende theoretische kenmerk dat beide indicatoren delen.

4. Doelstelling van de bivariate beschrijvende statistiek

Indien we de relatie tussen twee kenmerken willen beschrijven, maken we gebruik van de *bivariate beschrijvende statistiek*. Bivariate analyses beschrijven de relatie tussen *twee variabelen*. Afhankelijk van het meetniveau wordt een beschrijvende bivariate analysetechniek gekozen. We besteden in dit handboek zowel aandacht aan *associatiematen* voor *categorische* of *non-parametrische* variabelen als aan associatiematen voor en *metrische* variabelen.

Alhoewel we soms enkel in de frequentieverdeling van één variabele zijn geïnteresseerd, gaat onze aandacht in de criminologie vaker uit naar de relatie tussen twee of meer variabelen.

Bestaat er een verband tussen twee variabelen? Zo ja, dan willen criminologen volgende zaken weten:

- **(1) hoe sterk is deze relatie;**
- **(2) welke richting (+ of -) neemt deze relatie aan en;**
- **(3) welke vorm (lineair of curvilineair) neemt deze relatie aan?**

De doelstelling van de bivariate beschrijvende statistiek bestaat er in de belangrijkste aspecten van een *relatie tussen twee variabelen* samen te vatten op een eenduidige en economische manier. Wanneer we de waarden van de elementen in onze steekproef op twee variabelen kennen, kunnen we aan de hand van deze bruto-waarnemingen niets zeggen over de relatie tussen beide kenmerken. Daarom berekenen we vanuit deze bruto-waarnemingen een maat van associatie die de vorm aanneemt van een getal waaruit de sterkte en richting kan worden afgeleid.

Bivariate statistische associatiematen mogen niet zomaar op elke willekeurige set van variabelen worden toegepast. Zoals ook het geval was voor de univariate beschrijvende statistiek, hangt de keuze voor een bivariate beschrijvende associatiemaat af van het *meetniveau* van de twee variabelen in kwestie en van de doelstelling van de onderzoeksvraag. Dit betekent dat de criminoloog nu bij de keuze voor een associatiemaat met een kenmerk meer rekening dient te houden. De keuze voor een bepaalde associatiemaat kan verder bepaald worden door de onderzoeksvraag die de criminoloog zich stelt. Een criminoloog kan twee soorten onderzoeksvragen stellen die kunnen beantwoord worden aan de hand van de bivariate beschrijvende statistiek: *verklarende onderzoeksvragen* of *vragen naar samenhang*, zonder uitspraken te willen doen over de causale relatie zelf.¹ Verklarende onderzoeksvragen hebben doorgaans een causale ondertoon: de onderzoeker vermoedt dat gebeurtenissen van het type A een oorzaak zijn van gebeurtenissen van het type B. Typische verklarende vragen vinden we in de etiologische criminologie. In dat geval wordt de verklarende statistiek

¹ Het is zeer belangrijk om niet met de deur in huis te vallen en de statistische associatiematen te presenteren die de criminoloog kan gebruiken wanneer deze onderzoek doet. Met de bivariate beschrijvende statistiek kan de criminoloog geen hypothesen toetsen. Dit wil zeggen: de criminoloog kan geen extrapolatie maken van de onderzoeksbevindingen naar het universum. Dit is het domein van de inferentiële statistiek, en dat deel volgt later in deze syllabus.

gebruikt en wordt de predictie een instrument om een statistische effect-relatie vast te stellen. De effect-relatie kan echter geen waterdicht bewijs vormen voor de causale ondertoon.

5. Bivariate frequentieverdelingen voor lage en hoge meetniveaus

In de bivariate statistiek zijn we geïnteresseerd in bivariate frequentieverdelingen. Een bivariate frequentieverdeling is niet meer dan de frequentieverdeling van twee kenmerken. Aan de hand van een bivariate frequentieverdeling kunnen we nagaan of de frequentieverdeling van één variabele samenloopt met de frequentieverdeling van een ander kenmerk. We geven hier een eenvoudig voorbeeld: is het zo dat we kunnen zeggen dat er een associatie bestaat tussen het gebruik van harddrugs en het betrokken zijn bij georganiseerde diefstal? Er zijn meerdere verklaringen te geven voor de associatie tussen druggebruik en criminaliteit. Voor sommige individuen is het gebruik aanwezig voor de betrokkenheid bij criminaliteit en kan het gebruik een causaal mechanisme zijn. Gebruik vereist immers middelen, en daartoe kan criminaliteit een effectieve strategie zijn. Anderzijds kan een diepe betrokkenheid bij criminaliteit voor de nodige stress zorgen, en kan die stress een factor zijn die maakt dat de regelovertreder occasioneel verdovende middelen gebruikt om de stress te verlichten. Een derde verklaring is nog anders: zowel het druggebruik als de betrokkenheid bij criminaliteit hebben een gemeenschappelijke oorzaak: de zoektocht naar sensatie (thrill-seeking behaviour) en kicks. Hoe men deze associatie wil verklaren, de drie mogelijke scenario's veronderstellen allemaal dat er op zijn minst een associatie bestaat tussen de twee kenmerken.

