



# **Marketing en het Marktonderzoekproces**

**1998**

[www.johanblomme.net](http://www.johanblomme.net)

Johan Blomme  
Leenstrat 11  
8340–Damme

## Marketing en Marktonderzoek in een verander(en)de wereld

### 1. Maatschappelijke transformatie

De afgelopen decennia ontwikkelen moderne Westerse maatschappijen zich gestadig naar postindustriële, vooral op dienstverlening gerichte samenlevingen. Door mondialisering is de wereld in economisch opzicht één grote markt geworden. Kenmerkend voor de overgang van arbeidsintensieve naar kennisintensieve economieën is het steeds toenemend belang van *kennis als economische productiefactor*. De explosieve ontwikkeling van de informatie- en communicatietechnologie fungeert in dit opzicht als een katalyserende factor. Veranderingen die reeds op gang waren gebracht, voltrekken zich door de stuwende kracht van nieuwe technologieën in een hoger tempo.

Het is belangrijk te onderkennen dat er ingrijpende ontwikkelingen zijn die de contouren vormen voor een zicht ontwikkelende *kennis- en netwerkeconomie*. Aan het begin van de 21<sup>ste</sup> eeuw staan een viertal trends centraal die al sinds enige tijd werkzaam zijn.

Een eerste trend is de voortdurende *internationalisering van de economie*. Dit is geen nieuw verschijnsel, maar wel een proces dat steeds verder voortschrijdt. Allereerst zijn formele en informele barrières voor internationale transacties sterk verminderd. Voor de EU is de komst van de euro een belangrijke nieuwe mijlpaal in het internationaliseringsproces. Als gevolg van de vermindering van barrières zijn de stromen van goederen, diensten en kapitaal sterk toegenomen. Hoewel het grootste deel van internationale transacties zich nog steeds tussen de economische grootmachten afspeelt, doet een groeiend aantal landen mee aan het proces van internationalisering. Economische transacties voltrekken zich derhalve meer en meer in een mondiaal kader. De internationaliseringstrend geeft concurrentie een meer en meer grensoverschrijdend karakter en plaatst bedrijven, burgers en overheden in een meer dynamische omgeving.

Niet alleen op economisch vlak, ook op cultureel vlak vallen grenzen weg. De toegenomen welvaart heeft zich o.m. vertaald in een *culturele diversificatie*, met een pluriformiteit aan levensstijlen, culturele uitingen, vrijetijds- en consumptiepatronen. Deze verscheidenheid tekent zich niet alleen af op maatschappelijk niveau. Culturele verscheidenheid leidt eveneens tot een grotere verscheidenheid aan individuele identiteiten. Door *het proces van individualisering* zien individuen zichzelf minder als onderdeel van vaststaande collectiviteiten en meer als zelfstandige, zelfbewuste individuen met zelfgekozen sociale verbindingen en netwerken. In plaats van de verzuilde en inkomensafhankelijke ordening van de maatschappij, zien we een veelheid van nieuwe groepen en organisaties ontstaan, geordend langs categorieën als leeftijd, sekse, land van herkomst, milieubewustzijn, lifestyle, seksuele voorkeur, e.d. De individualiseringstrend heeft ook gevolgen voor het consumentengedrag en ligt daarmee ook aan de basis van veranderingen in de productieketen. De consument komt steeds meer achter het stuurwiel te zitten van ontwikkelingen in de productieketen. De consument heeft steeds individuelere behoeften, die ook nog snel veranderen, en kan door de internationalisering en het groeiende aanbod van goederen en diensten via het internet, uit een steeds groter arsenaal van leveranciers kiezen. Producenten moeten terdege rekening houdend met de veeleisende en geëmancipeerde consument.

Een derde belangrijke trend is die van de *informatisering*. In het bijzonder de ontwikkelingen op het vlak van informatie- en communicatietechnologieën (ICT) liggen aan de basis van een

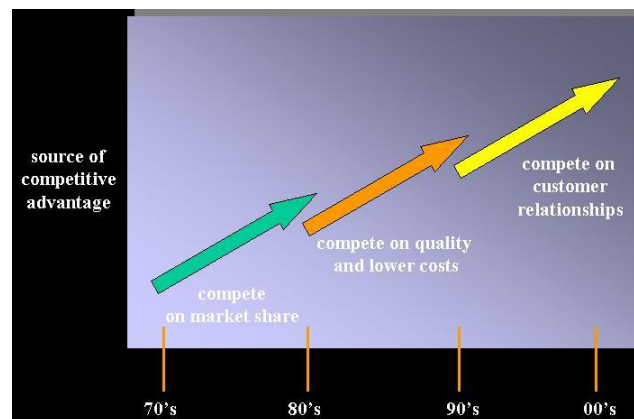
zich ontwikkelende *kennis- en netwerkeconomie*. Met deze laatste wordt verwezen naar een economie waarin het genereren van kennis en de capaciteit om informatie te verwerken en te verspreiden van fundamenteel belang zijn voor economisch succes. Kennis, het beschikken over informatie en het adequaat kunnen omgaan met informatie wordt steeds waardevoller en een doorslaggevende factor in concurrentie tussen bedrijven, tussen landen en tussen werknemers op de arbeidsmarkt. Als een consequentie hiervan wordt de beschikking over technologieën die de informatiestroom beter kunnen doen verlopen (ICT) van cruciaal belang.

Het is belangrijk te onderkennen dat naast internationalisering, individualisering en informatisering ook belangrijke *demografische veranderingsprocessen* plaatsgrijpen. Als gevolg van de aanhoudende stijging van de levensverwachting is er het proces van vergrijzing. Dit betekent dat een groeiend deel van de bevolking afhankelijk zal zijn van een al dan niet zelf opgebouwde ouderdagvoorziening. Anderzijds hebben de nieuwe ouderen vaak een relatief hoog besteedbaar inkomen, wat allerlei consumptiepatronen verandert. Het tweede proces binnen de demografische ontwikkeling is dit van de totstandkoming van een multiculturele samenleving als gevolg van de instroom van migranten.

## 2. Marketing

Hoe is, onder invloed van de verschillende transformaties die onder de noemer van de kennissamenleving schuilgaan, het vak marketing het laatste decennium veranderd? Aan het begin van de jaren negentig werd door de *Gartner Group* vooropgesteld dat het overwinnen van veranderingen de hoofdtaak van de bedrijfsvoering zal uitmaken. Kortere productlevenscyclussen, een toenemende concurrentiedynamiek en de vaststelling dat veel markten veranderd zijn van een verkoopmarkt in een koopmarkt, zijn slechts enkele factoren die aangeven dat de marktsituatie voor veel bedrijven een stuk moeilijker geworden is.

Bepalend voor de veranderingen is de afnemende rol van de traditionele marketing mix-instrumenten. Deze blijken onvoldoende houvast te bieden in de snel veranderende markten. Door te sturen vanuit speciaal daartoe gebouwde *marketing databases*, nemen de mogelijkheden op effectieve marktwerking aanzienlijk toe, mits de data voldoen aan de vereiste criteria. Daarbij stelt de techniek ons in toenemende mate in staat om in grote markten een individuele benadering toe te passen, aansluitend op de in een database opgebouwde kennis van klanten en prospects. Aanvankelijk ging het om gepersonaliseerde massacommunicatie, waarbij eenzelfde boodschap wordt verstuurd aan individueel geadresseerde personen. Maar ook massale persoonlijke communicatie vindt thans op grote schaal plaats. Hierbij wordt de boodschap individueel afgestemd op de kenmerken die van de ontvanger bekend zijn. Daardoor kan massamarketing in consumentenmarkten worden vervangen door *één-op-één-marketing*. Informatie op individueel prospect/klant-niveau verdringt aldus de meer generieke modellen en veronderstellingen. Er ontstaat m.a.w. grote aandacht voor "customer care": de relatie met de individuele klant. Het *managen van klantrelaties* wordt daarmee een essentieel onderdeel van marketing. Massamarketing is daarmee niet overleden, maar wel van sterk afnemend belang geworden. Klanten laten zich niet meer alles zeggen en welgevallen. Het individuele gedrag is moeilijk nog in doelgroepen samen te vatten. De relatie met de kritische en mondige klant komt bovenaan het lijstje van prioriteiten van de marketeer te staan.



Figuur 1 : Verschuivingen in het aandachtsveld binnen het vak marketing

De technologische ontwikkelingen stellen ons in staat klanten ook zelf te laten reageren naar de aanbieders, om hun wensen en keuzen kenbaar te maken. Het eenzijdig bombarderen vanuit aanbiederzijde wijkt voor een dialoog tussen aanbieder en afnemer. Het is niet meer de leverancier die roept, en de klant die dat allemaal over zich heen laat gaan ; de klant stelt vragen waarnaar het bedrijf luistert en waar het op reageert. Kwaliteit zal in de hele klantenbenadering kwantiteit als hoofdprioriteit verdringen. De laatste jaren is veel geschreven over de gewijzigde omgeving waaraan de marketing zich moet aanpassen. Individualisering, demassificatie, marktversplintering, fragmentatie, graascultuur, e.a. zijn tegenwoordig vaak gebruikte termen die marketeers hanteren om de hedendaagse consument te omschrijven. Er wordt ook verwezen naar kortere productlevenscyclussen, intensivering van marktsegmentatie, toenemende concurrentiedynamiek, e.d. Producten gaan steeds meer op elkaar lijken (en worden gemakkelijker gekopieerd) terwijl de concurrentie heviger wordt. Bedrijven moeten zich daarom in toenemende mate concentreren op de wijze waarop zij hun dienstverlening organiseren. Een goed functionerende organisatie is immers veel moeilijker te kopiëren. Kwaliteit zal in de hele (klanten)benadering kwantiteit als hoofdprioriteit verdringen. In deze context heeft Schultz (1990) het over de noodzaak van een nieuw type marketing. Hij omschrijft het als volgt : "In the past, traditional marketeers, in an effort to organize their marketing activities, tried to aggregate customer wants and needs. In other words, they tried to find common products and concepts and areas which would appeal to large numbers of people who could be served economically and efficiently with a product of service. Because customers and their social norms were rather homogeneous, that management approach resulted in the development of the mass market for consumer products. ... Today, there's a new look to the marketplace. A new look to consumers. ... Today and increasingly tomorrow, the market for many consumer products will be driven by time, not money. It is also driven by a desire by people to be different and unique, not more like others. ..., we're moving from what was historically mass marketing to a new form of specialized marketing" (Schultz, 1990 : 12).

Wat uit het bovenstaande kan afgeleid worden is dat het meer dan ooit belangrijk wordt om de markt en de ondernemingsmissie duidelijk te definiëren. Schultz (1990 : 17 e.v.) heeft het in dit opzicht over het belang van strategisch management. Bij dit laatste dienen twee uitgangspunten centraal te staan : "The organization's planning view must be external, that is, toward customers and markets ... Strategic marketing plans must be based on identifying and making use of some sustainable competitive advantage" (Schultz, 1990 : 19). Om de kansen goed te kunnen benutten en bedreigingen te kunnen weren, zullen ondernemingen

slagvaardig moeten opereren. Dit vereist een goede informatielogistiek en informatie-reductietechnieken zodat alleen de relevante en de juiste informatie op het juiste moment beschikbaar is. Om een individuele marketingstrategie mogelijk te maken en om een interactieve communicatie te onderhouden, is het noodzakelijk dat relevante gegevens op individueel niveau worden geregistreerd. Databases met individuele gegevens zijn noodzakelijk voor een gerichte benadering en voor de transformatie van doelgroep naar individuele relaties. Geavanceerde analysetechnieken maken het mogelijk om het gedrag van individuen te aggregeren tot variabel samen te stellen clusters zodat het individuele beleid ook leidt tot een voldoende omvang van verkoop. Hierbij moet niet de indruk ontstaan dat marketing gedreven wordt door de ontwikkeling van de techniek. Het is eerder zo dat de ontwikkeling van de techniek bijdraagt tot een andere denkwijze over het te voeren marketingbeleid en dat de consument andere eisen stelt aan de communicatie van een organisatie en meer ondersteuning verlangt bij het gebruik van producten.

De individuele relatie op basis van data in marketing databases kan tot stand gebracht worden omdat technologische ontwikkelingen niet alleen grootschalig datagebruik mogelijk maken, maar ook een nieuwe dimensie scheppen voor de communicatie tussen aanbieder en afnemer. Werd in de tijd van de overheersende massamarketing een medium vooral als advertentiemedium gebruikt, dan worden media in toenemende mate gebruikt om één-op-één-relaties te leggen met klanten en prospects. De massamedia zullen in de (direct) marketing een plaatsje moeten opschuiven voor de intussen steeds drukker bereden *informatiesnelweg*. Niet in het minst omdat het een bij uitstek *interactief medium* is, waardoor de gewenste "direct response" vanzelfsprekend wordt. Lester Wunderman die in de jaren zestig het begrip 'direct marketing' introduceerde als uitbreiding van het direct mail-concept, heeft het in dit opzicht over het gebruik van het wereldwijde Internet.

De voortschrijdende ontwikkeling van direct marketing, o.a. via Internet, is volgens Wunderman een veruitwendiging van een nieuw sociaal-economisch model, namelijk dat van de *informatiemaatschappij* waar we opnieuw tot een producent-consument-relatie komen zoals we die hadden vóór de industriële revolutie. De markten waren toen klein, zowel in omvang als in geografische uitgestrektheid. Dit maakte dat de producenten en verkopers nauwkeurig de koopgewoonten en productvoorkeuren van elke klant kenden. Aan dit op landbouw geënte economisch model kwam na 1840 een einde met de uitvinding van de machines. Die stelden de producent in staat op grote schaal te mechaniseren.

Het gevolg was seriewerk, daling van de prijzen, en uiteindelijk de ontwikkeling van grootschalige distributiesystemen zoals supermarkten en winkelketens en de aanwending van de massamedia om het aanbod bij een groot publiek kenbaar te maken.

De consument van de 21<sup>ste</sup> eeuw wil geen kuddebeest meer zijn. Hij wil benaderd worden als een uniek wezen en zal daarom veel belang hechten aan persoonlijke dienstverlening. Producten zullen diensten moeten worden. Ze zullen niet geëvalueerd worden op wat ze zijn, maar op wat ze voor de consument kunnen doen. De producent zal daarom opnieuw moeten leren elk van zijn klanten te begrijpen om van daaruit aan *loyalty marketing* te doen. Vanuit deze optiek gaan we in zekere zin terug naar de kleinschaligheid van de *face-to-face* verkoop, zoals we die hadden in de pre-industriële, agrarische maatschappij van begin negentiende eeuw. Dat betekent uiteraard niet dat we terug in de tijd reizen. De markten zijn wel degelijk omvangrijker geworden en zullen dat in se ook wel blijven. Alleen heeft de sterke opkomst van de computer het mogelijk gemaakt om aan de hand van databanken de markten opnieuw te verkleinen.

Samengevat is er sprake is van drie hoofdontwikkelingen.

1. Het sturen op basis van informatie in *marketing databases*, in plaats van generieke modellen.
2. Het gebruik van media om *transacties* te realiseren, in plaats van generieke (thema)advertenties.
3. Het sturen op *individuele klantrelaties* in plaats van op meer of minder gespecialiseerde doelgroepen.

Er zijn dus twee ontwikkelingen die ons in staat stellen om een klantrelatie op te bouwen : die vanuit de *informatiekant* en die vanuit de *mediakant*.

De *informatierevolutie* kan beschreven worden aan de hand van een viertal componenten. In de eerste plaats is er de digitalisering van zowat alles dat door de zintuigen kan opgevangen worden. In de tweede plaats is er de toename van de capaciteit van opslag, verwerking en overdracht van "enen en nullen". Er dient hier ook gewezen te worden op de toename van de capaciteit van datacompressie. Een derde component van de informatierevolutie wordt gevormd door het feit dat software meer en meer gebruiksvriendelijk wordt. Tenslotte creëren netwerken nieuwe communicatiestructuren.

Projecteren we de informatierevolutie op het gebruik van informatie voor commerciële processen, dan kunnen twee dimensies worden onderscheiden. De eerste dimensie is de ontwikkeling van informatie over de massa (mogelijke) klanten, naar gedefinieerde doelgroepen en tenslotte naar het kleinste denkbare niveau, het op naam en adres traceerbare individu. Dit laatste wordt gewoonlijk afgekort als NAW, wat staat voor naam-adres-woonplaats. Het is de dimensie *van algemeen naar specifiek en identificeerbaar*.

De tweede dimensie is de ontwikkeling van het type informatie dat wordt gebruikt. Aanvankelijk was het veelal informatie die werd afgeleid uit studies met uiteenlopende bronnen. Veelal was dat een combinatie van onderzoekgegevens, met oordelen van deskundigen leidend tot een eindoordeel. Ook dat eindoordeel kan elementen bevatten van persoonlijke interpretatie van de opstellers. Naast deze afgeleide informatie heeft zich het marktonderzoek ontwikkeld. Dit is overwegend informatie gebaseerd op opgegeven gedrag. Tegenwoordig verschuift het type informatie van afgeleide en opgegeven informatie naar *gedragsinformatie*. De informatietechnologie stelt ons in staat het gedrag vast te leggen op individueel niveau. Deze informatie is veruit het betrouwbaarst als het gaat om het voorspellen van toekomstig gedrag.

Nauw samenhangend met de informatierevolutie zijn de multimediale mogelijkheden in de marketing. Multimedia, interactieve media, elektronische snelwegen en interactieve marketing zijn termen die steeds vaker gebruikt worden om aan te geven dat de informatietechnologie nu gebruikt wordt als een middel om op een nieuwe manier te communiceren. De consument krijgt steeds meer vormen en methoden van communicatie tot zijn beschikking en ontwikkelt zo, op basis van zijn behoeften en wensen, zijn eigen wijze van interactie met de omgeving. Dit biedt voor de marketeer geheel nieuwe mogelijkheden om de consument te benaderen. De toepassingsmogelijkheden lijken schijnbaar eindeloos en de ontwikkelingen gaan razendsnel. De jaren tachtig werden sterk gedomineerd door steeds snellere, goedkopere en kleinere computers, alsmede de hieraan gerelateerde ontwikkeling van software. Programmatuur werd complexer en omvangrijker, maar tegelijk ook gebruiksvriendelijker en goedkoper. Hierdoor veranderde ook het marketingvak. Door de uitgebreidere communicatie-, databeheer- en analysemogelijkheden werd een dieper inzicht verkregen in de markt en de klant. In de jaren negentig zijn het vooral de communicatiemogelijkheden die tot een snelle verandering leiden.

Het is evenwel niet alleen de versnelde technologische vooruitgang die het ontwikkelingspad van de nieuwe media bepaalt. Ook de demassificatie van de

consumentenmarkt drukt zijn stempel op de ontwikkeling van nieuwe media. De door Faith Popcorn beschreven "cocooning" als een belangrijke trend in het hedendaags consumentengedrag, vormt een gunstige voedingsbodem voor de groei van nieuwe interactieve media.

Van nieuwe media in de echte zin van het woord kan niet gesproken worden. De meeste nieuwe interactieve mediavormen zijn ontstaan als uitbouw of integratie van de bestaande communicatievormen : televisie, telefoon en computer. Vandaar dat men de nieuwe media ook wel multimedia durft te noemen. Met nieuwe media wordt dus verwezen naar elk medium dat een directe, *interactieve communicatie* toelaat tussen klant of prospect en onderneming. Interactieve media worden in de eerste plaats gekenmerkt door informatieverstrekking op maat. De gebruiker gaat alleen in op datgene dat hem of haar interesseert. Een ander kenmerk van interactieve media is de vrijheid van timing. De gebruiker wordt niet op ongewenste momenten met een boodschap geconfronteerd. Het gebruik van het medium vindt plaats op het moment dat de gebruiker het zelf wil of op het moment dat daarvoor het meest geschikt is. Een interactief medium kan tenslotte de door gebruikers gegeven input opslaan. Het medium wordt dan een krachtig instrument om meer te weten te komen over het gedrag van klanten of prospects.

### 3. Marktonderzoek

Hoe kan, in het licht van de hierboven genoemde ontwikkelingen, de taak van marktonderzoek omschreven worden ? Als gevolg van de vermelde trends zijn vrijwel alle bedrijfsfuncties veranderd. Centraal in de verandering van de bedrijfsfuncties staat de overgang *van een bedrijfsorganisatie gericht op productie naar een organisatie gericht op het managen van kennis*. Kennis wordt daarmee voor het bedrijfsleven steeds meer de doorslaggevende concurrentiefactor. Om die reden dient zowel de functie als de organisatie van het marktonderzoek opnieuw moeten gedefinieerd te worden.

Ter aanduiding van hetgeen moet worden nagestreefd, is het begrip *kennismanagement* het meest geëigend. Kennismanagement betekent het beschikken over informatie en het vermogen om deze informatie, o.m. met behulp van ICT, adequaat aan te wenden. Naar het marktonderzoek toe betekent dit het volgende. De informatieverzameling door min of meer gestandaardiseerde ad hoc-onderzoeken moet vervangen worden door *een geplande, op de marketingtaken anticiperende opbouw van een informatiebasis*. Voor het marktonderzoek heeft dit als repercussie dat het sneller en flexibeler moet worden, en op nauwkeurig omschreven vragen moet gaan antwoorden. Ze voorziet de marketing permanent met informatie en wordt anderzijds door middel van feedback steeds geactualiseerd. De vandaag nog meer additieve, kwantitatieve verzameling van informatie moet door een integrale, kwalitatieve productie van informatie worden vervangen. Dit betekent vooral dat losse maatregelen van de informatiewinning (losse onderzoeken, opvragen databank, e.d.) al bij het ontwerpen zakelijk en in de tijd op elkaar moeten worden afgestemd. Informatie uit verschillende onderzoeken, van verschillende leveranciers, uit verschillende perioden moeten met elkaar in verband gebracht worden. Dit laatste is vooral van belang voor de scheidslijn tussen primair en secundair onderzoek. Uit tijds- en kostenoverwegingen zal de informatiebehoefte van de toekomst in versterkte mate door gebruik te maken van secundaire data moeten worden bevredigd. Voor primaire onderzoeken moet voor de volledigheid en voor de betere interpretatiemogelijkheden secundair beschikbare informatie gebruikt kunnen worden. Ze mogen niet alleen met het oog op de toevallig actuele vraagstelling worden beoordeeld, maar ook met het doel informatie voor toekomstige

vragen in te winnen. *Kennismanagement betekent informatieverstrekking tot ondersteuning van marketingbeslissingen.*

Gesteld kan worden dat de vanzelfsprekendheid van het marktonderzoek in toenemende mate zal veranderen. De praktijk van het marktonderzoek zal meer en meer onderhevig zijn aan kwalitatieve (informatie-)eisen van het marketing management :

- ◆ marktonderzoek moet geïntegreerd worden in het *marketing management* ;
- ◆ marktonderzoek moet toegevoegd worden aan het *kennismanagement*, niet (meer) met de functie van 'neuzenteller' maar met een zelfstandige taak binnen de eigen onderneming als marketing-adviseur ; dit zal er toe leiden dat marktonderzoek sterker dan voorheen zal betrokken worden bij de omzetting van empirische gegevens naar ondernemingsbeslissingen ;
- ◆ marktonderzoek zal op de toeleverende industrie, d.w.z. de marktonderzoekbureaus, grote invloed uitoefenen m.b.t. de verdere *ontwikkeling van onderzoeksmethoden* die aan de informatiebehoefte van de marketing voldoen.



## 1. Ontwikkelingen in marktonderzoek

In marketing als veld van (wetenschappelijk) onderzoek concentreert men zich op :

- ♦ het ontwikkelen van methoden en technieken om de problemen in het probleemgebied 'marketing' te kunnen beschrijven, verklaren of voorspellen ;
- ♦ het genereren van generaliseerbare kennis over bepaalde vraagstukken en verschijnselen die in het probleemgebied kunnen worden waargenomen.

Het wetenschappelijk onderzoek dat op de realisatie van deze doelstellingen is geënt, kan als volgt worden gekarakteriseerd:

- ♦ het onderzoek wordt verricht vanuit een relatief groot aantal disciplines ; hierbinnen heeft zich een aantal zgn. *wetenschappelijke specialismen* ontwikkeld ; deze wetenschappelijke specialismen zijn enerzijds betrokken op een gemeenschappelijk ervaringsobject, te weten het probleemgebied 'marketing', terwijl zij anderzijds deel uitmaken van een wetenschap die een eigen kenobject heeft (bv. commerciële bedrijfseconomie, consumentenrecht, psychologie van het consumentengedrag, arbeids- en industriële sociologie) ;
- ♦ bij het wetenschappelijk onderzoek maakt men veel gebruik van *kwantitatieve methoden en technieken* die in de econometrie en de sociale wetenschappen zijn ontwikkeld ; soms moet men bestaande methoden en technieken aanpassen vooraleer zij zich lenen voor toepassing op problemen in het probleemgebied 'marketing' ; de aandacht voor kwantitatieve methoden en technieken wordt weerspiegeld in het opzetten van verscheidene op dit terrein gespecialiseerde tijdschriften zoals de *Journal of Marketing Research* in 1963, de *Journal of Consumer Research* in 1974 en *Marketing Science* in 1982 ; volgens diverse auteurs zou de ontwikkeling en toepassing van kwantitatieve methoden en technieken zelfs de kern van marketing als veld van wetenschappelijk onderzoek vormen ;
- ♦ de uitkomsten van het onderzoek van vraagstukken in het probleemgebied 'marketing' zijn *fragmentarisch* van aard te noemen ; men richt zich meer op het analyseren van specifieke situaties (markten, specifieke problemen die vaak op één bepaalde klasse van marktinstrumenten betrekking hebben) en op het ontwikkelen, of beter het modificeren, van onderzoekstechnieken dan op het ontwikkelen van theorieën en generaliseerbare kennis ;
- ♦ naast het fragmentarisch, geïsoleerd onderzoek van marketingproblemen, dient ook gewezen te worden op de pogingen die ondernomen worden om de problemen in hun onderlinge samenhang te onderzoeken (in dit verband valt te denken aan de pogingen die ondernomen worden om te komen tot een *metatheorie* of "general theory of marketing").

### 1.1. Onderzoek gericht op het ontwikkelen van methoden en technieken

Met betrekking tot dit type onderzoek kan een onderscheid gemaakt worden tussen :

- ♦ onderzoek gericht op het verbeteren van methoden van marktonderzoek, d.w.z. de wijze waarop gegevens worden verkregen ;
- ♦ onderzoek gericht op de ontwikkeling of verbetering van technieken waarmee de verzamelde gegevens bewerkt en verwerkt kunnen worden (analysetechnieken) ;

- ♦ onderzoek gericht op het ontwikkelen van methoden ten behoeve van strategische (marketing)planning.

### 1.1.1. Het verbeteren van de methoden van marktonderzoek

Dit type onderzoek is o.m. gericht op het verhogen van de validiteit en betrouwbaarheid van gegevens die door middel van schriftelijke, telefonische en persoonlijke enquêtes (communicatie) worden verkregen. Vragen die hier beantwoord worden, hebben betrekking op o.m. het verminderen van de non-respons, het samenstellen van representatieve panels en het ontwikkelen van adequate schaaltechnieken. De validiteit en betrouwbaarheid van consumentenpanelgegevens wordt o.m. verbeterd door gebruik te maken van scanningsapparatuur ("homescan"). Scanning in de winkel, via het aflezen van de streepjescode ("bar code") die op verpakte producten is aangebracht door kassaterminals, kan gezien worden als een belangrijke ontwikkeling in het gebruik van observatiemethoden om marktonderzoekgegevens te genereren. Alhoewel de validiteit van de gegevens duidelijk verbetering behoeft, omdat de winkelpanelen waarmee de gegevens verkregen worden nog niet representatief zijn, is de betrouwbaarheid van de gegevens groot. Doordat de gegevens op een zeer gedisaggregeerd niveau (per variëteit, per dag, per winkel) gegenereerd worden, biedt het gebruik van scanning erg veel (nieuwe) mogelijkheden om responsmaatstaven (verkopen, marktaandeel) te bepalen. Scanninggegevens worden in onderzoek frequent gehanteerd om de effecten van acties (promoties) te bepalen. Promoties hebben veelal een effect dat zich slechts op korte termijn manifesteert. Deze effecten kunnen met behulp van scanninggegevens bepaald worden vanwege het niet-geaggregeerd karakter van deze gegevens. Stel dat een fabrikant samen met een detaillist of detailhandelsorganisatie een actie voor een bepaald merk j in winkel k voorbereidt. Het is dan van groot belang om te bepalen of deze actie een substitutie tussen merken binnen winkel k tot gevolg heeft of dat men in staat is om consumenten van merk j en andere merken i die in andere winkels dan winkel k dit product kopen te bewegen om merk j in winkel k te kopen (= methodiek om deze effecten van "brand substitution" en "store substitution" uit elkaar te halen). Scanninggegevens kunnen door middel van observatie in ongecontroleerde en in gecontroleerde experimenten worden verkregen. Op die manier is het mogelijk om met behulp van deze experimenten voorgenomen marketingprogramma's te evalueren. Deze experimenten staan bekend als "scantrack services".

### 1.1.2. Het ontwikkelen (verbeteren) van analysetechnieken

In de theorie en de in de praktijk van het marktonderzoek wordt een scala van technieken gehanteerd om aan de verzamelde gegevens informatie te ontfemen. In de praktijk wordt relatief veel gebruik gemaakt van chi-kwadraatanalyse, variantie-analyse en factoranalyse. Andere, wat minder frequent gehanteerde methoden zijn discriminantanalyse, AID, regressie- en correlatie-analyse, clusteranalyse, conjunct meten en meerdimensionale schaalanalyse. In het (wetenschappelijk) onderzoek wordt ook gebruik gemaakt van correspondentie-analyse. We zullen een voorbeeld geven waarmee we de toepassing van één van deze technieken kunnen illustreren.

Bij strategische planning gaat de aandacht o.m. uit naar de concurrentie. Met behulp van meerdimensionale schaalanalyse (MDS) kan de concurrentie in kaart gebracht worden. Daarbij kiest men voor een invalshoek waarbij men uitgaat van percepties van merkartikelen door consumenten. Consumenten beoordelen merkartikelen (maar bijvoorbeeld ook winkels) in het algemeen aan de hand van meerdere eigenschappen. De percepties van

consumenten m.b.t. de eigenschappen bepalen de posities van merkartikelen ten opzichte van elkaar. Met behulp van MDS kunnen de posities van merkartikelen worden weergegeven in een meerdimensionale ruimte, de zgn. merkenruimte ("brand map"). Daarbij worden de merkartikelen zodanig geplaatst dat merkartikelen die veel op elkaar lijken ook dicht bij elkaar liggen. Merkartikelen die weinig gelijkenis vertonen, worden daarentegen ver van elkaar geplaatst. In een MDS-studie wordt dienaangaande aan respondenten gevraagd hun voorkeuren m.b.t. de merken aan te geven (gaande van 'de meeste voorkeur' tot 'de minste voorkeur'). Op basis van deze data kunnen de respondenten in de merkenruimte geplaatst worden (de perceptuele ruimte). De positie van de respondent in de merkenruimte staat voor het ideaalpunt van de respondent. Het ideaalpunt is een denkbeeldig, hypothetisch merk met eigenschappen juist in die mate zoals dat door de betreffende respondent wordt verlangd (de preferentiële ruimte). Dit 'merkartikel' voldoet dus volledig aan de wensen van de respondent. Een merkenruimte waar de ideaalpunten aan zijn toegevoegd, noemen we de gemeenschappelijke ruimte (de gemeenschappelijke ruimte of "joint space"). Met behulp van de "joint space" kunnen uitspraken worden gedaan over o.m. marktsegmenten, 'gaten in de markt' en mogelijke substituties tussen merken. Ook correspondentie-analyse laat dergelijke uitspraken toe.

### 1.1.3. Het ontwikkelen van technieken ten behoeve van strategische planning

In het probleemgebied 'marketing' heeft men tal van concepten, methoden en technieken ontwikkeld en geïmplementeerd waarmee inhoud kan gegeven worden aan de strategische marketingconceptie en die gebruikt kunnen worden voor strategische marketingplanning. Deze methoden en technieken hebben o.m. betrekking op het aanduiden van sterkten en zwakten van een onderneming (S.W.O.T.-analyse) en het in kaart brengen van marktaandeelen (portfolio-analyse).

## 1.2. Onderzoek gericht op het genereren van generaliseerbare kennis

De hiervoor genoemde methoden en technieken kunnen worden gehanteerd om kennis m.b.t. vraagstukken in het probleemgebied 'marketing' te genereren. Het zou te ver voeren om op het resultaat van al deze inspanningen in te gaan. We volstaan met het noemen van enkele generalisaties.

Een belangrijk onderwerp is het segmenteren van markten in min of meer homogene groepen consumenten. Het onderzoek dat momenteel gericht is op het in kaart brengen van segmenten laat zich karakteriseren door het gebruik van segmentatievariabelen (variabelen waarmee men segmenten indeelt) die situatiegebonden (voor elk type product in principe andere segmentatievariabelen) en subjectief zijn. Voorbeelden van de laatst genoemde segmentatievariabelen zijn levensstijl en psychografische profielen.

In het veld van het wetenschappelijk onderzoek in het probleemgebied 'marketing' wordt veel aandacht geschonken aan de werking van reclame en het ontwikkelen van methoden om de effecten van reclamebeslissingen te bepalen. Daarbij gaat in de bedrijfspraktijk vrij veel aandacht uit naar het ontwikkelen van adequate responsmaatstaven zoals 'bekendheid' en 'herinnering' om effecten van reclame te bepalen.

## 2. HET MARKTONDERZOEKPROCES

Onder marketing research of marktonderzoek verstaan we *het systematisch verzamelen, analyseren en rapporteren van gegevens over problemen die verband houden met de marketing van goederen en diensten*. Marktonderzoekgegevens moeten systematisch worden verzameld en verwerkt. Systematisch houdt in : volgens een tevoren opgesteld plan. Onder het marktonderzoekproces verstaan we de stappen die de marktonderzoeker doet om het probleem waarmee hij/zij wordt geconfronteerd, op te lossen. Het geeft dus aan op welke wijze het onderzoek wordt aangepakt en uitgewerkt. Het onderzoekproces kan in drie fasen worden verdeeld, nl. de *planningsfase*, de *uitvoeringsfase* en de *rapporteringsfase*.

In marktonderzoek wordt de *wetenschappelijke methode* gevolgd. Het systematische van de wetenschappelijke methode bestaat hierin dat eenzelfde probleem tweemaal benaderd wordt. De eerste keer met de bedoeling informatie te verzamelen om een voorlopig besluit te trekken. Men noemt dit een hypothese. De tweede maal om genoeg gegevens te verzamelen om die hypothese(n) te kunnen aanvaarden of verwerpen. Het eerste noemt men exploratief onderzoek, het tweede conclusief onderzoek. Tijdens de planningsfase van het onderzoek, en met name bij het formuleren van de probleemstelling, onderneemt de marktonderzoeker in feite een *exploratief (verkenkend) onderzoek*. Men probeert zich immers in het probleem te oriënteren. Er bestaat geen vaste methode om exploratief onderzoek aan te vatten. Het is dan ook een erg soepele en weinig gestructureerde vorm van onderzoek. Creativiteit en flexibiliteit zijn noodzakelijk. Men werkt op een eerder informele manier. Door zich in het probleem te oriënteren krijgt de onderzoeker zicht op het probleem, kunnen (voorlopige) hypothesen opgesteld worden en wordt het eveneens duidelijk welke de informatiebehoefte zijn. Een hulpmiddel bij exploratief onderzoek is "desk research", waarbij secundaire gegevens (interne en externe) worden geraadpleegd. Voldoen secundaire gegevens niet aan de informatiebehoefte dan dienen *primaire gegevens* verzameld te worden.

Wanneer blijkt dat secundaire gegevens onvoldoende zijn voor de oplossing van een probleem, zijn we geneigd te zeggen : verzamel additionele gegevens. Alvorens we daartoe overgaan, moeten we ons de vraag stellen of de waarde van de additionele informatie wel opweegt tegen de kosten die nodig zijn om die informatie te verkrijgen. Het afwegen van de waarde versus de kosten van informatie kan plaatsvinden met behulp van de zogenaamde *Bayesiaanse analyse*. Het probleem van waarde versus kosten van informatie kunnen we als volgt in een vergelijking weergeven :

$$NVW = (VWO - VW) - C$$

NVW = de verwachte nettowaarde van het onderzoek

VWO = de verwachte waarde van de beslissing, gemaakt met behulp van onderzoek

VW = de verwachte waarde van de beslissing, gemaakt zonder onderzoek

C = de kosten van het onderzoek

Wanneer NVW positief is, kan onderzoek worden verricht. Er wordt in de formule duidelijk gesproken over verwachte waarden. Enige mate van onzekerheid blijft altijd over, ondanks het laten uitvoeren van een onderzoek. Het verschil tussen VWO en VW is dan ook op te vatten als een indicatie voor de mate waarin de onzekerheid door middel van onderzoek is gereduceerd. Met behulp van de Bayesiaanse analyse zullen we trachten een antwoord te geven op de vraag of NVW positief of negatief is.

Centraal staat de vraag of de kosten, verbonden aan het verwerven van extra gegevens wel worden goedge maakt door de waarde die aan de extra gegevens wordt ontleend. We onderscheiden in de Bayesiaanse analyse drie fasen, te weten : de *priori-analyse*, de *posteriori-analyse* en de *preposteriori-analyse*. In hetgeen volgt gaat onze aandacht enkel uit naar de *priori-analyse*. De reden hiervoor is dat de *posteriori-* en *preposteriori-analyse* in de praktijk nauwelijks worden gebruikt vanwege de vooronderstellingen die gemaakt moeten worden en de relatief complexe aard van de analyses.

We gaan er bij de Bayesiaanse analyse van uit dat de beslisser in staat is subjectieve waarschijnlijkheden toe te kennen aan de omgevingstoestanden. Tevens is hij in staat de resultaten te schatten bij elke combinatie van strategie en omgevingstoestand. Hij kan dus een resultatenmatrix opstellen. We kunnen de analyse toelichten aan de hand van een voorbeeld.

Een marketing manager wordt geconfronteerd met een dalende afzet van één van de producten van de onderneming. Door de directie wordt de mogelijkheid overwogen om een nieuw product op de markt te brengen. De marketing manager moet de directie nu adviseren of het nieuwe product al dan niet op de markt moet worden gebracht. In deze vereenvoudigde casus hebben we te maken met twee strategieën, nl. A1, het op de markt brengen van het nieuwe product en A2, het niet op de markt brengen van het nieuwe product. Bovendien onderscheiden we twee omgevingstoestanden, te weten S1, het introduceren wordt een succes en S2, het introduceren van het nieuwe product wordt een mislukking. De factoren die bepalen of het een succes wordt of een mislukking (bv. concurrentie-omstandigheden), laten we buiten beschouwing. De marketing manager is in staat een schatting te maken van de resultaten (in termen van nettowinsten) van het introduceren van het nieuwe product bij elke combinatie van A en S. Op grond hiervan kan onderstaande resultatenmatrix (in mio €) opgesteld worden.

Tabel 1 : Resultatenmatrix Bayesiaanse analyse

Strategieën	succes (S1)	mislukking (S2)
<b>introduceren (A1)</b>	30	-20
<b>niet introduceren (A2)</b>	0	0

Op grond van bovenstaande gegevens kan de marketing manager per strategie het verwachte resultaat berekenen, mits hij in staat is waarschijnlijkheden toe te kennen aan het optreden van de omgevingstoestanden. Stel dat hij op grond van zijn ervaring weet dat er een kans van 70% bestaat dat de introductie van een nieuw product een succes wordt en er een kans van 30% bestaat dat deze een mislukking wordt. Er kan dan gesteld worden dat  $P(S1) = 0,7$  en  $P(S2) = 0,3$ . De verwachte opbrengt wordt dan :

voor A1 :  $(0,7 * 30) + (0,3 * -20) = 15$  (mio €) ;

voor A2 : 0. Daar de verwachte opbrengt van strategie A1 het hoogst is, zal de manager deze strategie kiezen, dus het nieuwe product op de markt brengen.

In de Bayesiaanse analyse wordt veelvuldig met het begrip opportuiniteitsverlies in plaats van het resultaat gehanteerd. Een opportuiniteitsverlies treedt op als achteraf, wanneer de omgevingstoestand met zekerheid is opgetreden, blijkt dat niet de met die omgevingstoestand overeenkomende optimale strategie is ondernomen. Op basis van het Bayesiaanse beslissingscriterium wordt die strategie gekozen die het opportuiniteitsverlies minimaliseert. Uitgaande van ons voorbeeld vinden we in tabel 2 de opportuiniteitsverliezen.

Tabel 2 : Opportuiniteitsverliezen Bayesiaanse analyse

Strategieën	succes (S1)	mislukking (S2)
<b>introduceren (A1)</b>	0	20
<b>niet introduceren (A2)</b>	30	0

Zoals we reeds hebben opgemerkt, gaat de priori-analyse uit van subjectieve kansen. We kunnen ons afvragen hoe het toekennen van de priori-kansen aan het optreden van de omgevingstoestanden tot stand komt. Hiervoor zijn verschillende mogelijkheden :

- ◆ het gebruik van historische gegevens, indien aanwezig (bij de introductie van een nieuw product levert dit doorgaans geen problemen op ; men kan vaak als vervanging gegevens gebruiken van vergelijkbare, in het verleden geïntroduceerde producten ; het is evenwel de vraag of deze gegevens representatief zijn) ;
- ◆ afgaan op de ervaring/intuïtie van een beslisser ;
- ◆ het houden van een groepsdiscussie tussen deskundigen om te komen tot een gezamenlijk standpunt (hier de priori-kansen) ;
- ◆ het samenvoegen van individuele standpunten van deskundigen tot een algemeen standpunt (er vindt dus geen groepsdiscussie plaats).

Tot nu toe zijn er geen additionele gegevens verzameld. De marketing manager wil verder onderzoeken of het verzamelen van additionele gegevens de onzekerheid bij het nemen van de beslissing (A1 of A2) zal verminderen. Hij weegt daarom de waarde van de informatie af tegen de kosten die moeten worden gemaakt om de informatie te verkrijgen. We hebben het verwachte resultaat berekend zonder aanvullende informatie (nl. 15 mio € bij strategie A1). Wanneer we nu het verwachte resultaat berekenen bij perfecte informatie, kunnen we het bedrag berekenen dat we maximaal mogen uitgeven om extra informatie te verwerven. Hoe kan het verwachte resultaat bij perfecte informatie bepaald worden ? Daarbij gaan we uit van de eerder gegeven resultatenmatrix. Gegeven het feit dat de beslisser perfecte informatie kan krijgen, kan hij voor elke omgevingstoestand de optimale strategie bepalen.

