

Het onderzoeksproces

1.1 Een voorbeeld

Een onderwijskundige, opgeleid in de tijd dat de pedagogiekstudie voornamelijk wijsgerig van aard was, is geïnteresseerd in de vraag of motivatie prestatie beïnvloedt. In zijn kennissenkring bevinden zich een socioloog en een psychologe. Aan beiden stelt hij de vraag hoe zij zouden onderzoeken of gemotiveerd zijn leidt tot het leveren van betere prestaties dan niet gemotiveerd zijn. Zowel de psychologe als de socioloog geven een uitleg over de wijze waarop zij empirisch de vraag van de onderwijskundige zouden beantwoorden. Tevens besluiten zij onafhankelijk van elkaar een onderzoek naar de genoemde vraagstelling uit te voeren.

De socioloog – medewerker van de vakgroep Onderwijssociologie van Universiteit X – heeft een onbetaalde nevenfunctie als bestuurslid van een Scholengemeenschap. Na uitvoerig overleg met de overige bestuursleden, directie, staf en scholieren krijgt hij toestemming het volgende onderzoek uit te voeren. Uit het totale leerlingenbestand van duizend leerlingen die in het schoolregister zijn ingeschreven, trekt hij willekeurig tweehonderd namen. Van elke leerling gaat hij na wie hun klasseleraar is. Vervolgens benadert hij alle daarvoor in aanmerking komende klasseleraars met de vraag, of zij voor elke leerling het gemiddelde rapportcijfer willen berekenen van het kerstrapport. Tevens verzoekt hij hen met een soort rapportcijfer tussen 1 en 10 aan te geven, hoe gemotiveerd elke leerling is. Hoe hoger dit cijfer, des te gemotiveerder de leerling.

Op deze wijze krijgt de socioloog van elke leerling twee gegevens: het gemiddelde rapportcijfer en een cijfer dat de mate van motivatie aangeeft. Hij besluit deze gegevens als volgt te analyseren. Leerlingen met een cijfer lager dan 6 voor motivatie beschouwt hij als weinig gemotiveerde leerlingen; de overigen als wel gemotiveerd. Hij heeft nu twee subgroepen van leerlingen. Voor elke subgroep af-

zonderlijk berekent hij vervolgens het gemiddelde van de gemiddelde rapportcijfers. Bij de leerlingen die volgens hun klasseleraar wel gemotiveerd zijn, blijkt dat het gemiddelde cijfer over hun kerstrapport 7,3 is. Het gemiddelde cijfer van de weinig gemotiveerde leerlingen is daarentegen slechts 4,6.

Inmiddels heeft de psychologe dezelfde vraag op een geheel andere manier onderzocht. Als werkgroepbegeleidster van eerstejaars studenten in de psychologie voor het vak Inleiding Onderzoeksmethoden heeft zij een mogelijkheid aan proefpersonen voor onderzoek te komen. Alle eerstejaars studenten zijn namelijk verplicht in hun eerste studiejaar aan tien experimenten mee te doen.

De psychologe betreft alle twintig studenten van één werkgroep in haar onderzoek. Zij legt deze studenten tijdens een bijeenkomst een intelligentietest voor. Vervolgens gaat zij na hoe hoog iedere student op deze test heeft gescoord. Daarna deelt ze de groep op in twee subgroepen van elk tien studenten en wel zodanig dat de gemiddelde intelligentie in de ene groep precies gelijk is aan die in de andere groep. Tijdens de volgende bijeenkomst vertelt zij de twintig studenten dat zij vandaag verplicht zijn aan een experiment mee te doen. Zij vraagt de tien studenten uit de eerste subgroep met haar mee te gaan naar een ander lokaal. Daar legt zij uit dat een collega een aantal vraagstukken op het terrein van de logica heeft opgesteld waarvan hij de moeilijkheidsgraad wil vaststellen. Hij wil deze vraagstukken eerst uitproberen bij verschillende groepen studenten. Zij zou het – gezien het belang van het project waarin de vraagstukken een rol spelen – bijzonder op prijs stellen, als haar studenten een waardevolle bijdrage zouden leveren. Zij krijgen één uur de tijd om antwoorden op de vraagstukken te formuleren. Deze zullen worden nagekeken. Behalen zij een cijfer hoger dan 7, dan zullen deze tien studenten ter compensatie niet worden opgeroepen voor de volgende drie experimenten waaraan zij anders zouden moeten meedoen. De psychologe vertrekt, gaat naar de overige tien studenten, vertelt hen hetzelfde verhaal over haar collega, deelt de vraagstukken uit, maar zegt niets over een mogelijke beloning als de studenten meer dan een 7 halen voor het oplossen van de vraagstukken. Na een uur leveren beide groepen hun antwoorden in. Na correctie blijkt de eerste groep 7,3 behaald te hebben, de tweede groep 4,6.