Bivariate frequentieverdelingen worden voorgesteld aan de hand van een *contingentietabel* (*kruistabel*). Voor metrische variabelen kunnen we ze ook voorstellen aan de hand van een *puntenwolk*, of ook wel *scatterplot* genoemd.

De situatie voor contingentietabellen of kruistabellen

Voor nominale en ordinale kenmerken worden **contingentietabellen** of **kruistabellen** opgesteld. Een contingentietabel is een tabel waarin de categorieën van twee variabelen tegenover elkaar worden uitgezet en waarin de waargenomen frequentie van elke combinatie van categorieën vermeld staat. Een 2*2 tabel is een tabel met twee rijen en twee kolommen. Anders gezegd: er zijn twee categorieën bij elk van beide variabelen. Een 2*3 tabel is een tabel met twee rijen en drie kolommen. We zetten dus eerst de rijen en dan de kolommen in de notatie. We illustreren de gebruikelijke terminologie aan de hand van een 2*2 tabel voor de variabelen X en Y.

		Variabele X		Totaal
		Cat 1	Cat 2	
Variabele Y	Cat 1	A	B	A+B
	Cat 2	C	D	C+D
Totaal		A+C	B+D	A+B+C+D

A, B, C en D zijn **celfrequenties**: ze geven aan hoeveel keer een bepaalde combinatie van categorieën van variabelen voorkomen.

A+B en C+D noemen we de **rijtotalen** van de contingentietabel.

A+C en B+D noemen we de **kolomtotalen** van de contingentietabel.

A+B+C+D, ook wel de som van alle celfrequenties genoemd, is de **steekproefomvang**.

Een 2*2 kruistabel: de relatie tussen zelfgerapporteerde winkeldiefstal en cannabisgebruik

			ooit joint gerookt		Totaal
			nooit	ooit	
Winkel- diefstal	nooit	Count	2388	104	2492
		% <i>binnen winkeldiefstal</i>	95.8%	4.2%	100.0%
		% <i>binnen joint gerookt</i>	86.0%	42.6%	82.5%
	ooit	Count	390	140	530
		% <i>binnen winkeldiefstal</i>	73.6%	26.4%	100.0%
		% <i>binnen joint gerookt</i>	14.0%	57.4%	17.5%
Totaal	Count	2778	244	3022	
	% <i>binnen winkeldiefstal</i>	91.9%	8.1%	100.0%	
	% <i>binnen joint gerookt</i>	100.0%	100.0%	100.0%	

We zien in deze kruistabel dat elke cel (elke combinatie tussen de twee categorische variabelen) onderzoekseenheden bevat. We hebben naast de exacte geobserveerde eenheden eveneens de procentuele verdeling duidelijk gemaakt. Je ziet dat de celfrequenties (de aantallen) verschillend zijn voor elke combinatie van categorieën. Absolute cijfers zeggen zo weinig. Daarom presenteren we ook de percentages. Je kan die percentages in twee richtingen berekenen: het aantal jongeren die ooit cannabis hebben gerookt voor jongeren die winkeldiefstal hebben gerapporteerd en voor jongeren die geen winkeldiefstal hebben gerapporteerd. Omgekeerd kan je de percentages winkeldiefstal aflezen voor jongeren die cannabis hebben gerookt en voor jongeren die geen cannabis hebben gerookt. Je kan je nu

afvragen welke de juiste leesrichting is. Dat is een zeer terechte vraag, die we verderop behandelen. Hier volstaat het om vertrouwd te raken met de begrippen celfrequenties, rijtotalen en kolomtotalen.