Als S1 optreedt, wordt A1 gekozen : resultaat = 30 mio €.

Als S2 optreedt, wordt A2 gekozen.

Gegeven de kansen op het optreden van S1 en S2 - respectievelijk 0,7 en 0,3 - is het verwachte resultaat onder perfecte informatie te bepalen als  $(0,7 * 30) + (0,3 * 0) = 21$  mio €. Het verwachte resultaat zonder aanvullende informatie is 15 mio €. We kunnen dus

maximaal  $21 - 15 = 6$  mio € aan onderzoek uitgeven. Of anders gesteld : op grond van de priori-analyse heeft het geen zin meer dan 6 mio € uit te geven voor perfecte informatie.

Bovendien moeten we realistisch zijn : zelfs nadat additionele informatie verzameld is, blijft er een zekere mate van onzekerheid bestaan omtrent de vraag welke beslissing de meest geschikte is. De vraag stelt zich dan hoe we de verwachte waarde kunnen determineren nadat informatie verzameld werd waarvan we weten dat deze niet volledig perfect is. Dit vereist een posteriori- en preposteriori-analyse.

## 2.1. De probleemstelling

### 2.1.1. Algemene situering

De nauwgezette omschrijving van het probleem vormt een eerste en zeer belangrijke stap in het marktonderzoekproces. Men kan hierin twee fasen onderscheiden, hoewel deze in de praktijk meestal ongemerkt in elkaar overvloeien, nl. situatie-analyse en hypothesen-opbouw.

Bij *situatie-analyse* komt het erop aan het onderwerp nauwkeurig af te bakenen en alles wat niet terzake doet terzijde te schuiven. Daarna gaat men van het onderwerp alle belangwekkende aspecten stuk voor stuk onderzoeken, om zo de exacte informatiebehoefte vast te leggen. Het probleem moet dus allereerst in zijn volledige en juiste context geplaatst te worden.

Voorbeeld :

- situatie : dalend marktaandeel
- diagnose : klanten stappen over naar merk B
- marketing-probleem : hoe de klanten terugwinnen ?
- marktonderzoekprobleem : waarom krijgt merk B de voorkeur ?

De noodzaak van een grondige situatie-analyse is een belangrijke reden voor het bestaan van een informatiesysteem in ondernemingen. Wanneer een onderneming over een goed informatiesysteem beschikt, wordt het meestal snel duidelijk in welke richting zich het probleem situeert. Vandaar ook dat onderzoekers secundaire gegevens raadplegen ("desk research").

Onderzoekers beperken zich evenwel niet tot het beschrijven van de toestand : ze proberen er eveneens een verklaring voor te vinden. Dit houdt in dat hypothesen worden ontwikkeld omtrent het ontstaan van fenomenen (bv. veranderend consumentengedrag) en de mogelijkheden tot beïnvloeding ervan. De *hypothesen-opbouw*, die zowel het onderzoeksprobleem als de informatiebehoefte verder helpt definiëren, zal bepalen hoe het conclusief onderzoek er zal uitzien : bij wie wordt wat onderzocht en op welke wijze.

Er kan niet genoeg de nadruk op worden gelegd dat het nauwgezet omschrijven van de probleemstelling van het onderzoek van het allergrootste belang is voor het eventuele succes van het latere onderzoek. Bovendien is een correct gedefinieerde probleemstelling essentieel voor de briefing, d.w.z. de formulering van de opdracht bij de uitbesteding van het onderzoek aan een extern marktonderzoekbureau.

Zoals reeds eerder vermeld, dienen primaire gegevens verzameld te worden wanneer secundaire gegevens niet voldoen aan de informatiebehoefte. Maar in veel gevallen heeft de

marktonderzoeker nog niet voldoende gegevens voor handen om de harde conclusie te trekken dat een grootschalig onderzoek moet worden opgezet. Het kan dan zinvol zijn om - bij wijze van tussenstap - een *exploratief (verkenkend) onderzoek* uit te voeren, met als doel een voorlopige verklaring voor het probleem te vinden. Het gaat hier om een vorm van kwalitatief marktonderzoek : een kleinschalig onderzoek om, uitgaande van een beperkt aantal gegevens, vast te stellen welke aspecten een rol spelen bij een bepaald probleem. Een andere beweegreden voor exploratief onderzoek bestaat erin om, voorafgaand aan het grootschalig onderzoek, de onderzoeksaanpak die men voor ogen heeft (bv. het meetinstrument), uit te testen. Men spreekt in dit geval van een *proefonderzoek* of "pilot study". De conclusies van een verkennend onderzoek zijn oriënterend : ze geven indicaties die we niet zonder meer kunnen generaliseren. Een verkennend onderzoek kan echter wel *hypothesen-genererend* werken. In het algemeen levert een verkennend onderzoek de marktonderzoeker een bepaald inzicht in de situatie die in een later stadium met meer wetenschappelijke onderzoeksmethoden kan worden uitgewerkt en getoetst.

Aangaande verkennend onderzoek zoals hierboven bedoeld, worden twee typen onderscheiden : autoriteiten-onderzoek en case study. Het doel van *autoriteiten-onderzoek* (ook bekend als deskundigen-onderzoek of expert-interviews) bestaat erin kennis te inventariseren van mensen die iets van het probleem afweten. De meest uitgebreide vorm van autoriteitenonderzoek is het zgn. *delphi-onderzoek*, waarbij deskundigen in een aantal rondes onafhankelijk van elkaar hun mening geven. Door hen na afloop van elke ronde met de visie van anderen te confronteren, probeert men consensus te krijgen. In de praktijk gebruikt men echter veelal het gewone expert-interview. Een tweede techniek op het gebied van het exploratief onderzoek is het analyseren van een aantal cases of situaties die gelijkenis vertonen met het probleem dat aan de orde is. De onderzoeker tracht dan te achterhalen wat er in die gevallen gebeurd is en waarom.

## 2.1.2. De relatie tussen probleemstelling en theorie

De gebruikelijke term 'probleemstelling' wil niets anders zeggen dan een nauwkeurige formulering in de vorm van een uitspraak van de vraag (vragen) waarop het onderzoek een antwoord moet geven. In deze paragraaf gaan we nader in op soorten probleemstellingen en de wijze waarop probleemstellingen via onderzoek een oplossing kunnen krijgen, hetgeen ons brengt bij een algemene onderscheiding van enkele typen van onderzoek.

### 2.1.2.1. Soorten probleemstellingen

In de regel wordt een onderscheid gemaakt tussen drie soorten probleemstellingen, nl. beschrijvingsproblemen, verklaringsproblemen en voorspellingsproblemen. In veel onderzoek gaat het vooral om *beschrijvingsproblemen*. De onderzoeker wil een (gedeelte) van een sociaal systeem of van een sociaal proces beschrijven. *Verklaringsproblemen* zijn altijd "hoe komt het dat ..."-problemen. Als men een verklaringsprobleem wil oplossen, veronderstelt dat steeds dat daarnaast voorafgaand een beschrijvingsprobleem is opgelost : men weet hoe iets in elkaar zit, maar nu doet zich de vraag voor : hoe komt dat ? Dit betekent dat de onderzoeker een verklaring zal formuleren en deze via onderzoek zal toetsen. Bij een *voorspellingsprobleem* tenslotte willen we weten hoe een bepaalde situatie in de toekomst zal zijn.



### 2.1.2.2. Typen onderzoek

Wanneer een probleemstelling niet via primaire of secundaire bronnen, en ook niet via descriptie in eenvoudige, operationele termen kan opgelost worden, is een onderzoeker aangewezen op het verzamelen, verwerken en interpreteren/rapporteren van primaire onderzoeksgegevens. De onderzoeker wil naar een hoger abstractieniveau toe : we komen hierbij toe aan verschillende vormen van theorie-georiënteerd onderzoek. Naargelang van de relatie tussen probleemstelling en theorie kan een onderscheid gemaakt worden tussen een drietal typen van onderzoek : *beschrijvend*, *explorerend* en *toetsend* onderzoek. De keuze van het type onderzoek is belangrijk, omdat deze keuze directe consequenties heeft voor de wijze waarop de onderzoeker naar een antwoord zal zoeken voor de probleemstelling van het onderzoek. Voordat we hier nader op ingaan, moeten we proberen een antwoord te geven op de vraag, wat een theorie eigenlijk is.

#### 2.1.2.2.1. Theorieën

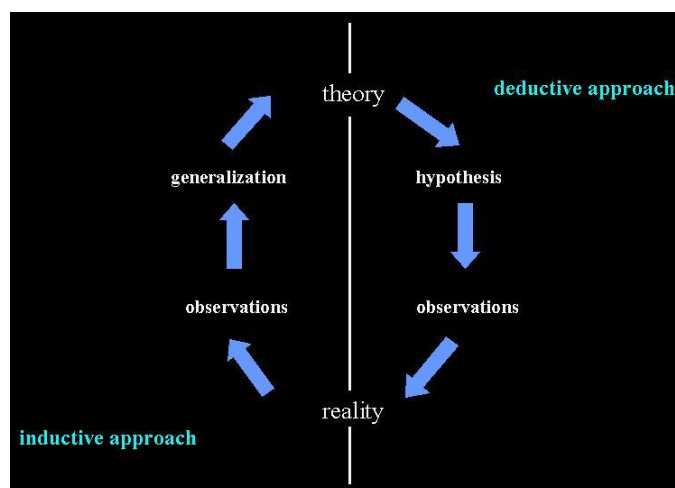
Een theorie is opgebouwd uit een min of meer samenhangend stel uitspraken over een bepaald gedeelte van de werkelijkheid. Eén zo'n uitspraak noemen we een hypothese of propositie. De meeste hypothesen hebben betrekking op relaties tussen kenmerken van (onderzoeks)eenheden. In een theorie hebben we te maken met kenmerken (variabelen) en relaties tussen kenmerken. Kenmerken kunnen op een continuüm van meer abstract naar meer concreet geplaatst worden. Kenmerken aan de abstracte kant van het continuüm worden soms "constructs", *concepten of theoretische begrippen* genoemd. Kenmerken aan de concrete zijde, en met name de meest concrete, direct in de werkelijkheid waarneembare, worden gewoonlijk *operationele kenmerken* genoemd. Een abstract kenmerk is een min of meer algemeen kenmerk waarin tal van concrete verschijnselen zijn onder te brengen. Voorbeelden hiervan zijn o.m. anomie, desintegratie, leiderschap, prestige, enz. Om in onderzoek te werken met dergelijke abstracte kenmerken, dient men in een stadium voorafgaand aan de analysefase indices samen te stellen : uit afzonderlijke vragen naar concreet waarneembare situaties wordt een index voor een algemeen kenmerk samengesteld. Het samenstellen van indicatoren wordt ook *instrumenteel-nomologisch onderzoek* genoemd. Een concreet of operationeel kenmerk geeft precies aan op welke wijze een verschijnsel zich in de werkelijkheid voordoet : het zijn concrete kenmerken, die onmiddellijk waargenomen of direct gemeten kunnen worden.

Tot nu toe spraken we over abstracte en concrete kenmerken. We zagen dat een theorie te maken heeft met kenmerken en met relaties tussen kenmerken. Een relatie geeft aan dat er een verband is tussen het voorkomen van twee of meer kenmerken (variabelen) in de empirische werkelijkheid. In de wetenschappen die zich bezighouden met menselijke houdingen en gedragingen zijn verbanden tussen variabelen in een overgrote meerderheid van de gevallen niet (exact) lineair : als we de waarden van variabelen tegen elkaar afzetten in een assenstelsel en we verbinden de overeenkomstige waarden dan wordt meestal geen rechte lijn gevonden. Wanneer er helemaal geen afwijkingen van de lijn zijn, spreken we van deterministische verbanden. In de gedragswetenschappen hebben we echter te maken met *probabilistische verbanden*. Vandaar ook dat het begrip 'probabiliteit' (kans) centraal staat.

Het doel van een theorie is drievoudig : *beschrijven*, *verklaren* en *voorspellen*. Met behulp van beschrijving en verklaring van een verschijnsel kan men andere, soortgelijke verschijnselen voorspellen. In de beschrijving worden bepaalde kenmerken en verschijnselen uitgeselecteerd. Men moet daartoe die verschijnselen een naam geven, definiëren. Hier zit tevens als een stuk verklaring in : *het onderbrengen van concrete waargenomen zaken in algemene begrippen*. Vervolgens worden die verschijnselen en kenmerken door middel van hypothesen met elkaar in verband gebracht. De onderzoeker formuleert de hypothesen met

een zeker inzicht, dat gebaseerd is op soortgelijke gevonden verbanden. Ook hier is de wisselwerking abstract-concreet duidelijk. *Het onderbrengen (subsumeren) van waargenomen kenmerken en verbanden in meer algemene, abstracte kenmerken en verbanden noemen we verklaren.* M.a.w. verklaren is het onderbrengen van de concrete situatie waarover de probleemstelling handelt, in een geheel van populaties en concepten waarop de theorie betrekking heeft.

Hoe komt nu een theorie tot stand? Aan de oorsprong van elke theorie ligt de waarneming van concrete verschijnselen. De verwondering over deze verschijnselen leidt tot interpretaties, verklaringen die een gemeenschappelijke noemer opleveren waaronder deze waargenomen verschijnselen zijn onder te brengen. Op grond van waarneming van concrete verschijnselen, bestaande brokjes theorie, onderzoeksresultaten van derden en niet in het minst eigen inzicht (intuïtie) van de onderzoeker wordt een aantal mogelijke theorieën bedacht. Dit denkproces noemen we *inductie*. De rudimentaire theorieën (veelal vage veronderstellingen en ideeën) leiden meestal tot een eerste onderzoek waarbij op uitgebreide schaal empirisch materiaal wordt verzameld. Uit een dergelijk *explorerend onderzoek* blijkt meestal welke veronderstellingen opgaan en welke niet. Hiermee wordt de inductieve fase afgesloten. Het resultaat is een theorie, waarvan het geldigheidskarakter evenwel nog niet vaststaat. Het is niet meer dan een vrij samenhangend geheel van veronderstellingen, waarvan niet aangetoond is, dat het op méér situaties, variabelen en eenheden betrekking heeft dan die welke in feite direct of indirect zijn waargenomen. We weten (nog) niet of we aan de theorie iets hebben, d.w.z. of de theorie kan dienen om nieuwe, niet-waargenomen verschijnselen te voorspellen. Het abstraherende, overkoepelende karakter moet nog worden aangemaakt. Het ligt voor de hand de theorie te proberen op andere situaties en/of op andere tijdstippen. Ook zal geprobeerd worden om uit de abstracte hypothesen *operationele hypothesen* af te leiden. Het afleiden van concrete hypothesen, of voorspellingen, uit de meer abstracte gedeelten van de theorie noemen we *deductie*. Het confronteren van deze proposities met de empirische werkelijkheid noemen we de *toetsing van de theorie*. Deze gehele wetenschappelijke procedure, nl. waarneming, inductie, formulering hypothesen, deductie en toetsing wordt aangeduid met de term *hypothetisch-deductief onderzoek*.



Figuur 2 : De empirische cyclus

Aan welke eisen moeten theorieën voldoen? In de eerste plaats moet een theorie *intern consistent* zijn. Dit impliceert dat geen onderdelen met elkaar in strijd mogen zijn, dat de afleidingen logisch zijn en dat de schakels in een keten zoveel mogelijk geëxpliciteerd worden. In de tweede plaats moet een theorie empirisch toetsbaar zijn, d.w.z. er moet minstens één operationele voorspelling in zijn opgenomen. De toetsbaarheidseis betekent

dat althans op enkele punten de mogelijkheid moet bestaan om de theorie te verwerpen, te *falsifiëren*. In de wetenschap stuurt men merkwaardig genoeg aan op falsifiëring, niet op verifiëring. Falsifiëring van een hypothese levert ons méér informatie, nl. méér informatie dan verifiëring. De verklaring hiervoor is de volgende. Indien een voorspelling logisch is afgeleid uit een hypothese, die weer deel uitmaakt van een theorie, en die voorspelling wordt gefalsificeerd, dan is die hypothese en die theorie in ieder geval niet geldig. Wanneer bv. een hypothese luidt dat in kleine werkgroepen problemen sneller een oplossing krijgen dan in grote werkgroepen, en we leiden hieruit de voorspelling af dat problemen sneller opgelost worden in teams van gemiddeld 5 werknemers dan in groepen van 10 werknemers, en uit een onderzoek blijkt het omgekeerde, dan is niet alleen de voorspelling maar ook de meer abstracte hypothese ongeldig. Wanneer de voorspelling echter geverifieerd wordt, dan weten we bv. dat de voorspelling geldig is, maar het is daarmee nog niet zeker of de abstracte hypothese en dus ook de theorie geldig is. In het zojuist gegeven voorbeeld kan de voorspelling geldig blijken, maar misschien alleen als het gaat om het oplossen van heel specifieke groepstaken. De theorie moet dan uitgebreid worden en misschien zelfs geheel anders gestructureerd worden. We hebben derhalve meer aan falsifiëring dan aan verifiëring. Voorts wordt aan een theorie de eis gesteld dat precies aangegeven moet worden op welk gebied, zowel wat variabelen als populatie betreft, de theorie betrekking heeft. Hoe groter het gebied, het domein van een theorie is, des te meer verschijnselen door die theorie verklaard worden en des te sterker die theorie geprefereerd wordt boven andere theorieën. Deze *reikwijdte van de theorie* is een belangrijk criterium voor de acceptatie ervan. Een vierde eis is dat een theorie *eenvoudig* moet zijn ("principle of parsimoniousness"). Voor elk onderzoek moet van tevoren duidelijk zijn bij welke uitslag er sprake is van geldigheid en ongeldigheid van de hypothesen en de theorie. Bovendien moeten de voorspellingen (operationele hypothesen) nauw omschreven zijn, om het andere onderzoekers mogelijk te maken de resultaten in een volgend onderzoek te falsifiëren (een theorie kan nooit 'bewezen' worden : daartoe zou men alle mogelijke voorspellingen bij alle mogelijke populaties en subpopulaties moeten toetsen ; aangezien dit zowel praktisch als principieel onhaalbaar is, stelt men zich tevreden met de toetsing van zoveel mogelijk voorspellingen ; hoe meer voorspellingen niet gefalsificeerd worden, hoe meer geconfirmeerd de theorie is, d.w.z. hoe sterker de theorie komt te staan).

Ten slotte de volgende opmerking. Veel onderzoekingen hebben betrekking op het verband tussen aan aantal onafhankelijke variabelen (oorzaken) en een afhankelijke variabele (gevolg). In dit verband wordt vaak in de rapportage van een dergelijk onderzoek de uitdrukking gebruikt : "met de onafhankelijke variabelen werd x % van de variantie in de afhankelijke variabele verklaard". Dit wil niets anders zeggen, dan dat de gebruikte theorie(ën) waaruit de onafhankelijke variabelen kwamen slechts voor een deel verklaren hoe het komt dat de afhankelijke variabele hoog of laag is. Het begrip 'variantie' dat een essentiële rol speelt in de analyse, vergt nog wat toelichting. Variantie heeft rechtstreeks te maken met het variabel zijn van een kenmerk. Hoe meer een kenmerk over de onderzoekseenheden varieert, hoe groter de variantie van dit kenmerk. Als 75% van de bevolking een krant koopt, en 25% niet, dan is de variantie kleiner dan wanneer er een 50-50 verdeling is. Hoe kleiner de variantie van een kenmerk, hoe minder er te verklaren valt.

#### 2.1.2.2.2. Vormen van theorie-georiënteerd onderzoek

##### 2.1.2.2.2.1. Onderzoek waarbij het accent ligt op descriptie

Dit type onderzoek doet zich voor, wanneer men zich in de probleemstelling tot doel stelt om een concreet maatschappelijk verschijnsel beter te begrijpen door het object van studie te beschrijven in termen die ontleend zijn aan de sociologische theorie. Het doel hiervan is om allerlei concrete waargenomen kenmerken onder te brengen in meer algemene,

overkoepelende trekken, die het object gemeen heeft met andere, reeds onderzochte maatschappelijke verschijnselen. Door deze werkwijze kunnen we de waargenomen werkelijkheid beter begrijpen en verklaren. In tegenstelling tot het explorerend onderzoek en het toetsend onderzoek blijft in het descriptief onderzoek het hoofddoel *de beschrijving van een concreet maatschappelijk verschijnsel*. Een descriptief onderzoek is immers niet gericht op het toetsen van werkhypothesen (toetsend onderzoek), noch op het ontwikkelen van een stuk theorie (explorerend onderzoek).

Descriptief onderzoek kan de vorm aannemen van het toepassen van één bepaalde visie op een concreet veld van waarneming. Vaak echter wordt ook gewerkt met niet één visie of afgerond begrippenapparaat, maar met allerlei 'losse' abstracte inzichten. Het meest kenmerkende van het descriptief onderzoek (naast de gerichtheid op een concreet maatschappelijk verschijnsel) is wellicht de geringe methodologische bagage : we gaan af op ons zgn. "Fingerspitzengefühl". Het spreekt vanzelf dat de interpretatie en verklaring die we geven meestal een voorlopig, en soms vrij willekeurig karakter hebben.

#### 2.1.2.2.2.2. Onderzoek waarbij het accent ligt op exploratie

In explorerend onderzoek is het doel *het ontwikkelen van een algemene theorie over een bepaald verschijnsel*. Uiteraard onderzoeken we het object bij bepaalde waarnemingseenheden, bij bepaalde groepen, maar het doel is het formuleren van een aantal samenhangende hypothesen van een deels abstract karakter. Het vormt een typisch onderdeel van het inductieproces om tot een verklarende en voorspellende theorie te komen.

Een *exploratief (explorerend) onderzoek* is nodig wanneer we nog niet beschikken over een uitgewerkte theorie over een verschijnsel. Explorerend onderzoek heeft vaak de functie van vooronderzoek ten behoeve van latere toetsende studies. Het dient daarbij twee doeleinden : ontdekken van verbanden en daardoor op het spoor komen van een samenhangend stel hypothesen, en het ontwerpen (nog niet het toetsen) van meetinstrumenten. In deze laatste functie wordt het ook als vooronderzoek voor beschrijvende studies toegepast. Explorerend onderzoek heeft dus naast een *hypothesen-formulerende* functie ook een *operationalisatie-onderzoekende* functie.

#### 2.1.2.2.2.3. Onderzoek waarbij het accent ligt op toetsing

Dit is de aangewezen vorm van onderzoek, indien de probleemstelling betrekking heeft op een bepaald sociaal verschijnsel, waarbij de onderzoeker van tevoren beschikking heeft over een theorie of meerdere alternatieve theorieën die met meer of minder grote zekerheid als mogelijke verklaring van het resultaat kan (kunnen) gelden. Voorafgaandelijk dienen dan ook beschrijvingsproblemen opgelost te zijn.

Bij het verzamelde empirisch materiaal wordt getoetst of de voorspellingen uitkomen (verificatie) of niet uitkomen (falsificatie), dan wel dat er geen uitspraak kan gedaan worden. Het toetsend onderzoek kan men ook *explanatie- of verklarend* onderzoek noemen, alhoewel dit minder juiste termen zijn aangezien een beschrijvend onderzoek ook meestal verklaart. In tegenstelling tot explorerend onderzoek ligt bij toetsend onderzoek de nadruk op deductie i.p.v. inductie.

Het onderscheid tussen de hier vermelde typen onderzoek is van relatieve betekenis : de drie behandelde typen dienen niet te exclusief gezien te worden. In menig onderzoek komen mengvormen voor. Het belangrijkste criterium is de rol die hypothesen spelen. In een descriptief onderzoek wordt eigenlijk niet met hypothesen gewerkt. In een toetsend

onderzoek worden de hypothesen geheel vooraf gesteld, terwijl explorerend onderzoek de hypothesen juist aan het einde, na de analyse, laat komen.

## 2.2. Het bepalen van de onderzoeksopzet

Het bepalen van de onderzoeksopzet omvat het plannen van de manier waarop de gegevens worden verzameld en verwerkt. Dit vraagt om drie beslissingen :

- ◆ het vaststellen van de steekproef ;
- ◆ het kiezen van de methode van gegevensverzameling ;
- ◆ het bepalen van de analysemethode.

We hebben gezien dat marktonderzoek, door het aandragen van de juiste informatie, organisaties in staat stelt betere beslissingen te nemen. Informatie uit de markt is met name essentieel om het marketing concept toe te passen. Een eenmalig marktonderzoek is voor talloze bedrijven echter niet voldoende om een afnemers- (c.q. gebruikers-) gerichte werkwijze in stand te houden. Zij volstaan niet met het op incidentele basis verzamelen van gegevens maar zijn meer geïnteresseerd in continu-onderzoek, dat zich dus over langere termijn uitstrekt. Voordat we verschillende samenwerkingsvormen (met marktonderzoekbureaus dan wel met andere opdrachtgevers) onder de loep nemen, verdient het onderscheid tussen ad hoc-onderzoek en continu-onderzoek enige toelichting.

De meeste enquêtes dragen een ad hoc-karakter. Ad hoc-onderzoek omvat het incidenteel verzamelen van gegevens bij een speciaal getrokken steekproef om een eenmalig probleem - voor een opdrachtgever - op te lossen. Zo zal een bedrijf, voordat het een nieuw product introduceert, in de regel eerst de markt waarop het zich richt willen doorlichten om de juiste marketing-mix te kunnen samenstellen. Onder continu-onderzoek verstaan we een meer permanente vorm van informatieverzameling, waarbij regelmatig gegevens worden geregistreerd die van belang zijn voor de commerciële besluitvorming. Zowel het *omnibus-onderzoek* als het *panel-onderzoek* is een bekende vorm van continu-onderzoek. De essentie van *continu -of longitudinaal onderzoek* is dat men op verschillende tijdstippen gegevens verzamelt op basis van dezelfde vragenlijst. Wint men de nodige informatie telkens bij dezelfde personen in, dan spreken we van panel-onderzoek. Een panel is dus een vaste groep respondenten die op regelmatige tijdstippen gegevens verstrekken. De term 'replicatie-onderzoek' reserveren we voor gevallen waarin steeds verschillende respondenten (op basis van dezelfde vragenlijst) gegevens verstrekken. Zowel bij panel- als bij replicatie-onderzoek is samenwerking niet ongebruikelijk. Met marktonderzoek is immers veel geld gemoeid. Om de kosten te drukken, worden de nodige gegevens soms in samenwerking met andere geïnteresseerde bedrijven verzameld. Diverse commerciële marktonderzoekbureaus bemiddelen daarbij. Zij vergaren regelmatig standaard marketinggegevens die geïnteresseerden in de bedrijfstak kunnen kopen, dikwijls op basis van vooraf gemaakte afspraken. We onderscheiden in principe twee mogelijke samenwerkingsvormen : *multi-sponsored onderzoek* en *multi-client onderzoek*.

Bij multi-sponsored onderzoek zijn, zoals de naam aangeeft, meerdere sponsors of opdrachtgevers betrokken. Hun samengaan blijft evenwel beperkt tot het maken van een deel van de vragenlijst die voor het verzamelen van de gegevens wordt gebruikt. Doorgaans ontvangen de deelnemers - exclusief en tegen betaling - alleen die resultaten die voor hen van belang zijn. Zo beschikken zij, voor een relatief laag bedrag, over gegevens uit een grotere steekproef, inclusief de nodige achtergrondinformatie over de respondenten.

Zijn de resultaten van een gezamenlijk onderzoek - uitgevoerd op initiatief van een onderzoeksbureau of groep van opdrachtgevers - in principe (tegen betaling) voor allen beschikbaar, dan spreken we van een multi-client onderzoek. Uiteraard hebben de gegevens van dergelijk multi-client onderzoek een minder vertrouwelijk karakter.

Gegeven de bovenstaande indeling kunnen we een omnibus-onderzoek het beste zien als een vorm van multi-sponsored marktonderzoek dat regelmatig plaatsvindt. Het initiatief gaat uit van marktonderzoekbureaus die zich daarin hebben gespecialiseerd. Zij spelen in op de gelijksoortige behoeften van veel bedrijven door het aanbieden van standaarddiensten ; de standaardisatie in dit type onderzoek heeft o.m. betrekking op de steekproeftrekking en de verwerkingsmethode. Het voornaamste verschil tussen een omnibus-onderzoek en een panel-onderzoek is dat bij een omnibus met een wisselende steekproef wordt gewerkt terwijl bij een panel met een vaste steekproef van respondenten wordt gewerkt.

Een voordeel van het herhaald ondervragen van dezelfde steekproef is dat men de ontwikkeling van bepaalde verschijnselen (bv. gedragsveranderingen, merkentrouw, e.a.) in de tijd kan volgen. Bovendien is men minder afhankelijk van het geheugen van de respondenten omdat zij hun (koop)gedrag zelf in een zgn. dagboek noteren of omdat de registratie met behulp van elektronische apparatuur plaatsvindt. Een nadeel van panels is het feit dat de respondenten, onder invloed van de voortdurende ondervraging zich op den duur atypisch kunnen gedragen, hun houding aanpassen of er een andere mening op nahouden dan wanneer zij geen paneeldeelneemers zouden zijn. De vrees voor deze conditionering is groter bij attitude-onderzoek dan bij gedragsregistratie. Een andere nadeel is de hoge kostprijs van een panel-onderzoek en het risico op grote uitval van deelnemers. Er zijn verschillende soorten panels. De meest gekende zijn consumenten- of verbruikerspanels (bv. GFK), detaillisten- of handelaarspanels (bv. Nielsen) en media-panels.

### 2.2.1. Het vaststellen van de steekproef

Marktonderzoekresultaten zijn vrijwel altijd gebaseerd op steekproefuitkomsten. Een steekproef definiëren we als een groep elementen die uit de populatie wordt getrokken om inzicht te verkrijgen in de eigenschappen van de hele populatie. *Steekproeftrekking* is het gebruiken van een bepaalde procedure bij de selectie van een deel van de populatie, met het uitdrukkelijke doel beschrijvingen of schattingen van bepaalde eigenschappen en karakteristieken van het geheel te verkrijgen. Om een steekproef te kunnen trekken, die een dwarsdoorsnede van de populatie vormt, moeten we eerst de populatie definiëren. Beschikken we over een administratieve registratie van de populatie, dan duiden we die registratie aan met de term *steekproefkader*. De relatieve grootte van de steekproef, dit is het aantal eenheden dat in de steekproef valt, gedeeld door het aantal in de populatie, wordt de *steekproeffractie* genoemd. Een steekproef kan getrokken worden met of zonder teruglegging.

Met teruglegging wordt indien een bepaald element gekozen is, dit element niet uit het steekproefkader verwijderd, zodat het dezelfde kans heeft om bij de volgende keuze weer gekozen te worden. Bij een steekproeftrekking zonder teruglegging wordt een gekozen element uit het steekproefkader verwijderd. Zonder teruglegging krijgen we een betere afspiegeling van de populatie. Bij een steekproef zonder teruglegging kan bv. één cijfer maar één keer getrokken worden, zodat de kans op een grote afwijking van het populatiegemiddelde kleiner is (alleen als de steekproeffractie klein is, zijn de verschillen tussen steekproeftrekking met en zonder teruglegging klein). In sociaal-wetenschappelijk onderzoek zal men altijd een steekproef zonder teruglegging trekken.

Er wordt vaak gesproken over een representatieve steekproef. Men bedoelt daarmee dat de steekproef een getrouwe weerspiegeling is van de populatie waaruit hij is getrokken. Dit is dan meestal op enkele variabelen gecontroleerd. Dat deze variabelen goed in de steekproef vertegenwoordigd zijn is natuurlijk belangrijk, maar men vergeet vaak dat de combinaties van deze variabelen dat ook moeten zijn wil men van representativiteit kunnen spreken. Van belang is dat we te doen hebben met een *waarschijnlijkheidssteekproef*, een *aselecte steekproef*. Een waarschijnlijkheidssteekproef is een steekproef die getrokken is met behulp van een aselecte procedure, waarbij elke eenheid een van tevoren bekende kans heeft om in de steekproef te komen. Deze kansen behoeven niet gelijk te zijn (cfr. infra). Het gebruiken van het *toevalselement*, het aselecte principe, dient om te voorkomen dat er factoren bij de keuze van de onderzoekseenheden een rol gaan spelen die ons onbekend zijn en die tot een vertekening van de steekproef t.o.v. de veronderstelde populatie zullen leiden. Een aselecte steekproef hoeft niet direct representatief te zijn voor een populatie, hoewel de kans op niet-representativiteit bij steekproeven van voldoende grootte niet groot is.

Een vorm van steekproeftrekking waarbij zeer uitdrukkelijk geen toevalselement gehanteerd wordt, is "purposive sampling". Hierbij wordt een poging gedaan om de steekproef zo samen te stellen dat deze representatief is naar enkele kenmerken van de populatie (zo kan men i.p.v. een representatieve steekproef te trekken van de Vlamingen een gebied of stad uit Vlaanderen uitkiezen waarbij men aanneemt dat dit gebied of deze stad op een aantal punten het meest gelijk is op de Vlaamse bevolking). In het algemeen moet men deze vorm van steekproeftrekking afraden omdat de samenstelling van de steekproef afhankelijk is van de persoonlijke keuze van de onderzoeker, zodat er gemakkelijk een systematische vertekening kan ontstaan.

Gebruiken we een waarschijnlijkheidssteekproef, dan hebben we weliswaar geen absolute garantie dat we een getrouwe weerspiegeling van de populatie verkrijgen, maar we kunnen dan wel de kans berekenen dat de waarde van een statistische grootheid van de steekproef, zoals bv. het gemiddelde, een bepaalde absolute afwijking van de populatiewaarde te zien geeft. Schattingen van populatiewaarden die op basis van een steekproef zijn verkregen, bevatten altijd onnauwkeurigheden die veroorzaakt worden door *steekproeffouten* ("error"). Steekproeffouten zijn het gevolg van de toevallige samenstelling van de steekproef. Aangezien zij door het toeval ontstaan, kan de kans op bepaalde afwijkingen berekend worden met behulp van de kansrekening. Naast steekproeffouten, komen er ook *niet-steekproeffouten* voor. Dit zijn afwijkingen die niet met behulp van de kansrekening te bepalen zijn maar bv. te wijten zijn aan een foutieve redactie van de vragenlijst, foutieve interviewerinstructie, weigeringen, fouten die gemaakt worden bij bewerkingen (coderen, e.d.), enz. Soms hebben deze fouten het karakter van toevallige afwijkingen (bv. sommige codeerfouten) en kunnen dan ook als "error" beschouwd worden. Vaak hebben niet-steekproeffouten een systematische vertekening ("bias") tot gevolg. Enkele voorbeelden :

- \* non-response : afwezigheid, ziekte, onwil van de te enquêteren personen, maar ook vergissingen van onderzoekers waardoor personen niet zijn benaderd ;
- \* waarnemingsfouten : afrondings-, meet- en afleesfouten ;
- \* verwerkingsfouten : fouten in de codering.

Behalve het begrip 'aselect' dat te maken heeft met de methode van steekproeftrekking en het begrip 'representativiteit' dat te maken heeft met het resultaat van een steekproef, dient ook nog een onderscheid gemaakt te worden tussen onafhankelijke en afhankelijke steekproeven. Twee gebeurtenissen zijn statistisch onafhankelijk als het voorkomen van de eerste gebeurtenis geen invloed uitoefent op het voorkomen van een tweede gebeurtenis. Als een steekproef aselect getrokken is, is er sprake van onafhankelijkheid in de steekproef.

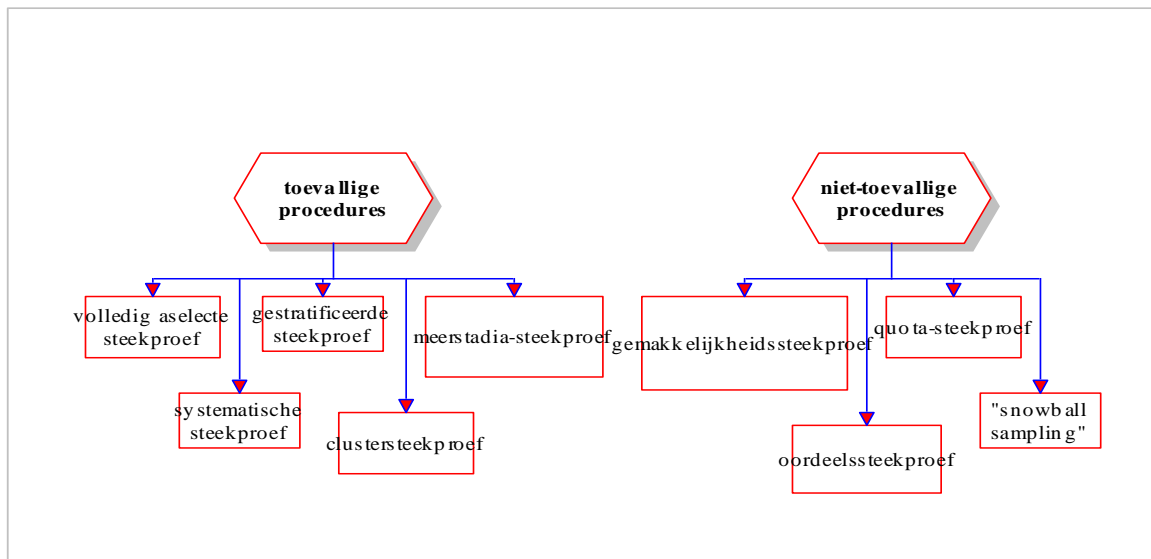
Met de score van respondent X in de steekproef kan geen uitspraak gedaan worden over de score van respondent Y in de steekproef. Dikwijls wordt er niet met één steekproef gewerkt, maar met meerdere steekproeven tegelijk. Dan dient eveneens gekeken te worden naar *onafhankelijkheid tussen steekproeven*. In het geval van onafhankelijkheid kan met de scores van respondenten uit de ene steekproef geen uitspraak gedaan worden over de scores van de respondenten uit de andere steekproef. In sommige studies wordt evenwel besloten tot het trekken van *afhankelijke steekproeven*. Dit gebeurt als men een grote invloed van externe factoren verwacht op de onderzoeksresultaten en men deze probeert uit te schakelen. Dezelfde personen worden zowel voor als na een 'behandeling' ondervraagd of er worden paren gemaakt van personen die met betrekking tot de externe factoren grote overeenstemming vertonen.

Enkele voorbeelden van afhankelijke steekproeven :

- veertig verkoopvertegenwoordigers worden naar een cursus verkooptechniek gestuurd en het management bepaalt de omzet per week van deze managers vóór en na de verkoopcyclus ;
- een geneesmiddelenfabrikant wil een nieuw geneesmiddel op de markt brengen ; alvorens een aanvraag in te dienen voor toelating, worden eerst de effectiviteit en de mogelijke bijwerkingen van het geneesmiddel onderzocht : een groep van proefpersonen krijgt het preparaat in verschillende doseringen toegediend, terwijl een controlegroep een placebo krijgt.

Het valt aan te raden te werken met onafhankelijke steekproeven, tenzij men niet zeker is van de externe factoren.

Het trekken van een steekproef uit de populatie kan op verschillende manieren gebeuren. Er wordt een onderscheid gemaakt tussen *toevallige procedures* en *niet-toevallige procedures*.



Figuur 3 : Methoden van steekproeftrekking



### 2.2.1.1. Toevalssteekproeven

Bij een volledig aselechte steekproef (ook enkelvoudige of gewone toevalssteekproef, "simple random sampling" genoemd) heeft elk element in de populatie precies dezelfde kans om in de steekproef terecht te komen. Het trekken van een zuivere "random sample" kan gebeuren door een willekeurige kolom te nemen uit een tabel met toevalscijfers achterin een statistiekboek en vervolgens de adressen uit de populatielijst die overeenkomen met de betreffende nummers in de steekproef op te nemen.

Zoals eerder vermeld, wordt een onderscheid gemaakt tussen het trekken *met teruglegging* en het trekken *zonder terugleggen*. Bij trekken zonder terugleggen is het niet mogelijk dat een element meer dan één keer wordt getrokken. Strikt genomen blijft in deze situatie de trekkingskans niet constant.

Evenals bij een volledig aselechte steekproef, wordt er bij het trekken van een *systematische steekproef* vanuit gegaan dat elk populatielid een even grote kans moet hebben om in de steekproef opgenomen te worden. De populatie wordt echter op een bepaalde manier geordend. Indien men bijvoorbeeld een steekproefomvang wenst te hebben die een tiende van de populatie bedraagt, dan neemt men bijvoorbeeld elk tiende element uit de populatie in de steekproef. Het voordeel van een systematische steekproef is dat deze gemakkelijk te trekken is. Een nadeel is echter dat wanneer sommige elementen een grotere kans hebben om in de steekproef opgenomen te worden, bezwaarlijk van een toevalssteekproef kan worden gesproken.

Marktonderzoekers maken in veel gevallen gebruik van *gestratificeerde steekproeven*. Immers, op grond van hun ervaring weten zij welke de relevante criteria zijn op grond van dewelke de populatie kan worden opgedeeld in herkenbare en zinvolle subgroepen. Bij een gestratificeerde steekproeftrekking gaat men er derhalve van uit dat de populatie op te delen is in een aantal subgroepen (strata) die ten opzichte van elkaar verschillen maar waarbij binnen elke afzonderlijke groep geen noemenswaardige verschillen ten opzichte van een bepaald kenmerk bestaan (= interne homogeniteit). Een eerste stap bij het trekken van een gestratificeerde steekproef is het bepalen van de juiste strata. Een populatie kan gestratificeerd worden naar kenmerken zoals geslacht, leeftijd, inkomen, opleidingsniveau, e.a. Er dient op gelet te worden dat een element tot slechts 1 stratum behoort. Het aantal elementen van elk stratum dat in de steekproef wordt opgenomen, hangt af van de relatieve omvang ervan in de populatie (*proportioneel gestratificeerde steekproef*). Het gebeurt soms dat men niet-proportioneel trekt (*disproportioneel gestratificeerde steekproef*). Binnen de homogene groepen van de gestratificeerde steekproef wordt aselekt getrokken of op systematische wijze.

De verdeling van de steekproef over de strata in het geval van een gestratificeerde steekproef, wordt de allocatie genoemd. Indien de strata, die duidelijk van elkaar (moeten) verschillen, intern relatief homogeen zijn, wordt gekozen voor een *evenredige allocatie*. In termen van de beschrijvende statistiek houdt dit in dat er m.b.t. het te bestuderen kenmerk significante verschillen bestaan tussen de strata wat het gemiddelde betreft maar dat de verschillen binnen de strata gering zijn (de standaardafwijkingen lopen niet ver uit elkaar).