Zowel de socioloog als de psychologe doen verslag van hun onderzoek. Bij het lezen ervan komt bij de onderwijskundige een groot aantal vragen op. Hij had immers aan beide kennissen dezelfde vraag gesteld: „Beïnvloedt motivatie prestatie?„. Beiden hebben deze vraag trachten te beantwoorden door middel van empirisch onderzoek. De onderzoeksmethode van de een verschilt nogal van die van de ander. De socioloog heeft zijn onderzoek uitgevoerd bij tweehonderd middelbare scholieren, de psychologe bij twintig eerstejaars studenten in de psychologie. De socioloog vertaalde het theoretische begrip ‘prestatie’ in ‘het gemiddelde cijfer van het kerstrapport’, de psychologe in ‘het cijfer voor het oplossen van een aantal vraagstukken op het terrein van de logica’.

Het begrip ‘motivatie’ werd door de socioloog vertaald in ‘motivatie volgens de

klasseleraar'. Aan deze vroeg hij immers met een cijfer aan te geven hoe gemotiveerd een leerling was. De psychologe daarentegen probeerde zelf de motivatie te beïnvloeden. Aan de ene groep vertelde zij dat de studenten bij een goed resultaat niet behoefden mee te doen aan drie experimenten waaraan zij normaliter zouden moeten deelnemen. De andere groep kreeg alleen de opdracht de vraagstukken te maken. Er werd geen beloning in het vooruitzicht gesteld bij goede resultaten.

Ondanks het feit dat beide wetenschappers de vraagstelling verschillend uitwerken, komen zij toch tot soortgelijke bevindingen. De socioloog vindt een gemiddeld rapportcijfer van 7,3 bij de leerlingen die volgens hun klasseleraar gemotiveerd zijn. De psychologe vindt een score van 7,3 bij de tien studenten die zij een beloning in het vooruitzicht heeft gesteld. De tien studenten die niets verteld is over een dergelijke beloning, scoren slechts 4,6. Van de leerlingen die volgens hun klasseleraar weinig gemotiveerd zijn, is het gemiddelde rapportcijfer 4,6. Mag de onderwijskundige nu zeggen, dat motivatie de prestatie beïnvloedt? De resultaten van beide onderzoeken wijzen in die richting, maar toch? De onderwijskundige besluit een methodoloog te raadplegen.

De methodoloog stelt dat het antwoord op zijn vraag niet eenvoudig te geven is. Hij merkt in de eerste plaats op dat het onderzoek van de psychologe beter is opgezet dan dat van de socioloog. Zonder aanvullende gegevens mag men volgens hem uit het sociologisch onderzoek niet concluderen dat motivatie prestatie beïnvloedt. Weliswaar heeft de groep leerlingen die gemotiveerd is gemiddeld een veel beter kerstrapport dan de groep die maar weinig gemotiveerd is, maar dat hoeft niet veroorzaakt te zijn door een verschil in motivatie. Misschien zijn de gemotiveerde leerlingen wel intelligenter dan de weinig gemotiveerde leerlingen en komen ze vanwege hun hogere intelligentie tot een beter kerstrapport dan de weinig gemotiveerde leerlingen. Het hoeft dus niet aan hun hogere motivatie te liggen. Omdat de socioloog geen gegevens over de intelligentie van zijn tweehonderd leerlingen heeft verzameld, kan hij geen definitieve conclusie trekken. Hij kan voorlopig alleen stellen dat de leerlingen die volgens hun klasseleraar gemotiveerd zijn, betere prestaties leveren dan leerlingen die door hun klasseleraar weinig gemotiveerd genoemd worden. Hij weet echter niet of dat verschil in prestatie beïnvloed wordt door een verschil in motivatie of door een verschil in intelligentie. Wellicht verschillen de gemotiveerde en de minder gemotiveerde leerlingen nog in andere kenmerken en zijn de verschillen in prestatie noch aan verschillen in motivatie, noch aan verschillen in intelligentie toe te schrijven, maar aan deze verschillen in andere kenmerken.