- De celfrequenties zijn 2388, 104, 390 en 140.
- De rijtotalen zijn 2492 en 530.
- De kolomtotalen zijn 2778 en 244.
- De steekproefomvang is 3022

Een 3*2 kruistabel: de relatie tussen dronkenschap en slagen en verwondingen

			ooit dronken geweest		Total
			nooit	ooit	
Slagen en verwondingen	Nooit	Count	1790	358	2148
		%	74.3%	58.2%	71.1%
	Enkele	Count	288	108	396
	keren	%	12.0%	17.6%	13.1%
	Drie	Count	330	149	479
	keer of meer	%	13.7%	24.2%	15.8%
Total		Count	2408	615	3023
		%	100.0%	100.0%	100.0%

De 2*2 kruistabel kan ook verder uitgebreid worden naar een r*k tabel. In het voorbeeld hierboven hebben we een tabel gepresenteerd waar drie rijen te zien zijn en twee kolommen.

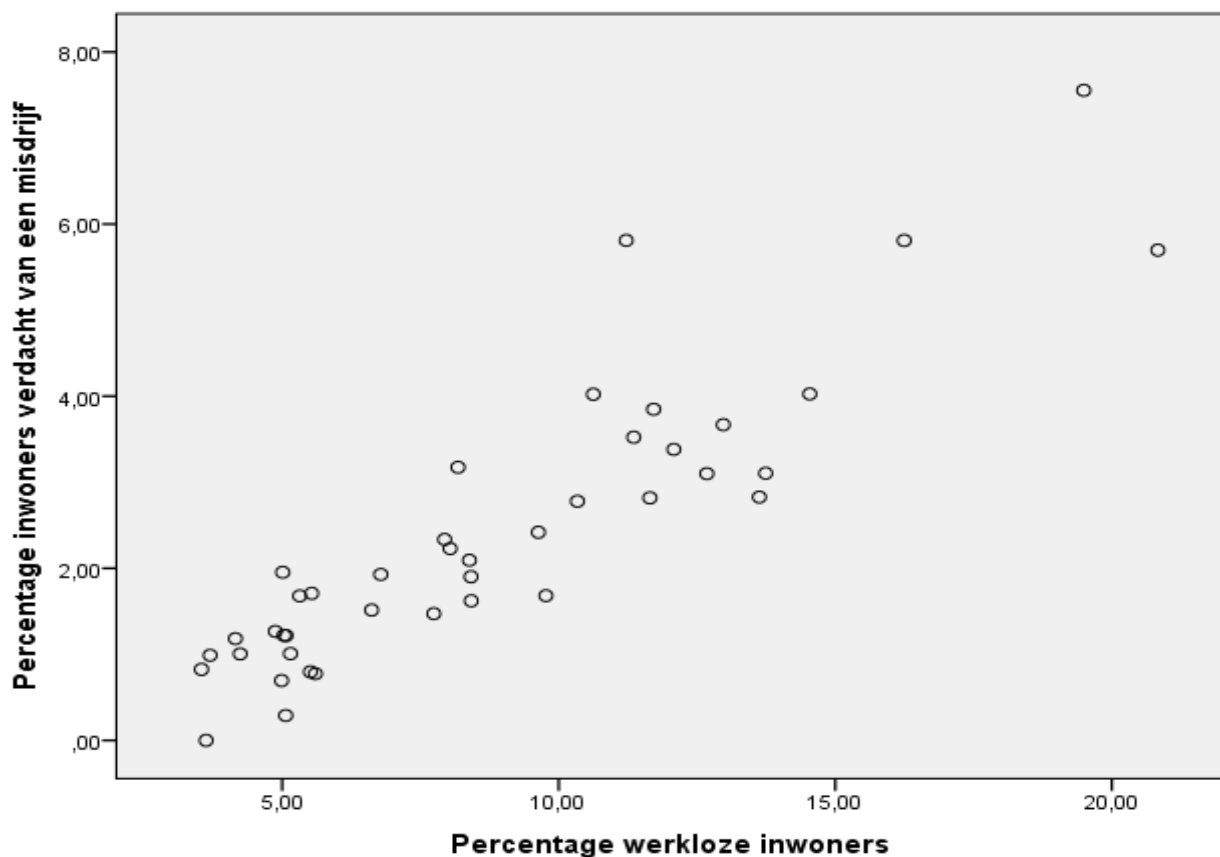
- De celfrequenties zijn nu 1790, 358, 288, 108, 330, 149.
- De rijtotalen zijn nu 2148, 396, 479
- De kolomtotalen zijn hier 2408 en 615
- De totale steekproef bestaat uit 3023 respondenten