Voorbeeld. Een bedrijf in de levensmiddelenbranche heeft de markt gesegmenteerd naar maandelijks inkomen in 3 groepen. De groepen bestaan uit resp. 1,8, 1,5 en 1,6 mio huishoudens. De drie groepen besteden maandelijks gemiddeld 450, 600 en 750 of meer € aan levens- en genotmiddelen. De standaardafwijking bedraagt 150 €. Voor het samenstellen van een gestratificeerde steekproef van 1000 gezinnen worden, bij toepassing van evenredige allocatie, uit de drie segmenten resp. 367, 306 en 327 huishoudens in het

onderzoek betrokken. Het gaat hier m.a.w. om een proportioneel gestratificeerde steekproef.

Indien de strata intern relatief heterogeen zijn (tussen de strata bestaan significante verschillen wat het gemiddelde betreft, maar binnen de strata zijn ook de standaardafwijkingen relatief groot), wordt gekozen voor optimale allocatie. Bij optimale allocatie komen relatief veel elementen uit de strata die groot en/of heterogeen zijn. Ook in het geval de variabele kosten per element grote verschillen oplevert tussen de strata, wordt gekozen voor optimale allocatie. In het geval van optimale allocatie wordt een optimale steekproefverdeling berekend met als regel : wijs de steekproef toe in verhouding van de bijdrage van iedere groep tot de totale standaardafwijking.

Voorbeeld. Een fabrikant wil de verkoop van soep in blik schatten en plant een steekproefonderzoek met 100 eenheden. Het merk wordt verkocht in 100 supermarkten met een gemiddelde dagverkoop van 100 blikken en een standaardafwijking van 50. In de 750 winkelpunten waar het merk verkocht wordt, wordt een gemiddelde dagverkoop van 10 blikken met een standaardafwijking van 2 gerealiseerd. Toepassing van de regel bij optimale allocatie betekent dat de winkelpunten voor 23 % bijdragen tot de totale standaardafwijking (nl.  $(750 * 2) / ((750 * 2) + (100 * 50))$ ). Er worden derhalve 77 supermarkten en 23 winkels in het onderzoek betrokken.

In het geval van een *clustersteekproef* of *trossteekproef* wordt de populatie ingedeeld in groepen die met betrekking tot de kenmerken die in het onderzoek een rol spelen zoveel mogelijk op elkaar gelijken. Elke groep of cluster representeert als het ware een miniversie van de gehele populatie. De groepen elementen worden in hun geheel waargenomen.

Voorbeeld. Een koekjesfabrikant produceert elke dag 1000 dozen met elk 25 pakjes koeken. De kwaliteitscontrole kiest aselekt 50 dozen waarvan alle pakjes worden gecontroleerd.

Tabel 3 : Vergelijking tussen gestratificeerde steekproef en clustersteekproef

criteria	gestratificeerde steekproef	Clustersteekproef
subgroepen	Enkele	Vele
intern	Homogeen	Heterogeen
extern	Heterogeen	Homogeen
steekproef	enkele elementen in iedere subgroep	subgroep in zijn geheel
voordelen	Nauwkeuriger	Goedkoper

Er kan ook gebruik gemaakt worden van een *getrapte steekproef* ("multistage sampling"), die opgebouwd is uit een aantal trappen of niveaus. Nauw verwant met de getrapte steekproef is de tweefasen-steekproef (in de tweede trap neemt men een steekproef uit de eerste steekproef) of steekproef met meerstadia-trekking.

Bij de keuze van een steekproeftype dient rekening gehouden te worden met de vereiste nauwkeurigheid en de kosten van de steekproeftrekking. Een gestratificeerde steekproef levert, bij eenzelfde steekproefomvang, een grotere nauwkeurigheid op dan een enkelvoudig aselekt steekproef. De nauwkeurigheid van een trossteekproef is, bij dezelfde steekproefomvang, beduidend geringer dan de enkelvoudig aselekt steekproef. Op

nauwkeurigheid scoort de getrapte steekproef beter dan de trossteekproef maar minder goed dan bij de enkelvoudig aselechte steekproef.

### 2.2.1.2. Niet-toevalssteekproeven

In het geval van een *gemakkelijkheidssteekproef* ("convenience sampling") wordt de steekproef op basis van het gemak getrokken/gekozen. Voorbeelden van "convenience sampling" zijn de 'man/vrouw op de straat'-interviews. Deze methode van steekproeftrekking is vooral bruikbaar bij exploratief-verkennend onderzoek en bij pretesten van vragenlijsten. Aangezien er geen sprake is van representativiteit dient men voorzichtig te zijn met uitspraken of voorspellingen op grond van de via convenience sampling bekomen onderzoeksbevindingen. Men spreekt van "judgement sampling" (*oordeelssteekproef*) wanneer de steekproefelementen worden gekozen op basis van hun geschiktheid omtrent het te onderzoeken fenomeen. Ook autoriteiten-onderzoek kan tot "judgement-sampling" worden gerekend.

Een *quota-steekproef* wordt gebruikt om een aselechte steekproef te imiteren. Eerst wordt bepaald hoe de steekproef er qua socio-demografische variabelen moet uitzien. Hoewel het hier in strikte zin niet gaat om een aselechte steekproef, is een voordeel van deze trekkingsmethode dat we de grootte van de quota zo kunnen bepalen dat de steekproef representatief is voor de kenmerken waarop wordt gequoteerd.

Bij "snowball sampling" worden de eerste elementen veelal aselekt getrokken. Deze dienen evenwel voornamelijk als informatiebron voor de te selecteren doelgroep. Deze methode wordt toegepast indien de te onderzoeken groep moeilijk kan worden getraceerd.

### 2.2.1.3. De steekproefomvang

Aangezien het in veel gevallen de bedoeling is om aan de hand van steekproefonderzoek uitspraken te doen over de populatie, dient de steekproef een voldoende omvang te hebben om generalisaties toe te laten.

#### 2.2.1.3.1. De steekproefomvang bij de schatting van een gemiddelde

Om de steekproefomvang te berekenen bij de schatting van een gemiddelde wordt gebruik gemaakt van de formule :

$$N \geq (z^2 * \sigma^2) / e^2$$

Voor het berekenen van de steekproefomvang in het geval van een gemiddelde, moeten drie factoren gekend zijn :

- ◆ het gewenste *betrouwbaarheidsniveau*, uitgedrukt onder de vorm van een z-waarde onder de standaardnormale verdeling ; meestal wordt een betrouwbaarheidsniveau van 95 % gebruikt, hetgeen correspondeert met een z-waarde van 1,96 ;
- ◆ de *populatie-standaardafwijking* ( $\sigma$ ) ; meestal is deze onbekend zodat deze geschat wordt op basis van een proefenquête of historische gegevens ; er kan ook uitgegaan worden van de variatiebreedte (VB), die in het geval van een normaal verdeelde variabele gelijk is aan  $6\sigma$  zodat  $\sigma = VB/6$  ;

- ♦ de *schattingsfout* ( $e$  ; "sampling error") ; bij een schatting van een populatiegemiddelde wordt een gemiddelde berekend uit een steekproef (puntschatting) ; de schattingsfout is het verschil tussen de puntschatting en de waarde van een populatieparameter ; door een maximale waarde voor  $e$  op te geven wordt de nauwkeurigheid van de schatting ingesteld ; de breedte van het betrouwbaarheidsinterval is gelijk aan  $2e$  (er wordt immers uitgegaan van een symmetrisch verdeelde kansvariabele (steekproefgemiddelde)).

Voorbeeld. Een telefoonmaatschappij vindt dat de tijd tussen de aanvraag van een telefoonaansluiting en de feitelijke installatie op het telefoonnet veel te lang is. Er wordt gepland een steekproefonderzoek uit te voeren onder nieuwe telefoongebruikers. Uit ervaring weet de telefoonmaatschappij dat elke aanvraag binnen 60 dagen gerealiseerd wordt. Hoe groot moet de steekproef zijn om met 95% betrouwbaarheid en met een fout van 2 dagen de gemiddelde wachttijd tussen aanvraag en aansluiting te schatten ? Bij een betrouwbaarheidsniveau van 95% is de  $z$ -waarde 1,96. De variatiebreedte is 60 dagen zodat de populatie-standaardafwijking op 10 dagen kan geraamd worden. De schattingsfout is 2 dagen (meer of minder) . Toepassing van laatstvermelde formule op deze gegevens wijst uit dat  $N \geq 96$ .

Bij eindige populaties moet aan de berekening van de steekproefomvang de eindigheidsfactor worden toegevoegd :

$$N \geq \frac{\sigma^2}{[(e^2/z^2) + (\sigma^2/n)]}$$

Voorbeeld. Een BMW-dealer wil de gemiddelde leeftijd schatten van de kopers van een model. Gedurende de afgelopen periode hebben 800 personen het model gekocht. Hoe groot moet de steekproef zijn om met een fout van 2 jaar bij 95% betrouwbaarheid de gemiddelde leeftijd van de kopers van het model te schatten ? Het werkelijke gemiddelde mag slechts 2 jaar van de puntschatting afwijken, dus  $e$  is gelijk aan 2. Als we veronderstellen dat mensen tussen 18 en 80 jaar in een BMW rijden, kan  $\sigma$  geschat worden op 10,33. Passen we de formule toe dan moet de steekproefomvang  $N \geq 91$ .

#### 2.2.1.3.2. De steekproefomvang bij een schatting van een fractie

Voor het schatten van een fractie gebruiken we de formule :

$$N \geq \frac{z^2 * p(1-p)}{e^2}$$

Ook hier moeten drie factoren gekend zijn :

- ♦ het *betrouwbaarheidsniveau*, aangeduid als een  $z$ -waarde ;
- ♦ de *populatiefractie* ( $p$ ) die onbekend is maar waarvan we de waarde proberen te schatten aan de hand van historische gegevens of door af te gaan op de ervaring ; bij gebrek aan een indicatie voor de populatiefractie stellen we  $p(1-p)$  gelijk aan 0,25 ;
- ♦ de *schattingsfout* ( $e$ ) is het verschil tussen puntschatting en de waarde van de populatieparameter ;  $e$  is dus gelijk aan het verschil tussen de steekproeffractie en de populatiefractie.

Voorbeeld. Een mediabureau maakt een schatting van de fractie Vlamingen die nooit naar TV4 kijkt. De fout mag niet meer dan 0,06 bedragen (6%) en het betrouwbaarheidsniveau wordt vastgelegd op 90 %. Welke steekproefomvang is hiervoor nodig ? Bij een betrouwbaarheidsniveau van 90% correspondeert een  $z$ -waarde van 1,645. Aangezien we de populatiefractie niet kennen, nemen we aan dat  $p(1-p)$  gelijk is aan 0,25. Toepassing van de formule wijst uit dat  $N \geq 188$ .

Bij eindige populaties moet de eindigheidsfactor in de formule opgenomen worden :

$$N \geq p(1-p) / \{(e^2/z^2) + p(1-p)/n\}$$

Voorbeeld. Een dagblad heeft het afgelopen jaar 3500 opzeggingen van abonnementen ontvangen. Om de fractie opzeggers die het dagblad niet meer wensen te ontvangen te kunnen schatten met een fout van 1% bij een betrouwbaarheidsniveau van 90% is een steekproefomvang nodig van  $N \geq 0,25 / \{(0,012/1,6452) + (0,25/3500)\}$  zodat de steekproef uit minstens 2311 Vlamingen moet bestaan.

In marktonderzoek gaat het er vaak om proporties te schatten : het percentage gebruikers van een bepaald product, het percentage respondenten dat het met iets eens of oneens is, enz. In dit geval wordt de standaardfout ( $s_e$ ) van de schatting berekend uit de formule  $s_e = \sqrt{(p * q)/N}$ . Als uit een onderzoek onder 1000 PC-gebruikers blijkt dat 20% ook een Internetaansluiting heeft, dan berekenen we dat  $s_e = \sqrt{(0.2 * 0.8/1000)}$  of 1,3%. Volgens de theorie van de normale verdeling (zie ook bijlage 2) wil dit zeggen dat als we 20% vinden, het werkelijk aantal Internetaansluitingen onder PC-gebruikers in 68% van de gevallen ligt tussen (20% - 1,3%) en (20% + 1,3%). We hebben 95% kans dat het percentage Internetgebruikers ligt tussen (20% - 2\*1,3%) en (20% + 2\*1,3%). Dat laatste betekent dat we in 95% van de gevallen een percentage Internetaansluitingen zullen vinden tussen 17,4% en 22,6% maar het betekent eveneens dat we in 5% van de gevallen een percentage kunnen vinden dat volledig buiten deze grenzen ligt (en dat we er compleet naast zitten). Er is sprake van twee percentages : een percentage om de betrouwbaarheid van de uitspraak aan te geven en een percentage om het interval aan te duiden. Het is ook mogelijk om de steekproefomvang af te stemmen op een voorafbepaald interval. Toegepast op het voorbeeld zoals hierboven beschreven zouden we bijvoorbeeld kunnen bepalen hoe groot de steekproef moet zijn als we met 95% zekerheid een interval wensen dat niet groter is dan 2% langs elke kant. Een zekerheid van 95% correspondeert met een interval van +/- 1.96 \*  $s_e$  of afgerond  $2s_e$ . Dat betekent dat  $2s_e = 2\%$  of  $s_e = 1\%$ . Als we de formule voor de berekening van de standaardfout invullen, berekenen we de steekproefomvang N door substitutie uit de formule  $s_e = \sqrt{(20 * 80)/N} = 1$  waaruit volgt dat een steekproefomvang van 1600 respondenten vereist is. Wanneer we dus 1600 PC-gebruikers ondervragen en tot een aansluitingspercentage van 20% komen, kunnen we met 95% zekerheid stellen dat het percentage PC-gebruikers met een Internetaansluiting in werkelijkheid ligt tussen 18% en 22% (bron : D. Luck & R. Rubin, Marketing research, Prentice-Hall, 1987) .

Tabel 4 : Schattingsinterval bij 95 % waarschijnlijkheid

N	Kans op voorkomen										
	1%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
50	2.8	6.2	8.5	10.1	11.4	12.3	13.0	13.5	13.9	14.1	14.2
100	2.0	4.4	6.0	7.1	8.0	8.7	9.2	9.5	9.8	9.9	10.0
150	1.6	3.6	4.9	5.9	6.6	7.1	7.5	7.8	8.0	8.1	8.2
200	1.4	3.1	4.3	5.1	5.7	6.1	6.5	6.8	7.0	7.0	7.1
300	1.1	2.5	3.5	4.1	4.6	5.0	5.3	5.5	5.7	5.8	5.8
500	0.89	2.0	2.7	3.2	3.6	3.9	4.1	4.3	4.4	4.5	4.5
1000	0.63	1.4	1.9	2.3	2.6	2.8	2.9	3.1	3.1	3.2	3.2
1500	0.51	1.1	1.6	1.9	2.1	2.3	2.4	2.5	2.5	2.6	2.6
2000	0.44	0.96	1.3	1.6	1.8	1.9	2.0	2.1	2.2	2.2	2.2
4000	0.31	0.69	0.95	1.1	1.3	1.4	1.4	1.5	1.5	1.6	1.6
5000	0.28	0.62	0.85	1.0	1.1	1.2	1.3	1.4	1.4	1.4	1.4
10000	0.20	0.44	0.60	0.71	0.80	0.87	0.92	0.95	0.98	0.99	1.0
25000	0.12	0.27	0.38	0.45	0.50	0.55	0.58	0.60	0.62	0.62	0.64
50000	0.08	0.17	0.24	0.29	0.32	0.35	0.37	0.38	0.39	0.40	0.40

In realiteit zullen marktonderzoekbureaus niet vaak overgaan tot berekeningen zoals hierboven uiteengezet. Om de berekeningen te vereenvoudigen, doen veel onderzoekers een beroep op tabellen die het schattingsinterval bij verschillende waarschijnlijkheden weergeven. Als voorbeeld hiervan geven we in tabel 4 een voorbeeld waarin het schattingsinterval bij een gegeven steekproefomvang en een kans op voorkomen wordt weergegeven. Uit tabel 4 leiden we af dat de standaardfout afhankelijk is van het verwacht aantal 'ja'-antwoorden. Naarmate de kans meer neigt naar een 50/50-verdeling, wordt ook de afwijking groter. Vooraleer dergelijke tabellen kunnen gebruikt worden, dient men m.a.w. een inschatting te maken van het percentage 'ja'-antwoorden. Is dit niet geval, dan gaat men uit van de meest conservatieve benadering van de schattingsfout, dus een 50/50-verdeling. Uit de tabel kan eveneens opgemaakt worden dat de standaardfout niet afhankelijk is van de verhouding steekproef-populatiegrootte. Een steekproef van 500 respondenten in Frankrijk vertoont dezelfde fouten als eenzelfde steekproef in Vlaanderen.

#### 2.2.1.4. Eisen aan steekproefonderzoek

Elke steekproef vertoont twee belangrijke aspecten : de *steekproeffout* ("sampling error") en de *vertekening* ("bias" of "nonsampling error").

Wanneer men slechts een gedeelte van de populatie onderzoekt, maakt men altijd een fout (steekproeffout). De statistiek leert ons dat het resultaat uit een steekproef een vaste kans heeft om in de buurt van het echte resultaat te liggen. Men spreekt van de betrouwbaarheid en nauwkeurigheid van de steekproef. Dit puur mathematisch probleem draait rond de steekproefgrootte. De vertekening of "nonsampling error" is een puur kwalitatief-logisch probleem. Door een verkeerde keuze van de onderzochte respondenten kan men een verkeerd beeld krijgen van de werkelijkheid. Wanneer er een fout gebeurde bij de selectie van de respondenten uit de steekproef, kan het gebeuren dat de elementen uit een gedeelte van het universum een systematisch hogere of lagere kans hebben om in de steekproef terecht te komen (vertekening). Dit is het probleem van de representativiteit van een steekproef.

Aan steekproefonderzoek kunnen derhalve een aantal eisen gesteld worden : representativiteit, betrouwbaarheid, nauwkeurigheid en validiteit.

##### 2.2.1.4.1. Representativiteit

Bij steekproeftrekking dient in de eerste plaats de populatie bepaald te worden ("sampling frame"). Een aselechte steekproeftrekking garandeert evenwel niet dat de uitkomsten van het onderzoek representatief zijn, d.w.z. een getrouwe weergave zijn op kleine schaal van de populatie. Door non-respons kunnen bepaalde groepen over- of ondervertegenwoordigd zijn in de steekproef. Men kan dit achteraf nog ontdekken door de representativiteit van de steekproef na te gaan. Daartoe vergelijkt men de samenstelling van de steekproef met die van het universum (hetgeen evenwel inhoudt dat men de samenstelling van het universum kent). Wanneer men tot de bevinding komt dat de steekproef niet helemaal het universum dekt, dient statistisch de conclusie getrokken te worden dat de steekproef geen representatief staal is van het universum. In de praktijk corrigeert men evenwel vaak de steekproef door middel van herweging (representativiteitscorrectie).

#### 2.2.1.4.2. Betrouwbaarheid en nauwkeurigheid

Kenmerkend voor steekproefonderzoek is dat wordt beoogd uitspraken te doen over het universum. Rekening houdend met de steekproeffout ("sampling error") loopt de onderzoeker evenwel het risico dat conclusies naar het universum toe toch (gedeeltelijk) onjuist kunnen zijn. Als uit een onderzoek onder 500 respondenten blijkt dat 40% product A positief beoordeelt, in welke mate kunnen we dan stellen dat 40% van de populatie positief staat tegenover product A? Dit probleem kan zuiver wiskundig worden opgelost. De zekerheid waarmee uitspraken van de steekproef naar de populatie toe kunnen worden gegeneraliseerd, hangt met name af van de steekproefgrootte en van de mate waarin de populatie-eenheden van elkaar verschillen (standaardafwijking).

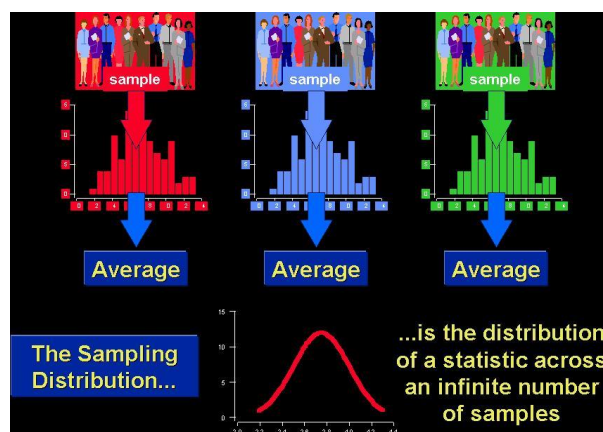
Er doen zich bij steekproefonderzoek in de eerste plaats *toevalsfouten* voor. We kunnen veronderstellen dat er een reeks condities zijn die ertoe leiden dat de scores van de steekproefeenheden in zekere mate vertekend worden. Fluctuaties in het meetresultaat ten gevolge van toevalsfouten (random-fouten) die een geringe, en elkaar gedeeltelijk neutraliserende, invloed uitoefenen, worden toevalsfouten genoemd. In de traditionele methodologische literatuur wordt de mate waarin een meting of meetinstrument vrij is van toevalsfouten de *betrouwbaarheid* van de meting of het meetinstrument genoemd. Als van meting tot meting vrij stabiele resultaten worden bekomen dan is de invloed van toevalsfouten gering. Immers, fluctuaties komen tot uiting wanneer metingen herhaald worden. Het risico op toevalsfouten in een steekproef dat nog aanvaardbaar wordt geacht, is bij afspraak aan te geven en wordt in de meeste gevallen op 5% (of lager) gesteld. Een betrouwbaarheid van 95% wil dus zeggen dat als men het onderzoek 100 keer op dezelfde wijze en hetzelfde moment zou herhalen, de uitkomsten in 95% van de gevallen een juist beeld zouden geven, d.w.z. binnen de aangegeven marges zouden liggen. Overeenkomstig de wet op de grote aantallen worden de betrouwbaarheidsmarges waarbinnen de werkelijkheid kan liggen kleiner naarmate we te maken hebben met een grotere steekproef.

Het doen van uitspraken over de populatie op grond van steekproefuitkomsten kan op twee manieren gebeuren, nl. door een schatting te maken van de populatieparameters aan de hand van steekproefuitkomsten (het *schattingsvraagstuk*, "confidence interval approach") of door *hypothesentoetsing* ("hypothesis-testing approach"). Zolang analyses worden gemaakt op basis van interne gegevens, zijn alle cijfers bekend (er is een continue informatiestroom). Universumparameters kunnen exact berekend worden. De *beschrijvende statistiek* houdt zich bezig met het overzichtelijk maken van dergelijke dataverzamelingen. Zodra echter een analyse moet gemaakt worden van een fenomeen buiten het bedrijf/de organisatie, dient gewoonlijk "ad hoc"-informatie verzameld te worden. In dergelijke gevallen wordt een steekproef getrokken, om aan de hand daarvan de universumparameters af te leiden. Het doen van uitspraken over een populatie op basis van een steekproef wordt *inferentiële statistiek* genoemd.

##### 2.2.1.4.2.1. Het schattingsvraagstuk

Meestal is in een statistisch onderzoek de waarde van een populatieparameter onbekend en moet worden geschat met een steekproefresultaat. Dit staat bekend als 'schatten'. De variabelen die worden gebruikt om een parameter te schatten worden 'schatters' genoemd. De waarde van deze variabele is een 'schatting'. Bekende schatters van het populatiegemiddelde en de populatie-standaardafwijking zijn resp. het steekproefgemiddelde en de standaardafwijking van de steekproef. Bestaat de schatting slechts uit één getal dan wordt gesproken van een puntschatting. Puntschattingen zijn enigszins onbevredigend. We weten bijvoorbeeld dat het steekproefgemiddelde gemiddeld gelijk is aan het populatiegemiddelde, maar toch zal het gemiddelde van een enkele

steekproef zelden exact gelijk zijn aan het populatiegemiddelde. Het steekproefgemiddelde zal iets te hoog of te laag zijn. Dat 'iets te hoog' of 'iets te laag' moet ook in de schatting worden opgenomen. We zullen trachten een gebied van waarden (= interval) vast te stellen, zodanig dat er een bepaalde mate van zekerheid is dat het interval de onbekende waarde van de parameter bevat. We spreken van een intervallschatting of betrouwbaarheidsinterval. Voor een intervallschatting wordt in de statistiek veelal uitgegaan van een betrouwbaarheid van 99%, 95% of 90%. Als men een steekproef trekt, het steekproefgemiddelde berekent en een interval rond het gemiddelde neemt, dan zal dit interval verschillen van een andere steekproef. Immers, geen enkele steekproef is exact gelijk. Stellen we ons voor dat er zeer veel steekproeven worden genomen en zeer veel intervallen worden berekend, dan zal in 95% van de intervallen de onbekende parameter liggen en in 5% van de gevallen niet. We spreken dan van een 95%-betrouwbaarheidsinterval. Een 90%-betrouwbaarheidsinterval zal dus in 90% van alle gevallen de parameter bevatten. De nauwkeurigheid van een steekproefonderzoek geeft de grenzen aan waarbinnen we mogen aannemen dat met een zekere waarschijnlijkheid (= betrouwbaarheid) de werkelijke waarde van de te schatten grootheid ligt. De nauwkeurigheidsmarge is dus een percentage dat weergeeft welke afwijking de onderzoeksuitkomsten kunnen vertonen ten opzichte van de kenmerken van de populatie. Als een steekproefonderzoek bijvoorbeeld aantoont dat 70% van de Vlaamse huishoudens over een platenspeler beschikt, dan duidt een nauwkeurigheidsmarge van 2% aan dat in feite tussen 68% en 72% van de Vlaamse huishoudens over een platenspeler beschikt.



Figuur 4 : De steekproefverdeling

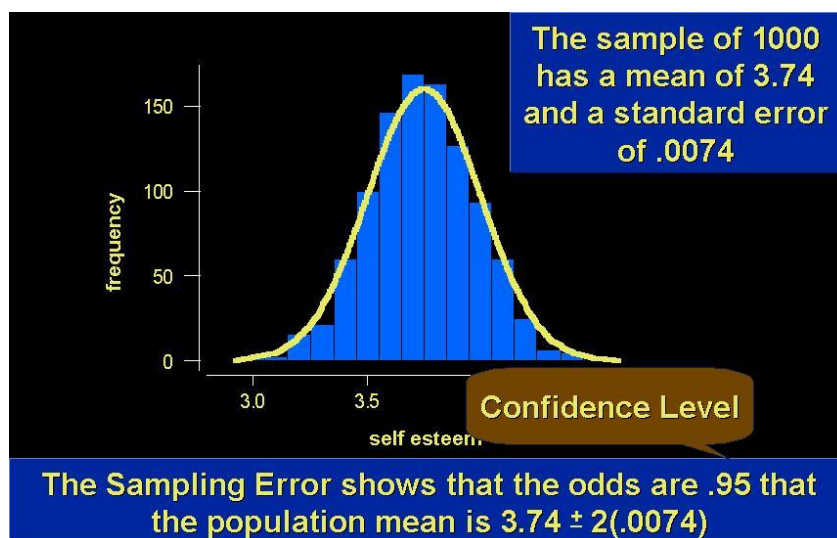
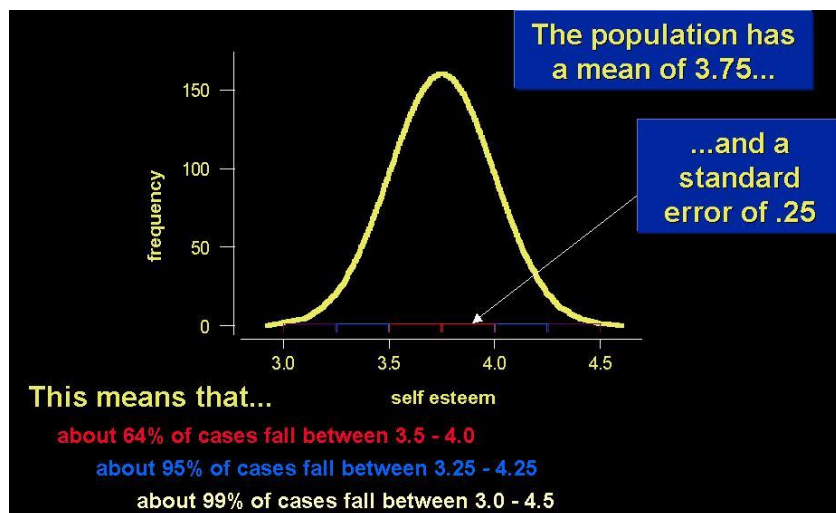
De statistische theorie die ten grondslag ligt aan het bovenstaande, is *de theorie van de normale verdeling*. Wanneer we een eigenschap van een populatie meten, zien we dat niet iedereen die eigenschap (bijvoorbeeld intelligentie) in dezelfde mate bezit. Als de populatie groot genoeg is, vertoont de frequentieverdeling meestal de beroemde klokvorm of *Gauss-curve*.

De breedte van het interval, d.w.z. de nauwkeurigheid van de schatting, wordt beïnvloed door drie factoren :

- ♦ *betrouwbaarheid* : met een andere betrouwbaarheid verandert de waarde van  $z$  of  $t$  in de formule ; een hogere betrouwbaarheid (99% in plaats van 95%) betekent een minder nauwkeurige schatting (breder interval) ;



- ♦ *standaardafwijking* : een grotere standaardafwijking betekent een minder nauwkeurige schatting (grotere standaardfout) ;
- ♦ *steekproefomvang* : een grotere steekproef betekent een nauwkeuriger schatting (kleinere standaardfout).



Figuur 5 : Intervalschatting

Het betrouwbaarheidsinterval van het gemiddelde op basis van een steekproefgemiddelde ( $\chi$ ) berekenen we uit de formule :

$$\chi \pm z^* s_e$$

Voorbeeld. Uit een onderzoek onder 100 personen blijkt dat het gemiddelde tweemaandelijks melkverbruik 16 liter per persoon bedraagt, met een standaardafwijking van 4 liter. De standaardfout van het gemiddelde ( $s_e = s_a/\sqrt{N}$ ) is dus  $4/\sqrt{100} = 0,4$  liter.

Op basis van deze steekproef kunnen we met 95% zekerheid besluiten dat het gemiddeld tweemaandelijks melkverbruik ligt tussen  $16 \pm (2 \cdot 0,4 \text{ l})$  of tussen 15,2 en 16,8 liter (er wordt gebruik gemaakt van de t-waarde, die in dit geval de waarde 2 aanneemt).

Tot nu toe zijn we ervan uitgegaan dat de variabele(n) in de steekproef continu is (zijn). Het is ook mogelijk dat de variabele waarin we geïnteresseerd zijn, slechts twee waarden kan aannemen (*binomiale variabele*). Een voorbeeld van een dergelijke variabele is het percentage huishoudens in Vlaanderen dat een filmcamera bezit. De mogelijke waarden die deze variabele kan aannemen, zijn p (het percentage huishoudens dat een filmcamera bezit) en  $q = 1 - p$  (het percentage huishoudens dat niet over een filmcamera beschikt). Als bijvoorbeeld in een steekproef blijkt dat 20% van de huishoudens in Vlaanderen een filmcamera bezit, kunnen we - zoals bij een continue variabele - een schatting maken van het percentage in de populatie. We kunnen zeggen dat het schattingsinterval van het populatiepercentage gelijk is aan :

$$p \pm z \cdot \sqrt{(p \cdot q) / N}$$

Voorbeeld. In een fietsbandenfabriek voldoet in een steekproef van 250, 20% van de fietsbanden niet aan de kwaliteitsnormen. Het 95%-betrouwbaarheidsinterval voor het onbekende percentage van alle fietsbanden dat niet aan de kwaliteitsnormen beantwoordt, ligt op basis hiervan tussen 15 en 25%.

De grootte van een steekproef kan worden afgeleid uit de nauwkeurigheid en de betrouwbaarheid die we eisen van de schatting van populatieparameters op basis van steekproefparameters. Stel dat we het gemiddelde gewicht van studenten op 3 kg nauwkeurig willen weten en we hierover een uitspraak willen hebben die een betrouwbaarheid heeft van 99,7%. De grenzen van het betrouwbaarheidsinterval (bij 99,7% betrouwbaarheid) zijn  $\pm 3s$ . We moeten evenwel werken met de grenzen  $\pm 3s/\sqrt{N}$ , waarbij N het steekproefgemiddelde is, de ongekende waarde. De nauwkeurigheidseis impliceert dat  $3s/\sqrt{N}=3$ . We zouden N kunnen berekenen als we de standaardafwijking van de populatie zouden kennen, maar deze waarde kennen we niet. Om die reden wordt voor de eenvoud uitgegaan van een veronderstelde standaardafwijking, bv. 10. De steekproef moet dus uit 100 studenten bestaan. Veronderstel dat we voor het voorbeeld van het bezit van een filmcamera een nauwkeurigheidseis van 5% aanhouden bij een betrouwbaarheid van 99,7%. De steekproefgrootte berekenen we in dit geval uit de formule  $0,05 = 3 \cdot \sqrt{(0,2 \cdot 0,8) / N}$ , waaruit afgeleid wordt dat  $N = 576$ .

#### 2.2.1.4.2.2. Hypothesentoetsing

Een bepaalde mening over de waarde van een populatieparameter staat weergegeven in een statistische hypothese. De onderzoeker beschikt slechts over steekproefgegevens en probeert de kans op een foute beslissing die daar een gevolg van kan zijn te minimaliseren. De vraag is nu hoe kan worden vastgesteld bij welke steekproefresultaat de hypothese bevestigd wordt en bij welk steekproefresultaat de hypothese moet worden verworpen. Dit staat bekend als het toetsingsprobleem.

Bij het formuleren van onderzoekshypothesen vertaalt de onderzoeker het probleem zoals dit in de eerste fase van het onderzoeksproces werd gedefinieerd, in een aantal specifieke onderzoeksvragen (formeel uitgedrukt als stellingen), die dan kunnen worden getoetst (d.w.z. aanvaardt of verworpen).

Wanneer men bij kruistabellen op het eerste gezicht verschillen tussen kolommen vaststelt, zal men ook eerst moeten nagaan, of deze door het toeval van de steekproeftrekking zijn ontstaan, dan wel op een werkelijk verschil wijzen. Dit betekent dat men een significantietoets uitvoert.

De hypothese die we willen toetsen noemen we de nulhypothese ( $H_0$ : de ontkenning van de onderzoekshypothese). Daarnaast formuleren we een alternatieve hypothese ( $H_1$ ). Er zijn twee soorten foute beslissingen mogelijk. Als we de nulhypothese ten onrechte verwerpen hebben we te maken met een *fout van de eerste soort* ("type I-error"). Indien we de nulhypothese accepteren, terwijl deze hypothese had moeten worden verworpen, maken we een *fout van de tweede soort* ("type II-error"). In essentie wordt bij hypothesentoetsing nagegaan of de waargenomen frequentie uit dezelfde populatie kan komen als de theoretische frequentie. Is dit geval, dan zijn de gevonden verschillen puur het toeval van steekproeftrekking. In het andere geval is er een betekenisvol (significant) verschil. Men hoopt er m.a.w. op de nulhypothese te moeten verwerpen. Indien de nulhypothese moet aanvaard worden, kunnen aan de gevonden verschillen geen conclusies verbonden worden.

Tabel 5 : Type I en type II-fout

	werkelijkheid	
uitspraak	HO juist	HO onjuist
HO is mogelijk juist (geaccepteerd)	goede beslissing	<b>ten onrechte geaccepteerd</b> type II-fout
HO is onjuist (verworpen)	<b>ten onrechte verworpen</b> type I-fout	goede beslissing

Toetsen kunnen geclassificeerd worden naar de eisen die gesteld worden aan het *meetniveau* van de variabelen (parametrische en niet-parametrische toetsen), de *populatie* (welke vorm heeft de verdeling van de populatie?) en de *steekproef* (is de steekproeftrekking aselekt? is er sprake van afhankelijkheid tussen steekproeven?).

Niet-parametrische toetsen (ook verdelingsvrije toetsen genoemd), zoals de chi kwadraat-toets, stellen geen eisen aan de verdeling van de populatie. De toetsen die eisen stellen aan de verdeling van de populatie zijn de parametrische toetsen. Toetsen die minder hoge eisen stellen aan de gegevens leiden in de regel tot vagere en meer algemene conclusies. De toetsen die hoge eisen stellen, hebben het grootste onderscheidingsvermogen. Zo heeft de t-toets voor verschillen tussen gemiddelden het hoogste onderscheidingsvermogen als aan de eisen van de toets voldaan is, d.w.z. normaal verdeelde populatie, minstens interval meetniveau, onafhankelijke aselekte steekproeven.

#### 2.2.1.4.2.2.1. Parametrische toetsen

##### 2.2.1.4.2.2.1.1. Het steekproefgemiddelde toetsen aan een andere waarde

De t-toets voor één steekproef wordt gebruikt om vast te stellen of het steekproefgemiddelde overeenkomt met een andere waarde. Die andere waarde kan een norm zijn of het nationale gemiddelde of de uitkomst van een ander onderzoek. Een voorbeeld. In een staalfabriek waar platen staal worden gesneden, is het noodzakelijk dat de gemiddelde lengte van een plaat niet afwijkt van de gestelde norm. Met de t-toets voor één steekproef kan worden vastgesteld of dit het geval is. Bij deze toets gaat het steeds om het gemiddelde

van een interval- of ratio-variabele waarbij het de veronderstelling is dat de waarnemingen afkomstig zijn uit een aselechte steekproef uit een normaal verdeelde populatie. De nulhypothese luidt dat het steekproefgemiddelde gelijk is aan het andere gemiddelde. De alternatieve hypothese stelt dat beide gemiddelden ongelijk zijn aan elkaar. Als echter het vermoeden bestaat dat het steekproefgemiddelde op een bepaalde manier zal afwijken van het andere gemiddelde, luidt de alternatieve hypothese dat het steekproefgemiddelde groter (of juist kleiner) is dan het andere gemiddelde.

De t-toets voor één steekproef geeft informatie over de mate van representativiteit van de steekproef. Als de steekproef representatief is, moet het gemiddelde ervan gelijk zijn aan het gemiddelde van de populatie. Dit mag evenwel niet worden omgedraaid. Als het steekproefgemiddelde niet afwijkt van het populatiegemiddelde hoeft de steekproef nog niet representatief te zijn. Heel belangrijk is of de verdeling van de steekproef overeenkomt met de verdeling van de populatie. Dit laatste kan gebeuren aan de hand van de chi-kwadraattoets. Veronderstel dat we weten hoeveel tennisschoenen per prijs categorie (goedkoop, midden en duur) in het afgelopen jaar in België zijn verkocht. Als de steekproef t.a.v. het kenmerk 'gekochte tennisschoenen' representatief is, betekent dit dat we in de steekproef dezelfde verdeling over de prijs categorieën moeten vinden als in geheel België. Stel dat we weten dat in het afgelopen jaar uit elke categorie evenveel schoenen zijn verkocht. De nulhypothese luidt dat de waargenomen verdeling overeenkomt met de verwachte verdeling, dus een gelijk aantal waarnemingen in elke prijs categorie. De alternatieve hypothese stelt dat dit niet het geval is en het aantal waarnemingen per prijs categorie verschilt. In de uitvoer van de chi-kwadraattoets staat de berekende chi-kwadraat, het aantal vrijheidsgraden (= aantal klassen minus 1) en de overschrijdingskans (p-waarde). Veronderstel dat we een overschrijdingskans van 0,0136 vinden. Dit betekent dat we bij een waarschijnlijkheidsgrens van 5% de nulhypothese verwerpen. Als niet wordt voldaan aan de veronderstellingen van de t-toets moet worden uitgeweken naar een niet-parametrische toets (cfr. infra).

Veronderstel dat we aan de hand van de t-toets wensen na te gaan of de gemiddelde leeftijd van tennisspelers in een steekproef (33,2 jaar) al of niet significant afwijkt van het landelijk gemiddelde van 32 jaar. De resultaten van de t-toets voor één steekproef middels gebruikmaking van het SPSS-programma kunnen er dan uitzien zoals in tabel 6. Het onderste gedeelte van de uitvoer geeft informatie over de verschillen tussen het steekproefgemiddelde en de testwaarde. Het gemiddelde verschil is 1,2 en daarachter staat het 95%-betrouwbaarheidsinterval van de verschillen. Tenslotte volgen de uitkomsten van de t-toets. Achtereenvolgens de t-waarde, het aantal vrijheidsgraden (= aantal waarnemingen minus 1) en de overschrijdingskans voor een tweezijdige toets. Deze overschrijdingskans (0,231), ook wel p-waarde genoemd, is groter dan de gebruikelijke grenswaarde van 0,05. Dit betekent dat we de nulhypothese niet verwerpen. Er is dus geen reden om aan te nemen dat het steekproefgemiddelde sterk afwijkt van het algemeen gemiddelde.

Tabel 6 : Uitvoer van de t-toets voor één steekproef

Variable	N	mean	sd	se of mean
"leeftijd"	150	33,200	12,226	0,998
mean difference	95% CI lower/upper	t-value	df	2-tail sig
1,20	-0,773 / 3,173	1,20	149	0,231

## 2.2.1.4.2.1.2. Gemiddelden toetsen bij twee groepen

De t-toets bij twee groepen wordt gebruikt om vast te stellen of de gemiddelden van twee groepen aan elkaar gelijk zijn. Het gaat om gemiddelden van een interval- of ratio-variabele waarbij het de veronderstelling is dat de waarnemingen voor beide groepen afkomstig zijn uit aselechte steekproeven uit een normaal verdeelde populatie. De nulhypothese luidt dat de twee gemiddelden gelijk zijn. De alternatieve hypothese kan op twee manieren worden geformuleerd. In het eerste geval veronderstellen we dat de gemiddelden niet gelijk zijn. Soms echter bestaat het vermoeden dat het gemiddelde van de ene groep groter zal zijn dan dat van de andere groep of omgekeerd en luidt de alternatieve hypothese dat het gemiddelde van groep 1 groter (of juist kleiner) is dan dat van groep 2.

Soms zijn de twee groepen aan elkaar gerelateerd. Stel dat we een vragenlijst hebben voorgelegd aan echtparen die tennissen. Dan kunnen we ons afvragen of mannen gemiddeld evenveel uitgeven aan tenniskledij als vrouwen. In dat geval moet de gepaarde t-toets uitgevoerd worden. Dit geldt eveneens in situaties waarin op twee tijdstippen dezelfde persoon is ondervraagd, bijvoorbeeld vóór en na een bepaalde gebeurtenis, en we ons afvragen of de gemiddelden van beide metingen gelijk zijn.