Het is ook mogelijk dat de verschillen in prestatie deels aan de verschillen in motivatie, deels aan verschillen in intelligentie en deels aan verschillen in andere kenmerken toe te schrijven zijn. Zonder aanvullende gegevens over die kenmerken kan de socioloog daarover geen uitspraken doen.

De psychologe heeft volgens de geraadpleegde methodoloog veel meer zekerheid. Ook zij heeft te maken met een groep gemotiveerde en een groep niet gemotiveerde studenten en ook bij haar blijkt de eerste groep betere prestaties te le-

veren dan de tweede. Zij kan in tegenstelling tot de socioloog wel stellen dat het verschil in prestatie in elk geval niet door een verschil in intelligentie veroorzaakt is. Bij het begin van haar onderzoek heeft zij immers twee groepen gevormd die gemiddeld even intelligent zijn. Levert de ene groep betere prestaties dan de andere, dan kan dat in elk geval niet meer worden toegeschreven aan het feit dat die ene groep intelligenter is dan de andere. Indien beide groepen ook in andere kenmerken niet verschillen en dus alleen verschillen op grond van het feit dat de ene groep gemotiveerd is en de andere niet, dan kan de psychologe de conclusie trekken dat uit haar onderzoek blijkt, dat dit verschil in motivatie oorzaak is van het verschil in prestatie.

Bovendien legt de methodoloog aan de onderwijskundige uit, dat het onderzoek van de psychologe om nog een reden beter is dan dat van de socioloog. Bij haar onderzoek is het namelijk duidelijk dat motivatie in tijd aan prestatie voorafgaat. Wil men kunnen zeggen dat het ene verschijnsel het andere beïnvloedt, dan moet het ene in ieder geval aan het andere voorafgaan. De psychologe manipuleert eerst met motivatie en laat daarna pas de prestatie leveren. De socioloog daarentegen is op een andere wijze te werk gegaan. Hij laat de klasseleraars voor beide kenmerken de score bepalen. Nu is het mogelijk, dat dezen een laag cijfer geven voor motivatie aan juist die leerlingen die een slecht kerstrapport hebben. Dat hoeft niet zo te zijn. Belangrijk in dit verband is alleen dat het gebeurd kan zijn. Is het in feite gebeurd, dan kan men niet meer stellen dat motivatie prestatie beïnvloedt. Eerder moet men de conclusie trekken dat de prestaties van leerlingen bepalend zijn voor het cijfer dat zij voor motivatie krijgen. Het zou volgens de methodoloog veel beter geweest zijn, als de socioloog in het begin van het schooljaar op een of andere wijze de motivatie van de tweehonderd leerlingen had vastgesteld en pas daarna hun prestatie had bepaald. Pas dan zou hij zekerheid gehad hebben dat motivatie in tijd aan prestatie voorafging. Zou de socioloog ook dan gevonden hebben dat gemotiveerde leerlingen betere cijfers halen dan niet gemotiveerde leerlingen, dan nog zou hij echter niet kunnen concluderen dat motivatie prestatie beïnvloedt. Het kan immers nog steeds zo zijn, dat zijn gemotiveerde leerlingen bijvoorbeeld intelligenter zijn dan de niet gemotiveerde leerlingen en juist daarom betere prestaties leveren. Zonder aanvullende gegevens over de intelligentie van de leerlingen kan men daarover geen uitsluitsel krijgen. Met aanvullende gegevens wel, maar daarop gaan we later nog in.

Resumerend somt de methodoloog nog eens de drie punten op waarom de psychologe wel kan stellen dat motivatie prestatie beïnvloedt. Op de eerste plaats is gebleken dat studenten die gemotiveerd zijn, betere prestaties leveren dan studenten die niet gemotiveerd zijn. Op de tweede plaats heeft zij ervoor gezorgd dat de studenten pas prestaties konden leveren, nadat er met motivatie gemanipuleerd was. Op de derde plaats heeft zij haar onderzoek zo opgezet, dat de twee groepen die ze vergeleek, in elk geval niet qua intelligentie van elkaar verschilden. Zouden ze ook niet verschillen in andere kenmerken, maar alleen in het kenmerk wel of niet gemotiveerd zijn, dan kan zij stellen dat het verschil in prestatie

veroorzaakt is door een verschil in motivatie. Op de wijze waarop men de groepen ook nog op andere kenmerken vergelijkbaar kan samenstellen, komen we later nog terug.