De situatie voor metrische variabelen

Voor metrische variabelen ligt het iets moeilijker om te werken met een contingentietabel. Het zou wel bijzonder onleesbaar zijn om twee metrische variabelen in kruistabel te willen gieten. Het gaat hier immers niet om een beperkt aantal categorieën, maar om variabelen die heel fijn zijn gemeten. Daarom is het allerm minst wenselijk om de ruwe data te willen presenteren aan de hand van een kruistabel. Betekent dit dan dat een kruistabel maken niet

wenselijk is? Neen. Het betekent dat we de ruwe gegevens zullen moeten hercoderen willen we de relatie tussen twee metrische kenmerken in een kruistabel gieten. Bij metrische variabelen hebben de onderzoekseenheden heel vaak zeer veel verschillende waarden. Dit komt natuurlijk omdat de variabelen op hoog meetniveau zijn gemeten: bij metrische variabelen is het meetniveau veel fijner en heeft het geen zin om een contingentietabel te maken waar elke waarde op de twee variabelen tegenover elkaar wordt uitgezet. Daarom worden de waarden tegenover elkaar geplaatst op een **scatterplot of puntenwolk**.

Een puntenwolk of scatterplot is een grafische voorstelling van de frequentieverdeling van twee variabelen die gemeten zijn op metrisch niveau. In een puntenwolk worden twee variabelen tegen elkaar uitgezet door de waarden van de variabele X op de **abscis** te plaatsen, en de waarden van de variabele Y op de **ordinaat** te zetten. Conventioneel zetten we altijd de Y-variabele op de ordinaat en de X-variabele onderaan op de abscis. Elke eenheid uit de steekproef kan nu in het **tweedimensionaal** (“er zijn twee assen”) **orthogonaal** (“loodrecht op elkaar staande assen”) vlak worden gesitueerd. Dit gebeurt aan de hand van de waarde van elke eenheid op de variabele X en Y. Deze waarden (x_j , y_j) worden de **coördinaten** genoemd.



Dit voorbeeld toont aan hoe we de relatie tussen twee metrische kenmerken presenteren. De puntenwolk is handig, omdat deze ons in een oogopslag al een hint geeft over de relatie tussen de twee kenmerken. Het voorbeeld dat hier gegeven wordt, betreft de relatie tussen twee buurtkenmerken: het percentage werklozen aan de ene kant en het percentage jongeren dat verdacht wordt van een misdrijf aan de andere kant. De gegevens zijn afkomstig van een onderzoek dat in alle Antwerpse buurten werd gevoerd.

6. Verantwoord kiezen tussen een reeks van associatiematen

Associatiematen zijn voor de kwantitatieve criminoloog-onderzoeker “*the tools of the trade*” zoals scalpels en andere operatiegereedschappen tot de gereedschapskist van de chirurg behoren. Het komt er steeds op aan de juiste associatiemaat te kiezen. De “juiste” maat is een maat die past bij de **onderzoeksvraag van de criminoloog**. De juiste associatiemaat is een maat die de onderzoeksvraag beantwoordt. Bij de keuze voor een best passende associatiemaat houdt men rekening met **twee regels**:

- *Ten eerste* is er het meetniveau van de afhankelijke en/of de onafhankelijke variabele. **Als beide verschillend zijn, kiest men doorgaans voor het meetniveau van de variabele die op het laagste niveau is gemeten.** Bijvoorbeeld: willen we de relatie nagaan tussen een nominaal en ordinaal kenmerk, dan moeten we een nominale associatiemaat gebruiken.
- *Ten tweede* is er de **verwachte theoretische richting van de associatie**: symmetrische of asymmetrische analysetechnieken worden dan gekozen in functie van de verwachte theoretische associatie. We spreken van een asymmetrische analyse wanneer we uitspraken willen doen in de trant van “X leidt tot Y” of “als X dan Y”. We veronderstellen als het ware een eenrichtingsverkeer; X leidt tot Y maar niet omgekeerd. Om het met een sprekend voorbeeld te stellen: geslacht kan nooit de afhankelijke variabele zijn in criminologisch onderzoek.

7. Leerdoelen

In dit hoofdstuk werden een aantal cruciale begrippen behandeld die echt van belang zijn wil men de bivariate verklarende statistiek en bij uitbreiding de multivariate beschrijvende statistiek begrijpen. Deze concepten maken niet enkel het hart uit van de statistiek, maar ook van het kwantitatieve verkennende en theorie-toetsende criminologische onderzoek. Het is dus van belang de betekenis van deze begrippen te kennen aangezien deze begrippen niet

enkel figureren in de statistiek. We verwachten dat de studenten de diverse invullingen die aan het causaliteitsbegrip werden gegeven kennen en vooral ook de beperkingen inzien van elke conceptualisering van causaliteit. De verschillende statistische termen die werden uitgelegd dienen gekend te zijn. Bedenk hierbij best een criminologisch voorbeeld.