Beschouwen we eerst het geval van twee *onafhankelijke steekproeven*. Nemen we terug het voorbeeld van tennissen waarbij we aan de hand van de t-toets nagaan of de gemiddelde uitgaven aan baanhuur verschillen tussen een groep enkel- en dubbelspelers (resp. groep 1 en groep 2). De uitvoer van de t-toets met het SPSS-programma kan eruit zien zoals in tabel 7. De bovenste helft van de uitvoer bevat een aantal kengetallen voor beide groepen (het aantal waarnemingen, het gemiddelde, de standaardafwijking en de standaardfout van het gemiddelde). Onder dit overzicht staat het verschil in gemiddelde van beide groepen. Samen met de standaardafwijking van de twee gemiddelden en het aantal waarnemingen bepaalt het verschil in gemiddelde de uitkomst van de t-toets. In het onderste deel van de uitvoer staat het resultaat van de t-toets. Er worden twee verschillende t-toetsen uitgevoerd die worden aangeduid met "equal variance" en "unequal variance". De keuze tussen beide methoden wordt bepaald door de aanname die wordt gemaakt over de variantie in beide populaties. Als we aannemen dat deze gelijk zijn, worden de uitkomsten van de "equal variance"-methode gebruikt en anders die van de "unequal variance"-methode.

Tabel 7 : De uitvoer van de t-toets bij twee onafhankelijke steekproeven

variable	N	Mean	sd	se of mean
enkelspel	22	1022,727	776.060	165.456
dubbelspel	48	626,250	463.133	66.847
mean difference				
396.4773				
Levene's test				
F=9,635				
p=0,030				
variance	t-value	2-tail sig	se of diff	95 % CI lower/upper
equal	2.66	0.010	148.851	99.383 / 693.571
unequal	2.22	0.035	178.450	30.853 / 762.102

Behalve het maken van een veronderstelling omtrent de (on)gelijkheid van beide varianties, kan dit ook getoetst worden. Hiervoor wordt de Levene's toets gebruikt. De nulhypothese van de Levene's toets is dat de varianties in beide populaties gelijk zijn. Bij deze toets wordt voor elke waarneming het verschil met het groepsgemiddelde berekend, waarna met deze verschillen een variantie-analyse wordt uitgevoerd. Het resultaat van deze toets staat in het midden van tabel 7. Zowel de F-waarde als de bijbehorende overschrijdingskans worden afgedrukt. Een zeer lage overschrijdingskans, bijvoorbeeld kleiner dan 0,05, leidt tot het verwerpen van de nulhypothese van gelijke varianties in beide populaties. In ons voorbeeld is de overschrijdingskans 0,03. We verwerpen dus de nulhypothese en gebruiken de "unequal variance"-methode (overigens wordt het verschil tussen "equal variance"-methode en "unequal variance"-methode bij grote steekproeven steeds kleiner).

De uitvoer van de "unequal variance"-methode staat in de onderste regel. In de derde kolom staat onder het kopje "2-tail sig" ("two-tailed significance") de overschrijdingskans voor een tweezijdige toets. De keuze tussen eenzijdig en tweezijdig toetsen wordt bepaald door de alternatieve hypothese. De nulhypothese is in beide gevallen dat de steekproefgemiddelden niet verschillen, de alternatieve hypothese luidt bij eenzijdig toetsen dat het gemiddelde in de ene groep hoger is dan in de andere groep. Bij tweezijdig toetsen luidt de alternatieve hypothese dat er verschil bestaat in gemiddelde tussen de beide groepen, ongeacht welke groep een hoger gemiddelde heeft. De door SPSS aangegeven overschrijdingskans is voor tweezijdig toetsen, bij een eenzijdige toets moet dit getal door 2 gedeeld worden.

In ons voorbeeld vergelijken we de uitgaven aan baanhuur tussen mensen die alleen in het enkelspel actief zijn en degenen die alleen in dubbelspel actief zijn. Het lijkt aannemelijk dat mensen die alleen dubbelen gemiddeld minder uitgeven aan baanhuur, omdat de dubbelspelers de baanhuur met vier personen kunnen delen. In dit geval voeren we dus een eenzijdige toets uit en daarom delen we de overschrijdingskans van 0,035 door 2. De overschrijdingskans is dus 0,0175. De overschrijdingskans is de kans dat een verschil als nu is aangetroffen tussen beide groepen, wordt aangetroffen in steekproeven met deze omvang als de gemiddelden in werkelijkheid gelijk zijn (ook wel p-waarde of fout van de eerste soort genoemd). Deze kans verwijst dus naar het ten onrechte verwerpen van de nulhypothese. Het is gebruikelijk de nulhypothese van gelijke groepsgemiddelden te verwerpen als de overschrijdingskans kleiner is dan 0,05. In dit geval is de overschrijdingskans duidelijk lager dan 0,05 en dus kunnen we concluderen dat we in dit geval de nulhypothese van gelijke gemiddelden moeten verwerpen. Omdat bovendien het gemiddelde van de enkelspelers hoger is dan dat van de dubbelspelers accepteren we de alternatieve hypothese : mensen die alleen in het enkelspel actief zijn geven gemiddeld significant meer geld uit aan baanhuur dan mensen die alleen dubbelen.

Evenals de gewone t-toets wordt ook de gepaarde t-toets gebruikt om vast te stellen of de gemiddelden van twee groepen aan elkaar gelijk zijn. Het verschil is dat ditmaal de twee groepen geen onafhankelijke steekproeven vormen. De waarnemingen uit beide groepen zijn aan elkaar gerelateerd. De *gepaarde t-toets* wordt vaak toegepast indien sprake is van voor- en nametingen. Zo kan dezelfde persoon twee keer dezelfde vragen zijn voorgelegd. De gepaarde t-toets wordt dan gebruikt om na te gaan of men (gemiddeld) van mening veranderd is. De gepaarde t-toets lijkt in veel opzichten op de t-toets bij onafhankelijke steekproeven.

Stel dat we de hypothese willen toetsen dat vrouwen (echtgenotes) gemiddeld meer geld besteden aan kenniskledij dan hun echtgenoot. De nulhypothese luidt dat de beide gemiddelden gelijk zijn. De alternatieve hypothese is in dit geval dat het gemiddelde voor echtgenotes hoger is dan dat voor echtgenoten. Dit is een vorm van eenzijdig toetsen (bij

een tweezijdige toets zou de alternatieve hypothesen hebben geluid : het gemiddelde van de echtgenoten en echtgenotes is ongelijk aan elkaar).

In tabel 8 hebben we de mogelijke uitvoer van de gepaarde t-toets opgenomen (eveneens gebruikmakend van het SPSS-programma). De bovenste helft van de uitvoer geeft een overzicht van de verschillen tussen echtgenoten en echtgenotes. In totaal zijn 150 paren ondervraagd en bedraagt de correlatie-coëfficiënt tussen groep 1 (echtgenoten) en groep 2 (echtgenotes) 0,79. Uit de overschrijdingskans (0,000 bij "2-tail sig") blijkt dat de correlatiecoëfficiënt ruimschoots van nul verschilt. De correlatiecoëfficiënt geeft een indruk van de effectiviteit van het paarsgewijs uitvoeren van de t-toets. Als de correlatie tussen beide variabelen laag is, had net zo goed elke waarneming van een man en diens vrouw als een aparte waarneming kunnen worden beschouwd (en hadden we een t-toets voor onafhankelijke steekproeven kunnen uitvoeren). Des te dichter de correlatiecoëfficiënt bij 1 ligt, des te groter is het voordeel om een gepaarde t-toets uit te voeren.

Tabel 8 : De uitvoer van de gepaarde t-toets

variable	number of pairs	corr	2-tail sig	Mean	sd	se of mean
mkleding				322.667	274.692	22.429
vkleding				474.000	297.288	24.273
	150	0.790	0.000			
paired differences						
mean	sd	se of mean	t-value	df	2-tail sig	
-151.333	186,359	15.216	-9.95	149	0.000	

Het onderste deel beschrijft de verschillen. Eerst verschijnen dezelfde drie kengetallen maar nu voor de gepaarde verschillen. Het minteken (-151,333) geeft aan dat de eerste groep (echtgenoten) lager scoort dan de tweede groep (echtgenotes). Een goede aanduiding van de uitkomst van de gepaarde t-toets geeft ook het 95%-betrouwbaarheidsinterval (niet opgenomen in tabel 8 ; ondergrens, resp. bovengrens -181,407 en -121,259). Als het getal nul zich in het interval bevindt, leidt dat tot acceptatie van de nulhypothese dat er geen verschillen bestaan tussen beide groepen. In het laatste deel van de uitvoer staat het resultaat van de gepaarde t-toets en wordt de overschrijdingskans bij tweezijdig toetsen afgedrukt. In ons voorbeeld zouden we de overschrijdingskans door twee moeten delen omdat we eenzijdig toetsen. Gezien het significantieniveau (0,000) is het duidelijk dat we de nulhypothese van gelijke gemiddelden verwerpen. Echtgenotes geven dus significant meer geld uit aan tenniskledij dan hun echtgenoten.

#### 2.2.1.4.2.2.2. Niet-parametrische toetsen

Bij de statistische toetsen die we tot dusver hebben besproken, werd steeds uitgegaan van een aantal vooronderstellingen, waaraan moest worden voldaan om de toets te mogen uitvoeren. De meeste toetsen vereisen dat de variabelen zijn gemeten op interval- of rationiveau en dat de steekproefverdeling normaal is, d.w.z. dat de variabelen afkomstig zijn uit een normaal verdeelde populatie en dat de steekproef voldoende groot is (minstens 30 "cases"). Statistische toetsen die een normale verdeling veronderstellen, worden parametrische toetsen genoemd.

Vaak echter wordt niet aan alle vooronderstellingen voldaan : de data zijn gemeten op nominale of ordinale schaal of we kunnen niet veronderstellen dat de steekproefverdeling normaal is. In dergelijke gevallen kan als alternatief een niet-parametrische of vrije verdelingstoets worden uitgevoerd. Aan niet-parametrische toetsen worden minder hoge eisen gesteld. Het nadeel is echter dat deze toetsen ook minder krachtig zijn, dit betekent dat een onjuiste nulhypothese minder vaak zal verworpen worden.

Behalve de chi-kwadraattoets (cfr. supra), is er de binomiale toets. Deze wordt gebruikt om te onderzoeken of de verzamelde gegevens bevestigen dat er een kans ( $p$ ) is op een bepaalde waarde van een variabele. Bij de toets worden twee groepen onderscheiden waarbij vergeleken wordt of de waargenomen frequenties in beide groepen met de verwachte frequenties overeenkomen bij een binomiale verdeling.

De Kolmogorov-Smirnov toets wordt gebruikt om te bepalen of de waargenomen verdeling overeenkomst met een bepaalde theoretische verdeling. Ordinale schaling van een variabele is voldoende. Veel toetsen, zoals de t-toets, veronderstellen dat de steekproef afkomstig is uit een normaal verdeelde populatie. Als hieraan getwijfeld wordt, kan met behulp van de Kolmogorov-Smirnov toets vastgesteld worden of de gegevens in de steekproef overeenstemmen met de normale verdeling.

De Mann-Whitney toets, ook wel aangeduid als Wilcoxon-toets, wordt gebruikt om na te gaan of de gemiddelden van twee steekproeven van elkaar verschillen en kan dus aangeduid worden als de niet-parametrische tegenhanger van de t-toets. Voor de Mann-Whitney toets is een ordinale schaal voldoende. De mediaantoets wordt gebruikt om vast te stellen of twee (of meer) onafhankelijke steekproeven afkomstig zijn uit dezelfde populatie of populaties met dezelfde mediaan. De mediaantoets veronderstelt dat een variabele minimaal ordinaal geschaald is. De tekentoets wordt gebruikt om te bepalen of twee gerelateerde steekproeven dezelfde mediaan hebben. De variabelen moeten minimaal ordinaal geschaald zijn.

#### 2.2.1.4.2.3. Methoden ter vaststelling van de betrouwbaarheid

De betrouwbaarheid van een meetprocedure kan, zoals in het voorafgaande reeds is omschreven, worden aangeduid als de mate waarin meetuitkomsten ongevoelig zijn voor toevalsfactoren. Centraal daarbij staat *de constantheid of stabiliteit van de meting*. Herhaalde toepassing op identieke elementen moet identieke resultaten opleveren. Deze omschrijving maakt reeds enigermate duidelijk, hoe bij een concrete meetprocedure kan worden vastgesteld, of en in welke mate zij betrouwbare resultaten geeft. Toch kunnen, afhankelijk van de wijze waarop de meetprocedure wordt herhaald, enkele methoden ter bepaling van de betrouwbaarheid worden onderscheiden, nl. test-retest betrouwbaarheid, instrument-betrouwbaarheid en onderzoeker-betrouwbaarheid.

##### 2.2.1.4.2.3.1. Test-retest betrouwbaarheid

De kern van deze methode is de toepassing (door dezelfde onderzoeker) op verschillende tijdstippen van hetzelfde meetinstrument op dezelfde onderzoekselementen. Geeft nu zo'n herhaalde toepassing van het instrument dezelfde resultaten, dan is er sprake van test-retest betrouwbaarheid.

Wil men een verantwoorde conclusie trekken uit de vergelijking van de twee meetuitkomsten, dan moet voldaan zijn aan twee voorwaarden : de twee metingen moeten



onafhankelijk van elkaar zijn, d.w.z. de eerste meting mag geen invloed hebben op het resultaat van de tweede meting en de onderzoekselementen mogen wat de onderzochte eigenschap betreft niet veranderd zijn in het tijdsinterval tussen de twee metingen.

Om aan de eerste voorwaarde te voldoen, zou een zo groot mogelijk tijdsinterval wenselijk zijn, zodat het effect (via bv. herinnering) van de eerste meting wordt geminimaliseerd. De tweede voorwaarde daarentegen vereist veelal juist een kort tijdsinterval tussen de beide metingen. Veel eigenschappen kunnen immers in de loop van de tijd veranderen. Bij een lang interval kan men dan niet meer uitmaken, of er van een echte verandering in de tijd sprake is, of alleen maar van een toevalsfluctuatie van het meetinstrument.

Afhankelijk van de onderzoekssituatie en van de aard van de te onderzoeken variabelen zal men de voor- en nadelen van een lang of kort interval tegen elkaar moeten afwegen. Heeft men te doen met onveranderlijke of relatief onveranderlijke variabelen, dan is een lang tijdsinterval tussen beide metingen te prefereren bij de vaststelling van de betrouwbaarheid van het instrument.

#### 2.2.1.4.2.3.2. Instrument-betrouwbaarheid

Hoewel alle methoden van betrouwbaarheidsbepaling te maken hebben met de betrouwbaarheid van het instrument, is de term instrument-betrouwbaarheid specifiek gereserveerd voor het onderzoek van die toevalsfouten, die terug te voeren zijn op de selectie van de indicatoren en de uitwerking ervan op het meetinstrument. Uitgangspunt bij de gedachtengang is dat er in principe een groot aantal verschijnselen is, die elk op zich empirische informatie kunnen verschaffen over de waarde van een theoretische eigenschap. Het maken van een keuze uit deze totaliteit van potentiële indicatoren en operationaliseren is dan ook onvermijdelijk. De term instrument-betrouwbaarheid verwijst nu naar de gevoeligheid van de meetuitkomst voor de specifieke set van indicatoren, die is geselecteerd. Worden immers de indicatoren onderling vervangbaar gedacht, dan zal er van instrument-betrouwbaarheid sprake zijn als er via verschillende indiceringen van een theoretisch begrip dezelfde meetresultaten worden bereikt.

Het toepassingsbereik van deze methode is voornamelijk dat van de *samengestelde variabelen*, d.w.z. meetprocedures waarbij een aantal indicatoren (items) samen resulteren in de waardebepaling van de variabele.

De methode ter vaststelling van de instrument-betrouwbaarheid kent een tweetal varianten, nl. de *parallel-test* en de "*split half*"-methode. Bij de *parallel-test* worden twee afzonderlijke procedures gebruikt ter meting van de waarde van een variabele. Beide instrumenten, met hun verschillende indiceringen, worden op dezelfde onderzoekselementen toegepast. Een gelijk meetresultaat van beide instrumenten impliceert dan betrouwbaarheid. Bij de "*split half*"-methode is er sprake van één meetinstrument, dat wel is opgebouwd uit een groot aantal indicatoren. Ter bepaling van de instrument-betrouwbaarheid worden de indiceringen verdeeld in twee gelijkwaardige delen, waarna wordt vastgesteld of de twee 'halve' meetinstrumenten dezelfde uitslag geven.

#### 2.2.1.4.2.3.3. Onderzoeker-betrouwbaarheid

De onderzoeker-betrouwbaarheid heeft betrekking op de gevoeligheid van het instrument voor toevalsfluctuaties die het gevolg zijn van factoren die bij de gebruiker van het instrument gelegen zijn. Van een betrouwbaar instrument mag immers verwacht worden,

dat het voor de uitkomst van de meting niets uitmaakt wie de meting verricht. Dit type betrouwbaarheid richt zich dus expliciet op de *intersubjectiviteit*. Betrouwbaarheid in deze zin is dus aanwezig, als het gebruik van hetzelfde instrument door verschillende onderzoekers bij constante onderzoekselementen dezelfde uitslag geeft.

Tenslotte willen we nog wijzen op systematische fouten die kunnen optreden door factoren, gelegen bij de onderzoeker. Het minimaliseren van meetfouten ten gevolge van toevalsfluctuaties mag bijzonder waardevol zijn, het ontslaat de onderzoeker niet van de plicht kritisch na te gaan of bepaalde eigenschappen van de gebruiker van het instrument het meetresultaat misschien systematisch vertekenen.

#### 2.2.1.4.3. Validiteit

##### 2.2.1.4.3.1. Begripsomschrijving

Validiteit heeft betrekking op de vraag of men inderdaad meet wat men wil meten. Voor de validiteit van een steekproefonderzoek is het niet alleen essentieel dat we de steekproef trekken uit de populatie waarnaar we de onderzoeksresultaten willen generaliseren, maar ook dat we - in het geval van een enquête - de vragen zo nauwkeurig mogelijk formuleren.

Terwijl de betrouwbaarheid van een onderzoek verwijst naar de afwezigheid van toevalsfouten (en betrekking heeft op de stabiliteit) heeft de validiteit van een onderzoek betrekking op de mate waarin het onderzoek vrij is van *systematische fouten*. In de enquêtetraject zijn vooral twee bronnen van systematische fouten bekend : nl. de *sociale wenselijkheid* ("social desirability", d.w.z. dat de antwoorden van respondenten medebepaald worden door hetgeen zij vinden dat een enquêteur een goed antwoord vindt) en de *zgn. instem-neiging* of *ja-knikkerij* ("acquiescence", d.w.z. de neiging van respondenten om in te stemmen met vragen ongeacht de inhoud ervan).

De *interne validiteit* heeft betrekking op de vraag of de juiste onderzoeksmethode(n) zijn gebruikt (bv. in een experiment : zijn veranderingen in de afhankelijke variabele een gevolg van wijzigingen in de "treatment"-variabele i.p.v. andere factoren ?). De *externe validiteit* slaat op de relevantie van het onderzoeksproject zelf. Hier is de vraag of de conclusies van een onderzoek met recht kunnen worden toegepast op andere situaties (een experiment kan in termen van interne validiteit volledig in orde zijn zonder dat het externe validiteit bezit).

##### 2.2.1.4.3.2. Checks op validiteit

We kunnen een kenmerk in zijn operationele vorm zien als een *meetinstrument*. Zo kunnen we de eenvoudige vraag naar de leeftijd een meetinstrument noemen. Eigenlijk moeten we echter het begrip 'meetinstrument' ruimer opvatten : het omvat behalve de vraag zelf alle zaken die invloed op de meetresultaten kunnen uitoefenen : de interviewer (je kunt een vraag op verschillende manieren stellen), de omstandigheden, de codeur, de ponser (kunnen fouten maken), e.d.

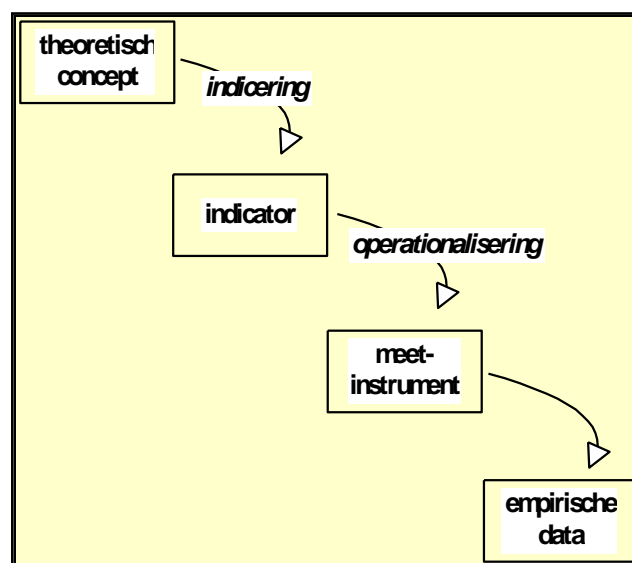
Bijna altijd doet zich in de relatie tussen de theoretisch bedoelde eigenschap en het operationeel kenmerk een vertekening voor : je meet lang niet alles wat onder het eerste valt. Het essentiële probleem is, dat meestal vertekening optreedt : je meet, behalve datgene wat je meten wil, ook nog andere zaken (bv. de fouten van de ponser). We noemen dit een validiteitsprobleem : een meetinstrument is valide als het dat kenmerk en alleen dat

kenmerk meet, wat de onderzoeker wil meten. Schiet het daarin tekort, dan daalt de validiteit, we zeggen ook wel dat er "bias" optreedt. Er is een vaste of *systematische vertekening*.

Ter verduidelijking van de geldigheidsproblematiek is het onderscheid tussen *theoretische concepten* en *empirische variabelen* van belang. Het geldigheidsprincipe houdt in dat de onderzoeksgegevens zodanig moeten zijn dat op legitieme wijze overgegaan kan worden van het niveau van de empirische variabelen naar dat van de theoretische concepten.

Binnen het proces van vertaling van theorie naar empirie wordt het onderscheid geïntroduceerd tussen *indicering* en *operationalisering*. De betekenis van deze termen en hun onderlinge betrokkenheid is weergegeven in figuur 6. Hoewel indicering en instrumentkeuze nauw met elkaar verbonden zijn en in de onderzoekspraktijk eigenlijk niet los van elkaar tot stand komen, kan het maken van een onderscheid tussen beide een beter zicht geven op de overwegingen die bij de beoordeling van de geldigheid van belang zijn. De fouten die bij de indicering en operationalisering kunnen gemaakt worden, hebben gevolgen voor de geldigheid van het meetresultaat, maar de betekenis van een foutieve indicering is een andere dan die van een foutieve operationalisering.

De geldigheid van meetprocedures heeft in eerste instantie betrekking op het *indiceren* en pas in tweede instantie op de *operationalisering*. Om te bereiken dat in de meetprocedure dat wordt gemeten wat men beoogt te meten, is vereist dat bij de indicering verschijnselen worden aangewezen die op adequate wijze het te meten theoretisch begrip dekken. De te kiezen indicatoren moeten dus in hun betekenis overeenstemmen met het theoretisch begrip.



Figuur 6 : Theoretische concepten en empirische variabelen

Niet alleen bij de indicering, ook bij de operationele definiëring kunnen fouten worden gemaakt die consequenties hebben voor de geldigheid van het meetinstrument. Indien voor een op zich geldige indicator meetprocedures worden gebruikt, die wat betreft de stimuloediening en de observatie van de response niet aan bepaalde vereisten voldoen, dan zal het meetresultaat ongeldig zijn. In zijn algemeenheid is dan ook vereist dat de waarnemingsprocedure registreert wat als indicator was aangewezen, en niet meer dan dat. Anders gezegd : er mogen in het instrument geen andere signalen meeklinken en in de

ontvangst van de response mag er geen ruis aanwezig zijn. Een op zich geldige indicator hoeft dus niet tot een goed meetresultaat te leiden omdat er ernstige feilen aan de operationele uitwerking kunnen kleven. Een operationalisering wordt als deugdelijk beschouwd als er geen systematische fouten optreden en als het instrument betrouwbaar is. Voorwaarde voor de geldigheid van een meetprocedure is derhalve dat aan deze beide vereisten is voldaan.

#### 2.2.1.4.3.3. Methoden ter vaststelling van de geldigheid

De geldigheidseis kan in kort schematisch worden omschreven als  $R = W$  : geregistreerde waarde (R) = werkelijke waarde (W). Wil men dit principe toepassen, dan veronderstelt men dat er methoden zijn waarmee kan worden vastgesteld dat de geregistreerde waarde (R) van een variabele gelijk is aan de werkelijke waarde (W). Men dient echter te bedenken dat van veel eigenschappen de werkelijke waarde bij de onderzochte elementen niet kan worden vastgesteld. Daartoe heeft men immers een meetinstrument nodig, en dat staat juist ter discussie. Dit betekent dan tevens dat een definitieve uitspraak over de geldigheid van een meetprocedure veelal niet mogelijk is. In een aantal gevallen zal een conclusie over de geldigheid minder problematisch zijn. Het gaat dan om eigenschappen waarbij de afstand tussen het latente theoretische concept en het manifeste empirische gegeven gering is. Het is dan min of meer voor de hand liggend, dat de meetprocedure zelf weinig of geen twijfels oproept. Eigenschappen zoals leeftijd, gezinssituatie, e.d. zullen met bepaalde procedures zeer adequaat kunnen worden vastgesteld. Men kan daarbij gebruik maken van administratieve bronnen ; voor leeftijd bijvoorbeeld de gegevens uit de burgerlijke stand.

Bij de meeste eigenschappen waarin de onderzoeker is geïnteresseerd, ligt het echter niet zo simpel. We kunnen bijvoorbeeld niet op dezelfde eenvoudige wijze zoals bij leeftijd aannemelijk maken dat iemands arbeidssatisfactie zonder meer kan worden afgeleid uit de duur van de tewerkstelling in een onderneming. Omdat hier de meetprocedure zelf ter discussie staat zal het niet mogelijk zijn de geldigheid ervan rechtstreeks en definitief aan te tonen. Voor deze situatie zal men derhalve op andere wijze moeten trachten iets te weten te komen over de geldigheid. Voor dat doel zijn een aantal *validatie-methoden* ontwikkeld, die op indirecte wijze indicaties kunnen opleveren voor de geldigheid van een meting.

##### 2.2.1.4.3.3.1. "Face"-validatie

Andere termen voor "face"-validiteit zijn : *logische geldigheid*, "sense validity". Met "face"-validiteit wordt verwezen naar de subjectieve beoordeling van de validiteit. Het verband tussen het begrip-zoals-bedoeld (theoretische eigenschap) en het begrip-zoals-bepaald (empirische variabele) wordt in deze methode beargumenteerd door erop te wijzen dat de 'inhoud' van de indicator overeenkomt met de 'inhoud' van de theoretische eigenschap. Het gaat hier dus om het aannemelijk maken dat de indicator 'lijkt op' het theoretisch begrip. De onderzoeker legt hier verantwoording af van de beslissing om een bepaald meetinstrument te kiezen. Het is dan ook mogelijk dat het kritisch forum van de onderzoeker zijn procedure in die zin evalueert.

Bij het begrip 'arbeidsbereidheid', bijvoorbeeld, kan het zoekgedrag van werklozen opgevat worden als een indicator die "face"-validiteit bezit t.a.v. het theoretische begrip (arbeidsbereidheid). De methode van face-validatie houdt o.m. een theoretische bezinning op en verheldering van de betekenis of inhoud van het bedoelde begrip in en het onderkennen en specificeren van de theoretische dimensies van het begrip (het blootleggen van de componenten van een verschijnsel). Het doordacht kiezen van indicatoren voor elk van de onderscheiden dimensies is van belang. Het gaat hier dus om de beargumenteerde

opvatting van de onderzoeker dat een bepaalde indicator hem op het spoor brengt van de theoretische inhoud van de eigenschap of dimensie. Het beoordelen van de wijze waarop de verschillende indicatoren worden samengevoegd tot een waardebeoordeling van de variabele is eveneens relevant. Deze samenvoeging dient zodanig te geschieden dat het relatieve gewicht van de indicatoren tot zijn recht komt en dat de relatie tussen de onderscheiden dimensies adequaat wordt weergegeven.

"Face"-validatie, het woord zegt het al, is het op uiterlijke betekenis of manifeste inhoud beoordelen van indicatoren als representant van een theoretisch concept of dimensie. Hoe zorgvuldig men echter ook te werk gaat, de methode bevat onmiskenbaar een subjectief moment. Deze subjectiviteit zal echter niet verworden tot louter willekeur omdat ook onderzoekers de "repressieve tolerantie" van hun sociaal systeem zullen ervaren. De keuze van indicatoren door een onderzoeker kan immers heel goed niet ernstig worden genomen door het publiek van collega's. Het is dan eigenlijk gedaan met de vermeende "face"-validiteit, omdat de indicering allerwege wordt aangevochten of, wat nog erger is, door niemand wordt overgenomen.

#### 2.2.1.4.3.3.2. Criterium-validiteit

De twee varianten die bij *criterium-validiteit* (ook *pragmatische validiteit* genoemd) worden onderscheiden (nl. soortgenoot-validatie en predictieve validatie), hebben gemeenschappelijk dat met behulp van een externe maatstaf (criterium) getracht wordt een uitspraak over de geldigheid van een meetprocedure te ondersteunen. De kern van deze methode bestaat dan ook uit het vergelijken van de meetresultaten van de te valideren meetprocedure met andere meetresultaten bij dezelfde elementen. De vergelijking heeft dan ofwel betrekking op de uitkomsten van een soortgelijk instrument (*soortgenoot-validatie*) ofwel op de bevindingen bij een verwant verschijnsel, dat met het bewuste instrument kan worden voorspeld (*predictieve validiteit*).

Bij de methode van de soortgenoot-validatie ("concurrent validity") worden de uitkomsten van de te beoordelen meetprocedure vergeleken met de resultaten van een reeds bestaand meetinstrument, dat dezelfde eigenschap meet en het predicaat 'geldig' reeds bezit. Als de uitkomst van beide meetprocedures gelijk is, kan het nieuwe instrument ook geldig worden genoemd.

Predictieve validatie wordt toegepast wanneer een variabele uitdrukkelijk bedoeld is om iets anders, een criterium-variabele, te voorspellen. De correlatie tussen voorspeller en criterium is dus van beslissende betekenis en kan dienen als een operationele definitie van predictieve validiteit. Een meetprocedure is dus geldig, als de uitkomsten ervan de onderzoeker in staat stellen om een andere eigenschap of gebeurtenis adequaat te voorspellen.

Bij predictieve validiteit is er dus nog een andere variabele nodig naast die waar het om gaat (de voorspeller). De andere variabele fungeert dan als criterium voor de geldigheid van de voorspeller. Opgemerkt moet worden dat deze validatie-methode strikt genomen niet voldoet aan de definitie van geldigheid. Het gaat er bij predictieve validatie immers niet zozeer om na te gaan of het instrument het theoretische begrip goed meet, maar of het instrument 'werkt'; d.w.z. of het kan voorspellen ongeacht wat het zelf precies meet.

#### 2.2.1.4.3.3.3. Begripsvalidatie

De Engelse term voor deze validatie-methode is "construct validation". Met "construct" wordt dan bedoeld dat de theoretische eigenschap, die geacht wordt te zijn afgebeeld in het

meetinstrument, niet zo maar kan worden aangewezen of geïdentificeerd met een bepaald verschijnsel ; het gaat juist om abstracties, om "constructs". Voor een bepaald theoretisch concept komt dit hierop neer, dat het in een theorie is gerelateerd aan een aantal andere eigenschappen. Indien men nu op goede gronden mag aannemen dat de theorie juist of houdbaar is, dan zal de operationalisering van een begrip *construct-validiteit* bezitten als de relaties, die dat begrip in theorie heeft met andere begrippen, ook in feite worden aangetroffen bij toepassing van het betreffende meetinstrument. Het meest kenmerkende van deze methode is dus, dat het totale patroon van relaties, dat een bepaalde eigenschap heeft met andere eigenschappen, als criterium wordt genomen voor de geldigheid van de operationalisering van die eigenschap. Het kritieke punt van deze methode is de assumptie dat de theorie juist of houdbaar is. Meestal immers is de houdbaarheid van de theorie een punt van onderzoek en dan is het niet geoorloofd om voor het aantonen van de geldigheid van het instrumentarium eerst te veronderstellen dat de theorie klopt en vervolgens met datzelfde instrumentarium aan te tonen dat de theorie (al of niet) klopt. De methode zal wel bruikbaar zijn bij theorieën die reeds in voldoende mate zijn getoetst. Immers, alleen dan heeft het zin om aan te nemen dat de theorie juist is. In dat geval kan men inderdaad op legitieme wijze concluderen tot de ongeldigheid van een nieuw meetinstrument, namelijk als het niet dezelfde relaties vaststelt, die reeds bleken te gelden.

## 2.2.2. De keuze van de methode tot gegevensverzameling

Er zijn twee methoden om additionele gegevens te verzamelen, nl. de *communicatie* (vraaggesprekmethode) en de *observatie*. Bij zowel de communicatie als de observatie wordt in marktonderzoek gebruik gemaakt van *kwantitatieve en kwalitatieve onderzoeksmethoden*.

### 2.2.2.1. Kwantitatieve en kwalitatieve onderzoeksmethoden

Als men zich verdiept in de verschillen en overeenkomsten tussen kwantificerend en kwalitatief-interpretierend onderzoek is een van de punten die de aandacht trekken dat in kwantificerend onderzoek *verklaren* als een van de meest voor de hand liggende doelstellingen van onderzoek wordt gezien, en dat de term dan ook in vrijwel elke onderzoeksrapportage opduikt, terwijl in het interpreterende onderzoek dit veel minder frequent het geval is ; hier bestaat een voorkeur voor zachtere termen zoals *interpreteren*, *beschrijven* en *begrijpen*.

Tot op zekere hoogte kan gesteld worden dat het onderscheid tussen interpreterend onderzoek en kwantificerend onderzoek terug te voeren is tot het onderscheid tussen *beschrijvingsproblemen* ("wat is vragen" : diagnose, inventarisatie) en *verklaringsproblemen* ("waarom vragen", "hoe-het-komt-vragen"). Door interpretatieve onderzoekers - wier werk weliswaar in belangrijke mate bestaat uit het geven van gedetailleerde beschrijvingen wordt evenwel niet zelden ontkend dat zij enkel beschrijvend bezig zijn (volledigheidshalve dient aangestipt te worden dat kwantitatieve onderzoekers zich niet alleen met verklaringsproblemen bezighouden ; een verwijzing naar bv. de werkzaamheden van het N.I.S. mag hier volstaan). Zij houden staande dat er in hun werk zowel van beschrijving als van verklaring sprake is ; het scherpe onderscheid tussen beschrijven en verklaren wordt door hen van de hand gewezen. De reden hiervoor is dat interpretatieve onderzoekers aanvoeren dat een nadere beschrijving van een situatie of van een proces fungeert als verklaring voor bv. gedrag. Interpretatief onderzoek kenmerkt zich door identificatie, beschrijving of door zogenaamd alledaagse verklaring. Typerend voor interpretatief onderzoek is de aandacht voor de *subjectieve betekenisverlening door de onderzochten* aan hun sociale en fysieke situatie, de subjectieve interpretatie en verklaring van sociale processen waarin zij betrokken zijn. Kijken we nu even hoe - overwegend - in het kwantificerende

onderzoek, en hoe in het interpretatieve onderzoek, vorm wordt gegeven aan het doel : verklaren.

In het kwantificerende onderzoek neemt verklaren vaak de vorm aan van het vinden van variantie in een of andere centrale afhankelijke variabele door voorafgaande variabelen die verondersteld worden als oorzaken te fungeren. Tegen het verklaren in termen van variabelen, reageren interpreterende onderzoekers door hun afschuw uit te drukken tegen positivistische opvattingen. Het begrip 'causaliteit' wordt daarbij vaak als irrelevant opzij geschoven en de creatieve, betekenisverlenende mens verrijst als een moreel en wetenschappelijk tegenbeeld aan de horizon. Op zichzelf is een dergelijke tegenstroom welkom, en in ieder geval zeer begrijpelijk als reactie op een in methoden en technieken verstand verklaren-in-termen-van-variantie. Inderdaad staat het begrip 'causaliteit' bij veel 'modelbouwers' centraal, en het vormt daardoor een steen des aanstoots voor interpretatieve onderzoekers. Zonder meer moet toegegeven worden dat het centrale begrip 'betekenisverlening', de interpretatie van de sociale werkelijkheid door de onderzochte, in het kwantificerende onderzoek volledig onder tafel blijft. Anderzijds dient evenwel opgemerkt te worden dat variabelen een invloed kunnen hebben op gedrag buiten alle interpretaties om. In interpretatieve benaderingen wordt immers nogal eens de onjuiste suggestie gewekt dat alle invloeden lopen via zin- en betekenisverlening door het individu. Ten slotte moet opgemerkt worden dat het kwantificerende onderzoek gericht is op de verklaring van het gedrag van sociale categorieën, terwijl de interpreterende onderzoeker veel meer geïnteresseerd is in de verklaring van het gedrag van deze persoon in deze situatie. Het hoeft dan ook geen verbazing te wekken dat in interpreterend onderzoek veel meer aandacht bestaat voor individuele betekenisverlening. De geringe interesse hiervoor in kwantificerende kring heeft niets te maken met het ontkennen van het belang van betekenisverlening en interpretatie, maar met de opvatting dat op geaggregeerd niveau verschillen in individuele interpretaties tegen elkaar wegvallen. Hoe staat het dan met het verklaren in het interpretatieve onderzoek? Sommige interpretatieve onderzoekers wijzen het verklaren als zodanig af. Men legt er de nadruk op dat het gedrag van mensen dient geïnterpreteerd te worden. Gedrag kan worden begrepen door te verwijzen naar de regels waarnaar mensen zich in het dagelijks leven gedragen. Men moet zich in de visie van deze interpretatieve onderzoekers niet bezig houden met causaal verklaren maar wel met interpreteren, d.w.z. het opsporen van de regels, voor wie ze gelden, hoe ze veranderen, in welke mate afwijkingen toegelaten zijn, e.d. De aanhangers van het "regelgeleide gedrag"-paradigma plegen bv. op te merken dat het verband tussen het op rood springen van een verkeerslicht en het stoppen van de verkeersstroom niet causaal verklaard kan worden. Het kan wel begrepen worden als we de betekenis, die mensen aan verkeerslichten en verkeersregels geven, kennen.

Soms wordt door aanhangers van een interpretatieve benadering het verklaren niet zozeer afgewezen, maar geeft men de voorkeur aan het begrip interpreteren omdat het verlenen van zin, van betekenis, centraal staat. Deze betekenisverlening door het individu ontstaat volgens de symbolisch interactionistische traditie in, en is gevolg van, de interactie, en leidt op haar beurt tot hernieuwde interactie. Vandaar ook dat rapportages van interpreterend onderzoek meestal het karakter aannemen van vloeiende procesbeschrijvingen.

In de methodologische literatuur is er een zekere traditie ontstaan om het *kwalitatieve onderzoek* tegenover het *survey-onderzoek* te plaatsen. Van het kwalitatieve onderzoek wordt dan gemeld dat het gebaseerd is op symbolisch-interactionistische beginselen. Deze voorstelling van zaken is evenwel aanmerkelijk te simpel, omdat geen rekening wordt gehouden met wezenlijk verschillende vormen die binnen het kwalitatieve onderzoek te onderscheiden zijn. Van grote betekenis in dit verband is het onderscheid tussen enerzijds het *kwalitatieve survey* en anderzijds het *kwalitatieve veldonderzoek*. In het kwalitatieve survey probeert de onderzoeker te achterhalen hoe een bepaalde categorie mensen hun situatie

ervaart. Uiteraard moet die situatie door de onderzoeker eerst nader gespecificeerd worden. De respondenten worden ondervraagd over hun situatie. In principe kan dit ook met de methoden die gebruikelijk zijn in het gewone survey en die samengevat kunnen worden onder het begrip 'gestandaardiseerde vragenlijst'. In het kwalitatieve survey werkt de onderzoeker echter niet met een gestandaardiseerde vragenlijst. De bedoeling is, dat de ondervraagden hun ervaringen formuleren in zelf gekozen begrippen in plaats van begrippen die door de onderzoeker zijn gekozen. In het kwalitatieve survey speelt de onderzoeker niet alleen de rol van passieve luisteraar, vol begrip en empathie, maar hij heeft ook een initiërende rol.

Hij werkt met zogenaamde "sensitizing concepts", *attenderende begrippen*. Dit zijn begrippen die in het dagelijks spraakgebruik een brede inhoud hebben. De bedoeling is nu dat de respondenten aan dit begrip een nadere invulling geven door een reflectie op hun eigen ervaringen. De onderzoeker laat de operationalisering van het begrip dus over aan de respondenten. Hier ligt m.a.w. een wezenlijk onderscheid met het gewone survey. Ook in het gewone survey werkt de onderzoeker met begrippen, maar deze begrippen heeft de onderzoeker zelf geoperationaliseerd alvorens de respondenten te benaderen.

Afgezien van dit elementaire onderscheid tussen een kwalitatief en kwantitatief survey zijn voor het overige de overeenkomsten veel groter dan de verschillen. In beide vormen van survey-onderzoek moet de onderzoekspopulatie nauwkeurig worden afgebakend. Als we dit achterwege laten of onzorgvuldig doen, dan weten we later niet wat het generalisatiebereik is van het onderzoek. In beide onderzoeksvormen moet de steekproef nauwkeurig worden getrokken, bij voorkeur op basis van aselectiviteit, waardoor het in de analysefase mogelijk wordt toevallige effecten te onderscheiden van systematische effecten. In beide vormen van onderzoek moeten de observatiegegevens gecodeerd worden teneinde ze toegankelijk te maken voor statistische analyse.