De methodoloog legt uit dat de psychologe in feite een *experimenteel onderzoek* heeft uitgevoerd en de socioloog een *correlatieel onderzoek*. De betekenis van deze termen wordt in het volgende nog toegelicht.

Met behulp van beide methoden kan men aantonen dat een kenmerk X met kenmerk Y samenhangt. In bovenstaand voorbeeld is er een samenhang tussen de kenmerken motivatie en prestatie; leerlingen/studenten die wel gemotiveerd zijn leveren betere prestaties dan leerlingen/studenten die weinig of niet gemotiveerd zijn. Een samenhang tussen twee kenmerken wil echter nog niet zeggen dat het ene kenmerk het andere beïnvloedt. Voordat men daartoe kan concluderen, moet nog aan twee andere voorwaarden voldaan zijn.

Op de eerste plaats moet men er zeker van zijn dat kenmerk X in tijd aan kenmerk Y voorafgaat. In ons voorbeeld zou men ook kunnen stellen dat slechte prestaties de motivatie doen verminderen. In dat geval gaat het kenmerk prestatie in tijd aan het kenmerk motivatie vooraf en niet omgekeerd. Een ander voorbeeld. Het blijkt dat de jeugdcriminaliteit in steden waarin veel agogen werken, groter is dan in steden waarin relatief minder agogen werkzaam zijn. Het zou te ver gaan hieruit te concluderen dat die agogen de jeugdcriminaliteit veroorzaken. Reëler lijkt het te veronderstellen dat juist vanwege de hoge criminaliteit in bepaalde steden er extra agogen worden aangesteld om de jeugd op te vangen. Nu is in experimenteel onderzoek gemakkelijker te voldoen aan de eis dat kenmerk X in tijd aan kenmerk Y vooraf moet gaan, dan in correlatieel onderzoek. Het is voor experimenteel onderzoek juist kenmerkend dat men eerst met kenmerk X manipuleert en pas daarna nagaat wat er met kenmerk Y gebeurt. Bij correlatieel onderzoek wordt niet gemanipuleerd. Men stelt bij een aantal personen vast hoe zij scoren op de kenmerken X en Y, en vervolgens gaat men na of kenmerk X met Y samenhangt. De waarden van de kenmerken X en Y worden derhalve op eenzelfde tijdstip vastgesteld. Men kan dan alleen op goede gronden beredeneren dat X aan Y vooraf gaat. Is kenmerk X bijvoorbeeld leeftijd en Y de mate waarin men progressieve politieke denkbeelden heeft, dan moet men wel concluderen dat X in tijd voor Y komt. Leeftijd kan wel iemands denkbeelden beïnvloeden. Het omgekeerde is niet denkbaar.

Naast samenhang en een duidelijke tijdvolgorde moet echter nog aan een derde voorwaarde voldaan zijn om te kunnen stellen dat kenmerk X mede van invloed is op kenmerk Y. Men moet er zeker van zijn dat Y ook werkelijk mede door X beïnvloed wordt. In het onderzoek van de socioloog was er die zekerheid niet. Het was denkbaar dat de verschillen in prestatie niet door verschillen in motivatie, maar door verschillen in bijvoorbeeld intelligentie veroorzaakt zijn. In het onderzoek van de psychologe was deze alternatieve verklaring niet mogelijk, omdat de groepen die zij vergeleek, niet verschilden wat intelligentie betreft. Maar de groepen mogen behalve in intelligentie ook niet verschillen in andere kenmerken. Pas dan kan men stellen dat het verschil in prestatie door het verschil

in motivatie veroorzaakt is. In experimenteel onderzoek is men veel beter in staat het onderzoek zo op te zetten, dat alternatieve verklaringen niet plausibel zijn. Men stelt immers zelf de groepen samen die men wil vergelijken. Als men nu deze groepen zo samenstelt dat zij in een aantal kenmerken die ook op Y van invloed kunnen zijn, vergelijkbaar zijn en alleen in motivatie verschillen, dan moet een verschil in prestatie wel het gevolg zijn van een verschil in motivatie.