Wat het coderen betreft, is er natuurlijk wel een verschil tussen een kwalitatief en een kwantitatief survey. Wanneer de in het onderzoek gehanteerde begrippen vooraf zijn geoperationaliseerd door de onderzoeker, dan is het coderen een uiterst eenvoudig transformatieproces. Wanneer echter de onderzoeker heeft gewerkt met attenderende begrippen die door de respondenten zelf zijn geoperationaliseerd, dan vraagt de codering meer aandacht. In feite betekent dit een inhoudsanalyse van geregistreerde informatie. De onderzoeker kan deze inhoudsanalyse zelf uitvoeren, maar de validiteit van de resultaten is waarschijnlijk hoger indien wordt gekozen voor een *intersubjectieve inhoudsanalyse*. Dit houdt in dat in eerste instantie door verschillende codeurs onafhankelijk van elkaar wordt gewerkt, en dat in een volgend stadium wordt getracht om daar waar de codeerresultaten niet identiek zijn door middel van een onderlinge discussie tot een consensus te komen. Het kwalitatieve survey moet duidelijk worden onderscheiden van het kwalitatieve veldonderzoek. In het kwalitatieve veldonderzoek gaat het niet om een categorie mensen die via steekproeftrekking uit een goed afgebakende populatie is verkregen. In het veldonderzoek bestuderen we niet alleen situatiedefinities zoals in het kwalitatieve survey, maar ook *interacties*. Daarom heeft het veldonderzoek altijd betrekking op een groep mensen, zodanig afgebakend in tijd en ruimte, dat interactie tussen hen mogelijk is. Bij deze vorm van onderzoek komen we terecht bij het symbolisch-interactionistisch perspectief. De interacties binnen de groep worden mede bestudeerd om te achterhalen op welke wijze situatiedefinities tot stand komen, die op hun beurt weer aanleiding geven tot nieuwe acties. Bij deze vorm van onderzoek ligt de relatie tussen kwantitatieve en kwalitatieve onderzoeksmethoden geheel anders dan bij het kwalitatieve survey. In het veldonderzoek zal de kwalitatieve component dominant zijn. Dat wil niet zeggen dat in het veldonderzoek niet gekwantificeerd zou kunnen worden. Het zou zeer wenselijk zijn als veldonderzoekers een statistisch geweten zouden ontwikkelen, niet alleen om te tellen, maar ook om ruwe schattingen te maken van de kansen dat bepaalde gebeurtenissen zich voordoen. Niettemin



blijft gelden dat veldonderzoekers de onderzoekssituatie niet naar hun hand kunnen en willen zetten.

In hetgeen voorafging zijn drie onderzoeksvormen de revue gepasseerd : het kwantitatieve survey, het kwalitatieve survey en het kwalitatieve veldonderzoek. Op een hoger abstractieniveau kunnen we deze drie onderzoeksvormen vergelijken op basis van de betekenis die wordt toegekend aan de onderdelen van het basisschema van het sociaal-wetenschappelijk onderzoek : *stimulus - interpretatie - respons*.

In het kwantitatieve survey ligt de nadruk in sterke mate op de componenten stimulus en respons. De onderzoeker verschaft de respondent een aantal stimuli en de respondent wordt geacht hierop te responderen. Natuurlijk is de onderzoeker zich er zeer wel van bewust dat de respondent geen antwoordautomaat is, en dat derhalve interpretatieprocessen een belangrijke rol spelen. Een van de redenen dat we in sociaal-wetenschappelijk onderzoek doorgaans lage correlaties tussen variabelen kunnen vinden, is ongetwijfeld gelegen in de grote mate van ruis die aanwezig is bij de toediening van stimuli. Het waakzame bewustzijn van de respondent speelt de survey-onderzoeker parten. Hierop zijn verschillende reacties mogelijk. Als onderzoeker kunnen we deze waakzaamheid negeren, een slechte oplossing, maar wel de meest toegepaste. We kunnen ook trachten de respondent op een dwaalspoor te zetten, een oplossing die vooral in het experiment toepassing vindt. Een derde mogelijkheid is, dat we proberen de respondent weerloos te maken door een bombardement van schaalitems, liefst met het verzoek aan de respondent niet te veel na te denken en vooral spontaan te antwoorden.

Er is echter ook een totaal andere oplossing mogelijk, namelijk door als interviewer zeer bewust in te spelen op het waakzame besef van de respondent. De interviewer treedt dan in overleg met de respondent over het juiste antwoord. Deze werkwijze past bij het kwalitatieve survey. In deze onderzoeksvorm wordt de tweede component van het basisschema, de interpretatie, tot een wezenlijk onderdeel van het eigenlijke onderzoek gerekend. Hoewel we zeker niet de illusie moeten koesteren van gegarandeerd valide onderzoeksresultaten, lijkt deze werkwijze toch perspectieven te bieden. Voorwaarde is echter, dat aan de training van interviewers bijzondere aandacht wordt besteed. Wanneer dit in onvoldoende mate gebeurt, lopen we het risico dat de oncontroleerbare ruis die de observatie in het kwantitatieve survey omgeeft, wordt ingeruild voor eveneens oncontroleerbare systematische foutenbronnen. Interviewerseffecten en sociale wenselijkheid spelen niet alleen een rol in het gestandaardiseerde vraaggesprek, maar evenzeer in het kwalitatieve interview. Misschien zelfs nog in sterkere mate. In het kwalitatieve interview heeft immers de interviewer meer betekenis voor de respondent dan in het gestandaardiseerde vraaggesprek. Het is bepaald niet ondenkbaar, dat naarmate de interviewer als persoon meer betekenis heeft voor de respondent, deze ook meer belang heeft bij het geven van 'prettige' antwoorden. Een goede interviewerstraining biedt hier wellicht een oplossing.

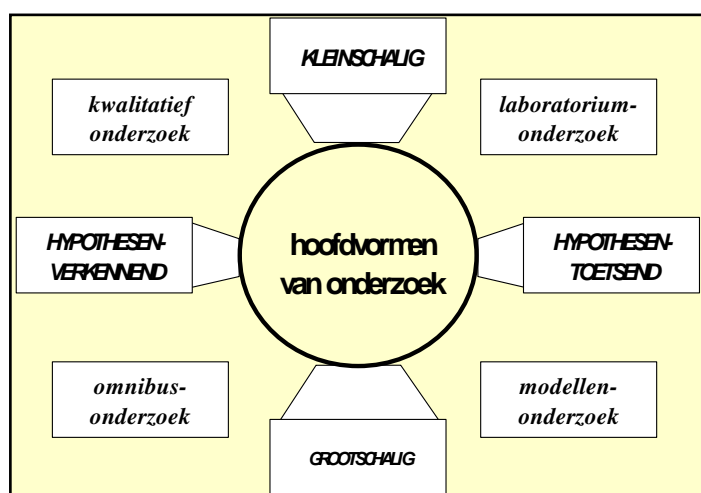
In het kwalitatieve veldonderzoek speelt de interpretatiecomponent een nog belangrijker rol. Zelfs in die mate, dat proponenten van deze onderzoeksmethode soms verkondigen dat zij het stimulus-respons-schema geheel verwerpen. Deze voorstelling van zaken is echter al te simplistisch ; interpretatie is alleen mogelijk als er wat te interpreteren valt, ofwel als er sprake is van stimuli. Het karakter van deze stimuli is in het veldonderzoek echter anders dan in de beide eerder genoemde onderzoeksvormen. De stimuli worden niet door de onderzoeker toegediend, maar het sociale proces waarin de informanten participeren, levert de stimuli. De onderzoeker zal trachten dit proces zo weinig mogelijk te verstoren wanneer hij de interpretatie die de informanten aan die stimuli registreert, en tevens hun reacties probeert te begrijpen in het licht van deze interpretaties.

Naast het zojuist genoemde onderscheid tussen de verschillende onderzoeksvormen is er een tweede aspect dat in de gebruikelijke discussies over onderzoeksmethoden enigszins onderbelicht blijft, namelijk *het rendementaspect*. Het verrichten van onderzoek kost geld. De vraag naar het rendement van onderzoek is daarom gewettigd. Het gaat hier om de eenvoudige constatering dat er een verschil dient te bestaan tussen de situatie bij de aanvang van het onderzoek en na afloop van het onderzoek. Met name bij veel kwalitatief onderzoek laat de vraag zich stellen of de onderzoeker zelf veel heeft geleerd van het onderzoek. Misschien is het leereffect wel aanwezig, maar vaak blijkt het niet uit het onderzoeksrapport. In hoeverre waren de inzichten die in het eindrapport werden gepubliceerd niet reeds bij de aanvang van het onderzoek aanwezig? Het kwantitatieve onderzoek staat in dit opzicht minder onder verdenking. Bij deze vorm van onderzoek is het immers gebruikelijk dat onderzoekers hun hypothesen vooraf formuleren, zodat ze weten wat de stand van kennis was voordat werd begonnen met de materiaalverzameling. Helaas is het formuleren van hypothesen in onderzoek dat voornamelijk op kwalitatieve methoden is gebaseerd, minder gebruikelijk. Het kwalitatieve onderzoek gaat soms te zeer gebukt onder het methodologisch adagium dat de onderzoeker zo onbevoorrecht mogelijk het veld in moet gaan.

Onderzoek is niet alleen het verrichten van nauwkeurige metingen of het empathisch begrijpen van mensen. Onderzoek is in de eerste plaats een onderdeel van een denkproces. Dit houdt in dat de onderzoeker, alvorens het veld in gaan, alle andere kennisbronnen mobiliseert. Het resultaat hiervan is tweeledig. In de eerste plaats bevordert dit een adequate conceptualisering van het onderzoeksprobleem. De kwantitatieve survey-onderzoekers leren onderkennen welke centrale begrippen zij moeten operationaliseren. De kwalitatieve onderzoekers komen op het spoor van de relevante attenderende begrippen waaraan zij tijdens het onderzoek een nadere invulling willen geven.

Het tweede effect van het mobiliseren van reeds aanwezige kennisbronnen is, dat de onderzoeker in tamelijk precieze bewoordingen kan aangeven wat de stand van kennis is bij de start van het onderzoek. Het is van belang dat de onderzoeker een "record" opbouwt van de vorderingen die tijdens het onderzoek zijn bereikt en aldus in de eindrapportage kan laten zien wat het rendement is geweest van onderzoek.

De in de praktijk vaak gehanteerde tegenstelling tussen kwantitatief en kwalitatief onderzoek, leidt tot een onvruchtbare discussie, omdat te veel technieken onder een begrip worden gevat. Tussen omnibus-onderzoek en zgn. "marketing model"-onderzoek, beide vormen van kwantitatief onderzoek, bestaat evenzeer een groot verschil als tussen kwalitatieve technieken als een groepsdiscussie en de Kelly-grid. Tevens blijven daarbij andere vormen van onderzoek zoals laboratorium-onderzoek en marktsimulatie-onderzoek buiten beschouwing. In figuur 7 wordt daarom een positionering gegeven van onderzoek met behulp van de dimensies *kleinschalig* versus *grootschalig* en *hypothesen-ontwikkeland* (exploratief) versus *hypothesen-toetsend*.



Figuur 7 : Een positionering van vier hoofdvormen van onderzoek

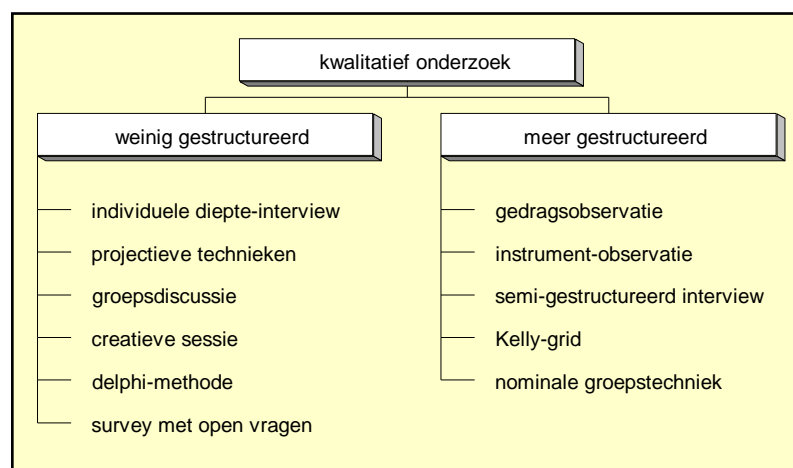
*Kwalitatief onderzoek* wordt gekenmerkt door het kleinschalig, inhoudelijk karakter ervan. De centrale vragen die daarbij door de marktonderzoeker gesteld worden, hebben betrekking op hoe vindt een bepaald gedrag plaats (aankoop, gebruik, e.d.), wat gebeurt er daarbij precies en waarom gebeurt het zo. Inzicht in *de individuele inhoudelijkheid van marktgerelateerd gedrag* kan daarbij verkregen worden.

*Laboratorium-onderzoek* wordt eveneens gekarakteriseerd doordat het gericht is op het 'hoe' en het 'waarom' van marktgedrag. Tevens wordt in een min of meer *gecontroleerde setting* nagegaan wanneer (onder welke soort omstandigheden) bepaalde reacties en gedragingen optreden. Hypothesen kunnen getoetst worden, gericht op *de verklaring van individueel marktgedrag*.

Onderzoek gericht op de constructie van *markt- of marketingmodellen* maakt doorgaans gebruik van grote databestanden. Daarom worden relaties tussen variabelen of verschijnselen expliciet weergegeven, welke bij toetsing aan de data antwoord kunnen geven op vragen naar hoe vaak en in welke mate verschijnselen zich voordoen. Aangezien deze modellen veelal op grootschalig, geaggregeerd niveau gespecificeerd worden, staat *de voorspelling van marktverschijnselen* als doelstelling voorop.

Uiteraard zijn er tal van tussenvormen van onderzoek. Zo vindt survey-onderzoek veelal plaats aan de hand van een ad hoc-onderzoeksmodel en neemt een positie in tussen omnibus- en modellen-onderzoek. Open vragen-surveys (cfr. kwalitatief survey) nemen een plaats in tussen kwalitatief onderzoek en omnibus-onderzoek.

In hetgeen volgt richten we ons op het linkerbovenkwadrant van het positioneringsschema. In figuur 8 onderscheiden we een aantal subgroepen van *kwalitatieve onderzoekstechnieken*.



Figuur 8 : Een positionering van technieken van kwalitatief onderzoek

#### 2.2.2.1.1. Diepte-interview

Het doel van deze techniek is het vinden van achterliggende factoren die het gedrag van de consument beïnvloeden. Na het inleiden van het thema wordt het interview ongestructureerd doorgevoerd. De respondent geniet een grote vrijheid bij het aansnijden van aspecten, die hem of haar m.b.t. het gestelde thema van belang lijken. De interviewer heeft de mogelijkheid tijdens het gesprek met directe of indirecte vragen te interveniëren.

De ongestructureerde vorm van het gesprek heeft voor- en nadelen. De voordelen van de techniek, vooral t.o.v. gestructureerde interviews is o.m. dat doordat het denken en de antwoorden van de respondenten niet in een nauw getrokken raamwerk geplaatst worden, voor de onderzoeker de mogelijkheid bestaat zich een beter beeld van de werkelijke houding van de participant te vormen. Het diepte-interview is flexibel.

Buiten het feit om dat hier geen generaliseerbare uitspraken mogelijk zijn, zijn er vooral twee nadelen te noemen : het gevaar bestaat dat de respondenten door de interviewer beïnvloed worden en de interpretatie gebeurt vanuit de specifieke achtergrond van de interviewer.

#### 2.2.2.1.2. Groepsdiscussie ("focus group")

Bij deze bekende kwalitatieve techniek leidt een moderator een groep van acht tot twaalf participanten door een discussie. De leiding van de discussie dient op een indirecte manier plaats te vinden. Net als in een diepte-interview probeert de moderator achterliggende motieven, die het koopgedrag beïnvloeden, te vinden.

Enkele voordelen van de groepsdiscussie t.o.v. het diepte-interview zijn :

- grotere spontaniteit van de participanten ;
- emotionele inspiratie ;
- generatie van meer ideeën door interactie.

De groepsdiscussie is te beschouwen als een exploratieve, verkennende onderzoekstechniek. De groepsdiscussie wordt o.m. toegepast als de meningen en gevoelens, die onder de consumenten t.o.v. producten, reclameboodschappen of andere communicatie-uitingen

leven, bij de producent grotendeels onbekend zijn. Een uitbreiding van het onderzoek met kwantitatieve technieken is op een later tijdstip mogelijk.

Een groepsdiscussie kan met name gebruikt worden voor :

- het ontwerpen van hypothesen ;
- het ontwerpen van vragenlijsten ;
- het vinden van belangrijke producteigenschappen ;
- het overbruggen van de kloof tussen marketing management en consument.

Maar ook aan de groepsdiscussie zijn nadelen verbonden, o.m.

- de beperkte mogelijkheid tot generaliseerbare uitspraken te komen ;
- groepsinvloeden kunnen inhiberend werken.

Groepsgesprekken en interviews hebben een complementair karakter. De eerste bieden het voordeel veel informatie op te leveren in een korte tijd tegen relatief beperkte kosten. Zij zijn vooral geschikt voor het verkrijgen van een globaal overzicht. Het sociale karakter van een groep heeft voordelen (onderlinge interacties, stimulus) en nadelen (controle, competentiestrijd). Groepen lenen zich niet goed voor het verkrijgen van details. Individuele diepte-interviews worden gebruikt om individuele processen in kaart te brengen en om details te verzamelen. In interviews is er minimale beïnvloeding door anderen. Privacy- of concurrentiegevoelige onderwerpen zullen dus eerder in interviews worden behandeld.

Bij algemene "usage and attitude"-studies zullen eerder groepsdiscussies dan interviews gebruikt worden. Is het echter belangrijk om precies te weten hoe het koopproces stap voor stap verloopt, dan zijn interviews aan te raden. In "business-to-business" onderzoek worden groepsdiscussies minder gebruikt dan interviews ; vaak zijn de respondenten immers concurrenten van elkaar en laat men in het bijzijn van 'collega's' het achterste van de tong niet zien. Binnen de methoden van het groepsgesprek en het diepte-interview worden diverse vraagtechnieken gehanteerd. In veel gevallen zal een kwalitatief onderzoeker iedere keer opnieuw het vraaginstrument maken. Er zijn natuurlijk ook meer gestandaardiseerde vraagtechnieken, zoals *laddering* (bij elk antwoord van de respondent wordt doorgevraagd naar de reden/het belang van dat antwoord, om naar dieper gelegen waarden en drijfveren te peilen) en "grouping" ("sorting" : merken of producteigenschappen laten groeperen en onderlinge verschillen en overeenkomsten laten aangeven om de onderscheiden dimensies op te sporen).

In groepsgesprekken en interviews wordt ook vaak gebruik gemaakt van *projectieve technieken*, waarbij de consument via een omweg naar de mening wordt gevraagd. Projectieve technieken werken vooral goed in imago-onderzoek, positionerings- en communicatie-onderzoek. Bekende projectieve technieken zijn :

- woordassociatie ;
- personificatie ;
- "thematic apperception test" ;
- fotosortering ;
- rollenspel ;
- "speech bubbles".

Non-verbale projectieve technieken zijn bijvoorbeeld het laten maken van tekeningen of collages over een product of merk. Bij non-verbale technieken legt de ondervraagde zelf uit wat het maaksel voorstelt en waarom men bepaalde kleuren of objecten heeft gebruikt.

Een belangrijk voordeel van het diepte-interview en de groepsdiscussie is de 'levensechtheid', het dicht bij de leefwereld van de consument blijven staan. Kwalitatief onderzoek is daarom vrij valide. Als nadeel geldt dat interviews soms onderling sterk verschillen. Betrouwbaarheid in de vorm van vergelijkbaarheid is dan ook geen sterke kant van kwalitatief onderzoek. Voor kwalitatief onderzoek moet betrouwbaarheid op een andere manier worden gedefinieerd.

#### 2.2.2.1.3. Creatieve sessie ("brainstorming")

Deze verbale groepstechniek wordt toegepast bij het ontwikkelen van nieuwe ideeën. In tegenstelling tot de groepsdiscussie, die de medewerking vraagt van een homogene groep van participanten, wordt hierbij de voorkeur gegeven aan een heterogene groep, veelal van deskundigen. Een creatieve sessie kan meer tijd in beslag nemen dan een groepsdiscussie en kan eveneens uit meerdere bijeenkomsten bestaan. Door de heterogene samenstelling van de groep, wordt het optreden van het *synergie-effect* verwacht, d.w.z. dat de creatieve sessie tot kwalitatief betere resultaten leidt dan het reflecteren van iedere deelnemer afzonderlijk. Vandaar ook dat de creatieve sessie vaak wordt aangewend bij de ontwikkeling van nieuwe productideeën.

Een voordeel van de creatieve sessie in vergelijking met de groepsdiscussie is de grotere kans op het optreden van het synergie-effect. Een speciaal nadeel is de beïnvloeding van de groep door dominante groepsleden.

#### 2.2.2.1.4. Delphi-methode

De Delphi-methode, oorspronkelijk ontwikkeld door de RAND-Corporation, heeft als belangrijkste karakteristiek dat door een serie van anonieme schriftelijke vragenlijsten over een onderwerp een consensus tussen experts wordt ontwikkeld. In de initiële vragenlijst wordt een brede vraagstelling opgenomen, waarna de antwoorden verzameld en samengevat weer terug naar de participanten gaan. Tussen de drie en vijf vragenlijsten worden zo gehanteerd. Daarmee wordt het proces zolang herhaald totdat een consensus bereikt is of het probleem voldoende beantwoord is. Een typische steekproef omvat 20 tot 30 experts, echter in de literatuur worden steekproeven van 4 tot over 100 gesignaleerd. De opeenvolgende vragenlijsten worden steeds verder gestructureerd, bijvoorbeeld door de voor- en nadelen van een probleem of oplossing te laten schalen.

Als voordelen van de Delphi-methode kunnen genoemd worden :

- moeilijk bereikbare personen kunnen ingeschakeld worden ;
- geen groepsinvloeden, conformiteitsdruk ;
- een groot aantal ideeën kunnen gegenereerd en geëvalueerd worden ;
- de schriftelijke procedure dwingt tot een zorgvuldiger formulering.

Nadelen zijn o.m. :

- de procedure is zeer tijdsintensief ;
- geen sociale beloning, waardering voor de participanten ;
- een hoge motivatie van de zijde van de participanten om de totale procedure af te maken ;
- een hoge schrijfvaardigheid is vereist ;
- de ongestructureerdheid van de initiële vragenlijsten kan als frustrerend ervaren worden.

#### 2.2.2.1.5. Gedragsobservatie

De gedragsobservatie is een niet-verbale onderzoekstechniek. Het voordeel van deze techniek ten opzichte van interviewtechnieken is, dat het werkelijke gedrag van mensen geobserveerd wordt en niet dat wat mensen zeggen te willen doen of gedaan hebben.

Vermeldenswaard in dit verband is de opkomst van de "behavioral process approach", de gedragsprocesbenadering in recent wetenschappelijk consumentenonderzoek. Daarbij wordt gesteld dat de geringe verklaringskracht van images, attitudes, e.d. en informatieverwerkingskarakteristieken, wanneer het gaat om bijvoorbeeld winkelkeuze en productkeuze, gelegen is in het feit dat bij de puur cognitieve, verbale benaderingen voorbijgegaan wordt aan de wijze waarop deze keuzen ingebed zijn in 'dagelijkse' gedragspatronen. Naarmate deze opvatting terrein gaat winnen, zal de vraag naar gedragsobservatietechnieken ook in toegepast onderzoek toenemen.

Bij deze observatietechnieken kan een onderscheid gemaakt worden tussen ongestructureerde observatie met behulp van "one-way screens", televisiecamera's en gestructureerde observatie met behulp van observatieschema's.

#### 2.2.2.1.6. Instrument-observatie

Onder instrument-observatie wordt hier verstaan het observeren van specifieke aspecten van gedrag, zoals oogbewegingen, huidweerstand, hartslag, e.d. met behulp van daartoe geëigende apparatuur. Wanneer er in de onderzoekssituatie gecontroleerd wordt voor storende factoren, is sprake van een laboratorium-experiment (in eenvoudige opzetten wordt gesproken van exploratief, kwalitatief onderzoek).

#### 2.2.2.1.7. Kelly-Grid

Deze techniek, ook wel "repertory-grid" genoemd, is gebaseerd op de door Kelly ontwikkelde *persoonlijke "construct"-theorie*. Kelly veronderstelt daarbij dat mensen in de loop van hun leven persoonlijke "constructs" of dimensies ontwikkelen, waarmee ze objecten bejegenen. In marktonderzoek worden deze dimensies vaak als attributen van producten beschouwd.

De procedure bij de Kelly-grid is als volgt te omschrijven : de respondent bekijkt enige tijd een lijst van stimuli (bv. producten, merken, reclameboodschappen) en schraapt die welke hem onbekend zijn. Vervolgens worden groepjes van drie stimuli voorgelegd en de opgave van de participant is aan te geven waarom hij twee stimuli als identiek (c.q. bij elkaar passend) en gelijktijdig verschillend van de derde beschouwt. Deze procedure wordt herhaald en eindigt als de respondent geen nieuw criterium ter beoordeling meer kan aandragen. Het doel van deze techniek is het vinden van "constructs" of dimensies, waarin de perceptie van objecten of stimuli door een groep van mensen vertaald kan worden. De mogelijkheid bestaat de dimensies in vijf- of zeven-puntenschalen te vertalen en de respondent de stimuli in de zo verkregen schalen in te doen passen. De met behulp van de Kelly-grid verzamelde data zijn geschikt als input voor multivariate analyses.

#### 2.2.2.1.8. Survey met open vragen

In tegenstelling tot het survey met gesloten vragen, dat samengaat met een grote, representatieve steekproef en een gestandaardiseerde vragenlijst, is het survey met open

vragen in beginsel kwalitatief van aard. Het doel is, net als bij vele andere technieken, inzicht in de attitudes en motivaties van mensen te verkrijgen. Natuurlijk bestaan tussen een survey met open vragen en het gestructureerde kwantitatieve survey gradaties.

Het hanteren van open vragen maakt het voor de respondent mogelijk in zijn eigen taal te reageren. Dit verhoogt bij nauwkeurig geformuleerde vragen de validiteit en sensitiviteit van de techniek.

#### 2.2.2.1.9. Nominale groepstechniek

Deze recente groepstechniek kent zes fasen. De fasering is als volgt. De groep wordt met een uitspraak omtrent het te behandelen probleem geconfronteerd. Iedere deelnemer zet zijn gedachten hierover op papier. Achtereenvolgens wordt iedere respondent gevraagd zijn ideeën op een bord neer te schrijven en deze gelijktijdig aan de groep toe te lichten. De geuite meningen worden geëvalueerd en van een code voorzien. Iedere deelnemer kiest nu de zijn inziens overtuigende meningen en rangschikt deze naar belangrijkheid. Tenslotte worden de resultaten geaggregeerd.

Op deze manier worden twee soorten output verkregen, nl. meningen die van belang zijn voor het probleem en gekwantificeerde metingen van het relatieve belang van de verschillende meningen. De nominale groepstechniek, die in de exploratieve fase van een onderzoek gebruikt wordt, heeft een zekere verwantschap met de Kelly-grid omdat ook hierin wordt gezocht naar dimensies waarmee objecten geëvalueerd worden.

#### 2.2.2.2. Communicatie

Bij de communicatie of vraaggesprekmethodode wordt een onderscheid gemaakt tussen *informele en formele methoden*. Tabel 9 verduidelijkt de voornaamste verschillen tussen beide.

Tabel 9 : Vraaggesprekmethoden

criteria	informeel	formeel
soorten	⇒ persoonlijk interview ⇒ Delphi-methode ⇒ groepsdiscussie ⇒ groepsinterview	enquête : ⇒ persoonlijk ⇒ telefonisch ⇒ per post
afname	geïndividualiseerd	gestandaardiseerd
vorm	vrij of semi-gestructureerd	gestructureerd
doel	kwalitatieve uitspraken	kwalitatieve uitspraken

#### 2.2.2.3. Observatie

Een andere manier om additionele gegevens te verkrijgen is observatie, het waarnemen van mensen en van materiële zaken. Evenals bij communicatie zijn we in staat een classificatie te maken van methoden om te observeren, nl. op basis van wat we waarnemen en met behulp



waarvan we waarnemen. De objecten die kunnen geobserveerd worden zijn mensen en/of materiële zaken (bv. prijzen, voorraden). Het observeren zelf kan worden gedaan door mensen of daartoe geconstrueerde instrumenten (bv. audimeter).

Het observeren van mensen wordt *micro-observatie* genoemd. Tevens onderscheiden we het observeren in de werkelijkheid (bv. het observeren van winkelgedrag) en in een laboratoriumsituatie (bv. het observeren van de deelnemers aan een groepsgesprek vanuit een speciale ruimte in een marktonderzoek- of reclamebureau).

In hetgeen volgt bespreken we de micro-observatie en het waarnemen in een experimentele situatie. In dit kader wordt een onderscheid gemaakt tussen een ongecontroleerd en een gecontroleerd experiment.

#### 2.2.2.3.1. Micro-observatie

Om het gedrag van mensen waar te nemen kan zowel van het interview als van de observatie gebruik gemaakt worden. Wanneer een detaillist informatie wenst te bekomen van het koopgedrag van zijn klanten dan bestaat een eerste mogelijkheid erin mensen te interviewen die op het punt staan de winkel binnen te treden. Een nadeel verbonden aan een dergelijk interview is dat het feitelijk gedrag van mensen niet altijd in overeenstemming is met het voorgenomen gedrag. Onverwachte impulsen kunnen immers het voorgenomen gedrag van de klant in de winkel verstoren. Daartoe neemt een andere onderzoeker het gedrag van de klant in de winkel waar. Samen geven de gegevens van het interview en de observatie een 'vrij zuiver' beeld van het gedrag van de consument.

Een van de voordelen van het waarnemen is dat we het feitelijk gedrag van mensen te weten komen. Dit voordeel bestaat niet bij het interview waarin we moeten afgaan op wat mensen zeggen. Een nadeel aan het observeren is uiteraard dat we uitsluitend uiterlijk waarneembare kenmerken van mensen in de beschouwing kunnen betrekken. De achterliggende motieven van consumentengedrag blijven buiten beschouwing. Bovendien is het waargenomen gedrag voor meerdere interpretaties vatbaar.

Voorbeelden van micro-observatie verricht door mensen :

- \* in de werkelijkheid :
  - het aantal passanten bij een winkel ;
  - observatie van winkelgedrag ;
  - observatie van het gedrag van de verkopers (o.m. door middel van schijnaankopen) ;
- \* in een laboratoriumsituatie (bv.groepsdiscussie, groepsinterview).

Voorbeelden van micro-observatie verricht door instrumenten :

- \* in de werkelijkheid : de televisiecamera ;
- \* in een laboratoriumsituatie :
  - oogcamera (meet veranderingen in de pupilomvang) ;
  - "one-way-screen" ;
  - tachitoscoop (meten van de herkenbaarheid van een reclameboodschap).

#### 2.2.2.3.2. Het waarnemen in een experiment

De verschillende uitvoeringen of werkwijzen ten aanzien van het verkrijgen van gegevens in een experiment worden *proefopzetten* genoemd. Alvorens in te gaan op de proefopzetten voor een ongecontroleerd en een gecontroleerd experiment, blijven we stilstaan bij de

*algemene criteria die bij de beoordeling van de geschiktheid van een (experimenteel) onderzoeksontwerp van belang zijn.*

#### 2.2.2.3.2.1. Criteria ter beoordeling van een onderzoeksontwerp

Met onderzoeksontwerpen beoogt men causale relaties tussen variabelen vast te stellen. Hieruit kan een algemeen centraal criterium worden afgeleid ter beoordeling van een onderzoeksontwerp : *een onderzoeksontwerp is adequaat als het in staat stelt de aan- of afwezigheid van de veronderstelde causale relaties op legitieme wijze vast te stellen.* Ter vaststelling van causaliteit kunnen een drietal vereisten worden afgeleid waaraan een onderzoeksontwerp moet voldoen :

- [1] het ontwerp moet het vaststellen van samenhang mogelijk maken ;
- [2] het ontwerp moet een beslissing over de richting van oorzakelijkheid toelaten ;
- [3] het ontwerp moet de invloed van storende factoren uitschakelen.

Het voldoen aan deze drie vereisten is echter nog niet toereikend om te kunnen generaliseren. Er kan immers strikt genomen slechts geconcludeerd worden tot een causale relatie die geldig is voor de feitelijk onderzochte groep op dat moment, gegeven de gebruikte onderzoeksinstrumenten. Onderzoek heeft echter een ruimere pretentie, hetgeen een tweede algemeen centraal criterium impliceert, nl. dat het onderzoeksontwerp op zijn *generaliseerbaarheid* beoordeeld moet worden.

*Theoretische adequaatheid* en *praktische uitvoerbaarheid*, tenslotte, completeren de centrale overwegingen ter beoordeling van de kwaliteit van een onderzoeksontwerp. Theoretische adequaatheid betekent dat het onderzoeksontwerp aansluit bij het theoretisch model. Praktische uitvoerbaarheid kan worden omschreven in termen van tijd, geld, e.d.

##### 2.2.2.3.2.1.1. Causaliteit

In vorig punt werden voor de beoordeling van de adequaatheid van een onderzoeksontwerp drie groepen criteria vermeld. Het eerste criterium heeft betrekking op de mate waarin het ontwerp het vaststellen van het al dan niet bestaan van een causale relatie toelaat. Deze hoofdvraag valt uiteen in drie deelvragen, nl. die naar het bestaan van samenhang, die naar de causale ordening en die naar het optreden van mogelijk storende factoren. Deze deeleisen komen in hetgeen volgt aan de orde.

#### [1] Samenhang

Van causaliteit is er sprake als variatie in de ene, onafhankelijke variabele variatie in de andere, afhankelijke variabele teweegbrengt. Het vergelijken van tenminste twee situaties die op de onafhankelijke variabele van elkaar verschillen, is dus een minimumvoorwaarde voor het kunnen trekken van causale conclusies. Het onderzoeken van één situatie is dus onvoldoende. Men kan het effect van een reclamecampagne op de verkoop van een product niet vaststellen door de waarde van de afhankelijke variabele (verkoop) uitsluitend te bezien in de periode tijdens en volgend op de reclamecampagne.

De gewenste variatie in de onafhankelijke variabele kan op twee manieren worden bepaald. In de eerste plaats kan men bij dezelfde onderzoeksgroep minstens twee keer in de tijd meten, waarbij de onafhankelijke variabele voor de verschillende tijdstippen verschillende waarden aanneemt (bv. voor en na een reclamecampagne). In de tweede plaats kan men op

hetzelfde tijdstip meten bij minstens twee groepen die verschillend scoren op de onafhankelijke variabele (bv. een groep die blootstaat aan de reclamecampagne en een groep waarvan men weet dat dit niet het geval is).

Het onderzoeksontwerp moet zo zijn ingericht dat de veronderstelde samenhang tussen onafhankelijke en afhankelijke variabele kan worden vastgesteld. Diverse omstandigheden vormen echter een belemmering voor het adequaat kunnen constateren of de bedoelde samenhang al dan niet aanwezig is. Zo is het logisch onmogelijk samenhang te constateren indien er slechts één waarde van de onafhankelijke variabele in het onderzoek betrokken wordt. In sommige onderzoekssituaties wordt dit extreme geval van een 'constante' onafhankelijke variabele echter dicht benaderd. De range van de onafhankelijke variabele mag dus niet beperkt zijn. Wanneer men bijvoorbeeld de relatie onderzoekt tussen schoolprestaties van ASO-leerlingen en universitaire slaagcijfers, dan kan men uit een gevonden zwakke samenhang niet zonder meer concluderen dat de capaciteiten die het ASO van leerlingen vraagt blijkbaar andere zijn dan die de universiteit vraagt. Deze zwakke samenhang kan tot stand komen doordat binnen de groep universiteitsstudenten die het ASO volgden de in het ASO behaalde eindcijfers te weinig variëren om een duidelijke samenhang te doen ontstaan. Omdat binnen de groep ASO-leerlingen die aan de universiteit gaat studeren de range van de onafhankelijke variabele (behaalde ASO-resultaten) beperkt is (bv. louter goede resultaten), kan de onafhankelijke variabele geen grote verschillen in de afhankelijke variabele teweegbrengen. Anderzijds mag ook de range van de afhankelijke variabele niet beperkt zijn. Onderzoek naar bijvoorbeeld de effecten van een reclamecampagne bij een groep die al overtuigd is van het door de campagne aangemoedigde doel, kan nooit grote verschuivingen in overtuiging te zien geven.

Maar ook indien de variabelen in voldoende mate (kunnen) variëren, is het vaststellen van samenhang nog niet zonder problemen. Zo is er de omstandigheid dat sterkte en richting van de samenhang op vele wijzen vastgesteld kunnen worden met behulp van verschillende technieken, die ieder hun eigen assumpties hebben. Uiteraard moet die techniek worden gekozen waarmee de causale hypothese zo adequaat mogelijk kan worden getoetst. Het onderzoeksontwerp dient dan zo te worden ingericht dat deze geprefereerde techniek kan worden toegepast. Enkele voorbeelden. Als het gebruik van variantie-analyse wordt beoogd, dienen de onafhankelijke variabelen onafhankelijk van elkaar te zijn. Indien men tussen opleiding en inkomen een curvilineair verband veronderstelt, dan kan men niet volstaan met twee onderzoeksgroepen, nl. een groep met een lage en een groep met een hoge opleiding, omdat via een dergelijke dichotomie de relatie noodzakelijk lineair zal zijn.

Een ander probleem betreft de vraag in welke mate een geconstateerde samenhang statistisch significant is. Wat algemener geformuleerd, gaat het hier om de vraag of het onderzoeksontwerp zodanig is dat uitspraken over de aan- of afwezigheid van samenhang mede zinvol gebaseerd kunnen worden op statistische toetsings- en schattingsprocedures. Deze procedures veronderstellen o.a. aselecte trekking van de onderzoekselementen, dan wel at random-toewijzing van de onderzoeksgroepen. Hiermee dient uiteraard in de ontwerpfasen van een onderzoek al rekening worden gehouden. Verder worden de uitkomsten van deze procedures beïnvloed door het aantal elementen in de onderzoeksgroepen en door de grootte van de variantie van de afhankelijke variabele. Naarmate de aantallen in de onderzoeksgroepen geringer zijn, neemt de kans op type II-fouten (het ten onrechte accepteren van de nulhypothese) toe. Hetzelfde doet zich voor wanneer de variantie in de afhankelijke variabele vrij groot is, of beter: wanneer het aandeel van de toevalsfactoren in deze variantie beduidend groter is dan het aandeel van de causale factoren. In de ontwerpfasen van een onderzoek moet dus gezorgd worden voor voldoende aantallen onderzoekselementen en moet zoveel mogelijk geprobeerd worden de toevalsvariantie te minimaliseren, o.m. door betrouwbare meetprocedures. Het

verminderden van de toevalsvariatie kan eveneens bewerkstelligd worden door het onder controle houden van storende factoren (cfr. infra).

## [2] Causale ordening

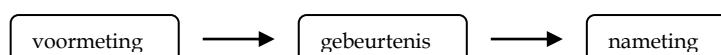
Onderzoek naar causale relaties dient de asymmetrie in deze relatie te verdisconteren. Het asymmetrische houdt in dat de gevonden samenhang tussen X en Y tot stand komt, doordat verschillen in de causale factor leiden tot verschillen in de gevolgvariabele Y, en niet omgekeerd.

De beslissing over de vraag : is X bepalend voor dan wel Y bepalend voor X, kan geldig genomen worden op grond van een tweetal soorten overwegingen. Het gaat dan om *overwegingen van logisch-theoretische aard* en om *argumenten die voortvloeiën uit de opzet van het onderzoek*. De eerste soort overwegingen hebben een a priori karakter : de onderzoeker weet op grond van 'logische' redenering aannemelijk te maken welke de causale richting is van de onderzochte factoren.

Een oplossing voor het probleem van de causale ordening is echter lang niet altijd op bevredigende wijze te vinden langs deze logisch-theoretische weg. In dit soort situaties is het van belang het onderzoek dusdanig op te zetten dat de beslissing over de causale ordening geen willekeurige blijft. Bij het beoordelen van onderzoeksontwerpen op hun vermogen een geldige beslissing te onderbouwen over de causale ordening van de variabelen, is het onderscheid van belang tussen *statische en dynamische onderzoeksstrategieën*. Synoniem voor dit onderscheid zijn de termen : *synchronische* en *diachronische* onderzoeksontwerpen.

In een *synchronisch onderzoeksontwerp* worden X en Y tegelijkertijd gemeten. Via survey-onderzoek stelt men bijvoorbeeld vast of er verschillen in pacifisme bestaan tussen kijkers die een bepaalde film hebben gezien en kijkers die deze film niet hebben gezien. Het is via deze opzet evenwel niet mogelijk uit te maken of eventuele verschillen in pacifisme het gevolg zijn van het al of niet zien van de film. Men blijft aangewezen op a priori vermoedens van logische of theoretische aard.

De grondvorm van het *diachronische, dynamische onderzoek* is die van het doen van verschillende metingen op verschillende tijdstippen bij dezelfde onderzoekselementen. Verschillen in deze na elkaar gedane metingen zal men dan willen toeschrijven aan de werking van de eerder gemeten causale factor. Bij dit ontwerptype wordt gebruik gemaakt van de parallelle die er bestaat tussen de onomkeerbaarheid van het tijdsverloop en de onomkeerbaarheid van de causale relatie : een latere gebeurtenis kan geen oorzaak zijn van hetgeen eerder heeft plaatsgevonden. Onderzoeksontwerpen van het diachronische type kunnen een doorslaggevende argumentatie bevatten over de causale richting. De onderzoeker kan daarbij met name veranderingen in de causale factor zodanig in het onderzoek introduceren, dat een omgekeerde interpretatie van de causale ordening niet meer mogelijk is.



Figuur 9 : Diachronisch onderzoeksontwerp

Aan de hand van bovenstaande opzet kan in principe een geldige beslissing over de causale ordening genomen worden. Het verschil tussen de voor- en nameting kan toegeschreven

worden aan de gebeurtenis. Toch kunnen nog problemen optreden, die verband houden met het optreden van "causal lags". Daarmee wordt bedoeld dat veranderingen in de onafhankelijke variabele niet onmiddellijk hoeven te worden gevolgd door veranderingen in de afhankelijke variabele. Het kan zijn dat er een zekere inwerkperiode nodig is voor de werking van de causale factor. Het kiezen van geschikte tijdstippen kan dan ook belangrijk zijn.

### [3] Afwezigheid van storende factoren

De afwezigheid van storende factoren wordt ook aangeduid met de term interne geldigheid. Een onderzoeksontwerp is dus intern geldig indien het dusdanig is, dat de variatie in de afhankelijke variabele kan worden toegeschreven aan de variatie in de onafhankelijke variabele (de causale factor) in plaats van aan het optreden van ongecontroleerde storende factoren. Onderzoek is er met name op gericht om andere (storende) factoren als alternatieve verklaring zoveel mogelijk uit te schakelen. In dit verband is het van belang een onderscheid te maken tussen *de bronnen van variatie in de afhankelijke variabele*, nl. causale factoren, gecontroleerde storende factoren, ongecontroleerde storende factoren en toevalsfactoren.

De *causale factoren* zijn de onafhankelijke variabelen in de werking waarvan de onderzoeker op grond van hypothesen is geïnteresseerd. Storende factoren zijn variabelen waarvan de effecten verward kunnen worden met die van causale factoren. Een mogelijk *storende variabele* is onder controle als er geen storende invloed meer van uitgaat omdat de onderzoeker de invloed ervan kan bepalen dan wel heeft uitgeschakeld. Bij *ongecontroleerde storende factoren* is verwarring met de invloed van causale factoren wel mogelijk. De hierboven omschreven drie groepen van factoren hebben met elkaar gemeen dat de invloed ervan systematisch is. Daardoor onderscheiden zij zich van de vierde groep, de *toevalsfactoren* waarvan de invloed op de afhankelijke variabele onsystematisch, in principe onvoorspelbaar en "in the long run" gelijk aan 0 is.

Onbetrouwbare meetprocedures kunnen een belangrijke bron van toevalsvariatie vormen. Ook beschouwt men wel het gecombineerde effect van allerlei, niet expliciet in het onderzoek opgenomen, systematische factoren als toevallig. Naast het opvoeren van de betrouwbaarheid van de metingen, is het identificeren en onder controle brengen van systematische factoren een manier om de toevalsvariatie in de afhankelijke variabele te minimaliseren. Hieronder worden een aantal soorten storende factoren vermeld.

In de eerste plaats zijn er *buitenexperimentele gebeurtenissen* ("history"). Het gaat hier om specifieke gebeurtenissen die, losstaand van het eigenlijke onderzoek, plaats vinden tussen het begin en het einde van het onderzoek en van invloed zijn op de afhankelijke variabele. Omdat de gebeurtenis nl. optreedt na de voormeting en voor de nameting, kunnen eventuele verschillen tussen beide metingen niet uitsluitend worden toegeschreven aan de causale factor maar ook aan de invloed van deze storende factoren. Hoe langer een experiment duurt, des te groter is de kans dat er veranderingen in de omgeving optreden. In principe kunnen alle gebeurtenissen die buiten de sfeer van het experiment liggen als 'deel van de historie' worden beschouwd (bv. wijzigingen in de concurrentiesituatie, modeveranderingen; het anders hanteren van de marketing-mix door de concurrentie, e.d.).

Een tweede bron van storende factoren zijn *spontane veranderingen* (het verzadigingseffect, "maturation"). Met spontane veranderingen worden die veranderingen bedoeld die als autonoom proces automatisch verlopen met de tijd. Het gaat dus niet zoals hierboven om de invloed van specifieke gebeurtenissen maar om automatisch verlopende en vanuit de onderzoeksoptiek spontaan optredende processen die van invloed zijn op de afhankelijke

variabele. Een voorbeeld ter verduidelijking : bij lange interviews kunnen de onderzochte personen vermoeid raken, honger of dorst krijgen, waardoor er verschillen tussen voor- en nameting kunnen ontstaan.

Een derde bron van storende factoren wordt het *testeffect* ("testing") genoemd. Dit is het effect van de ene meting op de andere. De voormeting van de afhankelijke variabele kan op allerlei manieren van invloed zijn op de scores van de nameting. Daardoor kunnen er verschillen of overeenkomsten tussen de scores op de voor- en nameting ontstaan die aanzien kunnen worden voor de effecten van de causale factor. Het is mogelijk dat een respondent zijn houding verandert juist onder invloed van vragen die gesteld worden om deze houding te meten. Hierdoor zal een later interview met deze respondent een andere houding te zien geven, afgezien van de werking van de causale factor. Ook verschijnselen als 'neiging tot consistentie' en 'testvaardigheid' worden tot "testing" gerekend. In het eerste geval gaat het om de neiging van een respondent bij de nameting dezelfde antwoorden te geven als bij de voormeting, ook al is in de tussentijd tussen de twee metingen de werkelijke houding veranderd onder invloed van de causale factor. Onder 'testvaardigheid' wordt verstaan het verschijnsel dat onderzoekspersonen bij de nameting bepaalde vragen gemakkelijker kunnen beantwoorden omdat ze bij de voormeting 'geoefend' hebben. Hun score op de test is de tweede keer beter dan de eerste keer, zonder dat noodzakelijkerwijs de kennis of vaardigheden die men wil meten, werkelijk verbeterd zijn. In feite kunnen 'neiging tot consistentie' en 'testvaardigheid' niet tot de bedreigingen van de interne geldigheid gerekend worden, maar tasten zij eerder de geldigheid van de metingen aan.

In de vierde plaats is er het *instrumenteel effect* ("instrumentation"). Indien de ene onderzoekssituatie (in de experimentele groep of bij de voormeting) een ander meetinstrument voor de afhankelijke variabele wordt gehanteerd dan in de andere situatie (in de controlegroep of bij de nameting) dan kunnen eventuele verschillen in meetresultaat het gevolg zijn van deze veranderingen in meetinstrumentarium.

De storende factor *regressie* ("regression") treedt veelvuldig op in die situaties waar de onderzoeksgroepen zijn samengesteld op grond van relatief hoge dan wel relatief lage scores op de afhankelijke variabele. Statistische regressie betekent dat eenheden die de ene keer extreem scoren, de andere keer gemiddeld minder extreem scoren, d.w.z. een regressie naar het gemiddelde ("regression to the mean") vertonen. De storende werking van deze regressie naar het gemiddelde wordt vaak niet onderkend door onderzoekers. Het doelbewust kiezen van een onderzoeksgroep die extreem hoog of laag scoort op de afhankelijke variabele, roept derhalve het gevaar op van een regressie-artefact. Een voorbeeld ter illustratie. In veel evaluatie-onderzoek wordt geprobeerd de effectiviteit vast te stellen van maatregelen, die genomen zijn om ongewenste situaties te verbeteren. Men constateert dat bepaalde kinderen op school achterblijven en besluit deze kinderen volgens een bepaalde methode bij te werken. Of de methode het beoogde resultaat heeft, wordt vervolgens geëvalueerd door de voormeting ( $M_1$ ) te vergelijken met de nameting ( $M_2$ ). Aangezien in deze onderzoeksopzet  $M_1$  een extreem ongunstige score is (achterblijven op school), kan men op grond van het regressiefenomeen verwachten dat  $M_2$  een minder ongunstige score zal zijn dan  $M_1$ . Het regressie-artefact is een concurrerende interpretatie voor de nogal gerechtvaardigd lijkende conclusie dat de maatregelen het beoogde effect tot gevolg hebben. Een oplossing kan worden gezocht in een uitbreiding van de onderzoeksopzet, nl. de toevoeging van een controlegroep, die in alle opzichten gelijk is aan de eerste groep, nl. even extreem ("matching"), maar niet onderworpen aan de te evalueren maatregelen. Ook voor deze controlegroep wordt een voor-en nameting verricht (nl.  $M_3$  en  $M_4$ ). De schijnbare, door regressie veroorzaakte verandering in de controlegroep ( $M_4 - M_3$ ) kan worden afgetrokken van de verandering in de oorspronkelijke groep ( $M_2 - M_1$ ). Men houdt dan als verschil de verandering ten gevolge van de maatregelen over. Er dient hier

echter aan toegevoegd te worden dat onderzoeksopzetten waarbij groepen vergelijkbaar worden gemaakt door matching, eveneens een groot gevaar voor regressie-artefacten inhouden, als de matching geschiedt zonder randomiseren (= indeling volgens toeval). In het zojuist beschreven voorbeeld werden immers twee extreme groepen geselecteerd die zich gedragen conform hetgeen op grond van het regressiefenomeen verwacht kan worden, nl. gemiddeld genomen bij de tweede meting minder extreem scoren. Het hoeft geen betoog dat het in dergelijke situaties zeer moeilijk is, echte en schijnbare verandering te scheiden.

Een andere storende factor is *differentiële groepssamenstelling* ("selection"). Deze komt tot stand doordat twee of meer groepen van bij de aanvang af van elkaar verschillen. De geconstateerde verschillen in scores op de afhankelijke variabele kunnen niet zonder meer aan de werking van de causale factor worden toegeschreven, omdat zij mede het resultaat zijn van de wijze waarop de onderzoeksgroepen samengesteld zijn. Het optreden van "selection" kan, indien praktisch uitvoerbaar, worden voorkomen door te randomiseren, d.w.z. door middel van een toevalsprocedure te bepalen tot welke onderzoeksgroep een element behoort.

De bevindingen van een onderzoek kunnen ook vertekend worden doordat aanvankelijk geselecteerde onderzoekselementen uitvallen (*uitval*, "mortality"). Als deze uitval toevallig verdeeld is, is er uit oogpunt van interne geldigheid niet veel aan de hand. Indien echter de uitval op een of andere wijze systematisch gerelateerd is aan de afhankelijke variabele, kunnen er verschillen of overeenkomsten in meetresultaat ontstaan die niets van doen hebben met de werking van de causale factor. Uitval is een probleem voor vrijwel alle onderzoekstypen. Veel bediscussieerd is de non-response in survey-onderzoek. Het probleem stelt zich echter ook bij (quasi-)experimentele opzetten.

De hier gegeven classificatie van storende factoren is niet uitsluitend. Daardoor kunnen er ook *interacties* optreden tussen deze storende factoren. Voor een goed begrip over interactie tussen storende factoren moet eerst kort worden ingegaan op het verschijnsel interactie in het algemeen. Van interactie is er sprake als richting en/of sterkte van de samenhang tussen twee factoren verschillend is binnen de categorieën van een derde variabele. Een dergelijke interactie wordt doorgaans een eerste-orde interactie genoemd. Zijn er vier variabelen in het geding, waarbij dan geconstateerd wordt dat de eerste-orde interactie verschillend is binnen categorieën van die vierde variabele, dan is er sprake van tweede-orde interactie. Analoog hieraan zijn hogere-orde interacties te definiëren.

Het optreden van *interacties* tussen storende factoren onderling en tussen storende factoren en de causale factor kan allerlei complicaties oproepen. We illustreren dit aan de hand van enkele voorbeelden. Nemen we onderstaande onderzoeksopzet. Veronderstel dat  $F_1$  een maat is om de houding ten opzichte van feminisme weer te geven,  $X$  is een vier weken durende cursus over het feminisme en  $F_2$  is een nameting van de houdingen t.o.v. feminisme. Het is uiteraard de bedoeling van de onderzoeker om het effect van  $X$  na te gaan op het verschil tussen  $F_2$  en  $F_1$ . Als mogelijke storende factoren treden o.m. op : testing (T) en buitenexperimentele gebeurtenissen (BG). Er is bijvoorbeeld tussen de twee metingen een tv-uitzending over feminisme geweest.

Tabel 10 : Causale onderzoeksopzet zonder controlegroep

voormeting	gebeurtenis (causale factor)	nameting
$F_1$	$X$	$F_2$

T en BG zijn derhalve alternatieve verklaringen voor het verschil tussen  $F_2$  en  $F_1$ , of, in formule :  $F_2 - F_1 = X + T + BG$ . Het vaststellen van het eventueel interacteren tussen X, T en BG zou alleen mogelijk zijn als deze drie factoren ook systematisch zouden variëren zodat bv. het effect van X op de afhankelijke variabele zou kunnen worden vastgesteld voor de situatie dat er wel en geen t.v.-uitzending heeft plaatsgevonden. Maar deze systematische variatie van beide storende factoren ontbreekt in bovenstaand ontwerp. Het effect van X kan enkel worden vastgesteld voor de situatie dat er wel een voormeting (T) en wel een t.v.-uitzending (BG) plaatsvond.

Bekijken we vervolgens onderstaande opzet :

Tabel 11 : Causale onderzoeksopzet met controlegroep

voormeting	gebeurtenis (causale factor)	nameting
$F_1$	X	$F_2$
$F_3$		$F_4$

De eerste rij in bovenstaande opzet (groep 1) is volledig dezelfde als in voorgaande opzet. Bij de tweede groep wordt wel een voor- en nameting inzake feminisme-houding verricht, maar groep 2 krijgt geen cursus over feminisme (X). Voor groep 1 berekenen we het verschil tussen beide metingen als  $V_1 = F_2 - F_1$ ; voor groep 2 is  $V_2 = F_4 - F_3$ . De verandering in de experimentele groep (groep 1),  $V_1$ , komt weer tot stand door X, T en BG. De verandering in de controlegroep (groep 2),  $V_2$ , komt tot stand door T en BG. Het verschil tussen deze veranderingen,  $V (= V_2 - V_1)$ , is dus schijnbaar geheel toe te schrijven aan het al of niet gehad hebben van de cursus feminisme, althans voor zover de effecten van T en BG in beide groepen gelijk zijn, met andere woorden voor zover er geen storende interacties optreden. Interacties tussen de storende factoren T en BG vormen geen probleem : het effect in beide groepen is identiek. Maar : interacties tussen de storende factoren en X kunnen wel problematisch zijn.

#### 2.2.2.3.2.1.2. Generaliseerbaarheid

De tot nu toe behandelde ontwerpvereisten hebben alle tot doel na te gaan of de in een bepaald onderzoek aanwezige variatie in de causale factor oorzaak is van de gevonden variatie in de afhankelijke variabele. Maar in het algemeen zal de onderzoeker niet tevreden zijn met de constatering dat er bij de feitelijk onderzochte elementen op dat moment sprake was van een causale invloed van X op Y. De onderzoeker tracht meer algemenere uitspraken te doen (externe geldigheid). Dergelijke generalisaties leveren op tal van punten problemen op. Enkele ervan komen in hetgeen volgt aan de orde, nl.

- onderzoekselementen en onderzoekspopulatie ;
- tijd en plaats ;
- operationalisering van de onderzoeksvariabelen en
- uitvoeringswijze van het onderzoek.



*[1] Onderzoekselementen en onderzoekspopulatie*

Bij externe geldigheid is allereerst de zeer belangrijke vraag aan de orde in hoeverre de feitelijk onderzochte onderzoekselementen als representatief kunnen worden beschouwd voor een grotere populatie, zodat de onderzoeksresultaten geldig mogen worden verklaard voor die grotere populatie. Het statistisch antwoord op deze vraag is het aselekt trekken van de onderzoekselementen uit een grotere populatie. Dit statistisch antwoord is, hoe belangrijk ook, echter niet in alle opzichten voldoende. Zo is er het probleem dat diverse, aanvankelijk geselecteerde onderzoekselementen in een latere onderzoeksfase kunnen uitvallen. Behalve een gevaar voor de interne geldigheid van het onderzoek vormt deze uitval ook een bedreiging voor de externe geldigheid : de onderzoeksresultaten kunnen alleen gegeneraliseerd worden naar die personen die mee willen werken met een onderzoek, die er de tijd voor hebben, etc.

Fundamenteler is nog het probleem dat op basis van een aselekte steekproef alleen gegeneraliseerd mag worden naar de populatie waaruit deze steekproef afkomstig is. De steekproef is immers niet altijd identiek met de theoretisch bedoelde populatie.

*[2] Tijd en plaats*

Resultaten van onderzoek dat op een bepaald moment op een bepaalde plaats uitgevoerd is, kunnen niet zonder meer van toepassing worden verklaard voor andere tijden en plaatsen. De generaliseerbaarheid van onderzoeksbevindingen over de tijd kan problemen opleveren omdat buitenexperimentele gebeurtenissen mogelijk hebben geleid tot een unieke constellatie van onderzoeksbevindingen. Analoge vragen kunnen gesteld worden rond de generaliseerbaarheid naar plaats. Enige zekerheid omtrent de mate van generaliseerbaarheid naar tijd en plaats, kan behalve door een nauwkeurige bestudering van de constellatie van relevante factoren die op dat moment en die plaats aanwezig waren, verkregen worden door het onderzoek op andere tijdstippen en plaatsen te herhalen. Daarbij komen evenwel vragen kijken betreffende de praktische toepasbaarheid.

*[3] Operationalisering van de onderzoeksvariabelen*

Het gaat hier om het probleem van zgn. "construct validity" (begripsgeldigheid) : de correspondentie tussen theoretische begrippen en operationele definities (speelt eveneens een rol bij interne geldigheid).

*[4] Uitvoeringswijze van het onderzoek*

Aan de hele wijze waarop het onderzoek in de praktijk opgezet en uitgevoerd wordt, zijn talloze facetten te onderscheiden die van invloed kunnen zijn op de resultaten. Zaken als testvaardigheid en neiging tot consistent scoren kunnen de "construct validity" aantasten. Deze begripsgeldigheid wordt bv. ook nog aangetast als onderzoekspersonen zich bij hun gedrag laten leiden door wat in hun ogen de interviewer of proefleider van hen verwacht. In het algemeen kan dus de onderzoekssetting zodanig zijn dat niet de theoretisch bedoelde variabele wordt gemeten maar 'iets anders'. De uitvoeringswijze van het onderzoek kan ook als storende factor optreden en als zodanig de interne geldigheid bedreigen. Kenmerken en optreden van interviewers en proefleiders kunnen rechtstreeks van invloed zijn op het gedrag van de onderzoekspersonen ; voormetingen kunnen respondenten aan het denken zetten ; proefpersonen kunnen extra hun best doen omdat ze onderzocht worden. In al deze gevallen hoeft niet de begripsgeldigheid in het gedrang te komen : de metingen zijn

wel valide, maar datgene wat gemeten wordt, verandert zelf onder invloed van het onderzoeksgebeuren. Daardoor kunnen er verschillen ontstaan tussen bv. voor- en nameting, die ten onrechte toegeschreven worden aan de werking van de causale factor. Tenslotte kan de uitvoeringswijze van een onderzoek ook, en daar gaat het uiteindelijk om, de externe geldigheid aantasten. Om een voorbeeld te geven : zijn de resultaten van een laboratorium-experiment te generaliseren naar zgn. "real life"-situaties ? Op dergelijke vragen kan een antwoord worden gezocht door de onderzoeksinstelling daadwerkelijk te laten variëren, door het onderzoek te herhalen in verschillende settings. Bovendien wordt de generaliseerbaarheid naar met name "real life"-situaties sterk bevorderd door het gebruik van "unobtrusive measures", d.w.z. metingen en onderzoekssituaties die de onderzoekspersonen min of meer natuurlijk voorkomen of die door hen zelfs helemaal niet als "onderzoek" ervaren worden.

#### 2.2.2.3.2.1.3. Theoretische adequaatheid en praktische uitvoerbaarheid

Theoretische adequaatheid is in feite een alles omvattende vereiste waaraan een onderzoeksontwerp moet voldoen. De onderzoeksopzet moet, vanuit dit gezichtspunt, dusdanig zijn dat het onderzoek uitsluitend geeft over de adequaatheid van de theoretische gedachtegang waarmee de onderzoeker gestart is. Het onderzoeksontwerp moet aansluiten bij het conceptueel model. Deze vereiste wordt hier vooral naar voren gebracht om expliciet te benadrukken dat onderzoek niet gericht is op het verzamelen van een aantal losstaande feiten, maar dat de functie ervan is het beantwoorden van een theoretische vraagstelling. In die zin vormen interne en externe geldigheid onderdelen van theoretische adequaatheid. In de gehele wijze waarop een onderzoek gestalte krijgt, vormen theoretische vragen het uitgangspunt. Daarnaast zullen ook altijd overwegingen van praktische uitvoerbaarheid een grote rol spelen. Uitvoerbaarheid kan men omschrijven als de beschikbaarheid van middelen om het voorgenomen onderzoek te realiseren.

#### 2.2.2.3.2.2. Het waarnemen in een ongecontroleerd experiment

We spreken over een ongecontroleerd experiment (of quasi-experiment) wanneer we slechts *de gevolgen van een gebeurtenis* meten. We meten alleen en grijpen dus zelf niet in. Het doel van het uitvoeren van een ongecontroleerd experiment is het opsporen van vermeende verbanden in de verzamelde gegevens. We willen het effect van de onafhankelijke variabele(n) op de afhankelijke variabele vaststellen.

Voorbeelden van het observeren van materiële zaken in een ongecontroleerd experiment :

\* door mensen :

- retail-audit van Nielsen ;
- consumer-audit van Attwood ;
- "pantry check" ;

\* door instrumenten :

- scanning ;
- verkeersteller ;
- audimeter.

Bij het ongecontroleerde experiment worden drie klassen van proefopzetten onderscheiden, nl.

- tijdserie-opzetten ;
- dwarsdoorsnede-opzetten ;
- opzetten die een combinatie inhouden van tijdserie- en dwarsdoorsnede-opzetten.

## 2.2.2.3.2.2.1. Tijdserie-opzetten

Hieronder verstaan we het verzamelen van gegevens over een (bij voorkeur vrij groot) aantal perioden. De metingen vinden dus plaats op verschillende achtereenvolgende tijdstippen (of perioden). Het kan hierbij gaan om het waarnemen van een verschijnsel of het waarnemen van gegevens met het doel de samenhang tussen verschijnselen, gebeurtenissen of variabelen vast te stellen.

## 2.2.2.3.2.2.1.1. "After only-design"

Er wordt gemeten nadat een bepaalde gebeurtenis heeft plaatsgevonden. We kunnen evenwel in het geheel niet nagaan wat het effect geweest is van de gebeurtenis.

Tabel 12 : After only-design

gebeurtenis	nameting
X	O

De algemene diagnose van bovenstaande opzet (ook "one shot case study" genoemd) is immers dat er geen enkele uitspraak over de effecten van de onafhankelijke variabele (X) op de afhankelijke variabele (O) gedaan kan worden, omdat binnen dit ontwerp *geen samenhang* kan worden vastgesteld. Daartoe dienen minimaal twee situaties met elkaar vergeleken te worden en hier wordt slechts 1 situatie bestudeerd.

Het ontwerp in tabel 13, ook "static group comparison" genoemd, voldoet aan de basisvoorwaarde voor het opsporen van causaliteit : het kunnen vaststellen van samenhang. De te bestuderen "case" wordt expliciet uitgebreid met (minstens) 1 vergelijkingsgroep, met een verschillende waarde voor de onafhankelijke variabele. Het verschil in waargenomen gevolg ( $O_1 - O_2$ ) is een maatstaf voor het effect van variabele 'X'. Wat het tweede criterium (de richting van de causale beïnvloeding) betreft, lijken er geen problemen te zijn. Het verschil tussen  $X_1$  en  $X_2$  komt in de tijd vóór het verschil ( $O_1 - O_2$ ), zodat ( $O_1 - O_2$ ) geen oorzaak kan zijn van het verschil tussen  $X_1$  en  $X_2$ . Nochtans kunnen er indirecte problemen met de richting van de causale beïnvloeding zijn, ook wanneer het verschil tussen  $X_1$  en  $X_2$  in de tijd vastgesteld is vóór het verschil tussen  $O_1$  en  $O_2$ . Het verschil tussen  $O_1$  en  $O_2$  kan immers reeds aanwezig zijn vóór de vaststelling van het verschil tussen  $X_1$  en  $X_2$  en kan juist de oorzaak zijn van het verschil tussen  $X_1$  en  $X_2$  (en van het latere verschil tussen  $O_1$  en  $O_2$ ).

Tabel 13 : Static group comparison

gebeurtenis	nameting
$X_1$	$O_1$
$X_2$	$O_2$

Aangaande de evaluatie van bovenstaande opzet naar interne geldigheid toe, zijn de belangrijkste storende factoren die bij deze opzet optreden in het algemeen differentiële groepssamenstelling (DG) en uitval (U) :  $(O_1 - O_2) = (X_1 - X_2) + DG + U$ . Differentiële groepssamenstelling wordt een probleem als de onderzoekselementen niet via een toevalsprocedure aan de diverse onderzoeksgroepen worden toegewezen. Uitval wordt minder problematisch voor zover deze at random plaatsvindt.

Buitenexperimentele gebeurtenissen zullen in het algemeen geen grote bedreiging voor de interne geldigheid van de onderzoeksopzet vormen. Omdat de onderzoeksperiode voor de beide groepen gelijk is, zullen de onderzoeksgroepen door dezelfde gebeurtenissen beïnvloed worden, zodat verschillen tussen de onderzoeksgroepen wat betreft de afhankelijke variabele niet uit die gebeurtenissen kunnen worden verklaard :  $(O_1 - O_2) = (X_1 - X_2) + (BG - BG) = (X_1 - X_2)$ . Er wordt evenwel verondersteld dat er geen groepsgebonden gebeurtenissen zijn. In dit laatste geval zouden eventueel interacties kunnen optreden tussen groepsgebonden factoren en de causale factor en/of tussen groepsgebonden gebeurtenissen en differentiële groepssamenstellingsfactoren.

Voor de factor 'spontane veranderingen' gelden dezelfde opmerkingen als voor de 'gemeenschappelijke' gebeurtenissen. Zij zijn onder controle tenzij er interactie-effecten zijn met de causale factor en/of de groepssamenstellingsfactoren.

De factor 'meten' treedt niet op omdat er geen sprake is van voormetingen. Het optreden van de factor 'regressie' is afhankelijk van de selectiewijze van de onderzoeksgroepen. Hoewel het weinig waarschijnlijk is dat beide groepen op basis van extreme scores op de afhankelijke variabele zijn samengesteld, is het wel mogelijk dat op indirecte wijze toch extreme groepen geselecteerd zijn. Voor zover de beide onderzoeksgroepen op gelijke wijze extreem zijn, zal de te verwachten regressie in beide groepen gelijk zijn en is zij daarmee dus onder controle. Maar indien de groepen in dit opzicht ongelijk van samenstelling zijn, kan regressie storend werken.

Over de externe geldigheid kan weinig gezegd worden. Omdat onderzoek met een dergelijke opzet zich nogal eens in de 'natuurlijke' omgeving van de respondent zal afspelen, zal de generaliseerbaarheid in zijn algemeenheid misschien niet al te problematisch zijn. Wellicht is de generaliseerbaarheid van de operationalisering van de onderzoeksvariabelen nog het zwakste punt. De onderzoeker heeft in de natuurlijke omgeving misschien weinig invloed op de wijze waarop hij een variabele zoals bv. aard van de leiding kan operationaliseren. De onderzoeker is dan afhankelijk van de manier waarop 'autoritair' en 'democratisch' leiderschap bv. gestalte hebben gekregen in diverse bedrijven. De praktische uitvoerbaarheid van een dergelijke opzet zal relatief groot zijn, terwijl tenslotte over de theoretische adequaatheid in zijn algemeenheid eigenlijk weinig zinnigs kan gezegd worden.

#### 2.2.2.3.2.2.1.2. "Before and after-design"

Er wordt gemeten zowel voor als na de gebeurtenis. Het effect van de gebeurtenis wordt gekwantificeerd als  $O_2 - O_1$ . In deze opzet (ook "one group pretest-posttest design" genoemd) wordt de situatie dus eerst opgenomen voordat de experimentele stimulus wordt toegediend ( $O_1$ ), vervolgens wordt de groep aan de bewuste stimulus (X) blootgesteld,

waarna de situatie van de groep opnieuw wordt gemeten ( $O_2$ ). Zoals wordt weergegeven, lijkt de variatie in de afhankelijke variabele ( $O_2 - O_1$ ) veroorzaakt te

Tabel 14 : Before and after design

Voormeting (t-1)	gebeurtenis	nameting (t+1)
$O_1$	X	$O_2$

worden door slechts één categorie van de causale factor (X) in plaats van door variatie in de causale factor. Dit is evenwel slechts schijn, omdat het ontwerp in feite te beknopt is weergegeven. Een consequentere voorstelling van zaken voor deze opzet is als volgt samen te vatten :  $X_1 \Rightarrow O_1 \Rightarrow X_2 \Rightarrow O_2$ .

Voor bovenstaande opzet geldt dat  $(O_2 - O_1) = (X_2 - X_1)$ . Deze diachronische opzet voorziet in de mogelijkheid geldige uitspraken te doen over causale relaties. Uit oogpunt van interne geldigheid is het zwakste punt van dit onderzoeksontwerp de factor buitenexperimentele gebeurtenissen. Er kunnen zich tussen  $O_2$  en  $O_1$  talloze gebeurtenissen voordoen die, buiten de causale factor om, van invloed zijn op de afhankelijke variabele, zodat we kunnen stellen :  $(O_2 - O_1) = (X_2 - X_1) + BG$ .

#### 2.2.2.3.2.2.1.3. "Before and after with control group-design"

Dit type onderzoeksontwerp (ook aangeduid als "nonequivalent control group design") is eveneens een diachronisch onderzoeksontwerp. In vergelijking met vorig ontwerp is hier een controlegroep aan het ontwerp toegevoegd. Deze controlegroep verschilt van de andere onderzochte (experimentele) groep wat betreft de waarde van de causale factor X, maar zou op vergelijkbare wijze onderhevig moeten zijn aan storende factoren. De wijze waarop het effect van de causale factor vastgesteld wordt, is hier iets ingewikkelder. Nu worden verschillen in waarneming in de ene groep vergeleken met verschillen in de andere groep :  $(O_2 - O_1) - (O_4 - O_3) = X + S$  (storende factoren). Door van de verschillen in de experimentele groep de verschillen in de vergelijkingsgroep af te trekken, probeert men een zuivere schatting te krijgen van het effect van de causale factor. Dit lukt slechts voor zover de (effecten van de) storende factoren in beide groepen gelijk zijn.

Tabel 15 : Before and after with control group-design

voormeting (t-1)	gebeurtenis	nameting (t+1)
$O_1$	X	$O_2$
$O_3$		$O_4$

#### 2.2.2.3.2.2.2. Dwarsdoorsnede-opzetten

Hierbij verrichten we waarnemingen bij een vrij groot aantal proefeenheden op één tijdstip. Verschillende groepen worden geconfronteerd met verschillende gebeurtenissen en de resultaten worden onderling vergeleken. Ook hier moet de opmerking gemaakt worden dat geen rekening wordt gehouden met storende factoren.

Tabel 16 : Dwarsdoorsnede-opzet

gebeurtenis	meting
$X_1$	$O_1$
$X_2$	$O_2$
$X_3$	$O_3$
.....	.....
$X_n$	$O_n$

#### 2.2.2.3.2.2.3. Combinatie van tijdserie-opzet en dwarsdoorsnede-opzet

We lichten deze opzet toe aan de hand van een voorbeeld van een consumentenpanel. Dit bestaat uit een aantal huishoudens, die gedurende een lange periode wekelijks de aankopen van bepaalde levensmiddelen noteren. Veronderstellen we dat we geïnteresseerd zijn in het effect van prijswijzigingen van melk op het gebruik. Het effect van de eerste prijsverandering kwantificeren we als  $(O_2 - O_1) - (O_4 - O_3)$ . Het effect van de tweede prijsverandering kwantificeren we als  $(O_5 - O_2) - (O_6 - O_4)$ , enz.

Tabel 17 : Combinatie tijdserie-opzet en dwarsdoorsnede-opzet

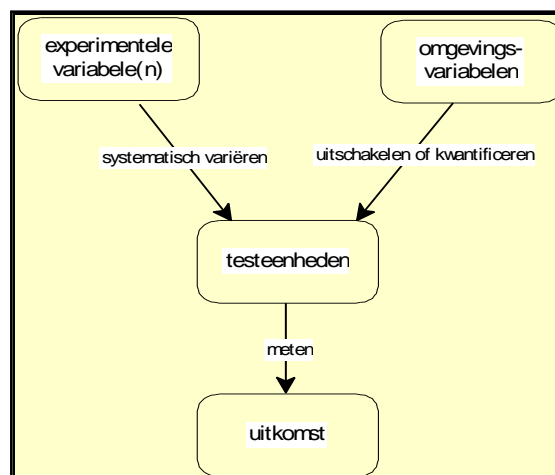
experimentele groep	$O_1$	$X_1$	$O_2$	$X_2$	$O_5$	$X_3$	$O_7$	.....	$O_m$
Controlegroep	$O_3$		$O_4$		$O_6$		$O_8$	.....	$O_{m+1}$

#### 2.2.2.3.2.3. Het waarnemen in een gecontroleerd experiment

De nadruk bij de hierboven geschetste proefopzetten ligt op het vinden van samenhangen tussen variabelen, zonder de pretentie dat er sprake is van een causaal verband. Hiervoor is het uitvoeren van een gecontroleerd experiment noodzakelijk.

Bij gecontroleerde experimenten gaat de 'interventie' van de onderzoeker verder dan alleen het meten van de gevolgen van een gebeurtenis. Er wordt ten minste één variabele gemanipuleerd om na te gaan wat het effect is van de experimentele variabele(n) op de afhankelijke variabele. Het doel van een gecontroleerd experiment is een zo zuiver mogelijk resultaat te krijgen. In voorgaand punt werd benadrukt dat er storende invloeden kunnen optreden bij een experiment. Bij het gecontroleerde experiment zijn we in staat expliciet rekening te houden met storende invloeden, hetgeen in wezen het belangrijkste onderscheid

is tussen een gecontroleerd en een ongecontroleerd experiment. Er wordt geprobeerd de storende invloeden te kwantificeren. Deze storende invloeden noemen we omgevingsvariabelen, d.w.z. variabelen die het resultaat van een experiment kunnen verstoren, c.q. beïnvloeden. De gang van zaken bij een gecontroleerd experiment wordt samengevat in figuur 10.



Figuur 10 : Het gecontroleerde experiment

Er zijn vijf mogelijkheden om de omgevingsvariabelen onder controle te houden :

- op voorhand uitschakelen van de invloed ervan ;
- randomisering toepassen ;
- stratificatie of blokvorming toepassen ;
- matching toepassen ;
- covariantie-analyse uitvoeren.

[1] *Het op voorhand uitschakelen van de invloed van omgevingsvariabelen*

Dit is de eenvoudigste oplossing, die echter lang niet altijd uitvoerbaar is. Veronderstel dat we een groep respondenten willen testen op hun ontvankelijkheid voor een reclamespot. Indien bijvoorbeeld het intelligentieniveau of de kerkelijke gezindte van de respondenten mogelijkwijs het experiment beïnvloeden, dan kan het experiment beperkt worden tot respondenten met een ongeveer gelijk intelligentiepeil of eenzelfde kerkelijke gezindte. We zoeken derhalve testeenheden die ten aanzien van de omgevingsvariabele(n) zo homogeen mogelijk zijn. Een probleem hiermee is evenwel dat de resultaten van het experiment niet zonder meer veralgemeend kunnen worden.

[2] *Het toepassen van randomisering*

Hieronder verstaan we het op volledig toevallige basis toewijzen van testeenheden aan de experimentele groep en aan de controle- groep. Dit is theoretisch de beste manier om alle invloeden van buiten onder controle te houden. We veronderstellen namelijk dat de

experimentele groep en de controlegroep door het toepassen van randomisering homogeen zijn ten aanzien van de invloeden van buiten (d.w.z. omgevingsinvloeden hebben op beide groepen een gelijke invloed).

*[3] Het toepassen van stratificatie of blokvorming*

Dit betekent dat we een invloed van buiten expliciet in het experiment betrekken door deze invloed op te nemen als een onafhankelijke variabele. Een onderzoeker kan immers geïnteresseerd zijn in de invloed van een omgevingsvariabele (bv. opleidingsniveau, geslacht, e.a.). In tegenstelling tot de hierboven vermelde methoden wordt de invloed van omgevingsvariabelen niet uitgeschakeld maar gekwantificeerd. Door toepassing van stratificatie of blokvorming kan de invloed van experimentele variabelen en omgevingsvariabelen gekwantificeerd worden (o.m. met behulp van variantie-analyse).

*[4] Het toepassen van matching*

Ook door deze methode wordt de invloed van omgevingsvariabelen in het experiment betrokken. De wijze waarop dit gebeurt verschilt echter. Veronderstel dat we willen matchen op de variabele leeftijd. Daartoe maken we bijvoorbeeld twee leef- tijdsklassen : 20 tot 40 jaar en 40 tot 60 jaar. Vervolgens nemen we twee respondenten van dezelfde leeftijdsgroep die we aselekt toewijzen aan de experimentele groep en de controlegroep. Door deze werkwijze zijn we er zeker van dat de leeftijdsgroepen gelijk verdeeld zijn over de experimentele en de controlegroep. We schakelen de invloed van de variabele leeftijd dus uit. Als we op een variabele matchen dan moeten we er echter wel zeker van zijn dat deze variabele het experiment zou verstoren indien we niet zouden matchen. Anders heeft het matchen uiteraard geen zin. Er dient ook op toegezien dat niet op te veel variabelen of op te veel niveaus van variabelen wordt gematcht. Anders houden we te weinig waarnemingen per combinatie van variabelen over.

*[5] Het toepassen van covariantie-analyse*

Deze methode is in staat om de invloeden van buiten die we niet onder controle hebben (niet-beheersbare invloeden) uit te sluiten. Met betrekking tot de verschillende proefopzetten bij het gecontroleerde experiment, wordt in de eerste plaats een onderscheid gemaakt tussen proefopzetten waarbij 1 of meerdere experimentele variabelen worden gehanteerd. Daarnaast wordt een onderscheid gemaakt tussen proefopzetten met 1 of meerdere *behandelingen*. Onder behandelingen verstaan we de verschillende manieren waarop de experimentele variabele(n) word(en)t gehanteerd. Veronderstel dat we willen onderzoeken of verschillende soorten verpakkingen van een artikel invloed hebben op de verkopen in bijvoorbeeld supermarkten. In de supermarkten van de controlegroep wordt de oude verpakking behouden, terwijl die van de experimentele groep een nieuwe verpakking gebruiken. Er is hier sprake van 1 behandeling (1 niveau van de experimentele variabele). Indien we meerdere nieuwe verpakkingen wensen te testen, dan hebben we te maken met meerdere behandelingen en daarom ook met meerdere experimentele groepen.

Zoals bij het ongecontroleerd experiment kan alleen na de gebeurtenis of zowel voor als na de gebeurtenis gemeten worden. We mogen ervan uitgaan dat de invloeden van buiten door randomisering onder controle worden gehouden en dat dus alle testeenheden voor het experiment homogeen zijn ten aanzien van alle invloeden waarvan kan worden verondersteld dat ze het testresultaat kunnen beïnvloeden. Op grond van deze



veronderstelling is vooraf meten niet nodig. Wordt evenwel verondersteld dat externe invloeden het experiment beïnvloeden ondanks randomisering, dan kan dit worden opgelost door stratificatie of matching toe te passen.

#### 2.2.2.3.2.3.1. Volstrekt aselechte opzet

Bij deze proefopzet kan het effect van 1 experimentele variabele op de afhankelijke variabele worden bepaald. Daartoe worden de behandelingen volledig aselechte (random) toegewezen aan de testeenheden. We gaan er bij deze proefopzet derhalve van uit dat het effect van externe invloeden op de testeenheden relatief gelijk is.

Stel dat we een onderzoek willen doen naar de invloed van de verpakking van sinaasappelen op de verkoop ervan. We betrekken 15 winkels (testeenheden) in het experiment en wijzen deze volledig aselechte toe aan de drie niveaus (behandelingen) van de experimentele variabele (luze verpakking, verpakking in plastic en losse verkoop). Vermits er drie niveaus zijn van de experimentele variabele worden de winkels in drie groepen verdeeld. Schematisch ziet deze volledig aselechte opzet (regelrechte meting) er uit zoals voorgesteld in tabel 18.

Tabel 18 : Volstrekt aselechte opzet

behandeling 1	behandeling 2	behandeling 3
$O_{11}$	$O_{12}$	$O_{13}$
$O_{21}$	$O_{22}$	$O_{23}$
$O_{31}$	$O_{32}$	$O_{33}$
$O_{41}$	$O_{42}$	$O_{43}$
$O_{51}$	$O_{52}$	$O_{53}$

Indien te verwachten valt dat de grootte van de winkels invloed kan uitoefenen op de resultaten van het experiment, dan kan deze proefopzet op dezelfde manier worden uitgevoerd door matching in plaats van randomisering te gebruiken. Wanneer we bijvoorbeeld een onderscheid maken tussen grote, middelgrote en kleine winkels, dan kunnen we grote winkels aselechte toewijzen aan de drie niveaus van de experimentele variabele, enz. Op die manier wordt het effect van de winkelgrootte uitgeschakeld. Wanneer we evenwel de invloed van de winkelgrootte wensen te kwantificeren dan dient gebruik gemaakt te worden van de aselechte blokkenopzet.

#### 2.2.2.3.2.3.2. Aselechte blokkenopzet

Hierbij gaan we ervan uit dat er sprake is van 1 experimentele variabele en van 1 invloed van buiten uit. Deze laatste wordt gekwantificeerd. Deze externe variabele noemen we een blokvariabele. De testeenheden worden gegroepeerd in blokken : de testeenheden binnen elk van de blokken worden verondersteld homogeen te zijn ten aanzien van de externe variabele. Vervolgens worden de testeenheden op aselechte wijze toegewezen aan de behandelingen van de experimentele variabele. In het geval van de aselechte blokkenopzet is

er derhalve sprake van twee invloeden op de afhankelijke variabele. Schematisch kan het zojuist besproken voorbeeld als volgt worden weergegeven :

Tabel 19 : Aselecte blokkenopzet

	Behandeling 1	behandeling 2	behandeling 3
blok 1	O <sub>11</sub>	O <sub>12</sub>	O <sub>13</sub>
blok 2	O <sub>21</sub>	O <sub>22</sub>	O <sub>23</sub>
blok 3	O <sub>31</sub>	O <sub>32</sub>	O <sub>33</sub>

Wanneer we veronderstellen dat er nog een tweede invloed van buiten meespeelt in het experiment, komen we bij de volgende opzet, nl. het Latijns vierkant.

### 2.2.2.3.2.3.3. Latijns vierkant

Bij deze proefopzet onderscheiden we twee varianten. De eerste is een uitbreiding van de vorige opzet, nl. het in de analyse betrekken van een tweede blokvariabele. De testeenheden worden nu op basis van twee omgevingsvariabelen in homogene groepen (blokken) verdeeld, terwijl we nog steeds experimenteren met slechts 1 variabele. In het gekozen voorbeeld zouden we bijvoorbeeld de dag van de week als tweede blokvariabele kunnen opnemen. We veronderstellen dan dat de dag van de week (als variabele) invloed heeft op de verkoop. Omdat we van een Latijns vierkant spreken, moeten we ervoor zorgen dat er evenveel behandelingen als blokken (per blokvariabele) zijn. Elke behandeling komt dan 1 keer voor in elke rij en elke kolom van de matrix.

Tabel 20 : Latijns vierkant

		blokvariabele 1 (bv. winkelgrootte)		
		1	2	3
blokvariabele 2	1	(A) O <sub>11</sub>	(B) O <sub>12</sub>	(C) O <sub>13</sub>
(bv. dagen v.d.	2	(B) O <sub>21</sub>	(C) O <sub>22</sub>	(A) O <sub>23</sub>
week)	3	(C) O <sub>31</sub>	(A) O <sub>32</sub>	(B) O <sub>33</sub>

A, B, C : behandelingen

Een speciaal geval van het Latijns vierkant is het schema met dubbele wisselingen. In een Latijns vierkant kan het voorkomen dat de effecten van opeenvolgende behandelingen niet onafhankelijk van elkaar zijn. Er kan een overdrachtseffect optreden. Doordat behandeling A bij een kleine winkel heeft plaatsgevonden op maandag/dinsdag kan dit gevolgen hebben voor de respons (verkoop) van behandeling B bij winkels van dezelfde grootte op woensdag/donderdag. Hieruit kunnen we afleiden dat we de matrix kolomsgewijs moeten lezen. Om het overdrachtseffect uit te schakelen passen we het Latijns vierkant vaker toe en wel zo dat alle mogelijke volgorden van behandelingen worden opgenomen. Schematisch krijgen we dan onderstaande opzet :

Tabel 21 : Schema met dubbele wisselingen

		blokvariabele 1					
		1	2	3	1	2	3
blokvariabele 2	1	A	B	C	A	B	C
	2	B	C	A	C	A	B
	3	C	A	B	B	C	A

A, B, C : behandelingen

De tweede versie van het Latijns vierkant gaat uit van meer dan 1 experimentele variabele. Een veronderstelling hierbij is wel dat de experimentele variabelen onafhankelijk van elkaar zijn. Indien we deze veronderstelling van onafhankelijkheid laten vallen, komen we bij de factoropzet.