Bij correlatieel onderzoek kan men de groepen die men wil vergelijken, niet zelf samenstellen. Het gevolg is dat de groepen zeer waarschijnlijk niet alleen verschillen in het kenmerk X, waarvan men wil nagaan of het van invloed is op Y, maar ook in andere kenmerken. Verschillen in Y kunnen dan het gevolg zijn van verschillen in X of van verschillen in andere kenmerken of van beide. Dit is reeds toegelicht aan de hand van het voorbeeld over motivatie en prestatie. Een ridicuul voorbeeld werkt in deze wellicht meer illustratief.

Mensen die vaak naar de dokter gaan, blijken eerder dood te gaan dan mensen die weinig bij een dokter komen. Er is dus een samenhang tussen de frequentie waarmee mensen naar doktoren gaan en de kans te sterven. Ook de tijdvolgorde vormt geen probleem. Als men dood is, kan men immers niet meer naar de dokter toe. Aan de derde voorwaarde is gelukkig echter niet voldaan. Het zal duidelijk zijn dat het frequent bezoeken van een arts geen doodsoorzaak is. Aannemelijker lijkt het dat mensen die vaak gaan, over het algemeen ernstiger ziek zijn dan mensen die weinig gaan. Het is juist dit verschil in de mate van ziek zijn dat oorzaak is van het verschil in sterftেকans. In vakjargon wordt dan opgemerkt dat de oorspronkelijke samenhang tussen 'frequentie doktersbezoek' en 'de kans te sterven' wordt verklaard door een derde kenmerk, namelijk de ernst van de ziekte. De kans dat zo'n derde kenmerk de feitelijke oorzaak is, is in correlatieel onderzoek veel groter dan in experimenteel onderzoek. In deze laatste vorm heeft men veel meer mogelijkheden deze 'derde kenmerken' onder controle te houden, omdat men zelf de groepen zo kan samenstellen dat zij op deze kenmerken vergelijkbaar zijn.

De onderwijskundige is enige tijd stil. Hij overdenkt het bovenstaande en zegt ten slotte aarzelend tegen de methodoloog: „Is het dan niet beter alleen maar experimenteel onderzoek te doen?” Het antwoord is ontkennend. Op de eerste plaats kan men niet manipuleren met alle kenmerken waarvan men bepaalde effecten verwacht. Zo kan men geïnteresseerd zijn in de effecten van gezinsgrootte, gezinsvolledigheid, opvoedingspatronen, wijze van gezagsverdeling tussen vader en moeder op de ontwikkeling van kinderen. Om ethische redenen is het niet geoorloofd met deze kenmerken te manipuleren. Men kan niet aan bepaalde groepen voorschrijven dat zij maar één kind mogen nemen of dat zij autoritair moeten opvoeden, en aan andere dat zij een groot gezin moeten stichten of hun kinderen anti-autoritair moeten benaderen. Men kan wel autoritair met anti-autoritair opvoedende ouders vergelijken en nagaan of bijvoorbeeld de cognitieve ontwikkeling van het kind in de laatste groep anders verloopt dan in de eerste. Stel dat dit het geval is, dan kan men niet zonder meer stellen dat het verschil in

opvoeding oorzaak is van het verschil in ontwikkeling. Wellicht wordt een anti-autoritaire opvoeding meer gepraktiseerd door ouders die gemiddeld een hogere opleiding hebben genoten dan door ouders wier opleidingsniveau lager is. Wellicht passen deze laatsten eerder een autoritaire opvoeding toe. Een verschil in cognitieve ontwikkeling bij de kinderen hoeft dan niet veroorzaakt te zijn door een verschil in opvoedingsstijl, maar kan ook toe te schrijven zijn aan het verschil in opleidingsniveau van de ouders. Misschien verschillen autoritaire en anti-autoritaire ouders nog wel in andere kenmerken van elkaar en zijn die kenmerken oorzaak van de geconstateerde verschillen in cognitieve ontwikkeling. Hoe het ook zij, wil men een antwoord kunnen geven op de vraag of de opvoedingsstijl oorzaak of mede-oorzaak is van cognitieve ontwikkeling, dan zal men er in ieder geval voor moeten zorgen tijdens het onderzoek gegevens te verzamelen over de andere kenmerken waarvan men veronderstelt dat zij mede of in plaats van opvoedingsstijl de verschillen in cognitieve ontwikkeling verklaren. We willen het op dit moment bij deze opmerking laten. Op de vraag hoe men de extra gegevens kan gebruiken om vast te stellen of deze ook of alleen van invloed zijn, gaan we nog gedetailleerd in.

Er is nog een tweede reden waarom onderzoekers in bepaalde situaties correlatieel onderzoek verrichten. Uit eigen ervaring weten we dat veel verschijnselen multicausaal bepaald zijn: zij worden door meer dan één factor beïnvloed. In experimenteel onderzoek is men vaak geïnteresseerd in de invloed van slechts één factor. De onderzoeker weet wel dat het verschijnsel dat hij onderzoekt, door meer factoren beïnvloed kan worden. Hij wil echter vaak de invloed van slechts één factor vaststellen. Bij correlatieel onderzoek is een onderzoeker meer geïnteresseerd in de invloed van verscheidene factoren op een bepaald verschijnsel. Het volgende voorbeeld dient ter illustratie.

In een onderzoek onder vrouwelijke gezinshoofden die een periodieke uitkering krachtens de Algemene Bijstandswet genoten, waren Van Bijsterveldt en ik geïnteresseerd in de vraag welke factoren van invloed zijn op het feit dat sommige vrouwen direct nadat zij alleen zijn komen te staan een beroep op deze wet doen, terwijl anderen daarmee veel langer wachten (Van Bijsterveldt en Janssens, 1974). Men kan zich voorstellen dat ter verklaring hiervoor meer factoren aan te wijzen zijn. Zo is het mogelijk dat vrouwen die direct bijstand aanvragen, beter op de hoogte zijn van de mogelijkheden daartoe of minder barrières voelen om naar de Sociale Dienst te gaan, of over geen andere bestaansmiddelen beschikken, of minder mogelijkheden op een alternatieve inkomensbron hebben of minder bereid zijn daarnaar te zoeken, dan vrouwen die langer wachten alvorens een aanvraag in te dienen. In de voorgaande zin zijn vijf mogelijke oorzaken genoemd. Men kan nu het relatieve belang van elke factor in een correlatieel onderzoek vaststellen. Op de vraag hoe, gaan we later nog in.

Een derde reden ten slotte om in bepaalde situaties voor correlatieel onderzoek te kiezen, is de overtuiging dat de resultaten ervan op veel meer mensen van toepassing zijn dan die van experimenteel onderzoek. Om dat toe te lichten gaan we terug naar het onderzoek van de psychologe. Daaruit bleek dat motivatie van in-

vloed was op prestatie. Maar mag men stellen dat deze conclusie geldt voor bouwvakkers, tweedejaars studenten, filosofen, ambtenaren, regisseurs, bejaarden, vrouwen en/of kinderen? Neen. Voorlopig is in het experimentele onderzoek alleen vastgesteld dat dit waarschijnlijk opgaat voor eerstejaars studenten in de psychologie en dit onder de conditie dat de twintig studenten die aan het experiment deelnamen, een goede afspiegeling vormen van de totale groep eerstejaars studenten in de psychologie. Geheel anders ligt de zaak bijvoorbeeld bij het onderzoek onder vrouwelijke gezinshoofden in bijstand. Ten tijde van het onderzoek bestond deze totale groep in Nederland uit \pm 30.000 vrouwen. Daarvan zijn er 1200 voor het onderzoek benaderd. Zijn deze 1200 representatief voor de totale groep van 30.000, dan mag men de resultaten die men bij deze 1200 vrouwen vindt, van toepassing verklaren op alle vrouwen die een uitkering krachtens de Algemene Bijstandswet genieten. Dit laatste overigens binnen bepaalde grenzen waarover we nog zullen spreken. De resultaten van dit correlatieve onderzoek hebben derhalve een veel groter bereik dan die van het experimenteel onderzoek. Dit hoeft overigens niet altijd het geval te zijn. Grosso modo kan men echter stellen dat men bij correlatieve onderzoek de resultaten op veel meer mensen van toepassing kan verklaren dan bij experimenteel onderzoek. Daar komt bovendien als voordeel bij dat men bij correlatieve onderzoek een veel groter aantal mogelijke oorzaken van een verschijnsel in het onderzoek betreft dan bij experimenteel onderzoek. In het laatste type onderzoek is het daarentegen weer relatief gemakkelijker van een veronderstelde oorzaak vast te stellen of het inderdaad een oorzakelijke factor is.