#### 2.2.2.3.2.3.4. Factoropzet

Bij deze proefopzet is het mogelijk het effect van twee of meerdere experimentele variabelen te testen. Ook de interactie tussen de experimentele variabelen wordt hierbij in beschouwing genomen. Onder interactie verstaan we : het effect van een verandering in 1 experimentele variabele op het resultaat is afhankelijk van het niveau van 1 of meerdere andere variabelen in het experiment. In het eerder gekozen voorbeeld zouden we naast de verpakking ook dagbladreclame (bv. een kleine versus een grote advertentie) als een tweede experimentele variabele kunnen opnemen. In dit geval krijgen we 6 behandelingen (3 van de experimentele variabele 'verpakking' en 2 van de experimentele variabele 'dagbladreclame'). Schematisch ziet de factoropzet er als volgt uit :

Tabel 22 : Factoropzet

		1 <sup>ste</sup> experimentele variabele		
		1	2	3
2 <sup>e</sup> experimentele variabele	1	(A) O11	(B) O12	(C) O13
	2	(D) O21	(E) O22	(F) O23

A, B, C, D, E, F : behandelingen

We onderscheiden zes behandelingen, waarbij per behandeling een aantal waarnemingen worden verricht. Het effect van de twee experimentele variabelen, evenals het interactie-effect tussen beide, kan worden vastgesteld door toepassing van variantie-analyse (bij proefopzetten waarbij meerdere experimentele variabelen worden onderscheiden, worden meestal geen blokvariabelen in de analyse betrokken).

Ook bij het gecontroleerde experiment treffen we experimenten in de werkelijkheid (veldexperimenten) en laboratoriumexperimenten aan. Voorbeelden van gecontroleerde experimenten zijn de volgende :

\* in de werkelijkheid :

- winkeltest ;
- testmarkt ;

\* in laboratoriumsituatie :

- producttest (bv. "paired comparison"-test) ;
- meting van de attentiewaarde (bv. Atpak-meetprocedure).

Tenslotte willen we het ongecontroleerde en het gecontroleerde experiment in de werkelijkheid met elkaar vergelijken. Het belangrijkste verschil is het feit dat we bij een ongecontroleerd experiment alleen meten bij de in het experiment betrokken testeenheden, terwijl we bij het gecontroleerde experiment bewust ingrijpen. Dit ingrijpen heeft betrekking op het manipuleren van ten minste 1 variabele en het aselekt (of via gelijkschakeling) toewijzen van testeenheden aan de experimentele groep(en) en controlegroep(en) (of in het algemeen aan behandelingen). Het gevolg van het ingrijpen is dat we, in tegenstelling tot wat bij het ongecontroleerde experiment het geval is, in staat zijn de omgevingsinvloeden onder controle te houden (c.q. te kwantificeren). Dit leidt ertoe dat de uitspraken gebaseerd op een gecontroleerd experiment de voorkeur genieten op die verkregen via een ongecontroleerd experiment. Beide experimenten zijn erop gericht samenhangen tussen variabelen aan te tonen. Bij een goed uitgevoerd gecontroleerd experiment (waarbij we ervoor zorgen dat de invloeden van buiten zijn uitgeschakeld of gekwantificeerd) is de garantie dat een causaal (oorzakelijk) verband aanwezig is, groter dan bij een ongecontroleerd experiment. We kunnen namelijk constateren wat het effect van een onafhankelijke (experimentele) variabele op de afhankelijke variabele is.

Tabel 23 : Een vergelijking van proefopzetten

proefopzetten	aantal experimentele variabelen	aantal blokvariabelen
volledig aselekt opzet	1	0
aselekt blokkenopzet	1	0
Latijns vierkant (1)	1	2
Latijns vierkant (2)	2 of meer (zonder interactie)	meestal geen
Factoropzet	2 of meer (met interactie)	meestal geen

## 2.3. Het bepalen van de analysemethode

### 2.3.1. Voorbewerking van gegevens

Nadat door middel van observatie, interview, enquête of experiment of door het eenvoudig verzamelen van het nodige materiaal van achter het bureau de gegevens zijn verzameld, is het materiaal in ruwe vorm aanwezig. Het gaat er nu om, een objectieve en goedkope manier te vinden om uit dit materiaal de probleemstelling(en) op te lossen. De probleemstelling(en) is (zijn) immers de exacte formulering van datgene waarop het onderzoek een antwoord dient te formuleren.

In de eerste plaats volgt een fase van technische verwerking : de gegevens dienen gecodeerd te worden waarna ze vervolgens in een bepaald formaat moeten binnengebracht worden in de computer of een opslagmedium (bv. diskette) zodat statistische analyses mogelijk zijn op het gegevensbestand middels daartoe speciaal ontworpen toepassingssoftware (bv. SAS, SPSS).

Alvorens de verzamelde data te analyseren, is het in veel gevallen noodzakelijk enkele voorbewerkingen te verrichten, te weten het controleren ("editing") en coderen ("coding") van de gegevens.

#### 2.3.1.1. Gegevenscontrole

Eerder werd erop gewezen dat zich bij het verzamelen van data fouten kunnen voordoen. We kunnen twee soorten fouten onderscheiden :

- ◆ *experimentele (toevallige) fouten* : de factoren die deze fouten veroorzaken, zijn niet te achterhalen omdat ze toevallig van aard zijn (deze fouten zijn te achterhalen door meer steekproeven te trekken uit dezelfde populatie en te kijken naar de verschillen tussen gemiddelden ; zijn deze verschillen erg groot, dan wijst dit op een zgn. systematische fout ; experimentele fouten kunnen verkleind worden door het vergroten van de steekproef : ze zijn dan het gevolg van het niet betrouwbaar zijn van de gegevens) ;
- ◆ *systematische fouten* : deze fouten worden veroorzaakt door fouten in het waarnemen en het interpreteren van gegevens, of door een verkeerde manier van steekproeftrekking (deze komen voort uit een te geringe validiteit van de data).

Ook de respondenten zelf veroorzaken fouten. Men kan deze fouten indelen in "response errors" en "non-response errors".

- ◆ *response errors* : deze fouten zijn het gevolg van semantische problemen die zich voordoen wanneer de betekenis van de vraag door de respondent anders begrepen wordt dan de bedoeling is (hetgeen natuurlijk zou moeten vermeden worden door de vragenlijst voldoende uit te testen) terwijl het anderzijds ook kan gaan om vrijwillige fouten ;
- ◆ *non-response errors* : weigeringen en afwezigheden kunnen eveneens leiden tot vertekeningen.

Het "editeren" bestaat er in de meeste gevallen in dat iedere vragenlijst wordt nageplozen op fouten (bv. opsporen van response errors).

### 2.3.1.2. Codering

Coderen bestaat uit het toekennen van een cijferwaarde aan iedere antwoordcategorie zodat de gegevens door de computer kunnen verwerkt worden. Door de projectleider wordt een codeboek opgesteld. Het coderen schept vooral problemen bij open vragen. Op basis van de codering bekomen we rechte tellingen (datamatrix).

Het eindresultaat van de codering is een *datamatrix*, een rechthoek waarvan de rijen de waarnemingseenheden en de kolommen de variabelen en de celvullingen de scores, codes van de eenheden op die variabelen zijn. Bij een datamatrix dient men zich goed te realiseren dat per kolom (dus per variabele) de codes een geheel ander onderling relatiepatroon kunnen hebben : twee verschillende codecijfers in kolom X kunnen wijzen op een gewoon verschil (nominaal verschil) ; soms gaat het om ordinale verschillen en in nog andere gevallen gaat het om interval- of ratio-verschillen.

Nu spreekt het vanzelf dat een goede codering (de verzameling symbolen, meestal getallen, die gebruikt worden om de antwoordcategorieën van een variabele af te beelden) aan twee eisen moet voldoen, nl. de codering moet én uitsluitend én uitputtend zijn. Het idee datamatrix is vooral belangrijk omdat men zich aan de hand van de datamatrix de verschillende analyseprocedures goed kan voorstellen. We zijn nu nl. toe aan de fase dat de datamatrix wordt geanalyseerd. Dit betekent dat er bewerkingen op worden uitgevoerd die noodzakelijk zijn om de probleemstelling(en) van het onderzoek op te lossen. Analyse komt in de eerste plaats neer op het weerhouden van de relevante variabelen en/of de comprimering van de vele waarden in de datamatrix tot enkele essentiële (= reductie en/of transformatie datamatrix).

## 2.3.2. Gegevensanalyse

### 2.3.2.1. Modelspecificatie

Wanneer een onderzoeker in het eerste stadium van de bestudering van een gebied (cfr. probleemstelling) nadenkt over wat voor aspecten het te bestuderen verschijnsel heeft, is de onderzoeker in feite bezig met de afbakening van een systeem, d.w.z. een geheel van min of meer afgebakende stukjes van de sociale realiteit. Geleid door ervaring, intuïtie, analogie-redenering en dergelijke streeft de onderzoeker ernaar om de 'belangrijke' variabelen en relaties in het onderzoek te betrekken en de 'minder belangrijke' er buiten te laten. Aldus bouwt de onderzoeker een vereenvoudigd model op van het systeem.

Een model is een systeem van elementen (eenheden, variabelen) en relaties tussen die elementen, dat zoveel mogelijk lijkt op het systeem waarvan het een model is. Een model is dus een hulpmiddel om de werkelijkheid beter te leren kennen. De eigenlijke modelspecificatie (de bepaling van de relaties tussen de variabelen) heeft een technisch karakter. Bij modellen gaat het immers om het vaststellen van de *eenheden*, de *variabelen* en de *relaties* tussen de variabelen.

#### 2.3.2.1.1. Het vaststellen van de eenheden

In uitspraken komen eenheden, variabelen en relaties tussen variabelen voor. In deze paragraaf gaan we nader in op de eenheden. De verzameling van alle eenheden waarover een uitspraak iets zegt, noemen we het domein of de reikwijdte van een uitspraak. De eenheden zijn niet altijd mensen of groepen. Het kunnen ook instituten, organisaties,

landen, culturen zijn. Maar ook cultuuruitingen : boeken, kranten, films, toespraken, uitzendingen e.d.

Wanneer we ons voornamelijk richten op individuele personen als eenheden, betekent dit niet dat we bij de modelafbakening alleen te maken hebben met kenmerken van individuen. Individuen worden immers beïnvloed door de groepen waar ze deel van uitmaken (en die groepen worden beïnvloed door de personen waaruit ze zijn opgebouwd). Als individuele personen de eenheden zijn, zullen we dus kenmerken van de groepen waar die personen deel van uitmaken in het model moeten betrekken. Zouden groepen de eenheden zijn, dan hebben we niet alleen met eigenschappen van die groepen, maar ook met eigenschappen van de individuen die van die groepen deel uitmaken, te maken.

In de meeste gevallen zullen dus kenmerken van eenheden van verschillend aggregatieniveau worden opgenomen. Betekent dit dat de dataverzameling zich ook altijd op verschillende aggregatieniveaus moet afspelen ? Dat we bij een onderzoek naar individueel gedrag niet alleen de personen zelf, maar op een of andere wijze ook de groepen waarvan ze deel uitmaken (bedrijf, sportclub, e.a.) moeten benaderen ? Het antwoord daarop is dat dit soms nuttig of mogelijk is, maar dat het tot op zekere hoogte mogelijk is om bv. in een personenenquête kenmerken van hogere niveaus mee te nemen. Zo kan men aan respondenten vragen hoe groot het bedrijf is waarin zij werken, wat de overwegende politieke gezindte is van hun vriendenkring, enz.

Bij de afbakening van een model, in het bijzonder de keuze van de variabelen, is de vraag van belang of het een fundamenteel dan wel een praktijkgericht onderzoek betreft. In een praktijkgericht onderzoek zoekt men altijd naar manipuleerbare variabelen; in een fundamenteel onderzoek hoeft men dat niet per se. Zo kan in een marktonderzoek een verband gevonden worden tussen regio en klantenontevredenheid. Voor een onderneming is het evenwel belangrijker te weten hoe het verband tussen regio en klantenontevredenheid tot stand komt, m.a.w. de interpretatie van het verband te vinden. Wellicht dat men dan stuit op een variabele zoals distributie, die beïnvloed kan worden.

#### 2.3.2.1.2. Het vaststellen van de variabelen

De eigenschappen van de eenheden noemen we variabelen. Een gebruikelijke definitie van 'variabele' luidt dan ook : een eigenschap of kenmerk waarop eenheden van elkaar kunnen verschillen.

Met betrekking tot de variabelen die in marktonderzoek veel aan de orde zijn, dient in de eerste plaats een onderscheid gemaakt te worden tussen de soorten eenheden waarop de variabelen betrekking hebben. Anderzijds kunnen variabelen onderscheiden worden naar de waarden ervan en tenslotte naar het meetniveau ervan.

##### 2.3.2.1.2.1. Variabelen naar de eenheden waarop ze betrekking hebben

Eigenschappen van individuen worden onderverdeeld in :

\* achtergrondvariabelen :

- fysiek-biologische variabelen (e.g. geslacht, leeftijd)
- psychologische variabelen (e.g. intelligentie, sociabiliteit)
- sociaal-structurele kenmerken (e.g. beroep, opleiding) ;

\* attituden, motivaties, opinies ;

\* gedragsvariabelen.

Eigenschappen van groepen worden, naargelang van de omvang van de groepen, onderverdeeld in :

- \* variabelen op micro-niveau (kleine groepen) ;
- \* variabelen op meso-niveau (groepen op meso-niveau) ;
- \* variabelen op macro-niveau (op het niveau van de maatschappij).

Er dient aangestipt te worden dat sommige eigenschappen van groepen worden berekend uit de waarden die de leden van de groepen op zo'n variabele hebben (bv. gemiddelde leeftijd van de Vlaamse bevolking). Andere eigenschappen van groepen kunnen niet herleid worden uit een eigenschap van individuen, maar 'bestaan' alleen op groepsniveau (bv. soort leiderschap, bedrijfsklimaat, e.a.).

Groepen, of algemener, verzamelingen van eenheden, hebben verschillende soorten kenmerken. In de eerste plaats zijn er *analytische kenmerken*. Dit zijn eigenschappen van verzamelingen die verkregen worden via een of andere algebraïsche operatie op een kenmerk van de elementen die deel uitmaken van die verzameling (bv. gemiddelde leeftijd van groepsleden, ziekteverzuim, e.a.). Analytische kenmerken worden veelal berekend op basis van absolute, globale kenmerken van personen. De eerder genoemde fysisch-biologische, psychologische en sociaal-structurele achtergrondkenmerken, attitudes, opinies en gedragingen zijn absolute kenmerken van personen.

*Structurele kenmerken* zijn eigenschappen van verzamelingen die verkregen worden via een of andere algebraïsche operatie op data over de (sociale) relaties van de elementen onderling (bv. de gemiddelde frequentie van sociale interactie). Wordt over alle of sommige sociale betrekkingen tussen elementen van een verzameling gegeneraliseerd, dan ontstaat een structureel kenmerk van een verzameling.

Vervolgens kan men een verzameling typeren door *globale kenmerken*. Globale kenmerken zijn niet gebaseerd op kenmerken van de elementen of van onderlinge relaties (bv. de mate van luchtvervuiling in een gemeente).

Een volgende groep kenmerken zijn *relationele kenmerken*. Deze hebben betrekking op sociale relaties tussen eenheden van eenzelfde aggregatieniveau (bv. een groep inpaksters kan goede of slechte relaties onderhouden met een of meer andere groepen inpaksters). Relationele kenmerken treffen we op lagere en hogere aggregatieniveaus aan. Immers, de structurele kenmerken van verzamelingen zijn gebaseerd op de sociale betrekkingen tussen de elementen, met andere woorden op relationele kenmerken van die elementen.

Tenslotte hebben we nog te maken met *contextuele kenmerken*. Dit zijn kenmerken die aan de eenheden van een lager niveau worden toegekend, maar die gebaseerd zijn op (analytische, structurele, globale of relationele) kenmerken van een hoger niveau (bv. het behoren tot een grote of kleine klas).

We gaan hieronder nog wat nader in op contextuele variabelen, niet omdat deze op een specifieke manier zouden 'werken', maar omdat op historische gronden in het gebruikelijke onderzoek wél veel gewerkt wordt met achtergrondvariabelen zoals geslacht, leeftijd, inkomen, e.a., terwijl de veel 'kleinere' sociale contexten waarin het individu leeft op de achtergrond blijven. Het is o.m. om die reden dat op de enquête als dataverzamelvorm veel kritiek is geleverd. Door de atomistische steekproeftrekking zou het individu losgemaakt worden uit de sociale context en daarmee zou de invloed van significante anderen ("significant others") op het gedrag van het individu over het hoofd gezien worden. Het is inmiddels mogelijk gebleken om ook in een enquête sommige kenmerken van die concrete sociale kaders te achterhalen. Al naar gelang het te verklaren individuele gedrag zullen we moeten letten op bepaalde contextuele kenmerken (bv. de aan- of afwezigheid van



een motiverend gezin als globaal kenmerk of het gemiddelde IQ van klasgenoten als analytisch kenmerk wanneer het gaat om het verklaren van schoolprestaties ; de aanwezigheid van reeds delinquente vrienden als analytisch kenmerk wanneer het gaat om het verklaren van delinquent gedrag ; het contactpatroon in de buurt als structureel kenmerk wanneer het gaat om het verklaren van verhuisgeneigdheid ; de gemiddelde motivatie van de werkgroepsleden als analytisch kenmerk als het gaat om het verklaren van de eigen prestaties, e.a.).

Met het vaststellen van een correlatie tussen een contextuele variabele en individueel gedrag zijn we er evenwel nog niet. We moeten ook kunnen begrijpen waarom dat verband bestaat. We zijn niet geneigd om met een geheimzinnige werking 'sui generis' van de groep op de elementen genoeg te nemen. De interpretatie die aan de werking van contextuele variabelen wordt gegeven, vindt steeds plaats in termen van het handelend individu of in termen van relaties tussen individuen. In de praktijk blijkt de correlatie tussen een contextuele variabele en individueel gedrag vaak bij uitsplitsing van de tabel te verdwijnen. Zo vonden onderzoekers een verband tussen het aspiratieniveau van leerlingen en de sociaal-economische status van de school. Dit verband bleek te verdwijnen wanneer de gevonden samenhang werd gecontroleerd voor de sociaal-economische status van de vaders van de vrienden. Uit het onderzoek bleek dat 'de invloed van de school' in feite betekent dat er op uiteenlopende scholen een uiteenlopende kans is om met vrienden van hoge sociale status in aanraking te komen, en dat het overnemen van hoge aspiraties geheel verloopt via beïnvloeding door vrienden. Het gaat dus niet om een of andere 'geheimzinnige werking' van het geheel op het individu : we kunnen 'eigenschappen van groepen of situaties' vertalen in termen van eigenschappen (meestal kansen) voor het individu zelf (d.w.z. de kansen die het individu vanuit zijn contexten krijgt aangereikt om gedragsalternatieven te ontwikkelen. Een contextueel kenmerk geeft als het ware deze kansen weer.

#### 2.3.2.1.2.2. Variabelen naar de waarden die ze aannemen

Categorieën of klassen van een eigenschap waarin via een uitspraak de eenheden geplaatst worden, noemen we de waarden van een eigenschap (variabele). We kunnen ons in de eerste plaats de vraag stellen hoeveel deelverzamelingen door een variabele worden aangebracht, of met andere woorden hoeveel waarden (klassen, categorieën) een variabele heeft. In de tweede plaats komt de vraag in welk opzicht de waarden van een variabele van elkaar verschillen. Dit laatste heeft betrekking op het meetniveau van variabelen (zie volgende paragraaf).

Een variabele heeft uiteraard minstens twee waarden. Heeft een variabele slechts twee waarden, dan spreken we van een dichotome variabele. Een variabele met drie waarden is een trichotome variabele. Voor variabelen met 4,5, ... k waarden gebruikt men het verzamelbegrip polytome variabelen.

Wanneer we tussen de waarden van een variabele in principe een oneindig aantal 'tussenstapjes' kunnen denken, spreken we van een continue variabele. Alle niet-continue variabelen zijn discrete variabelen. Vooral bij continue variabelen rijst de vraag naar het aantal te onderscheiden waarden. Hoe meer waarden er gebruikt worden, des te beter de eenheden van elkaar onderscheiden worden, met andere woorden des te meer informatie de frequentieverdeling verschaft. Echter, hier zijn twee bezwaren tegen in te brengen : in een rapport is overzichtelijkheid van groot belang en er is een grens aan het aanbrengen van fijne onderscheidingen omdat deze relevant moeten blijven.

### 2.3.2.1.2.3. Variabelen naar het meetniveau

Doorgaans worden vier meetniveaus onderscheiden, nl. nominaal, ordinaal, interval en ratio.

In het geval van een *nominale variabele* worden een eenheden gewoonweg geclassificeerd. Variabelen waarvan de waarden niet meer doen dan van elkaar verschillende klassen vormen, noemen we nominale variabelen.

Bij *ordinale variabelen* hebben de waarden een vaste onderlinge volgorde. Dit betekent dat de elementen die door een ordinale variabele beschreven worden niet alleen onderling van elkaar verschillen maar eveneens dat de een hoger, meer, positiever of iets dergelijks is dan de ander. Variabelen waarvan de waarden een vaste volgorde vertonen noemen we ordinale variabelen.

Variabelen waarbij de afstand tussen twee opeenvolgende waarden constant is, noemen we *interval-variabelen*.

Variabelen die behalve de vaste afstand tussen twee opeenvolgende waarden ook een natuurlijk nulpunt bezitten, noemen we *ratio-variabelen*.

### 2.3.2.1.3. Relaties

Zeer vaak hebben uitspraken niet betrekking op een variabele (*enkelvoudige uitspraken*), maar op de relatie (synoniemen : statistisch verband, associatie, correlatie) tussen twee of meer variabelen. Wanneer in een uitspraak minstens twee eigenschappen variabel zijn, spreken we van een *samengestelde uitspraak*. Bij *bivariate analyse* wordt gewerkt met twee kolommen van de datamatrix ; bij *multivariate analyse* wordt gewerkt met meer dan twee kolommen uit de datamatrix.

In marktonderzoek wordt in toenemende mate gebruik gemaakt van tal van statistische technieken en toetsen. *Analysetechnieken* (statistische technieken) worden gehanteerd om de samenhang tussen variabelen te bepalen. *Statistische toetsen* hebben tot doel te toetsen of de gevonden verbanden al of niet significant zijn (statistische afhankelijkheid, statistische onafhankelijkheid).

Het meten van associatie is een techniek die zich in een van de laatste etappes van een onderzoek, de zgn. analysefase, situeert. Maar om te begrijpen wat met associatie in de concrete praktijk wordt bedoeld, moeten nagenoeg alle andere etappes mede worden bekeken. De probleemstelling, de theorie- en hypothesenvorming, de onderzoeksvorm, de onderzoekspopulatie, de waarnemingstechnieken, e.d. bepalen de wijze waarop statistische verbanden worden opgespoord. De techniciteit van de analysefase mag de onderzoeker niet doen vergeten dat bij het meten van associaties of, meer algemeen, het doen van uitspraken over de samenhang tussen variabelen, een smeltkroes van problemen opduikt. In dit verband dient er o.m. op gewezen te worden dat statistische associatie niet hetzelfde is als associatie in de sociale werkelijkheid. In de sociale realiteit zijn verbanden tussen variabelen vrijwel nooit deterministisch, maar *probabilistisch*. In de probabilistische uitspraak over het verband tussen de variabelen X en Y wordt 'slechts' beweerd dat de relatieve (bv. procentuele) verdeling over de waarden van de variabele Y van eenheden die op X een bepaalde waarde hebben, anders is dan die van de eenheden die op X een andere waarde hebben. Nu is het nadeel van probabilistische uitspraken dat er een zeer groot aantal kruistabellen denkbaar is waarbij aan die uitspraak wordt voldaan. Probabilistische uitspraken laten dus veel meer situaties toe, sluiten veel minder uit dan deterministische

uitspraken. Probabilistische uitspraken zijn dus minder informatief dan deterministische. Het 'realistischer' zijn van probabilistische uitspraken betekent immers juist een geringere informativiteit vergeleken met deterministische uitspraken. Als algemene regel kan gesteld worden dat gestreefd moet worden naar *zo informatief mogelijk probabilistische uitspraken*.

Dit laatste brengt ons tot de *methodologische regels waaraan uitspraken moeten voldoen*, nl. *precisie, falsifieerbaarheid* en *informativiteit*. Uitspraken moeten duidelijk of precies zijn, omdat zij alleen dan door iedereen op dezelfde wijze begrepen worden. Falsifieerbaar betekent dat het mogelijk moet zijn om de eventuele onjuistheid van een uitspraak aan te tonen. Falsifieerbare kennis is dus eigenlijk kwetsbare kennis. Een uitspraak is informatiever naarmate deze meer mogelijkheden uitsluit en op een breder gebied van de empirische werkelijkheid betrekking heeft. Een gevolg hiervan is naarmate een uitspraak informatiever is, deze ook kwetsbaarder, onzeker wordt. Hieruit blijkt al dat we niet tegelijkertijd kunnen streven naar kennis die en informatief en zeker is (het streven naar zekerheid is overigens geen richtinggevend beginsel bij wetenschapsbeoefening). In het streven naar zo informatief mogelijke uitspraken wordt op een andere manier uitgedrukt wat wel eens als kenmerkend voor wetenschap wordt aangeduid, nl. streven naar algemeen generaliseerbare kennis.

Hoewel bij de analysefase van een onderzoek heel wat technische aspecten om de hoek komen kijken, wordt de keuze van een geschikte analysetechniek o.m. mede bepaald door de aard van de probleemstelling, de onderzoekspopulatie en het soort onderzoek. Zolang gebruik wordt gemaakt van interne bedrijfsgegevens of gegevens die betrekking hebben op een ganse populatie, zijn alle cijfers bekend (er is een continue informatiestroom) en kunnen de universumparameters exact berekend worden. De beschrijvende statistiek houdt zich bezig met het overzichtelijk maken van zulke volledige dataverzamelingen. Zodra gebruik wordt gemaakt van steekproeven (bv. ad hoc-onderzoek) moeten universumparameters afgeleid worden uit steekproefparameters. Als voorwaarden gelden daarbij dat de steekproef representatief is voor het universum en dat de steekproef voldoende groot is. Een steekproefuitkomst zal slechts bij benadering gelijk zijn aan de universumgegevens. Het is daarom gebruikelijk om betrouwbaarheidsmarges aan te geven voor de steekproefuitkomsten. Daarbij geldt dat hoe groter de steekproef, hoe betrouwbaarder de schatting, dus hoe kleiner de betrouwbaarheidsmarges (wet van de grote aantallen). Het doen van uitspraken over een populatie op basis van een steekproef wordt *inferentiële statistiek* genoemd.

Het onderscheid tussen beschrijvende en inferentiële statistiek houdt vanzelfsprekend verband met het type probleemstelling dat aan een onderzoek ten grondslag ligt (beschrijvingsproblemen, verklaringsproblemen, voorspellingsproblemen) en het soort onderzoek dat wordt verricht (instrumenteel-nomologisch onderzoek, beschrijvend onderzoek, toetsend onderzoek, verklarend onderzoek).

#### 2.3.2.1.3.1. De afbakening van het causaliteitsbegrip

Opvattingen over causaliteit zijn nauw verbonden met opvattingen over wetenschap. In wetenschapstheoretische discussies blijken vragen van belang als : ligt het in de aard der dingen dat zich causale processen afspelen (m.a.w. bestaat causaliteit) of is causaliteit 'slechts' een wijze van ordenen van de wereld door ons denken ? Ook voor de wetenschapstheorie belangrijke begrippen als wetmatigheid, determinisme, waarschijnlijkheid en toeval zijn nauw verbonden met het causaliteitsbegrip. Verder wordt ook bij verklaren en voorspellen het causaliteitsbegrip nogal eens verondersteld. Op al deze wetenschapstheoretische onderscheidingen zal hier niet verder worden ingegaan. Er zal vooral geprobeerd worden zo duidelijk mogelijk te maken wat een belangrijk deel van de

onderzoekers binnen de empirische (sociale) wetenschappen bedoelt met de uitspraak 'X is oorzaak van Y'.

*[1] Een causale relatie heeft een productiekarakter*

Variatie in X brengt variatie in Y teweeg. Dit idee maakt een causale relatie asymmetrisch. In de omschrijving van een causale relatie als 'variatie (of verschil) in X produceert variatie (of verschil) in Y' behoeven de termen 'variatie of verschillen in X en Y' enige toelichting. Sommige auteurs, w.o. Blalock (1964), omschrijven een causale relatie als: veranderingen in X produceren veranderingen in Y. Nu is het uiteraard mogelijk variatie te meten via verandering met behulp van longitudinaal onderzoek, d.w.z. diverse metingen te verrichten in de loop van de tijd bij hetzelfde onderzoekselement en na te gaan of er verschillen in de loop der tijd optreden in X en Y en of deze aan elkaar gerelateerd zijn. Maar eveneens kan men variatie of verschillen in X en Y vaststellen met behulp van zgn. cross-sectioneel onderzoek. Hoewel in een concrete onderzoekssituatie de ene wijze van meten van variatie duidelijk voordelen kan hebben boven de andere, volgt daaruit niet dat in de 'definitie' van causaliteit het begrip variatie tot één van beide beperkt moet blijven. Hiermede wordt impliciet ook de zienswijze afgewezen van degenen die het gebruik van het causaliteitsbegrip willen beperken tot die situaties waarin variatie tot stand komt door bewuste manipulatie van de onafhankelijke variabelen. Alleen in experimenteel onderzoek zou dan sprake kunnen zijn van het vaststellen van causale relaties.

*[2] Een causale relatie heeft een noodzakelijk karakter*

Daarmee wordt bedoeld dat steeds wanneer een bepaalde variatie in de ene variabele (de oorzaak) optreedt, er ook steeds een bepaalde variatie in de andere variabele (het gevolg) tot stand komt, m.a.w. de variatie in de ene variabele is een *voldoende voorwaarde* voor de variatie in de andere variabele. Daarbij gaat men er wel van uit dat alle omstandigheden gelijk zijn. Noodzakelijkheid kan zo in termen van *herhaalbaarheid* omschreven worden. Steeds wanneer dezelfde omstandigheden zich voordoen, zal een bepaalde variatie in de ene variabele een bepaalde variatie in de andere variabele tot gevolg hebben.

Dit wil niet zeggen dat alleen van causaliteit gesproken kan worden in geval van herhaalde gebeurtenissen. Ook bij eenmalige gebeurtenissen kan men spreken van oorzaak en gevolg. Het gaat om *conditionele herhaalbaarheid*: als hetzelfde complex van oorzakelijke omstandigheden zich weer zou voordoen, dan zouden dezelfde gevolgen plaatsvinden.

Bij de empirische vaststelling van het noodzakelijke karakter van een causale relatie stuit men wel op het zgn. inductieprobleem. Hoe stelt men bv. vast dat het complex van omstandigheden in twee gevallen volkomen identiek is? Het is onmogelijk dit met zekerheid empirisch aan te tonen.

*[3] Een causale relatie is een precies gedefinieerde functionele relatie*

Een ander belangrijk kenmerk van een causale relatie, in het verlengde liggend van het hierboven vermelde kenmerk van noodzakelijkheid, is dat de aard van de relatie steeds dezelfde is. Met causaliteit wordt bedoeld dat een bepaalde variatie in X ook steeds een bepaalde variatie in Y teweegbrengt, m.a.w. X legt vast hoe Y 'eruit zal zien'. Een causale relatie tussen X en Y impliceert daarom dat variatie in X variatie in Y teweegbrengt volgens een bepaalde welomschreven regel die de functionele relatie tussen X en Y specificeert. Daarbij geldt wel dat dezelfde functionele relatie slechts hoeft op te treden in gelijke omstandigheden.

De empirische bewijsvoering voor het bestaan van een causale relatie stuit op twee belangrijke problemen. Op de eerste plaats onttrekt het produktiekarakter zich aan de directe waarneming. Op de tweede plaats stuit men steeds op de onmogelijke opgave alle relevante factoren of omstandigheden op te sporen. Blalock prefereert dan ook over causaliteit te spreken in termen van een *causaal model*. In dit model wordt er van uitgegaan dat variatie in X op een bepaalde welomschreven wijze variatie in Y produceert. Binnen dit model wordt verder aangenomen dat omstandigheden naar believen herhaald kunnen worden en dat alle relevante omstandigheden bekend zijn. Uit dit model worden vervolgens een aantal empirische consequenties afgeleid. Indien het model niet klopt, wordt het verworpen of op een of meer onderdelen aangepast. Indien het model wel klopt, accepteren we het voorlopig. Strikte bewijsvoering voor de juistheid van het causale model is onmogelijk. Zoals elke hypothese kan het causale model alleen gefalsifieerd, niet geverifieerd worden.

### 2.3.2.1.3.2. Criteria ter toetsing van causale hypothesen

#### 2.3.2.1.3.2.1. Geldige criteria

Doorgaans worden in de literatuur (e.g. Hirschi & Selvin, 1967) drie criteria vermeld ter toetsing van causale hypothesen : er is *samenhang tussen X en Y, X gaat in tijd vooraf aan Y* en er zijn geen andere variabelen die de samenhang tussen X en Y tot stand brengen (of : de samenhang tussen X en Y is *geen schijnsamenhang* en *de werking van storende factoren is uitgeschakeld*).

Deze drie criteria worden ieder op zich als een noodzakelijke voorwaarde beschouwd : indien men kan aantonen dat de relatie X-Y aan een van de drie criteria niet voldoet, is de causale hypothese weerlegd. De drie criteria worden samen bij conventie, bij wijze van afspraak, ook wel als voldoende voorwaarde beschouwd : voor zover men kan aantonen dat de relatie X-Y aan de drie criteria voldoet, is de houdbaarheid van de causale hypothesen voorlopig aangetoond.

#### 2.3.2.1.3.2.2. Ongeldige criteria

Naast de drie genoemde criteria blijken onderzoekers in de praktijk nog diverse andere criteria te hanteren. Een aantal hiervan behandelen Hirschi en Selvin (1967) onder de hoofding "false criteria of causality". Bespreking van deze ongeldige criteria heeft niet alleen het voordeel dat een aantal foute redeneringen m.b.t. causale analyse duidelijk worden gemaakt, maar vooral geeft zij de gelegenheid op een aantal punten nader te preciseren wat er precies met een causale relatie bedoeld wordt.

##### [1] *Ongeldig criterium 1 : perfecte relatie*

(als een relatie tussen twee variabelen niet perfect is, dan is de relatie niet causaal)

Het eerder vermelde geldige criterium 'samenhang' wordt hier toegespitst op de eis van 'perfecte samenhang'. Een dergelijk standpunt houdt in dat causale hypothesen een deterministisch karakter hebben. In de opvatting over causaliteit wordt daarentegen aangenomen dat causale hypothesen een stochastisch of probabilistisch karakter hebben, m.a.w. de waarnemingen zullen niet exact op de regressielijn liggen maar in een puntenwolk er omheen (of in het geval van non-parametrische samenhangen : de kansen op Y verschillen naargelang van de categorieën van X). De waarden van de onafhankelijke variabele leggen de waarden van de afhankelijke wel vast, maar slechts met een bepaalde marge, met

bepaalde kansen. Niet-perfecte samenhangen zijn derhalve geen reden causale hypothesen te verwerpen, maken het bestaan van causale relaties niet onmogelijk.

[2] *Ongeldig criterium 2 : karakteristieke factor*

("als een factor niet karakteristiek is voor een bepaalde categorie, dan is deze factor geen oorzaak van het behoren tot deze categorie")

Dit argument komen we in allerlei gedaanten tegen. Roken is geen oorzaak van longkanker, want de overgrote meerderheid van de rokers heeft geen longkanker ; sociaal milieu is niet bepalend voor het wel of niet doorstuderen, want ook van de zgn. hogere klassen studeert slechts een zeer beperkt deel aan de universiteit. Op het eerste gezicht lijkt dit argument plausibel. Wanneer een factor niet typerend is, kan hij niet belangrijk zijn als oorzaak. Maar een (oorzakelijk) verband tussen twee variabelen kan niet vastgesteld worden door slechts één categorie van de onafhankelijke variabele te bekijken. Niet het feit dat slechts een minderheid van de rokers longkanker krijgt (bv. 5%) of dat slechts een minderheid van de hogere milieus doorstudeert (bv. 20%), is op zich relevant voor de vaststelling van samenhang. Het gaat om de vergelijking van deze percentages met het overeenkomstige percentage voor de niet-rokers, resp. de lagere sociale milieus. Indien zou blijken dat van de niet-rokers slechts 0,05% longkanker zou krijgen en van de lagere sociale milieus slechts 1% naar de universiteit gaat, dan is er vooralsnog geen reden, op basis van deze cijfers, de causale hypothese te verwerpen. Uiteraard geldt de onjuistheid van dit criterium ook in het omgekeerde geval : indien een factor wel karakteristiek is, is dit nog geen aanwijzing voor het bestaan van een causale relatie. Indien hogere inkomensgroepen een zeer hoge status hebben, zegt dit niets over de causale invloed van inkomen op status. Immers, indien de lagere inkomenscategorieën een even hoge status hebben, is er vooralsnog geen enkele reden de causale hypothese te aanvaarden.

[3] *Ongeldig criterium 3 : geen samenhang binnen een categorie*

("als er een relatie gevonden wordt tussen een onafhankelijke en een afhankelijke variabele binnen een categorie van een derde variabele, dan kan deze derde variabele nooit oorzaak zijn van de afhankelijke variabele")

Dit criterium is ongeldig omdat er over de invloed van een variabele die in een onderzoek niet varieert, op basis van dat onderzoek niets zinnigs kan gezegd worden.

[4] *Ongeldig criterium 4 : geen interveniërende variabelen*

("als er een relatie gevonden wordt tussen een onafhankelijke en een afhankelijke variabele en als een derde variabele intervenueert tussen beide, dan is de oorspronkelijke relatie niet causaal")

In feite stelt dit criterium dat indirecte oorzaken geen echte oorzaken zijn. Variatie in de onafhankelijke variabele veroorzaakt echter wel degelijk - zij het indirect - via de interveniërende variabele(n) variatie in de afhankelijke variabele. De conclusie zou dus moeten zijn dat er geen directe, maar indirecte causaliteit is, en niet dat er geen causaliteit is.

Het misverstand, waardoor dit criterium toch nog al eens gehanteerd wordt, vindt vermoedelijk zijn oorsprong in het feit dat de interveniërende variabele zich in zekere zin 'gedraagt' als een storende factor van het type schijnsamenhang. Wanneer een bepaalde relatie X-Y *schijnsamenhang* is, m.a.w. slechts tot stand komt door de werking van een storende factor, dan zal de samenhang X-Y verdwijnen wanneer de effecten van deze storende factor worden uitgeschakeld, bv. door de storende factor constant te houden. Dit

verdwijnen van de samenhang tussen X en Y kan dan als een bewijs gezien worden voor het niet causaal zijn van de relatie X-Y. Wanneer nu de effecten van de onafhankelijke variabele X op de afhankelijke variabele Y volledig lopen via de interveniërende (niet : de storende) variabele T, wanneer er dus geen eigen, direct is effect is van X op Y, dan zal eveneens de samenhang tussen X en Y verdwijnen bij constant houden van T, doordat dan a.h.w. de weg waarlangs het verband loopt, geblokkeerd wordt. In het geval van schijnsamenhang ("spurious correlation") is T een storende variabele die samenhang teweeg brengt tussen X en Y zonder dat variatie in X op een of andere wijze variatie in Y tot gevolg heeft (T is gemeenschappelijke oorzaak van X en Y). In het geval van interpretatie is T een *interveniërende variabele* die inzichtelijk maakt op welke wijze variatie in X variatie in Y produceert, nl. via variatie in T. De functie van interveniërende variabelen is dus het inzichtelijk maken van de wijze waarop de causale invloed van de ene variabele op de andere tot stand komt. Soms wordt hieruit afgeleid dat een relatie tussen twee variabelen pas dan 'echt' causaal genoemd mag worden, als het hele proces waarlangs de beïnvloeding plaatsvindt, aangegeven is, m.a.w. als alle interveniërende variabelen gevonden zijn. Dit is in feite een vrij onvruchtbare positie. In de eerste plaats wordt het aantal mogelijk interveniërende variabelen alleen beperkt door de fantasie van de onderzoeker. Met enige vindingrijkheid is tussen elke twee variabelen weer wel een derde te plaatsen, m.a.w. van elke directe relatie is een indirecte te maken, zodat het vinden van alle interveniërende variabelen bij voorbaat onmogelijk is. Maar, en dat is tevens de tweede reden voor het onvruchtbare van deze positie, het is ook vrij zinloos alle mogelijke interveniërende variabelen op te sporen omdat vanuit het gekozen theoretisch gezichtspunt, vanuit het causale model dus, slechts bepaalde interveniërende variabelen van belang zijn en andere niet. Die interveniërende variabelen moeten dus opgenomen worden die binnen de gekozen theoretische optiek van belang zijn. Dit 'moeten opgenomen worden' is dan, het zij nogmaals benadrukt, niet nodig om tot causaliteit te kunnen besluiten, maar wel om een theoretisch 'bevredigende' causale verklaring te geven.

[5] *Ongeldig criterium 5 : meetbare variabelen zijn geen oorzaak*

De betekenis van deze uitspraak is afhankelijk van de betekenis van het begrip 'meetbare variabelen'. Soms wordt er onder verstaan theoretische variabelen die gemakkelijk te operationaliseren zijn (bv. leeftijd, inkomen). Het is niet goed in te zien waarom het "toevallig" goed meetbaar zijn van een variabele tot gevolg heeft dat deze geen causale invloed kan uitoefenen. Iets anders ligt de situatie wanneer met meetbare variabelen de indicatoren of operationele definities bedoeld worden. In dit geval is de uitspraak zinloos. Immers, causaliteit is een begrip dat thuis hoort op theoretisch en niet op operationeel niveau.