De onderwijskundige is na deze uiteenzetting mateloos geïnteresseerd geworden en stelt nog een aantal vragen. Hij begrijpt nu wel de pros en de contras van beide benaderingen, maar weet nog niet wanneer men kan stellen dat het ene kenmerk met het andere samenhangt. Natuurlijk in zijn algemeenheid snapt hij ook wel dat, als gemotiveerde leerlingen betere prestaties behalen dan niet gemotiveerden, er samenhang is tussen motivatie en prestatie. Maar als die niet gemotiveerde leerlingen of studenten nu eens niet gemiddeld 4,6 hadden gehaald, maar 7,2, dus 0,1 minder dan de wel gemotiveerden, mag hij dan ook de conclusie trekken dat motivatie met prestatie samenhangt? Immers de gemotiveerden doen het dan toch ook beter dan niet gemotiveerden. Weliswaar niet veel beter, maar in elk geval toch iets.

De methodoloog antwoordt hierop dat men voor die vraag de hulp in moet roepen van een statisticus. In de statistiek zijn methoden ontwikkeld om bijvoorbeeld na te gaan of men aan een gevonden verschil veel, weinig of geen waarde moet hechten. Overigens hoeft men de statisticus in dit geval alleen te raadplegen, als niet alle mensen die voor het onderzoek in aanmerking komen, erin betrokken zijn. Zou de psychologe bijvoorbeeld voor alle eerstejaars studenten in de psychologie hebben willen nagaan of motivatie prestatie beïnvloedt, dan had zij alle studenten in haar onderzoek kunnen betrekken. De helft daarvan zou zij hebben moeten motiveren, de andere helft niet. Daarna zou zij voor beide sub-

groepen de prestaties hebben moeten vaststellen. Behalen de gemotiveerden betere prestaties dan de niet gemotiveerden, hoe klein het verschil tussen beiden wellicht ook is, dan moet zij concluderen dat in deze onderzoeksgroep motivatie met prestatie samenhangt. Is het verschil in prestatie gering tussen beide groepen, dan kan zij stellen dat de samenhang zwak is. Hoe groter het verschil, des te sterker de samenhang tussen motivatie en prestatie. Er is geen objectief criterium om te beslissen of een samenhang zwak of sterk is. Dit hangt af van het subjectieve oordeel van de onderzoekster en van degenen die van de onderzoeksresultaten gebruik gaan maken.

De psychologe heeft echter niet alle eerstejaars studenten in de psychologie in haar onderzoek betrokken. Zij heeft willekeurig twintig studenten voor haar onderzoek geselecteerd en vervolgens een verschil gevonden in prestatie tussen studenten die door haar gemotiveerd zijn en studenten die niet gemotiveerd zijn. Het verschil bedroeg $7,3 - 4,6 = 2,7$. Misschien zou ze geen verschil gevonden hebben, als ze toevallig twintig andere studenten in haar onderzoek betrokken had, of een verschil van 0,1 of van 1,3 of van 2,2. De statisticus redeneert in deze situatie als volgt. Misschien vindt men in die kleine groep van twintig wel een verschil, terwijl men, als de totale groep in het onderzoek betrokken was, geen verschil zou vinden. Hij gaat er nu vanuit dat er geen verschil is tussen gemotiveerde en niet gemotiveerde studenten qua prestatie. Dus, dat in werkelijkheid motivatie niet met prestatie samenhangt, dat in de totale groep de gemotiveerde studenten geen betere prestaties behalen dan niet gemotiveerde studenten. Wat is dan de kans dat de psychologe bij haar twintig studenten een verschil van 2,7 vindt? Is die kans erg klein, bijvoorbeeld kleiner dan vijf procent (of kleiner dan één procent afhankelijk van hoe streng het criterium is dat gehanteerd wordt), dan concludeert de statisticus dat het gevonden verschil van 2,7 *statistisch significant* is. Dat wil zeggen dat het erg waarschijnlijk is dat, ook als de psychologe alle eerstejaars studenten in de psychologie in haar onderzoek betrokken had, zij een verschil tussen gemotiveerde en niet gemotiveerde studenten gevonden zou hebben. De statisticus gaat dus na hoe groot de kans is, dat een verschil dat men vindt in een kleinere groep, ook geldt voor de totale groep waaruit die kleine groep voor het onderzoek is geselecteerd. Hij ziet een verschil bij de personen bij wie onderzoek is gedaan. Vervolgens berekent hij hoe groot de kans op zo'n verschil is, als er in werkelijkheid in de grote groep niet zo'n verschil zou zijn. Is de kans klein, dan concludeert hij dat het in de kleine groep gevonden verschil waarschijnlijk ook geldt voor de grote groep. In ons voorbeeld wil dat zeggen dat waarschijnlijk voor alle eerstejaars studenten in de psychologie geldt dat, als zij gemotiveerd worden, zij ook betere prestaties leveren dan studenten die niet gemotiveerd worden.