Op theoretisch niveau is er sprake van theoretische variabelen en van een causale relatie tussen deze variabelen. Op operationeel niveau is er sprake van indicatoren en van samenhang tussen deze indicatoren. In hoeverre we van het geconstateerde op empirisch niveau tot causaliteit kunnen besluiten, is afhankelijk van de geldigheid (validiteit) van de operationalisering en van het voldaan zijn aan de drie criteria voor causaliteit. Spreken over de (niet-)causale werking van een indicator wordt zo bezien echter onzinnig en komt neer op verwarring op twee niveaus : het theoretische en het operationele niveau.

[6] *Ongeldig criterium 6 : geen interactie*

(als een relatie tussen een onafhankelijke variabele en een afhankelijke variabele afhankelijk is van waarden van andere variabelen, is de onafhankelijke variabele geen oorzaak van de afhankelijke)

Dergelijke gevallen van differentiële samenhang, meestal interactie genoemd, zouden m.a.w. causaliteit uitsluiten. Hiertegen kunnen we stellen dat interactie geen bewijs is voor de afwezigheid van causaliteit, maar een specificatie daarvan inhoudt. De condities waaronder de causale relatie geldt, worden aangegeven. Idealiter worden de condities waaronder de veronderstelde causale relatie geldt, uitputtend in het causale model aangegeven. In feite lukt dit natuurlijk slechts ten dele en verder onderzoek dient er dan ook voor om dergelijke condities bloot te leggen, m.a.w. mogelijke interacties op te sporen.

### 2.3.2.2. Symmetrische en asymmetrische analysetechnieken

Een onderscheid kan gemaakt worden tussen twee soorten analysetechnieken. Zijn de variabelen van dezelfde soort, dan spreekt men van *symmetrische analysetechnieken*. In dergelijk geval wordt geen onderscheid gemaakt tussen onafhankelijke en afhankelijke variabelen. Wordt een dergelijk onderscheid wél gemaakt dan wordt gesproken van *asymmetrische analysetechnieken*. Wanneer we over causale analyse spreken, hebben we altijd te maken met variabelen van verschillend niveau en wordt dus een onderscheid gemaakt tussen de predictoren en de afhankelijke variabele(n).

Wanneer een gedeelte van de datamatrix in beslag wordt genomen door groepjes variabelen van hetzelfde niveau, is de eerste doelstelling van de analyse meestal om een dergelijk groepje variabelen te reduceren tot één nieuwe variabele, waarin specifieke eigenaardigheden van de diverse vraagstellingen tegen elkaar wegvallen, en dientengevolge als een valider maatinstrument (index) mag worden gezien dan elk van de afzonderlijke variabelen (indicatoren). We kunnen een groep van variabelen die op hetzelfde niveau liggen ook vanuit een andere en wat gecompliceerder doelstelling bekijken. In plaats van uit te gaan van de veronderstelling dat alle variabelen op hetzelfde concept slaan, kunnen we ook trachten te ontdekken of er een bepaalde structuur te ontdekken valt in de betreffende verzameling variabelen. Het is denkbaar dat sommige variabelen onderling hoog correleren, maar lage correlaties te zien geven met andere variabelen, die misschien intern een ander hoogcorrelerend subgroupje vormen. De matrix valt dan uiteen in enkele min of meer homogene submatrices.

Terwijl bij symmetrische analyseprocedures uitsluitend wordt uitgegaan van de mate van statistisch verband tussen twee of meer variabelen, kunnen we ons ook afvragen of de ene variabele niet de oorzaak van de andere zou kunnen zijn ; of dat er sprake zou kunnen zijn van één variabele als oorzaak en vele andere (gevolg)variabelen ; van meerdere variabelen als oorzaken van één gevolg ; van meerdere variabelen die al dan niet onafhankelijk van elkaar verschillende andere variabelen beïnvloeden, enz. De hypothesen waaruit een theorie is opgebouwd, worden vrijwel altijd geformuleerd in termen van oorzaak en gevolg. Dit ligt voor de hand, omdat in een theorie, wil deze bruikbare kennis van de werkelijkheid opleveren, deze kennis de vorm moet kunnen aannemen van voorspellingen. Hieruit volgt, dat zodra we ons bezighouden met theorie-exploratie en -toetsing, we met méér aspecten van de relatie tussen variabelen te maken hebben dan alleen met de mate van statistisch verband. De analyseprocedures die tot doel hebben om stelsels van causale hypothesen te exploreren of te toetsen kan men asymmetrische analyseprocedures noemen, omdat de oorzaak-gevolg relatie asymmetrisch van aard is.

In de schemata op volgende bladzijden wordt een indeling gegeven van analysetechnieken. Deze schemata kunnen als "flow charts" gebruikt worden bij de keuze van een geschikte analysetechniek (wat asymmetrische analysetechnieken betreft, beperken we ons voor de keuze van een geschikte analysetechniek tot de gevallen waarbij één afhankelijke variabele in het causale model is opgenomen).



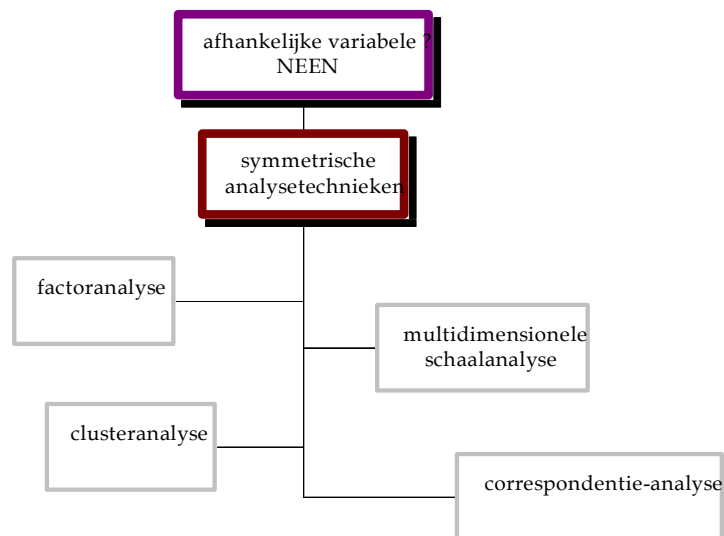
### 2.3.2.2.1. Interdependentietechnieken

In figuur 11 staan de analysetechnieken vermeld die gebruikt worden bij afwezigheid van een afhankelijke variabele. Het doel van de toepassing van *interdependentietechnieken* zoals deze ook genoemd worden, is het opsporen van relaties tussen variabelen om een mogelijk aanwezige, maar nog niet bekende structuur in de gegevens te ontdekken. Een vaak gebruikte techniek is *clusteranalyse*, die tot doel heeft groepen (segmenten) te construeren op basis van kenmerken. Elementen die deel uitmaken van een groep zijn zo homogeen mogelijk ten aanzien van de beschrijvende kenmerken terwijl tussen de groepen een maximale heterogeniteit naar deze kenmerken bestaat. Clusteranalyse is o.m. een geschikte techniek voor het afbakenen van doelgroepen en het opstellen van klantprofielen.

Ook *factoranalyse* is een exploratief-beschrijvende techniek om de dimensionaliteit in de data te analyseren. Het doel van factoranalyse is een hoeveelheid variabelen samen te vatten in een kleiner aantal onderliggende dimensies, die alle een lineaire combinatie zijn van de oorspronkelijke variabelen. Factoranalyse biedt o.m. de mogelijkheid om onderling sterk samenhangende variabelen te reduceren, hetgeen de interpretatie van de onderzoeksbevindingen ten goede komt.

De doelstelling van *multidimensionele schaalanalyse* is het 'plaatsen' van objecten (bv. merken) en/of subjecten (bv. consumenten) in een meerdimensionele ruimte. Uitgangspunt hierbij is de gedachte dat respondenten over het algemeen objecten beoordelen. Het is aan de onderzoeker om uit te vinden welke kenmerken de respondenten in gedachten hebben bij het vergelijken van merken of het geven van een voorkeur voor het ene merk boven het andere. Bij MDS beoordelen de respondenten de objecten niet op vooraf gespecificeerde variabelen (kenmerken), maar worden de objecten alleen met elkaar vergeleken. Het voordeel hiervan is dat de onderzoeker niet een 'uitputtende' lijst van relevante kenmerken hoeft te definiëren op grond waarvan de objecten moeten worden beoordeeld. In situaties waarin de onderzoeker onvoldoende duidelijk weet op grond waarvan zijn merk wordt vergeleken met andere merken, is het aan te bevelen dit door middel van een MDS-studie te achterhalen. In zijn algemeenheid kunnen we stellen dat MDS wordt gebruikt om de positie van een object (bv. een merk) t.o.v. concurrerende objecten te bepalen (produktpositionering). MDS wordt eveneens gebruikt om via de voorkeuren van respondenten voor objecten zgn. 'gaten in de markt' te ontdekken (marktsegmentatie). Bij een MDS-studie dient een onderscheid gemaakt te worden tussen het positioneren van objecten op basis van 'similarity'-data (gelijkenissen) en het bepalen van de zgn. 'ideaalpunten' van respondenten op basis van 'preference'-data (voorkeuren).

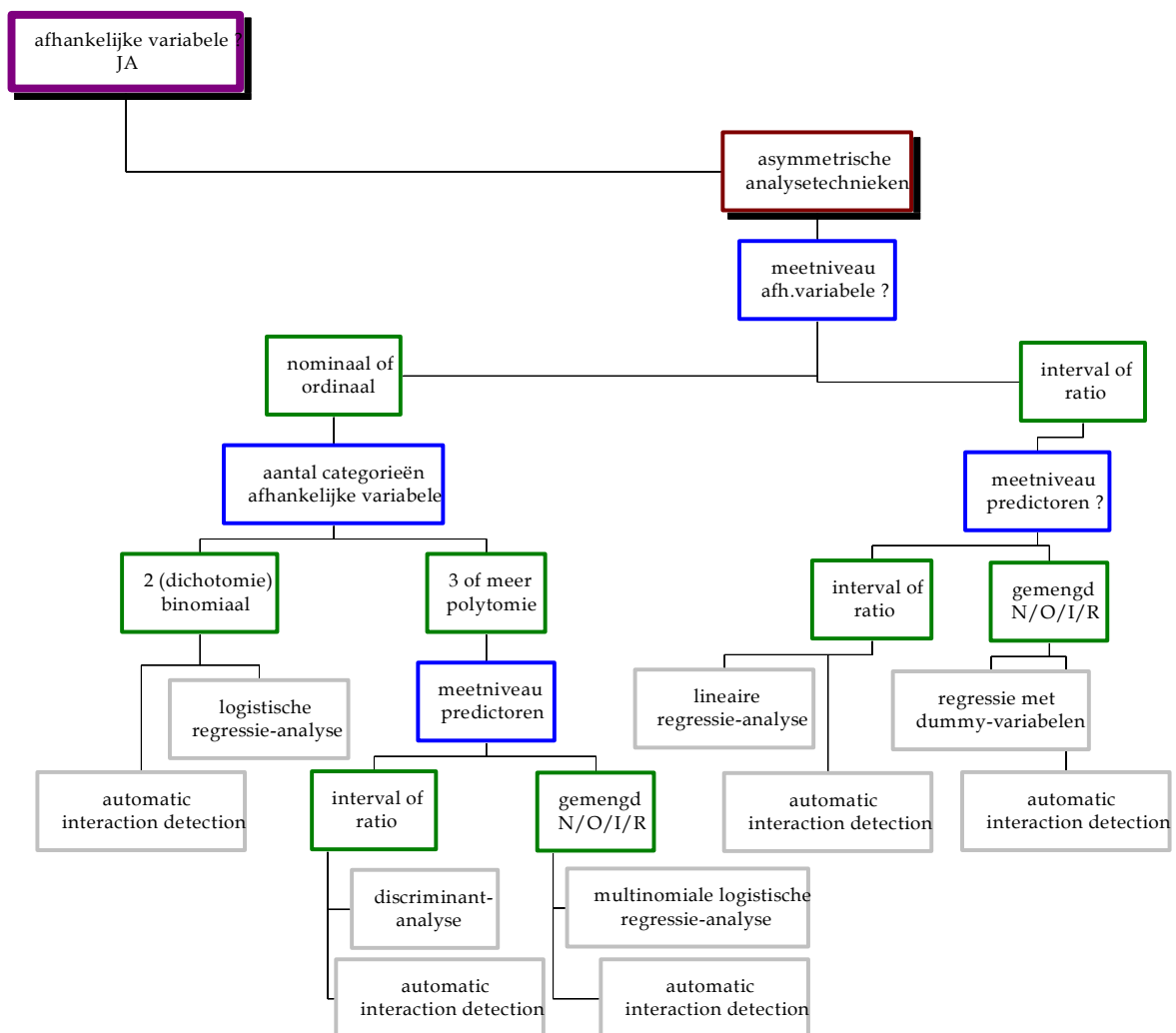
Het positioneren van objecten en/of subjecten in een meerdimensionele ruimte kan eveneens gebeuren door gebruikmaking van *correspondentie-analyse*. Correspondentie-analyse is een geheel van technieken die voorzien in de niet-lineaire analyse van categoriale variabelen.



Figuur 11 : Interdependentietechnieken

#### 2.3.2.2.2. Dependente technieken

*Dependentietechnieken* hebben tot doel de invloed van één of meerdere onafhankelijke variabelen (predictoren) na te gaan op een afhankelijke variabele (*criteriumvariabele*). Met predictieve modellen wordt beoogd ofwel het behoren tot een klasse (*classificatie*) ofwel een waarde te voorspellen (*regressie*). In het geval van classificatie is de klasse een categoriale variabele die uit twee of meerdere elkaar uitsluitende categorieën bestaat. Scoringsmodellen die voorzien in de predictie van de respons op een mailing, voorspellen het behoren van prospecten of klanten tot de klasse "ja" of "nee". Op analoge wijze kunnen klanten in kaart gebracht worden die de grootste kans hebben om binnen een bepaalde termijn te vertrekken. Door de klanten die binnen dit profiel passen extra aandacht te geven kan het verloop teruggedrongen worden. Regressie wordt toegepast in het geval de te verklaren waarde (de afhankelijke variabele) een veelheid aan (numerieke) waarden kan aannemen (continue variabele). Een voorbeeld hiervan is het voorspellen van de beursnotering van aandelen.



Figuur 12 : Dependentietechnieken

*Regressie-analyse* is een predictieve techniek die gebruik maakt van het optimaliseringsprincipe dat bekend staat als de methode van de kleinste kwadraten. De bedoeling hiervan is de waarde van een continue variabele te voorspellen aan de hand van een lineaire combinatie van onafhankelijke variabelen. Het verschil tussen de verwachte en geobserveerde waarden van de afhankelijke variabele geldt als criterium voor de beoordeling van het regressiemodel. Een bezwaar dat aan de toepassing van regressie-analyse kleeft, is dat de gegevens die ermee gemodelleerd worden vaak niet voldoen aan de lineariteitsassumptie van de techniek. Met regressie-analyse kunnen in principe enkel lineaire verbanden worden gemodelleerd. Interacties tussen variabelen worden over het hoofd gezien, tenzij deze als afzonderlijke variabelen aan de regressie-vergelijking worden toegevoegd. Bovendien blijken heel wat gegevens in de marketingpraktijk eerder van categoriale dan van parametrische aard te zijn. Het oplossen van de problemen die hiermee gepaard gaan, vereist vooral bij toepassing van regressie-analyse heel wat statistische expertise. Een ander zwak punt is de schaalbaarheid van regressie-analyse. Bij een toenemend aantal variabelen blijkt de kwaliteit van regressiemodellen achteruit te gaan. De oorzaak hiervan is dat regressie-analyse

gevoelig is voor multicollineariteit in de gegevensset (onafhankelijke variabelen die eenzelfde verband hebben met de voorspelde variabele).

De bezwaren die gepaard gaan bij het toepassen van regressie-analyse en de feitelijke vaststelling dat heel wat variabelen in de marketingpraktijk eerder van categoriale dan van parametrische aard te zijn, verklaren het toegenomen gebruik van technieken die voorzien in de niet-lineaire analyse van variabelen. Om die reden blijven we in hetgeen volgt stilstaan bij een tweetal technieken waarvoor een groeiende belangstelling bestaat, nl. beslissingsbomen en logistische regressie-analyse.

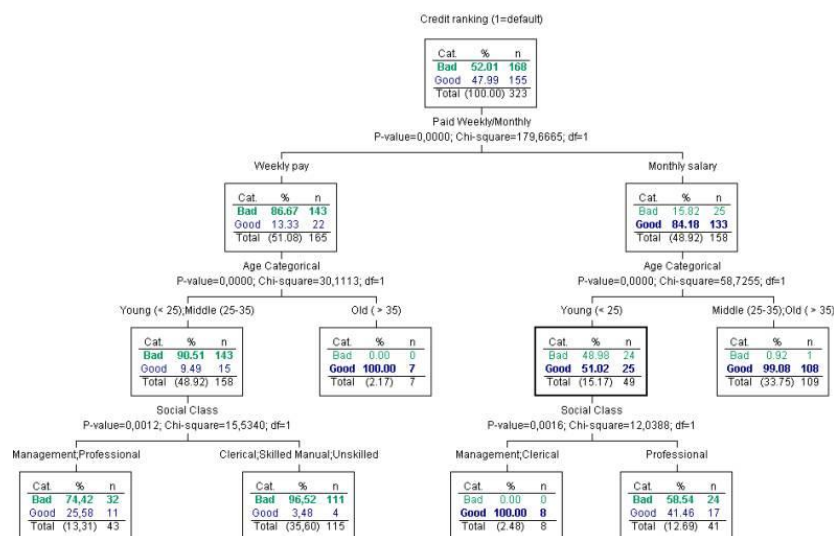
#### 2.3.2.2.1. Beslissingsbomen

Vrijwel alle statistische technieken voor het vaststellen van verbanden tussen meerdere variabelen zijn gebaseerd op tamelijk stringente vooronderstellingen. De invloed van het niet voldoen van de gebruikte data aan deze beperkingen en de fout die hierdoor optreedt, komen in de uiteindelijke resultaten niet (of onvoldoende) tot uitdrukking en zijn derhalve statistisch niet te testen. Een techniek die tegemoet komt aan deze bezwaren is *automatic interaction detection*. AID stelt minimale eisen aan het meetniveau van de variabelen : de afhankelijke variabele mag zowel interval- als dichotoom geschaald zijn en aan de onafhankelijke variabelen worden geen schaalessen gesteld. Bij AID worden geen vooronderstellingen gemaakt omtrent de functionele vorm van de relatie tussen afhankelijke en onafhankelijke variabelen. De grootste kracht van AID schuilt in de mogelijkheid tot het opsporen van interacties in de verzameling van verklarende variabelen, dus op het gebied waar andere technieken veelal het ontbreken hiervan veronderstellen. Vooral indien het om een complex samenstel van variabelen gaat, kunnen interactie-effecten in de uitkomsten van een AID-analyse via de boomstructuur onderkend worden. Eveneens kan uit een AID-analyse blijken welke variabelen voldoende verklaren van de variantie in de afhankelijke variabele voor een deel van de observaties, terwijl voor andere observaties andere of meerdere variabelen van belang zijn. Met AID krijgt de onderzoeker derhalve inzicht in de samenhang tussen een (groot) aantal variabelen, zonder dat er stringente vooronderstellingen behoeven te worden gemaakt. Vandaar dat AID moet gezien worden als een techniek die in een vroeg analysestadium gebruikt wordt. Een vereiste voor de toepassing van AID is wél dat de steekproef voldoende groot moet zijn om een voldoende aantal splitsingen te kunnen krijgen die elk op zich uit voldoende waarnemingen bestaan.

Een beslissingsboom is te vergelijken met een scoringsmodel waarbij de inputvariabelen (predictoren) gebruikt worden om waarnemingen te splitsen in verschillende groepen die discrimineren naar de te verklaren variabele (doelvariabele). Het gegevensbestand wordt opgedeeld in elkaar uitsluitende segmenten op basis van een doelvariabele. Deze segmentatie komt tot stand door na te gaan welke variabele en welke samenvoeging van waarden binnen deze variabele het hoogst mogelijk onderscheidend vermogen heeft t.a.v. de doelvariabele. Dit proces wordt iteratief toegepast totdat er geen onderverdeling meer te maken is die voldoende onderscheidend vermogen bezit.

Zoals vermeld, is een voordeel van beslissingsbomen dat de meest belangrijke variabelen en de combinaties (interacties) ertussen voor de voorspelling van de doelvariabele gedetecteerd worden. Op basis van beslissingsbomen worden regels opgesteld die aangeven aan welke voorwaarden een waarneming dient te voldoen om in een bepaald segment terecht te komen. Het segment geeft de voorspellende waarde aan voor de waarnemingen die daarbinnen vallen (te bedenken valt evenwel dat met boomstructuren enkel groepsvoorspellingen en geen individuele prognoses kunnen gemaakt worden).

In figuur 13 is een voorbeeld opgenomen van een beslissingsboom. Hierbij zijn voor de voorspelling van het kredietrisico als afhankelijke variabele een drietal predictoren (leeftijd, sociale klasse en betaalwijze) gebruikt.



Figuur 13 : Belissingsboom

### 2.3.2.2.2. Logistische regressie-analyse

Een tegenwoordig veel gebruikte classificatiemethode is *logistische regressie*. Deze techniek wordt toegepast voor het voorspellen van categoriale variabelen. Behalve de toepassing ervan bij responsanalyse, wordt de techniek van logistische regressie door bank- en verzekeringsinstellingen vaak aangewend voor het opstellen van modellen voor kredietacceptatie ("credit scoring"). De doelstelling hierbij is het optimaliseren van het percentage geaccepteerde aanvragen zodat het maximaal toegestane infectiepercentage niet wordt overschreden. Aan de hand van klantgegevens (geslacht, leeftijd, beroep, e.d.), het product (doorlopend krediet, persoonlijke lening) en gegevens van in het verleden verstrekte kredieten waarbij voor ieder krediet ook de afloop is vastgelegd, stelt het algoritme van logistische regressie de gebruiker in staat om profielen te ontdekken met een sterk verlaagde of verhoogde kans op wanbetaling. De kansen op wanbetaling worden afgeleid uit het logistische model dat de vorm aanneemt van een regressievergelijking. Door de waarden van de variabelen die in het model opgenomen zijn, te wegen overeenkomstig de parameters van de regressievergelijking worden logitscores bekomen. De logaritmische transformatie van deze laatste laat toe logitscores te vertalen in waarschijnlijkheden. De resultaten van een logistische regressie-analyse kunnen teruggeschreven worden naar de prospecten- of klantendatabase om deze te verrijken, waardoor voor iedere prospect of klant op basis van zijn/haar karakteristieken een kans op wanbetaling wordt berekend.

Logistische regressie-analyse is geschikt voor een variabele die dichotoom van aard is : er zijn twee categorieën. Indien een afhankelijke variabele meer dan 2 categorieën telt, wordt gebruik gemaakt van multinomiale logistische regressie-analyse. We zullen het logistische model illustreren aan de hand van de voorspelling van het al of niet ingaan van abonnees van een tijdschrift op een aanbod tot hernieuwing van hun leesperiode. Aan de

hand van logistische regressie-analyse is het o.m. mogelijk om het effect van twee condities in kaart te brengen, nl. het gebruik maken van een incentive enerzijds en het al of niet telefonisch contacteren van de abonnees tijdens de periode voorafgaand aan de hernieuwing van het abonnement anderzijds. Door het effect van deze twee variabelen, aangewend bij een eerdere hernieuwingactie, te kwantificeren, kan de verwachte respons bij herhaling voorspeld worden. In tabel 24 worden de responsfrequenties voor elk van de twee onderzochte condities weergegeven.

Tabel 24 : Responsfrequenties

Count

				BETALING		Total
				neen	ja	
<b>INCENTIVE</b>	<b>geen</b>	<b>REMINDER</b>	geen call	81	19	100
			call	734	266	1000
		<b>Total</b>		815	285	1100
	<b>incentive</b>	<b>REMINDER</b>	geen call	237	263	500
			call	214	286	500
		<b>Total</b>		451	549	1000
	<b>korting</b>	<b>REMINDER</b>	geen call	199	801	1000
			call	18	82	100
		<b>Total</b>		217	883	1100

#### 2.3.2.2.2.1. Modelfit

Bij het interpreteren van de resultaten van een logistische regressie-analyse dient eerst bekeken te worden of het model dat geschat wordt goed bij de data past. De meest belangrijke toets hiervoor is de Chi<sup>2</sup>-toets. De Chi<sup>2</sup>-toets vergelijkt de aannemelijkheidsratio van het geschatte model (hier gelijk aan 3723,445) met de aannemelijkheidsratio van het model met alleen maar een constante (hier gelijk aan 4419,015). Het verschil tussen deze twee aannemelijkheidsratio's is de *Chi-square* (hier gelijk aan 695,570). Een Chi<sup>2</sup> van 695,570 is significant hetgeen betekent dat het model met de variabelen 'incentive' en 'reminder' beter bij de data past dan een model zonder deze variabelen.

Logistische regressie-analyse geeft geen proportie verklaarde variantie (R<sup>2</sup>), omdat bij een dichotome afhankelijke variabele eigenlijk niet kan gesproken worden van variantie. Wel bestaan er pseudo R<sup>2</sup>-maten die vergelijkbaar zijn met de R<sup>2</sup> die bekomen wordt bij toepassing van lineaire regressie-analyse. Een dergelijke pseudo R<sup>2</sup>-maat is eenvoudig te berekenen door de aannemelijkheidsratio van het model met onafhankelijke variabelen af te trekken van de aannemelijkheidsratio van het model zonder onafhankelijke variabelen en het resultaat hiervan te delen door deze laatste aannemelijkheidsratio. In ons voorbeeld bekomen we een R<sup>2</sup>-waarde van 0,157. Opgemerkt moet worden dat pseudo R<sup>2</sup>-maten doorgaans een lage waarde aannemen.

Een andere manier om de fit van het model te bepalen, is door gebruik te maken van de *Hosmer en Lesmeshow Goodness-of-fit Test*. Hiermee wordt nagegaan of er significante verschillen bestaan tussen de kansen zoals die in de data waargenomen worden en de kansen zoals die door het model voorspeld worden. In ons voorbeeld zijn deze verschillen niet significant (Chi<sup>2</sup> = 0,736 en sig. = 0,865) zodat we kunnen concluderen dat het model goed bij de data past.

### 2.3.2.2.2.2. Interpretatie van de effecten

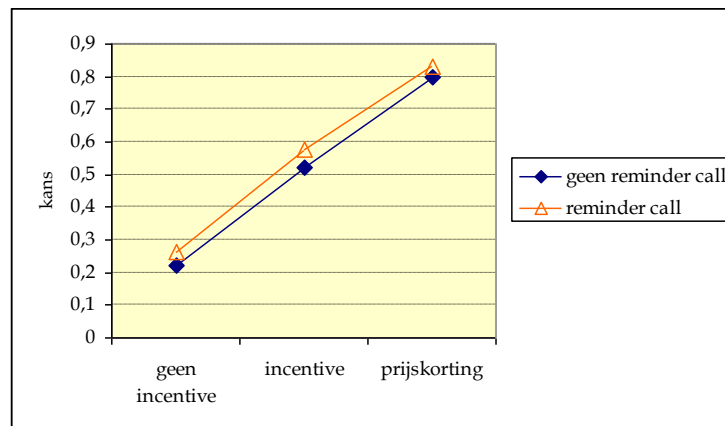
Nu we de fit van het model bepaald hebben, kunnen we overgaan naar de effecten van de onafhankelijke variabelen. In ons voorbeeld blijken de variabelen 'incentive' en 'reminder' beide een significante invloed te hebben op de kans om al of niet in te gaan op een hernieuwingsaanbod.

Zoals kan afgelezen worden uit tabel 25 worden de effecten van de beide variabelen weergegeven in de kolom met de logits (B) : de natuurlijke logaritme (ln) van de kansverhouding om al of niet in te gaan op het hernieuwingsaanbod. Hoe groter het getal, hoe groter het effect. Net als bij lineaire regressie-analyse betekent een positief getal een positief effect en een negatief getal een negatief effect. Het effect van de variabelen 'reminder' is bijvoorbeeld 0,219. Omdat we in het geval van logistische regressie-analyse de effecten van variabelen interpreteren in termen van kansverhoudingen (odds), kijken we naar de kolom met Exp(B). Bij een positief effect is de waarde van Exp(B) groter dan 1 ; bij een negatief effect ligt de waarde tussen 0 en 1. Met betrekking tot het effect van de variabele 'reminder' lezen we een waarde van 1,244 af ( $e^{0,219} = 1,244$ ). De interpretatie die aan deze *odds ratio* dient gegeven te worden is dat de kansen op hernieuwing van het abonnement gemiddeld met 24,4 % stijgen in het geval van een reminder call dan wanneer dit laatste niet het geval is.

Tabel 25 : Resultaten logistische regressie-analyse

	B	S.E.	Wald	df	Sig.	Exp(B)
INCENT			381.691	2	.000	
INCENT(1)	-2.635	.135	380.543	1	.000	.072
INCENT(2)	-1.297	.108	143.792	1	.000	.273
REMINDER	.219	.105	4.320	1	.038	1.244
Constant	1.385	.076	329.747	1	.000	

Met behulp van de regressievergelijking kunnen voor alle combinaties tussen de kenmerken de kansen op hernieuwing van het abonnement berekend worden. Het resultaat hiervan is weergegeven in figuur 14. Hieruit kan onmiddellijk opgemaakt worden dat het al of niet hernieuwen van een abonnement beduidend sterker beïnvloed wordt door het al of niet aanbieden van een incentive (in de vorm van een geschenk, resp. prijskorting) dan dit het geval is voor een 'reminder' in de vorm van een telefonische contactname. Van deze laatste variabele gaat een veel geringere invloed uit op het al of niet betalen van een hernieuwingsaanbod. Het effect van deze laatste variabele is anderzijds grotendeels constant binnen categorieën van de variabele 'incentive'.



*Figuur 14 : Logistische regressie van de kenmerken 'incentive' en 'reminder'*



## UITDAGINGEN VOOR HET NIEUWE MILLENNIUM

Het marktonderzoek heeft de afgelopen decennia een belangrijke evolutie doorgemaakt. Vanaf de jaren zeventig werd gaandeweg een scala aan *multivariate technieken* toegankelijk gemaakt voor de marktonderzoeker. Alleen al het kunnen toepassen van deze technieken gaf een marktonderzoeksbureau een concurrentievoordeel. De toegankelijkheid van statistische technieken werd nog ruimer met de opkomst van de PC in de jaren tachtig. Maar het beschikken over hard- en software bleek vanaf deze periode geen voldoende garantie meer voor concurrentievoordeel. Waar het vanaf de jaren tachtig meer en meer om gaat, is *het kunnen vertalen van uitkomsten van onderzoek in strategische adviezen*. Deze strategisch gerichte informatiefase geeft een ontwikkeling aan van dataverzameling naar informatiemanagement. Hiermee veranderde ook de taak van de marktonderzoeker. Meer en meer werd deze een *informatiemanager* die als taak heeft informatie uit verschillende bronnen te combineren (marktonderzoekgegevens, eigen klanten databases, aankoopgegevens, externe bronnen, ...). Met de explosie aan gegevens werd de rol om gegevens te structureren, te analyseren en te vertalen naar het bedrijfsbeleid steeds belangrijker.

De betekenis van het marktonderzoek op het vlak van informatiemanagement krijgt de afgelopen jaren evenwel een andere invulling. Onverminderd het belang van het strategisch denken, kan de vraag gesteld worden of er tegenwoordig niet meer om gaat gedrag optimaal te registreren en op basis van individuele informatie optimaal te communiceren met de consument. Met de ontwikkeling van databases die op het individu toegesneden gegevens bevatten is het mogelijk om voor elke bestaande of potentiële klant een marketing mix te specificeren waarvan men verwacht dat deze het best past bij de individuele klant. Het *nieuwe marketing paradigma* stelt dat organisaties relaties aangaan met consumenten, die een veel actievere rol spelen dan voorheen. Deze ontwikkeling heeft belangrijke consequenties voor marketingmodellen. Het gaat erom klanten aan te trekken en (vooral) te behouden. In beide gevallen dienen de preferenties van consumenten in kaart gebracht te worden. Modellen die toelaten consumentengedrag te voorspellen worden onontbeerlijk. Deze ontwikkeling verklaart mede de inzet van *datamining* technieken voor zowel het opmaken van klantprofielen als het ontwikkelen van scoringmodellen. Het toegenomen gebruik van technieken die geïnspireerd zijn op biologische processen (neurale netwerken) vindt eveneens zijn oorsprong in de mogelijkheden ervan gegevensverzamelingen met onderling vaak interacterende variabelen te modelleren.

Veranderingen op het vlak van de onderzoeksmethoden en -technieken zullen zich blijvend aandienen. *Clickstream data* bijvoorbeeld stellen nieuwe eisen. Hiermee wordt een belangrijke uitdaging voor het nieuwe millennium aangesneden : het *Internet*. Naast de postenquête ontwikkelen zich momenteel andere vormen van *self-administered questionnaires*, zoals Internet- en email-surveys. Veel communicatie verschuift van mondeling/telefonisch naar elektronisch. Een voorbeeld van deze omslag is de overgang van "push" naar "pull". Niet langer is het de producent die (ongevraagde) informatie richting ontvanger stuurt ("push"), maar het is de ontvanger die de gewenste informatie naar zich toetrekt ("pull").

De ontwikkeling van een klantgerichte marketingstrategie (*customer relationship management*) is afhankelijk van de invulling van zowel een interactieve als informatieve dimensie. De omslag van een aanbod- naar een vraaggeoriënteerde benadering impliceert in de eerste plaats dat bedrijven in staat moeten zijn om interactieve media aan te wenden in de communicatie met klanten en prospecten. Het principe van klantgerichtheid houdt in dat de onderneming elke klant individueel benadert en aan de individuele wensen van de consument tegemoet komt met een op maat gesneden aanbod. Dit is alleen mogelijk als de

onderneming voldoende inzicht heeft in de voorkeuren van klanten en prospecten. Waar het in toenemende mate om gaat, is kennis te verwerven in de zeer dynamische omgeving van de organisatie en deze om te buigen in een competitief voordeel. De mate waarin het marktonderzoek weet in te spelen op deze trend zal bepalend zijn voor het succes ervan.

### Geraadpleegde literatuur

- Blalock, H.M. (1972), *Social statistics*, London, McGraw-Hill.
- Cramer, J.S. & Hemlerijk, J. (1998), *Statistiek eenvoudig*, Amsterdam, Uitgeverij Nieuwezijds.
- Green, P.E., Tull, D.S. & Albaum, G. (1988), *Research for marketing decisions*, Englewood Cliffs, New Jersey, Prentice-Hall.
- Hirschi, T. & Selvin, H.C. (1967), *Principles of survey analysis*, New York, The free Press.
- Molenaar, C. (1993), *Het einde van de massamarketing*, Amsterdam, De Management Bibliotheek.
- Shavelson, R.J. (1981), *Statistical reasoning for the behavioral sciences*, Boston, Allyn and Bacon.
- Swanborn, P.G. (1979), *Aspecten van sociologisch onderzoek*, Meppel, Boom.
- Van Noort, W.C.J. (1989), *Statistiek voor marketingbeslissingen*, Leiden, Stenfert Kroese.
- Zwart, P.S. (1989), *Methoden van marktonderzoek*, Leiden, Stenfert Kroese.

## Inhoudsopgave

<b>Marketing en Marktonderzoek in een verander(en)de wereld</b>	<b>2</b>
1. Maatschappelijke transformatie .....	2
2. Marketing .....	3
3. Marktonderzoek .....	7
<b>1. Ontwikkelingen in marktonderzoek .....</b>	<b>9</b>
1.1. Onderzoek gericht op het ontwikkelen van methoden en technieken.....	9
1.1.1. Het verbeteren van de methoden van marktonderzoek .....	10
1.1.2. Het ontwikkelen (verbeteren) van analysetechnieken.....	10
1.1.3. Het ontwikkelen van technieken ten behoeve van strategische planning.....	11
1.2. Onderzoek gericht op het genereren van generaliseerbare kennis.....	11
<b>2. HET MARKTONDERZOEKPROCES .....</b>	<b>12</b>
2.1. De probleemstelling .....	15
2.1.1. Algemene situering .....	15
2.1.2. De relatie tussen probleemstelling en theorie .....	16
2.1.2.1. Soorten probleemstellingen.....	16
2.1.2.2. Typen onderzoek.....	17
2.1.2.2.1. Theorieën .....	17
2.1.2.2.2. Vormen van theorie-georiënteerd onderzoek .....	19
2.1.2.2.2.1. Onderzoek waarbij het accent ligt op descriptie .....	19
2.1.2.2.2.2. Onderzoek waarbij het accent ligt op exploratie .....	20
2.1.2.2.2.3. Onderzoek waarbij het accent ligt op toetsing .....	20
2.2. Het bepalen van de onderzoeksopzet.....	21
2.2.1. Het vaststellen van de steekproef.....	22
2.2.1.1. Toevalssteekproeven .....	25
2.2.1.2. Niet-toevalssteekproeven .....	27
2.2.1.3. De steekproefomvang .....	27
2.2.1.3.1. De steekproefomvang bij de schatting van een gemiddelde .....	27
2.2.1.3.2. De steekproefomvang bij een schatting van een fractie .....	28
2.2.1.4. Eisen aan steekproefonderzoek .....	30
2.2.1.4.1. Representativiteit .....	30
2.2.1.4.2. Betrouwbaarheid en nauwkeurigheid.....	31
2.2.1.4.2.1. Het schattingsvraagstuk .....	31
2.2.1.4.2.2. Hypothesentoetsing .....	34
2.2.1.4.2.2.1. Parametrische toetsen.....	35
2.2.1.4.2.2.1.1. Het steekproefgemiddelde toetsen aan een andere waarde .....	35
2.2.1.4.2.2.1.2. Gemiddelden toetsen bij twee groepen .....	37
2.2.1.4.2.2.2. Niet-parametrische toetsen.....	39
2.2.1.4.2.3. Methoden ter vaststelling van de betrouwbaarheid .....	40
2.2.1.4.2.3.1. Test-retest betrouwbaarheid .....	40
2.2.1.4.2.3.2. Instrument-betrouwbaarheid .....	41
2.2.1.4.2.3.3. Onderzoeker-betrouwbaarheid.....	41
2.2.1.4.3. Validiteit .....	42
2.2.1.4.3.1. Begripsomschrijving .....	42
2.2.1.4.3.2. Checks op validiteit.....	42
2.2.1.4.3.3. Methoden ter vaststelling van de geldigheid.....	44

2.2.1.4.3.3.1. "Face"-validatie.....	44
2.2.1.4.3.3.2. Criterium-validiteit.....	45
2.2.1.4.3.3.3. Begripsvalidatie .....	45
2.2.2. De keuze van de methode tot gegevensverzameling.....	46
2.2.2.1. Kwantitatieve en kwalitatieve onderzoeksmethoden .....	46
2.2.2.1.1. Diepte-interview .....	52
2.2.2.1.2. Groepsdiscussie ("focus group").....	52
2.2.2.1.3. Creatieve sessie ("brainstorming").....	54
2.2.2.1.4. Delphi-methode.....	54
2.2.2.1.5. Gedragsobservatie.....	55
2.2.2.1.6. Instrument-observatie.....	55
2.2.2.1.7. Kelly-Grid.....	55
2.2.2.1.8. Survey met open vragen.....	55
2.2.2.1.9. Nominale groepstechniek.....	56
2.2.2.2. Communicatie .....	56
2.2.2.3. Observatie .....	56
2.2.2.3.1. Micro-observatie .....	57
2.2.2.3.2. Het waarnemen in een experiment.....	57
2.2.2.3.2.1. Criteria ter beoordeling van een onderzoeksontwerp .....	58
2.2.2.3.2.1.1. Causaliteit .....	58
2.2.2.3.2.1.2. Generaliseerbaarheid.....	64
2.2.2.3.2.1.3. Theoretische adequaatheid en praktische uitvoerbaarheid .....	66
2.2.2.3.2.2. Het waarnemen in een ongecontroleerd experiment .....	66
2.2.2.3.2.2.1. Tijdserie-opzetten.....	67
2.2.2.3.2.2.1.1. "After only-design" .....	67
2.2.2.3.2.2.1.2. "Before and after-design" .....	68
2.2.2.3.2.2.1.3. "Before and after with control group-design" .....	69
2.2.2.3.2.2.2. Dwarsdoorsnede-opzetten .....	70
2.2.2.3.2.2.3. Combinatie van tijdserie-opzet en dwarsdoorsnede-opzet .....	70
2.2.2.3.2.3. Het waarnemen in een gecontroleerd experiment .....	70
2.2.2.3.2.3.1. Volstrekt aselechte opzet.....	73
2.2.2.3.2.3.2. Aselechte blokkenopzet .....	73
2.2.2.3.2.3.3. Latijns vierkant .....	74
2.2.2.3.2.3.4. Factoropzet.....	75
<b>2.3. Het bepalen van de analysemethode.....</b>	<b>77</b>
2.3.1. Voorbewerking van gegevens.....	77
2.3.1.1. Gegevenscontrole.....	77
2.3.1.2. Codering.....	78
2.3.2. Gegevensanalyse.....	78
2.3.2.1. Modelspecificatie .....	78
2.3.2.1.1. Het vaststellen van de eenheden .....	78
2.3.2.1.2. Het vaststellen van de variabelen .....	79
2.3.2.1.2.1. Variabelen naar de eenheden waarop ze betrekking hebben .....	79
2.3.2.1.2.2. Variabelen naar de waarden die ze aannemen.....	81
2.3.2.1.2.3. Variabelen naar het meetniveau .....	82
2.3.2.1.3. Relaties.....	82
2.3.2.1.3.1. De afbakening van het causaliteitsbegrip.....	83
2.3.2.1.3.2. Criteria ter toetsing van causale hypothesen .....	85
2.3.2.1.3.2.1. Geldige criteria.....	85
2.3.2.1.3.2.2. Ongeldige criteria.....	85
2.3.2.2. Symmetrische en asymmetrische analysetechnieken .....	88
2.3.2.2.1. Interdependentietechnieken .....	89
2.3.2.2.2. Dependente technieken .....	90
2.3.2.2.2.1. Beslissingsbomen.....	92
2.3.2.2.2.2. Logistische regressie-analyse.....	93
2.3.2.2.2.2.1. Modelfit.....	94
2.3.2.2.2.2.2. Interpretatie van de effecten.....	95

<b>UITDAGINGEN VOOR HET NIEUWE MILLENNIUM .....</b>	<b>97</b>
Geraadpleegde literatuur .....	99
Inhoudsopgave .....	100