We hebben reeds opgemerkt dat de statisticus in dat geval spreekt van statistisch significant. Dat wil alleen zeggen dat een verschil in een kleine groep waarschijnlijk ook geldt voor de grotere groep waaruit de kleine groep voor het onderzoek geselecteerd is. Op de eerste plaats moet in deze zin de nadruk gelegd worden op het woord *waarschijnlijk*. De statisticus kan het immers mis hebben.

Hij heeft berekend dat de kans om bij een groep van twintig studenten een verschil van 2,7 aan te treffen, zo klein is – kleiner dan vijf respectievelijk één procent – dat het erg onwaarschijnlijk is dat men geen verschil zou vinden, als alle studenten in het onderzoek betrokken waren geweest. Hij weet het echter niet zeker. Er is nog altijd een kans, zij het een kleine – kleiner dan vijf respectievelijk één procent – dat er in feite geen verschil is in de totale groep. Dat dus gemotiveerde studenten in feite geen betere prestaties zouden behalen dan niet gemotiveerden. De statisticus kan alleen met meer dan 95 procent of 99 procent zekerheid zeggen dat er waarschijnlijk een samenhang bestaat tussen motivatie en prestatie.

Absolute zekerheid heeft hij nooit en kan hij ook niet geven aan een onderzoeker. Op zich is dat geen probleem. De zekerheid waarmee uitspraken gedaan worden, lijkt voldoende hoog. Gevaarlijk is het daarentegen elk statistisch significant verschil als inhoudelijk relevant te beschouwen. In ons voorbeeld vonden we dat gemotiveerde studenten 7,3 scoorden en niet gemotiveerde studenten 4,6. In ons cijfersysteem beschouwen we deze scores als respectievelijk ruim voldoende en onvoldoende. Is het verschil tussen 7,3 en 4,6 statistisch significant – dat wil zeggen er is een grote kans dat men ook in de totale groep een verschil gevonden zou hebben – dan kan men stellen dat gelet op onze cijferwaardering het verschil ook inhoudelijk relevant is. Het verschil stelt iets voor, het maakt indruk.

Maar welke conclusie hadden we moeten trekken, als gemotiveerde studenten 7,3 hadden behaald en niet gemotiveerde studenten 6,8 en als ook dit verschil statistisch significant zou zijn? In elk geval mag men dan concluderen dat, zou men de totale groep onderzocht hebben, er ook daarin tussen gemotiveerde en niet gemotiveerde studenten een verschil in prestatie gevonden zou zijn. Zeer waarschijnlijk althans. Maar wat stelt het verschil inhoudelijk voor? Wat doet men er in de praktijk mee? Niets waarschijnlijk. Rondt men beide cijfers af, dan krijgen beide groepen een 7. Een statistisch significant verschil moet men derhalve niet altijd als inhoudelijk relevant beoordelen. Wanneer dan wel? Op deze vraag is geen ondubbelzinnig antwoord te geven. Dat hangt opnieuw af van de visie van de onderzoeker en van de visie van degenen die van de onderzoeksresultaten gebruik maken. Zij beoordelen, of een statistisch significant verschil inhoudelijk relevant is en daarmee kan men het al dan niet eens zijn. In het algemeen geldt overigens wel dat hoe groter het verschil is, des te meer waarde men eraan kan hechten.

Op zijn instituut had de onderwijskundige wel eens gehoord dat onderzoek doen knap ingewikkeld kon zijn. Hij had daaruit de conclusie getrokken dat mensen die zich daarmee bezig hielden, wel tot een absolute waarheid moesten komen. Uit het verhaal van de methodoloog is hem gebleken dat dat meestal niet het geval is. Ook onderzoekers kunnen slechts uitspraken doen die alleen met een grote mate van waarschijnlijkheid juist zijn. Hij begrijpt ook dat men van resultaten die onderzoekers vinden bij kleinere groepen, niet zo maar kan stellen dat die ook gelden voor de grotere groep waarvan die kleinere groepen een afspiegeling